

	ΔΙΑΤΜΗΜΑΤΙΚΟ ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ στα ΠΟΛΥΠΛΟΚΑ ΣΥΣΤΗΜΑΤΑ και ΔΙΚΤΥΑ ΤΜΗΜΑ ΜΑΘΗΜΑΤΙΚΩΝ ΤΜΗΜΑ ΒΙΟΛΟΓΙΑΣ ΤΜΗΜΑ ΓΕΩΛΟΓΙΑΣ ΤΜΗΜΑ ΟΙΚΟΝΟΜΙΚΩΝ ΕΠΙΣΤΗΜΩΝ ΑΡΙΣΤΟΤΕΛΕΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΟΝΙΚΗΣ	
---	---	---

ΜΕΤΑΠΤΥΧΙΑΚΗ ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Τίτλος Εργασίας

Ανάπτυξη συστήματος αυτόματης επισήμανσης εικόνων χρηστών διαδικτυακής κοινότητας με σκοπό την βελτίωση του συστήματος συστάσεων σχετικών προϊόντων

Αλέξανδρος Μ. Σούσος

ΕΠΙΒΛΕΠΩΝ: Ιωάννης Αντωνίου, Καθηγητής Α.Π.Θ.

ΣΥΝΕΠΙΒΛΕΠΩΝ: Χαράλαμπος Μπράτσας, ΕΔΙΠ Α.Π.Θ.

Θεσσαλονίκη, Δεκέμβριος 2019

	ΔΙΑΤΜΗΜΑΤΙΚΟ ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ στα ΠΟΛΥΠΛΟΚΑ ΣΥΣΤΗΜΑΤΑ και ΔΙΚΤΥΑ ΤΜΗΜΑ ΜΑΘΗΜΑΤΙΚΩΝ ΤΜΗΜΑ ΒΙΟΛΟΓΙΑΣ ΤΜΗΜΑ ΓΕΩΛΟΓΙΑΣ ΤΜΗΜΑ ΟΙΚΟΝΟΜΙΚΩΝ ΕΠΙΣΤΗΜΩΝ ΑΡΙΣΤΟΤΕΛΕΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΟΝΙΚΗΣ	
---	---	---

ΜΕΤΑΠΤΥΧΙΑΚΗ ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Τίτλος Εργασίας

Ανάπτυξη συστήματος αυτόματης επισήμανσης εικόνων χρηστών διαδικτυακής κοινότητας με σκοπό την βελτίωση του συστήματος συστάσεων σχετικών προϊόντων

Αλέξανδρος Μ. Σούσος

ΕΠΙΒΛΕΠΩΝ: Ιωάννης Αντωνίου, Καθηγητής Α.Π.Θ.

ΣΥΝΕΠΙΒΛΕΠΩΝ: Χαράλαμπος Μπράτσας, ΕΔΙΠ Α.Π.Θ.

Εγκρίθηκε από την Τριμελή Εξεταστική Επιτροπή την 20η Δεκεμβρίου 2019.

.....		
Ι. Αντωνίου Καθηγητής Α.Π.Θ.	Χ. Μπράτσας ΕΔΙΠ Α.Π.Θ.	Ν. Φαρμάκης Α. Καθηγητής Α.Π.Θ.

Θεσσαλονίκη, Δεκέμβριος 2019

.....
Αλέξανδρος Μ. Σούσος

Διπλωματούχος Ηλεκτρολόγος Μηχανικός και Μηχανικός Α.Π.Θ.

Copyright © Αλέξανδρος Μ. Σούσος, 2019

Με επιφύλαξη παντός δικαιώματος. All rights reserved.

Απαγορεύεται η αντιγραφή, αποθήκευση και διανομή της παρούσας εργασίας, εξ ολοκλήρου ή τμήματος αυτής, για εμπορικό σκοπό. Επιτρέπεται η ανατύπωση, αποθήκευση και διανομή για σκοπό μη κερδοσκοπικό, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν μήνυμα. Ερωτήματα που αφορούν τη χρήση της εργασίας για κερδοσκοπικό σκοπό πρέπει να απευθύνονται προς τον συγγραφέα.

Οι απόψεις και τα συμπεράσματα που περιέχονται σε αυτό το έγγραφο εκφράζουν τον συγγραφέα και δεν πρέπει να ερμηνευθεί ότι εκφράζουν τις επίσημες θέσεις του Α.Π.Θ.

	INTER-FACULTY MASTER PROGRAM on COMPLEX SYSTEMS and NETWORKS SCHOOL of MATHEMATICS SCHOOL of BIOLOGY SCHOOL of GEOLOGY SCHOOL of ECONOMICS ARISTOTLE UNIVERSITY of THESSALONIKI	
---	--	---

MASTER THESIS

Title

Development of an object detection system for automatic labeling of online-community users' images in order to improve the efficiency of product recommendations.

Alexandros M. Sousos

SUPERVISOR: Antoniou Ioannis, Professor, A.U.Th.

CO-SUPERVISOR: Bratsas Charalampos, Laboratory Teaching Staff, A.U.Th.

Θεσσαλονίκη, Δεκέμβριος 2019



Ανάπτυξη συστήματος αυτόματης επισήμανσης εικόνων
χρηστών διαδικτυακής κοινότητας με σκοπό τη βελτίωση του
συστήματος συστάσεων σχετικών προϊόντων

Αλέξανδρος Μ. Σούσος

Θεσσαλονίκη, Δεκέμβριος 2019





Περίληψη

Αντικείμενο της εργασίας είναι η μελέτη και η εφαρμογή αλγορίθμων για την παρακολούθηση, πρόβλεψη και ενίσχυση του συστήματος συστάσεων προϊόντων που προτείνονται στους χρήστες μίας διαδικτυακής πλατφόρμας κοινωνικής δικτύωσης και διαμοιρασμού φωτογραφιών. Για τον σκοπό αυτόν, προτείνεται μία ολοκληρωμένη πρόταση ενίσχυσης του συστήματος συστάσεων μέσω της ανίχνευσης αντικειμένων στις εικόνες των χρηστών. Η πρόταση ακολουθεί τεχνικές ανάπτυξης λογισμικού βάσει δεδομένων περιλαμβάνοντας μεταξύ άλλων τη δημιουργία κεντρικής αποθήκης δεδομένων, λήψη αποφάσεων βάσει δεδομένων, δοκιμές α/β και αυτοματοποιημένες αναφορές. Συγκεκριμένα, εξετάζονται αλγόριθμοι ανάλυσης και πρόβλεψης χρονοσειρών και χρησιμοποιείται ο αλγόριθμος Prophet για την παρακολούθηση και εκτίμηση των μελλοντικών τιμών του συνόλου των επιλεχθέντων συστάσεων. Στη συνέχεια, μελετώνται διάφοροι αλγόριθμοι στατιστικής μάθησης και η χρήση συνελικτικών νευρωνικών δικτύων για την ανίχνευση κλάσεων αντικειμένων στις εικόνες των χρηστών η οποία πραγματοποιείται με τον αλγόριθμο YOLOv3. Τέλος, εκτελείται δοκιμή α/β με τη χρήση του εργαλείου Google OptimizeTM (Google LLC, Mountain View, US-CA) για τον έλεγχο της υπόθεσης ότι η χρήση της αυτόματης επισήμανσης των εικόνων μέσω της ανίχνευσης αντικειμένων μπορεί να βελτιώσει τις επιλογές των διαφημιζόμενων προϊόντων (συστάσεων). Τα αποτελέσματα δείχνουν ότι η αυτόματη επισήμανση εικόνων χρηστών, με τις γενικές κλάσεις ανιχνευθέντων αντικειμένων, οδηγεί σε αξιοσημείωτη βελτίωση της αποτελεσματικότητας του συστήματος παραγωγής συστάσεων σχετικών προϊόντων.

Γλωσσάριο

Ανάπτυξη Βάσει Δεδομένων, Πρόβλεψη Χρονοσειρών, Μηχανική Όραση, Ανίχνευση Αντικειμένων, Συνελικτικά Νευρωνικά Δίκτυα (CNN)





Abstract

The purpose of this project is the review and implementation of algorithms to monitor, predict and enhance the recommendation system of products proposed to users of an online social networking and photo sharing platform. To this end, a comprehensive proposal to strengthen the recommendation system through object detection in user images is proposed. The proposal follows data-based software development techniques including, among others, the creation of a central data warehouse, data-driven decision-making, a/b testing and automated reporting. In particular, time series analysis and prediction algorithms are examined and the Prophet algorithm is used to monitor and estimate the future values of the selected recommendations. Subsequently, a/b testing is performed using the Google OptimizeTM tool to confirm the assessment that using auto-tagging images through object detection can improve the selection of advertised products (recommendations). Finally, various algorithms for statistical learning are being reviewed along with the use of convolutional neural networks to detect object classes in user images performed with the YOLOv3 algorithm. The results show that automatic tagging of user images, with the general classes of detected objects, leads to a significant improvement in the efficiency of the product recommendation system.

Keywords

Data-Driven Development, Time Series Forecasting, Computer Vision, Object Detection, Convolutional Neural Networks (CNN)



Αλέξανδρος Μ. Σούσος

Περιεχόμενα

1	Εισαγωγή	27
1.1	Σκοπός εργασίας	27
1.2	Εταιρική λογική	27
1.3	Σύνολο δεδομένων	29
2	Σύστημα παρακολούθησης και πρόβλεψης των επιτυχημένων συστάσεων	33
2.1	Ανάλυση χρονοσειρών	34
2.2	Πρόβλεψη χρονοσειρών	35
2.3	Αξιολόγηση προβλέψεων	36
2.3.1	Διάστημα εμπιστοσύνης	37
2.3.2	Μέσο απόλυτο σφάλμα	37
2.3.3	Μέσο σφάλμα τετραγώνου	38
2.3.4	Μέσο απόλυτο ποσοστιαίο σφάλμα	38
2.3.5	Μέσο ποσοστιαίο σφάλμα	39
2.4	Μέθοδοι πρόβλεψης	39
2.4.1	Μέθοδος Naive	39
2.4.2	Μέθοδος Seasonal Naive	40
2.4.3	Μέθοδοι εξομάλυνσης	41
2.4.4	Απλός κινητός μέσος	41
2.4.5	Απλή εκθετική εξομάλυνση	43
2.4.6	Διπλός κινητός μέσος	44
2.4.7	Διπλή εκθετική εξομάλυνση	45
2.4.8	Εκθετική εξομάλυνση με προσαρμογή στην τάση	46
2.4.9	Εκθετική εξομάλυνση με προσαρμογή στην τάση και την εποχικότητα	47
2.4.10	Μέθοδοι διάσπασης χρονοσειρών	49
2.4.11	Μέθοδος (S)ARIMA	50
2.4.12	Μέθοδος TBATS	52
2.4.13	Μέθοδος Prophet	53
2.5	Εφαρμογή με τη μέθοδο Prophet	61

3	Ανίχνευση αντικειμένων στις εικόνες χρηστών	69
3.1	Μηχανική όραση	69
3.2	Πεδία της μηχανικής όρασης	72
3.3	Βαθιά μάθηση	77
3.4	Νευρωνικά δίκτυα	78
3.5	Συνελικτικά νευρωνικά δίκτυα	85
3.6	Ανίχνευση αντικειμένων με συνελικτικά νευρωνικά δίκτυα	92
3.6.1	R-CNN	94
3.6.2	Fast R-CNN	98
3.6.3	Faster R-CNN	100
3.6.4	YOLO	105
3.6.5	YOLOv2	108
3.6.6	YOLOv3	113
3.7	Εφαρμογή ανίχνευσης αντικειμένων με το YOLOv3	117
4	Δοκιμή α/β	125
4.1	Εισαγωγή	125
4.2	Σχεδιασμός πειραμάτων	126
4.3	Προσεγγίσεις υλοποίησης	129
4.4	Υλοποίηση με την πλατφόρμα βελτιστοποίησης Google Optimize™	130
5	Μελλοντικές Επεκτάσεις	139
	Αναφορές	143

Εικόνες

1.1	Στιγμιότυπο ενός μέρους μίας σελίδας φωτογραφίας εσωτερικού χώρου από έναν χρήστη της διαδικτυακής πλατφόρμας, όπου φαίνονται οι συστάσεις προϊόντων κάτω από την εικόνα.	31
2.1	Παράδειγμα πρόβλεψης χρονοσειράς με την απλοϊκή μέθοδο (Naive) [47]. . .	40
2.2	Παράδειγμα πρόβλεψης χρονοσειράς με τη μέθοδο Average, που αποτελεί υποπερίπτωση της μεθόδου απλού κινητού μέσου [47].	42
2.3	Παράδειγμα πρόβλεψης χρονοσειράς με τη μέθοδο απλού κινητού μέσου [47].	43
2.4	Οι συνιστώσες μίας χρονοσειράς όπως προκύπτουν από τη μέθοδο διάσπασης χρονοσειρών [47].	49
2.5	Παράδειγμα πρόβλεψης χρονοσειράς με τη μέθοδο SARIMA [47].	53
2.6	Η Analyst-in-the-loop προσέγγιση του Prophet [89].	55
2.7	Λογιστική συνάρτηση ή λογιστική καμπύλη [64]	56
2.8	Προεπιλεγμένα πιθανά σημεία αλλαγής τάσης της μεθόδου Prophet [27]. . . .	57
2.9	Τα πρώτα τέσσερα αθροίσματα της σειράς Fourier για την προσέγγιση ενός τετραγωνικού κύματος.	58
2.10	Προσαρμογή εποχικότητας για τη δοθείσα χρονοσειρά με τάξη Fourier $N = 10$ στη μέθοδο Prophet.	59
2.11	Προσαρμογή εποχικότητας για τη δοθείσα χρονοσειρά με τάξη Fourier $N = 20$ στη μέθοδο Prophet.	60
2.12	Σύγκριση σφαλμάτων πρόβλεψης των μεθόδων auto.ARIMA, εκθετικής εξομάλυνσης, Prophet, sNaive και TBATS. Οι προβλέψεις με τη μέθοδο Prophet είχαν σημαντικά χαμηλότερο σφάλμα πρόβλεψης από τις άλλες αυτοματοποιημένες μεθόδους πρόβλεψης [89].	62
2.13	Μέτρηση της απόδοσης του Prophet με το κριτήριο αξιολόγησης MAPE στη δοθείσα χρονοσειρά, σύμφωνα με τον εσωτερικό διαγνωστικό έλεγχο του Prophet.	63
2.14	Η χρονοσειρά των εβδομαδιαίων συνόλων επιλογών προτάσεων από τους χρήστες ("shopclicks").	63
2.15	Απεικόνιση της τάσης της δοθείσας χρονοσειράς όπως ανιχνεύτηκε από το Prophet.	64

2.16	Απεικόνιση της ετήσιας εποχικότητας της δοθείσας χρονοσειράς όπως ανιχνεύτηκε από το Prophet (στο διάστημα Ιούνιος-Μάιος).	64
2.17	Απεικόνιση της ετήσιας εποχικότητας της δοθείσας χρονοσειράς όπως ανιχνεύτηκε από το Prophet (στο διάστημα Ιανουάριος-Δεκέμβριος).	65
2.18	Πρόβλεψη της δοθείσας χρονοσειράς σε ένα διαδραστικό διάγραμμα όπου απεικονίζονται οι τιμές των πραγματικών και των προβλεπόμενων τιμών μαζί με τα διαστήματα αβεβαιότητας που υπολογίζει το μοντέλο.	66
2.19	Η πρόβλεψη της χρονοσειράς των shopclicks για τους επόμενους δύο μήνες (8 εβδομάδες) από το Prophet.	67
2.20	Επισήμανση δύο περιοχών στην έξοδο του Prophet που υπάρχουν έντονες ακραίες τιμές. (Α) υποεκτίμηση, (Β) υπερεκτίμηση	68
3.1	Παράδειγμα χειρόγραφων ψηφίων από το σύνολο δεδομένων MNIST [100]. . .	72
3.2	Παράδειγμα ανίχνευσης εικόνας με τον αλγόριθμο YOLO στο σύνολο δεδομένων COCO [75].	73
3.3	Παράδειγμα κατάτμησης εικόνας με τη χρήση του αλγορίθμου Mask R-CNN [41].	74
3.4	Παράδειγμα μεταφοράς τεχνοτροπίας εικόνας [32].	74
3.5	Παραδείγματα από τον χρωματισμό εικόνας [104].	75
3.6	Δομή ενός νευρώνα του εγκεφάλου [48].	79
3.7	Τοπολογική απεικόνιση ενός απλού νευρωνικού δικτύου [51].	80
3.8	Μαθηματικό μοντέλο λειτουργίας του τεχνητού νευρώνα [48].	81
3.9	Συνάρτηση ενεργοποίησης με κατώφλι.	82
3.10	Συνάρτηση ενεργοποίησης RELU.	82
3.11	Συνάρτηση ενεργοποίησης tanh.	82
3.12	Συνάρτηση ενεργοποίησης logistic.	82
3.13	Ο “κύκλος” μεταξύ μεγάλων δεδομένων, αλγοριθμικών βελτιώσεων, και ισχυρότερων υπολογιστικών επιδόσεων [51].	85
3.14	Στο ραβδόγραμμα απεικονίζεται το σφάλμα ταξινόμησης για τις νικήτριες ομάδες του διαγωνισμού ImageNet για τα χρόνια 2010 έως 2017. Από το 2012 και έπειτα τα CNNs κυριαρχούν στις υλοποιήσεις που κερδίζουν τον διαγωνισμό [71].	86
3.15	Στο ραβδόγραμμα απεικονίζεται το σφάλμα τοποθέτησης για τις νικήτριες ομάδες του διαγωνισμού ImageNet (έτη 2010-2017) [71].	87
3.16	Βήμα (stride) και περιθώριο (padding) σε δισδιάστατα CNNs [81].	90
3.17	Απεικονίσεις των διαφόρων σταδίων επεξεργασίας σε ένα συνελικτικό επίπεδο [51].	91
3.18	Δύο τύποι προσεγγίσεων για την ανίχνευση αντικειμένων με χρήση CNN [105].	93

3.19	Μία γραφική αναπαράσταση του ορισμού του IoU [54].	93
3.20	Το διάγραμμα ροής του αλγορίθμου R-CNN αποτελείται από τρία στάδια: (1) εξαγωγή προτάσεων θέσης αντικειμένου, (2) υπολογισμός χαρακτηριστικών για κάθε πρόταση με τη χρήση CNN, και (3) ταξινόμηση κάθε περιοχής σε εξειδικευμένες κλάσεις με γραμμικούς ταξινομητές SVM [34].	95
3.21	Εφαρμογή του αλγορίθμου Selective Search [91]. Πάνω: απεικόνιση του αποτελέσματος της κατάτμησης (segmentation) του αλγορίθμου. Κάτω: απεικόνιση των προτάσεων περιοχών θέσης του αλγορίθμου.	96
3.22	Οπτική απεικόνιση της γενικής ιδέας της μείωσης του αριθμού ανιχνεύσεων σε μία εικόνα με την εφαρμογή του αλγορίθμου Non-Maximum Suppression (NMS) [73].	97
3.23	Ολόκληρη η εικόνα τροφοδοτεί ένα μοντέλο CNN για την ανίχνευση των περιοχών ενδιαφέροντος στους χάρτες χαρακτηριστικών. Κάθε περιοχή χωρίζεται χρησιμοποιώντας ένα επίπεδο ομαδοποίησης περιοχών ενδιαφέροντος (RoI Pooling) και τροφοδοτεί πλήρως συνδεδεμένα επίπεδα. Στη συνέχεια, χρησιμοποιείται ένας ταξινομητής Softmax για την ταξινόμηση του αντικειμένου και ένας γραμμικός παλινδρομητής για την προσαρμογή των συντεταγμένων του πλαισίου θέσης του αντικειμένου [33].	99
3.24	Η αρχιτεκτονική του Fast R-CNN [33].	100
3.25	Το RPN στο Faster R-CNN. Τα προκαθορισμένα k anchor boxes συνελίσσονται με κάθε παράθυρο ολίσθησης για να παράγουν διανύσματα σταθερού μήκους, που λαμβάνονται από τα επίπεδα cls και reg, για να αποκτήσουν αντίστοιχες εξόδους [77].	101
3.26	Η αρχιτεκτονική του Faster R-CNN. Ολόκληρη η εικόνα τροφοδοτεί ένα CNN για την παραγωγή anchor boxes ως προτάσεις περιοχών αντικειμένου μαζί με τον αντίστοιχο βαθμό εμπιστοσύνης για την παρουσία αντικειμένου. Λαμβάνονται ως είσοδος οι χάρτες χαρακτηριστικών και οι προτάσεις περιοχής, και για κάθε πλαίσιο παράγονται οι πιθανότητες για την ύπαρξη ενός αντικειμένου και ρυθμίζεται διορθωτικά η θέση του πλαισίου [77].	103
3.27	Καθορισμός προτάσεων περιοχών αντικειμένου στην αρχιτεκτονική Faster R-CNN. Η προκαθορισμένη επιλογή είναι η χρήση 9 anchor boxes, τριών σχημάτων σε διαφορετικές διαστάσεις τα οποία εξετάζονται σε κάθε θέση του χάρτη χαρακτηριστικών. Ο αριθμός αυτός μπορεί να αλλάξει ανάλογα με τις απαιτήσεις του κάθε προβλήματος [77].	104
3.28	Το ενοποιημένο μοντέλο ανίχνευσης αντικειμένων του YOLO [75].	105
3.29	Χωρισμός της εικόνας σε πλέγμα σύμφωνα με τον αλγόριθμο YOLO ($S = 13$) [75].	106
3.30	Η έξοδος του δικτύου του YOLO για $B = 20$ κλάσεις και $S = 7$ [75].	107

3.31 Παράδειγμα της αρχιτεκτονικής YOLO [34] για εικόνα εισόδου 448 x 448. Το YOLO έχει 24 συνελικτικά επίπεδα ακολουθούμενα από δύο ολικά συνδεδεμένα επίπεδα. Η χρήση 1 x 1 συνελικτικών επιπέδων επιλέγεται για να μειώσει τον χώρο των χαρακτηριστικών από τα προηγούμενα επίπεδα [75].	107
3.32 Αποτελέσματα mAP και FPS του YOLO για το VOC 2007 [75].	108
3.33 Συνδυασμός μοντέλων ανίχνευσης αντικειμένων για βελτίωση της ακρίβειας [75].	108
3.34 Σύγκριση αποτελεσμάτων mAP και FPS διαφόρων αλγορίθμων ανίχνευσης αντικειμένων στο PASCAL VOC 2007 Dataset [74].	109
3.35 Το ιεραρχικό δέντρο που προέκυψε από τον συνδυασμό των δύο συνόλων δεδομένων, COCO και ImageNet. Με μπλε χρώμα απεικονίζονται οι κλάσεις του COCO και με κόκκινο χρώμα οι κλάσεις του ImageNet [74].	111
3.36 Σύγκριση αποτελεσμάτων mAP και GPU Time διαφόρων αλγορίθμων ανίχνευσης αντικειμένων στο COCO Dataset [59].	112
3.37 Η αρχιτεκτονική του Darknet-53 που χρησιμοποιεί το YOLOv3 [58].	114
3.38 Η αρχιτεκτονική του YOLOv3 [21].	115
3.39 Παράδειγμα των Residual nets [42].	116
3.40 Σύγκριση αποτελεσμάτων mAP και ταχύτητας διαφόρων αλγορίθμων ταχείας ανίχνευσης αντικειμένων [59, 76]	117
3.41 Εικόνα χρήστη (A) της διαδικτυακής πλατφόρμας.	119
3.42 Εικόνα χρήστη (B) της διαδικτυακής πλατφόρμας.	119
3.43 Ανίχνευση κλάσεων αντικειμένων του YOLO με κατώφλι 0.1 σε φωτογραφία χρήστη της διαδικτυακής πλατφόρμας.	120
3.44 Ανίχνευση κλάσεων αντικειμένων του YOLO με κατώφλι 0.1 σε φωτογραφία χρήστη της διαδικτυακής πλατφόρμας.	120
3.45 Ανίχνευση κλάσεων αντικειμένων του YOLO με κατώφλι 0.6 σε φωτογραφία χρήστη της διαδικτυακής πλατφόρμας.	121
3.46 Ανίχνευση κλάσεων αντικειμένων του YOLO με κατώφλι 0.6 σε φωτογραφία χρήστη της διαδικτυακής πλατφόρμας	122
3.47 Αποτύπωση της εκτέλεσης του αλγορίθμου YOLO σε φωτογραφία χρήστη της διαδικτυακής πλατφόρμας (Εικόνα 1 από 2). Παρατηρούνται τα συνελικτικά επίπεδα, τα επίπεδα των residual nets, των downsamples και των upsamples που ενυπάρχουν στα 106 επίπεδα του Darknet-53, μαζί με τους χρόνους εκτέλεσης του κάθε επιπέδου.	123

3.48	Αποτύπωση της εκτέλεσης του αλγορίθμου YOLO σε φωτογραφία χρήστη της διαδικτυακής πλατφόρμας (Εικόνα 2 από 2). Παρατηρούνται τα συνελικτικά επίπεδα, τα επίπεδα των residual nets, των downsamples και των upsamples που ενυπάρχουν στα 106 επίπεδα του Darknet-53, μαζί με τους χρόνους εκτέλεσης του κάθε επιπέδου.	124
4.1	Γενική δομή μίας δοκιμής α/β στο διαδίκτυο.	128
4.2	Απεικόνιση της γενικής ιδέας της χρήσης βήτα κατανομών για την έκφραση μίας posterior κατανομής με βάση τα νέα δεδομένα (evidence) και την πρότερη γνώση (prior belief).	131
4.3	Παράδειγμα βήτα κατανομών για διάφορες τιμές των παραμέτρων α , β	131
4.4	Παράδειγμα εμφάνισης επικάλυψης μεταξύ των δύο posterior κατανομών.	132
4.5	Παράδειγμα διαγράμματος σχετικής απόδοσης των εκδόσεων A και B, με βάση τα δείγματα της προσομοίωσης Monte Carlo.	133
4.6	Γενικά στοιχεία και αποτελέσματα στατιστικής ανάλυσης α/β δοκιμής με το εργαλείο Google Optimize TM	134
4.7	Αποτελέσματα στατιστικής ανάλυσης του Google Optimize TM για τον αντικειμενικό στόχο "shopclicks".	135
4.8	Τα στατιστικά στοιχεία της ομάδας της πρωτότυπης έκδοσης του ιστοτόπου σύμφωνα με τη στατιστική ανάλυση του Google Optimize TM	135
4.9	Τα στατιστικά στοιχεία της ομάδας της παραλλαγμένης έκδοσης του ιστοτόπου σύμφωνα με τη στατιστική ανάλυση του Google Optimize TM	135
4.10	Ο ρυθμός μετατροπής (conversion rate) του αντικειμενικού στόχου "shopclicks" της παραλλαγμένης έκδοσης σε σύγκριση με την πρωτότυπη στην πορεία του χρόνου σύμφωνα με τη στατιστική ανάλυση του Google Optimize TM	137



Αλέξανδρος Μ. Σούσος



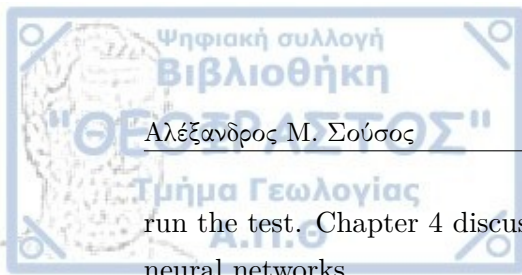
Summary

In a period of abundant content, it is increasingly difficult to find the information that is potentially desired. In this context, it is vital for internet based companies, which usually provide this abundance of content, to be able to serve users with the right content with personalized recommendations or other sophisticated solutions. The development of recommendation systems, although a field dominated by computer scientists and mathematicians, is also a field that needs the contribution of socio-economic sciences [25] as recommendation systems are the ones converting users' preferences into forecasts of their future preferences, shopping, interests.

This is especially important for e-commerce companies, but it is also very useful for users who have to search for their personal preferences among a huge number of products [16, 80]. This is especially clear for web companies that have a wealth of content (products, images, news, etc.) and are trying to take advantage of the long tail of their content. In the sense of "long tail", these are considered items that are not at the top of the list of users' interests, but because of their large number they can generate significant profits for companies [1, 2].

The purpose of this project is to propose a mechanism for providing valid recommendations to members of an online community. By providing users with an easier path for potentially desired products, it is expected to increase some of the company's core metrics such as more product clicks, more page views, or increased comments and frequency of user traffic.

In more detail, and in the context of the above, the project is structured as follows. Chapter 1 describes the data set, the corporate logic and the scope of the work. In Chapter 2, some basic concepts of time series are first developed, and then various time series analysis and forecasting algorithms are examined. To develop the system of monitoring and predicting the time series of users' choices, the Prophet method was used, in particular its implementation in the programming language R. For the implementation on the web platform of the object detection and classification, YOLOv3 was selected for its high accuracy in real-time execution. Having made a hypothesis that the use of automatic tagging of images can improve the number of advertised products being selected by users, it was considered appropriate to carry out an a/b test to prove it. Chapter 3 describes key elements of online a/b testing and uses Google's OptimizeTM optimization platform to



Αλέξανδρος Μ. Σούσος

run the test. Chapter 4 discusses various object detection algorithms using convolutional neural networks.



Ευχαριστίες

Ευχαριστώ ιδιαιτέρως τον επιβλέποντά μου, Ιωάννη Αντωνίου, για τη συνολική του συνεισφορά καθ' όλη τη διάρκεια των μεταπτυχιακών σπουδών μου, και συγκεκριμένα για τις δεξιότητες και τον τρόπο σκέψης, που αποτελούν εφόδια για την επαγγελματική μου σταδιοδρομία.

Επιπλέον, ευχαριστώ τον συνεπιβλέποντα, Χαράλαμπο Μπράτσα, για την ουσιαστική του υποστήριξη καθ' όλη τη διάρκεια της επεξεργασίας της και την όποια διευκόλυνση χρειάστηκε για την περάτωσή της.

Ευχαριστώ θερμά τους συμφοιτητές μου Έλλη Τσολακίδου, Δανάη Παπαθεοδώρου και Γιώργο Χριστόπουλο για την πολύτιμη βοήθεια τους καθ' όλη τη διάρκεια των μεταπτυχιακών σπουδών μου.

Επιπλέον, η εργασία δεν θα είχε πραγματοποιηθεί χωρίς τη βοήθεια των συναδέλφων μου στην εταιρεία, που επέδειξαν την υπομονή και ικανοποίησαν κάθε μου επιθυμία κατά της διάρκεια της συγγραφής της.

Τέλος, ευχαριστώ τους φίλους και την οικογένειά μου για την υποστήριξή τους σε αυτήν την προσπάθεια.



Αλέξανδρος Μ. Σούσος

Πρόλογος

Στην περίοδο αφθονίας περιεχομένου, η αποτελεσματική εύρεση των δυνητικά επιθυμητών πληροφοριών καθίσταται συνεχώς δυσχερέστερη. Σε αυτό το πλαίσιο, είναι ζωτικής σημασίας για τις εταιρείες με βάση το διαδίκτυο, που παρέχουν συνήθως αυτήν την αφθονία περιεχομένου, να βρίσκονται σε θέση να εξυπηρετούν τους χρήστες με το σωστό περιεχόμενο με εξατομικευμένες συστάσεις ή άλλες εκλεπτυσμένες λύσεις. Η ανάπτυξη των συστημάτων συστάσεων, παρ' ότι αποτελεί κύρια πεδίο της επιστήμης των υπολογιστών και των μαθηματικών, αποτελεί επίσης πεδίο σημαντικής συμβολής των κοινωνικοοικονομικών επιστημών [25], καθώς τα συστήματα συστάσεων είναι τα εργαλεία, τα οποία μετατρέπουν τις προτιμήσεις των χρηστών σε προβλέψεις των μελλοντικών προτιμήσεων, αγορών, συμπεριπόντων τους.

Αυτό είναι ιδιαίτερα σημαντικό για τα έσοδα των εταιρειών που βρίσκονται στον χώρο του ηλεκτρονικού εμπορίου, αλλά είναι επίσης πολύ χρήσιμο και για τους χρήστες, οι οποίοι καλούνται να αναζητήσουν για τις προσωπικές τους προτιμήσεις ανάμεσα σε έναν τεράστιο αριθμό προϊόντων [16, 80]. Αυτό είναι ιδιαίτερα σαφές για τις διαδικτυακές εταιρείες που διαθέτουν αφθονία περιεχομένου (όπως προϊόντα, εικόνες ή ειδήσεις) και προσπαθούν να επωφεληθούν από τη "μακριά ουρά" (long tail) του περιεχομένου τους. "Μακριά ουρά" θεωρούνται τα στοιχεία εκείνα που δεν βρίσκονται στην κορυφή της λίστας των ενδιαφερόντων των χρηστών, αλλά λόγω του μεγάλου αριθμού τους μπορούν να αποφέρουν σημαντικά κέρδη για τις εταιρείες [1, 2].

Σκοπός της παρούσας εργασίας είναι να προτείνει έναν μηχανισμό για την παροχή έγκυρων συστάσεων στα μέλη μίας διαδικτυακής κοινότητας. Παρέχοντας στους χρήστες ένα ευκολότερο μονοπάτι για τα δυνητικά επιθυμητά προϊόντα, αναμένεται να οδηγήσει στη βελτίωση της εμπειρίας χρήστη και ως συνέπεια, σε αύξηση κάποιων εκ των βασικών μετρικών της εταιρείας, όπως περισσότερες επικροτήσεις (clicks) σε προϊόντα, περισσότερες προβολές σελίδων ή αύξηση των σχολίων και της επισκεψιμότητας των χρηστών.

Αναλυτικότερα, για την επίτευξη του παραπάνω σκοπού, η εργασία αναπτύσσεται ως εξής. Στο κεφάλαιο 1 περιγράφεται το σύνολο δεδομένων, η εταιρική λογική και το αντικείμενο της εργασίας. Στο κεφάλαιο 2, αρχικά, αναπτύσσονται οι βασικές έννοιες των χρονοσειρών και στη συνέχεια, εξετάζονται διάφοροι αλγόριθμοι ανάλυσης και πρόβλεψης χρονοσειρών. Για την ανάπτυξη του συστήματος παρακολούθησης και πρόβλεψης της χρονοσειράς των επιλογών των χρηστών, χρησιμοποιήθηκε η μέθοδος Prophet και συγκεκριμένα η υλοποίηση

της στη γλώσσα προγραμματισμού R. Στο κεφάλαιο 3 εξετάζονται διάφοροι αλγόριθμοι ανίχνευσης αντικειμένων με τη χρήση συνελικτικών νευρωνικών δικτύων. Για την εφαρμογή της ανίχνευσης αντικειμένων και των κλάσεων τους στη διαδικτυακή πλατφόρμα επιλέχθηκε το YOLOv3 για τη μεγάλη του ακρίβεια σε πραγματικούς χρόνους εκτέλεσης. Στο κεφάλαιο 4, ακολουθώντας τη διατύπωση της υπόθεσης ότι η χρήση της αυτόματης επισήμανσης των εικόνων μπορεί να βελτιώσει τις επιλογές των διαφημιζόμενων προϊόντων από τους χρήστες, πραγματοποιήθηκε ο έλεγχος της υπόθεσης με την εφαρμογή μίας δοκιμής α/β. Περιγράφονται βασικά στοιχεία των δοκιμών α/β στο διαδίκτυο και χρησιμοποιείται η πλατφόρμα βελτιστοποίησης της GoogleTM, Google OptimizeTM για την εκτέλεση της δοκιμής.



Κεφάλαιο 1

Εισαγωγή

1.1 Σκοπός εργασίας

Ο σκοπός της παρούσας εργασίας είναι η παρουσίαση πρότασης προς μία πλατφόρμα διαδικτυακής διαφήμισης, σχετική με τη διακόσμηση εσωτερικών χώρων, για την αυτόματη επισήμανση εικόνων χρηστών με στόχο την αύξηση των επιτυχημένων συστάσεων που εμφανίζονται ως συνιστώμενα προϊόντα σχετικών με τις εικόνες (Εικόνα 1.1). Η χρήση ετικετών στις εικόνες χρηστών επιτρέπει στον μηχανισμό παραγωγής συστάσεων να παράξει συστάσεις προϊόντων με την ίδια ή παρόμοια επισήμανση. Ωστόσο, συχνά η χρήση των ετικετών παραλείπεται από τους χρήστες, με αποτέλεσμα ο μηχανισμός να καθίσταται άχρηστος και να επιστρέφει τυχαία προϊόντα προς σύσταση. Ως μηχανισμός, αναφέρεται το κομμάτι εκείνο του php κώδικα που αναλαμβάνει να επιστρέφει έναν πίνακα προτεινόμενων προϊόντων σχετικών με την εικόνα του χρήστη. Για την εύρεση των ετικετών των εικόνων, μελετήθηκαν οι τελευταίες εξελίξεις στον επιστημονικό τομέα της μηχανικής όρασης και συγκεκριμένα η χρήση συνελικτικών νευρωνικών δικτύων με σκοπό την εύρεση των αντικειμένων που εντοπίζονται μέσα στην εικόνα. Για την καταγραφή και απεικόνιση της εξέλιξης του ρυθμού επιλογής των προτεινόμενων προϊόντων από τους χρήστες, αξιοποιήθηκαν τεχνικές πρόβλεψης χρονοσειρών και συγκεκριμένα αναπτύχθηκε ένα πρόγραμμα στη γλώσσα R με τη χρήση της βιβλιοθήκης Prophet της FacebookTM Core Data Science Team (Facebook, Inc., Menlo Park, US-CA) [27]. Για τον έλεγχο της υπόθεσης ότι οι αυτόματες ετικέτες θα οδηγήσουν σε αύξηση της επιλογής συστάσεων, διενεργήθηκε δοκιμή α/β με τη χρήση του εργαλείου Google OptimizeTM (Google LLC, Mountain View, US-CA) [38]. Τέλος, παρουσιάζονται ιδέες για τη βελτίωση αλλά και επέκταση της παρούσας πρότασης.

1.2 Εταιρική λογική

Το αντικείμενο της εταιρείας είναι μία διαδικτυακή πλατφόρμα, υλοποιημένη σε DrupalTM <https://www.overleaf.com/project/5ddeb15bccd664000197e65f> CMS <https://www.overleaf.com/project/5ddeb15bccd664000197e65f> [24], της μεγαλύτερης γερμανόφωνης κοινότητας στο διαδίκτυο σχετικά με τη διακόσμηση εσωτερικού χώρου

(interior design). Ο σκοπός του κοινοτικού αυτού ιστοτόπου είναι να διευκολύνει και να ενθαρρύνει την ανάπτυξη ενός κοινωνικού δικτύου, καθώς και την επικοινωνία εντός αυτού του δικτύου, σχετικά με οτιδήποτε αφορά στη διακόσμηση και διαρρύθμιση ενός σπιτιού. Τα μέλη του ιστοτόπου εκθέτουν φωτογραφίες με τις προσωπικές τους επιλογές για την διακόσμηση των σπιτιών τους και ανταλλάσσουν απόψεις για όλες τις πτυχές της ζωής στο σπίτι. Ο ιστοτόπος περιστρέφεται γύρω από ιδέες και εμπνεύσεις πραγματικής ζωής για όλα τα σημαντικά θέματα σχεδίασης, είτε αυτά αφορούν στον σχεδιασμό ενός ολόκληρου διαμερίσματος, είτε αφορούν στην αναζήτηση νέων ιδεών διακόσμησης. Τα μέλη μπορούν να ανεβάσουν φωτογραφίες από το σπίτι τους, να εμπνεύσουν τους άλλους με τις ιδέες τους ή να θέσουν μία ερώτηση. Η εταιρεία έχει επίσης παρουσία σε μεγάλα κοινωνικά δίκτυα όπως το FacebookTM, το InstagramTM (της Facebook, Inc. επίσης) και το PinterestTM (Pinterest Inc., San Francisco, US-CA).

Όσον αφορά στις εικόνες χρηστών, οι ενέργειες που μπορεί να κάνει ένας χρήστης στην πλατφόρμα είναι οι εξής:

- Να μεταφορτώσει μία εικόνα,
 - να την επισημάνει με τυχόν προκαθορισμένες ετικέτες,
 - να τη συνδέσει με τυχόν εμπορικά σήματα ή προϊόντα που υπάρχουν στη φωτογραφία,
 - να προσθέσει μία σύντομη περιγραφή της εικόνας.
- Να επισημάνει ως "αγαπημένη" μία εικόνα άλλου χρήστη.
- Να σχολιάσει μία εικόνα.

Εκτός των εικόνων που μεταφορτώνουν οι χρήστες, η εταιρεία συνεργάζεται επίσης με πολλά καταστήματα και φιλοξενεί προϊόντα από μία μεγάλη γκάμα εμπορικών σημάτων που εκτίθενται σε διάφορα τμήματα του ιστοτόπου. Όσον αφορά στα ετειθέμενα προϊόντα, οι ενέργειες που μπορεί να κάνει ένας χρήστης στην πλατφόρμα είναι οι εξής:

- Να περιηγηθεί στον τομέα προϊόντων με εκατοντάδες χιλιάδες προϊόντα.
- Ο χρήστης δεν μπορεί να αγοράσει το προϊόν απευθείας από την πλατφόρμα αλλά επιλέγοντας ένα από αυτά ανακατευθύνεται σε ένα από τα πολλά συνεργαζόμενα ηλεκτρονικά καταστήματα που πουλούν το συγκεκριμένο προϊόν διαδικτυακά.
- Να επισημάνει ως "αγαπημένο" ένα προϊόν.

Επιπλέον, ο χρήστης μπορεί να προβεί σε ενέργειες που αφορούν άμεσα άλλους χρήστες όπως:

- Να "ακολουθήσει" άλλους χρήστες, έτσι ώστε να μπορεί να έχει μία εξατομικευμένη ροή για τις εικόνες του.
- Να προσθέσει εικόνες άλλων σε προσωπικά άλμπουμ.
- Να επικοινωνήσει με άλλον χρήστη με προσωπικό μήνυμα.

1.3 Σύνολο δεδομένων

Το σύνολο δεδομένων περιλαμβάνει 293.622 εικόνες χρηστών και 1.153.727 εικόνες προϊόντων (τελευταία ενημέρωση 19 Νοεμβρίου 2019) καθώς και μία χρονοσειρά με το σύνολο των επικροτήσεων (clicks) σε διαφημιζόμενα προϊόντα ανά εβδομάδα, από τον Ιανουάριο του 2015 έως τον Νοέμβριο του 2019. Το σύνολο δεδομένων παρήχθη απευθείας από τη βάση δεδομένων της πλατφόρμας μέσω ερωτημάτων SQL και κατόπιν συμφωνίας με την εταιρεία. Το σύνολο δεδομένων παραχωρήθηκε μόνο για την υλοποίηση αυτής της συγκεκριμένης εργασίας και δεν είναι διαθέσιμο για χρήση για άλλο σκοπό. Το σύνολο δεδομένων ή άλλες πληροφορίες σχετικά με την πλατφόρμα δεν χρησιμοποιήθηκαν ποτέ σε άλλες ερευνητικές ή εμπορικές εφαρμογές.

Οι βασικοί δείκτες απόδοσης της εταιρείας είναι έξι:

- αριθμός επισκέψεων (visits)
- προβολές σελίδων (pageviews)
- αριθμός μεταφορτωμένων εικόνων (image uploads)
- αριθμός επικροτήσεων (clicks) σε διαφημιζόμενα προϊόντα (shopclicks)
- αριθμός επισήμανσης εικόνων ως αγαπημένων (likes)
- αριθμός σχολίων (comments)

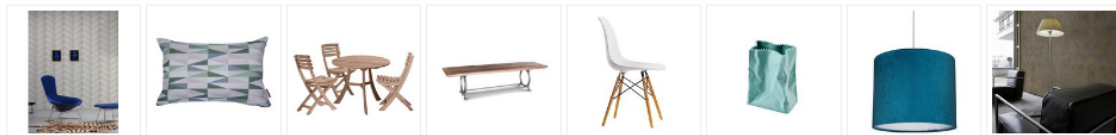
Δεδομένου ότι η εταιρεία μέχρι πρότινος παρήγαγε και αντλούσε δεδομένα από διάφορες πηγές και διαδικτυακά εργαλεία σε όλο το φάσμα του web development κρίθηκε αναγκαίο να αντληθούν αυτά τα δεδομένα σε ένα κοινό τόπο (data warehouse) με σκοπό την αξιοποίηση τους για τη λήψη αποφάσεων. Η απόφαση έγινε με γνώμονα τη δημιουργία ενός αρχιτεκτονικού μοντέλου για τη ροή δεδομένων από διάφορες πηγές προς ένα ενοποιημένο περιβάλλον υποστήριξης αποφάσεων. Το πλήθος των δεδομένων από τα διάφορα εν χρήσει εργαλεία που αποφασίστηκε να αντληθούν παρουσιάζονται συνοπτικά παρακάτω:

Google Analytics™	
Visitor	Page Views, Visit's Duration, Pages/Visit, Bounce Rate
Page Load Time	
Sources and Behavior	Google™, Direct, Pinterest™, Facebook™, Email
SEO	Google™ Visits, Pages per Session, Bounce Rate, /wohnen visits, /shop visits
Shopping Clicks	ImagePage Clicks, ProductPage Clicks, Categories Conversion Rate, Banner Clicks, ArticleInline clicks, TermPage Clicks
Device Categories	Mobile visits, Desktop visits, Tablet visits, Proportions
New Relic™	
Page Load Time	
Drupal™	
Conversions	
Flags	User2Node, User2User
CTR	views / node / week, product's clicks, uids, IPs, User Agents
History	uids, nids, timestamp
Social media	connected accounts
User Activity	logins, registers
User Location	
Comments	
New Images	
New Registrations	
Shopclicks	
Sistrix™	
Score	
Μέσα κοινωνικής δικτύωσης	
Δεδομένα μέσω κοινωνικής δικτύωσης όπως τάσεις και εξελίξεις αγοράς	

Για την άντλησή τους χρησιμοποιήθηκαν τα APIs των εργαλείων (όπου αυτά ήταν διαθέσιμα) με τη χρήση γλωσσών δέσμης ενεργειών (scripting languages) και προγραμματιστικών γλωσσών. Δεδομένου ότι κάποια από τα εργαλεία διατηρούν τα δεδομένα για ορισμένο μόνο χρονικό διάστημα, η μετακίνησή τους σε μία κοινή και "ιδιόκτητη" βάση δεδομένων δίνει επιπλέον το πλεονέκτημα της χρήσης όλου του πλήθους των δεδομένων για μελλοντική επεξεργασία. Όπως γίνεται αντιληπτό και από τον πίνακα που παρατέθηκε παραπάνω το μεγαλύτερο πλήθος δεδομένων βρισκόταν ήδη εντός του CMS Drupal™ σε μία

Ανάπτυξη συστήματος αυτόματης επισημάνσης εικόνων χρηστών διαδικτυακής κοινότητας με σκοπό τη βελτίωση του συστήματος συστάσεων σχετικών προϊόντων

γνωστή και λειτουργική δομή και έτσι θεωρήθηκε λογικό να μεταφερθούν και τα υπόλοιπα δεδομένα εντός του Drupal™, κάτι που βοηθάει στην εξαγωγή συμπερασμάτων και την άμεση χρήση τους στη διαδικτυακή πλατφόρμα.



Εικόνα 1.1: Στιγμιότυπο ενός μέρους μίας σελίδας φωτογραφίας εσωτερικού χώρου από έναν χρήστη της διαδικτυακής πλατφόρμας, όπου φαίνονται οι συστάσεις προϊόντων κάτω από την εικόνα.



Κεφάλαιο 2

Σύστημα παρακολούθησης και πρόβλεψης των επιτυχημένων συστάσεων

Για την παρακολούθηση αλλά και αξιολόγηση του εβδομαδιαίου αριθμού των επιλεγμένων προϊόντων υπήρχε η ανάγκη για ένα εργαλείο παρακολούθησης της πορείας τους με ένα απλό σύστημα πρόβλεψης των άμεσων μελλοντικών τιμών και ο εντοπισμός κάποιων περιπτώσεων ακραίων τιμών. Για τον σκοπό αυτόν χρησιμοποιήθηκε η συνολική εβδομαδιαία καταμέτρηση της επιλογής από τους χρήστες ενός συνιστώμενου προϊόντος από το 2015 έως τον Νοέμβριο του 2019. Το σύνολο αυτών των δεδομένων αποτελεί μία χρονοσειρά, καθώς εκφράζει την εξέλιξη των τιμών μίας μεταβλητής κατά τη διάρκεια ίσων διαδοχικών χρονικών περιόδων. Ειδικότερα, μία χρονοσειρά είναι μία ακολουθία από δεδομένα καταγεγραμμένα στην πορεία του χρόνου. Για τα δεδομένα από χρονοσειρές, η σειρά παίζει ρόλο. Συγκεκριμένα, οι τιμές σε μία χρονοσειρά εκφράζουν μία εξάρτηση από τον χρόνο. Επομένως αν αλλάξουμε τη σειρά μίας χρονοσειράς, αλλάζουμε και τη σημασία των δεδομένων.

μία άλλη κατηγορία που μπορούν να καταταχθούν οι χρονοσειρές είναι με βάση τον τρόπο προσδιορισμού των μελλοντικών δεδομένων, δηλαδή με βάση τη σχέση των διαδοχικών παρατηρήσεων που τις απαρτίζουν. Όταν οι παρατηρήσεις δεν είναι ανεξάρτητες μεταξύ τους, τότε οι μελλοντικές τιμές μπορούν να προσδιοριστούν από τις προηγούμενες. Ένα τέτοιο σύστημα ονομάζεται ντετερμινιστικό. Ωστόσο, στις περισσότερες χρονοσειρές στον πραγματικό κόσμο, οι μελλοντικές τιμές καθορίζονται μερικώς από το παρελθόν, επειδή εμπλέκεται και ο "τυχαίος παράγοντας", άρα σε αυτήν την περίπτωση έχουμε τις λεγόμενες στοχαστικές χρονοσειρές.

Μαθηματικά μία χρονοσειρά ορίζεται από τις τιμές $Y_1, Y_2, \dots$ κάποιας μεταβλητής Y (παραδείγματος χάρη θερμοκρασία ή τιμή κλεισίματος μετοχής) κατά τις χρονικές στιγμές $t_1, t_2, \dots$. Επομένως το Y είναι μία συνάρτηση του t , και αυτό συμβολίζεται ως $Y = F(t)$. Η γραφική παράσταση της συνάρτησης $Y = F(t)$ παρουσιάζει την εξέλιξη της μεταβλητής Y στον χρόνο.

2.1 Ανάλυση χρονοσειρών

Οι χρονοσειρές περιλαμβάνουν τις παρελθοντικές τιμές μίας μεταβλητής σε διαδοχικές καταστάσεις στον χρόνο. Έτσι, μπορούν να αναλυθούν ώστε να βγουν συμπεράσματα για τη συμπεριφορά της μεταβλητής. Η πληροφορία που εσωκλείουν οι παρελθοντικές τιμές της μεταβλητής μπορεί να οδηγήσει στην πρόβλεψη της ακολουθίας των μελλοντικών παρατηρήσεων. Η μελέτη του γραφήματος μίας χρονοσειράς οδηγεί στην αναγνώριση των βασικών ποιοτικών χαρακτηριστικών της, τα οποία είναι η βασική της στάθμη, η τάση, η εποχικότητα, η κυκλικότητα και η τυχαιότητα. Αυτά τα χαρακτηριστικά παίζουν καθοριστικό ρόλο στη μετέπειτα πορεία της διαδικασίας παραγωγής προβλέψεων.

Η τάση μπορεί να είναι ανοδική, φθίνουσα ή μηδενική και δηλώνει τη μακροπρόθεσμη μεταβολή του μέσου επιπέδου των τιμών της χρονοσειράς. Εάν μία χρονοσειρά εμφανίζει μία θετική μακροπρόθεσμη κλίση με την πάροδο του χρόνου, έχει ανοδική τάση. Αν αντίθετα, περιγράφει μία γενική αρνητική κλίση, έχει πτωτική τάση. Η γενική τάση μπορεί επίσης να αλλάξει κατεύθυνση. Μπορεί να υπάρχει τάση προς τα άνω ή τάση προς τα κάτω. Τέλος, μία σταθερή ή οριζόντια τάση ορίζει μία χρονοσειρά χωρίς θετικά ή αρνητικά μακροπρόθεσμα μοτίβα.

Η εποχικότητα είναι ένα μοτίβο διακύμανσης των τιμών της χρονοσειράς που προκαλείται από συμβάντα που σχετίζονται με κάποια συγκεκριμένη χρονική περίοδο και εμφανίζονται περιοδικά (όπως η εποχή του χρόνου, ο μηνιαίος κύκλος ή η ώρα της ημέρας). Η εποχικότητα έχει πάντα σταθερή συχνότητα. Δηλαδή, η έναρξη και η λήξη ενός εποχικού μοτίβου είναι χρονικά σταθερές.

Η κυκλικότητα είναι ένα χαρακτηριστικό παρόμοιο με την εποχικότητα, με τη διαφορά ότι η περιοδικότητα και η διάρκεια κύκλου δεν εμφανίζουν σταθερότητα. Τα κυκλικά πρότυπα δεν είναι επαναλαμβανόμενα. Πιο συγκεκριμένα, η κυκλικότητα εμφανίζεται κατά περιόδους, το μήκος των οποίων συνήθως ξεπερνά το ένα έτος, και χαρακτηρίζεται ως μία κυματοειδή μεταβολή λόγω εξωτερικών παραγόντων που συνήθως είναι δύσκολο να προβλεφθούν.

Οι τυχαίες κινήσεις (ή αλλιώς ακανόνιστες διακυμάνσεις) αποτελούν τον υπολειπόμενο παράγοντα αν απομονωθούν οι τρεις πρώτοι παράγοντες. Συνήθως παρουσιάζονται ως σποραδικές, ακανόνιστες κινήσεις λόγω τυχαίων παραγόντων και γεγονότων.

Οι τυχαίες κινήσεις επηρεάζουν τις τιμές των χρονοσειρών κατά ένα τυχαίο και μη συστηματικό τρόπο, ο οποίος δεν μπορεί να προσδιοριστεί. Οι ασυνέχειες αποτελούν υποσύνολο των ακανόνιστων διακυμάνσεων και είναι οι απομονωμένες και απότομες αλλαγές που παρατηρούνται σε μία χρονοσειρά. Ανάλογα με το αν έχουν περιοδική ή μόνιμη συμπεριφορά, χωρίζονται σε outliers και σε level-shifts, αντίστοιχα. Οι outliers αντιπροσωπεύουν ασυνήθιστες παρατηρήσεις που οφείλονται σε τυχαία ή απρόβλεπτα γεγονότα ενώ οι level-shifts αποτελούν τις απότομες αλλαγές στο μέσο επίπεδο των χρονοσειρών.

Σε μία συγκεκριμένη χρονοσειρά είναι δυνατόν να μην συνυπάρχουν και οι τέσσερις συνιστώσες αλλά μόνο κάποιες από αυτές. Η ανάλυση χρονοσειρών συνίσταται στην περιγραφή (εν γένει με μαθηματικό τρόπο) των συνιστωσών κινήσεων που υπάρχουν.

Για την ανάλυση των χρονοσειρών χρησιμοποιούμε τους ακόλουθους συμβολισμούς:

Y_t = Πραγματική τιμή της χρονοσειράς

T_t = Τάση

S_t = Εποχικότητα

C_t = Κυκλικότητα

I_t = Τυχαίες κινήσεις

όπου $t = 1, 2, 3, \dots, n$

Η εξέταση των στοιχείων αυτών γίνεται σύμφωνα με κάποιο μαθηματικό υπόδειγμα που φανερώνει τον τρόπο με τον οποίο οι παρατηρήσεις της χρονοσειράς προσδιορίζονται από τις συνιστώσες της χρονοσειράς. Τα χρησιμοποιούμενα υποδείγματα είναι το αθροιστικό μοντέλο (additive model) και το πολλαπλασιαστικό μοντέλο (multiplicative model).

Στο αθροιστικό μοντέλο οι πραγματικές τιμές της χρονοσειράς για κάθε περίοδο θεωρούνται ως το άθροισμα των τεσσάρων συνιστωσών και όλες οι συνιστώσες είναι εκφρασμένες στην ίδια μονάδα μέτρησης με εκείνη των παρατηρήσεων της χρονοσειράς:

$$Y_t = T_t + S_t + C_t + I_t \quad (2.1)$$

Αντίθετα στο πολλαπλασιαστικό μοντέλο οι πραγματικές τιμές της χρονοσειράς προσδιορίζονται από το γινόμενο των τεσσάρων συνιστωσών. Σε αυτήν την περίπτωση μόνο η τάση είναι εκφρασμένη στην ίδια μονάδα μέτρησης με εκείνη της χρονοσειράς Y_t ενώ τα στοιχεία C_t , S_t και I_t είναι δείκτες ανεξάρτητοι από μονάδες μέτρησης:

$$Y_t = T_t S_t C_t I_t \quad (2.2)$$

2.2 Πρόβλεψη χρονοσειρών

Η πρόβλεψη χρονοσειρών είναι η χρήση στατιστικών μεθόδων για την πρόβλεψη της μελλοντικής συμπεριφοράς βασισμένη σε ιστορικά δεδομένα. Πολλές εταιρείες εξερευνούν πλέον την πρόβλεψη χρονοσειρών ως ένα μέσο για τη λήψη καλύτερων επιχειρηματικών αποφάσεων. Η πρόβλεψη χρονοσειρών έχει συνήθως δύο στόχους. Πρώτον, ο προσδιορισμός των προτύπων που εξηγούν τη συμπεριφορά των χρονοσειρών. Δεύτερον, η χρήση αυτών των προτύπων για την πρόβλεψη νέων τιμών.

Η διαδικασία πρόβλεψης αποτελείται από πέντε βασικά βήματα [65]. Το πρώτο βήμα για την παραγωγή προβλέψεων είναι να καθοριστεί πλήρως η φύση του προβλήματος και η χρησιμότητα των προβλέψεων. Το επόμενο βήμα είναι η συλλογή όλων των ιστορικών στοιχείων που χαρακτηρίζουν το πρόβλημα, καθώς και η συλλογή όλων των δεδομένων

κάποιου άλλου συστήματος που μπορεί έμμεσα να βοηθήσει στον χαρακτηρισμό του κύριου προβλήματος. Το τρίτο βήμα είναι η ανάλυση της χρονοσειράς που δημιουργείται από τα ιστορικά δεδομένα, δηλαδή η μελέτη του γραφήματος της χρονοσειράς, ώστε να αναγνωριστούν ορισμένα βασικά χαρακτηριστικά της, όπως τα ποιοτικά χαρακτηριστικά ή η μορφή της. Με αυτόν τον τρόπο θα είναι ξεκάθαρο αν υπάρχει σημαντική τάση, εποχικότητα, κυκλικότητα ή ορισμένες ασυνήθιστες τιμές. Σε αυτό το βήμα μπορεί να γίνει και η απομόνωση των εποχικών ή άλλων συνιστωσών για την περαιτέρω μελέτη. Η επιλογή της κατάλληλης μεθόδου πρόβλεψης αποτελεί το επόμενο βήμα που αποτελεί και ένα διαρκώς μελετούμενο ερευνητικό αντικείμενο. Οι βασικοί παράγοντες που καθορίζουν την επιλογή της μεθόδου είναι τα διαθέσιμα στοιχεία, ο σκοπός πρόβλεψης, η περίοδος και ο ορίζοντας πρόβλεψης, το κόστος της μεθόδου, η ζητούμενη ακρίβεια, η απλότητα και ευκολία εφαρμογής ή γενικότερα το κόστος της μεθόδου. Το πέμπτο και τελευταίο στάδιο της διαδικασίας πρόβλεψης είναι η χρήση και η αξιολόγηση των μοντέλων που επιλέχθηκαν να χρησιμοποιηθούν.

Η πρόβλεψη χρονοσειρών έχει μία πλούσια οικογένεια αλγορίθμων. Κάθε ένας τους παίρνει μία διαφορετική υπόθεση για να προβλέψει νέες τιμές. Για την επιλογή της κατάλληλης μεθόδου χρησιμοποιούνται τα κριτήρια αξιολόγησης των μεθόδων προβλέψεων. Τα κριτήρια αυτά βασίζονται στις τιμές των αποκλίσεων των προβλεπόμενων τιμών από τις αντίστοιχες πραγματικές τιμές της χρονοσειράς.

Για μία μεταβλητή Y , η απόκλιση της προβλεπόμενης τιμής $\hat{Y}_t$ από την αντίστοιχη πραγματική τιμή Y_t για την περίοδο t , όπου $t = 1, 2, 3, \dots, n$, ονομάζεται σφάλμα πρόβλεψης (forecast error), συμβολίζεται με e_t και ορίζεται ως:

$$e_t = Y_t - \hat{Y}_t \quad (2.3)$$

Η παραπάνω σχέση εκφράζει για κάθε περίοδο t τη διαφορά μεταξύ της πραγματικής τιμής Y_t και της αντίστοιχης προβλεπόμενης τιμής $\hat{Y}_t$ που προήλθε από τη μέθοδο πρόβλεψης που χρησιμοποιήθηκε. Το σφάλμα της πρόβλεψης είναι ουσιαστικά η διαφορά της πραγματικής τιμής από την προβλεπόμενη. Επομένως, για να προσδιορίσουμε την αξιοπιστία μίας συγκεκριμένης μεθόδου πρόβλεψης, θα πρέπει να μελετήσουμε τη διαχρονική συμπεριφορά των τιμών των σφαλμάτων της πρόβλεψης.

