

1 Εισαγωγή

Στην παρούσα αναφορά θα γίνει παρουσίαση και ανάλυση των χαρακτηριστικών που χρησιμοποιήθηκαν στο μερικώς επιβλεπόμενο σύστημα, αλλά και στο μη επιβλεπόμενο σύστημα. Κάθε παράδειγμα μπορεί να περιγραφεί από ένα συνδυασμό τιμών που αναφέρονται επίσης ως μεταβλητές, χαρακτηριστικά, πεδία ή διαστάσεις. Οι τιμές αυτές μπορούν να είναι διαφορετικού τύπου όπως συνεχείς, δυαδικές ή κατηγορίες. Κάθε παράδειγμα μπορεί να αποτελείται μόνο από μια τιμή (μονοπαραγοντικό) ή και από περισσότερες (πολυπαραγοντική). Στην περίπτωση των πολυπαραγοντικών παραδειγμάτων, όλες οι τιμές μπορεί να είναι ίδιου τύπου ή μπορεί να είναι ένας συνδυασμός διαφορετικών τύπων [1]. Παράλληλα, κάθε παράδειγμα μπορεί να οριστεί βάση ακόμη δύο δομών ως προς τον ορισμό του προβλήματος [1].

1. *Τιμές Συσχετισμού* Τέτοιου είδους τιμές χρησιμοποιούνται για να περιγράψουν ένα γενικό πλαίσιο που χαρακτηρίζει ένα παράδειγμα. Στις χρονοσειρές, ο χρόνος είναι μια τιμή που παρέχει μια σχετικότητα, η οποία καθορίζει τη θέση ενός παραδείγματος σε μια ολόκληρη ακολουθία. Μία τιμή γενικού πλαισίου είναι η μηνιαία κατανάλωση ενός κατοίκου.
2. *Συμπεριφορικές Τιμές* Είναι οι τιμές που δεν προδίδουν ένα γενικό πλαίσιο για κάποιο παράδειγμα ή κάποια σχετικότητα. Ένα τέτοιο παράδειγμα θα μπορούσε να είναι η ετήσια παραγωγή ενέργειας σε όλο τον κόσμο.

2 Φύση Χαρακτηριστικών

Το μερικώς επιβλεπόμενο και μη επιβλεπόμενο σύστημα απαιτούν εισόδους που να δίνουν τη δυνατότητα να διαχωρίζονται σε δύο κλάσεις οι καταναλωτές. Για να γίνει αυτό απαιτείται η χρήση χαρακτηριστικών που να αντιπροσωπεύουν την κλάση, αλλά και χαρακτηριστικά που προσδίνουν γενικότητα στο κάθε παράδειγμα. Με αυτό τον τρόπο παρέχεται ένα περιθώριο στον αλγόριθμο, έτσι ώστε να μπορεί εύκολα να προσαρμόζεται σε καινούργια και ξεχωριστά παραδείγματα. Ένας απλοϊκός τρόπος να διαχωρίσουμε τα χαρακτηριστικά είναι σε χαρακτηριστικά γενίκευσης και σε χαρακτηριστικά διαχωρισμού κλάσεων. Όλα τα παρακάτω χαρακτηριστικά αποτελούν τιμές συσχέτισης.

Χαρακτηριστικά Γενίκευσης

Τα πλεονέκτημα των χαρακτηριστικών γενίκευσης είναι ότι βοηθούν στην κατάταξη του καταναλωτή σε σχέση με τους υπόλοιπους, ώστε να εξαχθούν πληροφορίες, όπως ο τύπος καταναλωτή (οικιακού ή βιομηχανικού) και το προφίλ κατανάλωσής του. Τέτοια χαρακτηριστικά πρέπει να περιορίζονται σε αριθμό όμως, καθώς ενδέχεται να δυσκολέυσουν τον διαχωρισμό με βάση το κριτήριο που θέτουμε παρέχοντας μεγάλο παράγοντα γενίκευσης. Τέτοιου είδους χαρακτηριστικά είναι τα παρακάτω:

1. *Ετήσια μέση τιμή ημίωρου* Βρίσκεται ο μέσος όρος ημίωρου κάθε μέρας και για όλες τις μέρες του έτους βρίσκεται ο ετήσιος μέσος όρος.
2. *Ετήσια τυπική απόκλιση ημίωρου* Βρίσκεται η τυπική απόκλιση κάθε μέρας και για όλες τις μέρες του έτους βρίσκεται ο ετήσιος μέσος όρος της τυπικής απόκλισης.
3. *Διαφορά Ετήσιου Ελάχιστου τάσης με όμοιους* Βάση αυτού του χαρακτηριστικού ορίζεται για όμοιους καταναλωτές το ελάχιστο της τάσης κατανάλωσής τους και στην συνέχεια βρίσκεται η απόλυτη διαφορά σε ημέρες.
4. *Διαφορά μέσης τιμής με όμοιους* Με αυτό το χαρακτηριστικό βρίσκεται η διαφορά του ετήσιου μέσου όρου κάθε καταναλωτή με την ομάδα καταναλωτών που ανήκει.
5. *Διαφορά τυπικής απόκλισης με όμοιους* Με αυτό το χαρακτηριστικό βρίσκεται η διαφορά της ετήσιας τυπικής απόκλισης κάθε καταναλωτή με την ομάδα καταναλωτών που ανήκει.