2.3 Αξιολόγηση προβλέψεων

Για την αξιολόγηση της προβλεπτικής ικανότητας των αλγορίθμων έχουν προταθεί διάφορα κριτήρια. Τα κριτήρια αυτά μπορεί να είναι εξαρτώμενα από την κλίμακα και τον τύπο των τιμών της χρονοσειράς ή μπορούν να εκφράζονται ποσοστιαία και να είναι απαλλαγμένα από την όποια μονάδα μέτρησης των τιμών της χρονοσειράς [15]. Τα δύο περισσότερο χρησιμοποιούμενα κλιμακο-εξαρτώμενα μέτρα είναι τα απόλυτα σφάλματα και τα τετραγωνικά

σφάλματα. Λόγω της εξάρτησής τους από την κλίμακα και μονάδα μέτρησης του σφάλματος, e_t , δεν μπορούν να χρησιμοποιηθούν για σύγκριση μεταξύ χρονοσειρών που εμπλέκουν διαφορετικούς τύπους τιμών. Τα ποσοστιαία σφάλματα δίνονται ως $p_t = 100 * e_t / y_t$. Έχουν το πλεονέκτημα να είναι απαλλαγμένα από τη μονάδα μέτρησης της τιμής της χρονοσειράς και συχνά χρησιμοποιούνται για να αξιολογήσουν τις επιδόσεις των αλγορίθμων πρόβλεψης εφαρμοσμένα σε διαφορετικά σύνολα δεδομένων. Το πιο διαδεδομένο κριτήριο σε αυτήν την κατηγορία είναι το μέσο απόλυτο ποσοστιαίο σφάλμα που εξετάζει τη συμπεριφορά της απόλυτης τιμής του σφάλματος της πρόβλεψης σε σχέση με την πραγματική τιμή της χρονοσειράς.

2.3.1 Διάστημα εμπιστοσύνης

Γύρω από την προβλεπόμενη τιμή μπορεί να διατυπωθεί ένα διάστημα εμπιστοσύνης με βάση τη διακύμανση του σφάλματος πρόβλεψης. Αν παραστήσουμε με σ^2 τη διακύμανση του σφάλματος πρόβλεψης, το $1-\alpha$ διάστημα εμπιστοσύνης για την αληθινή τιμή της χρονοσειράς ύστερα από h περιόδους, είναι:

$$\hat{Y}_{T+h} - z_{\alpha/2} * \sigma_h \leq Y_{T+h} \leq \hat{Y}_{T+h} + z_{\alpha/2} * \sigma_h \quad (2.4)$$

Η διακύμανση είναι συνάρτηση της διακύμανσης του όρου e_t , η τιμή του οποίου είναι άγνωστη. Οπότε και πρέπει να γίνει εκτίμησή του από το άθροισμα των τετραγώνων των σφαλμάτων:

$$\hat{\sigma}^2 = \frac{\sum \hat{e}^2}{T - p - q} \quad (2.5)$$

Το διάστημα εμπιστοσύνης της πρόβλεψης είναι ένα καλό μέτρο αξιολόγησης της πρόβλεψης το οποίο μας δείχνει την ακρίβεια της πρόβλεψης.

2.3.2 Μέσο απόλυτο σφάλμα

Το μέσο απόλυτο σφάλμα (Mean Absolute Error, MAE) ή μέση απόλυτη απόκλιση (Mean Absolute Deviation, MAD) ορίζεται ως το άθροισμα των απόλυτων τιμών του σφάλματος της πρόβλεψης διαιρούμενο με τον αριθμό των περιόδων n , στις οποίες έγιναν προβλέψεις, δηλαδή:

$$MAD = \frac{1}{n} \sum_{t=1}^n |Y_t - \hat{Y}_t| = \frac{1}{n} \sum_{t=1}^n |e_t| \quad (2.6)$$

Το μέσο απόλυτο σφάλμα εκφράζει τη μέση τιμή των απόλυτων αποκλίσεων των προβλεπόμενων τιμών της χρονοσειράς από τις αντίστοιχες πραγματικές. Η μονάδα μέτρησης του είναι η ίδια με εκείνη των τιμών της χρονοσειράς γεγονός που βοηθάει στην ερμηνεία του. Το γεγονός αυτό καθιστά το κριτήριο διαδεδομένο όταν συγκρίνονται μεταξύ τους μέθοδοι,

εφαρμοσμένοι σε ίδια χρονοσειρά ή σε διάφορες χρονοσειρές με τον ίδιο τύπο μονάδων. Στον υπολογισμό του λαμβάνονται υπόψη μόνο οι απόλυτες τιμές των σφαλμάτων καθιστώντας το ανεξάρτητο από θετικές (υπερεκτίμηση) ή αρνητικές τιμές (υποεκτίμηση). Επιπλέον, βασίζεται στην υπόθεση ότι η σοβαρότητα του σφάλματος ή το κόστος που δημιουργείται από το σφάλμα της πρόβλεψης σχετίζεται γραμμικά με το μέγεθος του σφάλματος.

2.3.3 Μέσο σφάλμα τετραγώνου

Το μέσο σφάλμα τετραγώνου (Mean Square Error, MSE) ορίζεται ως το άθροισμα των τετραγώνων των σφαλμάτων διαιρούμενο με τον αριθμό των χρονικών περιόδων n , στις οποίες έγιναν προβλέψεις, δηλαδή:

$$MSE = \frac{1}{n} \sum_{t=1}^n (Y_t - \hat{Y}_t)^2 = \frac{1}{n} \sum_{t=1}^n e_t^2 \quad (2.7)$$

Το MSE είναι η μέση τιμή των τετραγώνων των αποκλίσεων των προβλεπόμενων τιμών της χρονοσειράς από τις αντίστοιχες πραγματικές. Η μονάδα μέτρησης του MSE είναι εκφρασμένη στη μονάδα μέτρησης των τιμών των παρατηρήσεων, υψωμένη στο τετράγωνο. Συνήθως, χρησιμοποιούμε τη θετική τιμή της τετραγωνικής του ρίζας, που ονομάζεται τετραγωνική ρίζα μέσου σφάλματος τετραγώνου RMSE (Root Mean Squared Error):

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{n} \sum_{t=1}^n e_t^2} \quad (2.8)$$

Το RMSE εκφράζεται στην ίδια μονάδα μέτρησης με εκείνη των τιμών της χρονοσειράς. Η ύπαρξη προβλέψεων που απέχουν πολύ από τις αντίστοιχες πραγματικές τιμές γίνεται πολύ περισσότερο αισθητή με το κριτήριο MSE από ότι με το κριτήριο MAD, επειδή οι τιμές των σφαλμάτων της πρόβλεψης υψώνονται στο τετράγωνο. Συνεπώς το κριτήριο MSE είναι στατιστικά περισσότερο αξιόπιστο από το κριτήριο MAD και χρησιμοποιείται συχνότερα για την επιλογή της 'κατάλληλης' μεθόδου πρόβλεψης. Μία μέθοδος πρόβλεψης που ελαχιστοποιεί το MAE οδηγεί σε προβλέψεις για τη διάμεσο ενώ μία μέθοδος πρόβλεψης που ελαχιστοποιεί το RMSE θα οδηγήσει σε προβλέψεις για τη μέση τιμή. Επομένως, το RMSE χρησιμοποιείται ευρέως παρότι είναι πιο δύσκολη η ερμηνεία του.

$$MAE = \text{mean}(|e_t|) \quad (2.9)$$

$$RMSE = \sqrt{\text{mean}(e_t^2)} \quad (2.10)$$

2.3.4 Μέσο απόλυτο ποσοστιαίο σφάλμα

Το μέσο απόλυτο ποσοστιαίο σφάλμα (Mean Absolute Percentage Error, MAPE) εξετάζει τη συμπεριφορά της απόλυτης τιμής του σφάλματος της πρόβλεψης σε σχέση με την πραγματική

τιμή της χρονοσειράς. Ορίζεται ως το άθροισμα των απόλυτων τιμών των σφαλμάτων της πρόβλεψης προς τις αντίστοιχες πραγματικές τιμές της χρονοσειράς διαιρούμενο με τον αριθμό των χρονικών περιόδων n , στις οποίες έγιναν προβλέψεις, δηλαδή:

$$MAPE = \frac{1}{n} \sum_{t=1}^n \frac{|Y_t - \hat{Y}_t|}{Y_t} = \frac{1}{n} \sum_{t=1}^n \frac{|e_t|}{Y_t} \quad (2.11)$$

$$MAPE = \text{mean}(|p_t|) \quad (2.12)$$

Το κριτήριο αυτό είναι απαλλαγμένο από μονάδες μέτρησης και χρησιμοποιείται για να συγκρίνει την ακρίβεια μίας ή περισσότερων μεθόδων προβλέψεων και για περισσότερες από μία χρονοσειρές.

2.3.5 Μέσο ποσοστιαίο σφάλμα

Το μέσο ποσοστιαίο σφάλμα (Mean Percentage Error, MPE) είναι βασισμένο σε ποσοστά σφαλμάτων και έχει το μειονέκτημα να έχει τιμή άπειρη ή απροσδιόριστη όταν το $y_t = 0$ για οποιοδήποτε t στο χρονικό διάστημα ενδιαφέροντος. Επιπλέον έχει ακραία μεγάλες τιμές όταν το y_t πλησιάζει στο 0. Επιπλέον, ένα ακόμα μειονέκτημα είναι ότι υποθέτει ότι οι μετρήσεις της χρονοσειράς έχουν ένα σημασιολογικό μηδέν. Χρησιμοποιείται για τον προσδιορισμό της μεροληψίας της μεθόδου πρόβλεψης, δηλαδή αν οι προβλεπόμενες τιμές είναι συστηματικά μεγαλύτερες ή μικρότερες από τις αντίστοιχες πραγματικές. Αναμφισβήτητο όσο πιο κοντά στο μηδέν είναι η τιμή του MPE, τόσο πιο αμερόληπτη και καλή είναι η μέθοδος πρόβλεψης που χρησιμοποιήθηκε. Αντίθετα, μεγάλες απόλυτες τιμές του MPE φανερώνουν μεγάλη μεροληψία της μεθόδου [43].

$$MPE = \frac{1}{n} \sum_{t=1}^n \frac{Y_t - \hat{Y}_t}{Y_t} = \frac{1}{n} \sum_{t=1}^n \frac{e_t}{Y_t} \quad (2.13)$$

2.4 Μέθοδοι πρόβλεψης

Στη συνέχεια αναλύονται κάποιοι διαδεδομένοι μέθοδοι πρόβλεψης χρονοσειρών [10, 14, 47].

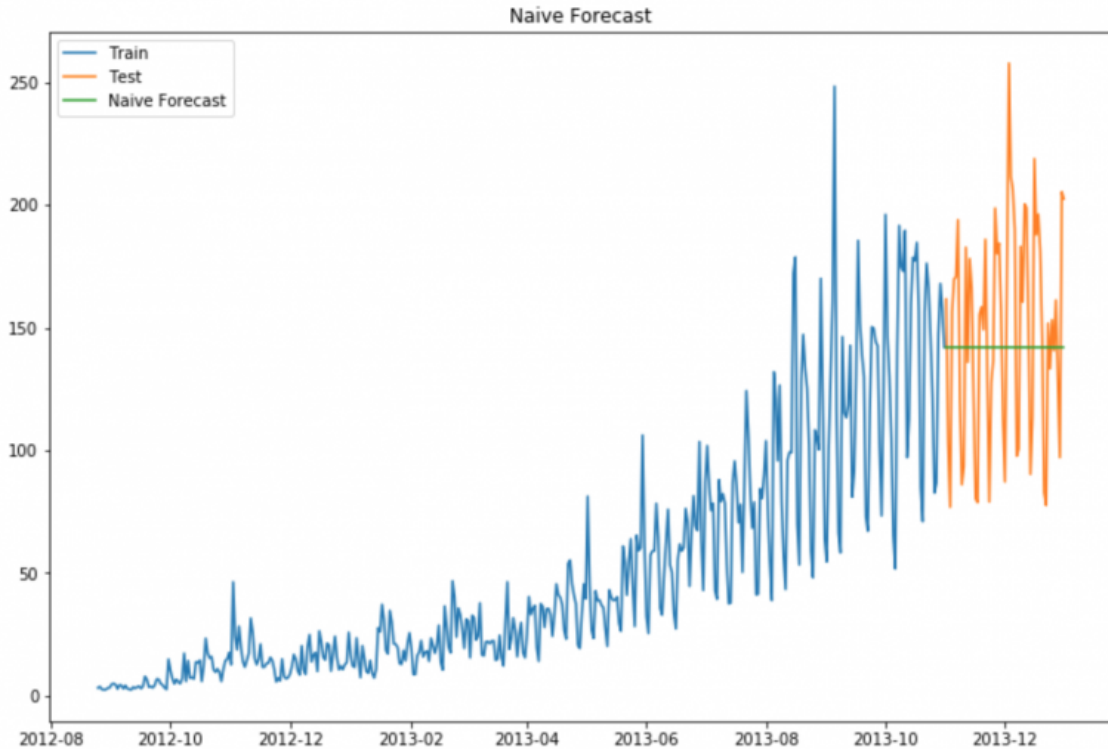
2.4.1 Μέθοδος Naive

Η απλοϊκή μέθοδος (Naive), αποτελεί την απλούστερη στατιστική μέθοδο, καθώς θεωρεί ότι η πρόβλεψη για την κάθε χρονική περίοδο t ισούται με την τιμή της προηγούμενης περιόδου $t-1$ των δεδομένων, δηλαδή υποθέτει ότι το επόμενο συμβάν θα είναι ίσο με το πιο πρόσφατο.

$$\hat{y}_{t+1} = y_t \quad (2.14)$$

ή πιο γενικά:

$$\hat{Y}(t+h|t) = Y(t) \quad (2.15)$$



Εικόνα 2.1: Παράδειγμα πρόβλεψης χρονοσειράς με την απλοϊκή μέθοδο (Naive) [47].

Τα πλεονεκτήματα της μεθόδου είναι η απλότητα και το χαμηλό κόστος υλοποίησης και επεξεργασίας της.

Ωστόσο, με τέτοιες απλές μεθόδους, σίγουρα δεν εκτιμάται μία ακριβής πρόβλεψη. Ως εκ τούτου, η χρήση τους στον επαγγελματικό κόσμο αποφεύγεται καθώς οι προβλέψεις συνήθως αποκλίνουν αρκετά των πραγματικών τιμών. Αυτό συμβαίνει διότι οι μέθοδοι αυτές δεν λαμβάνουν υπόψη πολλές διακυμάνσεις που συνήθως υπάρχουν σε δεδομένα χρονοσειρών. Η μέθοδος είναι αποδοτική όταν ο μέσος όρος, η τάση ή τα εποχικά φαινόμενα είναι σταθερά και η τυχαιότητα είναι μικρή. Οι μέθοδοι πρόβλεψης συνήθως εκμεταλλεύονται τα μοτίβα των χρονοσειρών για να παράγουν αξιόπιστες προβλέψεις. Παρ' όλ' αυτά, η απλοϊκή μέθοδος συνήθως χρησιμοποιείται ως σημείο αναφοράς απόδοσης άλλων πιο πολύπλοκων μεθόδων.

2.4.2 Μέθοδος Seasonal Naive

Η μέθοδος Seasonal Naive είναι μία τροποποίηση της απλοϊκής μεθόδου, καθώς λαμβάνει υπόψη την εποχικότητα των δεδομένων. Θεωρώντας ότι η χρονοσειρά έχει μία εποχική συνιστώσα και ότι η περίοδος της εποχικότητας είναι T , οι προβλέψεις που δίνονται από το SNaive model είναι:

$$\hat{Y}(t+h|t) = Y(t+h-T) \quad (2.16)$$

Επομένως οι προβλέψεις για τα επόμενα T χρονικά βήματα είναι ίσα με τα προηγούμενα T βήματα. Για παράδειγμα, αν το T ισούται με ένα έτος, τότε, η πρόβλεψη για μία τιμή του επόμενου χρόνου ισούται όχι με την τελευταία παρατήρηση, αλλά με την αντίστοιχη παρατήρηση (χρονική στιγμή) του προηγούμενου χρόνου. Η μέθοδος αυτή αν και μπορεί να ανιχνεύσει την εποχικότητα δεν λαμβάνει υπόψη την τάση της χρονοσειράς και ως εκ τούτου συχνά χρησιμοποιείται ως μοντέλο αναφοράς για τη σύγκριση με άλλες μεθόδους.

2.4.3 Μέθοδοι εξομάλυνσης

Οι μέθοδοι εξομάλυνσης (smoothing methods) είναι τεχνικές στις οποίες η δημιουργία των προβλέψεων προέρχεται από την εξομάλυνση της διαχρονικής εξέλιξης των τιμών της χρονοσειράς, ώστε να αναγνωριστεί καλύτερα ο τρόπος συμπεριφοράς της στην πορεία του χρόνου. Οι μέθοδοι αυτές μπορούν να προσδιορίσουν τις μελλοντικές τιμές βασισμένες στις διαθέσιμες παρατηρήσεις αλλά ανεξάρτητα της σχέσης που μπορεί να έχει η μεταβλητή με άλλες.

Αρχικά θα αναπτύξουμε τη μέθοδο του απλού κινητού μέσου m -περιόδων, σύμφωνα με την οποία η πρόβλεψη είναι ο αριθμητικός μέσος όρος των m πιο πρόσφατων παρατηρήσεων της χρονοσειράς. Στη συνέχεια θα παρουσιάσουμε τη μέθοδο της απλής εκθετικής εξομάλυνσης σύμφωνα με την οποία, η πρόβλεψη είναι ένας σταθμισμένος μέσος όρος των παρατηρήσεων της χρονοσειράς, οι οποίες όσο πιο πρόσφατες είναι, έχουν και μεγαλύτερη βαρύτητα. Έπειτα θα περιγράψουμε τις μεθόδους του διπλού κινητού μέσου, της διπλής εκθετικής εξομάλυνσης του Brown και του Holt, οι οποίες χρησιμοποιούνται όταν οι παρατηρήσεις της χρονοσειράς εμφανίζουν τάση. Οι τρεις αυτές μέθοδοι, σε αντίθεση με τις δύο προηγούμενες, μας δίνουν τη δυνατότητα να διενεργήσουμε προβλέψεις για περισσότερες από μία μελλοντικές περιόδους. Τέλος, θα αναλύσουμε τη μέθοδο Winters, η οποία εφαρμόζεται όταν στις παρατηρήσεις της χρονοσειράς υπάρχουν τα στοιχεία της τάσης και της εποχικότητας. Στόχος όλων αυτών των μεθόδων εξομάλυνσης είναι η αναγνώριση του τρόπου δημιουργίας των παρατηρήσεων της χρονοσειράς, ώστε να πραγματοποιηθούν οι καλύτερες δυνατές προβλέψεις.

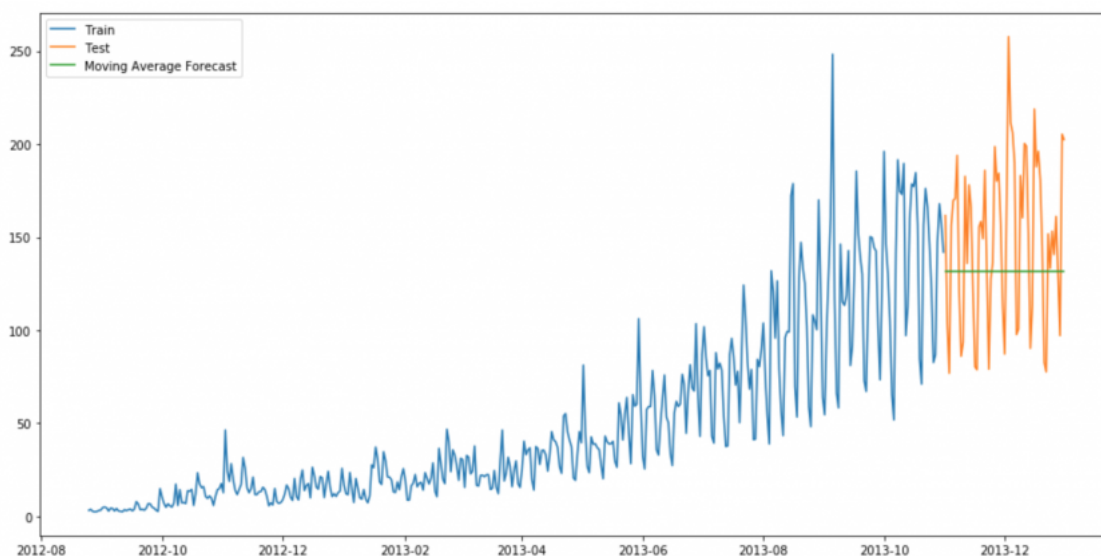
2.4.4 Απλός κινητός μέσος

Η μέθοδος αυτή θεωρεί ότι οι πιο πρόσφατες παρατηρήσεις της χρονοσειράς είναι οι πιο αντιπροσωπευτικές για την πρόβλεψη των μελλοντικών τιμών. Χρησιμοποιεί ως πρόβλεψη την τιμή του αριθμητικού μέσου όρου των m πρόσφατων παρατηρήσεων της χρονοσειράς. Η τιμή του μέσου όρου αναπροσαρμόζεται με την είσοδο κάθε νέας παρατήρησης και για αυτό η μέθοδος φέρει την ονομασία κινητός μέσος.

Οι προβλέψεις μίας χρονοσειράς Y_t , για $t = 1, 2, \dots, n$, δημιουργούνται με τη μέθοδο του απλού κινητού μέσου ως εξής:

$$\hat{Y}_{t+1} = \frac{1}{m} \sum_{j=1}^m Y_{t-j+1} = \frac{1}{m} (Y_t + Y_{t-1} + \dots + Y_{t-m+1}) = \hat{Y}_t + \frac{Y_t}{m} - \frac{Y_{t-m}}{m} \quad (2.17)$$

όπου $\hat{Y}_{t+1}$ είναι η πρόβλεψη για την περίοδο $(t+1)$ και m ο αριθμός των περιόδων που χρησιμοποιούνται για τον υπολογισμό της τιμής του μέσου όρου.



Εικόνα 2.2: Παράδειγμα πρόβλεψης χρονοσειράς με τη μέθοδο *Average*, που αποτελεί υποπερίπτωση της μεθόδου απλού κινητού μέσου [47].

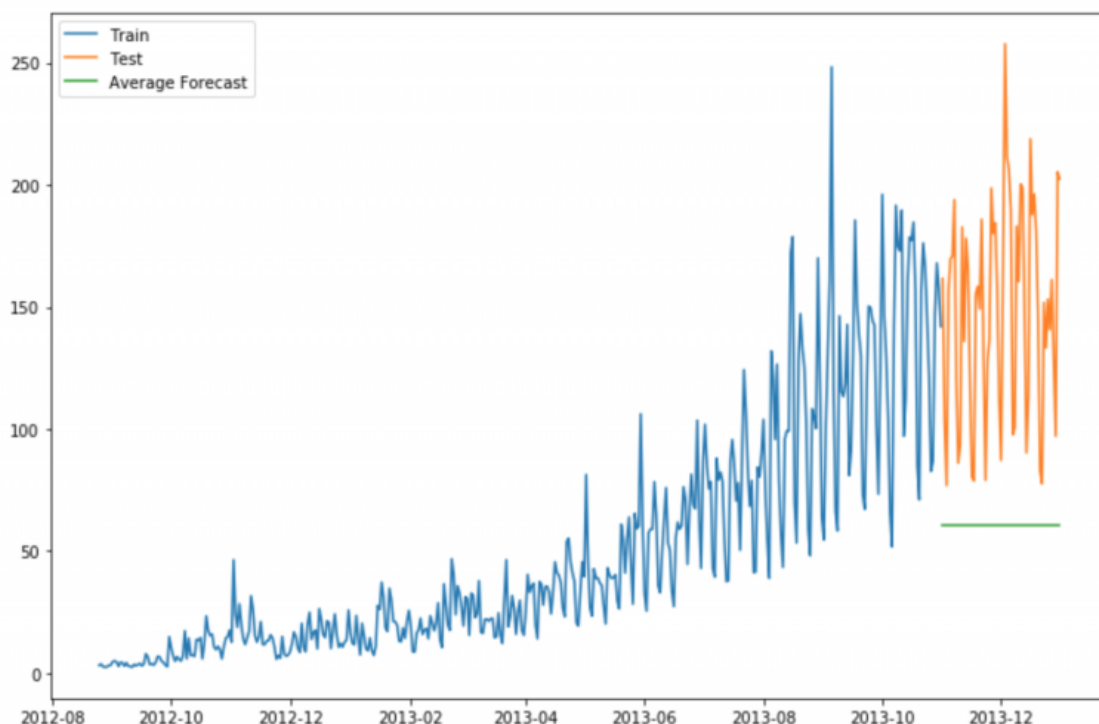
Για $m = 1$ η πρόβλεψη της επόμενης περιόδου είναι ίση με την πραγματική τιμή της προηγούμενης περιόδου

$$Y_{t+1} = Y_t \quad (2.18)$$

Για $m = n$ ίσο με το σύνολο όλων των παρατηρήσεων η μέθοδος υποθέτει ότι ένα μελλοντικό γεγονός περιγράφεται καλύτερα από τον μέσο όρο όλων των γεγονότων του παρελθόντος και επομένως επιστρέφει τον συνολικό μέσο όρο (*Average method*).

$$\hat{y}_{x+1} = \frac{1}{x} \sum_i^x y_i \quad (2.19)$$

Μία βελτιωμένη έκδοση της Μεθόδου Απλού Κινητού Μέσου είναι η Μέθοδος Σταθμισμένου Κινητού Μέσου. Στη μέθοδο Απλού Κινητού Μέσου, σταθμίζουμε εξίσου τις m τελευταίες παρατηρήσεις του παρελθόντος. Μπορούμε όμως να αντιμετωπίσουμε καταστάσεις όπου κάθε παρατήρηση από το παρελθόν επηρεάζει την πρόβλεψη με διαφορετικό



Εικόνα 2.3: Παράδειγμα πρόβλεψης χρονοσειράς με τη μέθοδο απλού κινητού μέσου [47].

τρόπο. Μία τέτοια τεχνική, η οποία ζυγίζει διαφορετικά τις παρελθοντικές παρατηρήσεις, ονομάζεται Μέθοδος Σταθμισμένου Κινητού Μέσου.

$$\hat{y}_i = \frac{1}{m}(w_1 * y_{i-1} + w_2 * y_{i-2} + w_3 * y_{i-3} + \dots + w_m * y_{i-m}) \quad (2.20)$$

2.4.5 Απλή εκθετική εξομάλυνση

Στη μέθοδο της απλής εκθετικής εξομάλυνσης (simple exponential smoothing), η πρόβλεψη είναι ένας σταθμισμένος μέσος όρος των παρατηρήσεων της χρονοσειράς, οι οποίες όσο πιο πρόσφατες είναι έχουν και μεγαλύτερη βαρύτητα. Σε αντίθεση με τη μέθοδο του απλού κινητού μέσου, η απλή εκθετική εξομάλυνση προσπαθεί να δώσει διαφορετική βαρύτητα σε κάθε παρατήρηση ανάλογα με το πόσο αυτή απέχει από την προβλεπόμενη περίοδο. Συγκεκριμένα, η μέθοδος δίνει μεγαλύτερη βαρύτητα στις πιο πρόσφατες παρατηρήσεις, από ότι στις απομακρυσμένες. Για να το πετύχει αυτό χρησιμοποιεί ως τιμή πρόβλεψης έναν σταθμισμένο μέσο όρο των παρατηρήσεων της χρονοσειράς.

$$\hat{Y}_{t+1} = \alpha Y_t + \alpha(1 - \alpha)Y_{t-1} + \alpha(1 - \alpha)^2 Y_{t-2} + \dots \quad (2.21)$$

όπου η παράμετρος α ονομάζεται σταθερά εξομάλυνσης (smoothing constant) και λαμβάνει τιμές μεταξύ 0 και 1 δηλαδή

$$0 \leq \alpha \leq 1 \quad (2.22)$$

Η σχέση μπορεί να γραφεί με την αναδρομική της μορφή ως εξής:

$$\hat{Y}_{t+1} = \alpha Y_t + (1 - \alpha) \hat{Y}_t \quad (2.23)$$

Όσο πιο μεγάλη είναι η τιμή της παραμέτρου α , τόσο μεγαλύτερη βαρύτητα δίνεται στις πιο πρόσφατες παρατηρήσεις και πολύ μικρή έως μηδαμινή βαρύτητα στις πιο απομακρυσμένες. Μπορεί να παρατηρηθεί ότι όταν $\alpha = 0$, τότε η κάθε πρόβλεψη ισούται με το αρχικό επίπεδο και όταν $\alpha = 1$, τότε το μοντέλο ταυτίζεται με την απλοϊκή μέθοδο. Το συγκεκριμένο μοντέλο υποθέτει ότι τα δεδομένα χαρακτηρίζονται από σταθερό μέσο όρο και κατ' επέκταση από απουσία τάσης. Χρησιμοποιείται κυρίως σε χρονοσειρές που παρουσιάζουν υψηλή τυχαιότητα ή θόρυβο.

2.4.6 Διπλός κινητός μέσος

Η μέθοδος διπλού κινητού μέσου (double moving average) επιτρέπει στην πρόβλεψη να λάβει υπόψη την όποια ανοδική ή πτωτική πορεία που μπορεί να εκφράζει κάποια γραμμική τάση. Για την πρόβλεψή της, η μέθοδος υπολογίζει από τον απλό κινητό μέσο έναν δεύτερο κινητό μέσο ενώ στη διαμόρφωση των προβλέψεων λαμβάνεται υπόψη και η γραμμική τάση. Για τον λόγο αυτόν η μέθοδος ονομάζεται πολύ συχνά και μέθοδος του γραμμικού κινητού μέσου (linear moving average). Η εφαρμογή της μεθόδου του διπλού κινητού μέσου βασίζεται στην ακόλουθη διαδικασία:

- i. Υπολογίζεται ο απλός κινητός μέσος m -περιόδων, M_t , ως:

$$M_{t+1} = \frac{1}{m} \sum_{j=1}^m Y_{t-j+1} \quad (2.24)$$

- ii. Υπολογίζεται ο διπλός κινητός μέσος m -περιόδων, M'_t , ως:

$$M'_{t+1} = \frac{1}{m} \sum_{j=1}^m M_{t-j+1} \quad (2.25)$$

- iii. Υπολογίζεται η διαφορά α_t ως:

$$\alpha_t = 2M_t - M'_t \quad (2.26)$$

- iv. Υπολογίζεται ο παράγοντας προσαρμογής για την τάση, b_t , ως:

$$b_t = \frac{2}{m-1} (M_t - M'_t) \quad (2.27)$$

v. Υπολογίζεται η πρόβλεψη $\hat{Y}_{t+h}$ για την h μελλοντική περίοδο ως:

$$\hat{Y}_{t+h} = \alpha_t + hb_t \quad (2.28)$$

όπου h είναι ένας ακέραιος θετικός αριθμός.

Η μέθοδος αυτή, σε αντίθεση με τις δύο προηγούμενες μεθόδους προβλέψεων που παρουσιάσαμε, μπορεί για $h > 1$ να χρησιμοποιηθεί για τη διενέργεια προβλέψεων για περισσότερες από μία μελλοντικές περιόδους, ενώ για $h = 1$ δίνει την πρόβλεψη για την επόμενη περίοδο. Βέβαια, η χρήση της προϋποθέτει την ύπαρξη μεγαλύτερου αριθμού παρατηρήσεων, ιδιαίτερα μάλιστα όταν η τιμή του m είναι σχετικά μεγάλη.

2.4.7 Διπλή εκθετική εξομάλυνση

Η διπλή εκθετική εξομάλυνση ή μέθοδος Brown [11] είναι μία ακόμα μέθοδος που επιχειρεί να προβλέψει τις μελλοντικές παρατηρήσεις μίας χρονοσειράς λαμβάνοντας υπόψη την όποια τάση αυτή η χρονοσειρά μπορεί να διαθέτει. Η βασική φιλοσοφία της μεθόδου αυτής είναι παραπλήσια με εκείνη της μεθόδου του διπλού κινητού μέσου, με την έννοια ότι η εξομάλυνση των παρατηρήσεων της χρονοσειράς γίνεται δύο φορές, και στη συνέχεια λαμβάνεται υπόψη και η όποια τάση της χρονοσειράς. Η εφαρμογή της μεθόδου της διπλής εκθετικής εξομάλυνσης στηρίζεται στην ακόλουθη διαδικασία:

- i. Εξομαλύνονται οι αρχικές παρατηρήσεις της χρονοσειράς με τη μέθοδο της απλής εκθετικής εξομάλυνσης ως ακολούθως:

$$A_t = \alpha Y_t + (1 - \alpha)A_{t-1} \quad (2.29)$$

όπου α είναι η σταθερά εξομάλυνσης, για $0 \leq \alpha \leq 1$, A_t οι εξομαλυνθείσες τιμές της χρονοσειράς που προκύπτουν από την πρώτη εξομάλυνση, για $t = 2, 3, \dots, n$, ενώ για $t = 1$ ορίζεται ως αρχική συνθήκη $A_1 = Y_1$.

- ii. Εξομαλύνονται οι εξομαλυνθείσες τιμές A_t της χρονοσειράς με τη μέθοδο της απλής εκθετικής εξομάλυνσης ως ακολούθως:

$$A'_t = \alpha A_t + (1 - \alpha)A'_{t-1} \quad (2.30)$$

όπου A'_t είναι οι εξομαλυνθείσες τιμές της χρονοσειράς που προκύπτουν από τη δεύτερη εξομάλυνση, για $t = 2, 3, \dots, m$ ενώ για $t = 1$, $A'_1 = A_1$.

- iii. Υπολογίζεται η διαφορά α_t ως:

$$\alpha_t = 2A_t - A'_t \quad (2.31)$$

iv. Υπολογίζεται ο παράγοντας προσαρμογής για την τάση, b_t , ως:

$$b_t = \frac{\alpha}{1 - \alpha}(A_t - A'_t) \quad (2.32)$$

v. Υπολογίζεται η πρόβλεψη $\hat{Y}_{t+h}$ για την h μελλοντική περίοδο ως:

$$\hat{Y}_{t+h} = \alpha_t + hb_t \quad (2.33)$$

όπου h είναι ένας ακέραιος θετικός αριθμός.

Η μέθοδος αυτή μπορεί να εφαρμοστεί για τη διαμόρφωση προβλέψεων για περισσότερες από μία μελλοντικές περιόδους σε αντίθεση με τη μέθοδο της απλής εκθετικής εξομάλυνσης, η οποία παρέχει προβλέψεις μόνο για την επόμενη χρονική περίοδο. Η τιμή της σταθεράς εξομάλυνσης επιλέγεται έτσι ώστε να ελαχιστοποιεί το κριτήριο MSE ή άλλο κριτήριο. Σημειώνουμε ότι ο αριθμός των παρατηρήσεων που απαιτούνται για την εφαρμογή της μεθόδου αυτής είναι αρκετά μικρότερος από τον αντίστοιχο αριθμό της μεθόδου του διπλού κινητού μέσου.

2.4.8 Εκθετική εξομάλυνση με προσαρμογή στην τάση

Η εκθετική εξομάλυνση με προσαρμογή στην τάση (exponential smoothing adjusted for trend ή μέθοδος Holt [44] καθώς εισήχθη από τον Holt το 1957), αποτελεί μία επέκταση του μοντέλου απλής εκθετικής εξομάλυνσης και είναι και αυτή μία μέθοδος ικανή να εντοπίσει την τάση στις παρατηρήσεις της χρονοσειράς. Χρησιμοποιείται σε χρονοσειρές που χαρακτηρίζονται από τη συνιστώσα της τάσης και καθώς η συνιστώσα αυτή παρατηρείται συχνά στα δεδομένα, είναι συνήθως πιο αποτελεσματική μέθοδος από την απλή εκθετική εξομάλυνση. Η μέθοδος, σε αντίθεση με τη μέθοδο της διπλής εξομάλυνσης έχει δύο παραμέτρους, την παράμετρο α για την εξομάλυνση των τιμών της χρονοσειράς και την παράμετρο β για την εξομάλυνση της τάσης. Η εφαρμογή της μεθόδου Holt βασίζεται στην ακόλουθη διαδικασία:

i. Η εξομάλυνση των τιμών της χρονοσειράς γίνεται με την ακόλουθη σχέση:

$$A_t = \alpha Y_t + (1 - \alpha)(A_{t-1} + T_{t-1}) \quad (2.34)$$

όπου α είναι η σταθερά για την εξομάλυνση των τιμών της χρονοσειράς, για $0 \leq \alpha \leq 1$, A_t οι εξομαλυνθείσες τιμές της χρονοσειράς, για $t = 2, 3, \dots, n$, ενώ για $t = 1$ ορίζεται ως αρχική συνθήκη $A_1 = Y_1$.

ii. Η εξομάλυνση της τάσης γίνεται ως εξής:

$$T_t = \beta(A_t - A_{t-1}) + (1 - \beta)T_{t-1} \quad (2.35)$$

όπου β , για $0 \leq \beta \leq 1$, είναι η σταθερά για την εξομάλυνση της τάσης, T_t οι εξομαλυνθείσες τιμές της τάσης, για $t = 2, 3, \dots, n$ ενώ για $t = 1$ ορίζεται ως αρχική συνθήκη $T_1 = 0$.

iii. Η πρόβλεψη $\hat{Y}_{t+h}$ για την h μελλοντική περίοδο προσδιορίζεται ως:

$$\hat{Y}_{t+h} = A_t + hT_t \quad (2.36)$$

όπου $h = 1, 2, 3, \dots$

Οι τιμές των παραμέτρων α, β επιλέγονται έτσι ώστε να ελαχιστοποιούν την τιμή κάποιου κριτηρίου (όπως MSE). Οι συντελεστές εξομάλυνσης παίρνουν τιμές στο διάστημα $[0, 1]$. Εδώ είναι σημαντικό να επισημανθεί ότι η βέλτιστη τιμή του συντελεστή της βάσης α είναι μεγαλύτερη από τη βέλτιστη τιμή του συντελεστή της τάσης β και ο λόγος είναι ότι η τιμή της βάσης είναι συνήθως πολύ μεγαλύτερη από την τιμή της τάσης. Στην πράξη, η μέθοδος Holt προτιμάται έναντι της μεθόδου Brown διότι παρέχει καλύτερα αποτελέσματα. Η εφαρμογή των δύο μεθόδων έχει δείξει ότι με τη μέθοδο Holt γίνεται καλύτερη εκτίμηση των τιμών της τάσης από ότι με τη μέθοδο Brown που είναι πολύ ευαίσθητη στις τυχαίες διακυμάνσεις της χρονοσειράς. Αυτό συμβαίνει διότι με τη μέθοδο Holt οι τιμές της τάσης εξομαλύνονται απευθείας ενώ με τη μέθοδο Brown η τάση προσδιορίζεται από τον παράγοντα b_t αφού προηγουμένως εξομαλυνθούν δύο φορές οι τιμές της χρονοσειράς.

2.4.9 Εκθετική εξομάλυνση με προσαρμογή στην τάση και την εποχικότητα

Η εκθετική εξομάλυνση με προσαρμογή στην τάση και την εποχικότητα εφαρμόζεται σε χρονοσειρές όπου παρατηρούνται εποχικές διακυμάνσεις, οι οποίες επαναλαμβάνονται ανά σταθερά χρονικά διαστήματα με την ίδια ή περίπου ίδια μορφή. Για την εξέταση της εποχικότητας χρησιμοποιούνται μέθοδοι εξομάλυνσης οι οποίες τη λαμβάνουν άμεσα υπόψη. Οι μέθοδοι αυτές μειώνουν το σφάλμα της πρόβλεψης σε τέτοιες χρονοσειρές, παρέχοντας έτσι καλύτερες προβλέψεις. Η εκθετική εξομάλυνση με προσαρμογή στην τάση και στην εποχικότητα (μέθοδος Winters [98]), αποτελεί επέκταση της μεθόδου Holt και φέρει τρεις παραμέτρους, α, β και γ , οι οποίες χρησιμοποιούνται για την εξομάλυνση των τιμών της χρονοσειράς, της τάσης και της εποχικότητας αντίστοιχα. Για τον λόγο αυτόν χρησιμοποιεί μία επιπλέον σχέση από τη μέθοδο Holt για την εξομάλυνση της εποχικότητας. Η εφαρμογή της μεθόδου Winters στηρίζεται στην ακόλουθη διαδικασία:

i. Η εξομάλυνση των τιμών της χρονοσειράς γίνεται με την ακόλουθη σχέση:

$$A_t = \alpha \frac{Y_t}{S_{t-L}} + (1 - \alpha)(A_{t-1} + T_{t-1}) \quad (2.37)$$

όπου α είναι η σταθερά για την εξομάλυνση των τιμών της χρονοσειράς, για $0 \leq \alpha \leq 1$, A_t οι εξομαλυνθείσες τιμές της χρονοσειράς, ενώ S_t είναι ο εποχικός συντελεστής

της περιόδου t και L η περιοδικότητα της εποχικότητας, δηλαδή $L = 12$ για μηνιαία δεδομένα, $L = 4$ για τριμηνιαία δεδομένα και ούτω καθεξής.

ii. Η εξομάλυνση της τάσης γίνεται όπως και στη μέθοδο Holt, δηλαδή ως εξής:

$$T_t = \beta(A_t - A_{t-1}) + (1 - \beta)T_{t-1} \quad (2.38)$$

όπου β , για $0 \leq \beta \leq 1$, είναι η σταθερά για την εξομάλυνση της τάσης ενώ T_t οι εξομαλυνθείσες τιμές της τάσης.

iii. Η εξομάλυνση της εποχικότητας γίνεται ως ακολούθως:

$$S_t = \gamma \frac{Y_t}{A_t} + (1 - \gamma)S_{t-L} \quad (2.39)$$

όπου γ , για $0 \leq \gamma \leq 1$, είναι η σταθερά για την εξομάλυνση της εποχικότητας.

iv. Η πρόβλεψη $\hat{Y}_{t+h}$ για τις h μελλοντικές περιόδους του πρώτου έτους προσδιορίζεται ως:

$$\hat{Y}_{t+h} = (A_t + hT_t)S_{t+h-L} \quad (2.40)$$

όπου $h = 1, 2, \dots, L$ και για τις h μελλοντικές περιόδους του δεύτερου έτους από τη σχέση:

$$\hat{Y}_{t+h} = (A_t + hT_t)S_{t+h-2L} \quad (2.41)$$

για $h = L + 1, L + 2, \dots, 2L$ και ούτω καθεξής.

Οι αρχικές συνθήκες των σχέσεων υπολογίζονται με τον ακόλουθο τρόπο:

a. Για $t = 1, 2, \dots, L - 1$ δεν προσδιορίζονται οι τιμές A_t , ενώ για $t = L$ το A_L ορίζεται ως:

$$A_L = \frac{Y_1 + Y_2 + \dots + Y_L}{L} \quad (2.42)$$

b. Για $t = 1, 2, \dots, L - 1$ δεν προσδιορίζονται οι τιμές T_t , ενώ για $t = L$, τίθεται $T_L = 0$.

c. Για $t = 1, 2, \dots, L$ οι τιμές των εποχικών συντελεστών S_t υπολογίζονται ως εξής:

$$S_t = \frac{Y_t}{A_L} \quad (2.43)$$

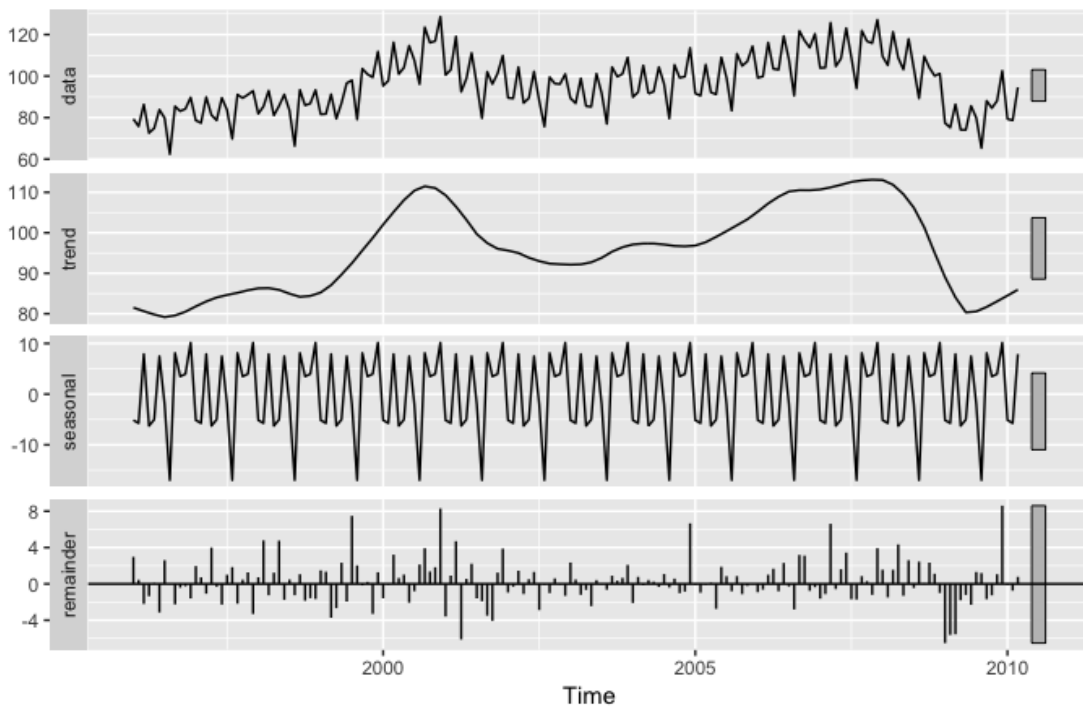
Όταν προσδιοριστούν οι παραπάνω αρχικές τιμές για το πρώτο έτος, η μέθοδος Winters, σύμφωνα με την παραπάνω διαδικασία, μπορεί να προβλέψει τις μελλοντικές τιμές της χρονοσειράς για περισσότερες από μία περιόδους. Οι τιμές των παραμέτρων α , β και γ επιλέγονται έτσι ώστε να ελαχιστοποιούν κάποιο κριτήριο αξιολόγησης προβλέψεων.

2.4.10 Μέθοδοι διάσπασης χρονοσειρών

Οι μέθοδοι διάσπασης χρονοσειρών χρησιμοποιούνται επιπρόσθετα σε κάποιο μοντέλο πρόβλεψης. Τα μοτίβα μίας χρονοσειράς μπορούν να διακριθούν σε τέσσερα κύρια είδη, τα οποία ονομάζονται συνιστώσες της χρονοσειράς. Αυτές είναι οι συνιστώσες της τάσης, της εποχικότητας, της κυκλικότητας και της τυχαιότητας. Η διάσπαση χρονοσειρών (time series decomposition) στηρίζεται στην υπόθεση ότι οι τιμές μίας χρονοσειράς σχηματίζονται από τις παραπάνω συνιστώσες που τη συνθέτουν. Για τη δημιουργία των προβλέψεων με τη μέθοδο αυτή, η χρονοσειρά διασπάται στις ανωτέρω τέσσερις συνιστώσες και προσδιορίζεται η επιρροή που έχει καθεμία από αυτές στη διαμόρφωση των τιμών της μεταβλητής.

$$Y_t = T_t + S_t + C_t + I_t \quad (2.44)$$

όπου T_t είναι η συνιστώσα της τάσης, S_t είναι η εποχική συνιστώσα, C_t η συνιστώσα της κυκλικότητας και I_t είναι η συνιστώσα των υπολοίπων.



Εικόνα 2.4: Οι συνιστώσες μίας χρονοσειράς όπως προκύπτουν από τη μέθοδο διάσπασης χρονοσειρών [47].

Υπάρχουν διάφορες τεχνικές για να επιτευχθεί μία τέτοια αποσύνθεση [47]. Μία απλή μέθοδος που δείχνει πως η αποσύνθεση της χρονοσειράς μπορεί να βοηθήσει στην πρόβλεψη των μελλοντικών τιμών της είναι η εξής:

- i. Αποσύνθεση της χρονοσειράς με κάποια τεχνική αποσύνθεσης (όπως STL [18]): $Y_t = T_t + S_t + I_t$.

- ii. Υπολογισμός της συνιστώσας εποχικότητας $Y_t - S_t$. Χρήση ενός μοντέλου για την πρόβλεψη της συνιστώσας εποχικότητας.
- iii. Προσθήκη των προβλέψεων της εποχικής συνιστώσας στις προβλέψεις της χρονοσειράς της τάσης της χρονοσειράς.

2.4.11 Μέθοδος (S)ARIMA

Τα ολοκληρωμένα αυτοπαλίνδρομα υποδείγματα κινητού μέσου όρου (Autoregressive Integrated Moving Average) είναι στοχαστικά μαθηματικά μοντέλα τα οποία χρησιμοποιούνται για την περιγραφή, ανάλυση και πρόβλεψη της διαχρονικής εξέλιξης κάποιου φυσικού μεγέθους που εξαρτάται από μη ντετερμινιστικούς παράγοντες. Χρησιμοποιούνται για προβλέψεις των μελλοντικών τιμών, λαμβάνοντας υπόψη τις παρελθοντικές τιμές και με ικανότητα μοντελοποίησης μεγάλου εύρους δεδομένων εποχικότητας. Για να εφαρμοστούν σε μία χρονοσειρά, πρέπει να είναι στάσιμη και διακριτή. Τα στοχαστικά μοντέλα περιλαμβάνουν τον τυχαίο παράγοντα (τυχαίο σφάλμα ή σφάλμα πρόβλεψης), τις τιμές του μεγέθους οι οποίες εμφανίζονται σε προηγούμενες χρονικές στιγμές και ίσως κάποιους άλλους στοχαστικούς παράγοντες. Το υπόδειγμα που προκύπτει είναι ένας γραμμικός συνδυασμός των παραπάνω ποσοτήτων ενώ ο εντοπισμός των τιμών των τριών αυτών παραγόντων μπορεί να προσεγγιστεί αποτελεσματικά. Οι Box-Jenkins [102] πρότειναν μία οικογένεια αλγεβρικών υποδειγμάτων πρόβλεψης από τα οποία επιλέγεται το πιο κατάλληλο για την πρόβλεψη μίας χρονοσειράς. Σε αυτά τα υποδείγματα, οι προβλέψεις στηρίζονται στις προηγούμενες χρονικές τιμές της χρονοσειράς και στα πρότυπα συμπεριφοράς που εμφανίζει η χρονοσειρά. Τα μοντέλα πολλαπλής γραμμικής παλινδρόμησης περιγράφονται από την εξίσωση:

$$Y = b_0 + b_1X_1 + b_2X_2 + \dots + b_pX_p + e \quad (2.45)$$

όπου Y η εξαρτημένη μεταβλητή και $X_1, X_2, \dots, X_p$ οι ανεξάρτητες μεταβλητές. Η εξίσωση:

$$Y_t = b_0 + b_1Y_{t-1} + b_2Y_{t-2} + \dots + b_pY_{t-p} + e_t \quad (2.46)$$

είναι μία εξίσωση πολλαπλής παλινδρόμησης με τη διαφορά ότι οι ανεξάρτητες μεταβλητές είναι τιμές της χρονοσειράς σε προηγούμενες χρονικές περιόδους. Το παραπάνω υπόδειγμα ονομάζεται αυτοπαλίνδρομο υπόδειγμα (AutoRegression model) και συμβολίζεται $AR(p)$. Εδώ θεωρείται πως η μία τιμή της χρονοσειράς είναι εξαρτώμενη από τις προηγούμενες τιμές με κάποια καθυστέρηση (lag) p . Συνεπώς τα υποδείγματα AR εκφράζουν το Y_t σαν γραμμική συνάρτηση των p παρελθουσών πραγματικών τιμών του Y_t . Επίσης η εξίσωση:

$$Y_t = b_0 + e_t - b_1e_{t-1} - b_2e_{t-2} - \dots - b_qe_{t-q} \quad (2.47)$$

είναι μία εξίσωση πολλαπλής παλινδρόμησης με τη διαφορά ότι οι ανεξάρτητες μεταβλητές είναι τιμές των σφαλμάτων προηγούμενων χρονικών περιόδων (διαφορά προβλεφθείσας τιμής

από την πραγματική). Το παραπάνω υπόδειγμα ονομάζεται υπόδειγμα κινητού μέσου όρου (Moving Average) και συμβολίζεται $MA(q)$. Συνεπώς τα υποδείγματα MA εκφράζουν το Y_t σαν γραμμικό συνδυασμό παρελθόντων σφαλμάτων πρόβλεψης. Τα αυτοπαλίνδρομα υποδείγματα $AR(p)$ μπορούν να συνδυαστούν αποτελεσματικά με τα υποδείγματα κινητού μέσου όρου $MA(q)$ και να σχηματίσουν μία χρήσιμη ομάδα υποδειγμάτων χρονοσειρών τα οποία ονομάζονται αυτοπαλίνδρομα υποδείγματα κινητού μέσου όρου (AutoRegressive Moving Average models) και συμβολίζονται $ARMA(p,q)$. Τα υποδείγματα αυτά μπορούν να χρησιμοποιηθούν μόνο για σταθερές (στάσιμες) χρονοσειρές. Ωστόσο μπορούν να επεκταθούν και σε μη σταθερές (μη στάσιμες) χρονοσειρές με τη χρήση της μεθόδου της διαφορίσης. Σε αυτήν την περίπτωση ονομάζονται ολοκληρωμένα αυτοπαλίνδρομα μοντέλα κινητού μέσου όρου (AutoRegressive Integrated Moving Average models) και συμβολίζονται ως $ARIMA(p,d,q)$ όπου p είναι η τάξη του αυτοπαλίνδρομου υποδείγματος, d η τάξη της διαφορίσης για την επίτευξη σταθερότητας (μετατροπή της χρονοσειράς σε στάσιμη) και q η τάξη του κινητού μέσου όρου υποδείγματος. Το μοντέλο λευκού θορύβου (white noise) εκφράζεται ως $ARIMA(0,0,0)$ ενώ το μοντέλο τυχαίου περιπάτου ως $ARIMA(0,1,0)$. Τα συγκεκριμένα μοντέλα έχουν μελετηθεί εκτεταμένα από τους Box και Jenkins [102], οι οποίοι έχουν προτείνει μία μεγάλη ομάδα αλγεβρικών μοντέλων πρόβλεψης, όπου αναλόγως της χρονοσειράς εφαρμόζεται και το κατάλληλο μοντέλο ώστε να βελτιστοποιείται η τελική πρόβλεψη. Η μέθοδος που ακολουθείται για την ανάλυση χρονοσειρών με τη μέθοδο $ARIMA$ περιλαμβάνει τέσσερα στάδια:

- i. Ταυτοποίηση υποδείγματος
- ii. Εκτίμηση υποδείγματος
- iii. Διαγνωστικός έλεγχος υποδείγματος
- iv. Πρόβλεψη χρονοσειράς

Η ταυτοποίηση του υποδείγματος πραγματοποιείται με την εξέταση των αυτοσυσχετίσεων και των μερικών αυτοσυσχετίσεων των ακατέργαστων δεδομένων. Μέσω της διαδικασίας αυτής ελέγχουμε αν η χρονοσειρά είναι τυχαία, στάσιμη, αν εμφανίζει κάποιο πρότυπο τάσης ή κάποιο εποχικό πρότυπο. Εάν η χρονοσειρά δεν είναι στάσιμη τότε η χρονοσειρά διαφορίζεται μέχρις ότου να αποκτήσει σταθερότητα και στη συνέχεια παράγονται νέα διαγράμματα αυτοσυσχετίσεων και μερικών αυτοσυσχετίσεων. Από αυτά τα διαγράμματα δείχνουν εάν υπάρχει κάποιο πρότυπο τάσης ή κάποιο εποχικό μοτίβο. Μέσα από αυτήν τη διαδικασία παράγεται ένα δοκιμαστικό υπόδειγμα $ARIMA$. Ουσιαστικά σε αυτό το στάδιο περιλαμβάνεται ο καθορισμός του αριθμού d (η τάξη διαφορίσης που χρησιμοποιείται για τη μετατροπή μίας διαδικασίας σε στάσιμη σε περίπτωση που δεν είναι), του αριθμού p (τάξη της αυτοπαλίνδρομης διαδικασίας AR) και του αριθμού q (τάξη της διαδικασίας κινητού μέσου MA). Η εκτίμηση του υποδείγματος περιλαμβάνει την εκτίμηση των παραμέτρων

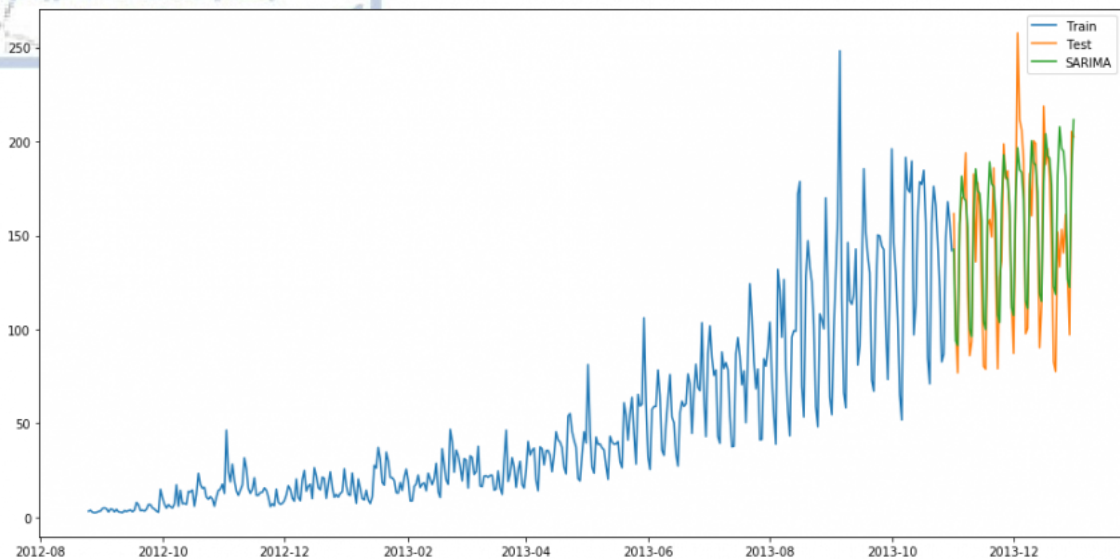
φ , θ που ορίζουν το υπόδειγμα ARIMA οι οποίοι ελαχιστοποιούν το MSE. Η διαδικασία επαναλαμβάνεται όπου οι τελικές τιμές των εκτιμήσεων των παραμέτρων είναι σημαντικά διαφορετικές από τις αρχικές τιμές της διαδικασίας εκτίμησης. Συνήθως οι εκτιμήσεις συγκλίνουν σε μία άριστη τιμή για τις παραμέτρους μετά από σχετικά μικρό αριθμό επαναλήψεων. Σε αυτό το στάδιο πραγματοποιείται ο καθορισμός των p παραμέτρων της αυτοπαλίνδρομης διαδικασίας και των q παραμέτρων της διαδικασίας κινητού μέσου. Επίσης εκτελείται έλεγχος της τάξεως συγκρίνοντας το υπόδειγμα ARIMA με άλλο μεγαλύτερης τάξεως (κριτήρια AIC, SBC). Ο διαγνωστικός έλεγχος του υποδείγματος πραγματοποιείται για τον έλεγχο της καταλληλότητας του υποδείγματος. Για να ελέγξουμε την καταλληλότητα του υποδείγματος κατασκευάζουμε το διάγραμμα των αυτοσυσχετίσεων των καταλοίπων της χρονοσειράς. Εάν οι συντελεστές αυτοσυσχέτισης δεν είναι στατιστικά σημαντικοί (βρίσκονται μέσα στα όρια των δύο σίγμα) τότε το υπόδειγμα κρίνεται επαρκές. Ένας δεύτερος έλεγχος για την επάρκεια του υποδείγματος είναι ο έλεγχος Ljung-Box (Q test). Το κριτήριο ελέγχου είναι:

$$Q = n(n+2) \sum_i^k \frac{r_i^2}{n-k} \quad (2.48)$$

όπου n είναι το μήκος της χρονοσειράς, k είναι οι πρώτες k αυτοσυσχετίσεις που έχουν ελεγχθεί, p είναι η τάξη της διαδικασίας AR, q είναι η τάξη της διαδικασίας MA και r είναι ο εκτιμημένος συντελεστής αυτοσυσχέτισης του i -οστού κατάλοιπου όρου. Εάν η υπολογισθείσα τιμή του Q είναι μικρότερη από το X^2 για $k-p-q$ βαθμούς ελευθερίας τότε το υπόδειγμα θεωρείται επαρκές. Ένα κριτήριο για την επιλογή του καλύτερου υποδείγματος είναι το Akaike's Information Criterion ή AIC. Αν $m = p + q + P + Q$ είναι το πλήθος των όρων τότε επιλέγουμε τις τιμές p, q, P, Q που ελαχιστοποιούν το AIC. Χρησιμοποιούνται επίσης παρόμοια κριτήρια όπως είναι τα BIC (Bayesian Information Criterion) και FPE (Final Prediction Error). Αφού επιλέξουμε το καλύτερο υπόδειγμα που περιγράφει τη χρονοσειρά το χρησιμοποιούμε για την πρόβλεψη των μελλοντικών τιμών της χρονοσειράς. Όλες οι μελλοντικές προβλέψεις της χρονοσειράς βασίζονται μόνο σε τιμές του Y_t που έχουν προβλεφθεί εφόσον οι μελλοντικές τιμές του e_t είναι άγνωστες. Εάν το υπόδειγμα περιγράφει με ορθό τρόπο τη χρονοσειρά και δίνει προβλέψεις οι οποίες αποκλίνουν ελάχιστα από τις πραγματικές τιμές της χρονοσειράς τότε υιοθετείται ως υπόδειγμα πρόβλεψης της χρονοσειράς. Σε αντίθετη περίπτωση επαναλαμβάνεται η ίδια διαδικασία από την αρχή. Στη συνέχεια ακολουθούν κάποιες εφαρμογές ανάλυσης χρονοσειρών με τη μέθοδο ARIMA.

2.4.12 Μέθοδος TBATS

Η μέθοδος TBATS είναι ένα μοντέλο βασισμένο στη μέθοδο της εκθετικής εξομάλυνσης. Χρησιμοποιεί έναν συνδυασμό μεταξύ όρων Fourier, το μοντέλο χώρου καταστάσεων των μεθόδων εκθετικής εξομάλυνσης και τον μετασχηματισμό Box-Cox με έναν τελείως



Εικόνα 2.5: Παράδειγμα πρόβλεψης χρονοσειράς με τη μέθοδο SARIMA [47].

αυτοματοποιημένο τρόπο. Ο μετασχηματισμός Box-Cox δημιουργήθηκε από τους Box και Cox [9] και είναι ικανός να αντιμετωπίσει μη γραμμικά δεδομένα. Σε συνδυασμό με το μοντέλο ARMA (AutoRegressive Moving Average model), μπορούν να δημιουργήσουν ανεξαρτησία μεταξύ των δεδομένων μίας χρονοσειράς. Έπειτα, το 2011 οι De Livera, Hyndman, Snyder [63] παρουσίασαν μία εναλλακτική μέθοδο, ώστε να παράγουν προβλέψεις για πολύπλοκες εποχικές χρονοσειρές και απέδειξαν ότι είναι ιδιαίτερα αποτελεσματική. Αυτή η μέθοδος είναι γνωστή ως BATS (exponential smoothing state space model with Box-Cox transformation, ARMA errors, Trend and Seasonal components). Ωστόσο, η μέθοδος BATS δεν ήταν αποδοτική όταν η εποχικότητα είχε μεγαλύτερη συχνότητα και ήταν ιδιαίτερα πολύπλοκη. Οπότε το 2011, προτάθηκε η μέθοδος TBATS (Trigonometric exponential smoothing state space model with Box-Cox transformation, ARMA errors, Trend and Seasonal components), όπου η τριγωνομετρική έκφραση της εποχικότητας μπορεί να δώσει στη χρονοσειρά την ευελιξία να αντιμετωπίζει σύνθετες εποχικότητες και επίσης μειώνει τις παραμέτρους του μοντέλου όταν η συχνότητα είναι πολύ υψηλή. Επιπροσθέτως, η TBATS είναι ικανή να χειρίζεται μη ακέραιες εποχικές συχνότητες και τα ληφθέντα δεδομένα δε χρειάζονται κανονικοποίηση, συγκριτικά με την BATS. Οι παράμετροι εφαρμογής της μεθόδου είναι οι $\omega, p, q, \phi, (m_1, k_1) \dots (m_j, k_j)$, όπου ω είναι η παράμετρος Box-Cox, ϕ η παράμετρος απόσβεσης (damping parameter), p, q είναι οι παράμετροι του ARMA, $m_1 \dots m_j$ οι εποχικές περίοδοι που χρησιμοποιούνται στο μοντέλο και $k_1 \dots k_j$ οι αντίστοιχοι όροι Fourier.

2.4.13 Μέθοδος Prophet

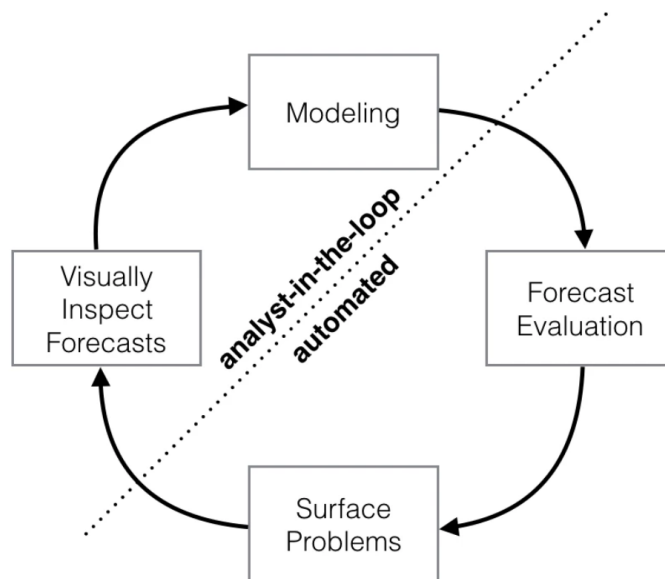
Η μέθοδος Prophet [88] δημιουργήθηκε από την ομάδα FacebookTM Core Data Science (Facebook Inc., Menlo Park, US-CA), είναι μία μέθοδος πρόβλεψης που επιτρέπει πολλαπλές

εποχικότητες και έχει υλοποιηθεί σε πακέτα στις γλώσσες R [89] και Python. Η παρατήρηση ότι η παραγωγή αξιόπιστων προβλέψεων στον επαγγελματικό τομέα απαιτεί εξειδικευμένο και έμπειρο προσωπικό, οδήγησε τους συγγραφείς στη δημιουργία της μεθόδου. Το κίνητρό τους ήταν να κάνουν εφικτή την πρόβλεψη χρονοσειρών στον επαγγελματικό κόσμο από προσωπικό, όχι απαραίτητα εξειδικευμένο στην ανάλυση και πρόβλεψη χρονοσειρών. Η μέθοδος Prophet επιτυγχάνει να προβλέψει με επιτυχία τις εποχικές επιδράσεις που προκαλούνται από την ανθρώπινη συμπεριφορά (για παράδειγμα εβδομαδιαίοι, μηνιαίοι και ετήσιοι κύκλοι, δημόσιες αργίες και άλλες ειδικές συνθήκες). Επιπλέον, μπορεί να ανιχνεύσει αλλαγές στην τάση που προκαλούνται από δραστηριότητα της αγοράς (όπως η εισαγωγή ενός νέου προϊόντος ή ενός νέου καταστήματος). Ένα πλεονέκτημα της μεθόδου είναι και η διαχείριση των ακραίων και των τιμών που πιθανώς απουσιάζουν από τις μετρήσεις της χρονοσειράς. Οι συγγραφείς υποστηρίζουν ότι ακόμα και με τις προεπιλεγμένες ρυθμίσεις του πακέτου, παράγονται προβλέψεις μεγάλης ακρίβειας. Όπως και τα αθροιστικά γενικευμένα μοντέλα (GAM), το Prophet προσπαθεί να μοντελοποιήσει τη χρονοσειρά με διάφορες γραμμικές και μη γραμμικές συναρτήσεις του χρόνου. Η εποχικότητα του μοντέλου ως συστατικό πρόσθετης ύλης είναι η ίδια η προσέγγιση που λαμβάνεται με την εκθετική εξομάλυνση. Η διαμόρφωση GAM έχει το πλεονέκτημα ότι αποσυντίθεται εύκολα και μπορεί να υιοθετηθεί νέους παράγοντες αν χρειαστεί, όπως για παράδειγμα όταν αναγνωριστεί μία νέα πηγή εποχικότητας.

Το πακέτο Prophet έχει αρκετές εύκολα ερμηνεύσιμες παραμέτρους που επιτρέπουν τη βελτίωση της ποιότητας του μοντέλου πρόβλεψης. Οι παράμετροι αυτοί είναι κατανοητοί ακόμα και από μη ειδικό προσωπικό. Το σχετικό άρθρο [88], τιτλοφορείται ως Πρόβλεψη σε Κλίμακα ("Forecasting at Scale"), χωρίς ωστόσο ο όρος "σε κλίμακα" να αναφέρεται, ως συνήθως, στην αντιμετώπιση υπολογιστικών προβλημάτων ή προβλημάτων υποδομής. Σύμφωνα με τους συγγραφείς το Prophet μπορεί να κλιμακωθεί σε τρεις τομείς:

- i. Προσβασιμότητα σε ένα ευρύ ακροατήριο αναλυτών χωρίς εξειδίκευση στις χρονοσειρές.
- ii. Εφαρμογή σε ένα ευρύ φάσμα διαφόρων προβλημάτων πρόβλεψης.
- iii. Αυτόματη αξιολόγηση επιδόσεων για ένα μεγάλο αριθμό προβλέψεων και επισήμανση προβλημάτων της ανάλυσης, εάν αυτά υπάρχουν.

Συγκεκριμένα, το Prophet εισάγει την έννοια του "analyst-in-the-loop" σύμφωνα με την οποία το Prophet ξεκινά με τη μοντελοποίηση μίας χρονοσειράς χρησιμοποιώντας τις παραμέτρους του αναλυτή, παράγει προβλέψεις και στη συνέχεια τις αξιολογεί. Αν ανιχνεύσει κάποιο πρόβλημα στη διαδικασία ή κακή απόδοση το Prophet φέρνει στην επιφάνεια το πρόβλημα ενημερώνοντας τον αναλυτή ώστε να προβεί σε αλλαγές στη ρύθμιση του μοντέλου σύμφωνα με τις αξιολογήσεις.



Εικόνα 2.6: Η Analyst-in-the-loop προσέγγιση του Prophet [89].

Στον πυρήνα του το Prophet είναι ένα αθροιστικό μοντέλο (additive regression model) και η διαδικασία χρησιμοποιεί ένα μοντέλο αποδιαμόρφωσης χρονοσειράς με τρία βασικά στοιχεία του μοντέλου: τάση, εποχικότητα και διακοπές.

Στην απλούστερη μορφή του:

$$y(t) = g(t) + s(t) + h(t) + \epsilon(t) \quad (2.49)$$

Οι όροι $g(t)$, $s(t)$ και $h(t)$ αντιστοιχούν στην τάση, την εποχικότητα και τις αργίες. Ο τελευταίος $\epsilon(t)$ είναι ο όρος των σφαλμάτων. Η εφαρμογή του μοντέλου προσεγγίζεται ως μία άσκηση προσαρμογής καμπύλης και επομένως δεν λαμβάνει ρητά υπόψη τη δομή της χρονικής εξάρτησης στα δεδομένα. Αυτό επιτρέπει επίσης να υπάρχουν παρατηρήσεις με ακανόνιστη απόσταση μεταξύ τους.

Υπάρχουν δύο επιλογές για τη μοντελοποίηση της τάσης. Η μία επιλογή χρησιμοποιεί ένα μη γραμμικό, αυξητικό μοντέλο με κορεσμό (Saturating Growth model), ενώ η άλλη χρησιμοποιεί ένα γραμμικό μοντέλο (Piecewise Linear model). Η μοντελοποίηση της εποχικότητας βασίζεται σε σειρές Fourier.

Τάση. Το $g(t)$ μοντελοποιεί την τάση που περιγράφει μακροχρόνιες αυξήσεις ή μειώσεις των τιμών στα δεδομένα. Το Prophet, όπως και αναφέρθηκε, προσφέρει δύο μοντέλα για τη μοντελοποίηση της τάσης. Η επιλογή του ενός ή του άλλου εξαρτάται από τον τύπο του προβλήματος πρόβλεψης.

Το μη γραμμικό, αυξητικό μοντέλο με κορεσμό εκφράζεται στη μορφή του λογιστικού αυξητικού μοντέλου: [64]:

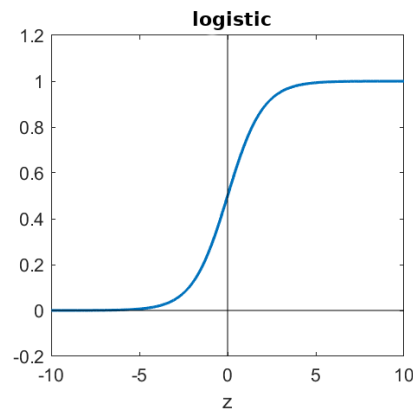
$$g(t) = \frac{C}{1 + e^{-k(t-m)}} \quad (2.50)$$

όπου

C είναι η φέρουσα χωρητικότητα (η μέγιστη τιμή της καμπύλης)

k είναι ο ρυθμός αύξησης (που εκφράζει το πόσο απότομη είναι η καμπύλη)

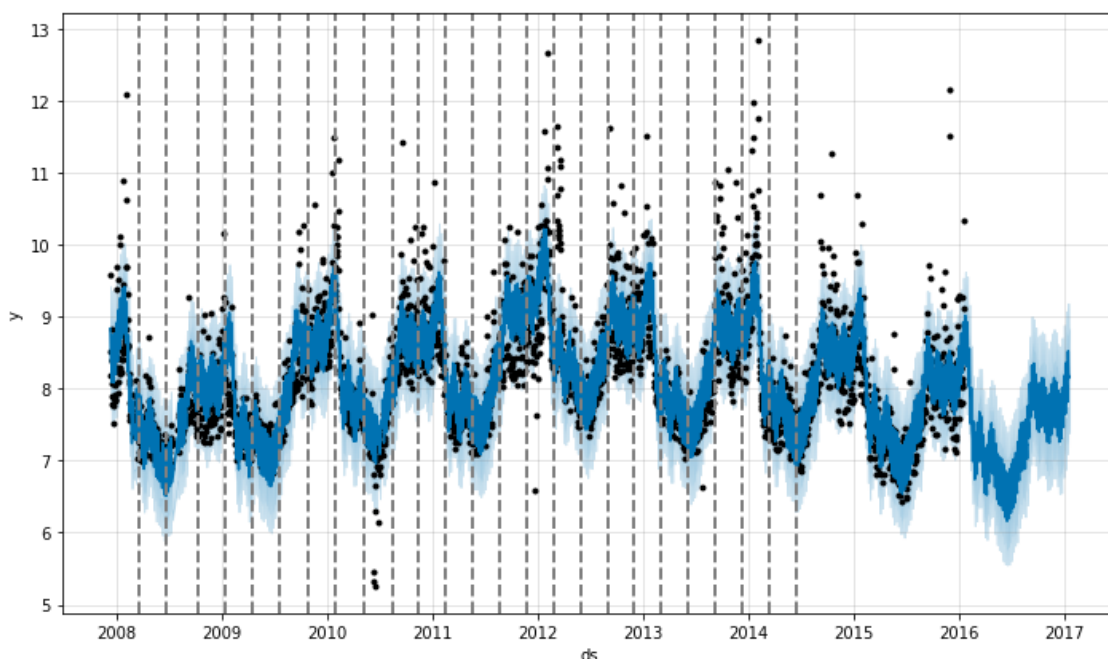
m είναι μία παράμετρος αντιστάθμισης (η x -τιμή του σιγμοειδούς σημείου)



Εικόνα 2.7: Λογιστική συνάρτηση ή λογιστική καμπύλη [64]

Η λογιστική συνάρτηση επιτρέπει τη μη γραμμική μοντελοποίηση της αύξησης των τιμών της χρονοσειράς σε συνάρτηση με τον κορεσμό που επέρχεται όσο προσεγγίζεται η μέγιστη τιμή, δηλαδή όταν ο ρυθμός αύξησης των τιμών μειώνεται όσο αυξάνονται οι τιμές. Στην πραγματικότητα οι παράμετροι C και k δεν είναι απαραίτητα σταθεροί και μπορούν να μεταβάλλονται στον χρόνο. Το πακέτο Prophet υποστηρίζει τόσο την αυτόματη όσο και τη μη αυτόματη ρύθμισή τους αν αυτό κριθεί αναγκαίο. Το πακέτο μπορεί να επιλέξει υποψήφια σημεία αλλαγών της τάσης μοντελοποιώντας τα παρεχόμενα ιστορικά δεδομένα. Από προεπιλογή το Prophet χρησιμοποιεί ένα γραμμικό μοντέλο για την πρόβλεψη της τάσης. Κατά την πρόβλεψη της τάσης, υπάρχει συνήθως κάποιο μέγιστο επιτεύξιμο σημείο, όπως το συνολικό μέγεθος της αγοράς ή το συνολικό μέγεθος του πληθυσμού. Αυτό ονομάζεται φέρουσα ικανότητα ή φέρουσα χωρητικότητα και η πρόβλεψη επιδέχεται κορεσμό σε αυτό το σημείο. Το Prophet επιτρέπει επίσης τη δημιουργία προβλέψεων της τάσης με τη χρήση ενός λογιστικού αυξητικού μοντέλου, με συγκεκριμένη φέρουσα ικανότητα. Η λογιστική συνάρτηση έχει ένα νοητό ελάχιστο ίσο με το 0 και η πρόβλεψη θα κορεστεί στο 0 με τον ίδιο τρόπο που υφίσταται τον κορεσμό στη φέρουσα χωρητικότητα. Είναι επίσης δυνατό να καθοριστεί ένα διαφορετικό ελάχιστο κορεσμού. Επίσης, το Prophet επιτρέπει στους αναλυτές να ορίσουν με μη αυτόματο τρόπο τα σημεία αλλαγής τάσης και τις τιμές της φέρουσας χωρητικότητας για διαφορετικά χρονικά σημεία. Για παράδειγμα, οι αναλυτές ενδέχεται να έχουν γνώση σχετικά με ημερομηνίες προηγούμενων γεγονότων που επηρέασαν

εμφανώς μερικούς βασικούς μετρικούς δείκτες. Ωστόσο, από προεπιλογή, το Prophet θα εντοπίζει αυτόματα αυτά τα σημεία αλλαγής και θα επιτρέψει στην τάση να προσαρμοστεί κατάλληλα. Οι κατακόρυφες γραμμές στο σχήμα 2.8 δείχνουν πού έχουν τοποθετηθεί τα πιθανά σημεία αλλαγής.



Εικόνα 2.8: Προεπιλεγμένα πιθανά σημεία αλλαγής τάσης της μεθόδου Prophet [27].

Εάν η διαμέριση και τα σημεία αλλαγής είναι γνωστά, τότε εφαρμόζεται η γραμμική παλινδρόμηση ανεξάρτητα σε κάθε ένα κομμάτι της διαμέρισης. Εάν η διαμέριση δεν είναι γνωστή, τότε χρησιμοποιείται το άθροισμα των τετραγωνικών εκτιμήσεων των σφαλμάτων (Sum of Squares Estimate of Errors [78]) για τη βέλτιστη επιλογή των σημείων αλλαγής.

Το δεύτερο μοντέλο για την προσαρμογή της τάσης είναι απλό γραμμικό μοντέλο [72] με σταθερό ρυθμό αύξησης. Χρησιμοποιείται για χρονοσειρές που δεν παρουσιάζουν κορεσμό στον ρυθμό αύξησης των τιμών τους.

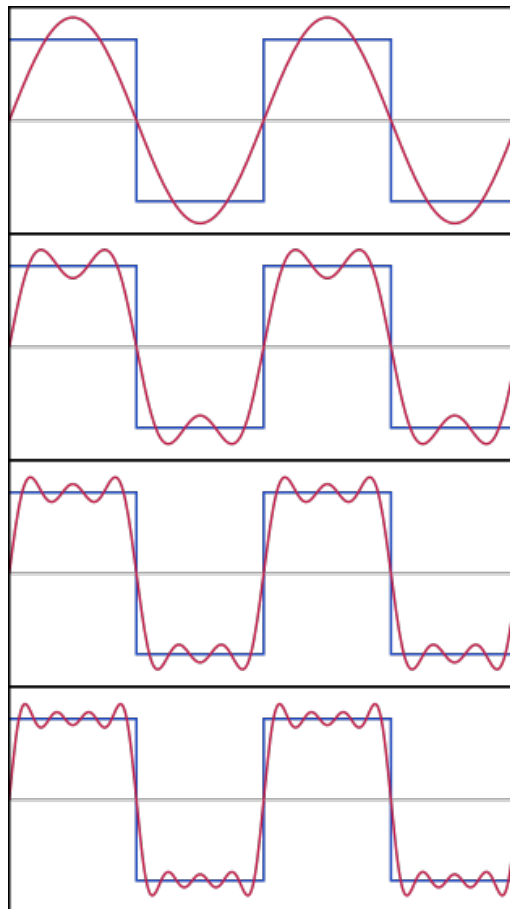
Εποχικότητα. Η συνιστώσα της εποχικότητας $s(t)$ εκφράζει περιοδικές αλλαγές εξαιτίας εβδομαδιαίων ή/και ετήσιων εποχικών κύκλων και μοντελοποιείται με το μερικό άθροισμα σειρών Fourier. Στην παρακάτω εικόνα μπορεί να γίνει αντιληπτό πως ένα μερικό άθροισμα Fourier μπορεί να προσεγγίσει ένα περιοδικό σήμα.

Οι επιδράσεις της εποχικότητας προσεγγίζονται από την ακόλουθη συνάρτηση:

$$s(t) = \sum_{n=1}^N \left(\alpha_n \cos\left(\frac{2\pi n t}{P}\right) + \beta_n \sin\left(\frac{2\pi n t}{P}\right) \right) \quad (2.51)$$

όπου,

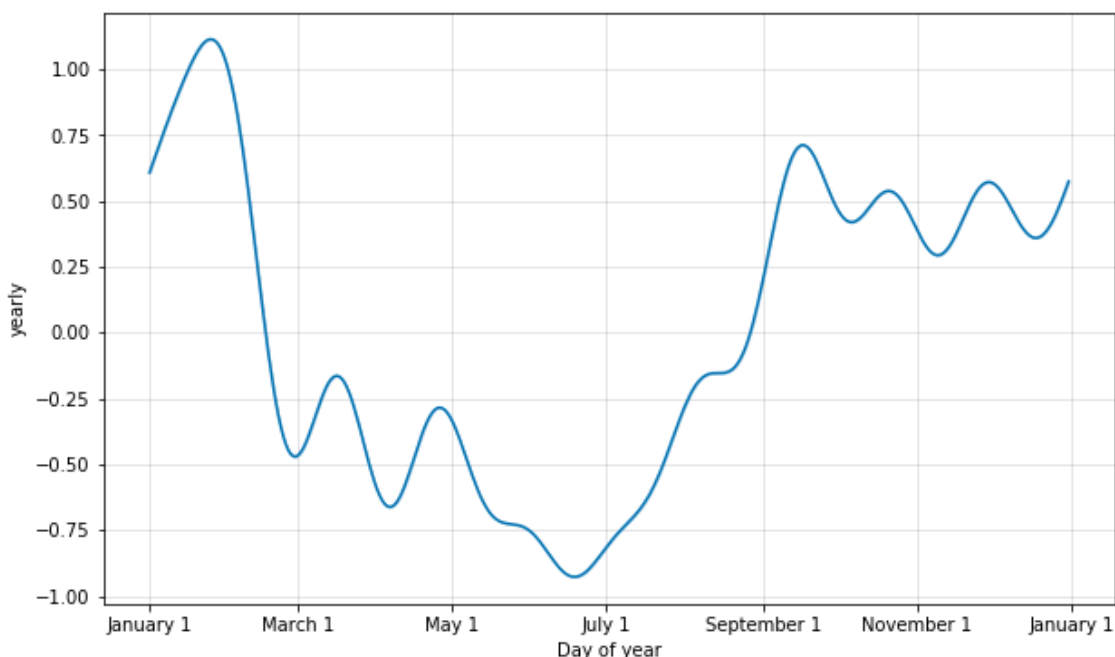
P είναι η περίοδος (365.25 για ετήσια δεδομένα και 7 για εβδομαδιαία δεδομένα).



Εικόνα 2.9: Τα πρώτα τέσσερα αθροίσματα της σειράς Fourier για την προσέγγιση ενός τετραγωνικού κύματος.

Οι παράμετροι $[(\alpha_1, \beta_1) \dots (\alpha_N, \beta_N)]$ πρέπει να εκτιμηθούν για ένα δεδομένο N ώστε να μοντελοποιήσουν την εποχικότητα. Η τάξη της σειράς Fourier N , καθορίζει το αν οι αλλαγές με υψηλή συχνότητα επιτρέπεται να συμπεριληφθούν στο μοντέλο. Είναι μία παράμετρος δηλαδή που ρυθμίζει το πόσο γρήγορες μπορούν να είναι οι αλλαγές της εποχικότητας. Για μία χρονοσειρά που ο αναλυτής θεωρεί ότι οι διακυμάνσεις υψηλής συχνότητας είναι απλά θόρυβος και δεν θα πρέπει να περιληφθούν στη μοντελοποίηση, αρκεί μία χαμηλή τιμή για την παράμετρο N . Εάν όχι, τότε μία υψηλή τιμή της παραμέτρου N , θα προσπαθήσει να μοντελοποιήσει και τις διακυμάνσεις αυτές.

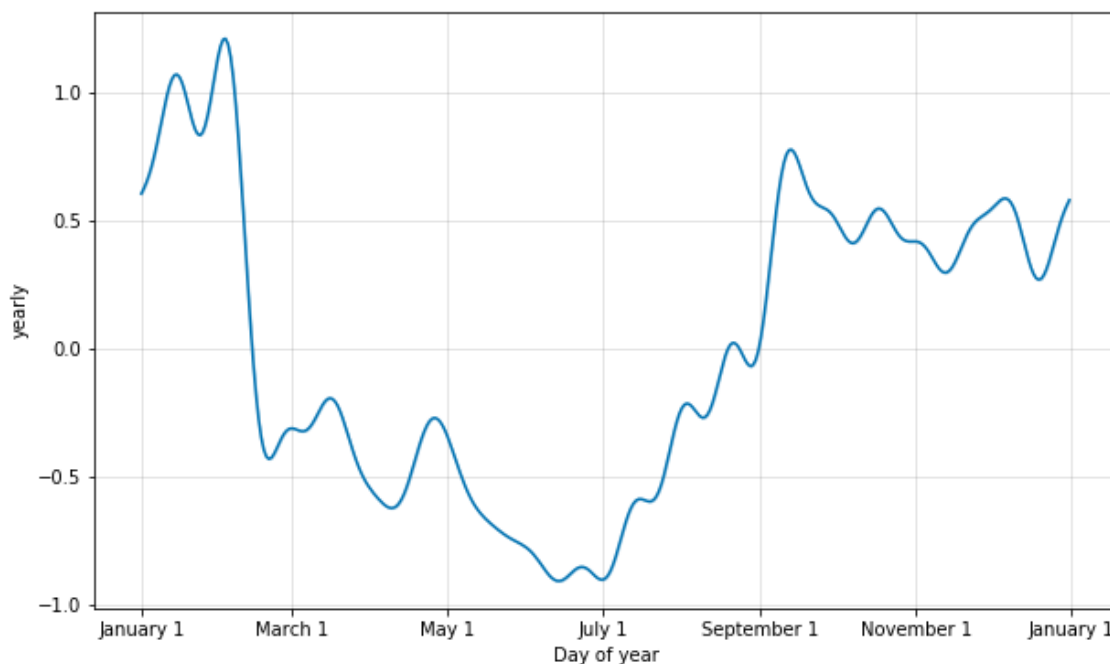
Η προεπιλεγμένη τάξη Fourier, που χρησιμοποιεί το Prophet για ετήσιες εποχικότητες είναι 10, όπου παράγεται η προσαρμογή που απεικονίζεται στην Εικόνα 2.10. Όταν η εποχικότητα είναι λιγότερο "ομαλή" και περιλαμβάνει αλλαγές υψηλής συχνότητας, η αύξηση της τάξης N σε 20, προκαλεί την προσαρμογή που φαίνεται στην Εικόνα 2.11. Η αύξηση του αριθμού τάξης Fourier, επιτρέπει τη μοντελοποίηση συντομότερων κύκλων εποχικότητας, ωστόσο μπορεί να οδηγήσει σε υπερτροφοδότηση του μοντέλου.



Εικόνα 2.10: Προσαρμογή εποχικότητας για τη δοθείσα χρονοσειρά με τάξη Fourier $N = 10$ στη μέθοδο Prophet.

Το Prophet, από προεπιλογή χρησιμοποιεί το αθροιστικό μοντέλο για τη μοντελοποίηση των εποχικοτήτων, δηλαδή η επίδραση της εποχικότητας αθροίζεται στην τάση για να εξαχθεί η πρόβλεψη. Εάν η εποχικότητα δεν είναι ένας σταθερός αθροιστικός παράγοντας, αλλά μεγαλώνει με την τάση, τότε πρέπει να χρησιμοποιηθεί το πολλαπλασιαστικό μοντέλο. Το Prophet δίνει επίσης τη δυνατότητα στον χρήστη να ορίσει εξατομικευμένες εποχικότητες.

Αργίες και συμβάντα. Η συνιστώσα $h(t)$ μοντελοποιεί την επίδραση των αργιών,



Εικόνα 2.11: Προσαρμογή εποχικότητας για τη δοθείσα χρονοσειρά με τάξη Fourier $N = 20$ στη μέθοδο Prophet.

εορτών, σημαντικών ημερομηνιών, μεγάλων γεγονότων, δηλαδή προβλέψιμες ακανόνιστες ημέρες του χρόνου όπου συνήθως παρατηρούνται αιχμές. Για την εκμετάλλευση αυτής της συνιστώσας, μία λίστα με τις ημερομηνίες αυτές (παρελθοντικές και μελλοντικές) πρέπει να παρασχεθούν από τον χρήστη. Σε αυτήν την περίπτωση το Prophet δημιουργεί ένα “παράθυρο” γύρω από αυτές τις μέρες που εξετάζεται και μοντελοποιείται ξεχωριστά υπολογίζοντας ειδικές παραμέτρους για αυτές τις μέρες.

Σφάλματα. Ο όρος σφάλματος ϵ αντανακλά την πληροφορία που δεν μπορεί να συμπεριληφθεί στο μοντέλο. Συνήθως μοντελοποιείται ως θόρυβος με κανονική κατανομή (κανονικός θόρυβος).

Διαστήματα Αβεβαιότητας. Το Prophet από προεπιλογή επιστρέφει διαστήματα αβεβαιότητας για την πρόβλεψη $\hat{Y}$. Πίσω από τα διαστήματα αβεβαιότητας υπάρχουν διάφορες σημαντικές εικασίες. Υπάρχουν τρεις σειρές αβεβαιότητας στην πρόβλεψη, αβεβαιότητα στην τάση, αβεβαιότητα στην εποχικότητα και ο επιπλέον θόρυβος παρατήρησης. Η μεγαλύτερη πηγή αβεβαιότητας για την πρόβλεψη είναι οι πιθανές μελλοντικές αλλαγές στην τάση. Το Prophet είναι ικανό να τις ανιχνεύσει και να τις προσαρμόσει. Είναι αδύνατο βέβαια να γνωρίζει κανείς εκ των προτέρων τι αλλαγές τάσης θα συμβούν στο μέλλον, έτσι η πιο λογική αντιμετώπιση είναι να υποτεθεί ότι το μέλλον θα έχει παρόμοιες αλλαγές τάσης με το παρελθόν. Συγκεκριμένα, το Prophet υποθέτει ότι η μέση συχνότητα και σημαντικότητα αλλαγών της τάσης στο μέλλον, θα είναι όμοιες με τις αλλαγές που παρατηρούνται στο παρελθόν.

Η αβεβαιότητα στην πρόβλεψη της τάσης εκτιμάται σύμφωνα με την παραδοχή ότι υπάρχουν S σημεία αλλαγής σε ένα ιστορικό T σημείων, κάθε ένα από τα οποία έχει ρυθμό αλλαγής $\delta_j \sim \text{Laplace}(0, \tau)$. Το τ (rate scale parameter) υπολογίζεται ως εκτίμηση της μέγιστης πιθανοφάνειάς του:

$$\lambda = \frac{1}{S} \sum_{j=1}^S |\delta_j| \quad (2.52)$$

Τα μελλοντικά σημεία αλλαγής τάσης επιλέγονται τυχαία με τέτοιο τρόπο ώστε η μέση συχνότητα για τα σημεία αλλαγής να ταιριάζει με αυτήν του παρελθόντος.

Γενικά, η διαδικασία που ακολουθείται είναι η εξής:

- Χρήση των προσαρμοσμένων μεγεθών αλλαγών τάσης για την εκτίμηση της παραμέτρου Laplace Rate Scale.
- Υπολογισμός του μέσου ρυθμού εμφάνισης των σημείων αλλαγής στο παρελθόν.
- Επιλογή του αριθμού των μελλοντικών σημείων αλλαγής τάσης από τη διωνυμική κατανομή με ρυθμό ίσο με το μέσο ρυθμό εμφάνισης των σημείων αλλαγής από το προηγούμενο βήμα.
- Για κάθε ένα από τα σημεία αλλαγής τάσης, επιλέγεται ένα μέγεθος αλλαγής τάσης από την κατανομή Laplace του πρώτου βήματος.

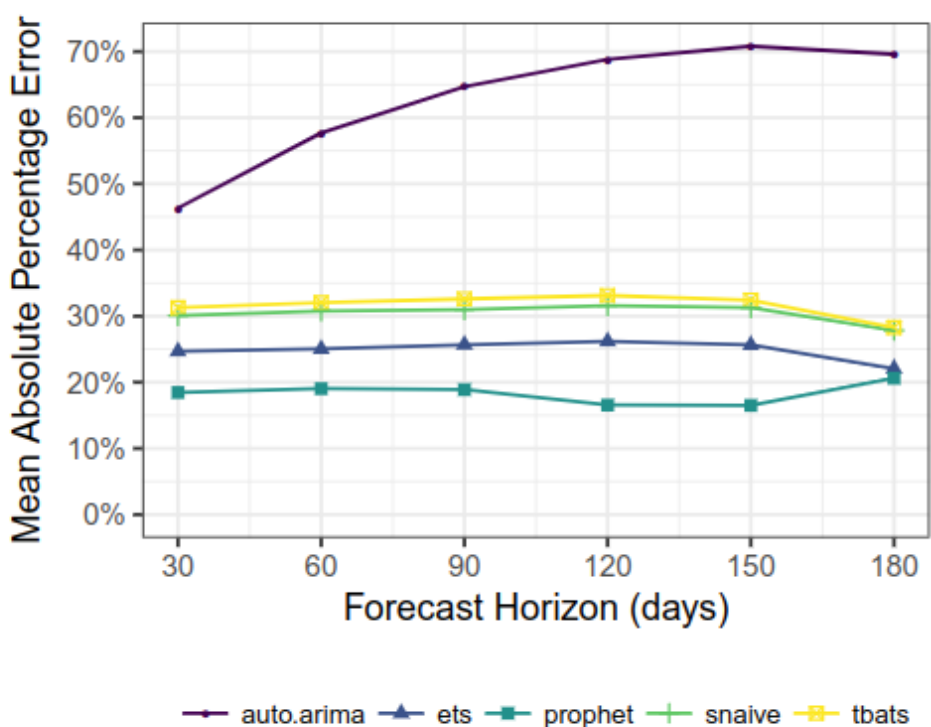
Η ανωτέρω διαδικασία επαναλαμβάνεται πολλές φορές (1000 από προεπιλογή) και υπολογίζονται τα διαστήματα που περιλαμβάνουν το 80% των τιμών.

Ο ρόλος των σημείων αλλαγής τάσης πριν την πρόβλεψη είναι σημαντικός. Μειώνοντας τα σημεία αλλαγής τάσης θα μειωθεί και το πλήθος των μεγεθών αλλαγής τάσης που θα προσαρμοστούν στο μοντέλο. Τα διαστήματα αβεβαιότητας είναι ένας χρήσιμος δείκτης για το επίπεδο της αβεβαιότητας και ειδικά ένας δείκτης για την υπερτροφοδότηση του μοντέλου. Όσο το τ (ουσιαστικά η μέση τιμή των απόλυτων τιμών των αλλαγών, παράμετρος "changepoint_prior_scale" στο Prophet) αυξάνεται, το μοντέλο έχει μεγαλύτερη ευελιξία στο να προσαρμοστεί στο παρελθόν, μειώνοντας έτσι τα σφάλματα στην παρατηρούμενη χρονοσειρά. Ωστόσο, όταν αυτή η ευελιξία επεκταθεί στις μελλοντικές προβλέψεις θα οδηγήσει σε μεγαλύτερα διαστήματα αβεβαιότητας. Από προεπιλογή το Prophet υπολογίζει την αβεβαιότητα μόνο για την τάση και τον θόρυβο παρατήρησης και όχι για την εποχικότητα.

2.5 Εφαρμογή με τη μέθοδο Prophet

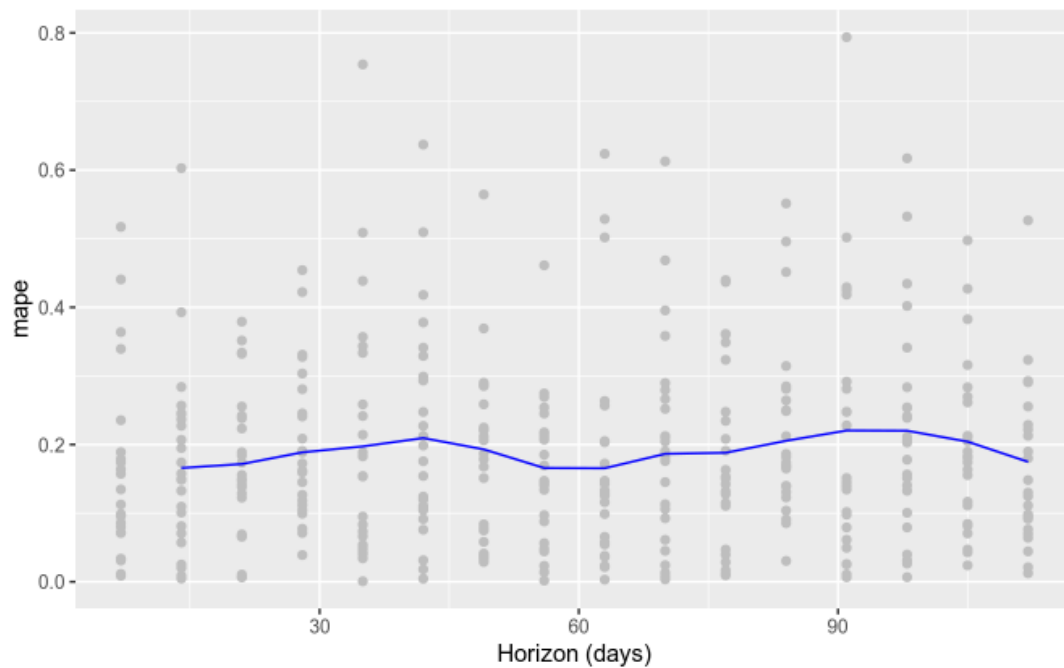
Για την ανάπτυξη του συστήματος παρακολούθησης και πρόβλεψης της χρονοσειράς των επιλογών των χρηστών, χρησιμοποιήθηκε η μέθοδος Prophet και συγκεκριμένα η υλοποίηση της στη γλώσσα προγραμματισμού R [89]. Οι συγγραφείς της μεθόδου [88] αξιολόγησαν και συνέκριναν τη μεθοδό τους με άλλες βασικές μεθόδους χρησιμοποιώντας το κριτήριο του

Μέσου Απόλυτου Ποσοστιαίου Σφάλματος (MAPE). Το Prophet παρουσίασε σημαντικά χαμηλότερο σφάλμα πρόβλεψης από τα άλλα μοντέλα. Τα αποτελέσματα συνοψίζονται στην Εικόνα 2.12.

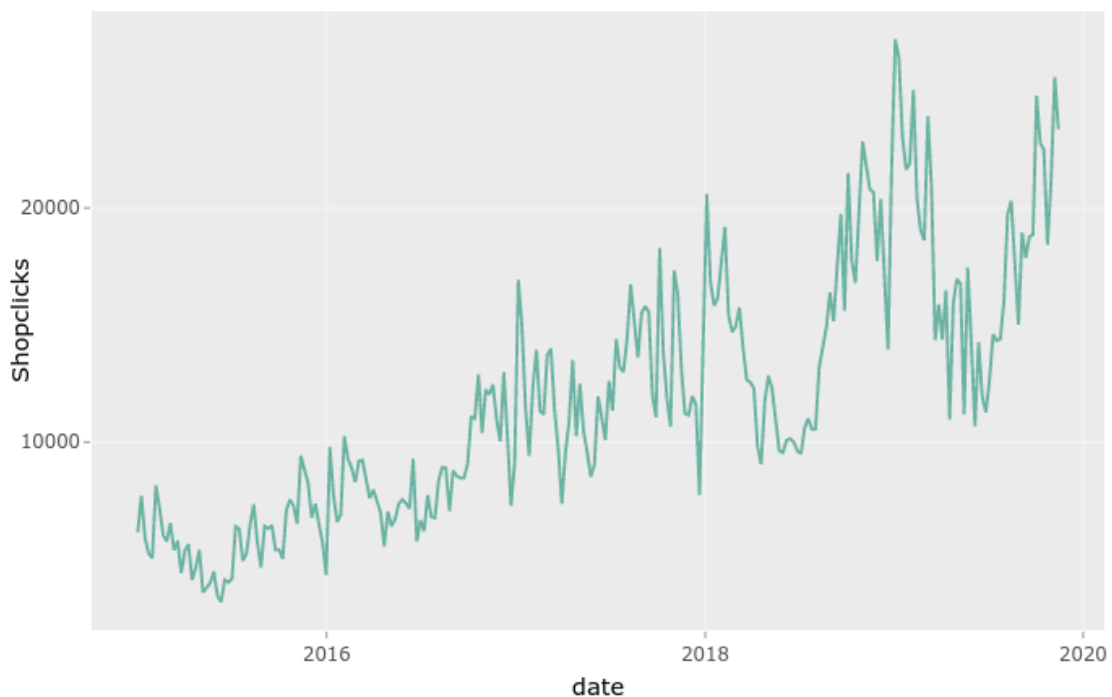


Εικόνα 2.12: Σύγκριση σφαλμάτων πρόβλεψης των μεθόδων *auto.ARIMA*, *εκθετικής εξομάλυνσης*, *Prophet*, *sNaive* και *TBATS*. Οι προβλέψεις με τη μέθοδο *Prophet* είχαν σημαντικά χαμηλότερο σφάλμα πρόβλεψης από τις άλλες αυτοματοποιημένες μεθόδους πρόβλεψης [89].

Επιπρόσθετα, ένας λόγος που επιλέχθηκε το Prophet ήταν η ευκολία παραμετροποίησης του, η ικανότητα δηλαδή που δίνει σε προγραμματιστές με διαφορετικές καταβολές και ειδίκευση να μπορούν να ρυθμίζουν τις παραμέτρους της μεθόδου όταν η αυτόματη πρόβλεψη δεν παράγει ένα ικανοποιητικό αποτέλεσμα. Η ρύθμιση των παραμέτρων στις μεθόδους που προαναφέρθηκαν προϋποθέτει την πλήρη κατανόηση του τρόπου με τον οποίο λειτουργούν τα υποκείμενα μοντέλα χρονοσειρών. Για παράδειγμα, οι πρώτοι παράμετροι που θα χρειαστεί να ρυθμιστούν στην αυτοματοποιημένη μέθοδο ARIMA είναι η μέγιστη τάξη διαφορίσης, οι παράμετροι της αυτοπαλινδρόμησης και οι παράμετροι του κινητού μέσου. Μία τέτοια ειδίκευση στην ανάλυση χρονοσειρών δεν συναντάται εύκολα στο προσωπικό μίας εταιρείας διαδικτύου και ένας τυπικός προγραμματιστής δεν ξέρει συνήθως πως να προσαρμόσει αυτές τις παραμέτρους. Το πακέτο Prophet παρέχει παραμέτρους που είναι εύκολο να κατανοηθούν και να ρυθμιστούν, ακόμα και από μη εξειδικευμένο προσωπικό, για την αντιμετώπιση πληθώρας σεναρίων στον επαγγελματικό κόσμο.

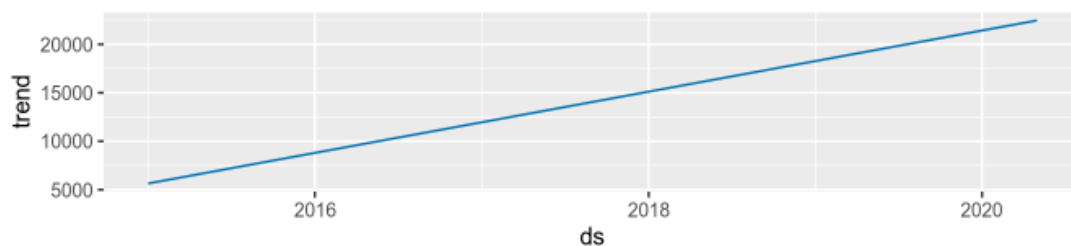


Εικόνα 2.13: Μέτρηση της απόδοσης του Prophet με το κριτήριο αξιολόγησης MAPE στη δοθείσα χρονοσειρά, σύμφωνα με τον εσωτερικό διαγνωστικό έλεγχο του Prophet.

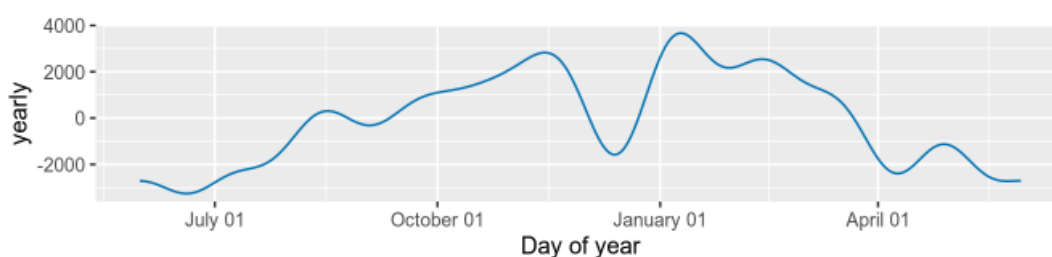


Εικόνα 2.14: Η χρονοσειρά των εβδομαδιαίων συνόλων επιλογών προτάσεων από τους χρήστες ("shopclicks").

Στην Εικόνα 2.14 απεικονίζεται η παρατηρούμενη χρονοσειρά. Στη συνέχεια, για την εφαρμογή του μοντέλου πρόβλεψης στη χρονοσειρά, το μοντέλο παραμετροποιήθηκε χρησιμοποιώντας γραμμικό μοντέλο ανάπτυξης της τάσης, αφήνοντας τον αριθμό και τη θέση αλλαγής τάσης προς αυτόματο υπολογισμό και ορίζοντας προσαρμογή ετήσιας εποχικότητας με τάξη σειρών Fourier ίση με 10. Στην Εικόνα 2.18 απεικονίζεται η πρόβλεψη της χρονοσειράς σε ένα διαδραστικό διάγραμμα όπου απεικονίζονται οι τιμές των πραγματικών και των προβλεπόμενων τιμών μαζί με τα διαστήματα αβεβαιότητας που υπολογίζει το μοντέλο. Στις Εικόνες 2.16 και 2.17 διαφαίνονται οι ετήσιοι εποχικοί κύκλοι με πιο έντονη παρουσία την πτώση των τιμών της χρονοσειράς την περίοδο των εορτών (Χριστουγέννων, Πάσχα) αλλά και την πτώση κατά τους θερινούς μήνες όπου η καλοκαιρία επιτρέπει τις υπαίθριες δραστηριότητες και μειώνει το ενδιαφέρον της κοινότητας για τη διακόσμηση εσωτερικών χώρων. Στην Εικόνα 2.13 παρουσιάζεται η αξιολόγηση του προβλεπτικού μηχανισμού μέσω του υπολογισμού του μέσου απόλυτου ποσοστιαίου σφάλματος. Στις Εικόνες 2.15, 2.16 και 2.17 παρουσιάζονται οι δύο κύριες συνιστώσες που συνθέτουν το μοντέλο πρόβλεψης στην παρούσα εφαρμογή, η τάση και η εποχικότητα.

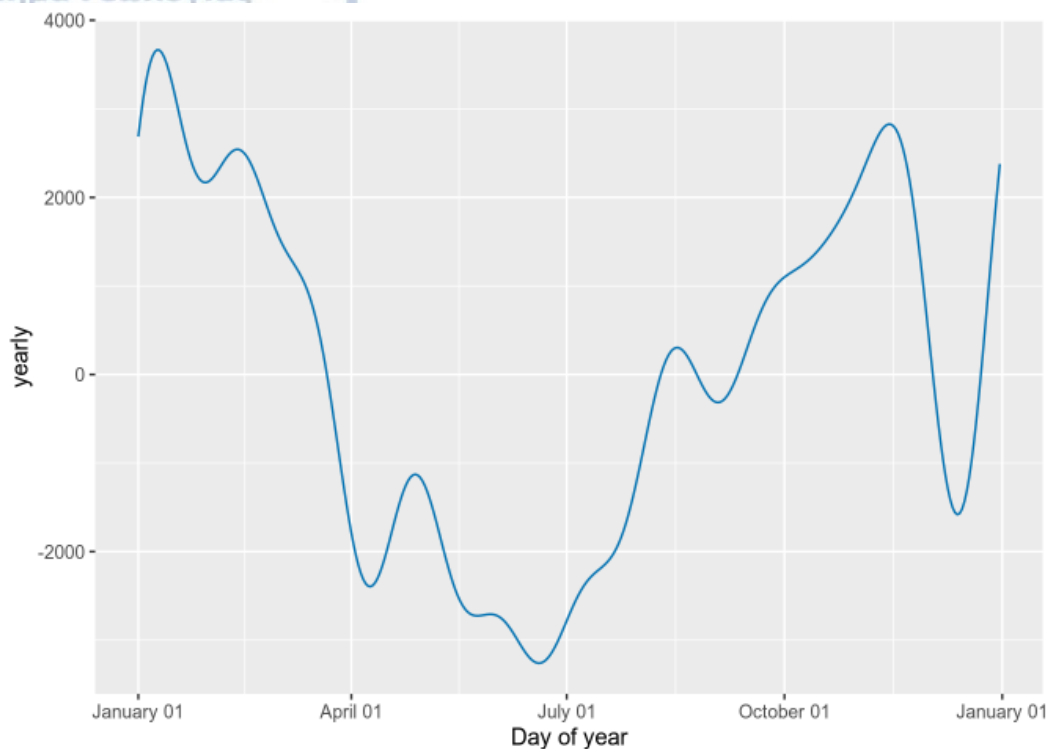


Εικόνα 2.15: Απεικόνιση της τάσης της δοθείσας χρονοσειράς όπως ανιχνεύτηκε από το Prophet.



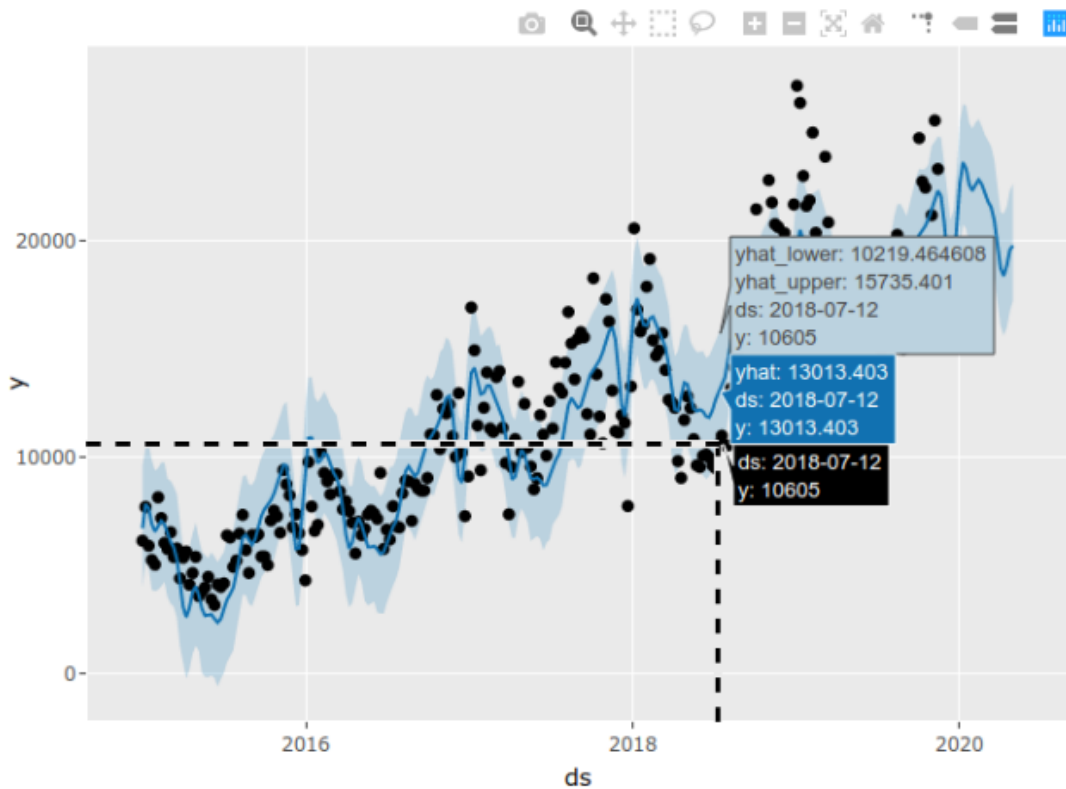
Εικόνα 2.16: Απεικόνιση της ετήσιας εποχικότητας της δοθείσας χρονοσειράς όπως ανιχνεύτηκε από το Prophet (στο διάστημα Ιουνίου-Μαΐος).

Είναι ενδιαφέρον να παρατηρήσουμε τις θετικές αιχμές που εντοπίζονται στα τέλη του χειμώνα του 2019 και τις αρνητικές αιχμές που εντοπίζονται στα μέσα της άνοιξης του 2019 (επισημασμένες περιοχές A και B στην Εικόνα 2.20). Η εξήγηση για τις θετικές αιχμές είναι δύο άρθρα (ένα σχετικό με προϊόντα IKEATM (IKEA, Delft, NL) που συγκέντρωσε 4100 επικροτήσεις (clicks) και 14000 επισκέψεις σε μία εβδομάδα και ένα σχετικό με την πασχάλινή διακόσμηση των σπιτιών) που γράφτηκαν εκείνη την περίοδο και προκάλεσαν

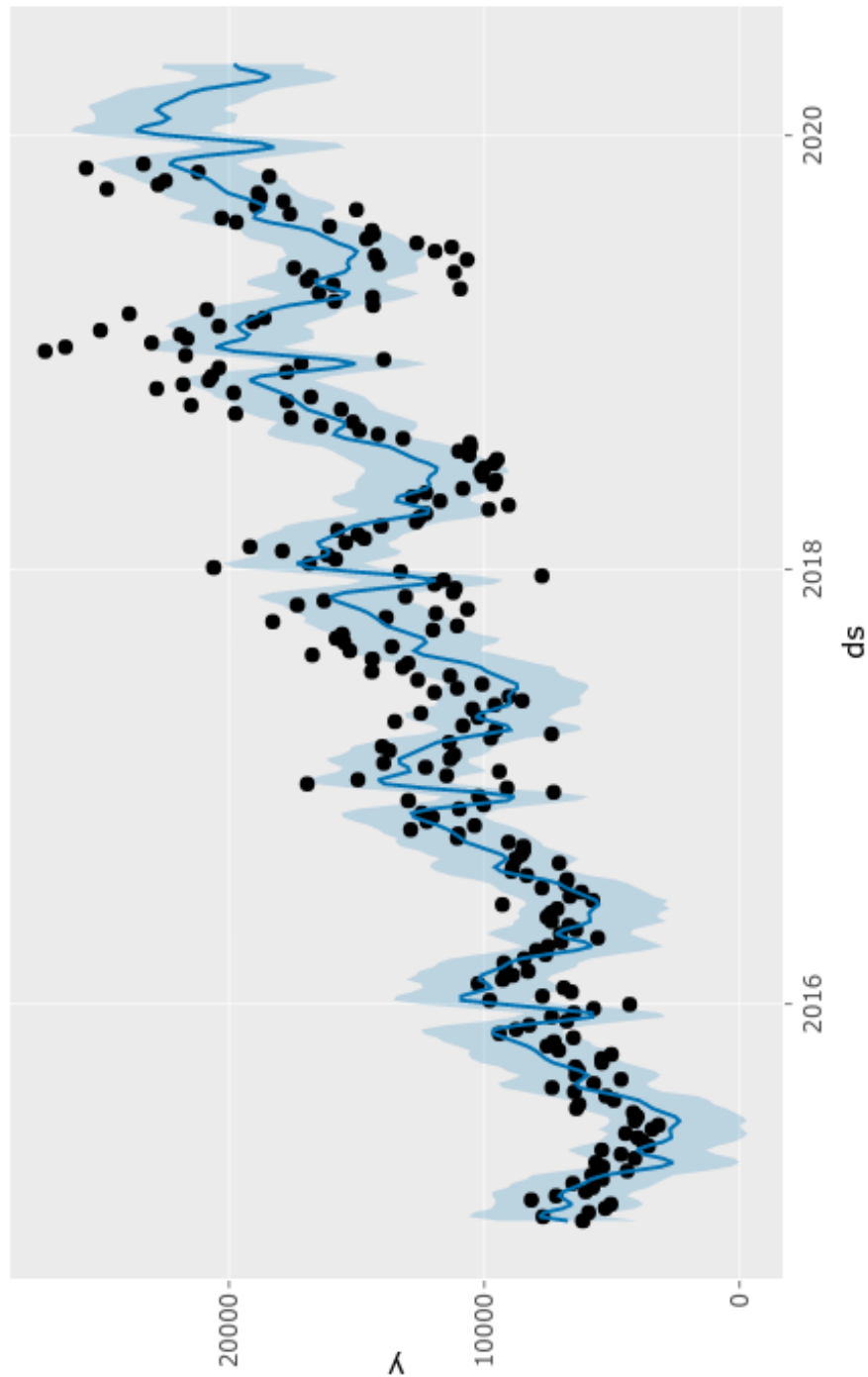


Εικόνα 2.17: Απεικόνιση της ετήσιας εποχικότητας της δοθείσας χρονοσειράς όπως ανιχνεύτηκε από το Prophet (στο διάστημα Ιανουάριος-Δεκέμβριος).

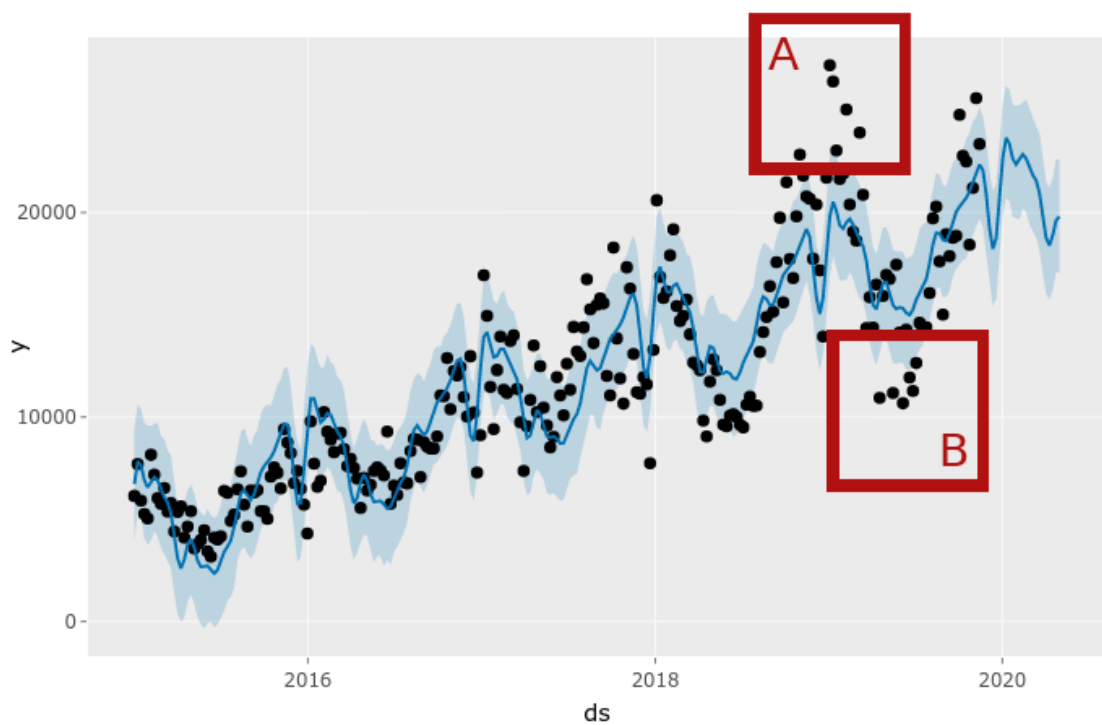
αυξημένη επισκεψιμότητα σε συνδυασμό με μία αύξηση της κίνησης μέσω PinterestTM (Pinterest Inc., San Francisco, US-CA) κατά 20% μέσω προωθητικών ενεργειών. Επιπλέον το νέο εβδομαδιαίο newsletter προσέλκυσε στην πλατφόρμα 240 χρήστες εκείνη την εβδομάδα, περίπου πέντε με έξι φορές περισσότερους από τον εβδομαδιαίο μέσο όρο. Οι αρνητικές αιχμές που παρουσιάζονται στα μέσα της άνοιξης και συνεχίζονται και μέχρι τα τέλη της οφείλονται στον αξιοσημείωτα αλλά και απρόσμενα καλό καιρό που επικράτησε στην Γερμανία το διάστημα εκείνο, με τις γιορτές του Πάσχα να περιλαμβάνονται σε αυτό. Επιπρόσθετα, μία αλλαγή στη διαδικτυακή πλατφόρμα, η οποία έπληξε άμεσα τη διαθεσιμότητα των προβαλλόμενων προϊόντων αλλά και την κίνηση στον ιστότοπο, την προερχόμενη από την GoogleTM (Google LLC, Mountain View, US-CA), η οποία υποχώρησε κατά 14%, χωρίς κάποια αλλαγή στον αλγόριθμό της, οδήγησε στην κατακρήμνιση των τιμών της χρονοσειράς για μερικές εβδομάδες.



Εικόνα 2.18: Πρόβλεψη της δοθείσας χρονοσειράς σε ένα διαδραστικό διάγραμμα όπου απεικονίζονται οι τιμές των πραγματικών και των προβλεπόμενων τιμών μαζί με τα διαστήματα αβεβαιότητας που υπολογίζει το μοντέλο.



Εικόνα 2.19: Η πρόβλεψη της χρονοσειράς των shopclicks για τους επόμενους δύο μήνες (8 εβδομάδες) από το Prophet.



Εικόνα 2.20: Επισήμανση δύο περιοχών στην έξοδο του *Prophet* που υπάρχουν έντονες ακραίες τιμές. (A) υποεκτίμηση, (B) υπερεκτίμηση

Κεφάλαιο 3

Ανίχνευση αντικειμένων στις εικόνες χρηστών

Η ανίχνευση αντικειμένων είναι ένα πεδίο της μηχανικής όρασης. Κύριο αντικείμενό της είναι η ανίχνευση αντικειμένων συγκεκριμένων κλάσεων μέσα σε μία εικόνα ή βίντεο καθώς και ο εντοπισμός της θέσης τους μέσα στην εικόνα.

3.1 Μηχανική όραση

Η μηχανική όραση ασχολείται με την αυτόματη εξαγωγή, ανάλυση και κατανόηση χρήσιμων πληροφοριών από μία μόνο εικόνα ή μία ακολουθία εικόνων. Περιλαμβάνει την ανάπτυξη μίας θεωρητικής και αλγοριθμικής βάσης για την αυτόματη οπτική κατανόηση. Ως επιστημονικός κλάδος, η όραση υπολογιστή ασχολείται με τη θεωρία που σχετίζεται με τα τεχνητά συστήματα που εξάγουν πληροφορίες από εικόνες. Τα δεδομένα εικόνας μπορούν να λάβουν πολλές μορφές, όπως ακολουθίες βίντεο, προβολές από πολλές κάμερες ή πολυδιάστατα δεδομένα από έναν ιατρικό σαρωτή. Ως τεχνολογικός κλάδος, η όραση υπολογιστή επιδιώκει να εφαρμόσει τις σχετικές θεωρίες και μοντέλα για την κατασκευή συστημάτων ηλεκτρονικής όρασης.

Τα παραδείγματα εφαρμογών ηλεκτρονικής όρασης στον πραγματικό κόσμο πληθαίνουν όλο και περισσότερο. Μερικά από αυτά είναι τα εξής:

- Συστήματα αυτόματης επιθεώρησης, όπως σε εγκαταστάσεις παραγωγής
- Υποβοήθηση ανθρώπων σε εργασίες αναγνώρισης, όπως αναγνώριση ειδών [95]
- Διαδικασίες ελέγχου, όπως βιομηχανικά ρομπότ
- Ανίχνευση συμβάντων, όπως οπτική παρακολούθηση ή καταμέτρηση ατόμων
- Αλληλεπίδραση, όπως είσοδος σε μία συσκευή αλληλεπίδρασης μεταξύ υπολογιστή και ανθρώπου
- Μοντελοποίηση αντικειμένων ή και περιβάλλοντος, όπως ανάλυση ιατρικών εικόνων ή τοπογραφικών μοντέλων
- Πλοήγηση, όπως σε αυτόνομο όχημα ή κινητό ρομπότ

- Οργάνωση πληροφοριών, όπως δημιουργία ευρετηρίου βάσης δεδομένων εικόνων
- Υποστήριξη της δημιουργίας οπτικών εφέ στην παραγωγή κινηματογραφικών ταινιών
- Παρακολούθηση
- Ανίχνευση και μέτρηση οργανισμών στις βιολογικές επιστήμες [12]

Η οργάνωση ενός συστήματος ηλεκτρονικής όρασης εξαρτάται σε μεγάλο βαθμό από την εφαρμογή. Ορισμένα συστήματα είναι αυτόνομες εφαρμογές που επιλύουν ένα συγκεκριμένο πρόβλημα μέτρησης ή ανίχνευσης, ενώ άλλα αποτελούν ένα υποσύστημα ενός μεγαλύτερου σχεδίου, το οποίο για παράδειγμα περιλαμβάνει, επίσης, υποσυστήματα για τον έλεγχο μηχανικών ενεργοποιητών, προγραμματισμού, βάσεων δεδομένων και μηχανικών διεπαφών. Η συγκεκριμένη υλοποίηση ενός συστήματος υπολογιστικής όρασης εξαρτάται επίσης από το αν η λειτουργικότητά του είναι προκαθορισμένη ή το αν κάποιο μέρος του θα έχει δυνατότητα μάθησης ή δυνατότητα τροποποίησης κατά τη λειτουργία. Πολλές λειτουργίες είναι μοναδικές για την κάθε εφαρμογή. Υπάρχουν, ωστόσο, και τυπικές λειτουργίες που βρίσκονται στα περισσότερα συστήματα οράσεως υπολογιστών.

- **Απόκτηση εικόνας** - Μία ψηφιακή εικόνα παράγεται από έναν ή περισσότερους αισθητήρες εικόνας, οι οποίοι, εκτός από τους διάφορους τύπους φωτοευαίσθητων φωτογραφικών αισθητήρων, περιλαμβάνουν αισθητήρες εύρους πεδίου, συσκευές τομογραφίας, ραδιοεντοπιστής ή/και κάμερες υπερήχων. Ανάλογα με τον τύπο του αισθητήρα, τα δεδομένα εικόνας που προκύπτουν είναι μία συνηθισμένη εικόνα δύο διαστάσεων, ένας τρισδιάστατος όγκος ή μία ακολουθία εικόνων. Οι τιμές των μεμονωμένων εικονοστοιχείων αντιστοιχούν συνήθως στην ένταση φωτός σε μία ή περισσότερες φασματικές ζώνες (μονόχρωμες ή πολύχρωμες εικόνες), αλλά μπορούν επίσης να σχετίζονται και με διάφορες μετρήσεις, όπως βάθος, απορρόφηση ή ανάκλαση ηχητικών ή ηλεκτρομαγνητικών κυμάτων ή πυρηνικού μαγνητικού συντονισμού [22].
- **Προεπεξεργασία** - Προτού μία μέθοδος ηλεκτρονικής όρασης δύναται να εφαρμοστεί στα δεδομένα εικόνας προκειμένου να εξαχθεί κάποια συγκεκριμένη πληροφορία, είναι συνήθως απαραίτητη η επεξεργασία των δεδομένων προκειμένου να διασφαλιστεί ότι ικανοποιούν τις προϋποθέσεις που απαιτούνται από τη μέθοδο. Παραδείγματα αποτελούν:

Επαναδειγματοληψία για τη διασφάλιση ότι το σύστημα συντεταγμένων εικόνας είναι σωστό.

Μείωση θορύβου για να διασφαλιστεί ότι ο θόρυβος του αισθητήρα δεν εισάγει ψευδείς πληροφορίες.

Βελτίωση αντίθεσης για να διασφαλιστεί η ανίχνευση σχετικών πληροφοριών.

Κλιμάκωση της αναπαράστασης για τη βελτίωση των δομών εικόνας σε τοπικά κατάλληλες κλίμακες.

- **Εξαγωγή χαρακτηριστικών** - Τα χαρακτηριστικά της εικόνας εξάγονται σε διάφορα επίπεδα πολυπλοκότητας [22]. Τυπικά παραδείγματα τέτοιων χαρακτηριστικών είναι:

Γραμμές, άκρες και ακμίες.

Τοπικά σημεία ενδιαφέροντος όπως γωνίες, κηλίδες ή σημεία.

Πιο περίπλοκα χαρακτηριστικά μπορεί να σχετίζονται με την υφή, το σχήμα ή την κίνηση.

- **Ανίχνευση / τμηματοποίηση** - Σε κάποιο σημείο της επεξεργασίας λαμβάνεται απόφαση σχετικά με το ποια σημεία ή περιοχές της εικόνας είναι κατάλληλα για περαιτέρω επεξεργασία [22]. Ως παραδείγματα:

Επιλογή ενός συγκεκριμένου συνόλου σημείων ενδιαφέροντος.

Τμηματοποίηση μίας ή πολλαπλών περιοχών εικόνας που περιέχουν ένα συγκεκριμένο αντικείμενο ενδιαφέροντος.

Τμηματοποίηση εικόνας σε αρχιτεκτονική ένθετων σκηνών (spatial-taxon scene hierarchy) που περιλαμβάνει προσκήνιο, ομάδες αντικειμένων, μεμονωμένα αντικείμενα ή προεξέχοντα μέρη αντικειμένου [5].

Τμηματοποίηση ενός ή και πολλαπλών καρέ βίντεο [62].

- **Επεξεργασία σε υψηλό επίπεδο** - Σε αυτό το βήμα η είσοδος είναι τυπικά ένα μικρό σύνολο δεδομένων, για παράδειγμα ένα σύνολο σημείων ή μία περιοχή εικόνας που υποτίθεται ότι περιέχει ένα συγκεκριμένο αντικείμενο [22]. Η υπόλοιπη επεξεργασία αφορά, για παράδειγμα:

Επαλήθευση ότι τα δεδομένα ικανοποιούν τις προϋποθέσεις που ορίζονται από το μοντέλο και την εκάστοτε εφαρμογή.

Εκτίμηση παραμέτρων ειδικών της εφαρμογής, όπως η παρουσίαση αντικειμένου ή το μέγεθος αντικειμένου.

Αναγνώριση εικόνας - ταξινόμηση ενός ανιχνευθέντος αντικειμένου σε διαφορετικές κατηγορίες.

Σύγκριση και συνδυασμός δύο διαφορετικών προβολών του ίδιου αντικειμένου.

- **Λήψη αποφάσεων** - Λήψη της τελικής απόφασης [22], για παράδειγμα:

Θετικό ή αρνητικό στις εφαρμογές αυτόματης επιθεώρησης.

Αντιστοιχία ή μη σε εφαρμογές αναγνώρισης.

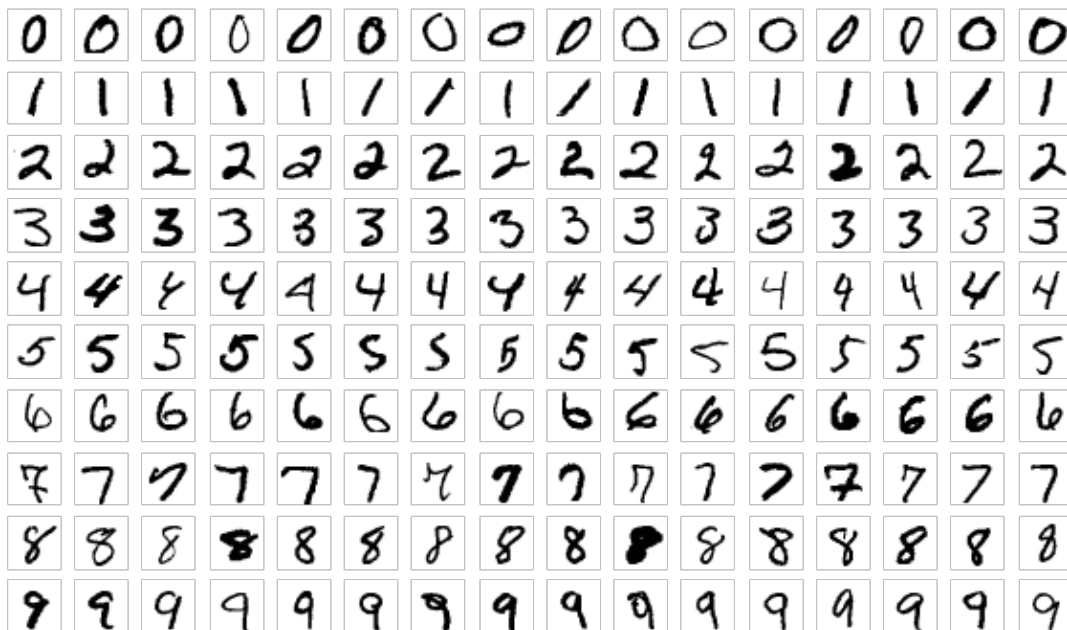
Σήμανση για περαιτέρω εξέταση από εξειδικευμένο προσωπικό κυρίως σε εφαρμογές που άπτονται θέματα ιατρικά, στρατιωτικά, ασφάλειας και αναγνώρισης.

3.2 Πεδία της μηχανικής όρασης

Το κλασικό πρόβλημα στην όραση υπολογιστή και τη μηχανική όραση είναι το να προσδιοριστεί αν τα δεδομένα εικόνες περιέχουν ή όχι κάποιο συγκεκριμένο αντικείμενο, χαρακτηριστικό ή δραστηριότητα. Συνοπτικά, τα κυριότερα αντικείμενα της μηχανικής όρασης είναι τα εξής:

Ταξινόμηση εικόνων. Η ταξινόμηση εικόνων αφορά την αντιστοίχιση ετικέτας σε μία εικόνα ή φωτογραφία. Το πρόβλημα αυτό αναφέρεται επίσης και ως “ταξινόμηση αντικειμένου” και ίσως πιο γενικά ως “αναγνώριση εικόνων”.

Ταξινόμηση εικόνων με εντοπισμό. Η ταξινόμηση εικόνων με εντοπισμό περιλαμβάνει την αντιστοίχιση μίας ετικέτας κλάσης σε μία εικόνα και την εμφάνιση της θέσης του αντικειμένου στην εικόνα με ένα πλαίσιο οριοθέτησης (σχεδιάζοντας ένα πλαίσιο γύρω από το αντικείμενο).

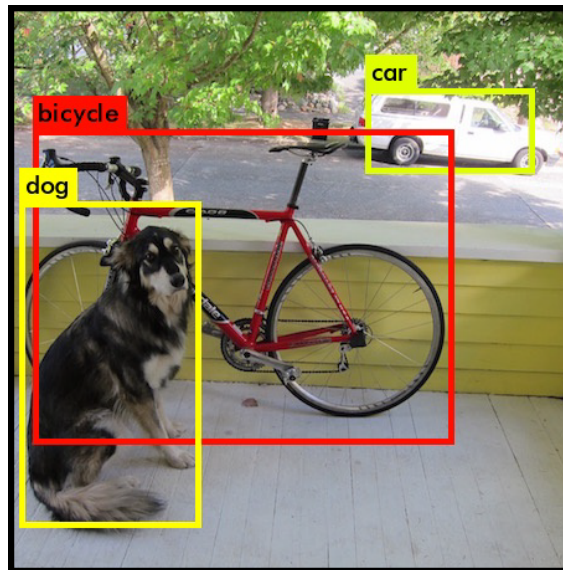


Εικόνα 3.1: Παράδειγμα χειρόγραφων ψηφίων από το σύνολο δεδομένων MNIST [100].

Ανίχνευση αντικειμένων. Ανίχνευση (ή αναγνώριση) αντικειμένων είναι η εργασία της ταξινόμησης εικόνων με εντοπισμό αλλά για όλα τα αντικείμενα που μπορεί να περιέχονται σε μία εικόνα. Η ανίχνευση αντικειμένων είναι μία τεχνολογία πληροφορικής που σχετίζεται με την υπολογιστική όραση και την επεξεργασία εικόνων και ασχολείται με την αναγνώριση σημασιολογικών αντικειμένων μίας συγκεκριμένης κλάσης (όπως

άνθρωποι, κτίρια ή αυτοκίνητα) σε ψηφιακές εικόνες και βίντεο. Καλά ερευνημένοι τομείς αναγνώρισης αντικειμένων περιλαμβάνουν αναγνώριση προσώπου και αναγνώριση πεζών. Η ανίχνευση αντικειμένων έχει εφαρμογές σε πολλούς τομείς της υπολογιστικής όρασης, συμπεριλαμβανομένης της ανάκτησης εικόνων και της βιντεοεπιτήρησης.

Το YOLO είναι η πλέον σύγχρονη λύση για γρήγορη ανίχνευση αντικειμένων σε πραγματικό χρόνο στις εικόνες.



Εικόνα 3.2: Παράδειγμα ανίχνευσης εικόνας με τον αλγόριθμο YOLO στο σύνολο δεδομένων COCO [75].

Κατάτμηση εικόνας. Η κατάτμηση (ή τμηματοποίηση ή κατακερματισμός) εικόνας ή σημασιολογική κατάτμηση είναι η εργασία ανίχνευσης αντικειμένων όπου σχεδιάζεται το περίγραμμα του κάθε αντικειμένου που ανιχνεύεται στην εικόνα. Η κατάτμηση εικόνων είναι ένα γενικότερο πρόβλημα τμηματοποίησης του συνόλου της εικόνας.

Σε αντίθεση με την ανίχνευση αντικειμένων που περιλαμβάνει τη χρήση ενός πλαισίου οριοθέτησης για τον προσδιορισμό των αντικειμένων, η κατάτμηση εικόνας προσδιορίζει τα συγκεκριμένα εικονοστοιχεία της εικόνας που ανήκουν στο αντικείμενο.

Μεταφορά τεχνοτροπίας εικόνας. Μεταφορά τεχνοτροπίας είναι η εργασία της εκμάθησης της τεχνοτροπίας από μία ή περισσότερες εικόνες και της εφαρμογής αυτής σε μία νέα εικόνα. Αυτή η εργασία μπορεί να θεωρηθεί ως ένας τύπος εφαρμογής φωτογραφικού φίλτρου σε μία εικόνα. Παράδειγμα αποτελεί η εφαρμογή της τεχνοτροπίας συγκεκριμένων διάσημων έργων τέχνης σε νέες εικόνες [35].

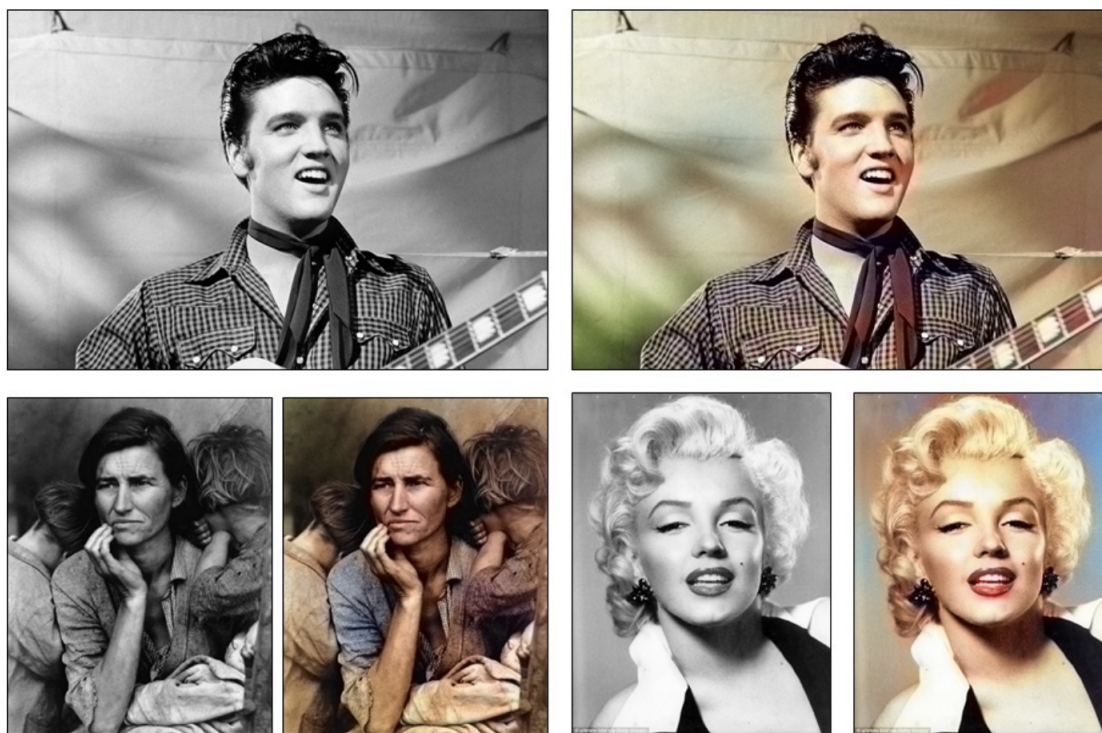
Χρωματισμός εικόνας. Ο χρωματισμός εικόνας αφορά τη μετατροπή μίας εικόνας από την κλίμακα του γκρι σε μία πλήρη έγχρωμη εικόνα. Αυτή η εργασία μπορεί επίσης να θεωρηθεί ως ένας τύπος φωτογραφικού φίλτρου. Παραδείγματα περιλαμβάνουν τον χρωματισμό παλιών ασπρόμαυρων φωτογραφιών και ταινιών.



Εικόνα 3.3: Παράδειγμα κατάτμησης εικόνας με τη χρήση του αλγορίθμου Mask R-CNN [41].



Εικόνα 3.4: Παράδειγμα μεταφοράς τεχνοτροπίας εικόνας [32].



Εικόνα 3.5: Παραδείγματα από τον χρωματισμό εικόνας [104].

Ανακατασκευή εικόνας. Η ανακατασκευή (ή αποκατάσταση) εικόνων είναι η εργασία της πλήρωσης ελλιπών ή φθαρμένων τμημάτων μίας εικόνας. Παραδείγματα περιλαμβάνουν την ανακατασκευή (αποκατάσταση) παλαιών, κατεστραμμένων φωτογραφιών και ταινιών. Ο στόχος της αποκατάστασης εικόνας είναι κυρίως η απομάκρυνση του θορύβου (όπως θόρυβος αισθητήρα ή κίνησης) από τις εικόνες. Η απλούστερη δυνατή προσέγγιση για την απομάκρυνση του θορύβου είναι διάφοροι τύποι φίλτρων όπως φίλτρα χαμηλής διέλευσης ή μεσαία φίλτρα. Οι πιο εξελιγμένες μέθοδοι υποθέτουν ένα μοντέλο της εμφάνισης των τοπικών δομών εικόνας, για να τις διακρίνουν από τον θόρυβο. Αρχικά αναλύοντας τα δεδομένα εικόνας σε σχέση με τις τοπικές δομές εικόνας, όπως γραμμές ή ακμές, και στη συνέχεια ελέγχοντας το φιλτράρισμα με βάση τις τοπικές πληροφορίες από το βήμα της ανάλυσης, λαμβάνεται συνήθως ένα καλύτερο επίπεδο απομάκρυνσης θορύβου σε σύγκριση με τις απλούστερες προσεγγίσεις.

Αναδημιουργία σκηνής. Λαμβάνοντας υπόψη μία ή περισσότερες εικόνες μίας σκηνής, η αναδημιουργία σκηνής στοχεύει στον υπολογισμό ενός τρισδιάστατου μοντέλου της. Στην απλούστερη περίπτωση, το μοντέλο μπορεί να είναι ένα σύνολο τρισδιάστατων σημείων. Πιο εξελιγμένες μέθοδοι παράγουν ένα πλήρες τρισδιάστατο μοντέλο επιφανειών. Η εμφάνιση της τρισδιάστατης απεικόνισης που δεν απαιτεί κίνηση ή σάρωση και οι σχετικοί αλγόριθμοι επεξεργασίας επιτρέπουν την ταχεία πρόοδο σε αυτό το πεδίο. Η ανίχνευση σε τρεις διαστάσεις μπορεί να χρησιμοποιηθεί για την απόκτηση τρισδιάστατων εικόνων από

Εικόνα επαυξημένης ανάλυσης (Image Super-Resolution). Η εικόνα επαυξημένης ανάλυσης είναι η δημιουργία μίας νέας έκδοσης μίας εικόνας με μεγαλύτερη ανάλυση και λεπτομέρεια από την αρχική εικόνα. Συχνά τα μοντέλα που αναπτύσσονται για την επαυξημένη ανάλυση μπορούν να χρησιμοποιηθούν και για την αποκατάσταση εικόνας καθώς επιλύουν σχετικά προβλήματα.

Σύνθεση εικόνας. Η σύνθεση εικόνας είναι η δημιουργία στοχοθετημένων τροποποιήσεων υφιστάμενων εικόνων ή εντελώς νέων εικόνων. Αυτός είναι ένας πολύ ευρύς και με ραγδαία εξέλιξη τομέας. Μπορεί να περιλαμβάνει μικρές τροποποιήσεις εικόνας και βίντεο, όπως για παράδειγμα:

- Αλλαγή του στυλ ενός αντικειμένου σε μία σκηνή.
- Προσθήκη ενός αντικειμένου σε μία σκηνή.
- Προσθήκη προσώπου σε μία σκηνή.

Μπορεί επίσης να περιλαμβάνει τη δημιουργία εντελώς νέων εικόνων, όπως η δημιουργία προσώπων, δωματίων ή ρούχων (με τη χρήση Generative Adversarial Networks, GAN).

Προβλήματα παραγωγής κειμένου από εικόνα και εικόνας από κείμενο. Αυτά περιλαμβάνουν συνήθως τη σύνθεση εικόνων από κείμενο αλλά και την παραγωγή κειμένου (όπως περιγραφή εικόνων) από εικόνες. Αξιοσημείωτα παραδείγματα μετατροπής εικόνας σε κείμενο και κειμένου σε εικόνα:

- Υποτιτλισμός εικόνας: Δημιουργία κειμενικής περιγραφής μίας εικόνας (Show and Tell: A Neural Image Caption Generator, 2014 [93]).
- Περιγραφή εικόνας: Δημιουργία μίας κειμενικής περιγραφής του κάθε αντικειμένου σε μία εικόνα καθώς και των ενεργειών που τυχόν απεικονίζονται σε μία εικόνα (Deep Visual-Semantic Alignments for Generating Image Descriptions, 2015 [49]).
- Μετατροπή κειμένου σε εικόνα: Σύνθεση μίας εικόνας στη βάση ενός περιγραφικού κειμένου (AttnGAN: Fine-Grained Text to Image Generation with Attentional Generative Adversarial Networks, 2017 [99]).

Πρόσθετα προβλήματα. Παρακάτω περιγράφονται ορισμένα επιπλέον εξειδικευμένα προβλήματα υπολογιστικής όρασης.

- Ανάκτηση εικόνας βάσει περιεχομένου - εύρεση όλων των εικόνων σε ένα ευρύτερο σύνολο εικόνων που έχουν ένα συγκεκριμένο περιεχόμενο. Το περιεχόμενο μπορεί να προσδιοριστεί με διάφορους τρόπους, για παράδειγμα όσον αφορά στην ομοιότητα σε σχέση με μία εικόνα στόχου (εύρεση όλων των εικόνων των παρόμοιων με την

εικόνα X) ή με όρους κριτηρίων αναζήτησης υψηλού επιπέδου που δίνονται ως εισαγωγή κειμένου (όπως εύρεση όλων των εικόνων που περιέχουν πολλά σπίτια, έχουν ληφθεί κατά τη διάρκεια του χειμώνα, και δεν περιέχουν αυτοκίνητα).

- Υπολογισμός θέσης - εκτίμηση της θέσης ή του προσανατολισμού ενός συγκεκριμένου αντικειμένου σε σχέση με την κάμερα. Ένα παράδειγμα εφαρμογής αυτής της τεχνικής θα ήταν η παροχή βοήθειας σε έναν ρομποτικό βραχίονα στην ανάκτηση αντικειμένων από έναν ιμάντα μεταφοράς γραμμής συναρμολόγησης ή η λήψη μερών της συναρμολόγησης από ένα κουτί.
- Οπτική αναγνώριση χαρακτήρων (OCR) - αναγνώριση χαρακτήρων σε εικόνες τυπωμένου ή χειρόγραφου κειμένου, συνήθως με σκοπό την κωδικοποίηση του κειμένου σε μορφή περισσότερο επιδεκτική επεξεργασίας ή ευρετηρίασης (όπως ASCII).
- Ανάγνωση δισδιάστατου κώδικα - όπως πίνακες δεδομένων και κώδικες QR.
- Αναγνώριση προσώπου.
- Τεχνολογία αναγνώρισης σχήματος (SRT).

Επί του παρόντος, οι καλύτεροι αλγόριθμοι για τέτοιες εργασίες βασίζονται στη βαθιά μάθηση και συγκεκριμένα σε συνελικτικά νευρωνικά δίκτυα. Μία απεικόνιση των δυνατοτήτων τους δίνεται από το ImageNet Large Scale Visual Recognition Challenge. Ο διαγωνισμός αυτός αποτελεί ένα σημείο αναφοράς στην ταξινόμηση και ανίχνευση αντικειμένων, με εκατομμύρια εικόνες και εκατοντάδες τάξεις αντικειμένων. Η απόδοση των συνελικτικών νευρωνικών δικτύων, στις δοκιμές ImageNet, προσομοιάζει πλέον αυτήν των ανθρώπων [79]. Οι καλύτεροι αλγόριθμοι εξακολουθούν να δυσκολεύονται με αντικείμενα που είναι μικρά ή λεπτά. Έχουν επίσης προβλήματα με εικόνες που έχουν παραμορφωθεί με φίλτρα (ένα όλο και συχνότερο φαινόμενο με τις σύγχρονες ψηφιακές φωτογραφικές μηχανές). Αντίθετα, τέτοιου είδους εικόνες σπάνια δυσκολεύουν τον άνθρωπο. Οι άνθρωποι, ωστόσο, τείνουν να έχουν προβλήματα με άλλα θέματα. Για παράδειγμα, δεν είναι καλοί στην ταξινόμηση αντικειμένων σε λεπτομερείς κλάσεις, όπως συγκεκριμένη φυλή σκύλου ή είδος πουλιού, που τα συνελικτικά νευρωνικά δίκτυα χειρίζονται σχετικά εύκολα.

3.3 Βαθιά μάθηση

Η βαθιά μάθηση (deep learning) [51] είναι ένα πεδίο της τεχνητής νοημοσύνης που εστιάζει στη δημιουργία μοντέλων μεγάλων νευρωνικών δικτύων ικανών στην ακριβή εξαγωγή συμπερασμάτων βασισμένων σε δεδομένα. Η βαθιά μάθηση ενδείκνυται ιδιαίτερα σε περιπτώσεις όπου τα δεδομένα είναι πολύπλοκα και όπου είναι διαθέσιμα μεγάλα σύνολα δεδομένων.

Ένας αλγόριθμος μηχανικής μάθησης είναι μία διαδικασία αναζήτησης σχεδιασμένης να επιλέγει την καλύτερη συνάρτηση, από ένα σύνολο πιθανών συναρτήσεων, ώστε να εκφράσει τις σχέσεις μεταξύ των χαρακτηριστικών ενός συνόλου δεδομένων.

Η μηχανική μάθηση περιλαμβάνει μία διαδικασία εκπαίδευσης. Κατά τη διάρκεια της εκπαίδευσης, ένας αλγόριθμος μηχανικής μάθησης επεξεργάζεται ένα σύνολο δεδομένων και επιλέγει τη συνάρτηση που καλύτερα αναπαριστά τα μοτίβα στα δεδομένα. Η εξαγόμενη συνάρτηση θα κωδικοποιηθεί σε ένα υπολογιστικό πρόγραμμα. Η κωδικοποιημένη συνάρτηση είναι γνωστή ως μοντέλο, και η ανάλυση των δεδομένων ώστε να εξαχθεί η συνάρτηση συχνά αναφέρεται ως εκπαίδευση του μοντέλου.

Στο πλαίσιο της βαθιάς μάθησης, η εξαγόμενη συνάρτηση από ένα σύνολο δεδομένων κατά τη διάρκεια της εκπαίδευσης αναπαρίσταται ως ένα μοντέλο νευρωνικού δικτύου και αντιστρόφως ένα μοντέλο νευρωνικού δικτύου κωδικοποιεί μία συνάρτηση ως ένα υπολογιστικό πρόγραμμα. Ο όρος βαθιά μάθηση περιγράφει μία οικογένεια από μοντέλα νευρωνικών δικτύων που έχουν πολλαπλά επίπεδα από προγράμματα επεξεργασίας απλών πληροφοριών, γνωστά ως νευρώνες.

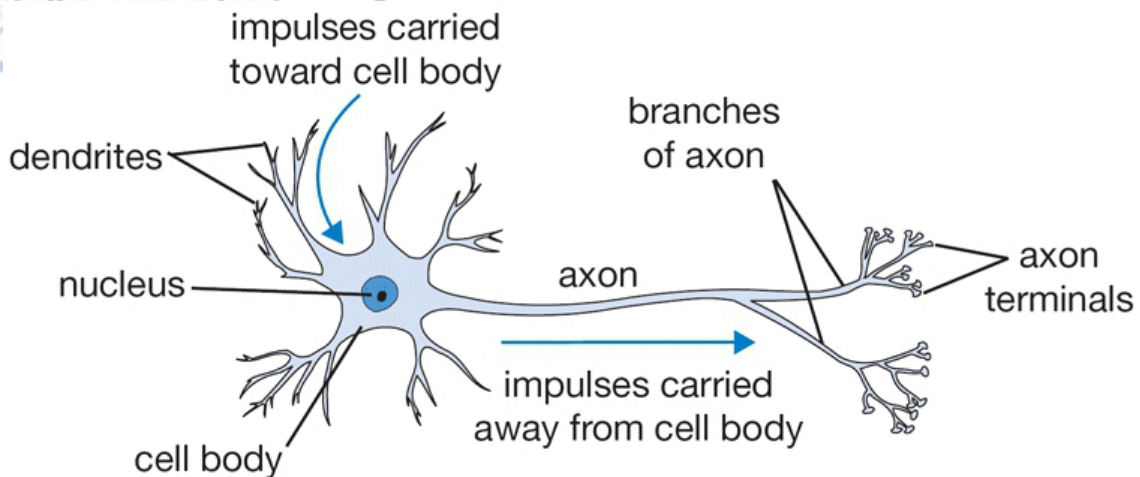
3.4 Νευρωνικά δίκτυα

Ένα νευρωνικό δίκτυο είναι ένα υπολογιστικό μοντέλο εμπνευσμένο από τη δομή του ανθρώπινου εγκεφάλου. Ο ανθρώπινος εγκέφαλος αποτελείται από έναν τεράστιο αριθμό νευρικών κυττάρων, ονομαζόμενα νευρώνες.

Οι νευρώνες έχουν μία απλή δομή τριών μερών που αποτελείται από: ένα κυτταρικό σώμα, ένα σύνολο ινών που ονομάζονται δενδρίτες και μία ενιαία μακρά ίνα που ονομάζεται άξονας. Το σχήμα 3.6 απεικονίζει τη δομή ενός νευρώνα και πως αυτός συνδέεται με άλλους νευρώνες του εγκεφάλου. Οι δενδρίτες και ο άξονας συνδέονται με το κυτταρικό σώμα και οι δενδρίτες ενός νευρώνα συνδέονται με τους άξονες άλλων νευρώνων. Οι δενδρίτες λειτουργούν ως κανάλια εισόδου στον νευρώνα και λαμβάνουν σήματα που αποστέλλονται από άλλους νευρώνες μέσω των αξόνων τους. Ο άξονας λειτουργεί ως κανάλι εξόδου ενός νευρώνα και έτσι οι άλλοι νευρώνες, των οποίων οι δενδρίτες συνδέονται με τον άξονα, δέχονται τα σήματα που στέλνονται κατά μήκος του άξονα ως είσοδο.

Η λειτουργία των νευρώνων είναι απλή. Εάν τα εισερχόμενα ερεθίσματα είναι αρκετά ισχυρά, ο νευρώνας μεταδίδει έναν ηλεκτρικό παλμό, που ονομάζεται δυναμικό δράσης, κατά μήκος του άξονά του στους άλλους νευρώνες που είναι συνδεδεμένοι με αυτόν. Συνεπώς, ένας νευρώνας λειτουργεί σαν ένας διακόπτης, ο οποίος δέχεται ένα σύνολο εισόδων και είτε εξάγει ένα δυναμικό δράσης είτε όχι.

Ένα τεχνητό νευρωνικό δίκτυο αποτελείται από ένα δίκτυο μονάδων επεξεργασίας απλών πληροφοριών, ονομαζόμενων νευρώνων. Η δύναμη των νευρωνικών δικτύων να αναπαραστήσουν σύνθετες διασυνδέσεις δεν είναι το αποτέλεσμα περίπλοκων μαθηματικών μοντέλων αλλά κυρίως προκύπτει από τις αλληλεπιδράσεις ενός μεγάλου συνόλου απλών

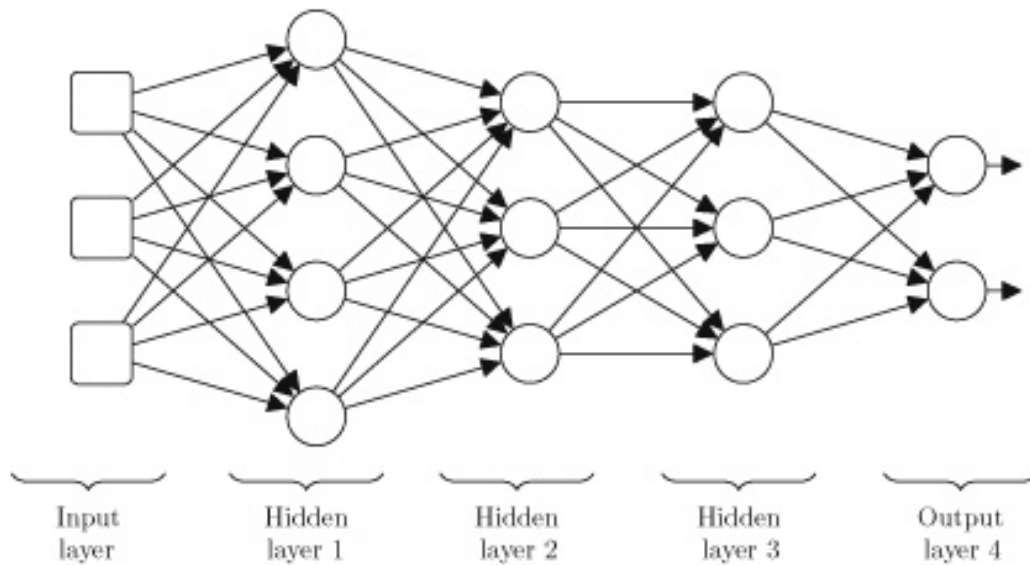


Εικόνα 3.6: Δομή ενός νευρώνα του εγκεφάλου [48].

νευρώνων.

Η Εικόνα 3.7 απεικονίζει τη δομή ενός νευρωνικού δικτύου. Είναι συνήθης η απεικόνιση των νευρώνων ενός νευρωνικού δικτύου οργανωμένων σε επίπεδα. Το εικονιζόμενο δίκτυο έχει πέντε επίπεδα: ένα επίπεδο εισόδου, τρία κρυφά επίπεδα και ένα επίπεδο εξόδου. Ένα κρυφό επίπεδο είναι απλά ένα επίπεδο που δεν είναι ούτε η είσοδος ούτε η έξοδος. Τα δίκτυα βαθιάς μάθησης είναι νευρωνικά δίκτυα που έχουν πολλά κρυφά επίπεδα νευρώνων. Ο ελάχιστος αριθμός κρυφών επιπέδων που απαιτείται ώστε να θεωρούνται βαθιά είναι δύο. Ωστόσο, τα περισσότερα δίκτυα βαθιάς μάθησης έχουν πολλά περισσότερα από δύο κρυφά επίπεδα. Το σημαντικό σημείο είναι ότι το βάθος ενός δικτύου μετράται ως ο αριθμός των κρυφών επιπέδων συν το επίπεδο εξόδου. Στην Εικόνα 3.7, τα τετράγωνα στο επίπεδο εισόδου αντιπροσωπεύουν περιοχές στη μνήμη που χρησιμοποιούνται για την υποδοχή εισόδων στο δίκτυο. Αυτές οι περιοχές μπορούν να θεωρηθούν ως αισθητικοί νευρώνες. Στους νευρώνες αυτούς δεν πραγματοποιείται επεξεργασία πληροφοριών και η έξοδος καθενός από αυτούς τους αισθητικούς νευρώνες είναι απλώς η τιμή των δεδομένων των αποθηκευμένων στη θέση μνήμης. Οι κύκλοι στο σχήμα απεικονίζουν τους νευρώνες επεξεργασίας πληροφοριών στο δίκτυο. Κάθε ένας από αυτούς τους νευρώνες δέχεται ένα σύνολο αριθμητικών τιμών ως είσοδο και τις αντιστοιχίζει σε μία μόνο τιμή εξόδου. Κάθε είσοδος σε ένα νευρώνα επεξεργασίας είναι είτε η έξοδος ενός αισθητικού νευρώνα είτε η έξοδος ενός άλλου νευρώνα επεξεργασίας.

Τα βέλη στην Εικόνα 3.7 απεικονίζουν τον τρόπο ροής της πληροφορίας μέσω του δικτύου από την έξοδο ενός νευρώνα στην είσοδο ενός άλλου νευρώνα. Κάθε σύνδεση σε ένα δίκτυο συνδέει δύο νευρώνες και κάθε σύνδεση έχει κατεύθυνση, πράγμα που σημαίνει ότι οι πληροφορίες που μεταφέρονται κατά μήκος μίας σύνδεσης ρέουν μόνο προς μία κατεύθυνση. Κάθε μία από τις συνδέσεις σε ένα δίκτυο έχει ένα βάρος συσχετισμένο με αυτό. Το βάρος μίας σύνδεσης είναι απλά ένας αριθμός, αλλά αυτά τα βάρη είναι πολύ σημαντικά. Το βάρος



Εικόνα 3.7: Τοπολογική απεικόνιση ενός απλού νευρωνικού δικτύου [51].

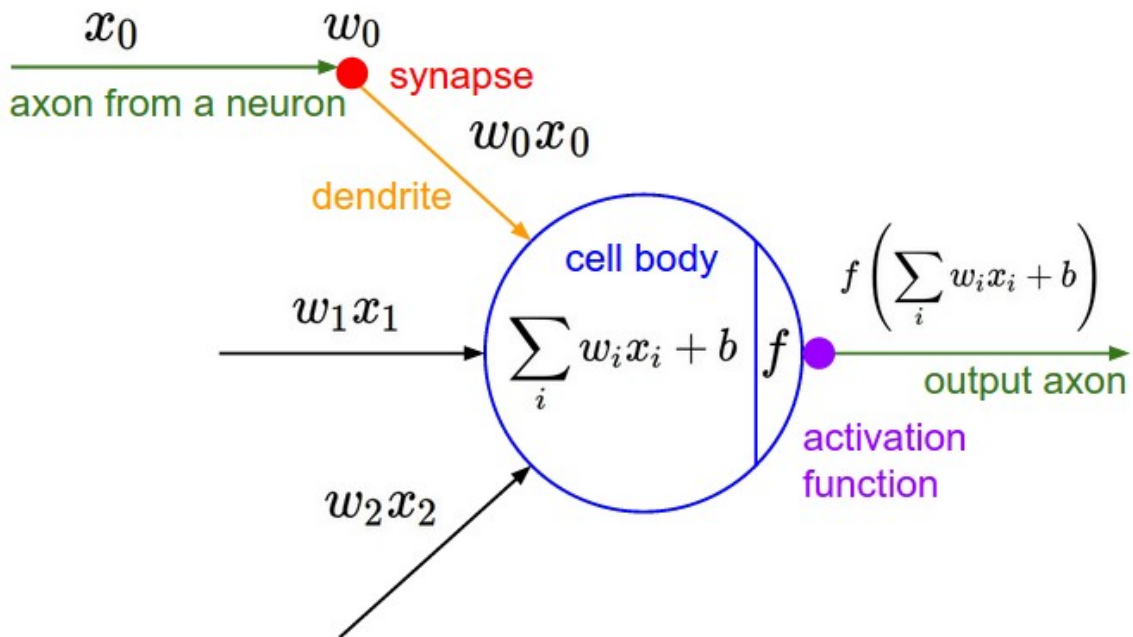
μίας σύνδεσης επηρεάζει το πώς ένας νευρώνας επεξεργάζεται τις πληροφορίες που λαμβάνει κατά τη σύνδεση και στην πραγματικότητα, η κατάρτιση ενός τεχνητού νευρωνικού δικτύου, ουσιαστικά, σχετίζεται με την αναζήτηση για το καλύτερο (ή βέλτιστο) σύνολο βαρών.

Ένας νευρώνας υλοποιεί επίσης μία διαδικασία δύο σταδίων για την αντιστοίχιση των εισόδων σε μία έξοδο. Το πρώτο στάδιο επεξεργασίας περιλαμβάνει τον υπολογισμό ενός σταθμισμένου αθροίσματος των εισόδων στον νευρώνα. Στη συνέχεια, το αποτέλεσμα του υπολογισμού του σταθμισμένου αθροίσματος διαβιβάζεται σε ένα δεύτερο στάδιο που αντιστοιχίζει τα αποτελέσματα του αθροίσματος στην τελική τιμή εξόδου του νευρώνα. Όταν σχεδιάζουμε έναν νευρώνα, μπορούμε να χρησιμοποιήσουμε πολλούς διαφορετικούς τύπους συναρτήσεων για αυτό το δεύτερο στάδιο ή την επεξεργασία.

Τυπικά η τιμή εξόδου ενός νευρώνα είναι γνωστή ως η τιμή ενεργοποίησής του, έτσι αυτό το δεύτερο στάδιο στο οποίο αντιστοιχίζεται το αποτέλεσμα του σταθμισμένου αθροίσματος στην τιμή ενεργοποίησης του νευρώνα, είναι γνωστό ως λειτουργία ενεργοποίησης. Η Εικόνα 3.8 απεικονίζει πώς αυτά τα στάδια επεξεργασίας αντικατοπτρίζονται στη δομή ενός τεχνητού νευρώνα. Στην Εικόνα 3.8, το σύμβολο Σ αντιπροσωπεύει τον υπολογισμό του σταθμισμένου συνόλου και το σύμβολο f αντιπροσωπεύει τη συνάρτηση ενεργοποίησης που επεξεργάζεται το σταθμισμένο άθροισμα και παράγει την έξοδο από τον νευρώνα.

Ο νευρώνας στην Εικόνα 3.8 λαμβάνει n εισόδους σε n διαφορετικές συνδέσεις εισόδου και κάθε σύνδεση έχει ένα συσχετισμένο βάρος. Ο υπολογισμός του σταθμισμένου αθροίσματος ενέχει τον πολλαπλασιασμό των εισόδων με τα αντίστοιχα βάρη και την άθροιση των τιμών που προκύπτουν. Μαθηματικά ο υπολογισμός αυτός γράφεται ως εξής:

$$z = (x_1 \times w_1) + (x_2 \times w_2) + \dots + (x_n \times w_n) \quad (3.1)$$



Εικόνα 3.8: Μαθηματικό μοντέλο λειτουργίας του τεχνητού νευρώνα [48].

Ο υπολογισμός αυτός μπορεί επίσης να γραφτεί σε πιο συμπαγή μαθηματική μορφή ως:

$$z = \sum_{i=1}^n x_i \times w_i \quad (3.2)$$

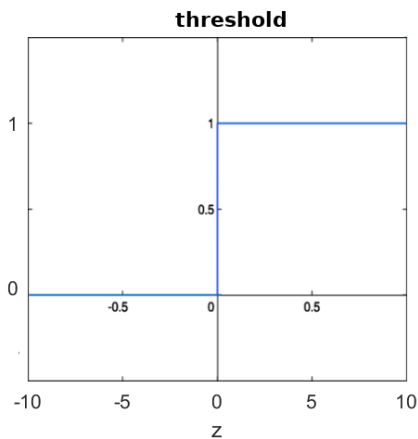
Το δεύτερο στάδιο της επεξεργασίας μέσα σε ένα νευρώνα είναι η μεταβίβαση του αποτελέσματος του σταθμισμένου αθροίσματος, της z τιμής, μέσω μίας συνάρτησης ενεργοποίησης. Οι εικόνες που ακολουθούν εμφανίζουν σχηματικά μία σειρά πιθανών συναρτήσεων ενεργοποίησης. Η Εικόνα 3.9 ορίζει μία συνάρτηση ενεργοποίησης κατωφλίου. Οι συναρτήσεις ενεργοποίησης κατωφλίου ήταν κοινές στην αρχική έρευνα νευρωνικών δικτύων. Οι Εικόνες 3.12 και 3.11 απεικονίζουν τη λογιστική συνάρτηση και τη συνάρτηση $\tanh$. Οι μονάδες που έκαναν χρήση αυτών των λειτουργιών ενεργοποίησης ήταν δημοφιλείς σε δίκτυα μέχρι πολύ πρόσφατα. Η Εικόνα 3.10 εμφανίζει το γράφημα συνάρτησης ενεργοποίησης ανόρθωσης (ή θετικής γραμμικής συνάρτησης). Αυτή η συνάρτηση ενεργοποίησης είναι πολύ δημοφιλής στα σύγχρονα δίκτυα βαθιάς μάθησης. Το 2011 μάλιστα αποδείχθηκε ότι η συνάρτηση ενεργοποίησης ανόρθωσης επιτρέπει καλύτερη εκμάθηση σε βαθιά δίκτυα [35].

Οι εξισώσεις των αναφερόμενων συναρτήσεων είναι οι εξής:

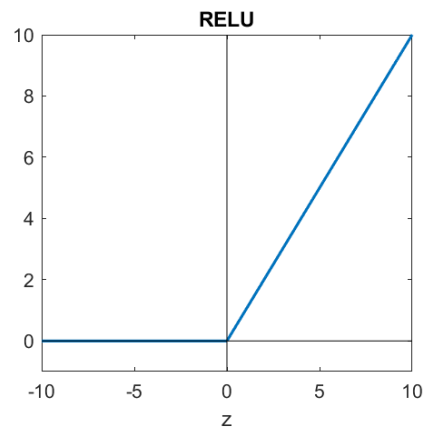
- $\text{threshold}(z) = \begin{cases} 1 & z \geq 0 \\ 0 & z < 0 \end{cases}$
- $\text{logistic}(z) = \sigma(z) = \frac{1}{1+e^{-z}}$

- $\tanh(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}}$

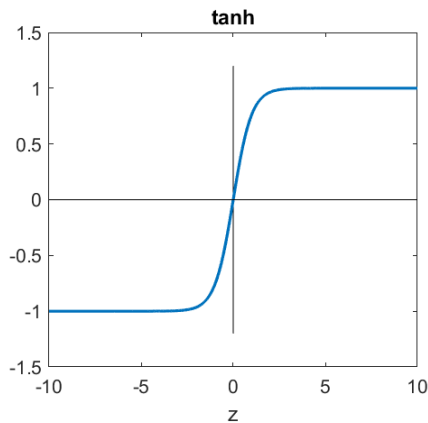
- $\text{rectifier}(z) = \max(0, z) = \begin{cases} 0 & z \leq 0 \\ z & z > 0 \end{cases}$



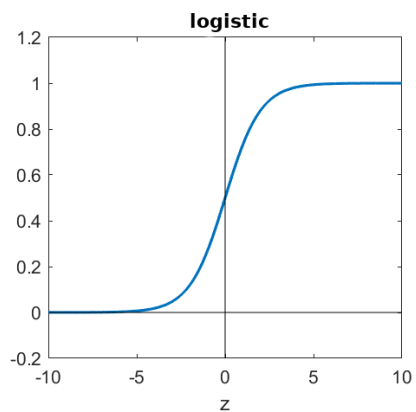
Εικόνα 3.9: Συνάρτηση ενεργοποίησης με κατώφλι.



Εικόνα 3.10: Συνάρτηση ενεργοποίησης RELU.



Εικόνα 3.11: Συνάρτηση ενεργοποίησης tanh.



Εικόνα 3.12: Συνάρτηση ενεργοποίησης logistic.

Οι Εικόνες 3.11 και 3.12 δείχνουν ότι ούτε η συνάρτηση $\tanh$ ούτε η logistic είναι γραμμικές συναρτήσεις. Όλες μάλιστα οι συναρτήσεις ενεργοποίησης εφαρμόζουν μη γραμμική αντιστοίχιση στην έξοδο του σταθμισμένου αθροίσματος. Στην πραγματικότητα, είναι η εισαγωγή μη γραμμικής αντιστοίχισης στην επεξεργασία ενός νευρώνα που είναι και ο λόγος για τον οποίο χρησιμοποιούνται οι συναρτήσεις ενεργοποίησης.

Η συνέπεια αυτού είναι ότι εάν όλοι οι νευρώνες εντός ενός δικτύου περιορίζονταν σε γραμμικές αντιστοιχίσεις το συνολικό δίκτυο θα περιοριζόταν σε μία γραμμική αντιστοίχιση

από τις εισόδους στις εξόδους. Ωστόσο, πολλές από τις σχέσεις στον κόσμο που ενδεχομένως πρέπει να μοντελοποιηθούν είναι μη γραμμικές, και αν αυτή η μοντελοποίηση γίνει χρησιμοποιώντας ένα γραμμικό μοντέλο, τότε το μοντέλο θα είναι πολύ ανακριβές.

Καταρχήν, η χρήση οποιασδήποτε μη γραμμικής συνάρτησης ως συνάρτησης ενεργοποίησης δίνει τη δυνατότητα σε ένα νευρωνικό δίκτυο να μάθει μία μη γραμμική αντιστοίχιση από τις εισόδους στις εξόδους. Ωστόσο, οι περισσότερες από τις συναρτήσεις ενεργοποίησης που απεικονίζονται στις Εικόνες 3.9, 3.11, 3.12 και 3.10 έχουν ωραίες μαθηματικές ιδιότητες που είναι χρήσιμες κατά την εκπαίδευση ενός νευρωνικού δικτύου και αυτό είναι που τις καθιστά δημοφιλείς στην έρευνα αυτών των δικτύων.

Το γεγονός ότι η εισαγωγή μη γραμμικότητας στην επεξεργασία των νευρώνων επιτρέπει στο δίκτυο να μάθει μία μη γραμμική αντιστοίχιση μεταξύ των εισόδων και των εξόδων είναι άλλη μία έκφανση του γεγονότος ότι η συνολική συμπεριφορά του δικτύου προκύπτει από τις αλληλεπιδράσεις της επεξεργασίας που πραγματοποιείται από τους μεμονωμένους νευρώνες μέσα στο δίκτυο. Τα νευρωνικά δίκτυα λύνουν προβλήματα χρησιμοποιώντας μία στρατηγική διαίρεσης και κατάκτησης. Καθένας από τους νευρώνες σε ένα δίκτυο επιλύει ένα στοιχείο του μεγαλύτερου προβλήματος και το συνολικό πρόβλημα επιλύεται συνδυάζοντας αυτές τις λύσεις.

Μέσα σε ένα νευρωνικό δίκτυο, ορισμένοι νευρώνες μπορεί να χρησιμοποιούν διαφορετικές συναρτήσεις ενεργοποίησης από άλλους νευρώνες στο δίκτυο. Γενικά ωστόσο, όλοι οι νευρώνες εντός ενός δεδομένου επιπέδου ενός δικτύου θα είναι του ίδιου τύπου (όλοι θα χρησιμοποιούν την ίδια συνάρτηση ενεργοποίησης). Οι νευρώνες οι αναφερόμενοι και ως μονάδες διακρίνονται με βάση τη συνάρτηση ενεργοποίησης ως μονάδες κατωφλίου, λογιστικές μονάδες και ανορθωμένες γραμμικές μονάδες (Rectified Linear Units, ReLUs). Ως μονάδες κατωφλίου ονομάζονται οι νευρώνες που χρησιμοποιούν μία συνάρτηση ενεργοποίησης κατωφλίου. Οι μονάδες που χρησιμοποιούν μία λογιστική συνάρτηση ενεργοποίησης είναι γνωστές ως λογιστικές μονάδες και οι νευρώνες που χρησιμοποιούν την ανορθωμένη γραμμική συνάρτηση ενεργοποίησης είναι γνωστές ως ReLUs. Για παράδειγμα, ένα δίκτυο μπορεί να έχει ένα επίπεδο ReLU συνδεδεμένο σε ένα επίπεδο λογιστικών μονάδων. Η απόφαση σχετικά με το ποιες συναρτήσεις ενεργοποίησης θα χρησιμοποιηθούν στους νευρώνες σε ένα δίκτυο λαμβάνεται από τον υπεύθυνο δεδομένων που σχεδιάζει το δίκτυο. Για να ληφθεί αυτή η απόφαση, ένας υπεύθυνος δεδομένων μπορεί να εκτελέσει μία σειρά πειραμάτων για να ελέγξει ποιες συναρτήσεις ενεργοποίησης δίνουν την καλύτερη απόδοση σε ένα σύνολο δεδομένων.

Τα στοιχεία ενός νευρωνικού δικτύου που έχουν ρυθμιστεί προσωπικά από τον υπεύθυνο δεδομένων πριν από τη διαδικασία μάθησης ονομάζονται υπερπαραμέτροι (hyperparameters). Ο όρος υπερπαραμέτρος (hyperparameter) χρησιμοποιείται για να περιγράψει τα ρυθμιζόμενα με μη αυτόματο τρόπο μέρη ενός μοντέλου, ώστε να διαφοροποιηθούν από τις παραμέτρους του μοντέλου που ορίζονται αυτόματα από τον αλγόριθμο μηχανικής μάθησης κατά τη

διάρκεια της εκπαίδευσης. Οι παράμετροι ενός νευρωνικού δικτύου είναι τα βάρη που χρησιμοποιούνται στους υπολογισμούς του σταθμισμένου συνόλου των νευρώνων του δικτύου.

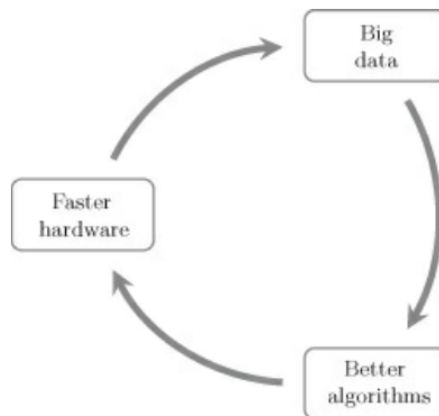
Υπάρχει ένας σημαντικός αριθμός τρόπων με τους οποίους ένα σύνολο νευρώνων μπορεί να συνδεθεί μαζί. Για παράδειγμα οι νευρώνες οργανώνονται σε επίπεδα και κάθε νευρώνας σε ένα επίπεδο μπορεί να συνδέεται άμεσα με όλους τους νευρώνες στο επόμενο επίπεδο του δικτύου. Αυτά τα δίκτυα είναι γνωστά ως δίκτυα feedforward επειδή δεν υπάρχουν βρόχοι στις συνδέσεις του δικτύου. Όλες οι συνδέσεις έχουν κατεύθυνση από την είσοδο προς την έξοδο. Επιπλέον, συνδέονται πλήρως, επειδή κάθε νευρώνας συνδέεται με όλους τους νευρώνες στο επόμενο επίπεδο. Είναι δυνατός και συχνά χρήσιμος ο σχεδιασμός και η εκπαίδευση δικτύων που δεν είναι feedforward ή/και δικτύων που δεν είναι πλήρως συνδεδεμένα. Όταν γίνεται σωστά, η προσαρμογή της αρχιτεκτονικής των δικτύων μπορεί να γίνει κατανοητή ως η ενσωμάτωση στην αρχιτεκτονική του δικτύου, πληροφοριών σχετικά με τις ιδιότητες του προβλήματος που το δίκτυο προσπαθεί να μάθει να μοντελοποιεί. Ένα πολύ επιτυχημένο παράδειγμα ενσωμάτωσης εξειδικευμένης γνώσης ενός πεδίου σε ένα δίκτυο είναι ο σχεδιασμός των συνελικτικών νευρωνικών δικτύων (CNN) για την ανίχνευση αντικειμένων σε εικόνες.

Αν και βελτιωμένες μέθοδοι αρχικοποίησης των βαρών και νέες συναρτήσεις ενεργοποίησης έχουν συμβάλει στην ανάπτυξη της βαθιάς μάθησης, τα τελευταία χρόνια οι δύο πιο σημαντικοί παράγοντες που εξελίσσουν τη βαθιά μάθηση είναι η επιτάχυνση της ισχύος των υπολογιστών και η τεράστια αύξηση μεγέθους των συνόλων δεδομένων. Από υπολογιστική άποψη, σημαντική εξέλιξη για τη βαθιά μάθηση σημειώθηκε στα τέλη της δεκαετίας του 2000 με την εκτεταμένη χρήση γραφικών μονάδων επεξεργασίας (GPUs) για την επιτάχυνση της εκπαίδευσης των δικτύων. Ένα νευρωνικό δίκτυο μπορεί να γίνει κατανοητό ως μία ακολουθία πολλαπλασιασμών πινάκων σε συνδυασμό με μη γραμμικές συναρτήσεις ενεργοποίησης, και οι GPUs είναι βελτιστοποιημένες για πολύ γρήγορο πολλαπλασιασμό πινάκων. Κατά συνέπεια οι GPUs είναι ιδανικό υλικό για να επιταχύνει την κατάρτιση νευρωνικών δικτύων και η χρήση τους έχει συμβάλει σημαντικά στην ανάπτυξη του πεδίου.

Ωστόσο, ακόμη και με αυτούς τους ισχυρότερους επεξεργαστές υπολογιστών, η βαθιά μάθηση δεν θα ήταν εφικτή εάν δεν υπήρχαν επίσης μαζικά σύνολα δεδομένων. Η ανάπτυξη του Διαδικτύου και των διαφόρων κοινωνικών μέσων, ο πολλαπλασιασμός των "έξυπνων" τηλεφώνων και των αισθητήρων του «διαδικτύου των πραγμάτων» είχε ως αποτέλεσμα τη σημαντική αύξηση της ποσότητας δεδομένων. Αυτή η αύξηση των δεδομένων ήταν εξαιρετικά σημαντική για τη βαθιά μάθηση επειδή τα μοντέλα νευρωνικών δικτύων κλιμακώνονται καλύτερα με περισσότερα δεδομένα (και στην πραγματικότητα μπορεί να δυσκολεύονται με μικρότερα σύνολα δεδομένων).

Αυτό με τη σειρά του οδήγησε στην ανάγκη για νέα (πιο πολύπλοκα) υπολογιστικά

μοντέλα προκειμένου να παραδοθούν αυτές οι νέες εφαρμογές. Ακολούθως, ο συνδυασμός μεγάλων δεδομένων και πιο περίπλοκων αλγορίθμων απαιτεί γρηγορότερο υλικό για να καταστεί δυνατή η διαχείριση του απαραίτητου υπολογιστικού φόρτου εργασίας. Η Εικόνα 3.13 απεικονίζει τον “κύκλο” μεταξύ μεγάλων δεδομένων, αλγοριθμικών βελτιώσεων (παραδείγματος χάρη καλύτερη αρχικοποίηση βαρών ή ReLUs), και ισχυρότερων υπολογιστικών επιδόσεων.



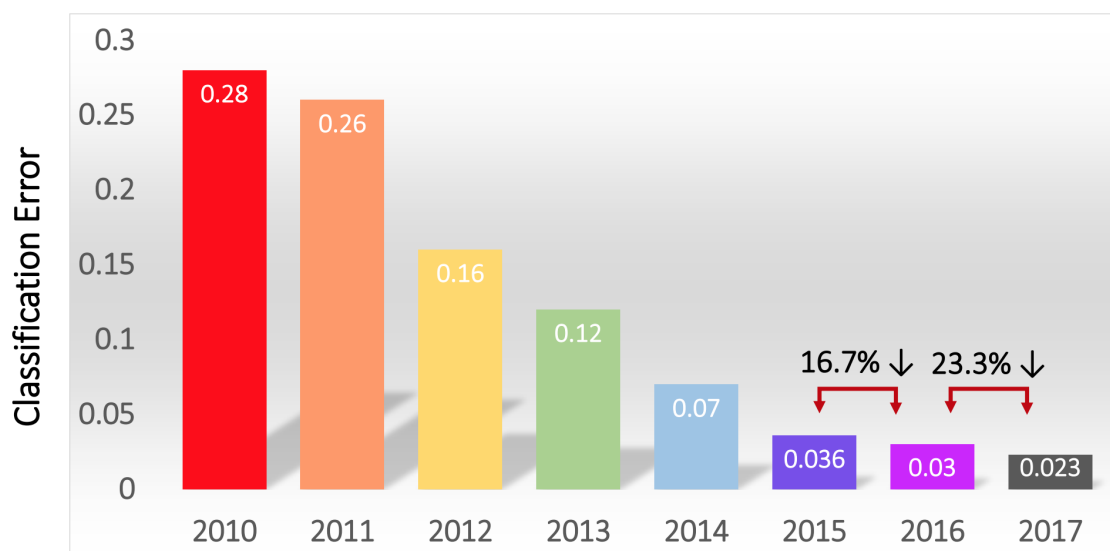
Εικόνα 3.13: Ο “κύκλος” μεταξύ μεγάλων δεδομένων, αλγοριθμικών βελτιώσεων, και ισχυρότερων υπολογιστικών επιδόσεων [51].

3.5 Συνελικτικά νευρωνικά δίκτυα

Στη δεκαετία του 1960, οι Hubel και Wiesel πραγματοποίησαν μία σειρά πειραμάτων στον οπτικό φλοιό των γατών [45], [46]. Στα τέλη της δεκαετίας του 1970 και στις αρχές της δεκαετίας του 1980, ο Kunihiro Fukushima εμπνεύστηκε από την ανάλυση του οπτικού φλοιού των γατών από τους Hubel και Wiesel και ανέπτυξε μία αρχιτεκτονική νευρωνικών δικτύων για την αναγνώριση οπτικών μοτίβων που ονομάστηκε neocognitron (Fukushima 1980). Ο σχεδιασμός του neocognitron βασίστηκε στην παρατήρηση ότι ένα δίκτυο αναγνώρισης εικόνων θα πρέπει να είναι σε θέση να αναγνωρίζει εάν υπάρχει ένα οπτικό χαρακτηριστικό σε μία εικόνα ανεξάρτητα από την τοποθεσία του στην εικόνα (spatially invariant visual feature detection). Το 1989, ο LeCun ανέπτυξε την αρχιτεκτονική συνελικτικών νευρωνικών δικτύων (Convolutional Neural Networks, CNN) ειδικά για την επεξεργασία εικόνων [57]. Η αρχιτεκτονική CNN μοιράστηκε πολλά από τα χαρακτηριστικά σχεδιασμού που βρέθηκαν στο neocognitron. Ωστόσο, ο LeCun έδειξε πώς αυτοί οι τύποι δικτύων θα μπορούσαν να εκπαιδεύονται χρησιμοποιώντας οπισθοδιάδοση (backpropagation).

Τα CNN έχουν αποδειχθεί εξαιρετικά επιτυχημένα στην επεξεργασία εικόνων και σε άλλες εργασίες. Ένα ιδιαίτερα διάσημο CNN είναι το δίκτυο AlexNet, το οποίο κέρδισε το ImageNet Large Scale Visual Recognition Challenge (ILSVRC) το 2012 [55]. Ο στόχος του διαγωνισμού ILSVRC, όπως αναφέρθηκε και νωρίτερα, είναι ο εντοπισμός αντικειμένων σε φωτογραφίες. Η επιτυχία του AlexNet στον διαγωνισμό ILSVRC δημιούργησε ενθουσιασμό για τα CNN και σε συνέχεια του AlexNet πολλές άλλες αρχιτεκτονικές βασισμένες σε CNN έχουν κερδίσει τον διαγωνισμό (Εικόνες 3.14, 3.15). Τα CNN παραμένουν ένας από τους πιο δημοφιλείς τύπους βαθιών νευρωνικών δικτύων.

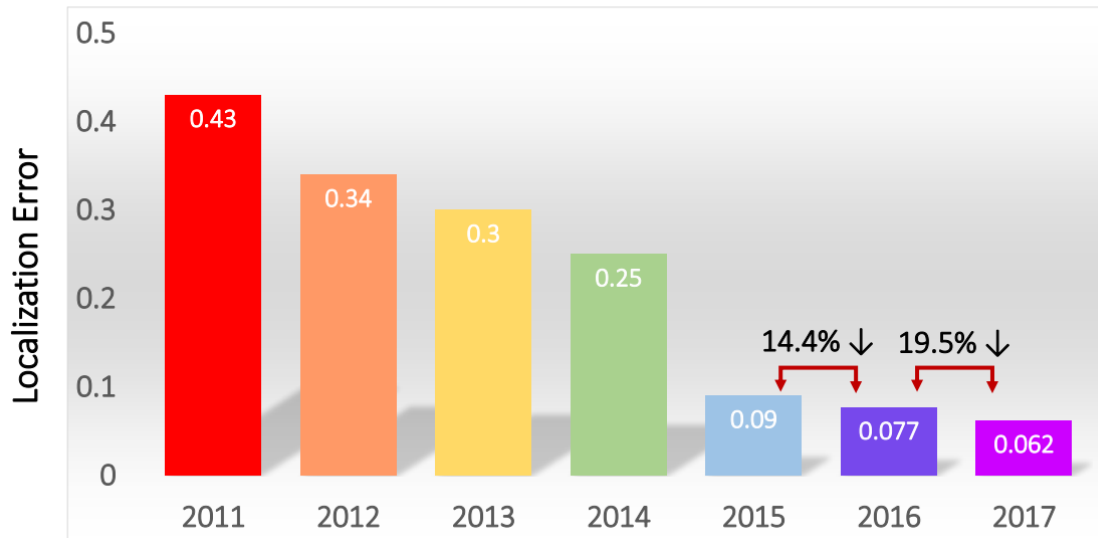
Classification Results (CLS)



Εικόνα 3.14: Στο ραβδόγραμμα απεικονίζεται το σφάλμα ταξινόμησης για τις νικήτριες ομάδες του διαγωνισμού ImageNet για τα χρόνια 2010 έως 2017. Από το 2012 και έπειτα τα CNNs κυριαρχούν στις υλοποιήσεις που κερδίζουν τον διαγωνισμό [71].

Η προσαρμογή της δομής ενός δικτύου στα ειδικά χαρακτηριστικά των δεδομένων ενός πεδίου μπορεί να μειώσει τον χρόνο εκπαίδευσης του δικτύου και να βελτιώσει την ακρίβεια του δικτύου. Η προσαρμογή μπορεί να γίνει με διάφορους τρόπους, όπως περιορισμός των συνδέσεων μεταξύ νευρώνων σε γειτονικά επίπεδα (αντί των πλήρως συνδεδεμένων επιπέδων), διαμοιρασμός των βαρών μεταξύ των νευρώνων ή εισαγωγή συνδέσεων με αντίστροφη ροή (backward) στο δίκτυο. Η προσαρμογή με αυτούς τους τρόπους αντιστοιχεί στην “εμφύτευση” της γνώσης του πεδίου στο δίκτυο. Μία άλλη, συναφής, οπτική είναι ότι βοηθά το δίκτυο περιορίζοντας το σύνολο των πιθανών συναρτήσεων που μπορεί να μάθει, καθοδηγώντας με αυτόν τον τρόπο το δίκτυο να βρει μία χρήσιμη λύση. Δεν είναι πάντοτε σαφές πώς να ταιριάζει μία δομή δικτύου σε έναν τομέα, αλλά για ορισμένους τομείς όπου τα δεδομένα έχουν μία πολύ τακτική δομή (παραδείγματος χάρη διαδοχικά δεδομένα όπως

Localization Results (LOC)



Εικόνα 3.15: Στο ραβδόγραμμα απεικονίζεται το σφάλμα τοποθέτησης για τις νικήτριες ομάδες του διαγωνισμού ImageNet (έτη 2010-2017) [71].

κείμενο ή δεδομένα σε πλέγμα όπως εικόνες) υπάρχουν γνωστές αρχιτεκτονικές δικτύου που έχουν αποδειχθεί επιτυχείς.

Τα συνελικτικά νευρωνικά δίκτυα (CNNs) σχεδιάστηκαν για εργασίες αναγνώρισης εικόνων και αρχικά εφαρμόστηκαν στην πρόκληση της αναγνώρισης χειρόγραφων ψηφίων [31], [57]. Ο βασικός σχεδιαστικός στόχος των CNNs ήταν να δημιουργηθεί ένα δίκτυο όπου οι νευρώνες στο αρχικό επίπεδο του δικτύου θα εξάγουν τα τοπικά οπτικά χαρακτηριστικά και οι νευρώνες σε επόμενα επίπεδα θα συνδυάζουν αυτά τα χαρακτηριστικά για να σχηματίσουν πιο πολύπλοκα χαρακτηριστικά. Ένα τοπικό οπτικό χαρακτηριστικό είναι ένα χαρακτηριστικό του οποίου η έκταση περιορίζεται σε ένα μικρό κομμάτι, ένα σύνολο γειτονικών εικονοστοιχείων, μίας εικόνας. Για παράδειγμα, όταν εφαρμόζονται στην εργασία αναγνώρισης προσώπου, οι νευρώνες στα αρχικά επίπεδα ενός CNN μαθαίνουν να ενεργοποιούνται σε απλά τοπικά χαρακτηριστικά (όπως γραμμές σε συγκεκριμένη γωνία ή τμήματα καμπυλών), νευρώνες βαθύτερα στο δίκτυο συνδυάζουν αυτά τα χαρακτηριστικά χαμηλού επιπέδου σε χαρακτηριστικά που αντιπροσωπεύουν μέρη του σώματος (όπως μάτια ή μύτη) και οι νευρώνες στα τελικά επίπεδα του δικτύου συνδυάζουν τις προηγούμενες ενεργοποιήσεις ώστε να αναγνωρίσουν συνολικά τα πρόσωπα σε μία εικόνα. Χρησιμοποιώντας αυτήν την προσέγγιση, ο θεμελιώδης στόχος στην αναγνώριση εικόνων είναι η εκμάθηση των λειτουργιών ανίχνευσης χαρακτηριστικών που μπορούν με ακρίβεια να αναγνωρίσουν την παρουσία ή την απουσία τοπικών οπτικών χαρακτηριστικών σε μία εικόνα. Η διαδικασία της μάθησης βρίσκεται στον πυρήνα των νευρωνικών δικτύων και επιτυγχάνεται με την αναζήτηση της κατάλληλης δέσμης βαρών για τις συνδέσεις στο δίκτυο. Τα CNNs μαθαίνουν

τις λειτουργίες ανίχνευσης χαρακτηριστικών για τοπικά οπτικά χαρακτηριστικά με αυτόν τον τρόπο. Ωστόσο, μία σχετική πρόκληση είναι ο σχεδιασμός της αρχιτεκτονικής του δικτύου, ώστε το δίκτυο να εντοπίζει την παρουσία ενός τοπικού οπτικού χαρακτηριστικού σε μία εικόνα ανεξάρτητα από το πού βρίσκεται στην εικόνα (translation invariant). Για παράδειγμα, ένα σύστημα αναγνώρισης προσώπου θα πρέπει να είναι σε θέση να αναγνωρίζει τη μορφή ενός ματιού σε μία εικόνα είτε το μάτι βρίσκεται στο κέντρο της εικόνας είτε στην πάνω δεξιά γωνία της εικόνας. Αυτή η ανάγκη αποτελεί πρωταρχική σχεδιαστική αρχή των CNNs για την επεξεργασία εικόνων [57]. Είναι δυνατό να σχεδιαστεί ένα δίκτυο αναγνώρισης χαρακτηριστικών ανεξαρτήτως του πού αυτά εμφανίζονται στην εικόνα (translation invariant), δημιουργώντας ένα σύνολο νευρώνων που μοιράζονται τα ίδια βάρη και έχοντας οργανωθεί έτσι ώστε κάθε νευρώνας να επιθεωρεί ένα διαφορετικό τμήμα της εικόνας, και συνολικά τα ληφθέντα από τους νευρώνες πεδία να καλύπτουν ολόκληρη την εικόνα.

Το σενάριο της αναζήτησης μίας εικόνας σε ένα σκοτεινό δωμάτιο με έναν φακό που έχει μία στενή δέσμη χρησιμοποιείται μερικές φορές για να εξηγήσει πώς ένα CNN ψάχνει μία εικόνα για τοπικά χαρακτηριστικά. Σε κάθε στιγμή μπορείτε να τοποθετήσετε τον φακό σε μία περιοχή της εικόνας και να επιθεωρήσετε αυτήν την τοπική περιοχή. Σε αυτό το μεταφορικό σενάριο, η περιοχή της εικόνας που φωτίζεται από τον φακό ανά πάσα στιγμή είναι ισοδύναμη με το λαμβανόμενο πεδίο ενός μόνο νευρώνα και έτσι η εστίαση του φακού σε μία θέση ισοδυναμεί με την εφαρμογή της λειτουργίας ανίχνευσης χαρακτηριστικών σε αυτήν την τοπική περιοχή. Η διαδικασία της διαδοχικής εστίασης σε μία εικόνα και η εφαρμογή σε κάθε θέση εστίασης της ίδιας συνάρτησης είναι η ουσία της συνέλιξης μίας συνάρτησης σε μία εικόνα. Μέσα σε ένα CNN, αυτή η διαδοχική αναζήτηση σε μία εικόνα υλοποιείται χρησιμοποιώντας ένα σύνολο νευρώνων που μοιράζονται βάρη και των οποίων η ένωση των δεκτικών πεδίων καλύπτει ολόκληρη την εικόνα.

Η Εικόνα 3.17 απεικονίζει τα διαφορετικά στάδια επεξεργασίας που βρίσκονται συχνά σε ένα CNN. Ο πίνακας 6×6 που βρίσκεται στα αριστερά του σχήματος αντιπροσωπεύει την εικόνα που είναι η είσοδος στο CNN. Ο πίνακας 4×4 που βρίσκεται ακριβώς δεξιά της εισόδου αντιπροσωπεύει ένα επίπεδο νευρώνων που μαζί ερευνούν ολόκληρη την εικόνα για την παρουσία ενός συγκεκριμένου τοπικού χαρακτηριστικού. Κάθε νευρώνας σε αυτό το επίπεδο συνδέεται με ένα διαφορετικό 3×3 πεδίο υποδοχής (περιοχή) στην εικόνα και όλοι οι νευρώνες εφαρμόζουν τον ίδιο πίνακα βαρών στις εισόδους τους.

Το πεδίο υποδοχής του νευρώνα $[0,0]$ (πάνω αριστερά) στο πρώτο αυτό επίπεδο συνέλιξης σημειώνεται με το γκρι τετράγωνο που καλύπτει μία περιοχή 3×3 στο πάνω αριστερό μέρος της εικόνας εισόδου. Τα διακεκομμένα βέλη από κάθε μία από τις θέσεις αυτής της γκριζας περιοχής αντιπροσωπεύουν τις εισόδους στον νευρώνα $[0,0]$. Το πεδίο υποδοχής του γειτονικού νευρώνα $[0,1]$ υποδεικνύεται από το τετράγωνο 3×3 , το έντονα περιγεγραμμένο στην εικόνα εισόδου. Παρατηρήστε ότι τα δεκτικά πεδία αυτών των δύο νευρώνων

επικάλυπτονται. Το ποσό επικάλυψης των δεκτικών πεδίων ελέγχεται από μία υπερπαράμετρο που ονομάζεται μήκος του βήματος (stride length). Σε αυτήν την περίπτωση, το μήκος του βήματος είναι ένα, το οποίο σημαίνει ότι κάθε μετακίνηση στο επίπεδο, αντιστοιχεί σε ανάλογη μετατόπιση του δεκτικού από τον νευρώνα πεδίου. Εάν η τιμή της υπερπαράμετρου για το μήκος βήματος αυξηθεί, το ποσό επικάλυψης μεταξύ των δεκτικών πεδίων μειώνεται.

Τα δεκτικά πεδία και των δύο αυτών νευρώνων ($[0,0]$ και $[0,1]$) είναι πίνακες των τιμών των εικονοστοιχείων και τα βάρη που χρησιμοποιούνται από αυτούς τους νευρώνες είναι επίσης πίνακες. Στη μηχανική όραση, ο πίνακας των βαρών που εφαρμόζεται σε μία είσοδο είναι γνωστός ως πυρήνας (ή μάσκα συνέλιξης). Η λειτουργία της διαδοχικής διέλευσης ενός πυρήνα σε μία εικόνα και σε κάθε τοπική περιοχή, η στάθμιση κάθε εισόδου και η πρόσθεση του αποτελέσματος στους τοπικούς γείτονές της, είναι γνωστή ως συνέλιξη. Παρατηρήστε ότι μία λειτουργία συνέλιξης δεν περιλαμβάνει μία μη γραμμική συνάρτηση ενεργοποίησης (αυτό εφαρμόζεται σε μεταγενέστερο στάδιο της επεξεργασίας). Ο πυρήνας ορίζει τη λειτουργία ανίχνευσης χαρακτηριστικών που εφαρμόζουν όλοι οι νευρώνες στη συνέλιξη. Η συνέλιξη ενός πυρήνα σε μία εικόνα ισοδυναμεί με τη διέλευση ενός τοπικού ανιχνευτή οπτικών χαρακτηριστικών στην εικόνα και την καταγραφή όλων των θέσεων της εικόνας όπου το ζητούμενο οπτικό χαρακτηριστικό ήταν παρών. Το αποτέλεσμα από αυτήν τη διαδικασία είναι ένας χάρτης όλων των τοποθεσιών της εικόνας όπου εμφανίστηκε το σχετικό οπτικό χαρακτηριστικό. Για τον λόγο αυτόν, η έξοδος μίας διαδικασίας συνέλιξης είναι γνωστή και ως χάρτης χαρακτηριστικών. Όπως σημειώθηκε παραπάνω, η λειτουργία συνέλιξης δεν περιλαμβάνει μία μη γραμμική λειτουργία ενεργοποίησης (περιλαμβάνει μόνο ένα σταθμισμένο άθροισμα των εισόδων).

Συνεπώς, είναι τυπικό να εφαρμόζεται μία λειτουργία μη γραμμικότητας σε έναν χάρτη χαρακτηριστικών. Συχνά, αυτό γίνεται με την εφαρμογή ReLU σε κάθε θέση ενός χάρτη χαρακτηριστικών. Η ανορθωμένη γραμμική συνάρτηση ενεργοποίησης ορίζεται ως: $rectifier(z) = \max(0, z)$. Περνώντας μία ανορθωμένη γραμμική συνάρτηση ενεργοποίησης πάνω από έναν χάρτη χαρακτηριστικών απλά αλλάζει όλες τις αρνητικές τιμές σε 0. Στην Εικόνα 3.17, η διαδικασία ενημέρωσης ενός χάρτη χαρακτηριστικών με την εφαρμογή μίας ανορθωμένης γραμμικής συνάρτησης ενεργοποίησης σε κάθε ένα από τα στοιχεία του, αναπαρίσταται από το επίπεδο επισημασμένο ως “nonlinearity”.

Τα CNNs συχνά απορρίπτουν τις πληροφορίες τοποθεσίας, ευνοώντας τη γενίκευση της ικανότητας του δικτύου να κάνει ταξινόμηση εικόνων (image classification). Συνήθως, αυτό επιτυγχάνεται με τη μείωση της δειγματοληψίας του ενημερωμένου χάρτη χαρακτηριστικών χρησιμοποιώντας ένα επίπεδο ομαδοποίησης (pooling). Η ομαδοποίηση είναι παρόμοια με τη λειτουργία συνέλιξης που περιγράφεται παραπάνω, καθώς η ομαδοποίηση ενέχει την επανειλημμένη εφαρμογή της ίδιας συνάρτησης σε έναν χώρο εισόδου. Για την ομαδοποίηση, ο χώρος εισόδου είναι συχνά ένας χάρτης χαρακτηριστικών των οποίων τα στοιχεία έχουν ενημερωθεί χρησιμοποιώντας μία συνάρτηση ReLU.



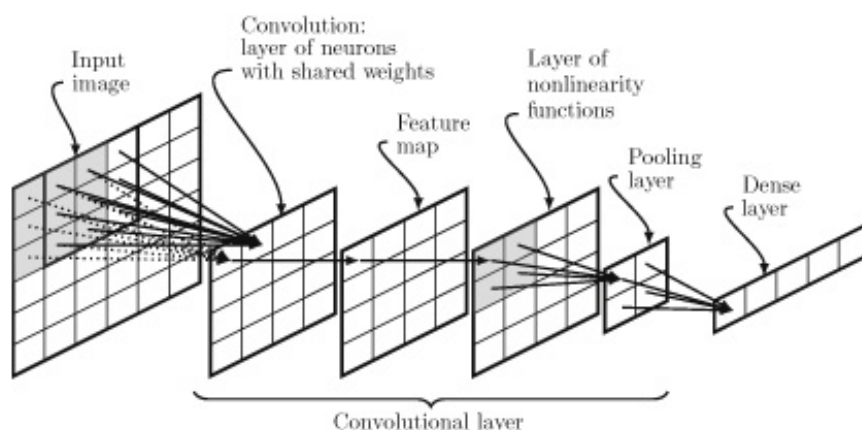
Εικόνα 3.16: Βήμα (*stride*) και περιθώριο (*padding*) σε διδιάστατα CNNs [81].

Χρησιμοποιούνται διάφορες συναρτήσεις ομαδοποίησης με πιο κοινή την max pooling, η οποία επιστρέφει τη μέγιστη τιμή οποιασδήποτε από τις εισόδους της. Ο υπολογισμός της μέσης τιμής των εισόδων χρησιμοποιείται επίσης ως συνάρτηση ομαδοποίησης.

Η ακολουθία λειτουργιών της εφαρμογής μίας συνέλιξης, ακολουθούμενη από την εφαρμογή μίας μη γραμμικής συνάρτησης στον χάρτη χαρακτηριστικών και στη συνέχεια τη μείωση της δειγματοληψίας, χρησιμοποιώντας τη λειτουργία ομαδοποίησης, είναι σχετικά τυπική στα περισσότερα CNNs. Συχνά αυτές οι τρεις λειτουργίες θεωρούνται από κοινού ότι ορίζουν ένα συνελκτικό επίπεδο σε ένα δίκτυο.

Με αυτόν τον τρόπο, ένα CNN υποστηρίζει την ανίχνευση οπτικών χαρακτηριστικών ανεξαρτήτως θέσης. Ωστόσο, αυτό έχει τον περιορισμό ότι η συνέλιξη μπορεί να αναγνωρίσει μόνο έναν τύπο χαρακτηριστικού. Τα CNN γενικεύουν πέρα από ένα χαρακτηριστικό, εκπαιδεύοντας παράλληλα πολλαπλά επίπεδα συνέλιξης (ή φίλτρα), με κάθε φίλτρο να χρησιμοποιεί έναν μοναδικό πυρήνα για τη λειτουργία ανίχνευσης χαρακτηριστικών. Επισημαίνεται ότι το επίπεδο συνέλιξης στην Εικόνα 3.17 απεικονίζει ένα μόνο φίλτρο.

Οι έξοδοι πολλαπλών φίλτρων μπορούν να ενσωματωθούν με διάφορους τρόπους. Ένας τρόπος για την ενσωμάτωση πληροφοριών από διαφορετικά φίλτρα είναι ο συνδυασμός των χαρτών χαρακτηριστικών, που δημιουργούνται από τα ξεχωριστά φίλτρα, σε έναν ενιαίο χάρτη πολλαπλών φίλτρων. Αυτός ο χάρτης πολλαπλών φίλτρων εισάγεται στη συνέχεια σε ένα επόμενο συνελκτικό επίπεδο. Ένας άλλος τρόπος για την ενσωμάτωση πληροφοριών από διαφορετικά φίλτρα είναι η χρησιμοποίηση ενός πυκνά συνδεδεμένου επιπέδου νευρώνων. Το τελικό επίπεδο στην Εικόνα 3.17 απεικονίζει ένα πυκνό επίπεδο. Αυτό το πυκνό επίπεδο λειτουργεί με τον ίδιο τρόπο όπως ένα τυπικό επίπεδο σε ένα πλήρως συνδεδεμένο feedforward δίκτυο. Κάθε νευρώνας στο πυκνό επίπεδο συνδέεται με όλα τα στοιχεία που εξάγονται από κάθε ένα από τα φίλτρα και κάθε νευρώνας μαθαίνει ένα σύνολο βαρών, μοναδικό για αυτόν, που εφαρμόζει στις εισόδους.



Εικόνα 3.17: Απεικονίσεις των διαφόρων σταδίων επεξεργασίας σε ένα συνελκτικό επίπεδο [51].

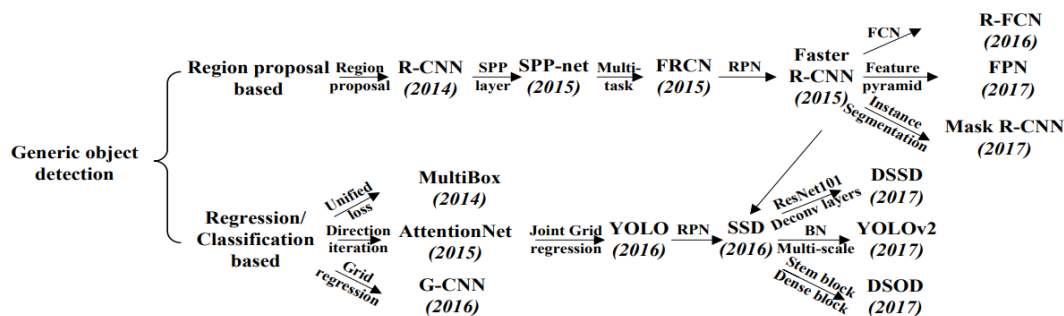
Το AlexNet CNN [55], το οποίο κέρδισε το 2012 τον διαγωνισμό ILSVRC, είχε πέντε συνελικτικά επίπεδα, ακολουθούμενα από τρία πυκνά επίπεδα. Το πρώτο επίπεδο συνέλιξης είχε ενενήντα έξι διαφορετικούς πυρήνες (ή φίλτρα) και περιλάμβανε μία λειτουργία μη γραμμικότητας ReLU και λειτουργία ομαδοποίησης. Το δεύτερο επίπεδο συνέλιξης είχε 256 πυρήνες και περιλάμβανε επίσης λειτουργίες μη γραμμικότητας ReLU και ομαδοποίησης. Το τρίτο, το τέταρτο και το πέμπτο συνελικτικό επίπεδο δεν περιελάμβαναν στάδιο μη γραμμικότητας ή ομαδοποίησης και είχαν 384, 384 και 256 πυρήνες, αντίστοιχα. Ακολουθώντας του πέμπτου συνελικτικού επιπέδου, το δίκτυο είχε τρία πυκνά επίπεδα με 4096 νευρώνες το καθένα. Συνολικά, το AlexNet είχε εξήντα εκατομμύρια βάρη και 650.000 νευρώνες. Αν και τα εξήντα εκατομμύρια βάρη είναι ένας μεγάλος αριθμός, το γεγονός ότι πολλοί από τους νευρώνες μοιράζονταν τα βάρη μείωνε στην ουσία τον αριθμό των βαρών στο δίκτυο. Αυτή η μείωση του αριθμού των απαιτούμενων βαρών είναι ένα από τα πλεονεκτήματα των δικτύων CNN. Το 2015, η Microsoft Research (Microsoft Corporation, Redmond, US-WA) ανέπτυξε ένα δίκτυο CNN με την ονομασία ResNet, το οποίο κέρδισε τον διαγωνισμό ILSVRC 2015 [40]. Η αρχιτεκτονική ResNet επέκτεινε την τυπική αρχιτεκτονική του CNN χρησιμοποιώντας συνδέσεις παράκαμψης (skip-connections). Μία σύνδεση παράκαμψης παίρνει την έξοδο από ένα επίπεδο στο δίκτυο και την τροφοδοτεί κατευθείαν σε ένα επίπεδο που μπορεί να είναι πολύ βαθύτερα στο δίκτυο. Χρησιμοποιώντας συνδέσεις παράκαμψης, είναι δυνατή η εκπαίδευση πολύ βαθιών δικτύων. Το μοντέλο ResNet μάλιστα, που αναπτύχθηκε από την Microsoft Research, είχε βάθος 152 επιπέδων.

3.6 Ανίχνευση αντικειμένων με συνελικτικά νευρωνικά δίκτυα

Τα πλαίσια των μεθόδων ανίχνευσης γενικών αντικειμένων μπορούν να κατηγοριοποιηθούν κυρίως σε δύο τύπους (Εικόνα 3.18). Ο ένας ακολουθεί την παραδοσιακή διαδικασία ανίχνευσης αντικειμένων, δημιουργώντας αρχικά προτάσεις περιοχών (region proposals) θέσης των αντικειμένων και στη συνέχεια ταξινομώντας κάθε πρόταση σε διαφορετικές κατηγορίες αντικειμένων. Οι τεχνικές που βασίζονται στις προτάσεις περιοχών, μία διαδικασία δύο σταδίων, ταιριάζουν σε κάποιο βαθμό με τον μηχανισμό προσοχής του ανθρώπινου εγκεφάλου, ο οποίος αφού πρώτα σαρώσει το περιβάλλον, επικεντρώνεται στις περιοχές που παρουσιάζουν ενδιαφέρον. Ο άλλος τύπος χειρίζεται την ανίχνευση αντικειμένων ως πρόβλημα παλινδρόμησης ή ταξινόμησης, υιοθετώντας ένα ενιαίο πλαίσιο για την επίτευξη τελικών αποτελεσμάτων (κατηγορίες και τοποθεσίες) άμεσα.

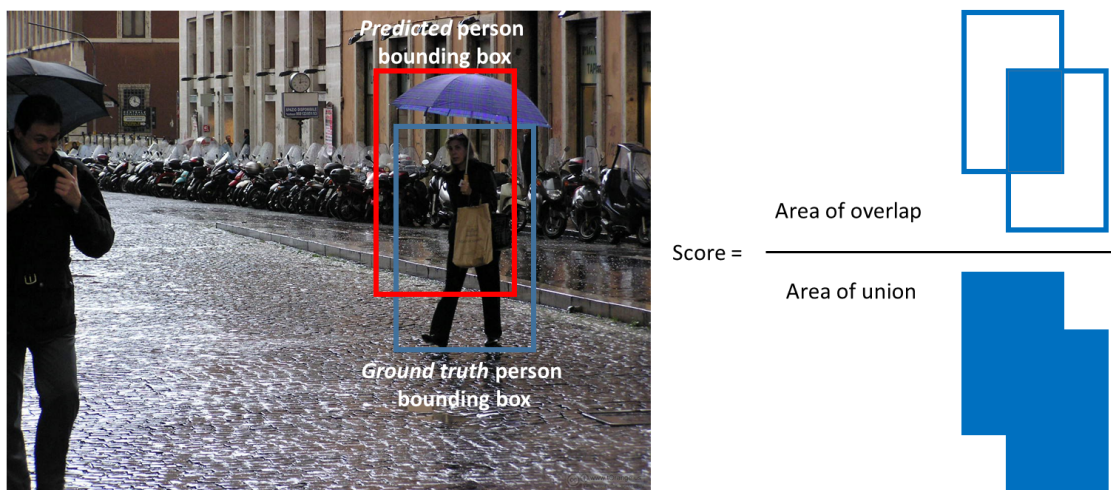
Οι μέθοδοι που βασίζονται στις προτάσεις περιοχών [34] περιλαμβάνουν κυρίως τις SPP-net [39], Fast R-CNN [33], Faster R-CNN [77], R-FCN [20], FPN [60] and Mask R-CNN [41]. Οι μέθοδοι παλινδρόμησης / ταξινόμησης περιλαμβάνουν τις MultiBox [26], AttentionNet [101], G-CNN [67], YOLO [75], SSD [61], YOLOv2 [74], DSSD [29] and DSOD [82].

Μέτρα αξιολόγησης. Το πρόβλημα της ανίχνευσης αντικειμένων είναι ένα πρόβλημα



Εικόνα 3.18: Δύο τύποι προσεγγίσεων για την ανίχνευση αντικειμένων με χρήση CNN [105].

παλινδρόμησης και ταξινόμησης. Πρώτα, για να αξιολογήσουμε την ακρίβεια της τοποθεσίας των προβλεπόμενων πλαισίων πρέπει να αφαιρεθούν τα πλαίσια με χαμηλή εμπιστοσύνη (συνήθως τα μοντέλα παράγουν πολύ περισσότερα πλαίσια από ότι τα πραγματικά αντικείμενα στην εικόνα). Στη συνέχεια γίνεται χρήση της Intersection over Union (IoU) περιοχής, που είναι μία τιμή μεταξύ 0 και 1. Η περιοχή αυτή αντιστοιχεί στην επικαλυπτόμενη περιοχή μεταξύ του προβλεπόμενου πλαισίου και του πραγματικού πλαισίου του αντικειμένου. Όσο υψηλότερο είναι το IoU, τόσο καλύτερη είναι η προβλεπόμενη θέση του πλαισίου για ένα δεδομένο αντικείμενο. Συνήθως, περιλαμβάνονται μόνο όσα προτεινόμενα πλαίσια έχουν ένα IoU μεγαλύτερο από κάποιο προκαθορισμένο κατώφλι.



Εικόνα 3.19: Μία γραφική αναπαράσταση του ορισμού του IoU [54].

Τα τελευταία χρόνια, η πιο συχνά χρησιμοποιούμενη αξιολόγηση για την ικανότητα ταξινόμησης των ανιχνευτών αντικειμένων είναι η Μέση Ακρίβεια (Average Precision, AP). Το AP είναι η μέση ακρίβεια στη δυαδική ταξινόμηση και συνήθως αξιολογείται ανά κλάση αντικειμένου. Για να συγκρίνουμε την απόδοση σε όλες τις κατηγορίες αντικειμένων, η μέση τιμή AP (mean Average Precision, mAP) υπολογίζεται για όλες τις κατηγορίες αντικειμένων

και χρησιμοποιείται συνήθως ως η τελική μέτρηση της απόδοσης. Αυτή είναι ένας αριθμός από 0 έως 100 και όσο ψηλότερη είναι η τιμή τόσο καλύτερη η ακρίβεια. Κάθε προτεινόμενο πλαίσιο οριοθέτησης έχει μία σχετική βαθμολογία (πιθανότητα του πλαισίου να περιέχει ένα αντικείμενο). Με βάση τις προβλέψεις υπολογίζεται μία καμπύλη Precision-Recall (PR) για κάθε κατηγορία, μεταβάλλοντας το κατώτατο επιτρεπόμενο όριο βαθμολογίας. Η μέση ακρίβεια (AP) είναι η περιοχή που περικλείεται από την καμπύλη PR. Πρώτα υπολογίζεται το AP για κάθε κλάση και στη συνέχεια υπολογίζεται κατά μέσο όρο για τις διάφορες κατηγορίες (mAP).

Η βαθμολογία mAP υπολογίζεται συνήθως για ένα σταθερό IoU. Το mAP που βασίζεται σε 0.5 IoU (mAP@0.5) έχει γίνει το καθιερωμένο μετρικό για τα προβλήματα ανίχνευσης αντικειμένων.

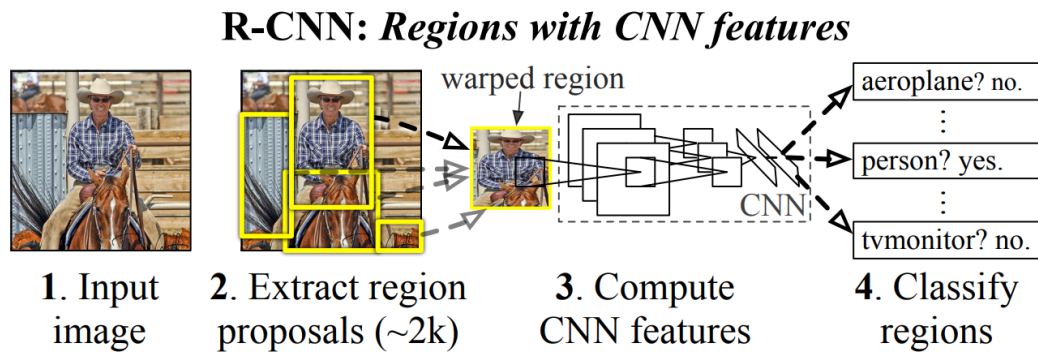
Παρακάτω παρατίθενται τα βασικά στοιχεία τεσσάρων βασικών τεχνικών ανίχνευσης αντικειμένων με τη χρήση συνελικτικών νευρωνικών δικτύων βασισμένων στις προτάσεις περιοχών θέσης αντικειμένου (R-CNN, Fast R-CNN, Faster R-CNN) και συνελικτικών δικτύων μονής λήψης (YOLO).

3.6.1 R-CNN

Ο αλγόριθμος R-CNN δημιουργήθηκε το 2014 από τον R. Girshick et al. [34]. Από τότε, η ανίχνευση αντικειμένων άρχισε να εξελίσσεται με μία άνευ προηγουμένου ταχύτητα. Το μοντέλο R-CNN συνδυάζει τη μέθοδο επιλεκτικής αναζήτησης για την εύρεση προτάσεων περιοχής και τη βαθιά μάθηση των συνελικτικών νευρωνικών δικτύων για την εύρεση της κλάσης του αντικειμένου στις περιοχές αυτές. Το R-CNN υιοθετεί τη μέθοδο επιλεκτικής αναζήτησης (selective search) για τη δημιουργία δύο χιλιάδων προτάσεων θέσης αντικειμένου για κάθε εικόνα [91]. Η μέθοδος βασίζεται σε μία απλή, από κάτω προς τα πάνω, ομαδοποίηση και επιτυγχάνει να παρέχει, γρήγορα, ακριβή υποψήφια πλαίσια αντικειμένων, μειώνοντας τον χώρο αναζήτησης στην ανίχνευση αντικειμένων [28, 23].

Οι περιοχές αυτές, αφού ανασχηματιστούν σε ένα σταθερό μέγεθος εικόνας (το μέγεθος που απαιτείται από το CNN), δίνονται ως είσοδος σε ένα CNN για την εξαγωγή χαρακτηριστικών από αυτές και συγκεκριμένα ένα διάνυσμα χαρακτηριστικών 4096 διαστάσεων. Στη συνέχεια ο χάρτης των χαρακτηριστικών δίνεται ως είσοδος στα πλήρως συνδεδεμένα επίπεδα (fully connected layers) του νευρωνικού δικτύου για την ταξινόμηση που αντικειμένου και την εύρεση του περιγεγραμμένου του πλαισίου. Για την πρόβλεψη της παρουσίας ενός αντικειμένου μέσα σε μία περιοχή, αλλά και για την πρόβλεψη της κλάσης του, το διάνυσμα των χαρακτηριστικών τροφοδοτείται σε πολλαπλούς ταξινομητές για να υπολογιστεί η πιθανότητα να ανήκουν σε κάθε κατηγορία. Κάθε μία από αυτές τις κλάσεις έχει έναν ταξινομητή SVM εκπαιδευμένο να συμπεράνει μία πιθανότητα για την ανίχνευση του αντικειμένου για ένα δεδομένο διάνυσμα χαρακτηριστικών. Αυτό το διάνυσμα τροφοδοτεί επίσης έναν γραμμικό παλινδρομητή (linear regressor) για να προσαρμόσει τα

πλαίσια οριοθέτησης για μία πρόταση περιοχής και έτσι να μειώσει τα σφάλματα εντοπισμού. Η Εικόνα 3.20 απεικονίζει το διάγραμμα ροής του R-CNN που αναλύεται σε τρία στάδια.



Εικόνα 3.20: Το διάγραμμα ροής του αλγορίθμου R-CNN αποτελείται από τρία στάδια: (1) εξαγωγή προτάσεων θέσης αντικειμένου, (2) υπολογισμός χαρακτηριστικών για κάθε πρόταση με τη χρήση CNN, και (3) ταξινόμηση κάθε περιοχής σε εξειδικευμένες κλάσεις με γραμμικούς ταξινομητές SVM [34].

Προτάσεις περιοχών θέσης αντικειμένου. Υπάρχουν πολλές διαφορετικές προσεγγίσεις για τη δημιουργία προτάσεων θέσης αντικειμένου. Το R-CNN χρησιμοποιεί τον αλγόριθμο επιλεκτικής αναζήτησης (selective search) (Εικόνα 3.21). Η επιλεκτική αναζήτηση είναι μία προσέγγιση που προσπαθεί να ομαδοποιήσει τα εικονοστοιχεία και να δημιουργήσει προτάσεις με βάση τις δημιουργούμενες ομάδες.

Άλλες προσεγγίσεις χρησιμοποιούν πιο πολύπλοκα οπτικά χαρακτηριστικά που εξάγονται από την εικόνα για να προτείνουν περιοχές (παραδείγματος χάρη με βάση τα χαρακτηριστικά από ένα μοντέλο βαθιάς μάθησης) ή υιοθετούν μία brute-force προσέγγιση παραγωγής περιοχών (όπως sliding window). Αυτές οι περιοχές δημιουργούνται χωρίς να λαμβάνονται υπόψη τα χαρακτηριστικά εικόνας.

Ένα σημαντικός συμβιβασμός που γίνεται με την παραγωγή προτάσεων περιοχής είναι ο αριθμός των περιοχών έναντι της υπολογιστικής πολυπλοκότητας. Όσο περισσότερες περιοχές προταθούν, τόσο πιο πιθανό θα είναι να βρεθεί το υπό ανίχνευση αντικείμενο. Από την άλλη πλευρά, εάν παραχθούν όλες οι πιθανές προτάσεις σε μία εικόνα, δεν θα είναι δυνατή η εκτέλεση των ανιχνευτών αντικειμένων σε έναν λογικό χρόνο (σίγουρα όχι σε πραγματικό χρόνο). Σε ορισμένες περιπτώσεις, είναι πιθανό να χρησιμοποιηθούν πληροφορίες που αφορούν συγκεκριμένα προβλήματα, για να μειωθεί ο αριθμός των προτεινόμενων περιοχών.

Εξαγωγή χαρακτηριστικών βασισμένη στα συνελικτικά νευρωνικά δίκτυα. Σε αυτό το στάδιο, κάθε προτεινόμενη περιοχή αλλάζει κλίμακα και περικόπτεται



Εικόνα 3.21: Εφαρμογή του αλγορίθμου *Selective Search* [91]. Πάνω: απεικόνιση του αποτελέσματος της κατάτμησης (*segmentation*) του αλγορίθμου. Κάτω: απεικόνιση των προτάσεων περιοχών θέσης του αλγορίθμου.

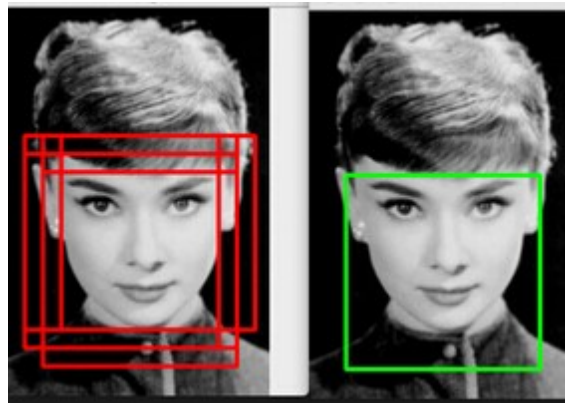
σε μία σταθερή ανάλυση [55], η οποία εισάγεται στο CNN για να εξάγει χαρακτηριστικά 4096 διαστάσεων ως τελική αναπαράσταση. Λόγω της μεγάλης ικανότητας μάθησης και της ιεραρχικής δομής των CNN, μπορεί να επιτευχθεί μία υψηλού επιπέδου, σημασιολογική και ισχυρή αναπαράσταση χαρακτηριστικών για κάθε πρόταση περιοχής.

Ο στόχος της εξαγωγής χαρακτηριστικών είναι ουσιαστικά η μετατροπή μίας εικόνας μεταβλητού μεγέθους σε ένα σταθερό σύνολο οπτικών χαρακτηριστικών. Τα μοντέλα ταξινόμησης εικόνων τυπικά κατασκευάζονται χρησιμοποιώντας ισχυρές μεθόδους εξαγωγής οπτικών χαρακτηριστικών. Είτε βασίζονται σε παραδοσιακές μεθόδους υπολογιστικής όρασης (όπως προσεγγίσεις με βάση κάποιο φίλτρο ή μέθοδο ιστογράμματος) είτε σε βαθιές μεθόδους μάθησης, ο στόχος είναι η εξαγωγή χαρακτηριστικών από την εικόνα εισόδου, τέτοιων που να είναι αντιπροσωπευτικά για το εκάστοτε πρόβλημα, και η χρήση των χαρακτηριστικών για τον καθορισμό της κλάσης της εικόνας. Στο πλαίσιο ανίχνευσης αντικειμένων, χρησιμοποιούνται συνήθως τυποποιημένα μοντέλα ταξινόμησης εικόνων για να εξαγάγουν οπτικά χαρακτηριστικά. Προκειμένου όμως να βελτιωθούν τα μοντέλα πρόβλεψης, συχνά δοκιμάζονται διάφορα μοντέλα ταξινόμησης.

Ταξινόμηση και εντοπισμός. Τα διάφορα προτεινόμενα πλαίσια θέσης αντικειμένων, βαθμολογούνται με κάποια τιμή, χρησιμοποιώντας τους προ-εκπαιδευμένους γραμμικούς ταξινομητές SVM για πολλαπλές κλάσεις. Οι βαθμολογημένες περιοχές στη

συνέχεια ρυθμίζονται με παλινδρόμηση (bounding-box regression) και φιλτράρονται με τον Non-Maximum Suppression (NMS) αλγόριθμο για να παράξουν τα τελικά πλαίσια θέσης αντικειμένων.

Η γενική ιδέα του NMS είναι να μειωθεί ο αριθμός των ανιχνεύσεων σε μία εικόνα στον πραγματικό αριθμό αντικειμένων που υπάρχουν. Αν το αντικείμενο στο πλαίσιο είναι αρκετά μεγάλο και έχουν δημιουργηθεί περισσότερες από 2000 προτάσεις αντικειμένων, είναι πολύ πιθανό ότι μερικές από αυτές θα έχουν σημαντική αλληλεπικάλυψη μεταξύ τους και με το αντικείμενο.



Εικόνα 3.22: Οπτική απεικόνιση της γενικής ιδέας της μείωσης του αριθμού ανιχνεύσεων σε μία εικόνα με την εφαρμογή του αλγορίθμου Non-Maximum Suppression (NMS) [73].

Ο αλγόριθμος NMS παίρνει ως είσοδο μία λίστα προτεινόμενων πλαισίων B , τους αντίστοιχους βαθμούς εμπιστοσύνης S και ένα κατώφλι επικάλυψης N και δίνει ως έξοδο μία νέα φιλτραρισμένη λίστα προτεινόμενων πλαισίων D . Ο αλγόριθμος μπορεί να συνοψιστεί ως:

- Επιλογή της πρότασης με την υψηλότερη βαθμολογία εμπιστοσύνης, αφαίρεση από τη λίστα B και προσθήκη της στην τελική λίστα προτάσεων D (αρχικά η D είναι άδεια).
- Σύγκριση της επιλεγμένης πρότασης με όλες τις προτάσεις. Υπολογισμός της τιμής του IoU της παρούσας πρότασης προς κάθε άλλη πρόταση. Εάν η τιμή IoU είναι μεγαλύτερη από το κατώτατο όριο N , η πρόταση αφαιρείται από τη λίστα B .
- Έπειτα, επιλέγεται η πρόταση με την υψηλότερη εμπιστοσύνη από τις υπόλοιπες προτάσεις στη λίστα B , αφαιρείται από αυτήν και προστίθεται στη λίστα D .
- Υπολογισμός της τιμής του IoU της παρούσας πρότασης προς κάθε άλλη πρόταση της λίστας B και εξάλειψη των πλαισίων που έχουν υψηλότερο IoU από το όριο.
- Η διαδικασία αυτή επαναλαμβάνεται έως ότου δεν υπάρχουν άλλες προτάσεις στη λίστα B .

Παρά τις βελτιώσεις του αλγορίθμου σε σχέση με τις παραδοσιακές μεθόδους και τη σημασία της εισαγωγής του CNN για την ανίχνευση αντικειμένων στην πράξη, εξακολουθεί να έχει ορισμένα μειονεκτήματα.

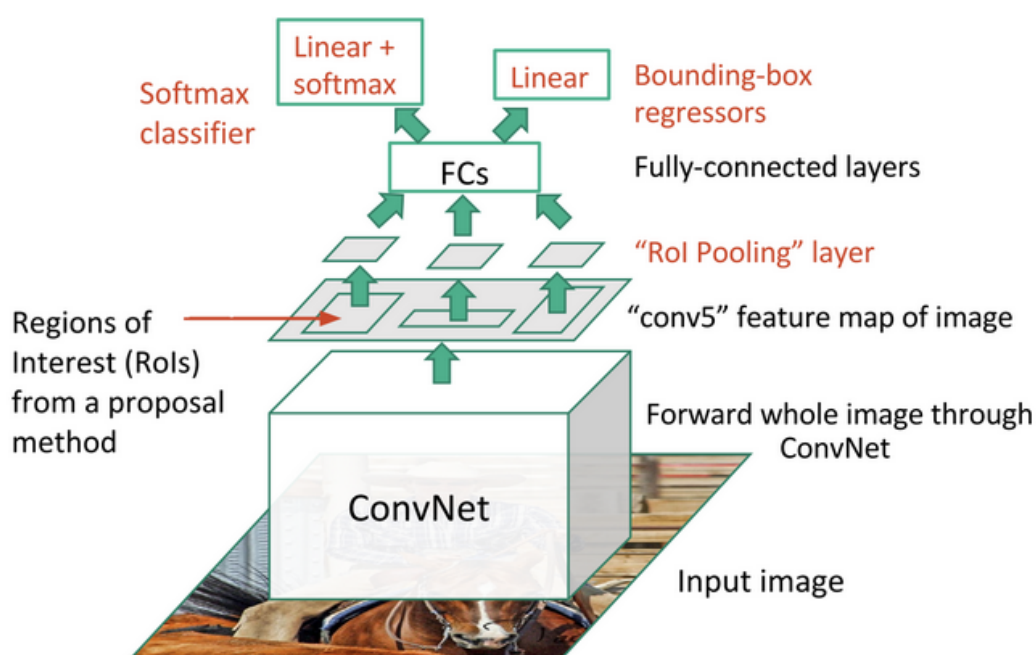
- Λόγω της ύπαρξης του πλήρως συνδεδεμένου επιπέδου, το CNN απαιτεί μία εικόνα εισόδου σταθερού μεγέθους (για παράδειγμα 227×227), η οποία οδηγεί άμεσα στον επανυπολογισμό ολόκληρου του CNN για κάθε αξιολογημένη περιοχή, κάτι που έχει ως αποτέλεσμα μεγάλους χρόνους εκπαίδευσης.
- Η εκπαίδευση των R-CNN είναι μία διαδικασία πολλών σταδίων. Αρχικά, το συνελικτικό νευρωνικό δίκτυο ρυθμίζεται και εισάγονται σε αυτό οι προτάσεις περιοχών. Στη συνέχεια, ο ταξινομητής Softmax που εκπαιδεύτηκε από τη ρύθμιση, αντικαθίσταται από έναν SVM ταξινομητή για την ταξινόμηση των χαρακτηριστικών που εξήχθησαν από το συνελικτικό δίκτυο. Τέλος, ρυθμίζονται (με παλινδρόμηση) τα πλαίσια των προτεινόμενων περιοχών.
- Η εκπαίδευση είναι δαπανηρή σε χώρο και χρόνο. Τα χαρακτηριστικά που εξάγονται από τις διάφορες προτάσεις περιοχής αποθηκεύονται στον δίσκο. Θα χρειαστεί πολύς χρόνος για την επεξεργασία ενός σχετικά μικρού εκπαιδευτικού συνόλου σε πολύ βαθιά δίκτυα (όπως για παράδειγμα το VGG16) ενώ ταυτόχρονα απαιτείται μεγάλη μνήμη αποθήκευσης για το σύνολο αυτών των χαρακτηριστικών.
- Αν και η επιλεκτική αναζήτηση μπορεί να δημιουργήσει προτάσεις περιοχών με σχετικά υψηλές ανακλήσεις, οι περισσότερες προτάσεις περιοχής που έχουν ληφθεί εξακολουθούν να είναι περιττές και αυτή η διαδικασία είναι χρονοβόρα (περίπου δύο δευτερόλεπτα για να εξαχθούν 2000 προτάσεις περιοχής).

3.6.2 Fast R-CNN

Το νευρωνικό δίκτυο Fast R-CNN, που αναπτύχθηκε το 2015 από τους R. Girshick και συνεργάτες [33], επιλύει το πρόβλημα των R-CNN εφαρμόζοντας τη μέθοδο εξαγωγής χαρακτηριστικών απευθείας στην αρχική εικόνα. Παράλληλα, με χρήση μίας μεθόδου παραγωγής προτεινόμενων περιοχών ενδιαφέροντος, εξάγονται οι περιοχές που συνδυάζονται με τον χάρτη χαρακτηριστικών της εικόνας. Το αποτέλεσμα είναι η παραγωγή των χαρτών χαρακτηριστικών όλων των περιοχών ενδιαφέροντος σε μόλις δύο βήματα. Οι χάρτες αυτοί υπόκεινται πρώτα σε μία τεχνική που ονομάζεται ομαδοποίηση περιοχών ενδιαφέροντος (Regions of Interest (RoI) pooling) για να αποκτήσουν ίσες διαστάσεις και στη συνέχεια δίνονται ως είσοδος στα πλήρως συνδεδεμένα επίπεδα του νευρωνικού δικτύου για την ταξινόμηση και την εύρεση του πλαισίου. Το Fast R-CNN, ακολουθώντας την παραπάνω διαδικασία, αποφεύγει την εφαρμογή της συνάρτησης εξαγωγής χαρακτηριστικών για κάθε

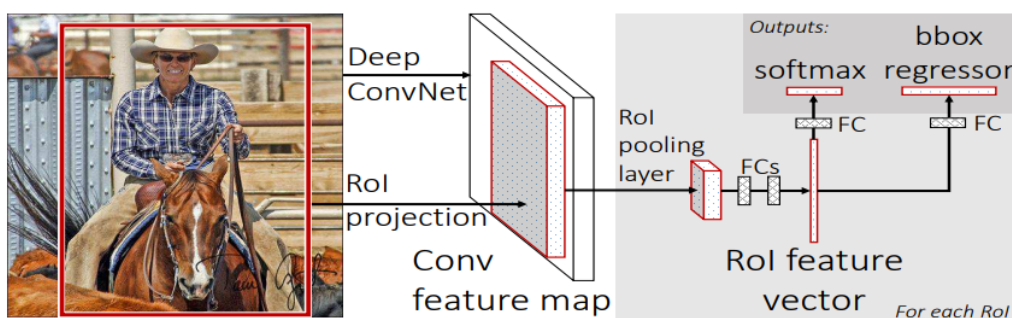
περιοχή ενδιαφέροντος με αποτέλεσμα ο συνολικός χρόνος εκτέλεσης της μεθόδου για την ανίχνευση του αντικειμένου να ελαττώνεται σημαντικά.

Ο στόχος του Fast R-CNN είναι να μειώσει τον χρόνο που σχετίζεται με τον μεγάλο αριθμό μοντέλων που είναι απαραίτητα για την ανάλυση όλων των προτάσεων περιοχής. Ένα κύριο CNN με πολλαπλά συνελκτικά στρώματα παίρνει ολόκληρη την εικόνα ως είσοδο αντί να χρησιμοποιεί CNN για κάθε πρόταση περιοχής (R-CNN). Οι περιοχές ενδιαφέροντος ανιχνεύονται με τη μέθοδο επιλεκτικής αναζήτησης που εφαρμόζεται στους παραγόμενους χάρτες χαρακτηριστικών. Το μέγεθος χαρτών χαρακτηριστικών μειώνεται χρησιμοποιώντας ένα επίπεδο RoI pooling για την απόκτηση έγκυρων περιοχών ενδιαφέροντος με σταθερό ύψος και πλάτος ως υπερπαραμέτρους. Κάθε περιοχή ενδιαφέροντος τροφοδοτείται στα πλήρως συνδεδεμένα επίπεδα δημιουργώντας ένα διάνυσμα χαρακτηριστικών. Το διάνυσμα χρησιμοποιείται για την πρόβλεψη του παρατηρούμενου αντικειμένου με έναν ταξινομητή Softmax ενώ ένας γραμμικός παλινδρομητής αναλαμβάνει την προσαρμογή των προτεινόμενων πλαισίων θέσης αντικειμένων. Τα καλύτερα Fast R-CNNs έχουν επιτύχει αξιοσημείωτα αποτελέσματα για τα σύνολα δεδομένων δοκιμής PASCAL VOC (mAp 68.8 %), PASCAL VOC 2010 (mAp 70.0 %) και PASCAL VOC 2012 (mAp 68.4 %).



Εικόνα 3.23: Ολόκληρη η εικόνα τροφοδοτεί ένα μοντέλο CNN για την ανίχνευση των περιοχών ενδιαφέροντος στους χάρτες χαρακτηριστικών. Κάθε περιοχή χωρίζεται χρησιμοποιώντας ένα επίπεδο ομαδοποίησης περιοχών ενδιαφέροντος (RoI Pooling) και τροφοδοτεί πλήρως συνδεδεμένα επίπεδα. Στη συνέχεια, χρησιμοποιείται ένας ταξινομητής Softmax για την ταξινόμηση του αντικειμένου και ένας γραμμικός παλινδρομητής για την προσαρμογή των συντεταγμένων του πλαισίου θέσης του αντικειμένου [33].

Η αρχιτεκτονική του Fast R-CNN παρουσιάζεται στην Εικόνα 3.24. Αρχικά, συνελκτικώς επίπεδα επεξεργάζονται όλη την εικόνα και παράγουν χάρτες χαρακτηριστικών. Στη συνέχεια, εξάγεται ένα διάνυσμα χαρακτηριστικών σταθερού μήκους για κάθε πρόταση περιοχής μέσω ενός επιπέδου ομαδοποίησης περιοχών ενδιαφέροντος. Ακολούθως, κάθε διάνυσμα χαρακτηριστικών τροφοδοτείται σε μία ακολουθία πλήρως συνδεδεμένων επιπέδων πριν τελικά διακλαδωθεί σε δύο επίπεδα εξόδου. Το ένα επίπεδο είναι υπεύθυνο για την ταξινόμηση του αντικειμένου ενώ το άλλο επίπεδο αναλαμβάνει τον βελτιωμένο επανακαθορισμό της θέσης των προτεινόμενων πλαισίων.

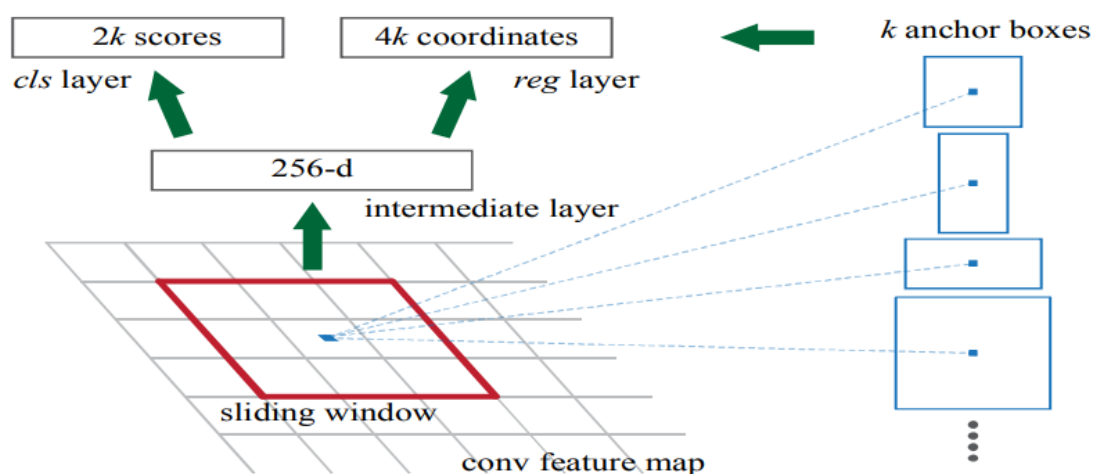


Εικόνα 3.24: Η αρχιτεκτονική του Fast R-CNN [33].

3.6.3 Faster R-CNN

Το νευρωνικό δίκτυο Faster R-CNN δημοσιεύτηκε το 2016 και αποτελεί μετεξέλιξη του R-CNN και του Fast R-CNN [77]. Όπως προαναφέρθηκε, στο πεδίο της ανίχνευσης αντικειμένων, σε αντίθεση με την απλή ταξινόμηση, απαιτείται και χωρικός εντοπισμός του εκάστοτε αντικειμένου στην εικόνα. Ως εκ τούτου, οι παραδοσιακές αρχιτεκτονικές των συνελκτικών δικτύων, που καταλήγουν σε πλήρως συνδεδεμένα επίπεδα με αριθμό ίσο με τις υποψήφιες κλάσεις, δεν μπορούσαν να εφαρμοστούν αυτούσιες, αφού ο αριθμός των υποψήφιων περιοχών και αντικειμένων δεν είναι σταθερός και μπορεί να ποικίλλει από εικόνα σε εικόνα. Στην πρώτη εκδοχή του Faster R-CNN, δηλαδή το R-CNN, η λύση στο πρόβλημα ήταν ο περιορισμός των υποψήφιων περιοχών ενδιαφέροντος σε δύο χιλιάδες ανά εικόνα. Η επιλογή των περιοχών αυτών γινόταν μέσω του αλγορίθμου επιλεκτικής αναζήτησης και ήταν εξαιρετικά χρονοβόρα. Η προσέγγιση που ακολουθήθηκε στο Fast R-CNN είναι ελαφρώς διαφορετική, με το σύστημα αντί να τροφοδοτεί τις υποψήφιες περιοχές ενδιαφέροντος στο δίκτυο, να παρέχει ολόκληρη την εικόνα εισόδου με στόχο την κατασκευή ενός χάρτη χαρακτηριστικών, αξιοποιώντας τις ιδιότητες της συνέλιξης. Χρησιμοποιώντας πάλι τη μέθοδο επιλεκτικής αναζήτησης στον χάρτη χαρακτηριστικών, εξάγονται οι περιοχές ενδιαφέροντος οι οποίες επανασχεδιάζονται σε σταθερό μέγεθος για να τροφοδοτηθούν στα επόμενα επίπεδα της αρχιτεκτονικής. Μέσω αυτής της βελτίωσης ο χρόνος εκπαίδευσης περιορίστηκε αισθητά, ωστόσο η διαδικασία εξαγωγής και περιορισμού του αριθμού των

περιοχών ενδιαφέροντος εξακολουθούσε να καταλαμβάνει μεγάλο μέρος της. Η δυσκαμψία στην εκπαίδευση του δικτύου οφείλεται σε μεγάλο βαθμό στον αλγόριθμο επιλεκτικής αναζήτησης για την επιλογή των υποψήφιων περιοχών. Η εξάλειψη αυτού του προβλήματος γίνεται με τον αλγόριθμο Faster R-CNN, όπου ο ρόλος του αλγορίθμου επιλεκτικής αναζήτησης επιτυγχάνεται με τη χρήση ενός νέου δικτύου, του Region Proposal Network (RPN) [77]. Το RPN ενεργεί με σχεδόν μηδενικό "κόστος" διανέμοντας χαρακτηριστικά ολόκληρης της εικόνας στο δίκτυο ανίχνευσης αντικειμένων. Το RPN παίρνει μία εικόνα αυθαίρετου μεγέθους για να παράξει ένα σύνολο από προτάσεις αντικειμένων, όντας υπεύθυνο για την επιλογή των υποψήφιων περιοχών. Μετά την εξαγωγή του χάρτη χαρακτηριστικών, το RPN καλείται να εντοπίσει τις περιοχές. Σε αντίθεση με την επιλεκτική αναζήτηση, που αποτελεί μία ντετερμινιστική διαδικασία που δεν επιδέχεται μάθηση, το RPN είναι ένα υποδίκτυο που επιδέχεται συνεχιζόμενης εκπαίδευσης. Βασικό εργαλείο στο έργο αυτό αποτελούν τα λεγόμενα anchor boxes, τα οποία είναι πλαίσια διαφορετικών διαστάσεων που εξετάζονται σε κάθε θέση. Το RPN ακολουθεί την τεχνική του κυλιόμενου παραθύρου (sliding window) καθώς σε κάθε περιοχή του χάρτη χαρακτηριστικών εξετάζει όλα τα anchor boxes και τα βαθμολογεί ως υποψήφιες περιοχές. Ουσιαστικά αποτελεί συνδυασμό δύο βασικών τεχνικών: πρώτον, της εξαγωγής διαφορετικών χαρτών χαρακτηριστικών για διαφορετικές διαστάσεις των εικόνων (πιο πλήρεις λύσεις αλλά χρονοβόρες), και δεύτερον, της χρήσης διαφορετικών φίλτρων που διατρέχουν την εικόνα μέσω της τεχνικής του κυλιόμενου παραθύρου.



Εικόνα 3.25: Το RPN στο Faster R-CNN. Τα προκαθορισμένα k anchor boxes συνελίσσονται με κάθε παράθυρο ολίσθησης για να παράγουν διανύσματα σταθερού μήκους, που λαμβάνονται από τα επίπεδα cls και reg, για να αποκτήσουν αντίστοιχες εξόδους [77].

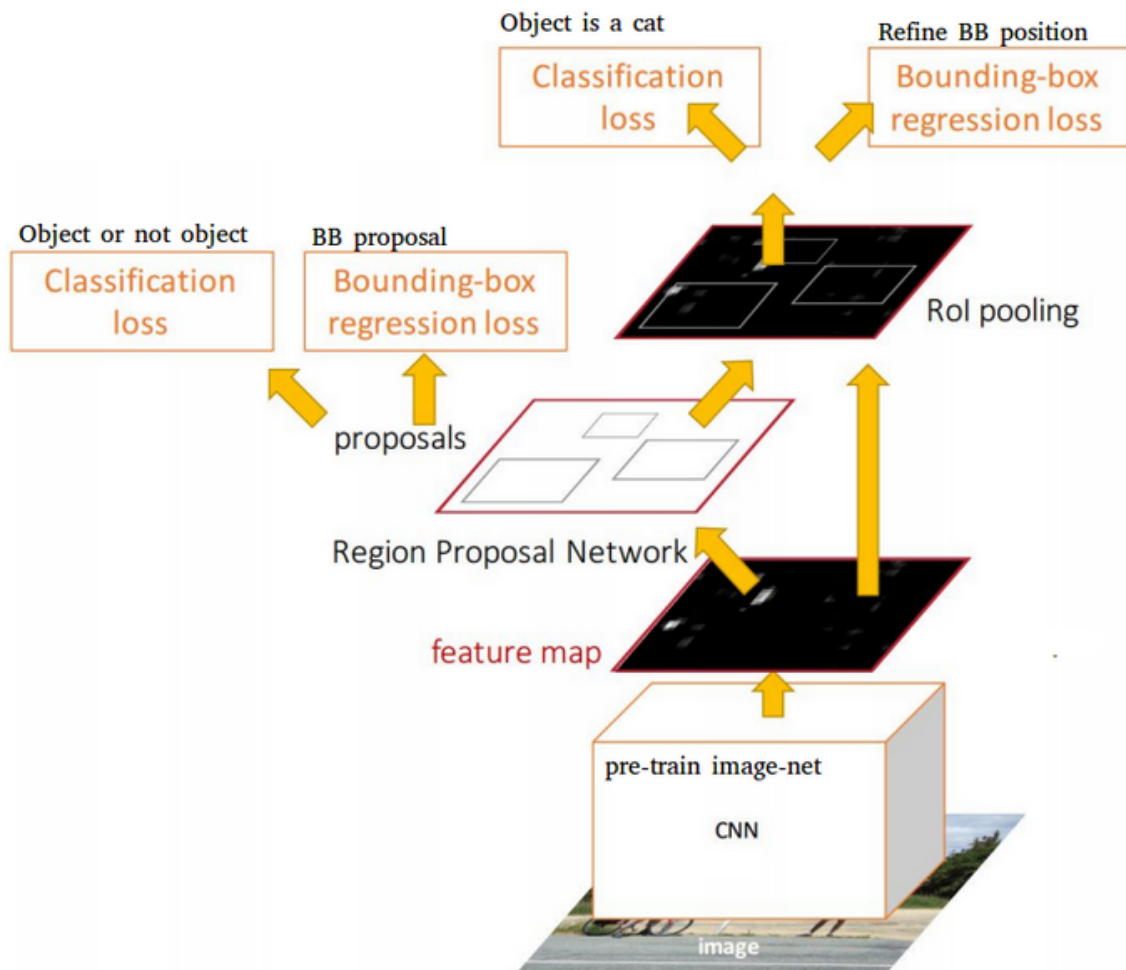
Στόχος του RPN είναι να αποφανθεί αν σε ένα υποψήφιο anchor box περιλαμβάνεται αντικείμενο ή όχι. Αντικείμενο θεωρείται ότι υπάρχει όταν η υποψήφια περιοχή εξασφαλίζει

ένα όριο IoU σε σχέση με τα αληθινά όρια ($\text{IoU} > 0.7$ θετικά δείγματα, $\text{IoU} < 0.3$ αρνητικά, τα υπόλοιπα δεν συνεισφέρουν). Επειδή αρκετές περιοχές είναι επικαλυπτόμενες, χρησιμοποιείται ο αλγόριθμος Non-Maximum Suppression (NMS) για να περιοριστούν οι υποψήφιες περιοχές.

Στον αλγόριθμο Fast R-CNN το CNN λαμβάνει ως είσοδο ολόκληρη την εικόνα και παράγει χάρτες χαρακτηριστικών (Εικόνα 3.25). Ένα παράθυρο μεγέθους 3×3 ολισθαίνει σε όλους τους χάρτες χαρακτηριστικών και εξάγει ένα διάνυσμα χαρακτηριστικών το οποίο συνδέεται με δύο πλήρως συνδεδεμένα επίπεδα. Το ένα είναι υπεύθυνο για την προσαρμογή των προτεινόμενων πλαισίων και το άλλο για την ταξινόμηση του πλαισίου. Οι προτάσεις περιοχών προτείνονται από τα πλήρως συνδεδεμένα επίπεδα. Καθορίζεται αρχικά το μέγιστο των περιοχών k , οπότε η έξοδος του επιπέδου παλινδρόμησης έχει μέγεθος $4k$ (συντεταγμένες των πλαισίων, ύψος και πλάτος) και η έξοδος του επιπέδου ταξινόμησης έχει μέγεθος $2k$ (για την ανίχνευση ενός αντικειμένου ή όχι στο πλαίσιο). Οι k προτάσεις περιοχών που ανιχνεύονται από το συρόμενο παράθυρο καλούνται άγκυρες. Αυτές οι άγκυρες και οι χάρτες χαρακτηριστικών που υπολογίζονται από το αρχικό μοντέλο CNN, τροφοδοτούνται στη συνέχεια σε ένα μοντέλο Fast R-CNN.

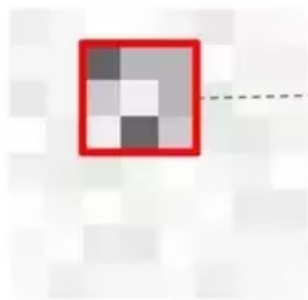
Αφού ολοκληρωθεί η διαδικασία επιλογής υποψήφιων περιοχών, το υπόλοιπο δίκτυο παραμένει παρόμοιο με το Fast R-CNN. Δηλαδή, η συνάρτηση σφάλματος προσομοιάζει αυτήν που χρησιμοποιείται στο Fast R-CNN και οφείλει να λαμβάνει υπόψη τόσο την ύπαρξη αντικειμένου όσο και τις συντεταγμένες των ορίων κάθε πλαισίου. Η διαδικασία εκπαίδευσης είναι η ακόλουθη. Πρώτα εκπαιδεύεται το δίκτυο RPN και οι προτάσεις τροφοδοτούνται στο κυρίως δίκτυο το οποίο επίσης εκπαιδεύεται. Στη συνέχεια, με ένα μοντέλο θετικής ανατροφοδότησης έχουμε επανάληψη της εκπαίδευσης του RPN δικτύου με βάση το δίκτυο που εκπαιδεύτηκε μόλις, διατηρώντας σταθερά τα συνελικτικά επίπεδα τα οποία πλέον είναι κοινά στα δύο επιμέρους δίκτυα. Τέλος, το κυρίως δίκτυο εκπαιδεύεται εκ νέου χρησιμοποιώντας τα νέα βάρη, διατηρώντας επίσης σταθερά τα συνελικτικά επίπεδα. Αρχικά, το Faster R-CNN χρησιμοποίησε ως κεντρικό δίκτυο υποστήριξης το VGG-16 [77, 83] το οποίο έχει 13 κοινά συνελικτικά επίπεδα. Πλέον, με την ανάπτυξη και διάδοση των GPUs και τη δυνατότητα βαθύτερων δικτύων, το Faster R-CNN μπορεί να υλοποιηθεί με χρήση δικτύων ResNet-50 και ResNet-101 [40] για ακόμα καλύτερα αποτελέσματα.

Χρησιμοποιώντας το RPN για την αποφυγή της μεθόδου επιλεκτικής αναζήτησης, επιταχύνονται οι διαδικασίες εκπαίδευσης και δοκιμών, καθιστώντας το Faster R-CNN έως και 34 φορές ταχύτερο από το Fast R-CNN. Με την πρόταση του Faster R-CNN, οι αρχιτεκτονικές των CNN για την ανίχνευση αντικειμένων με βάση τις προτάσεις περιοχών μπορούν να εκπαιδευτούν ενιαία από την αρχή ως το τέλος. Επίσης, επιτυγχάνεται δυνατότητα επεξεργασίας 5 FPS (Frames Per Second) σε μία GPU. Το RPN χρησιμοποιεί ένα προεκπαιδευμένο στο σύνολο δεδομένων ImageNet μοντέλο για ταξινόμηση και ρυθμίζεται περαιτέρω στο σύνολο δεδομένων PASCAL VOC. Στη συνέχεια, οι προκύπτουσες προτάσεις



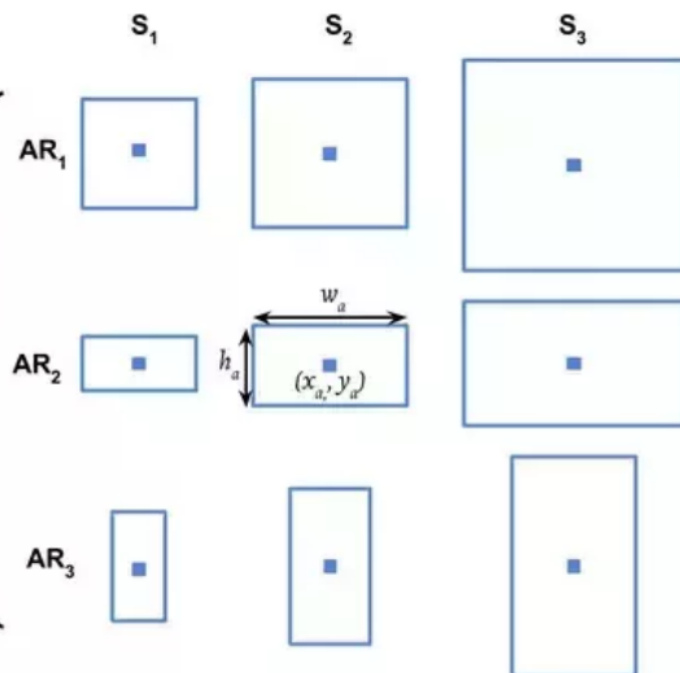
Εικόνα 3.26: Η αρχιτεκτονική του Faster R-CNN. Ολόκληρη η εικόνα τροφοδοτεί ένα CNN για την παραγωγή anchor boxes ως προτάσεις περιοχών αντικειμένου μαζί με τον αντίστοιχο βαθμό εμπιστοσύνης για την παρουσία αντικειμένου. Λαμβάνονται ως είσοδος οι χάρτες χαρακτηριστικών και οι προτάσεις περιοχής, και για κάθε πλαίσιο παράγονται οι πιθανότητες για την ύπαρξη ενός αντικειμένου και ρυθμίζεται διορθωτικά η θέση του πλαισίου [77].

Generate 9 anchors for each **sliding window** on conv. feature map



w_a : anchor's width
 h_a : anchor's height
 x_a, y_a : anchor's center

@vmirly



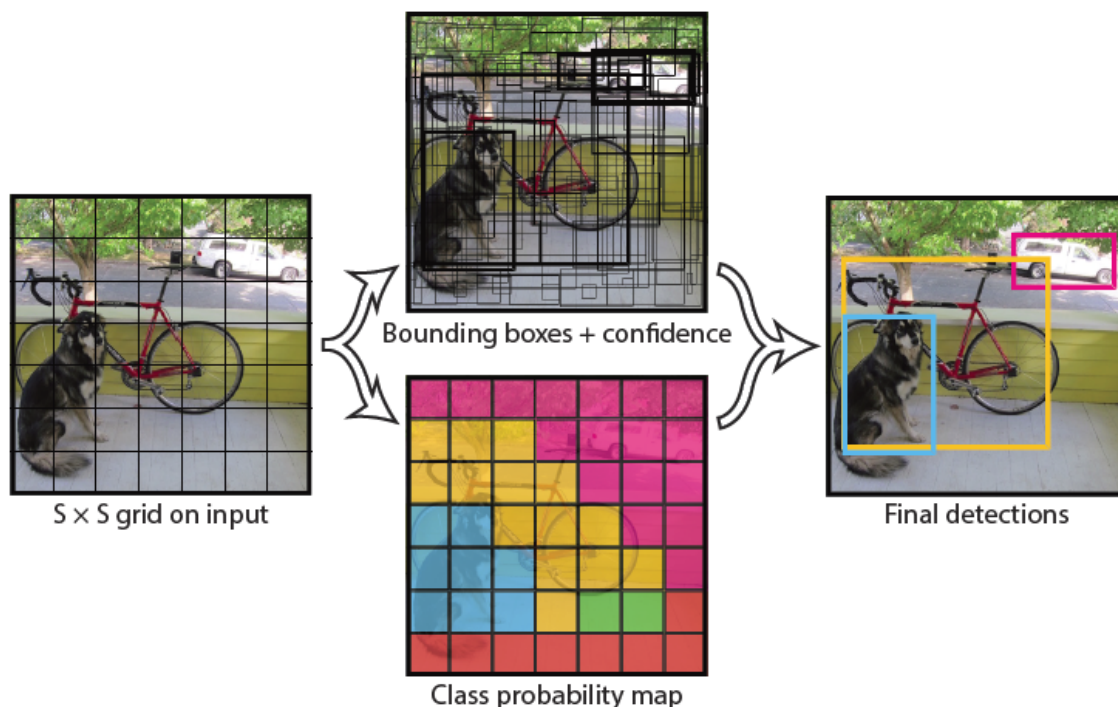
Εικόνα 3.27: Καθορισμός προτάσεων περιοχών αντικειμένου στην αρχιτεκτονική Faster R-CNN. Η προκαθορισμένη επιλογή είναι η χρήση 9 anchor boxes, τριών σχημάτων σε διαφορετικές διαστάσεις τα οποία εξετάζονται σε κάθε θέση του χάρτη χαρακτηριστικών. Ο αριθμός αυτός μπορεί να αλλάξει ανάλογα με τις απαιτήσεις του κάθε προβλήματος[77].

περιοχών με άγκυρες χρησιμοποιούνται για την εκπαίδευση του Fast R-CNN σε μία επαναληπτική διαδικασία. Ο αλγόριθμος Faster R-CNN παρουσίασε βαθμολογία mAP 78,8% για το σύνολο δεδομένων PASCAL VOC και 75,9% στο σύνολο δεδομένων PASCAL VOC 2012.

3.6.4 YOLO

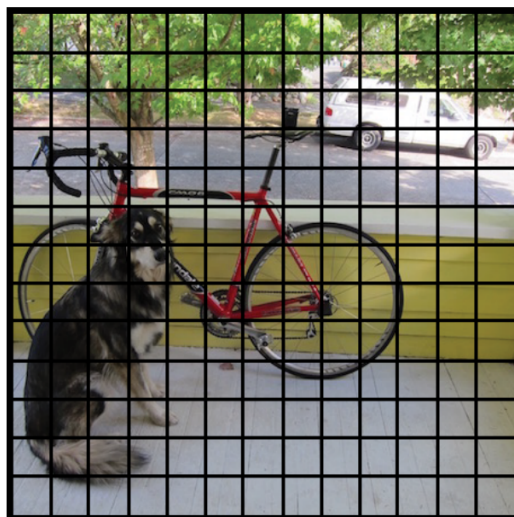
Οι ανιχνευτές που βασίζονται σε προτάσεις περιοχών εικόνας είναι αρκετά ακριβείς αλλά ο χρόνος εκτέλεσής τους συνεχίζει να είναι υψηλός. Ένας τρόπος για να βελτιωθεί ο χρόνος εκτέλεσης της μεθόδου είναι η μείωση του χρόνου που απαιτείται για την επεξεργασία και εξαγωγή αποτελέσματος από κάθε περιοχή ενδιαφέροντος. Συγκεκριμένα, με τους ανιχνευτές μονής λήψης (Single Shot Detectors, SSDs) μπορεί να εξαχθεί η κλάση και η θέση ενός αντικειμένου κατευθείαν από τον χάρτη χαρακτηριστικών της εικόνας. Με αυτόν τον τρόπο παρακάμπτονται τα βήματα της παραγωγής και επεξεργασίας των περιοχών ενδιαφέροντος, ελαττώνοντας σημαντικά τον χρόνο εκτέλεσης του αλγορίθμου.

Διαδεδομένο ανιχνευτή μονής λήψης αποτελεί ο YOLO, ο οποίος χρησιμοποιεί το νευρωνικό δίκτυο DarkNet [75] για την εξαγωγή των χαρακτηριστικών, ακολουθούμενο από συνελικτικά επίπεδα. Το δίκτυο αυτό προτάθηκε ως μία λύση για το πρόβλημα της ανίχνευσης αντικειμένων σε πραγματικό χρόνο διατηρώντας όσο το δυνατόν μεγαλύτερη μέση ακρίβεια.



Εικόνα 3.28: Το ενοποιημένο μοντέλο ανίχνευσης αντικειμένων του YOLO[75].

Επειδή η επεξεργασία εισόδου γίνεται μία φορά, η αρχιτεκτονική του νευρωνικού δικτύου αποτελείται από επίπεδα συνέλιξης, υπο-δειγματοληψίας και τέλος, από δύο ολικά συνδεδεμένα επίπεδα. Τα επίπεδα αυτά είναι συνδεδεμένα σε σειρά. Η αναγνώριση της κλάσης του αντικειμένου αλλά και του περιβλήματός του γίνεται την ίδια χρονική στιγμή. Η εικόνα εισόδου διαιρείται σε ένα πλέγμα $S \times S$ (Εικόνα 3.29). Αν το κέντρο ενός αντικειμένου πέσει σε ένα κελί του πλέγματος, αυτό το κελί είναι υπεύθυνο για την ανίχνευση αυτού του αντικειμένου. Κάθε κελί προβλέπει τα πλαίσια οριοθέτησης B και τον βαθμό εμπιστοσύνης για αυτά τα πλαίσια. Ο βαθμός εμπιστοσύνης αντανακλά τη βεβαιότητα του μοντέλου για την παρουσία οποιουδήποτε αντικειμένου στο πλαίσιο, $P(\text{Object})$.



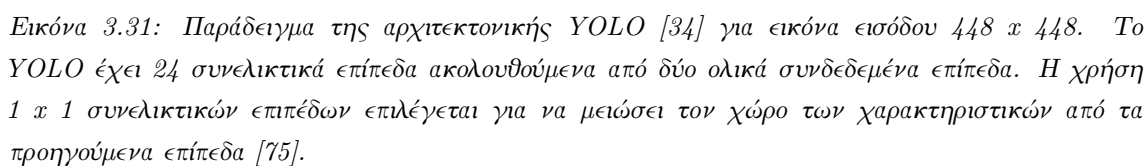
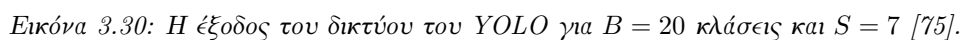
Εικόνα 3.29: Χωρισμός της εικόνας σε πλέγμα σύμφωνα με τον αλγόριθμο YOLO ($S = 13$) [75].

Κάθε πλαίσιο οριοθέτησης αποτελείται από πέντε προβλέψεις: x , y , w , h , και εμπιστοσύνη. Οι συντεταγμένες (x, y) αντιπροσωπεύουν το κέντρο του πλαισίου σε σχέση με τα όρια του κελιού του δικτύου. Το πλάτος w και το ύψος h προβλέπονται σε σχέση με ολόκληρη την εικόνα. Η εμπιστοσύνη αντιπροσωπεύει την τιμή IoU μεταξύ του προβλεπόμενου πλαισίου και του πραγματικού πλαισίου του αντικειμένου. Κάθε κελί προβλέπει, επίσης, πιθανότητες κλάσης υπό όρους, $P(\text{Class}|\text{Object})$. Η έξοδος του δικτύου για συνολικό αριθμό κλάσεων ίσο με 20, $B = 2$ και $S = 7$, παρουσιάζεται στην Εικόνα 3.30.

Όπως παρατηρούμε το μέγεθος της εξόδου ισούται με: $7 \times 7 \times (2 \times 5 + 20) = 1470$.

Το μοντέλο αποτελείται από 24 συνελικτικά στρώματα ακολουθούμενα από δύο πλήρως συνδεδεμένα στρώματα. Τα επικαλυπτόμενα στρώματα εναλλαγής 1×1 μειώνουν τον χώρο των χαρακτηριστικών από τα προηγούμενα επίπεδα.

Στην Εικόνα 3.31 παρουσιάζεται η αρχιτεκτονική του δικτύου.



Ως συναρτήσεις ενεργοποίησης, αντί για τις πλέον διαδεδομένες ReLU χρησιμοποιείται μία παραπλήσια μορφή που ονομάζεται leaky ReLU:

$$\phi(x) = \begin{cases} 0,1x & x \leq 0 \\ x & x > 0 \end{cases} \quad (3.3)$$

Τα αποτελέσματα αξιολόγησης του YOLO για το σύνολο δεδομένων VOC 2007 που απεικονίζονται στην Εικόνα 3.32 δείχνουν ότι το YOLO επιτυγχάνει 63.4% mAP και 45 FPS. Σε σύγκριση με τα DPM, R-CNN, Fast R-CNN και Faster R-CNN, το YOLO μπορεί να επιτύχει ανίχνευση αντικειμένων σε πραγματικό χρόνο με παρόμοιες τιμές mAP.

Real-Time Detectors	Train	mAP	FPS
100Hz DPM [30]	2007	16.0	100
30Hz DPM [30]	2007	26.1	30
Fast YOLO	2007+2012	52.7	155
YOLO	2007+2012	63.4	45
Less Than Real-Time			
Fastest DPM [37]	2007	30.4	15
R-CNN Minus R [20]	2007	53.5	6
Fast R-CNN [14]	2007+2012	70.0	0.5
Faster R-CNN VGG-16[27]	2007+2012	73.2	7
Faster R-CNN ZF [27]	2007+2012	62.1	18
YOLO VGG-16	2007+2012	66.4	21

Εικόνα 3.32: Αποτελέσματα mAP και FPS του YOLO για το VOC 2007 [75].

Ο συνδυασμός των YOLO και Fast R-CNN μπορεί να προσδώσει στο μοντέλο αυξημένη ακρίβεια, συνδυάζοντας τα πλεονεκτήματα των επιμέρους μοντέλων. Πράγματι, μετά τον συνδυασμό του μοντέλου, επιτυγχάνεται 75,0% mAP το οποίο έχει σχετικά μεγαλύτερη ακρίβεια σε σύγκριση με άλλους συνδυασμούς (Εικόνα 3.33).

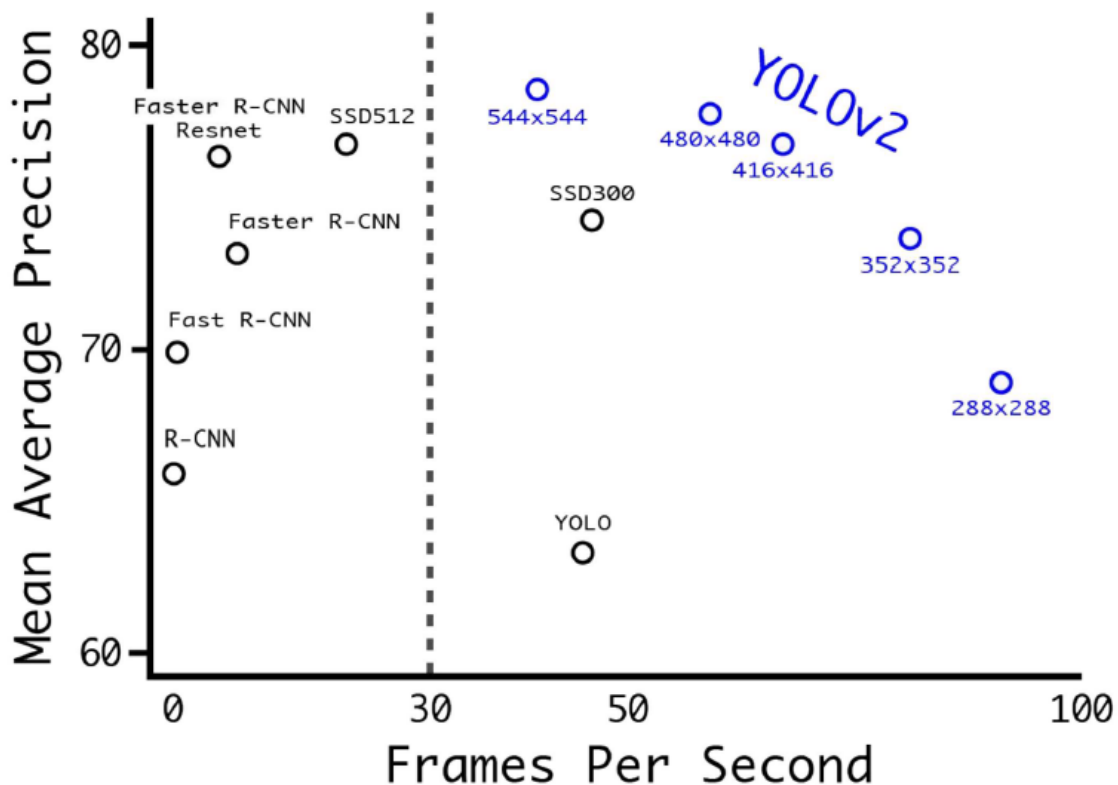
	mAP	Combined	Gain
Fast R-CNN	71.8	-	-
Fast R-CNN (2007 data)	66.9	72.4	.6
Fast R-CNN (VGG-M)	59.2	72.4	.6
Fast R-CNN (CaffeNet)	57.1	72.1	.3
YOLO	63.4	75.0	3.2

Εικόνα 3.33: Συνδυασμός μοντέλων ανίχνευσης αντικειμένων για βελτίωση της ακρίβειας [75].

3.6.5 YOLOv2

Η αρχική έκδοση του YOLO [75] παρά τα ενθαρρυντικά και προπαντός γρήγορα αποτελέσματα, δυσκολευόταν στην ανίχνευση αντικειμένων που εμφανίζονταν πολλές φορές σε κάποιο πλαίσιο, ενώ ο καθορισμός των πλαισίων προκαλούσε έναν μεγάλο αριθμό σφαλμάτων. Αυτοί οι περιορισμοί αντιμετωπίστηκαν στην επόμενη έκδοση του αλγορίθμου, YOLO9000 (YOLOv2) [74], όπου περιέχει μία σειρά από βελτιώσεις. Η βελτίωση αυτή του

YOLO μπορεί να εντοπίζει αντικείμενα σε μία εικόνα που ανήκουν σε 9000 διαφορετικές κατηγορίες. Οι διαδικασίες μάθησης και εντοπισμού γίνονται πιο γρήγορα προσφέροντας μάλιστα μεγαλύτερη ακρίβεια. Χαρακτηριστική είναι η τιμή 76.8% mAP στα 67 FPS (Εικόνα 3.34). Ωστόσο, για την επίτευξη αυτών των αποτελεσμάτων, οι συγγραφείς αντί να επεκτείνουν τις διαστάσεις του δικτύου, προτίμησαν να χρησιμοποιήσουν διάφορες τεχνικές για τη βελτίωση και τη διευκόλυνση της εκπαίδευσής του.



Εικόνα 3.34: Σύγκριση αποτελεσμάτων mAP και FPS διάφορων αλγορίθμων ανίχνευσης αντικειμένων στο PASCAL VOC 2007 Dataset [74].

Η λίστα των βελτιώσεων της έκδοσης YOLOv2 έναντι της πρώτης περιλαμβάνει:

Batch Normalization. Χρησιμοποιείται σε όλα τα επίπεδα σύνθεσης του YOLOv2 και προσφέρει 2% βελτίωση mAP.

Ταξινομητής υψηλής ανάλυσης. Αφού εκπαιδευτεί σε εικόνες μεγέθους 224x224, το YOLOv2 χρησιμοποιεί επίσης 448x448 εικόνες για την τελειοποίηση του δικτύου ταξινόμησης, βελτίωση που προσφέρει 4% αύξηση στο mAP.

Συνελίξεις με anchor boxes. Το YOLOv2 αφαιρεί όλα τα πλήρως συνδεδεμένα στρώματα και χρησιμοποιεί anchor boxes για την πρόβλεψη πλαισίων οριοθέτησης. Επίσης, αφαιρείται ένα επίπεδο ομαδοποίησης (pooling) για να αυξηθεί η ανάλυση της εξόδου. Χωρίς τη χρήση των anchor boxes, το ενδιάμεσο μοντέλο πέτυχε 69,5% mAP και 81% ανάκληση. Με τη χρήση των anchor boxes, λαμβάνεται 69,2% mAP και ανάκληση 88%. Παρόλο που

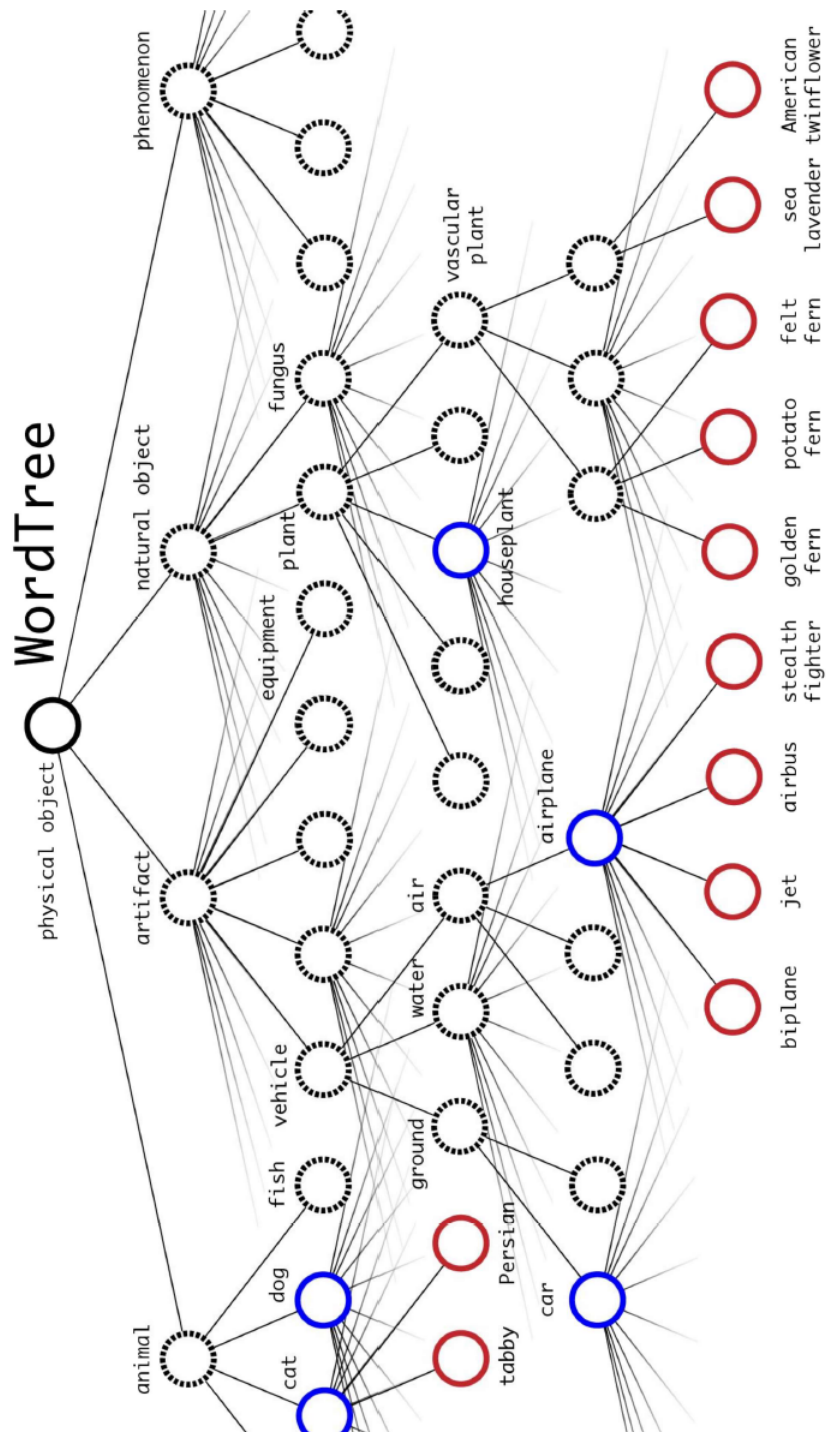
το mAP μειώνεται λίγο, η ανάκληση αυξάνεται σε μεγάλο βαθμό.

Συστάδες διαστάσεων. Τα anchor boxes έχουν τη δυνατότητα να προσαρμόζονται και αυτά στα εκάστοτε αντικείμενα μέσα από τη διαδικασία εκμάθησης. Όμως, με τυχαία αρχικοποίηση οι αρχικές συνθήκες των βαρών εντοπισμού δημιουργούν πρόβλημα για την εκπαίδευση του συστήματος. Η χρήση κλασικής ευκλείδειας ομαδοποίησης με βάση την απόσταση k-means δεν είναι αρκετά καλή επειδή τα μεγαλύτερα πλαίσια δημιουργούν μεγαλύτερα λάθη από τα μικρότερα. Για αυτόν τον λόγο προτείνεται να εφαρμοστεί πρώτα ένας αλγόριθμος συσταδοποίησης (clustering k-means) στο σύνολο δεδομένων ώστε να χρησιμοποιηθούν τα anchor boxes με τη μεγαλύτερη τιμή IoU. Το δίκτυο αποδίδει πολύ καλύτερα με αυτήν την προσέγγιση οδηγώντας σε υψηλότερα IoU. Επειδή υπάρχει η πιθανότητα ένα αντικείμενο να φέρει υψηλή πιθανότητα σε διαφορετικά anchor boxes, εφαρμόζεται και εδώ ο αλγόριθμος NMS (Non-Maximum Suppression) για τον περιορισμό των anchor boxes.

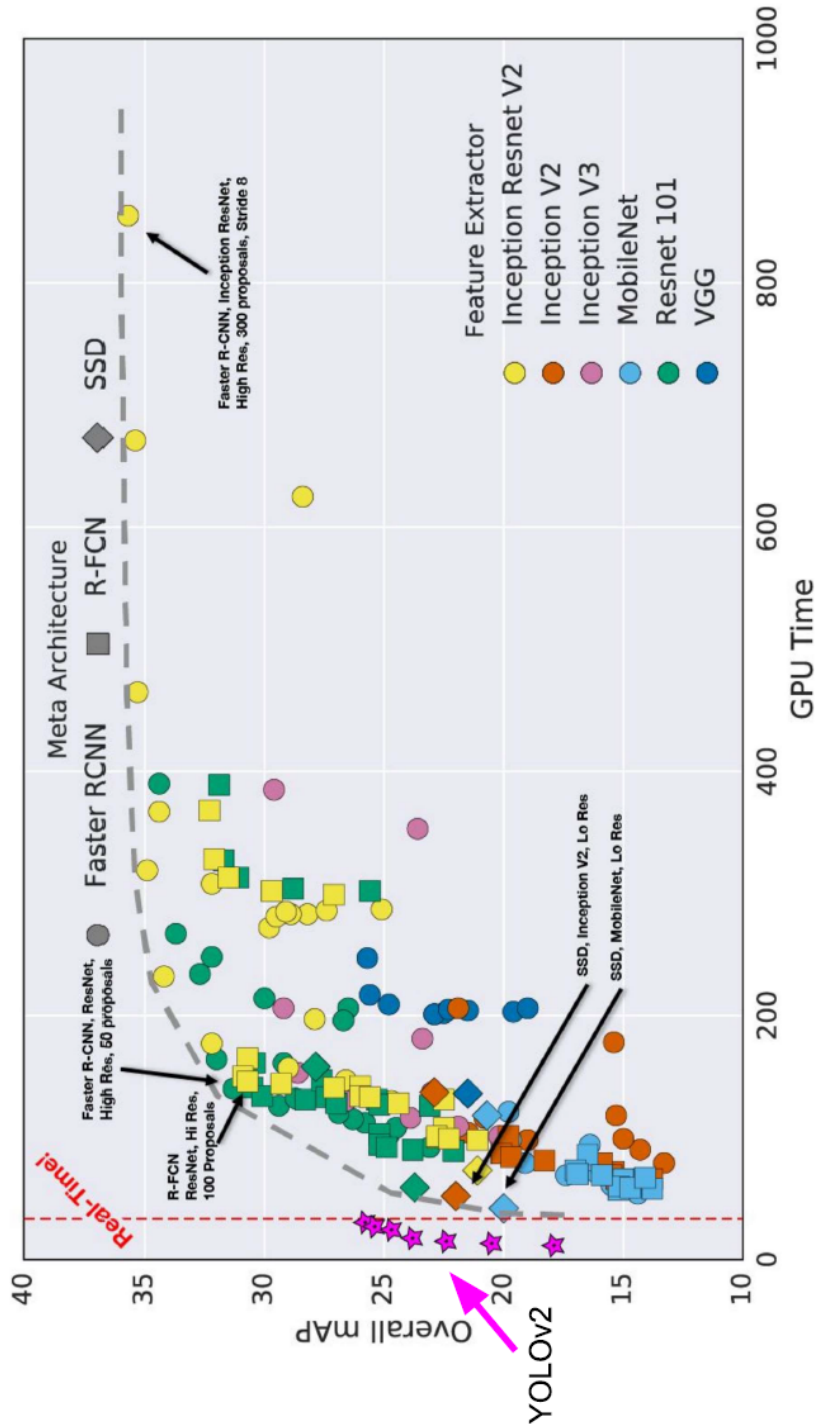
Εξαγωγή χαρακτηριστικών μικρών αντικειμένων. Η έξοδος χαρτών χαρακτηριστικών 13×13 είναι επαρκής για την ανίχνευση μεγάλων αντικειμένων. Για την καλή ανίχνευση μικρών αντικειμένων, οι χάρτες χαρακτηριστικών $26 \times 26 \times 512$ από τα προηγούμενα επίπεδα αντιστοιχίζονται σε χάρτη χαρακτηριστικών $13 \times 13 \times 2048$, και στη συνέχεια συνδυάζονται με τους αρχικούς χάρτες χαρακτηριστικών 13×13 για ανίχνευση. Με αυτόν τον τρόπο επιτυγχάνεται αύξηση 1% στην τιμή της mAP.

WordTree. Πρόσθετα, οι συγγραφείς επιχείρησαν να συνδυάσουν τις κλάσεις από πολλά σύνολα δεδομένων για ταξινόμηση και ανίχνευση. Η δυνατότητα εντοπισμού πολλών κλάσεων οφείλεται στη χρήση ενός δέντρου που εμπεριέχει ιεραρχικά τις κλάσεις αυτές. Επομένως, κάθε αντικείμενο έχει πολλές κλάσεις και η τελική του κλάση καθορίζεται από κάποιο φύλλο του δέντρου. Με αυτόν τον τρόπο, αρκεί να εντοπιστεί ότι το αντικείμενο είναι ενός τύπου και μετά αναζητείται η σχέση μεταξύ των υποτύπων του ώσπου ο αλγόριθμος να καταλήξει σε κάποιο φύλλο του δέντρου. Τα σύνολα δεδομένων που συνδυάστηκαν είναι το COCO με 100 χιλιάδες εικόνες και 80 κλάσεις, όπου οι κλάσεις είναι πιο γενικές όπως "σκύλος" ή "βάρκα", και το ImageNet με 13 εκατομμύρια εικόνες, 22 χιλιάδες κλάσεις, όπου οι κλάσεις είναι πιο συγκεκριμένες όπως "Norfolk terrier", "Yorkshire terrier" ή "Bedlington terrier", αντί για την γενικότερη κλάση "σκύλος". Οι κλάσεις όπως το "σκύλος" και "Norfolk terrier" δεν αλληλοαποκλείονται. Για τον συνδυασμό τους χρησιμοποιείται το WordTree, ένα ιεραρχικό δέντρο συσχετισμού των κλάσεων και των δύο συνόλων δεδομένων. Στην Εικόνα 3.35 απεικονίζεται το ιεραρχικό δέντρο που προέκυψε από τον συνδυασμό των δύο συνόλων δεδομένων με την εφαρμογή του WordTree.

Σημαντική είναι, επίσης, η επίδοση αυτής της βελτιωμένης έκδοσης του YOLO, στην εφαρμογή του στο σύνολο δεδομένων COCO (Εικόνα 3.36), όπου γίνεται σαφής η δυνατότητα του YOLOv2 να αποδώσει σε πραγματικό χρόνο.



Εικόνα 3.35: Το ιεραρχικό δέντρο που προέκυψε από τον συνδυασμό των δύο συνόλων δεδομένων, COCO και ImageNet. Με μπλε χρώμα απεικονίζονται οι κλάσεις του COCO και με κόκκινο χρώμα οι κλάσεις του ImageNet [74].



Εικόνα 3.36: Σύγκριση αποτελεσμάτων mAP και GPU Time διαφόρων αλγορίθμων ανίχνευσης αντικειμένων στο COCO Dataset [59].

3.6.6 YOLOv3

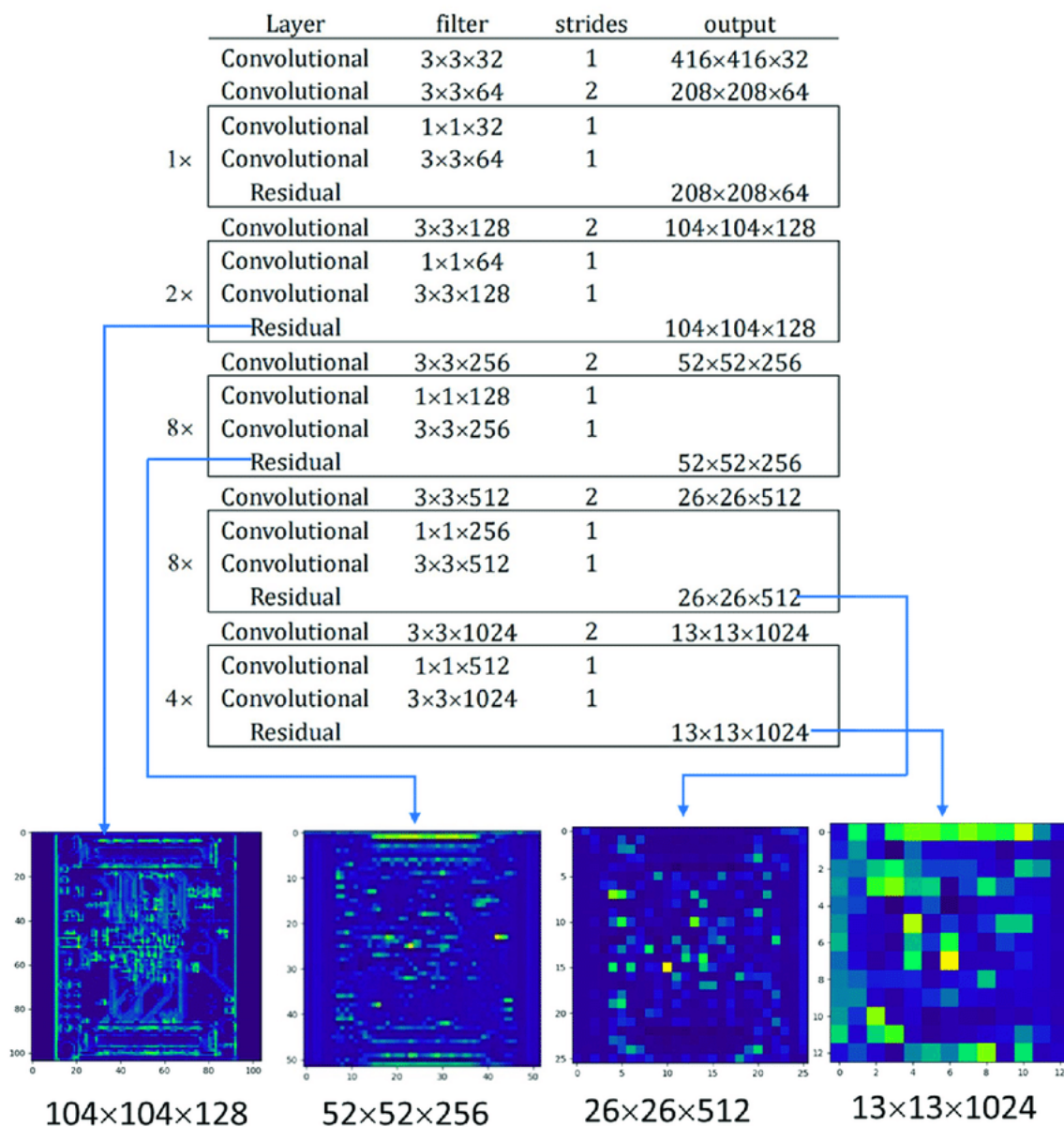
Το YOLOv3 δεν είναι ταχύτερο από το YOLOv2, προσφέρει όμως μεγαλύτερη ακρίβεια. Στο YOLOv2 χρησιμοποιήθηκε μία προσαρμοσμένη βαθιά αρχιτεκτονική Darknet-19, ένα αρχικό δίκτυο 19 επιπέδων, συνοδευόμενο από 11 ακόμη επίπεδα για ανίχνευση αντικειμένων. Με μία αρχιτεκτονική 30 στρωμάτων, το YOLOv2 συχνά αντιμετώπιζε προβλήματα με ανιχνεύσεις μικρών αντικειμένων. Αυτό αποδόθηκε στην απώλεια των λεπτομερών χαρακτηριστικών, καθώς τα επίπεδα περιορίζουν το μέγεθος της εισόδου. Για να αντιμετωπιστεί αυτό, το YOLOv2 συνδύασε χάρτες χαρακτηριστικών από προηγούμενα επίπεδα για να καταγράψει χαρακτηριστικά χαμηλού επιπέδου. Το YOLOv3 χρησιμοποιεί μία παραλλαγή του Darknet, η οποία αρχικά έχει εκπαιδευτεί σε 53 επίπεδα στο Imagenet (Εικόνα 3.37). Για τη λειτουργία της ανίχνευσης, πάνω σε αυτό στοιβάζονται 53 πρόσθετα επίπεδα, δημιουργώντας μία πλήρως συνελικτική αρχιτεκτονική 106 επιπέδων. Αυτό προσδίδει στο YOLOv3 μεγαλύτερη ακρίβεια αλλά ταυτόχρονα συνεπάγεται και μικρότερη ταχύτητα συγκριτικά με το YOLOv2.

Στην Εικόνα 3.38 απεικονίζεται το διάγραμμα της αρχιτεκτονικής δικτύου του YOLOv3. Πρόκειται για ένα δίκτυο που υιοθετεί 106 επίπεδα συνέλιξης. Δεν χρησιμοποιεί πλήρως συνδεδεμένο επίπεδο. Αυτή η δομή καθιστά δυνατή την αντιμετώπιση εικόνων με οποιαδήποτε μεγέθη. Επίσης, δεν χρησιμοποιούνται επίπεδα ομαδοποίησης (pooling). Αντί αυτών, χρησιμοποιείται ένα συνελικτικό επίπεδο με βήμα (stride) 2 για να μειώσει το μέγεθος του χάρτη χαρακτηριστικών. Επιπλέον, η χρήση της δομής ResNet και η δομή Feature Pyramid Network (FPN) αποτελούν κρίσιμα στοιχεία στη βελτίωση της ακρίβειας του YOLOv3.

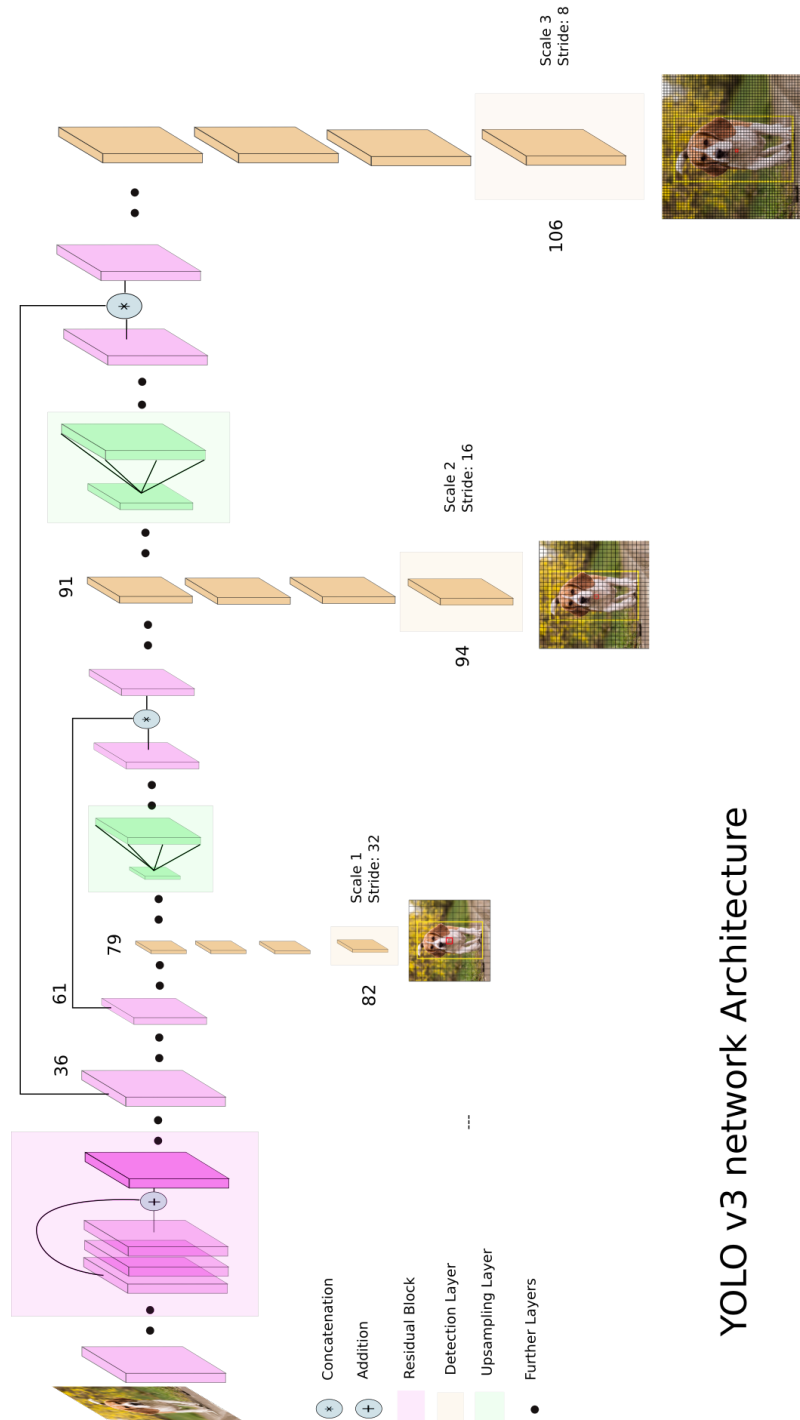
Στην Εικόνα 3.38 απεικονίζονται οι τρεις φάσεις της διαδικασίας ανίχνευσης. Κάθε μία από τις φάσεις χρησιμοποιεί το ένα τρίτο των anchor boxes και διεξάγει ανίχνευση για διαφορετικά μεγέθη αντικειμένων. Πρώτα γίνεται η ανίχνευση αντικειμένων μεγάλων διαστάσεων έχοντας τους μικρότερους χάρτες χαρακτηριστικών. Στη συνέχεια, με χρήση αντίστροφης δειγματοληψίας, οι χάρτες χαρακτηριστικών μεγαλώνουν και στις επόμενες δύο φάσεις γίνεται η ανίχνευση για αντικείμενα μεσαίων και μικρών διαστάσεων, αντίστοιχα. Επισημαίνεται ότι στα πρώτα βήματα πραγματοποιούνται προσθέσεις χαρτών για καλύτερη εκμάθηση.

Τρεις κλίμακες: χειρισμός αντικειμένων διαφορετικών μεγεθών. Υπάρχει ένα ευρύ φάσμα μεγεθών αντικειμένων στις εικόνες και το δίκτυο πρέπει να είναι σε θέση να αναγνωρίζει το σύνολο αυτών. Καθώς το δίκτυο πηγαίνει βαθύτερα, ο χάρτης χαρακτηριστικών του γίνεται όλο και μικρότερος. Δηλαδή, όσο βαθύτερα στο δίκτυο τόσο πιο δύσκολη η ανίχνευση των μικρότερων αντικειμένων. Είναι καλύτερα να εντοπίζονται τα αντικείμενα σε διαφορετικούς χάρτες χαρακτηριστικών πριν τα μικρά αντικείμενα καταλήξουν να εξαφανίζονται. Η ανίχνευση αντικειμένων πραγματοποιείται σε διαφορετικούς χάρτες χαρακτηριστικών προκειμένου να εξάγονται διάφορες κλίμακες.

Ωστόσο, τα χαρακτηριστικά είναι απολύτως σχετικά σε διαφορετικό βάθος του δικτύου.



Εικόνα 3.37: Η αρχιτεκτονική του Darknet-53 που χρησιμοποιεί το YOLOv3 [58].



YOLO v3 network Architecture

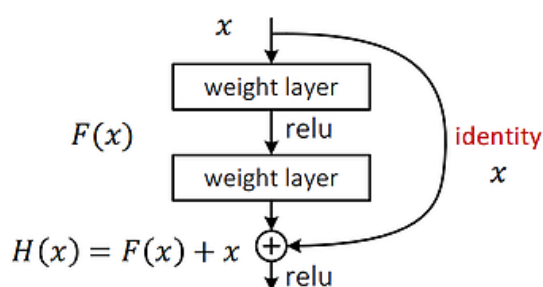
Εικόνα 3.38: Η αρχιτεκτονική του YOLOv3 [21].

Με την αύξηση του βάθους δικτύου, τα χαρακτηριστικά αλλάζουν από χαρακτηριστικά χαμηλού επιπέδου (όπως άκρες, χρώματα ή ακατέργαστες θέσεις 2D) σε χαρακτηριστικά υψηλού επιπέδου (σημειολογικά σημαντικές πληροφορίες όπως σκύλος, γάτα ή αυτοκίνητο). Έτσι, κάνοντας προβλέψεις σε χάρτες χαρακτηριστικών διαφορετικού βάθους, το δίκτυο είναι σε θέση να κάνει ανίχνευση αντικειμένων πολλών μεγεθών. Στο YOLOv3 αυτό επιτυγχάνεται με την ενσωμάτωση μίας FPN δομής.

ResNet-like δομή: ένας καλύτερος τρόπος για την κατανόηση χαρακτηριστικών. Στο YOLOv3, μία δομή ResNet, που ονομάζεται Residual Blocks χρησιμοποιείται για την εκμάθηση χαρακτηριστικών. Τα Residual Blocks αποτελούνται από διάφορα επίπεδα συνέλιξης και διαδρομές συντόμευσης (shortcut paths). Ένα παράδειγμα μίας διαδρομής συντόμευσης απεικονίζεται στην Εικόνα 3.39.

Deep Residual Learning

• Residual net



Εικόνα 3.39: Παράδειγμα των Residual nets [42].

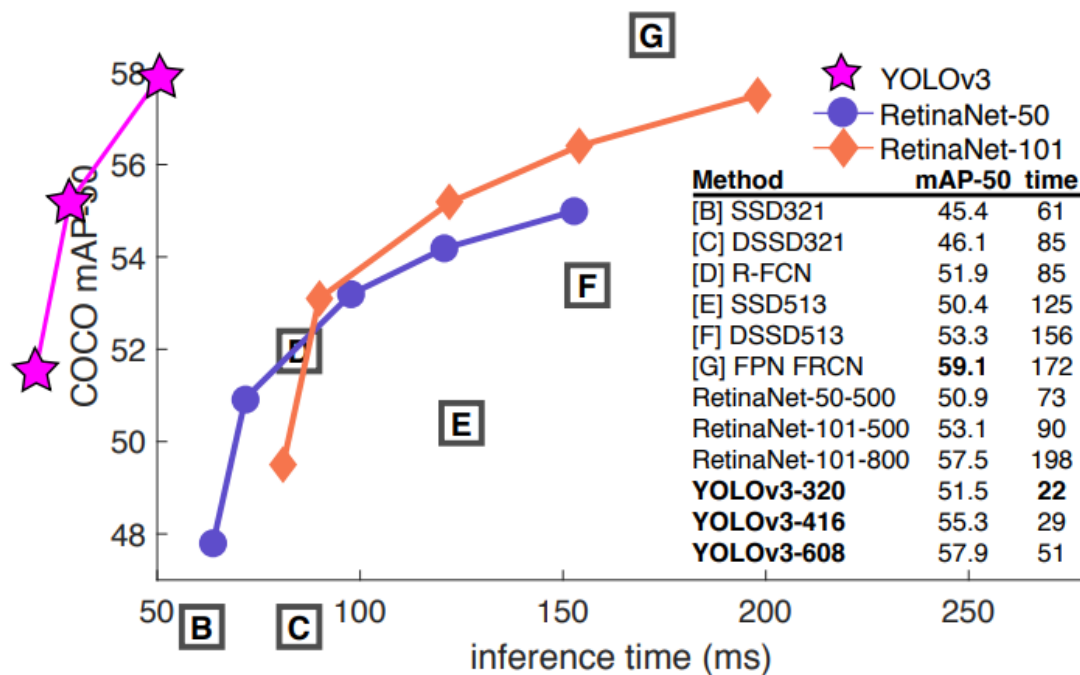
Το χαμπύλο βέλος στα δεξιά της Εικόνας 3.39 αντιπροσωπεύει μία διαδρομή συντόμευσης. Χωρίς αυτό, το δίκτυο είναι ένα κλασικό δίκτυο CNN, το οποίο μαθαίνει τα χαρακτηριστικά διαδοχικά σε σειρά. Καθώς το δίκτυο πηγαίνει βαθύτερα, η επιτυχής εξαγωγή των χαρακτηριστικών γίνεται πιο δύσκολη. Αν όμως προστεθεί η εικονιζόμενη συντόμευση, τα επίπεδα μέσα στο δίκτυο μαθαίνουν τι πρέπει να προσθέσουν στην παλιά λειτουργία για να δημιουργήσουν καλύτερα χαρακτηριστικά. Ως αποτέλεσμα, ένα πολύπλοκο χαρακτηριστικό $H(x)$ διαμορφώνεται τώρα ως $H(x) = F(x) + x$, όπου x είναι το παλιό χαρακτηριστικό που προέρχεται από τη συντόμευση και $F(x)$ είναι η νεοαποκτηθείσα πληροφορία. Αυτό διευκολύνει το δίκτυο να μάθει τα χαρακτηριστικά, ακόμα και σε βαθύτερα σημεία του δικτύου. Αυτή η τεχνική αλλάζει τον στόχο της μάθησης, ο οποίος είναι η εκμάθηση του συμπληρωματικού "υπολειμματικού" που χρησιμοποιείται για να προστεθεί στο παλιό χαρακτηριστικό, γεγονός που απλοποιεί την πολυπλοκότητα της μάθησης.

Λογιστική παλινδρόμηση για βαθμούς εμπιστοσύνης. Το YOLOv3 προβλέπει μία βαθμολογία εμπιστοσύνης για κάθε πλαίσιο χρησιμοποιώντας λογιστική

παλινδρόμηση, ενώ τα YOLO και YOLOv2 χρησιμοποιούν το άθροισμα των τετραγωνικών σφαλμάτων. Η γραμμική παλινδρόμηση της πρόβλεψης αντιστάθμισης οδηγεί σε μείωση του mAP.

Κατάργηση ταξινόμησης Softmax: ταξινόμηση πολλαπλών κλάσεων. Το στρώμα Softmax αντικαθίσταται από ένα επίπεδο συνέλιξης 1×1 με λογιστική λειτουργία. Κατά την πρόβλεψη της εμπιστοσύνης της κλάσης, το YOLOv3 χρησιμοποιεί πολλαπλούς ανεξάρτητους λογιστικούς ταξινομητές για κάθε κατηγορία αντί για ένα επίπεδο Softmax. Αυτό είναι πολύ χρήσιμο, δεδομένου ότι μία εικόνα μπορεί να έχει πολλαπλές κλάσεις (δεν είναι όλες οι κλάσεις αμοιβαίως αποκλειόμενες).

Τα αποτελέσματα του YOLOv3 όπως φαίνεται στην Εικόνα 3.40 είναι συγκρίσιμα σε ακρίβεια σε σχέση με άλλους σύγχρονους αλγόριθμους αναγνωρισμένους ως προς την απόδοσή τους αλλά με σημαντικά υψηλότερη ταχύτητα.



Εικόνα 3.40: Σύγκριση αποτελεσμάτων mAP και ταχύτητας διάφορων αλγορίθμων ταχείας ανίχνευσης αντικειμένων [59, 76]

3.7 Εφαρμογή ανίχνευσης αντικειμένων με το YOLOv3

Για την εφαρμογή την ανίχνευσης αντικειμένων και των κλάσεων τους στη διαδικτυακή πλατφόρμα επιλέχθηκε το YOLOv3 για τη μεγάλη του ακρίβεια σε πραγματικούς χρόνους εκτέλεσης. Εκμεταλλευόμενη την ταχύτητα του αλγορίθμου, η εταιρεία μπορεί να προβεί σε οποιεσδήποτε ενέργειες τη στιγμή που μία φωτογραφία μεταφορτώνεται, ανάλογα με τις ανιχνευμένες σε αυτήν κλάσεις (όπως προτεινόμενες στον χρήστη ετικέτες για τη

φωτογραφία του σε πραγματικό χρόνο). Το YOLOv3 μπορεί να ρυθμιστεί για την ανίχνευση οποιουδήποτε τύπου αντικειμένου. Στην πραγματικότητα, δίνει τη δυνατότητα ρύθμισης της αρχιτεκτονικής του νευρωνικού δικτύου μέσω ενός αρχείου παραμετροποίησης (yolov3.cfg). Αυτό το αρχείο παραμετροποίησης μπορεί να ρυθμίσει τα εξής στοιχεία του δικτύου:

- Αρχιτεκτονικές CNN (επίπεδα και συναρτήσεις ενεργοποίησης)
- Anchor Boxes
- Αριθμός κλάσεων
- Ρυθμός εκμάθησης
- Τεχνική βελτιστοποίησης
- Μέγεθος εικόνας εισόδου
- Κατώτατο όριο βαθμού πιθανότητας
- Μεγέθη παρτίδων

Κάθε παραμετροποίηση έχει και τα αντίστοιχα προκαθορισμένα βάρη. Για την εφαρμογή, έγινε χρήση των βαρών yolov3.weights σε συνδυασμό με την πλήρη παραμετροποίηση yolov3.cfg. Τα βάρη αυτά είναι τα καθορισμένα για το σύνολο των 9000+ κλάσεων του συνόλου δεδομένων Open Images [69]. Το Open Images είναι ένα σύνολο δεδομένων αποτελούμενο από περίπου 9 εκατομμύρια εικόνες που έχουν επισημανθεί με ετικέτες, περιγράμματα αντικειμένων και οπτικές σχέσεις. Το σύνολο εκπαίδευσης της έκδοσης v4 περιλαμβάνει 14.6 εκατομμύρια πλαίσια για 600 κλάσεις αντικειμένων σε 1.74 εκατομμύρια εικόνες, καθιστώντας το το μεγαλύτερο υπάρχον σύνολο δεδομένων με επισημάνσεις και θέσεις αντικειμένων. Τα πλαίσια, ως επί το πλείστον, σχεδιάστηκαν με μη αυτόματο τρόπο από επαγγελματίες για την εξασφάλιση ακρίβειας και συνέπειας. Οι εικόνες ανήκουν σε ένα πολύ διαφορετικό φάσμα κατηγοριών και περιέχουν μεταξύ άλλων πολύπλοκες σκηνές με πολλά αντικείμενα (8.4 αντικείμενα ανά εικόνα κατά μέσο όρο) [56].

Το Darknet αποτυπώνει τα αντικείμενα που ανιχνεύει, τον βαθμό εμπιστοσύνης και τον χρόνο ανίχνευσής τους. Από προεπιλογή δεν εμφανίζει απευθείας τις ανιχνεύσεις, αλλά τις αποθηκεύει σε ένα αρχείο predictions.png.

Εκτελώντας το Darknet με χρήση της CPU χρειάζονται περίπου 12 με 30 δευτερόλεπτα ανά εικόνα. Εάν χρησιμοποιηθεί η έκδοση GPU, η ανίχνευση γίνεται πολύ γρηγορότερα, σε πραγματικό χρόνο.

Αλλαγή του ορίου ανίχνευσης. Από προεπιλογή, το YOLO εμφανίζει μόνο αντικείμενα που έχουν ανιχνευθεί με βαθμό εμπιστοσύνης 0.25 ή μεγαλύτερο. Η τιμή αυτή μπορεί να ρυθμιστεί. Αν τεθεί στο 0 τότε θα εμφανίσει όλα τα ανιχνευθέντα αντικείμενα ανεξαρτήτως βαθμού εμπιστοσύνης. Συνήθως η τιμή αυτή καθορίζεται ανάλογα με το



Εικόνα 3.41: Εικόνα χρήστη (Α) της διαδικτυακής πλατφόρμας.



Εικόνα 3.42: Εικόνα χρήστη (Β) της διαδικτυακής πλατφόρμας.

πρόβλημα της ανίχνευσης και της απαιτούμενης για αυτό ακρίβειας. Παρατηρείται ότι για τιμή κατώφλιου εμπιστοσύνης ίσο με 0.1 εμφανίζονται και κάποιες ψευδώς θετικές ανιχνεύσεις. Αυξάνοντας το κατώφλι σε μεγαλύτερες τιμές εξαλείφονται οι ψευδώς θετικές ανιχνεύσεις αλλά και κάποιες αληθώς θετικές με χαμηλή εμπιστοσύνη. Στην πράξη, χρησιμοποιήθηκε ένα κατώφλι ίσο με 0.6, ενώ η περαιτέρω βελτιστοποίηση της τιμής του αποτελεί αντικείμενο μελλοντικής επέκτασης της παρούσας πρότασης.

Στις Εικόνες 3.43, 3.44, 3.45, 3.46 φαίνεται η εφαρμογή του YOLO σε δύο φωτογραφίες (Εικόνα 3.41 και 3.42) χρηστών της διαδικτυακής πλατφόρμας για διαφορετικά κατώφλια. Οι κλάσεις των αντικειμένων που ανιχνεύτηκαν για την Εικόνα 3.41 με κατώφλι 0.1 είναι οι:

vase: 51%	vase: 36%	vase: 11%
microwave: 59%	microwave: 37%	diningtable: 71%
pottedplant: 98%	chair: 100%	chair: 100%
chair: 100%	chair: 100%	chair: 98%
chair: 28%	bowl: 65%	bowl: 27%
spoon: 24%	cup: 99%	cup: 97%
cup: 86%	cup: 62%	cup: 61%
cup: 50%	cat: 98%	

και οι κλάσεις αντικειμένων που ανιχνεύθηκαν στην Εικόνα 3.42 με κατώφλι 0.1 είναι οι:

vase: 94%

cup: 17%

chair: 28%

pottedplant: 26%

person: 81%

vase: 39%

diningtable: 18%

pottedplant: 51%

sofa: 100%

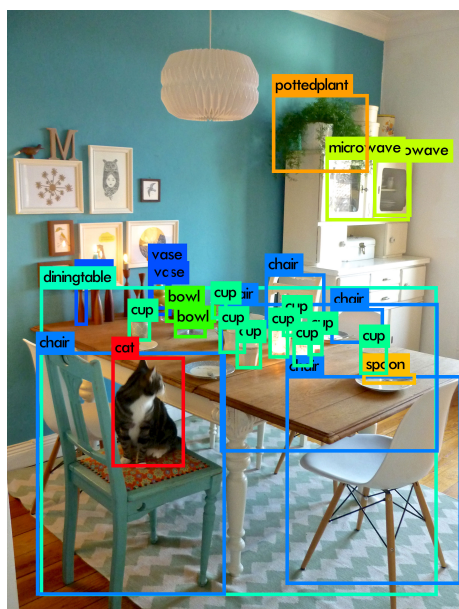
vase: 26%

diningtable: 17%

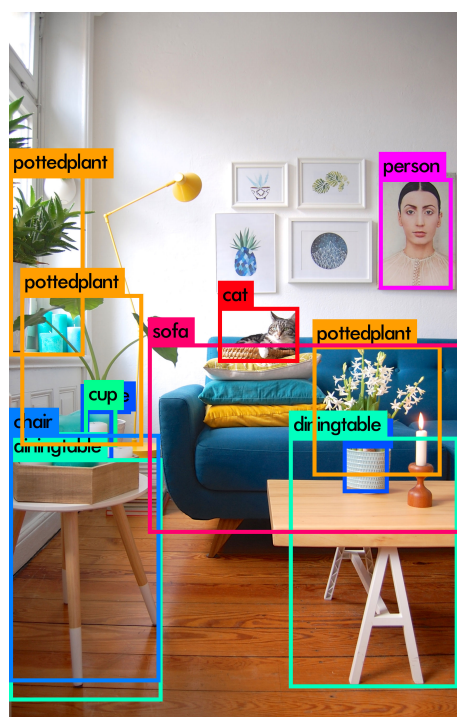
pottedplant: 32%

cat: 99%

Οι ανιχνευμένες κλάσεις για κατώφλι 0.1 φαίνονται περιγεγραμμένες με χρωματιστά πλαίσια στις Εικόνες 3.43 και 3.44 και για το κατώφλι που τελικώς επιλέχθηκε (0.6) στις Εικόνες 3.45 και 3.46.



Εικόνα 3.43: Ανίχνευση κλάσεων αντικειμένων του YOLO με κατώφλι 0.1 σε φωτογραφία χρήστη της διαδικτυακής πλατφόρμας.

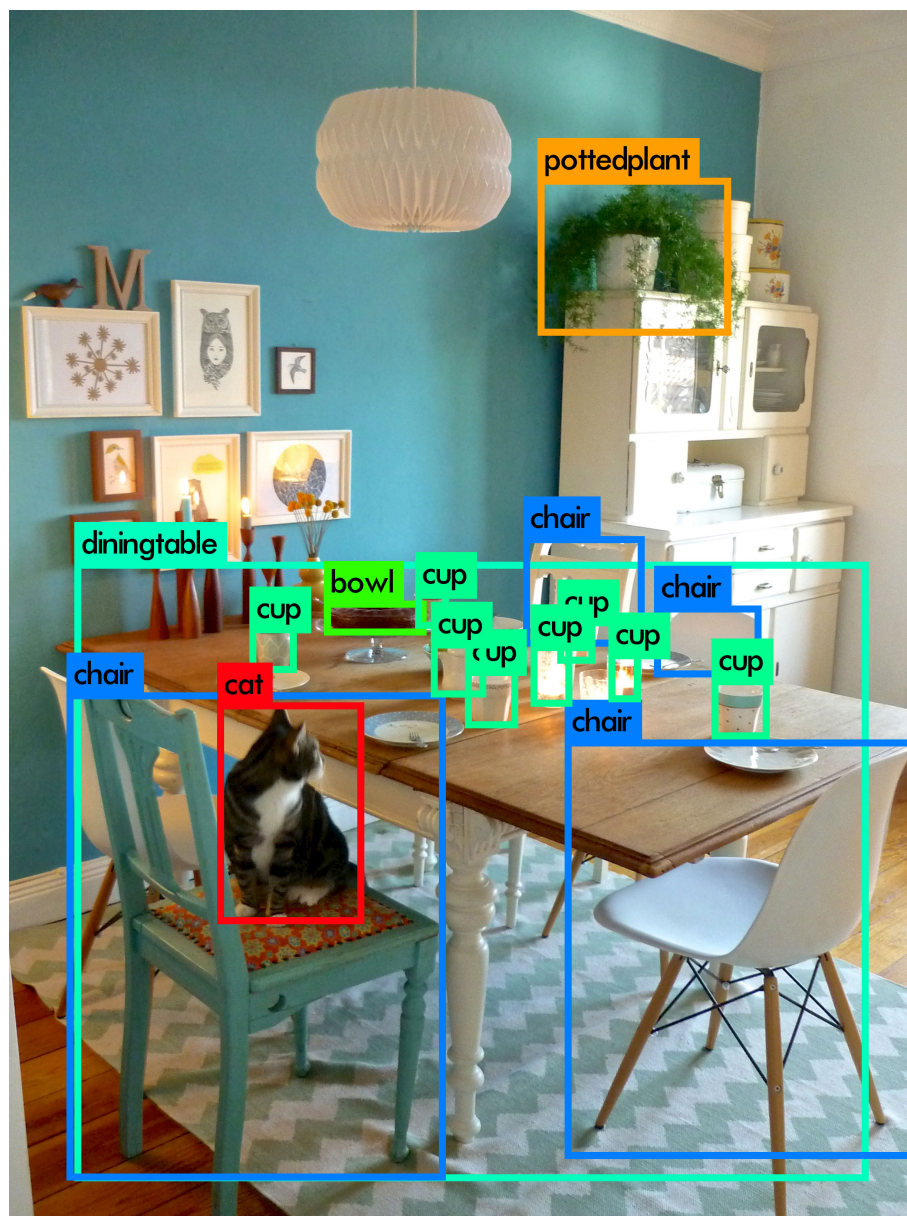


Εικόνα 3.44: Ανίχνευση κλάσεων αντικειμένων του YOLO με κατώφλι 0.1 σε φωτογραφία χρήστη της διαδικτυακής πλατφόρμας.

Στις Εικόνες 3.47, 3.48 αποτυπώνεται η εκτέλεση του αλγορίθμου. Παρατηρούνται τα συνελικτικά επίπεδα, τα επίπεδα των residual nets, των downsamples και των upsamples που ενυπάρχουν στα 106 επίπεδα του Darknet-53, μαζί με τους αντίστοιχους χρόνους εκτέλεσης του κάθε επιπέδου.



Εικόνα 3.45: Ανίχνευση κλάσεων αντικειμένων του YOLO με κατώφλι 0.6 σε φωτογραφία χρήστη της διαδικτυακής πλατφόρμας.



Εικόνα 3.46: Ανίχνευση κλάσεων αντικειμένων του YOLO με κατώφλι 0.6 σε φωτογραφία χρήστη της διαδικτυακής πλατφόρμας

layer		filters	size	input		output	
0	conv	32	3 x 3 / 1	608 x 608 x 3	->	608 x 608 x 32	0.639 BFLOPs
1	conv	64	3 x 3 / 2	608 x 608 x 32	->	304 x 304 x 64	3.407 BFLOPs
2	conv	32	1 x 1 / 1	304 x 304 x 64	->	304 x 304 x 32	0.379 BFLOPs
3	conv	64	3 x 3 / 1	304 x 304 x 32	->	304 x 304 x 64	3.407 BFLOPs
4	res	1		304 x 304 x 64	->	304 x 304 x 64	
5	conv	128	3 x 3 / 2	304 x 304 x 64	->	152 x 152 x 128	3.407 BFLOPs
6	conv	64	1 x 1 / 1	152 x 152 x 128	->	152 x 152 x 64	0.379 BFLOPs
7	conv	128	3 x 3 / 1	152 x 152 x 64	->	152 x 152 x 128	3.407 BFLOPs
8	res	5		152 x 152 x 128	->	152 x 152 x 128	
9	conv	64	1 x 1 / 1	152 x 152 x 128	->	152 x 152 x 64	0.379 BFLOPs
10	conv	128	3 x 3 / 1	152 x 152 x 64	->	152 x 152 x 128	3.407 BFLOPs
11	res	8		152 x 152 x 128	->	152 x 152 x 128	
12	conv	256	3 x 3 / 2	152 x 152 x 128	->	76 x 76 x 256	3.407 BFLOPs
13	conv	128	1 x 1 / 1	76 x 76 x 256	->	76 x 76 x 128	0.379 BFLOPs
14	conv	256	3 x 3 / 1	76 x 76 x 128	->	76 x 76 x 256	3.407 BFLOPs
15	res	12		76 x 76 x 256	->	76 x 76 x 256	
16	conv	128	1 x 1 / 1	76 x 76 x 256	->	76 x 76 x 128	0.379 BFLOPs
17	conv	256	3 x 3 / 1	76 x 76 x 128	->	76 x 76 x 256	3.407 BFLOPs
18	res	15		76 x 76 x 256	->	76 x 76 x 256	
19	conv	128	1 x 1 / 1	76 x 76 x 256	->	76 x 76 x 128	0.379 BFLOPs
20	conv	256	3 x 3 / 1	76 x 76 x 128	->	76 x 76 x 256	3.407 BFLOPs
21	res	18		76 x 76 x 256	->	76 x 76 x 256	
22	conv	128	1 x 1 / 1	76 x 76 x 256	->	76 x 76 x 128	0.379 BFLOPs
23	conv	256	3 x 3 / 1	76 x 76 x 128	->	76 x 76 x 256	3.407 BFLOPs
24	res	21		76 x 76 x 256	->	76 x 76 x 256	
25	conv	128	1 x 1 / 1	76 x 76 x 256	->	76 x 76 x 128	0.379 BFLOPs
26	conv	256	3 x 3 / 1	76 x 76 x 128	->	76 x 76 x 256	3.407 BFLOPs
27	res	24		76 x 76 x 256	->	76 x 76 x 256	
28	conv	128	1 x 1 / 1	76 x 76 x 256	->	76 x 76 x 128	0.379 BFLOPs
29	conv	256	3 x 3 / 1	76 x 76 x 128	->	76 x 76 x 256	3.407 BFLOPs
30	res	27		76 x 76 x 256	->	76 x 76 x 256	
31	conv	128	1 x 1 / 1	76 x 76 x 256	->	76 x 76 x 128	0.379 BFLOPs
32	conv	256	3 x 3 / 1	76 x 76 x 128	->	76 x 76 x 256	3.407 BFLOPs
33	res	30		76 x 76 x 256	->	76 x 76 x 256	
34	conv	128	1 x 1 / 1	76 x 76 x 256	->	76 x 76 x 128	0.379 BFLOPs
35	conv	256	3 x 3 / 1	76 x 76 x 128	->	76 x 76 x 256	3.407 BFLOPs
36	res	33		76 x 76 x 256	->	76 x 76 x 256	
37	conv	512	3 x 3 / 2	76 x 76 x 256	->	38 x 38 x 512	3.407 BFLOPs
38	conv	256	1 x 1 / 1	38 x 38 x 512	->	38 x 38 x 256	0.379 BFLOPs
39	conv	512	3 x 3 / 1	38 x 38 x 256	->	38 x 38 x 512	3.407 BFLOPs
40	res	37		38 x 38 x 512	->	38 x 38 x 512	
41	conv	256	1 x 1 / 1	38 x 38 x 512	->	38 x 38 x 256	0.379 BFLOPs
42	conv	512	3 x 3 / 1	38 x 38 x 256	->	38 x 38 x 512	3.407 BFLOPs
43	res	40		38 x 38 x 512	->	38 x 38 x 512	
44	conv	256	1 x 1 / 1	38 x 38 x 512	->	38 x 38 x 256	0.379 BFLOPs
45	conv	512	3 x 3 / 1	38 x 38 x 256	->	38 x 38 x 512	3.407 BFLOPs
46	res	43		38 x 38 x 512	->	38 x 38 x 512	
47	conv	256	1 x 1 / 1	38 x 38 x 512	->	38 x 38 x 256	0.379 BFLOPs
48	conv	512	3 x 3 / 1	38 x 38 x 256	->	38 x 38 x 512	3.407 BFLOPs
49	res	46		38 x 38 x 512	->	38 x 38 x 512	
50	conv	256	1 x 1 / 1	38 x 38 x 512	->	38 x 38 x 256	0.379 BFLOPs
51	conv	512	3 x 3 / 1	38 x 38 x 256	->	38 x 38 x 512	3.407 BFLOPs
52	res	49		38 x 38 x 512	->	38 x 38 x 512	
53	conv	256	1 x 1 / 1	38 x 38 x 512	->	38 x 38 x 256	0.379 BFLOPs

Εικόνα 3.47: Αποτύπωση της εκτέλεσης του αλγορίθμου YOLO σε φωτογραφία χρήση της διαδικτυακής πλατφόρμας (Εικόνα 1 από 2). Παρατηρούνται τα συνελκτικά επίπεδα, τα επίπεδα των residual nets, των downsamples και των upsamples που ενυπάρχουν στα 106 επίπεδα του Darknet-53, μαζί με τους χρόνους εκτέλεσης του κάθε επιπέδου.

53	conv	256	1 x 1 / 1	38 x 38 x 512	->	38 x 38 x 256	0.379	BFLOPs
54	conv	512	3 x 3 / 1	38 x 38 x 256	->	38 x 38 x 512	3.407	BFLOPs
55	res	52		38 x 38 x 512	->	38 x 38 x 512		
56	conv	256	1 x 1 / 1	38 x 38 x 512	->	38 x 38 x 256	0.379	BFLOPs
57	conv	512	3 x 3 / 1	38 x 38 x 256	->	38 x 38 x 512	3.407	BFLOPs
58	res	55		38 x 38 x 512	->	38 x 38 x 512		
59	conv	256	1 x 1 / 1	38 x 38 x 512	->	38 x 38 x 256	0.379	BFLOPs
60	conv	512	3 x 3 / 1	38 x 38 x 256	->	38 x 38 x 512	3.407	BFLOPs
61	res	58		38 x 38 x 512	->	38 x 38 x 512		
62	conv	1024	3 x 3 / 2	38 x 38 x 512	->	19 x 19 x 1024	3.407	BFLOPs
63	conv	512	1 x 1 / 1	19 x 19 x 1024	->	19 x 19 x 512	0.379	BFLOPs
64	conv	1024	3 x 3 / 1	19 x 19 x 512	->	19 x 19 x 1024	3.407	BFLOPs
65	res	62		19 x 19 x 1024	->	19 x 19 x 1024		
66	conv	512	1 x 1 / 1	19 x 19 x 1024	->	19 x 19 x 512	0.379	BFLOPs
67	conv	1024	3 x 3 / 1	19 x 19 x 512	->	19 x 19 x 1024	3.407	BFLOPs
68	res	65		19 x 19 x 1024	->	19 x 19 x 1024		
69	conv	512	1 x 1 / 1	19 x 19 x 1024	->	19 x 19 x 512	0.379	BFLOPs
70	conv	1024	3 x 3 / 1	19 x 19 x 512	->	19 x 19 x 1024	3.407	BFLOPs
71	res	68		19 x 19 x 1024	->	19 x 19 x 1024		
72	conv	512	1 x 1 / 1	19 x 19 x 1024	->	19 x 19 x 512	0.379	BFLOPs
73	conv	1024	3 x 3 / 1	19 x 19 x 512	->	19 x 19 x 1024	3.407	BFLOPs
74	res	71		19 x 19 x 1024	->	19 x 19 x 1024		
75	conv	512	1 x 1 / 1	19 x 19 x 1024	->	19 x 19 x 512	0.379	BFLOPs
76	conv	1024	3 x 3 / 1	19 x 19 x 512	->	19 x 19 x 1024	3.407	BFLOPs
77	conv	512	1 x 1 / 1	19 x 19 x 1024	->	19 x 19 x 512	0.379	BFLOPs
78	conv	1024	3 x 3 / 1	19 x 19 x 512	->	19 x 19 x 1024	3.407	BFLOPs
79	conv	512	1 x 1 / 1	19 x 19 x 1024	->	19 x 19 x 512	0.379	BFLOPs
80	conv	1024	3 x 3 / 1	19 x 19 x 512	->	19 x 19 x 1024	3.407	BFLOPs
81	conv	255	1 x 1 / 1	19 x 19 x 1024	->	19 x 19 x 255	0.189	BFLOPs
82	yolo							
83	route	79						
84	conv	256	1 x 1 / 1	19 x 19 x 512	->	19 x 19 x 256	0.095	BFLOPs
85	upsample		2x	19 x 19 x 256	->	38 x 38 x 256		
86	route	85 61						
87	conv	256	1 x 1 / 1	38 x 38 x 768	->	38 x 38 x 256	0.568	BFLOPs
88	conv	512	3 x 3 / 1	38 x 38 x 256	->	38 x 38 x 512	3.407	BFLOPs
89	conv	256	1 x 1 / 1	38 x 38 x 512	->	38 x 38 x 256	0.379	BFLOPs
90	conv	512	3 x 3 / 1	38 x 38 x 256	->	38 x 38 x 512	3.407	BFLOPs
91	conv	256	1 x 1 / 1	38 x 38 x 512	->	38 x 38 x 256	0.379	BFLOPs
92	conv	512	3 x 3 / 1	38 x 38 x 256	->	38 x 38 x 512	3.407	BFLOPs
93	conv	255	1 x 1 / 1	38 x 38 x 512	->	38 x 38 x 255	0.377	BFLOPs
94	yolo							
95	route	91						
96	conv	128	1 x 1 / 1	38 x 38 x 256	->	38 x 38 x 128	0.095	BFLOPs
97	upsample		2x	38 x 38 x 128	->	76 x 76 x 128		
98	route	97 36						
99	conv	128	1 x 1 / 1	76 x 76 x 384	->	76 x 76 x 128	0.568	BFLOPs
100	conv	256	3 x 3 / 1	76 x 76 x 128	->	76 x 76 x 256	3.407	BFLOPs
101	conv	128	1 x 1 / 1	76 x 76 x 256	->	76 x 76 x 128	0.379	BFLOPs
102	conv	256	3 x 3 / 1	76 x 76 x 128	->	76 x 76 x 256	3.407	BFLOPs
103	conv	128	1 x 1 / 1	76 x 76 x 256	->	76 x 76 x 128	0.379	BFLOPs
104	conv	256	3 x 3 / 1	76 x 76 x 128	->	76 x 76 x 256	3.407	BFLOPs
105	conv	255	1 x 1 / 1	76 x 76 x 256	->	76 x 76 x 255	0.754	BFLOPs
106	yolo							

Εικόνα 3.48: Αποτύπωση της εκτέλεσης του αλγορίθμου YOLO σε φωτογραφία χρήση της διαδικτυακής πλατφόρμας (Εικόνα 2 από 2). Παρατηρούνται τα συνελκτικά επίπεδα, τα επίπεδα των residual nets, των downsamples και των upsamples που ενυπάρχουν στα 106 επίπεδα του Darknet-53, μαζί με τους χρόνους εκτέλεσης του κάθε επιπέδου.

Κεφάλαιο 4

Δοκιμή α/β

4.1 Εισαγωγή

Μία δοκιμή α/β ή/και δοκιμή $\alpha/\beta/n$, είναι ένα τυχαιοποιημένο πείραμα χρησιμοποιώντας δύο ή περισσότερες παραλλαγές της ίδιας ιστοσελίδας. Η παραλλαγή Α είναι το πρωτότυπο. Οι παραλλαγές Β έως n περιέχουν ένα ή περισσότερα στοιχεία που έχουν τροποποιηθεί από το πρωτότυπο (για παράδειγμα ένα διαφορετικού χρώματος πλαίσιο). Οι παραλλαγές Β έως n μπορεί να περιέχουν άλλοτε άλλης έκτασης τροποποιήσεις. Μπορεί να αφορούν σε ένα μόνο στοιχείο μίας μόνο ιστοσελίδας, να περιέχουν πολλαπλές τροποποιήσεις ή να αποτελούν μία εντελώς διαφορετική ιστοσελίδα.

Οι δοκιμές α/β περιλαμβάνουν το έργο της στατιστικής σύγκρισης των δύο αυτών εκδόσεων, για να διαπιστωθεί ποια αποδίδει καλύτερα. Στην ιδανική περίπτωση, αυτό οδηγεί σε περισσότερο πληροφορημένες λήψεις αποφάσεων και επιτρέπει γρήγορους κύκλους ανατροφοδότησης και συνεχή βελτίωση. Για τη διεξαγωγή του πειράματος, δημιουργείται αρχικά μία υπόθεση σχετικά με τη συμπεριφορά των χρηστών (που συχνά βασίζεται σε ψυχολογικές θεωρίες). Για τον έλεγχο αυτής της υπόθεσης, δημιουργείται ένα πείραμα, όπου δοκιμάζεται η τρέχουσα υλοποίηση ("Control") και μία ή περισσότερες παραλλαγές μίας νέας υλοποίησης ("Treatment"). Πριν ξεκινήσει η δοκιμή, λαμβάνεται μία απόφαση σχετικά με τους αντικειμενικούς στόχους που θα αξιολογηθούν κατά το πείραμα.

Οι χρήστες αποκτούν πρόσβαση στον ιστότοπο ή την εφαρμογή και συγκεκριμένα στις σελίδες με το υπό δοκιμή χαρακτηριστικό. Έπειτα, καθορίζεται ποια έκδοση του πειράματος θα προσφερθεί στον χρήστη και αυτή αποτελεί την έκδοση στην οποία θα έχει πρόσβαση ο χρήστης καθ' όλη τη διάρκεια του πειράματος. Συνήθως, τα πειράματα είναι ενεργά για χρονική περίοδο τουλάχιστον μία εβδομάδας και συχνά για μεγαλύτερες περιόδους. Κατά τη χρονική αυτή περίοδο, τα συμβάντα συμπεριφοράς των χρηστών καταγράφονται και αποθηκεύονται για στατιστική ανάλυση. Τα αποτελέσματα αναλύονται και στις περισσότερες περιπτώσεις η παραλλαγή του πειράματος με τις καλύτερες επιδόσεις σταδιακά εφαρμόζεται στον ιστότοπο, διαθέσιμη για όλους τους χρήστες. Στην Εικόνα 4.1, απεικονίζεται η γενική δομή μίας διαδικτυακής δοκιμής α/β .

4.2 Σχεδιασμός πειραμάτων

Σύμφωνα με το “Alternatives controlled experiments on the web: survey and practical guide” (Kohavi et al. 2009) [52], το οποίο παρέχει μία ανασκόπηση πολλών εναλλακτικών αρχιτεκτονικής δοκιμών α/β , τα τρία κύρια συστατικά ενός πειράματος αποτελούν ο αλγόριθμος τυχαιότητας, η μέθοδος ανάθεσης και η ροή των δεδομένων.

Υπάρχουν αρκετές παράμετροι στον σχεδιασμό ενός πειράματος με τις πιο σημαντικές να είναι ο υπολογισμός επαρκούς δείγματος, η επιλογή του κατάλληλου δείγματος και η συλλογή των σωστών μετρήσεων.

Μέγεθος Δείγματος. Το μέγεθος του δείγματος καθορίζεται από το ποσοστό των χρηστών που εκτίθενται σε μία εκδοχή του πειράματος και το χρονικό διάστημα εκτέλεσης του πειράματος. Καθώς ένα πείραμα διαρκεί περισσότερο, όλο και περισσότεροι επισκέπτες ομαδοποιούνται στις διάφορες εκδοχές του πειράματος και αυξάνονται τα μεγέθη δειγμάτων. Οι ερευνητές μπορούν να επιλέξουν το σχετικό ποσοστό των επισκεπτών που θα ομαδοποιούνται στην πρωτότυπη εκδοχή ή σε κάποια παραλλαγή, αριθμός που σχετίζεται με το χρονικό διάστημα εφαρμογής του πειράματος. Μία μεθοδολογικά καλή πρακτική είναι η εφαρμογή του πειράματος για τουλάχιστον μία εβδομάδα (για τη καταγραφή ενός πλήρους εβδομαδιαίου κύκλου) και κατά περίπτωση επέκταση του πειράματος για μεγαλύτερα χρονικά διαστήματα. Όταν υπάρχουν υπόνοιες ότι η καινοτομία μιας αλλαγής είναι αυτή που προκαλεί μία αρχική διαφοροποίηση στον αντικειμενικό στόχο, δηλαδή η αρχική επίδραση της παραλλαγής δεν είναι η ίδια με τη μακροπρόθεσμη επίδραση, τότε το πείραμα θα πρέπει να διεξαχθεί για αρκετό διάστημα ώστε να εκτιμηθεί το ασυμπτωτικό αποτέλεσμα της παραλλαγής.

Καταγραφή Παρατηρήσεων. Η σωστή καταγραφή των παρατηρήσεων και των μετρήσεων (δηλαδή η καταγραφή γεγονότων) είναι κρίσιμη για την επιτυχία του πειράματος. Όποτε είναι δυνατόν και οικονομικά εφικτό, θα πρέπει να συγκεντρώνονται όσο το δυνατόν περισσότερες παρατηρήσεις μπορούν να απαντήσουν πιθανά ζητήματα ενδιαφέροντος, είτε σχετίζονται με τον χρήστη είτε σχετίζονται με άλλες μεταβλητές παραμέτρους του πειράματος.

Κατανομή δείγματος - Τυχαιοποίηση (“Bucketing”). Οι χρήστες κατανέμονται σε ομάδες, που εκτίθενται στις διαφορετικές συνθήκες του πειράματος. Η βάση της διαδικασίας κατανομής είναι ουσιαστικά μία συνάρτηση του τύπου $\text{Math.random}() > 0.5$. Συχνά, αυτός ο διαχωρισμός είναι 50/50 (ή ομοιόμορφη κατανομή του δείγματος μεταξύ των παραλλαγών), αλλά πολλές φορές η κατανομή αυτή οφείλει να διαφέρει, όπως για παράδειγμα να ακολουθεί μία κατανομή 80/10/10. Μπορεί επίσης να χρησιμοποιηθούν κριτήρια επιλογής ή/και αποκλεισμού για το ποιοι χρήστες ή ποια περιβάλλοντα (για παράδειγμα έκδοση περιηγητή ιστού ή συσκευής) μπορούν να συμμετέχουν σε ένα πείραμα ή μία συγκεκριμένη

ομάδα κατανομής αυτού. Ορισμένες αλλαγές σε έναν ιστότοπο μπορεί να αφορούν σε όλους τους χρήστες που επισκέπτονται έναν ιστότοπο (για παράδειγμα μία αλλαγή στην αρχική σελίδα). Ωστόσο, στην πλειοψηφία των πειραμάτων, η παραλλαγή που εισάγεται αφορά ουσιαστικά σε ένα υποσύνολο επισκεπτών (για παράδειγμα μία αλλαγή στη διαδικασία πληρωμής, την οποία ξεκινά μόνο το 10% των επισκεπτών). Σε αυτές τις περιπτώσεις, είναι καλύτερο να συμπεριλαμβάνεται στο πείραμα μόνο το υποσύνολο του πληθυσμού που ενδεχομένως επηρεάστηκε (επισκέφθηκε την συγκεκριμένη σελίδα, για παράδειγμα). Η σωστή επιλογή δείγματος μειώνει τη μεταβλητότητα της εκτίμησης του αποτελέσματος του πειράματος, οδηγώντας σε ακριβέστερες εκτιμήσεις και ασφαλέστερα συμπεράσματα.

Ανάλυση δεδομένων - εξαγωγή συμπερασμάτων. Σε συνέχεια της ολοκλήρωσης του πειράματος και της συλλογής των καταγεγραμμένων δεδομένων, ακολουθεί η κατάλληλη ανάλυσή τους προς εξαγωγή των σχετικών συμπερασμάτων. Συνήθως, και ανάλογα με τον προηγούμενο σχεδιασμό του πειράματος, τα αποτελέσματα αυτά μπορούν να χρησιμοποιηθούν για την εξαγωγή αποφάσεων ανά τύπο συσκευής ή τύπο χρηστών, παίρνοντας τα αντίστοιχα τμήματα (segmentation) συσκευών και χρηστών που συμμετείχαν στο πείραμα. Μία κοινή προσέγγιση στην απόφαση του αν μία παραλλαγή είναι καλύτερη από την πρωτότυπη έκδοση είναι η χρήση παραμετρικών μεθόδων στον στατιστικό έλεγχο υποθέσεων, θεωρώντας την κατανομή των δεδομένων ως κανονική εφόσον το μέγεθος του δείγματος είναι ικανό [52]. Όταν η κανονικότητα των δεδομένων δεν πληρείται, εναλλακτικές αποτελούν οι μετασχηματισμοί των δεδομένων [8] και οι μη παραμετρικές μέθοδοι ή οι μέθοδοι αναδειγματοληψίας / μετάθεσης για τον προσδιορισμό του πόσο ασυνήθιστο είναι το παρατηρούμενο δείγμα κατά τη μηδενική υπόθεση [36]. Όταν διεξάγεται μία στατιστική δοκιμή για το αν μία παραλλαγή έχει αποτέλεσμα ή όχι (όπως μία δοκιμή για το αν οι μέσοι όροι των μετρήσεων της παραλλαγής και του πρωτότυπου είναι ίδιοι) παράγεται η τιμή σημαντικότητας p -value της στατιστικής δοκιμής. Πιο συγκεκριμένα, η τιμή σημαντικότητας p είναι η πιθανότητα να προκύψει ένα αποτέλεσμα ίσο ή περισσότερο ακραίο από το παρατηρούμενο, υποθέτοντας ότι η μηδενική υπόθεση είναι αληθής [7]. Μία άλλη εναλλακτική είναι η χρήση του θεωρήματος του Bayes στον υπολογισμό της εκ των υστέρων (posterior) ή δεσμευμένης πιθανότητας ότι μία παραλλαγή είχε ένα θετικό αντίκτυπο έναντι των πιθανοτήτων ότι δεν είχε αντίκτυπο [86].

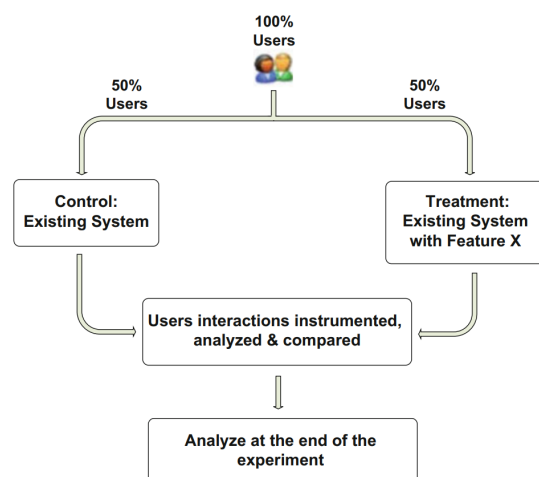
Διαστήματα εμπιστοσύνης. Τα περισσότερα συστήματα αναφοράς εμφανίζουν το αποτέλεσμα του πειράματος (πραγματική και ποσοστιαία αλλαγή) μαζί με κατάλληλα διαστήματα εμπιστοσύνης. Για εύλογα μεγάλα μεγέθη δειγμάτων, που γενικά θεωρούνται χιλιάδες χρήστες σε κάθε παραλλαγή, οι μέσες τιμές μπορούν να θεωρηθούν ότι ακολουθούν κανονικές κατανομές [53].

Μείωση απόκλισης. Η αύξηση του μεγέθους του δείγματος είναι ένας τρόπος για τη βελτίωση της εκτίμησης του αποτελέσματος. Ωστόσο, οι ερευνητές αναζητούν συνεχώς τρόπους για να αυξήσουν την εμπιστοσύνη των πειραμάτων τους ενώ μειώνουν

ή δεν αυξάνουν τη διάρκεια των δοκιμών. Ένας τρόπος για να επιτευχθεί αυτό είναι η χρήση πρόσθετων μεταβλητών όπως χαρακτηριστικά του χρήστη πριν από το πείραμα, δημογραφικά στοιχεία, εξοπλισμός και λογισμικό του χρήστη ή ταχύτητα σύνδεσης.

Διαγνωστικοί έλεγχοι και παρακολούθηση. Για την αξιολόγηση της επιτυχίας ή αποτυχίας μίας νέας έκδοσης θα χρειαστεί επίσης να ληφθούν δεδομένα σχετικά με τη συμπεριφορά των χρηστών στο πείραμα. Τα δεδομένα αυτά συνήθως περιλαμβάνουν την καταγραφή συμβάντων και ενεργειών του χρήστη ως μετρήσεις αξιολόγησης του πειράματος. Αυτά τα συμβάντα καταγράφονται και διατηρούνται σε μία κεντρική τοποθεσία για μετέπειτα ανάλυση. Για να διασφαλιστεί ότι τα πειραματικά αποτελέσματα είναι αξιόπιστα, κάθε σύστημα πειραματισμού θα πρέπει να έχει ενσωματωμένα κάποια διαγνωστικά εργαλεία. Οι αριθμοί των χρηστών σε κάθε παραλλαγή, οι μετρήσεις και τα αποτελέσματα ενός πειράματος με την πάροδο του χρόνου, είναι μόνο κάποια από τα στοιχεία που θα δώσουν την ευκαιρία στον ερευνητή να προλάβει έγκαιρα μη αναμενόμενα προβλήματα ή δυσλειτουργία του πειράματος.

Απομακρυσμένη διαμόρφωση. Η απομακρυσμένη διαμόρφωση βοηθάει στη δημιουργία και διαχείριση πειραμάτων, συνήθως από κάποια εξωτερική πλατφόρμα χωρίς την άμεση αλλαγή του κώδικα του ιστοτόπου. Στην πλειονότητα των περιπτώσεων, η ανάπτυξη μίας νέας έκδοσης μίας εφαρμογής είναι μία δυσκίνητη διαδικασία και συχνά, το προσωπικό που διαχειρίζεται τα πειράματα δεν είναι οι ίδιοι οι προγραμματιστές. Ιδανικά, οι διαχειριστές ή οι προωθητές των προϊόντων μπορούν να διαχειριστούν τη διαμόρφωση ενός ή περισσότερων πειραμάτων μέσω μίας ξεχωριστής διαδικασίας που είναι πιο φιλική προς χρήση για αυτούς. Εργαλεία όπως το Google OptimizeTM (Google LLC, Mountain View, US-CA) προσφέρουν έναν οπτικό επεξεργαστή (WYSIWYG) που επιτρέπει την εύκολη σχεδίαση ενός νέου πειράματος ακόμα και από μη εξειδικευμένο προσωπικό.



Εικόνα 4.1: Γενική δομή μίας δοκιμής α/β στο διαδίκτυο.

4.3 Προσεγγίσεις υλοποίησης

Υπάρχουν πολλές τεχνικές προσεγγίσεις για τη δημιουργία μίας δοκιμής α/β. Οι περισσότεροι χρησιμοποιούμενες είναι οι εξής:

Εκδόσεις Ελέγχου. Για τη δοκιμή μίας νέας παραλλαγής του ιστοτόπου, μπορεί να αναπτυχθεί μία νέα έκδοσή του σε ένα διαφορετικό branch από το κύριο branch (master branch) στο οποίο βρίσκεται από προεπιλογή ο κώδικάς του. Στη συνέχεια, ένα υποσύνολο χρηστών μπορεί να δρομολογηθεί προς τη νέα έκδοση (sticky sessions). Η εφαρμογή αυτής της στρατηγικής προϋποθέτει μία καλά διαχειριζόμενη υποδομή, ειδικά όταν δοκιμάζονται παράλληλα πολλαπλές δοκιμές, δηλαδή όταν προσφέρονται ταυτόχρονα πολλές διαφορετικές εκδόσεις του ιστοτόπου. Επιπλέον, απαιτεί τη σωστή κατανομή των χρηστών στις διάφορες εκδόσεις του ιστοτόπου, ώστε ο αριθμός τους να είναι επαρκής για αξιόπιστα αποτελέσματα. Πλεονέκτημα αυτής της τεχνικής είναι ότι ένα αποτυχημένο πείραμα δεν εισάγει ανάγκη για τεχνικές αλλαγές στον κώδικα του ιστοτόπου, καθώς η αποτυχημένη έκδοση δεν βρίσκεται στο master branch και οι υπό δοκιμή αλλαγές απλά διαγράφονται.

Διαχωρισμός διευθύνσεων URL. Με στόχο την αποτροπή προβλημάτων SEO (Search Engine Optimization), συστήνεται η χρησιμοποίηση διαφορετικών διευθύνσεων URL για την ανακατεύθυνση των χρηστών σε διαφορετικές παραλλαγές του πειράματος. Το πλεονέκτημα αυτής της προσέγγισης είναι ότι κατά τη διάρκεια του πειράματος δεν επηρεάζεται αρνητικά οποιαδήποτε τιμή SEO μίας διεύθυνσης URL.

Πλευρά διακομιστή. Οι χρήστες ομαδοποιούνται στον διακομιστή κατά την πρόσβαση σε μία σελίδα. Στη συνέχεια, ορίζεται ένα cookie για να εξασφαλιστεί ότι ο χρήστης θα παραμείνει στην ίδια έκδοχή της σελίδας κατά τη διάρκεια του πειράματος. Καθώς, μάλιστα, η ιστοσελίδα υλοποιείται στην πλευρά του διακομιστή, ο αντίκτυπος στην εμπειρία του χρήστη, αναφορικά με την απόδοση, είναι αμελητέος.

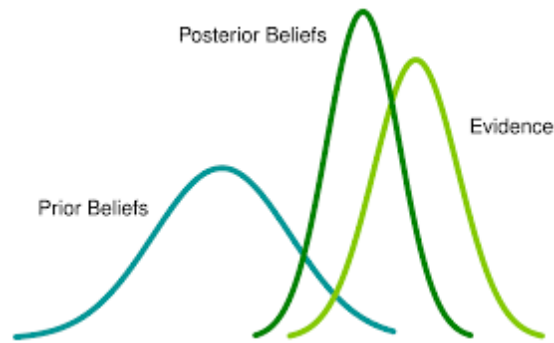
Πλευρά χρήστη (client-side). Στις περιπτώσεις όπου δεν είναι δυνατή η πρόσβαση στον διακομιστή ή απαιτείται μέγιστη ευελιξία, η δοκιμή α/β στην πλευρά του περιηγητή του χρήστη αποτελεί, επίσης, επιλογή. Σε αυτήν την περίπτωση, αφότου διαμορφωθεί η πρωτότυπη έκδοση της σελίδας στον περιηγητή του χρήστη, εφαρμόζονται οι όποιες αλλαγές αντιστοιχούν στην ομάδα του πειράματος στην οποία έχει περιληφθεί ο χρήστης. Αυτή η επιλογή βρίσκει εφαρμογή όταν δεν υπάρχει πρόσβαση σε εξειδικευμένο προσωπικό και γίνεται χρήση εξωτερικών εργαλείων για την εκτέλεση πειραμάτων. Ωστόσο, είναι συχνά η χειρότερη επιλογή όσον αφορά στην απόδοση, καθώς η εφαρμογή των αλλαγών περιμένει την προηγούμενη διαμόρφωση της πρωτότυπης έκδοσης.

4.4 Υλοποίηση με την πλατφόρμα βελτιστοποίησης Google Optimize™

Υποθέτοντας ότι η χρήση της αυτόματης επισήμανσης των εικόνων μπορεί να βελτιώσει τις επιλογές των διαφημιζόμενων προϊόντων από τους χρήστες, κρίθηκε σκόπιμο να πραγματοποιηθεί μία δοκιμή α/β για να τον έλεγχο της υπόθεσης. Για την εργασία αυτή χρησιμοποιήθηκε η πλατφόρμα βελτιστοποίησης της Google™, Google Optimize™ καθώς προσφέρει μία σειρά πλεονεκτημάτων. Συγκεκριμένα, η πλατφόρμα προσφέρει ενσωμάτωση με το Google Analytics™ (Google LLC, Mountain View, US-CA) για την ανάλυση των αντικειμενικών στόχων που έχουν οριστεί ως τέτοιοι στο Google Analytics™ και δίνει τη δυνατότητα δημιουργίας νέων παραλλαγών που εφαρμόζονται απευθείας στον περιηγητή του χρήστη. Για τη δυνατότητα αυτή, προσφέρει έναν οπτικό επεξεργαστή της ιστοσελίδας (WYSIWYG), όπου οι περισσότερες αλλαγές μπορούν να πραγματοποιηθούν από μη εξειδικευμένο προσωπικό. Επιπλέον, υποστηρίζει δοκιμές α/β ή πολλαπλών παραλλαγών και χρησιμοποιεί Bayesian στατιστικές μεθόδους για να μοντελοποιήσει την απόδοση σε πραγματικές συνθήκες και να παρέχει ακριβέστερα αποτελέσματα. Μάλιστα, πραγματοποιεί τη στατιστική ανάλυση και επιτρέπει την παρακολούθηση της πορείας του πειράματος και των διαφόρων παραλλαγών σε πραγματικό χρόνο καθ' όλη τη διάρκεια της διαδικασίας. Επιπλέον, η πλατφόρμα δίνει τη δυνατότητα στον χρήστη να διαχειρίζεται πολλαπλά πειράματα που ενδέχεται να εκτελούνται ταυτόχρονα στον ιστότοπο.

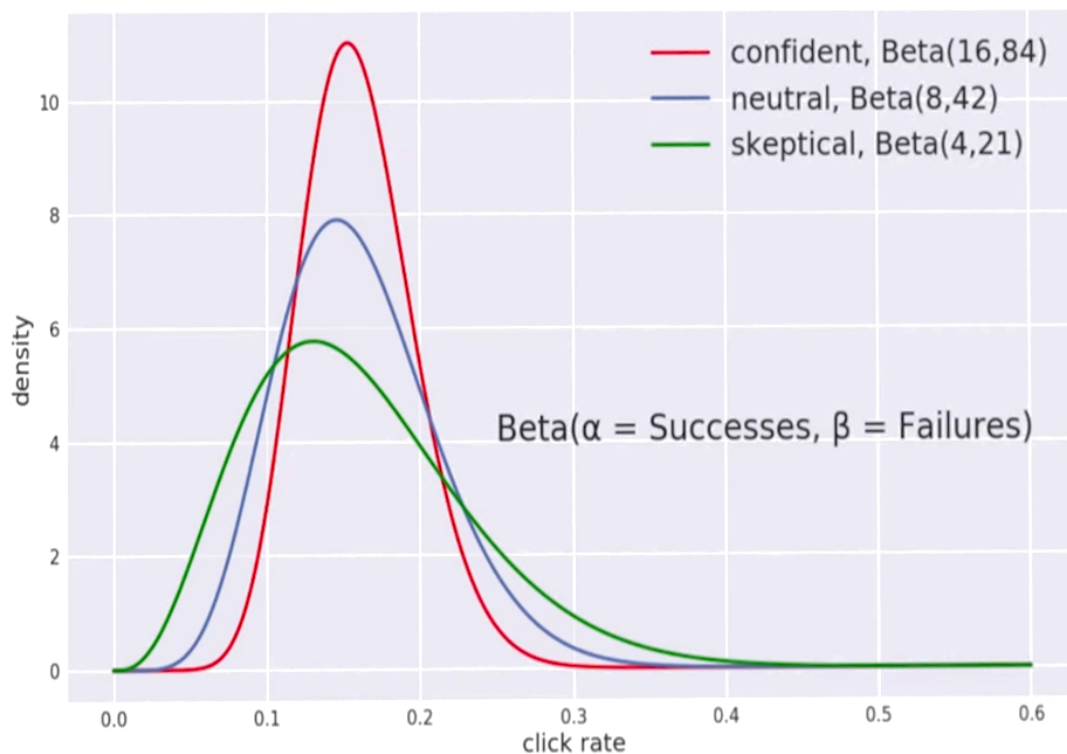
Το Google Optimize™ χρησιμοποιεί Bayesian ανάλυση (Bayesian inference) για την ανάλυση και εξαγωγή αποτελεσμάτων. Η Bayesian ανάλυση είναι μία μέθοδος στατιστικής ανάλυσης που επιτρέπει στο Google Optimize™ να βελτιστοποιεί διαρκώς τα αποτελέσματά του όσο λαμβάνονται νέα δεδομένα. Αν και υπολογιστικά ακριβή, η Bayesian ανάλυση επιτρέπει τον υπολογισμό πιθανοτήτων απευθείας ώστε να είναι σε θέση να απαντήσει ερωτήσεις όπως “πόσο πιθανό είναι αυτή η εκδοχή να είναι καλύτερη από την υπάρχουσα;” ή “πόσο πιθανό είναι αυτή η εκδοχή να είναι η καλύτερη συνολικά;”, οι οποίες είναι πιο κοντά στην ανθρώπινη σκέψη. Αποφεύγει τα προβλήματα που σχετίζονται με τις προσεγγίσεις ελέγχου υποθέσεων, όπως την κοινή παρανόηση ότι οι τιμές σημαντικότητας αντιπροσωπεύουν τις πιθανότητες. Επιπλέον, η χρήση Bayesian στατιστικών μεθόδων επιτρέπει τον τερματισμό ενός πειράματος τη στιγμή που διαπιστωθεί ότι το περιθώριο απόκτησης νέας γνώσης σε συνέχεια του πειράματος είναι μικρό. Η προσέγγιση του Google Optimize™ μπορεί να ερευνήσει τόσο την απόδοση των συνδυασμών, του ενός απέναντι στους άλλους, όσο και την απόδοση μίας παραλλαγής σε διάφορους συνδυασμούς. Συνολικά, είναι δυνατό να ελέγξουμε όλους τους συνδυασμούς, αλλά και να έχουμε αποτελέσματα πολύ πιο γρήγορα από μία αντίστοιχη δοκιμή α/β. Επειδή οι Bayesian μέθοδοι υπολογίζουν άμεσα τη σχετική απόδοση όλων των εκδοχών μαζί και όχι μόνο κατά ζεύγη, οι ερευνητές δεν χρειάζεται να εκτελούν πολλαπλές συγκρίσεις παραλλαγών για να κατανοήσουν τον τρόπο

με τον οποίο ο καθένας είναι πιθανόν να αποδώσει.



Εικόνα 4.2: Απεικόνιση της γενικής ιδέας της χρήσης βήτα κατανομών για την έκφραση μίας *posterior* κατανομής με βάση τα νέα δεδομένα (*evidence*) και την πρότερη γνώση (*prior belief*).

Αν και τα ακριβή μοντέλα που χρησιμοποιεί ο αλγόριθμος της GoogleTM, δεν είναι γνωστά, το γενικότερο πλαίσιο της Bayesian μεθοδολογίας για τη δοκιμή α/β απεικονίζεται στην Εικόνα 4.2. Το πρώτο στάδιο οποιασδήποτε bayesian ανάλυσης είναι η κατανόηση και ο σχεδιασμός της πρότερης γνώσης για τα δεδομένα. Αυτή η γνώση μοντελοποιείται με βήτα κατανομές $Beta(\alpha, \beta)$ με την παράμετρο α να διαμορφώνεται από τις επιτυχίες και την παράμετρο β από τις αποτυχίες των πειραμάτων σε παρελθοντικά δεδομένα (Εικόνα 4.3).



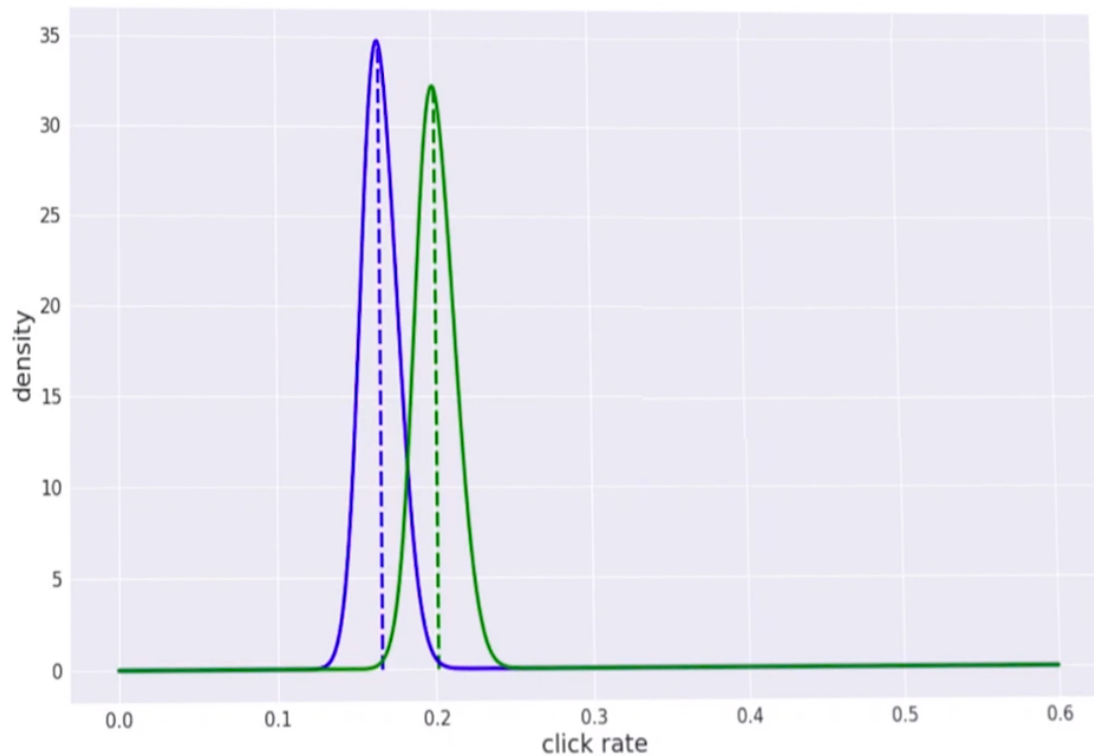
Εικόνα 4.3: Παράδειγμα βήτα κατανομών για διάφορες τιμές των παραμέτρων α , β .

Θεωρούνται δύο όμοιες κατανομές βήτα $Beta(\alpha, \beta)$, μία για κάθε παραλλαγή. Αυτές οι

κατανομές μοντελοποιούν την πρότερη γνώση και το αντίστοιχο διάστημα εμπιστοσύνης. Η γνώση αυτή συνδιαμορφώνεται από διάφορες πηγές όπως προηγούμενες δοκιμές ή δεδομένα της αγοράς. Στη συνέχεια, νέοι χρήστες εκτίθενται στις δύο παραλλαγές και οι δύο κατανομές αναπροσαρμόζονται με βάση τα νέα δεδομένα που εισέρχονται στο πείραμα καθώς οι χρήστες κατανέμονται στις δύο παραλλαγές και παράγουν δεδομένα. Σε βάθος χρόνου, με βάση τις επιτυχίες (που αυξάνουν τον παράγοντα α) και τις αποτυχίες (που αυξάνουν τον παράγοντα β) της δοκιμής, οι δύο κατανομές αναδιαμορφώνονται από τα πραγματικά δεδομένα.

$$Beta_A(\alpha + W_A, \beta + L_A), \quad (4.1)$$

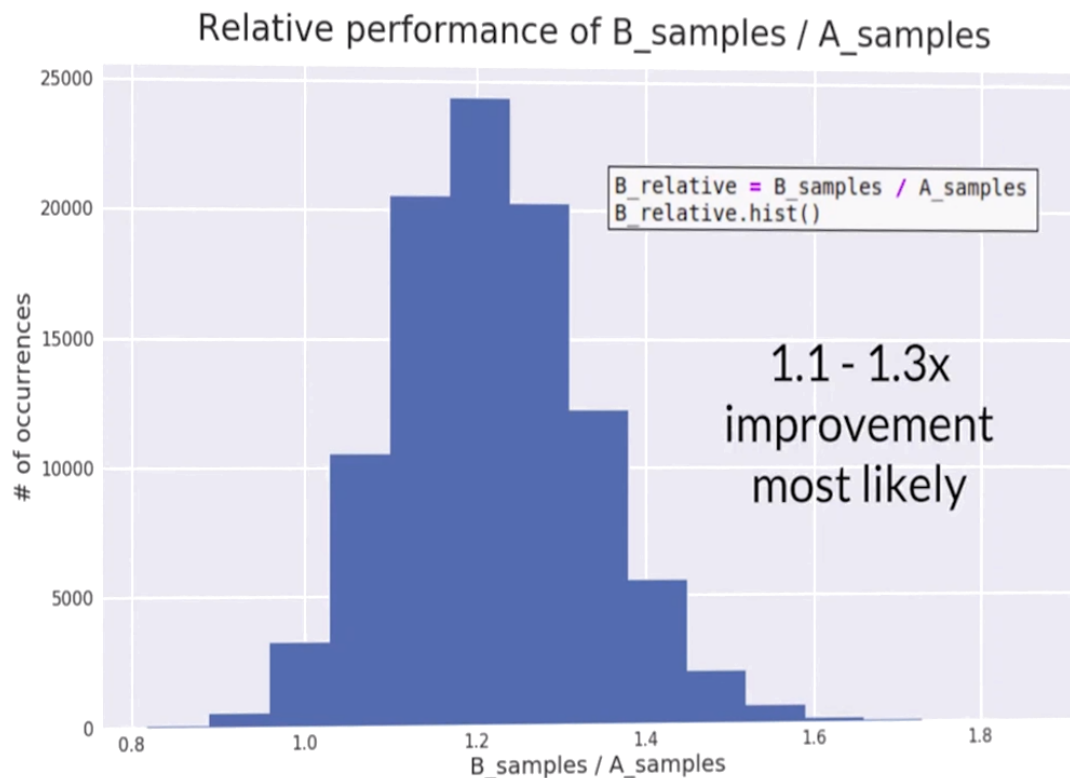
$$Beta_B(\alpha + W_B, \beta + L_B) \quad (4.2)$$



Εικόνα 4.4: Παράδειγμα εμφάνισης επικάλυψης μεταξύ των δύο posterior κατανομών.

Η αναδιαμόρφωση της πρότερης γνώσης από τα πραγματικά δεδομένα, οδηγεί στη δημιουργία της posterior κατανομής. Αν υπάρχει επικάλυψη μεταξύ των δύο posterior κατανομών (Εικόνα 4.4) (αν δεν είναι επαρκή τα δεδομένα, ή αν η δοκιμή δεν οδήγησε σε σαφή αποτελέσματα), τότε για τη λήψη της απόφασης, πραγματοποιείται συνήθως μία προσομοίωση Monte Carlo. Λαμβάνονται n δείγματα από την $Beta_A$ και την $Beta_B$. Ο αριθμός των επαναλήψεων στις οποίες το δείγμα X_B της κατανομής B ήταν μεγαλύτερο από το δείγμα X_A της κατανομής A είναι ο αριθμός B_{wins} . Το ποσοστό, επομένως, των επαναλήψεων στις οποίες ισχύει $X_B > X_A$ είναι B_{wins}/n . Αν το ποσοστό αυτό είναι για παράδειγμα 98%, αυτό

σημαίνει ότι στο 98% των επαναλήψεων η έκδοση B ήταν καλύτερη από την A, το οποίο θα μπορούσε να αντιστοιχηθεί σε έναν έλεγχο υποθέσεων με μία τιμή $p = 0.02$ και επομένως τα αποτελέσματα είναι στατιστικά σημαντικά. Ένα ενδιαφέρον διάγραμμα που μπορεί να γίνει από τα δείγματα της προσομοίωσης Monte Carlo είναι η σχετική βελτίωση που εκτιμάται για τη μία έκδοση έναντι της άλλης μέσω του ιστογράμματος της τιμής των $B_samples / A_samples$ (Εικόνα 4.5).

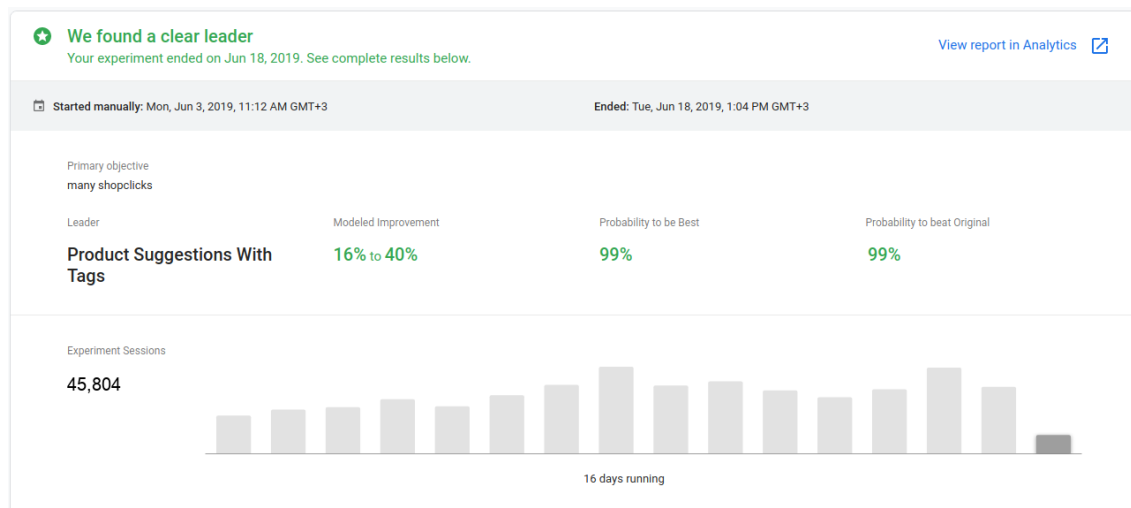


Εικόνα 4.5: Παράδειγμα διαγράμματος σχετικής απόδοσης των εκδόσεων A και B, με βάση τα δείγματα της προσομοίωσης Monte Carlo.

Στην πλατφόρμα βελτιστοποίησης της GoogleTM, κάθε παραλλαγή παρέχεται σε παρόμοιους χρόνους, έτσι ώστε η απόδοσή της να μπορεί να παρατηρηθεί και να μετρηθεί ανεξάρτητα από άλλους εξωτερικούς παράγοντες. Επίσης, το πείραμα μπορεί να περιοριστεί σε συγκεκριμένο στοχευμένο κοινό. Η πλατφόρμα καταγράφει τα δεδομένα απόδοσης για κάθε παραλλαγή σε σχέση με έναν αντικειμενικό στόχο και προσδιορίζει την επικρατέστερη. Πρόσθετα, καθ' όλη τη διάρκεια του πειράματος η πλατφόρμα Google OptimizeTM προσφέρει διαγνωστικό έλεγχο για τη λειτουργία του πειράματος και ενημερώνει τον ερευνητή για τυχόν δυσλειτουργίες. Επιπλέον, τα δεδομένα εξάγονται από τους υποκείμενους πίνακες δεδομένων του Google AnalyticsTM στην ολότητά τους και όχι δειγματοληπτικά.

Το πείραμα εφαρμόστηκε μόνο στις σελίδες των εικόνων χρηστών (targeting rule) με καθορισμό της κατανομής των χρηστών 1:1 για κάθε έκδοχή (traffic allocation). Η πρωτότυπη εκδοχή των σελίδων των εικόνων χρηστών, δεν έκανε χρήση των ετικετών

για την παραγωγή συστάσεων, ενώ η παραλλαγμένη εκδοχή χρησιμοποιούσε τις ετικέτες των αντικειμένων που περιείχονταν στην εικόνα, όπως αυτές εξήχθησαν από το YOLO, για να παράξει τις συστάσεις προϊόντων κάτω από την εικόνα του χρήστη. Ως πρωτεύων αντικειμενικός στόχος για την ανάλυση του πειράματος ορίστηκε ο αριθμός των επιλεγμένων προϊόντων από τους χρήστες (objective goal). Στις Εικόνες 4.6 έως 4.10 φαίνεται η αναφορά του εργαλείου Google OptimizeTM για την έκβαση του πειράματος.



Εικόνα 4.6: Γενικά στοιχεία και αποτελέσματα στατιστικής ανάλυσης α/β δοκιμής με το εργαλείο Google OptimizeTM.

Η αναφορά δείχνει την τρέχουσα κατάσταση του πειράματος για τον δεδομένο αντικειμενικό στόχο. Στην Εικόνα 4.6 φαίνονται τα αποτελέσματα για τον αντικειμενικό στόχο “many shopclicks”, τα οποία αφορούν στις επιλογές των διαφημιζόμενων προϊόντων από τους χρήστες. Το Google OptimizeTM αποφάνθηκε ότι τα δεδομένα που συνέλεξε σε διάστημα 16 ημερών ήταν αρκετά για να συμπεράνει ότι μία από τις παραλλαγές του πειράματος απέδωσε καλύτερα από την πρωτότυπη εκδοχή. Συγκεκριμένα, συμπέρανε ότι η παραλλαγμένη εκδοχή έχει 99% πιθανότητα να είναι καλύτερη της πρωτότυπης (probability to beat original) και συνολικά η καλύτερη (probability to be best) δεδομένου ότι δεν υπήρχαν άλλες παραλλαγές στη δοκιμή. Αυτό σημαίνει ότι η δεδομένη παραλλαγή θα οδηγήσει σε ένα ποσοστό επίτευξης των αντικειμενικών στόχων που έχουν τεθεί για το πείραμα, καλύτερο συγκρινόμενο με αυτό της πρωτότυπης εκδοχής. Συνεπώς, ήταν πλέον ασφαλής ο τερματισμός του πειράματος και προτεινέτο η εφαρμογή της παραλλαγμένης εκδοχής για όλους τους χρήστες (ως προεπιλογή).

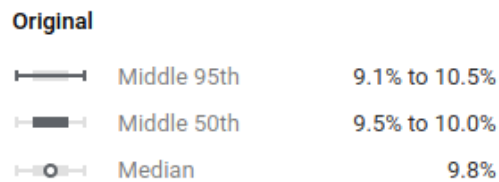
Στο αριστερό μέρος της καρτέλας αναφοράς φαίνονται οι συνολικές συνεδρίες των χρηστών, δηλαδή οι συνολικοί χρήστες που συμμετείχαν στο πείραμα, είτε εκτέθηκαν στην παραλλαγμένη έκδοση είτε στην πρωτότυπη. Συνολικά καταγράφηκαν 45804 συνεδρίες. Ο αριθμός της βελτίωσης του μοντέλου (model improvement) 16%-40% αναφέρεται στη διαφορά του ρυθμού μετατροπής της παραλλαγής και της βάσης (baseline) για τον δεδομένο

Ανάπτυξη συστήματος αυτόματης επισήμανσης εικόνων χρηστών διαδικτυακής κοινότητας με σκοπό τη βελτίωση του συστήματος συστάσεων σχετικών προϊόντων

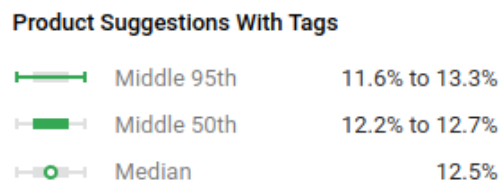
αντικειμενικό στόχο. Η βελτιστοποίηση χρησιμοποιεί Bayesian ανάλυση για να καθορίσει τον τρόπο με τον οποίο οι παραλλαγές θα αποδώσουν στο μέλλον και η βελτίωση που παρουσιάζεται είναι το Bayesian ισοδύναμο ενός διαστήματος εμπιστοσύνης.

many shopclicks					
☐ Variant ↑	Modeled Improvement	Probability to be Best	Probability to beat Original	Modeled Conversion Rate	Experiment Conversions
☒ Original 22,482 sessions	Baseline	1%	Baseline	9.1% - 10.5%	2,198
☒ Product Suggestions With... 23,322 sessions	16% to 40%	99%	99%	11.6% - 13.3%	2,905

Εικόνα 4.7: Αποτελέσματα στατιστικής ανάλυσης του Google OptimizeTM για τον αντικειμενικό στόχο "shopclicks".

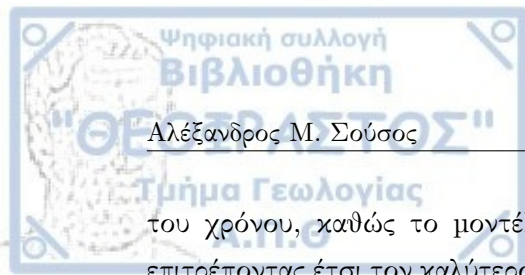


Εικόνα 4.8: Τα στατιστικά στοιχεία της ομάδας της πρωτότυπης έκδοσης του ιστοτόπου σύμφωνα με τη στατιστική ανάλυση του Google OptimizeTM.



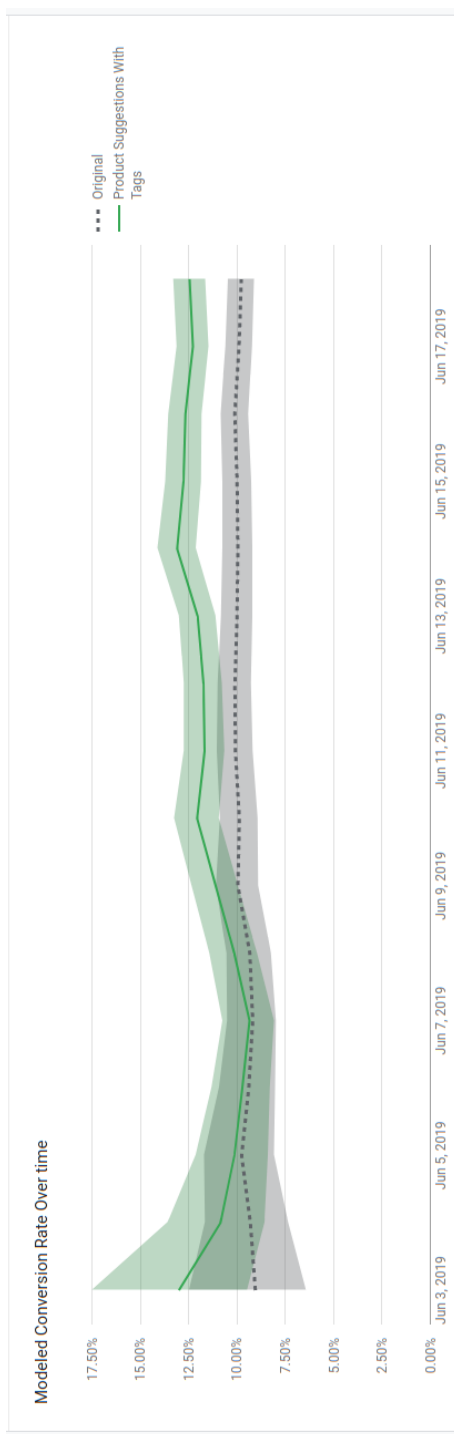
Εικόνα 4.9: Τα στατιστικά στοιχεία της ομάδας της παραλλαγμένης έκδοσης του ιστοτόπου σύμφωνα με τη στατιστική ανάλυση του Google OptimizeTM.

Κάτω από τα περιγραφικά δεδομένα και την ανάλυση του πειράματος, εμφανίζεται το διάγραμμα που δείχνει τον ρυθμό μετατροπής της παραλλαγμένης έκδοσης σε σύγκριση με την πρωτότυπη σε συνάρτηση με τον χρόνο (Εικόνα 4.10). Στο διάγραμμα απεικονίζονται οι επιδόσεις των παραλλαγών αναφορικά με τον αντικειμενικό στόχο. Οι έγχρωμες περιοχές αντιπροσωπεύουν το εύρος απόδοσης που η πρωτότυπη έκδοση και οι παραλλαγές είναι πιθανό να έχουν στο 95% του χρόνου. Η γραμμή στη μέση μίας περιοχής δείχνει τη μέση τιμή της περιοχής. Στην αρχή ενός πειράματος υπάρχει μεγαλύτερη αβεβαιότητα για την απόδοση κάθε παραλλαγής, με αποτέλεσμα μεγαλύτερα διαστήματα (μεγαλύτερες σκιασμένες περιοχές). Στην πλειονότητα των περιπτώσεων, το εύρος κάθε παραλλαγής θα περιοριστεί με την πάροδο



Αλέξανδρος Μ. Σούσος

του χρόνου, καθώς το μοντέλο βελτιστοποίησης λαμβάνει υπόψη περισσότερα δεδομένα, επιτρέποντας έτσι τον καλύτερο προσδιορισμό του τρόπου με τον οποίο η παραλλαγή αποδίδει έναντι του πρωτότυπου.



Εικόνα 4.10: Ο ρυθμός μετατροπής (conversion rate) του αντικειμενικού στόχου "shopclicks" της παραλλαγμένης έκδοσης σε σύγκριση με την πρωτότυπη στην πορεία του χρόνου σύμφωνα με τη στατιστική ανάλυση του Google OptimizeTM.



Αλέξανδρος Μ. Σούσος

Κεφάλαιο 5

Μελλοντικές Επεκτάσεις

Όπως υποδείχθηκε από τα αποτελέσματα της δοκιμής α/β σχετικά με την εφαρμογή της ανίχνευσης αντικειμένων στις εικόνες των χρηστών της διαδικτυακής πλατφόρμας, το σύστημα παραγωγής προτάσεων προϊόντων βελτιώθηκε (Εικόνα 4.6). Το προτεινόμενο σύστημα της παρούσας εργασίας είναι μόνο η αρχή για τη βελτίωση των συστάσεων σχετικών προϊόντων κάτω από τις εικόνες χρηστών. Οι δυνατότητες για περαιτέρω βελτίωση του συστήματος είναι πολλές και στηρίζονται, κυρίως, στις τελευταίες εξελίξεις του επιστημονικού τομέα της μηχανικής όρασης, οι οποίες επιτρέπουν την καλύτερη "κατανόηση" μίας εικόνας.

Μία προφανής βελτίωση του υπάρχοντος συστήματος είναι η εκπαίδευση του συστήματος σε συγκεκριμένες κλάσεις για την αναγνώριση του συγκεκριμένου μοντέλου των προϊόντων που εμφανίζονται σε μία εικόνα, αντί για τη γενική κλάση του αντικειμένου. Η εκπαίδευση ενός νευρωνικού δικτύου μπορεί επίσης να επιτευχθεί για μία πληθώρα άλλων σημασιολογικών κατηγοριοποιήσεων, όπως η εξαγωγή κλάσεων που περιγράφουν την υφή του ανιχνευμένου προϊόντος, την εξαγωγή κλάσεων για την περιγραφή του γενικότερου σκηνικού που περιγράφει μία εικόνα (για παράδειγμα "σαλόνι" ή "κουζίνα") ή την εξαγωγή κλάσεων σχετικών με το στυλ του απεικονιζόμενου δωματίου (για παράδειγμα "σκανδιναβικό στυλ"). Παραδείγματος χάρη, η υφή μπορεί είτε να χρησιμοποιηθεί αυτόνομα, ως ένα φίλτρο βοήθειας του χρήστη για την πλοήγηση στον κατάλογο των προϊόντων του ιστοτόπου, είτε να ενσωματωθεί στο σύστημα παραγωγής συστάσεων προϊόντων κάτω από τις εικόνες χρηστών, φιλτράροντας ουσιαστικά τις τελικές συστάσεις [17, 6, 70, 3].

Για την εκπαίδευση ενός προσαρμοσμένου μοντέλου, απαιτούνται επισημασμένα δεδομένα. Τα επισημασμένα δεδομένα στην διαδικασία της ανίχνευσης αντικειμένων είναι εικόνες με ορισμένες τις συντεταγμένες του πλαισίου οριοθέτησης του αντικειμένου και την αντίστοιχη κλάση ταξινόμησης του. Πριν την εκπαίδευση ενός μοντέλου είναι σημαντικό να κατανοηθούν σωστά τα σενάρια για τα οποία το μοντέλο θα αναπτυχθεί. Είναι σημαντικό να διατίθεται ένας μεγάλος αριθμός αντιπροσωπευτικών εικόνων ανά κατηγορία (ενδεχομένως και περισσότερες από 1000 για κάποιες εφαρμογές). Αντιπροσωπευτικές εικόνες σε αυτό το πλαίσιο νοούνται οι εικόνες που ανταποκρίνονται σε όλο το φάσμα των σεναρίων στο οποίο το μοντέλο θα χρησιμοποιηθεί. Για παράδειγμα, σε ένα μοντέλο που θα χρησιμοποιηθεί σε αυτοκίνητο για να

ανιχνεύει τις πινακίδες ή τους φωτεινούς σηματοδότες των δρόμων, θα πρέπει να υπάρχουν αντιπροσωπευτικές εικόνες που έχουν ληφθεί σε διάφορες συνθήκες καιρού, φωτεινότητας, θέσης κάμερας και λοιπές παραμέτρους. Εάν το μοντέλο δεν έχει αρκετά δεδομένα για να μάθει τα απαιτούμενα στη χρήση γενικά μοτίβα, δεν θα αποδώσει καλά στην πραγματική χρήση. Ο όγκος των επισημασμένων εικόνων και το κόστος σε ανθρώπινο δυναμικό για την επισημάνση του απαιτούμενου αριθμού αντιπροσωπευτικών εικόνων αποτελούν συνήθως περιοριστικούς ή και αποτρεπτικούς παράγοντες για την εκπαίδευση εξειδικευμένων ανιχνευτών.

Μία ακόμα βελτίωση του συστήματος, η οποία μπορεί να ενταχθεί, επίσης, στη βελτίωση της εμπειρίας χρήστη, είναι η δυνατότητα επισημάνσης συγκεκριμένων σημείων της εικόνας από τους χρήστες βάσει προτάσεων περιοχών από το σύστημα ανίχνευσης αντικειμένων. Η λειτουργία αυτή, μπορεί να προσφέρει στον χρήστη τη δυνατότητα να σχολιάσει μέσω επισημάνσεων συγκεκριμένες περιοχές της εικόνας. Σε μεταγενέστερο χρόνο, οι επισημάνσεις αυτές μπορούν να χρησιμοποιηθούν για την περαιτέρω εκπαίδευση ενός συστήματος μέσω μετάδοσης γνώσης (transfer learning) [90]).

Πρόσθετα, μία μελλοντική επέκταση της παρούσας εργασίας είναι η χρήση των συνελκτικών νευρωνικών δικτύων σε συνδυασμό με κάποιον αλγόριθμο εύρεσης ομοιότητας εικόνων, ώστε το σύστημα, αναζητώντας στη βάση δεδομένων της εταιρείας, να προσφέρει στον χρήστη το πλέον όμοιο προϊόν με αυτό που εντοπίστηκε από το YOLO. Η αναζήτηση αυτή μπορεί να επιτευχθεί με κάποιον αλγόριθμο ομοιότητας εικόνας (όπως Nearest Neighbour Search [68]). Σκοπό αυτής της πρότασης αποτελεί η εξαγωγή αντικειμένων επίπλων από μία εικόνα που περιλαμβάνει πολλαπλά αντικείμενα επίπλων και στη συνέχεια η εύρεση παρόμοιων αντικειμένων στη βάση δεδομένων της εταιρείας, δηλαδή η ανάκτηση αντικειμένων εσωτερικού χώρου από τη βάση δεδομένων, τα οποία μοιράζονται οπτικές και αισθητικές ομοιότητες με το απεικονιζόμενο αντικείμενο [97]. Ένα σύστημα εύρεσης όμοιων εικόνων συνήθως περιλαμβάνει έναν εξαγωγέα χαρακτηριστικών από τις εικόνες και έναν αλγόριθμο σύγκρισης απόστασης των διανυσμάτων χαρακτηριστικών [19]. Οι κλάσεις που εξάγονται από την εφαρμογή του YOLO μπορούν να χρησιμοποιηθούν για επιπρόσθετο λεξιλογικό έλεγχο ομοιότητας [87] (για παράδειγμα μέσω του αλγορίθμου "bag of words" [4]). Τα τελικά διανύσματα χαρακτηριστικών ενός ανιχνευτή αντικειμένων (για παράδειγμα το προτελευταίο επίπεδο στο CNN του YOLO), μπορούν να χρησιμοποιηθούν για τον υπολογισμό της ομοιότητας των εικόνων [96, 87]. Η διαδικασία αυτή, της εύρεσης όμοιων αν όχι των ίδιων ακριβώς προϊόντων, είναι γνωστή ως οπτική αναζήτηση (visual search [94]). Η οπτική αναζήτηση που συνδυάζει την αξιοποίηση των λεξιλογικών πληροφοριών που ενδεχομένως συνοδεύουν τις εικόνες, οδηγεί σε καλύτερα αποτελέσματα.

Επιπλέον, περαιτέρω βελτίωση του συστήματος μπορεί να επιτευχθεί με τη χρήση των τελευταίων εξελίξεων και σε άλλα πεδία της μηχανικής όρασης. Με τη χρήση CNN και αναδρομικών νευρωνικών δικτύων (Recurrent Neural Networks, RNN), είναι δυνατή

η παραγωγή ενός κείμενου που να περιγράφει μία εικόνα. Το κείμενο αυτό μπορεί να χρησιμοποιηθεί για λόγους προσβασιμότητας (ως αυτόματα παραγόμενο εναλλακτικό κείμενο για τις εικόνες της πλατφόρμας) ή και για επιπλέον λεξιλογική ανάλυση για την εύρεση όμοιων εικόνων. Πρόσθετα, η εξέλιξη στο πεδίο της μεταφοράς τεχνοτροπίας εικόνων επιτρέπει την αυτόματη εναλλαγή στυλ ενός δωματίου, ενδεχομένως ως φίλτρο στη φωτογραφία του χρήστη [32, 92]. Επιπρόσθετα, η εισαγωγή και εξέλιξη των Παραγωγικών Αντιπαραθετικών Δικτύων (Generative Adversarial Networks, GANs) [37, 50] επιτρέπουν, πλέον, την παραγωγή νέων αληθοφανών μη πραγματικών εικόνων, βασισμένων στην υπάρχουσα βάση δεδομένων και μπορούν να χρησιμοποιηθούν για τη σύνθεση εικονικών δωματιών.

Τέλος, τα συνιστώμενα συστήματα επιδέχονται περαιτέρω βελτίωσης, βάση όχι μόνο των χαρακτηριστικών των εικόνων αλλά και των διασυνδέσεων των χρηστών (συστήματα συστάσεων collaborative filtering έναντι των content-based). Ένα τέτοιο σύστημα μπορεί να οδηγήσει σε εξατομικευμένες συστάσεις που για να παραχθούν αξιοποιούν την πληροφορία των συνδέσεων μεταξύ των χρηστών και των προσωπικών τους επιλογών. Για παράδειγμα, στην περίπτωση της διαδικτυακής πλατφόρμας, ένας χρήστης Α μπορεί να "ακολουθήσει" άλλους χρήστες, έτσι ώστε να μπορεί να έχει μία εξατομικευμένη ροή για τις εικόνες τους. Αυτή η ενέργεια μπορεί να ερμηνευτεί ως το ενδιαφέρον του χρήστη Α για τις στυλιστικές επιλογές των φωτογραφιών των άλλων χρηστών και επομένως οι επιλογές των συνδεδεμένων χρηστών να συνυπολογιστούν στην παραγωγή συστάσεων προς τον χρήστη Α. Επιπρόσθετα, τα προτεινόμενα προϊόντα έχουν και αυτά κάποιες εσωτερικές διασυνδέσεις ανάλογα με την κατηγορία τους, το χρώμα τους, την επωνυμία τους ή οποιοδήποτε άλλο χαρακτηριστικό μπορεί να εξαχθεί και να χρησιμοποιηθεί ως κριτήριο ομοιότητας μεταξύ των προϊόντων. Ένα τέτοιο πολύπλοκο σύστημα μπορεί να αναχθεί σε ένα διμερές δίκτυο, όπου στο ένα επίπεδο βρίσκονται οι κόμβοι των προϊόντων και στο άλλο επίπεδο οι κόμβοι των χρηστών [13, 66, 85, 30, 103]. Συνδέσεις υπάρχουν τόσο μεταξύ των προϊόντων (ομοιότητα προϊόντων) όσο και μεταξύ των χρηστών (διασυνδέσεις χρηστών) αλλά και μεταξύ των επιπέδων συμβολίζοντας τις επιλογές προϊόντων από χρήστες [103].

Οι παραδοχές (για τους κόμβους και τις συνδέσεις του δικτύου) που πρέπει να γίνουν για τη δημιουργία ενός τέτοιου πολύπλοκου συστήματος είναι κρίσιμες για την απόδοση του συστήματος και η ρύθμιση του για τη σωστή αξιοποίηση όλων των διαθέσιμων πληροφοριών, είναι πολύπλοκη, εξειδικευμένη ανά περίπτωση και απαιτεί πολλαπλές δοκιμές για τη βελτιστοποίησή του.



Αλέξανδρος Μ. Σούσος



Αναφορές

- [1] Anderson, C. “The Long Tail”. In: *Wired Magazin* (Oct. 2004). ISSN: 1059-1028. URL: <https://www.wired.com/2004/10/tail/> (visited on 11/19/2019).
- [2] Anderson, C. “Long Tail : Why the Future of Business Is Selling Less of More”. In: *Hyperion* (2008), p. 282.
- [3] Andrearczyk, V. and Whelan, P. F. “Texture segmentation with Fully Convolutional Networks”. In: *CoRR* (Mar. 15, 2017). arXiv: 1703.05230. URL: https://www.researchgate.net/publication/315096628_Texture_segmentation_with_Fully_Convolutional_Networks (visited on 12/04/2019).
- [4] *Bag-of-words model*. In: *Wikipedia*. Page Version ID: 928524653. Nov. 29, 2019. URL: https://en.wikipedia.org/w/index.php?title=Bag-of-words_model&oldid=928524653 (visited on 12/04/2019).
- [5] Barghout, L. “Visual Taxometric Approach to Image Segmentation Using Fuzzy-Spatial Taxon Cut Yields Contextually Relevant Regions”. In: *Information Processing and Management of Uncertainty in Knowledge-Based Systems*. Ed. by A. Laurent et al. Cham: Springer International Publishing, 2014, pp. 163–173. ISBN: 978-3-319-08855-6.
- [6] Bell, S. et al. “Material recognition in the wild with the Materials in Context Database”. In: *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). ISSN: 1063-6919. June 2015, pp. 3479–3487. DOI: 10.1109/CVPR.2015.7298970.
- [7] Biau, D. J., Jolles, B. M., and Porcher, R. “P value and the theory of hypothesis testing: an explanation for new researchers”. In: *Clinical Orthopaedics and Related Research* 468.3 (Mar. 2010), pp. 885–892. ISSN: 1528-1132. DOI: 10.1007/s11999-009-1164-4.
- [8] Bickel, P. J. and Doksum, K. A. “An Analysis of Transformations Revisited”. In: *Journal of the American Statistical Association* 76.374 (June 1981), pp. 296–311. ISSN: 0162-1459, 1537-274X. DOI: 10.1080/01621459.1981.10477649. URL: [http:](http://)

[//www.tandfonline.com/doi/abs/10.1080/01621459.1981.10477649](http://www.tandfonline.com/doi/abs/10.1080/01621459.1981.10477649) (visited on 12/02/2019).

- [9] Box, G. E. P. and Cox, D. R. "An Analysis of Transformations". In: *Journal of the Royal Statistical Society. Series B (Methodological)* 26.2 (1964), pp. 211–252. URL: <http://www.jstor.org/stable/2984418>.
- [10] Brockwell, P. J. and Davis, R. A. *Introduction to time series and forecasting*. 2nd ed. Springer texts in statistics. New York: Springer, 2002. 434 pp. ISBN: 978-0-387-95351-9.
- [11] Brown, R. G. *Exponential Smoothing for Predicting Demand*. Google-Books-ID: Eo_rMgEACAAJ. Little, 1956. URL: https://books.google.de/books?id=Eo_rMgEACAAJ.
- [12] Bruijning, M. et al. "trackdem: Automated particle tracking to obtain population counts and size distributions from videos in r". In: *Methods in Ecology and Evolution* 9.4 (2018), pp. 965–973. ISSN: 2041-210X. DOI: 10.1111/2041-210X.12975. URL: <https://besjournals.onlinelibrary.wiley.com/doi/abs/10.1111/2041-210X.12975> (visited on 12/02/2019).
- [13] Castillejo, E., Almeida, A., and López-de-Ipiña, D. "Social Network Analysis Applied to Recommendation Systems: Alleviating the Cold-User Problem". In: *Ubiquitous Computing and Ambient Intelligence*. Ed. by J. Bravo, D. López-de-Ipiña, and F. Moya. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer, 2012, pp. 306–313. ISBN: 978-3-642-35377-2. DOI: 10.1007/978-3-642-35377-2_42.
- [14] Chatfield, C. *Time-Series Forecasting*. Chapman and Hall/CRC, Oct. 25, 2000. ISBN: 978-0-429-12635-2. DOI: 10.1201/9781420036206. URL: <https://www.taylorfrancis.com/books/9780429126352> (visited on 11/19/2019).
- [15] Chatfield, C. *The Analysis of Time Series : An Introduction, Sixth Edition*. Chapman and Hall/CRC, July 29, 2003. ISBN: 978-0-429-20870-6. DOI: 10.4324/9780203491683. URL: <https://www.taylorfrancis.com/books/9780429208706> (visited on 11/19/2019).
- [16] Chen, P.-Y. S., Wu, S.-y., and Yoon, J. "The Impact of Online Recommendations and Consumer Feedback on Sales". In: *ICIS*. 2004, p. 58.
- [17] Cimpoi, M. et al. "Deep Filter Banks for Texture Recognition, Description, and Segmentation". In: *International Journal of Computer Vision* 118.1 (May 1, 2016), pp. 65–94. ISSN: 1573-1405. DOI: 10.1007/s11263-015-0872-3. URL: <https://doi.org/10.1007/s11263-015-0872-3> (visited on 12/26/2020).
- [18] Cleveland, R. B. et al. "STL: A seasonal-trend decomposition". In: *Journal of official statistics* 6.1 (1990), pp. 3–73.

- [19] Content-based image retrieval. In: *Wikipedia*. Page Version ID: 928223886. Nov. 27, 2019. URL: https://en.wikipedia.org/w/index.php?title=Content-based_image_retrieval&oldid=928223886 (visited on 12/04/2019).
- [20] Dai, J. et al. "R-FCN: object detection via region-based fully convolutional networks". In: *Proceedings of the 30th International Conference on Neural Information Processing Systems*. NIPS'16. Red Hook, NY, USA: Curran Associates Inc., Dec. 5, 2016, pp. 379–387. ISBN: 978-1-5108-3881-9. (Visited on 12/25/2020).
- [21] Dai, Y. et al. "Efficient Foreign Object Detection Between PSDs and Metro Doors via Deep Neural Networks". In: *IEEE Access* PP (Mar. 6, 2020), pp. 1–1. DOI: 10.1109/ACCESS.2020.2978912.
- [22] Davies, E. R. *Machine Vision: Theory, Algorithms, Practicalities*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2004. ISBN: 978-0-12-206093-9.
- [23] Deng, J. et al. "ImageNet: A large-scale hierarchical image database". In: *2009 IEEE Conference on Computer Vision and Pattern Recognition*. 2009, pp. 248–255. DOI: 10.1109/CVPR.2009.5206848.
- [24] Drupal CMS, D. *Drupal - Open Source CMS*. Drupal.org. Apr. 5, 2018. URL: <https://www.drupal.org/home> (visited on 12/09/2019).
- [25] Ellenberg, J. "This Psychologist Might Outsmart the Math Brains Competing for the Netflix Prize". In: *Wired* (Feb. 25, 2008). ISSN: 1059-1028. URL: <https://www.wired.com/2008/02/mf-netflix/> (visited on 11/19/2019).
- [26] Erhan, D. et al. "Scalable Object Detection Using Deep Neural Networks". In: *2014 IEEE Conference on Computer Vision and Pattern Recognition*. 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Columbus, OH, USA: IEEE, June 2014, pp. 2155–2162. ISBN: 978-1-4799-5118-5. DOI: 10.1109/CVPR.2014.276. URL: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6909673> (visited on 12/02/2019).
- [27] Facebook Data Science Core Team. *Trend ChangePoints*. Prophet. 2019. URL: http://facebook.github.io/prophet/docs/trend_changepoints.html (visited on 11/30/2019).
- [28] Felzenszwalb, P. F. et al. "Object detection with discriminatively trained part-based models". In: *IEEE transactions on pattern analysis and machine intelligence* 32.9 (Sept. 2010), pp. 1627–1645. ISSN: 1939-3539. DOI: 10.1109/TPAMI.2009.167.
- [29] Fu, C.-Y. et al. "DSSD : Deconvolutional Single Shot Detector". In: *arXiv e-prints* 1701 (Jan. 23, 2017), arXiv:1701.06659. arXiv: 1701.06659. URL: <http://adsabs.harvard.edu/abs/2017arXiv170106659F> (visited on 12/26/2020).

- [30] Fu, Y., Liu, Q., and Cui, Z. "A Collaborative Recommend Algorithm Based on Bipartite Community". In: *The Scientific World Journal* 2014 (2014). ISSN: 2356-6140. DOI: 10.1155/2014/295931. URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4009125/> (visited on 12/04/2019).
- [31] Fukushima, K. "Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position". In: *Biological Cybernetics* 36.4 (Apr. 1, 1980), pp. 193–202. ISSN: 1432-0770. DOI: 10.1007/BF00344251. URL: <https://doi.org/10.1007/BF00344251> (visited on 11/20/2019).
- [32] Gatys, L., Ecker, A., and Bethge, M. "A Neural Algorithm of Artistic Style". In: *Journal of Vision* 16.12 (Aug. 31, 2016). Publisher: The Association for Research in Vision and Ophthalmology, pp. 326–326. ISSN: 1534-7362. DOI: 10.1167/16.12.326. URL: <https://jov.arvojournals.org/article.aspx?articleid=2550311> (visited on 04/04/2021).
- [33] Girshick, R. "Fast R-CNN". In: *2015 IEEE International Conference on Computer Vision (ICCV)*. 2015 IEEE International Conference on Computer Vision (ICCV). ISSN: 2380-7504. Dec. 2015, pp. 1440–1448. DOI: 10.1109/ICCV.2015.169.
- [34] Girshick, R. et al. "Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation". In: *2014 IEEE Conference on Computer Vision and Pattern Recognition*. 2014 IEEE Conference on Computer Vision and Pattern Recognition. ISSN: 1063-6919. June 2014, pp. 580–587. DOI: 10.1109/CVPR.2014.81.
- [35] Glorot, X., Bordes, A., and Bengio, Y. "Deep Sparse Rectifier Neural Networks". In: *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*. Ed. by G. Gordon, D. Dunson, and M. Dudík. Vol. 15. Proceedings of Machine Learning Research. Fort Lauderdale, FL, USA: PMLR, Apr. 11, 2011, pp. 315–323. URL: <http://proceedings.mlr.press/v15/glorot11a.html>.
- [36] Good, P. I. *Permutation, Parametric, and Bootstrap Tests of Hypotheses*. 3rd ed. Springer Series in Statistics. New York: Springer-Verlag, 2005. ISBN: 978-0-387-20279-2. DOI: 10.1007/b138696. URL: <https://www.springer.com/gp/book/9780387202792> (visited on 12/02/2019).
- [37] Goodfellow, I. et al. "Generative Adversarial Networks". In: *Advances in Neural Information Processing Systems* 3 (June 10, 2014). DOI: 10.1145/3422622.
- [38] Google Marketing Tool, G. *Website, A/B Testing & Optimization Tools - Google Optimize*. Google Marketing Platform. URL: <https://marketingplatform.google.com/about/optimize/> (visited on 12/09/2019).

- [39] He, K. et al. "Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 37.9 (Sept. 2015). Conference Name: IEEE Transactions on Pattern Analysis and Machine Intelligence, pp. 1904–1916. ISSN: 1939-3539. DOI: 10.1109/TPAMI.2015.2389824.
- [40] He, K. et al. "Deep Residual Learning for Image Recognition". In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). ISSN: 1063-6919. June 2016, pp. 770–778. DOI: 10.1109/CVPR.2016.90.
- [41] He, K. et al. "Mask R-CNN". In: *2017 IEEE International Conference on Computer Vision (ICCV)*. 2017 IEEE International Conference on Computer Vision (ICCV). ISSN: 2380-7504. Oct. 2017, pp. 2980–2988. DOI: 10.1109/ICCV.2017.322.
- [42] He, K. et al. "Deep Residual Learning for Image Recognition". In: *arXiv:1512.03385 [cs]* (Dec. 10, 2015). arXiv: 1512.03385. URL: <http://arxiv.org/abs/1512.03385> (visited on 12/02/2019).
- [43] Herlocker, J. L. et al. "Evaluating collaborative filtering recommender systems". In: *ACM Transactions on Information Systems* 22.1 (Jan. 1, 2004), pp. 5–53. ISSN: 10468188. DOI: 10.1145/963770.963772. URL: <http://portal.acm.org/citation.cfm?doid=963770.963772> (visited on 11/19/2019).
- [44] Holt, C. C. "Forecasting seasonals and trends by exponentially weighted moving averages". In: *International Journal of Forecasting* 20.1 (1957), pp. 5–10. ISSN: 0169-2070. DOI: 10.1016/j.ijforecast.2003.09.015. URL: <http://www.sciencedirect.com/science/article/pii/S0169207003001134> (visited on 12/03/2019).
- [45] Hubel, D. H. and Wiesel, T. N. "Receptive fields, binocular interaction and functional architecture in the cat's visual cortex". In: *The Journal of Physiology* 160.1 (Jan. 1962), pp. 106–154.2. ISSN: 0022-3751. URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1359523/> (visited on 12/02/2019).
- [46] Hubel, D. H. and Wiesel, T. N. "Receptive fields and functional architecture in two nonstriate visual areas (18 and 19) of the cat". In: *Journal of Neurophysiology* 28.2 (1965), pp. 229–289. ISSN: 1522-1598(Electronic),0022-3077(Print).
- [47] Hyndman, R. J. and Athanasopoulos, G. *Forecasting: principles and practice*. 2. Edition. Heathmont, Victoria: OTexts, 2018. 382 pp. ISBN: 978-0-9875071-1-2.
- [48] Karpathy, A. *CS231n Convolutional Neural Networks for Visual Recognition*. CS231n: Convolutional Neural Networks for Visual Recognition. 2019. URL: <http://cs231n.github.io/> (visited on 11/30/2019).

- [49] Karpathy, A. and Li, F. F. "Deep Visual-Semantic Alignments for Generating Image Descriptions". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 39 (Feb. 15, 2014).
- [50] Karras, T., Laine, S., and Aila, T. "A Style-Based Generator Architecture for Generative Adversarial Networks". In: *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. ISSN: 2575-7075. June 2019, pp. 4396–4405. DOI: 10.1109/CVPR.2019.00453.
- [51] Kelleher, J. D. *Deep Learning*. Google-Books-ID: p_fTxQEACAAJ. MIT Press, 2019. 272 pp. ISBN: 978-0-262-35489-9.
- [52] Kohavi, R. et al. "Controlled experiments on the web: survey and practical guide". In: *Data Mining and Knowledge Discovery* 18.1 (Feb. 2009), pp. 140–181. ISSN: 1384-5810, 1573-756X. DOI: 10.1007/s10618-008-0114-1. URL: <http://link.springer.com/10.1007/s10618-008-0114-1> (visited on 11/19/2019).
- [53] Kohavi, R. et al. "Seven rules of thumb for web site experimenters". In: *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '14*. the 20th ACM SIGKDD international conference. New York, New York, USA: ACM Press, 2014, pp. 1857–1866. ISBN: 978-1-4503-2956-9. DOI: 10.1145/2623330.2623341. URL: <http://dl.acm.org/citation.cfm?doid=2623330.2623341> (visited on 12/02/2019).
- [54] Kovashka, A. *Assistant Professor, Department of Computer Science University of Pittsburgh*. URL: <http://people.cs.pitt.edu/~kovashka/> (visited on 04/05/2021).
- [55] Krizhevsky, A., Sutskever, I., and Hinton, G. E. "ImageNet Classification with Deep Convolutional Neural Networks". In: *Advances in Neural Information Processing Systems 25*. Ed. by F. Pereira et al. Curran Associates, Inc., 2012, pp. 1097–1105. URL: <http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf> (visited on 11/18/2019).
- [56] Kuznetsova, A. et al. "The Open Images Dataset V4: Unified Image Classification, Object Detection, and Visual Relationship Detection at Scale". In: *International Journal of Computer Vision* 128 (Mar. 13, 2020). DOI: 10.1007/s11263-020-01316-z.
- [57] LeCun, Y. et al. "Backpropagation Applied to Handwritten Zip Code Recognition". In: *Neural Comput.* 1.4 (Dec. 1989), pp. 541–551. ISSN: 0899-7667. DOI: 10.1162/neco.1989.1.4.541. URL: <http://dx.doi.org/10.1162/neco.1989.1.4.541> (visited on 12/02/2019).

- [58] Li, J. et al. "Application Research of Improved YOLO V3 Algorithm in PCB Electronic Component Detection". In: *Applied Sciences* 9 (Sept. 8, 2019), p. 3750. DOI: 10.3390/app9183750.
- [59] Lin, T. et al. "Focal Loss for Dense Object Detection". In: *2017 IEEE International Conference on Computer Vision (ICCV)*. 2017 IEEE International Conference on Computer Vision (ICCV). ISSN: 2380-7504. Oct. 2017, pp. 2999–3007. DOI: 10.1109/ICCV.2017.324.
- [60] Lin, T.-Y. et al. "Feature Pyramid Networks for Object Detection". In: *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, HI: IEEE, July 2017, pp. 936–944. ISBN: 978-1-5386-0457-1. DOI: 10.1109/CVPR.2017.106. URL: <http://ieeexplore.ieee.org/document/8099589/> (visited on 12/26/2020).
- [61] Liu, W. et al. "SSD: Single Shot MultiBox Detector". In: *Computer Vision – ECCV 2016*. Ed. by B. Leibe et al. Lecture Notes in Computer Science. Cham: Springer International Publishing, 2016, pp. 21–37. ISBN: 978-3-319-46448-0. DOI: 10.1007/978-3-319-46448-0_2.
- [62] Liu, Z. et al. "Joint Video Object Discovery and Segmentation by Coupled Dynamic Markov Networks". In: *IEEE Transactions on Image Processing* 27.12 (Dec. 2018), pp. 5840–5853. ISSN: 1057-7149, 1941-0042. DOI: 10.1109/TIP.2018.2859622. URL: <https://ieeexplore.ieee.org/document/8423204/> (visited on 12/02/2019).
- [63] Livera, A. M. D., Hyndman, R. J., and Snyder, R. D. "Forecasting Time Series With Complex Seasonal Patterns Using Exponential Smoothing". In: *Journal of the American Statistical Association* 106.496 (2011), pp. 1513–1527. ISSN: 0162-1459. DOI: 10.1198/jasa.2011.tm09771. URL: <https://doi.org/10.1198/jasa.2011.tm09771> (visited on 11/30/2019).
- [64] *Logistic function*. In: *Wikipedia*. Page Version ID: 928587055. Nov. 30, 2019. URL: https://en.wikipedia.org/w/index.php?title=Logistic_function&oldid=928587055 (visited on 11/30/2019).
- [65] Makridakis, S. G., Wheelwright, S. C., and Hyndman, R. J. *Forecasting*. 3rd edition. New York: John Wiley & Sons, Jan. 16, 1998. 656 pp. ISBN: 978-0-471-53233-0.
- [66] Mirza, B., Keller, B., and Ramakrishnan, N. "Studying Recommendation Algorithms by Graph Analysis". In: *J. Intell. Inf. Syst.* 20 (Mar. 1, 2003), pp. 131–160. DOI: 10.1023/A:1021819901281.

- [67] Najibi, M., Rastegari, M., and Davis, L. S. "G-CNN: An Iterative Grid Based Object Detector". In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). ISSN: 1063-6919. June 2016, pp. 2369–2377. DOI: 10.1109/CVPR.2016.260.
- [68] *Nearest neighbour algorithm*. In: *Wikipedia*. Page Version ID: 893121690. Apr. 19, 2019. URL: https://en.wikipedia.org/w/index.php?title=Nearest_neighbour_algorithm&oldid=893121690 (visited on 12/04/2019).
- [69] *Open Images Dataset V5*. URL: <https://storage.googleapis.com/openimages/web/index.html> (visited on 12/01/2019).
- [70] *OpenSurfaces - A Richly Annotated Catalog of Surface Appearance*. URL: <http://opensurfaces.cs.cornell.edu/> (visited on 12/04/2019).
- [71] Park, E. et al. "Large Scale Visual Recognition Challenge (ILSVRC) 2017". In: *image-net.org* (2017), p. 26. URL: http://image-net.org/challenges/talks_2017/ILSVRC2017_overview.pdf.
- [72] *Piecewise linear function*. In: *Wikipedia*. Page Version ID: 925737862. Nov. 12, 2019. URL: https://en.wikipedia.org/w/index.php?title=Piecewise_linear_function&oldid=925737862 (visited on 12/03/2019).
- [73] RafieTarabay. *Deep Learning*. IBM Developer Recipes. Oct. 31, 2019. URL: <https://developer.ibm.com/recipes/tutorials/deep-learning/> (visited on 04/05/2021).
- [74] Redmon, J. and Farhadi, A. "YOLO9000: Better, Faster, Stronger". In: *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). ISSN: 1063-6919. July 2017, pp. 6517–6525. DOI: 10.1109/CVPR.2017.690.
- [75] Redmon, J. et al. "You Only Look Once: Unified, Real-Time Object Detection". In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). ISSN: 1063-6919. June 2016, pp. 779–788. DOI: 10.1109/CVPR.2016.91.
- [76] Redmon, J. and Farhadi, A. "YOLOv3: An Incremental Improvement". In: *Computer Vision and Pattern Recognition 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (2018)*, p. 5. DOI: 10.1109/CVPR40276.2018. URL: <http://cvpr2018.thecvf.com/>.
- [77] Ren, S. et al. "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 39.6 (June 2017). Conference Name: IEEE Transactions on Pattern

Analysis and Machine Intelligence, pp. 1137–1149. ISSN: 1939-3539. DOI: 10.1109/TPAMI.2016.2577031.

- [78] *Residual sum of squares*. In: *Wikipedia*. Page Version ID: 921781584. Oct. 17, 2019. URL: https://en.wikipedia.org/w/index.php?title=Residual_sum_of_squares&oldid=921781584 (visited on 12/03/2019).
- [79] Russakovsky, O. et al. “ImageNet Large Scale Visual Recognition Challenge”. In: *International Journal of Computer Vision* 115.3 (Dec. 1, 2015), pp. 211–252. ISSN: 1573-1405. DOI: 10.1007/s11263-015-0816-y. URL: <https://doi.org/10.1007/s11263-015-0816-y> (visited on 12/26/2020).
- [80] Sastry, N. R. “How To Tell Head From Tail in User-Generated Content Corpora”. In: *ICWSM*. 2012.
- [81] Shankar, S. et al. “Refining Architectures of Deep Convolutional Neural Networks”. In: *Conference on Computer Vision and Pattern Recognition (CVPR), 2016*. CVRP 2016. 2016.
- [82] Shen, Z. et al. “Dsod: Learning deeply supervised object detectors from scratch”. In: *Proceedings of the IEEE international conference on computer vision*. 2017, pp. 1919–1927.
- [83] Simonyan, K. and Zisserman, A. “Very Deep Convolutional Networks for Large-Scale Image Recognition”. In: *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*. Ed. by Y. Bengio and Y. LeCun. Apr. 10, 2015. URL: <http://arxiv.org/abs/1409.1556>.
- [84] Soltani, A. A. et al. “Synthesizing 3D Shapes via Modeling Multi-view Depth Maps and Silhouettes with Deep Generative Networks”. In: *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, HI: IEEE, July 2017, pp. 2511–2519. ISBN: 978-1-5386-0457-1. DOI: 10.1109/CVPR.2017.269. URL: <http://ieeexplore.ieee.org/document/8099752/> (visited on 12/01/2019).
- [85] Stan, J., Muhlenbach, F., and Largeron, C. *Recommender Systems Using Social Network Analysis: Challenges and Future Trends*. Jan. 1, 2014. DOI: 10.1007/978-1-4614-6170-8_35.
- [86] Stone, J. V. *Bayes’ Rule: A Tutorial Introduction to Bayesian Analysis*. Sebtel Press, June 1, 2013. 182 pp. ISBN: 978-0-9563728-4-0.

- [87] Tautkute, I. et al. "What looks good with my sofa: Multimodal search engine for interior design". In: *2017 Federated Conference on Computer Science and Information Systems (FedCSIS)*. 2017 Federated Conference on Computer Science and Information Systems (FedCSIS). Sept. 2017, pp. 1275–1282. DOI: 10.15439/2017F56.
- [88] Taylor, S. J. and Letham, B. *Forecasting at scale*. preprint. PeerJ Preprints, Sept. 27, 2017. DOI: 10.7287/peerj.preprints.3190v2. URL: <https://peerj.com/preprints/3190> (visited on 11/30/2019).
- [89] Taylor, S. and Letham, B. *prophet: Automatic Forecasting Procedure*. May 14, 2019. URL: <https://CRAN.R-project.org/package=prophet> (visited on 11/21/2019).
- [90] *Transfer learning*. In: *Wikipedia*. Page Version ID: 929223648. Dec. 4, 2019. URL: https://en.wikipedia.org/w/index.php?title=Transfer_learning&oldid=929223648 (visited on 12/04/2019).
- [91] Uijlings, J. R. R. et al. "Selective Search for Object Recognition". In: *International Journal of Computer Vision* 104.2 (Sept. 1, 2013), pp. 154–171. ISSN: 1573-1405. DOI: 10.1007/s11263-013-0620-5. URL: <https://doi.org/10.1007/s11263-013-0620-5> (visited on 11/19/2019).
- [92] Ulyanov, D., Vedaldi, A., and Lempitsky, V. "Improved Texture Networks: Maximizing Quality and Diversity in Feed-Forward Stylization and Texture Synthesis". In: *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). ISSN: 1063-6919. July 2017, pp. 4105–4113. DOI: 10.1109/CVPR.2017.437.
- [93] Vinyals, O. et al. "Show and tell: A neural image caption generator". In: *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Boston, MA, USA: IEEE, June 2015, pp. 3156–3164. ISBN: 978-1-4673-6964-0. DOI: 10.1109/CVPR.2015.7298935. URL: <http://ieeexplore.ieee.org/document/7298935/> (visited on 12/02/2019).
- [94] *Visual search*. In: *Wikipedia*. Page Version ID: 928982495. Dec. 2, 2019. URL: https://en.wikipedia.org/w/index.php?title=Visual_search&oldid=928982495 (visited on 12/04/2019).
- [95] Wäldchen, J. and Mäder, P. "Plant Species Identification Using Computer Vision Techniques: A Systematic Literature Review". In: *Archives of Computational Methods in Engineering* 25.2 (2018), pp. 507–543. ISSN: 1134-3060. DOI: 10.1007/s11831-016-9206-z. URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6003396/> (visited on 12/02/2019).

- [96] Wan, J. et al. "Deep Learning for Content-Based Image Retrieval: A Comprehensive Study". In: *Proceedings of the ACM International Conference on Multimedia - MM '14*. the ACM International Conference. Orlando, Florida, USA: ACM Press, 2014, pp. 157–166. ISBN: 978-1-4503-3063-3. DOI: 10.1145/2647868.2654948. URL: <http://dl.acm.org/citation.cfm?doid=2647868.2654948> (visited on 11/19/2019).
- [97] Wang, Y., Gao, W., and Wang, Y. "Application of furniture images selection based on neural network". In: 6TH INTERNATIONAL CONFERENCE ON COMPUTER-AIDED DESIGN, MANUFACTURING, MODELING AND SIMULATION (CDMMS 2018). Busan, South Korea, 2018, p. 040016. DOI: 10.1063/1.5039090. URL: <http://aip.scitation.org/doi/abs/10.1063/1.5039090> (visited on 11/19/2019).
- [98] Winters, P. R. "Forecasting Sales by Exponentially Weighted Moving Averages". In: *Management Science* 6.3 (Apr. 1, 1960), pp. 324–342. ISSN: 0025-1909. DOI: 10.1287/mnsc.6.3.324. URL: <https://pubsonline.informs.org/doi/abs/10.1287/mnsc.6.3.324> (visited on 12/03/2019).
- [99] Xu, T. et al. "AttnGAN: Fine-Grained Text to Image Generation with Attentional Generative Adversarial Networks". In: *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. ISSN: 2575-7075. June 2018, pp. 1316–1324. DOI: 10.1109/CVPR.2018.00143.
- [100] Yann LeCun, Courant Institute, NYU, Corinna Cortes, Google Labs, New York, and Christopher J.C. Burges, Microsoft Research, Redmond. *MNIST handwritten digit database, Yann LeCun, Corinna Cortes and Chris Burges*. THE MNIST DATABASE. URL: <http://yann.lecun.com/exdb/mnist/> (visited on 04/04/2021).
- [101] Yoo, D. et al. "AttentionNet: Aggregating Weak Directions for Accurate Object Detection". In: *2015 IEEE International Conference on Computer Vision (ICCV)*. 2015 IEEE International Conference on Computer Vision (ICCV). ISSN: 2380-7504. Dec. 2015, pp. 2659–2667. DOI: 10.1109/ICCV.2015.305.
- [102] Young, W. L. "The Box-Jenkins approach to time series analysis and forecasting : principles and applications". In: *RAIRO - Operations Research* 11.2 (1977), pp. 129–143. ISSN: 0399-0559, 1290-3868. DOI: 10.1051/ro/1977110201291. URL: <http://www.rairo-ro.org/10.1051/ro/1977110201291> (visited on 11/30/2019).
- [103] Zanin, M. et al. "Complex Networks in Recommendation Systems". In: *Proceedings of the 2nd WSEAS International Conference on Computer Engineering and Applications*. CEA'08. event-place: Acapulco, Mexico. Stevens Point, Wisconsin,

USA: World Scientific, Engineering Academy, and Society (WSEAS), 2008, pp. 120–124. ISBN: 978-960-6766-33-6.

- [104] Zhang, R., Isola, P., and Efros, A. A. “Colorful Image Colorization”. In: *Computer Vision – ECCV 2016*. Ed. by B. Leibe et al. Lecture Notes in Computer Science. Cham: Springer International Publishing, 2016, pp. 649–666. ISBN: 978-3-319-46487-9. DOI: 10.1007/978-3-319-46487-9_40.
- [105] Zhao, Z.-Q. et al. “Object Detection with Deep Learning: A Review”. In: *arXiv:1807.05511 [cs]* (Apr. 16, 2019). arXiv: 1807.05511. URL: <http://arxiv.org/abs/1807.05511> (visited on 11/19/2019).