Χαρακτηριστικά Διαχωρισμού

Τα χαρακτηριστικά διαχωρισμού επικεντρώνονται στην όξυνση των διαφορών μεταξύ των καταναλωτών διαφορετικών κλάσεων. Λειτουργούν, λοιπόν σαν οδηγοί για τον αλγόριθμο ώστε να κάνουν πιο εμφανείς τις διαφορές των κλάσεων. Το πλεονέκτημα τους είναι ο παράγοντας εξειδίκευσης που παρέχουν στον αλγόριθμο διευκολύνοντας τον να αναγνωρίζει με διαφορετικούς τρόπους κάθε κλάση. Το μειονέκτημα είναι πως λόγω της εξειδικευμένης τους φύσης μπορεί να μην εφαρμόζονται απόλυτα από όλους τους καταναλωτές ή στην χειρότερη περίπτωση να περιγράφουν μια σπάνια συμπεριφορά που δεν ενδιαφερόμαστε να διαχωρίσουμε.

1. *Κινούμενος μέσος όρος μηνιαίου μέσου όρου* Πρόκειται για υπό συνθήκη χαρακτηριστικό που αν παρατηρήσει κάποια σημαντική πτώση των καταναλώσεων τότε ψάχνει για την μέγιστη και την καταγράφει.
2. *Κινούμενος μέσος όρος μηνιαίας τυπικής απόκλισης* Πρόκειται για υπό συνθήκη χαρακτηριστικό που αν παρατηρήσει κάποια σημαντική πτώση της τυπικής απόκλισης τότε ψάχνει για την μέγιστη και την καταγράφει.
3. *Συμμετρική διαφορά κατανάλωσης* Πρόκειται για υπό συνθήκη χαρακτηριστικό που παρατηρεί μια γενική συμπεριφορά όμοιων καταναλωτών ως προς τη χρονική στιγμή της ελάχιστης κατανάλωσης και ψάχνει για κάποια σημαντική πτώση της κατανάλωσης ανάμεσα σε 2 συμμετρικές χρονικές στιγμές με άξονα συμμετρίας την εκάστοτε χρονική στιγμή ελαχίστου.
4. *Συμμετρική διαφορά τυπικής απόκλισης* Πρόκειται για υπό συνθήκη χαρακτηριστικό που παρατηρεί μια γενική συμπεριφορά όμοιων καταναλωτών ως προς τη χρονική στιγμή της ελάχιστης κατανάλωσης και ψάχνει για κάποια σημαντική πτώση της τυπικής απόκλισης ανάμεσα σε 2 συμμετρικές χρονικές στιγμές με άξονα συμμετρίας την εκάστοτε χρονική στιγμή ελαχίστου.
5. *Τμηματική διαφορά κατανάλωσης με όμοιους καταναλωτές* Πρόκειται για υπό συνθήκη χαρακτηριστικό που παρατηρεί μια γενική συμπεριφορά όμοιων καταναλωτών ως προς τη χρονική στιγμή της ελάχιστης κατανάλωσης και ψάχνει για κάποια σημαντική πτώση της κατανάλωσης ανάμεσα στον καταναλωτή και τους όμοιούς του μετά την χρονική στιγμή της ελάχιστης κατανάλωσης.

6. *Τμηματική διαφορά τυπικής απόκλισης με όμοιους καταναλωτές* Πρόκειται για υπό συνθήκη χαρακτηριστικό που παρατηρεί μια γενική συμπεριφορά όμοιων καταναλωτών ως προς τη χρονική στιγμή της ελάχιστης κατανάλωσης και ψάχνει για κάποια σημαντική πτώση της τυπικής απόκλισης ανάμεσα στον καταναλωτή και τους όμοιούς του μετά την χρονική στιγμή της ελάχιστης κατανάλωσης.

3 Δοκιμή Χαρακτηριστικών

Αφού ορίσουν τα χαρακτηριστικά που εκτιμάται ότι μπορούν να βοηθήσουν στον διαχωρισμό των κλάσεων, έπεται φυσικά η δοκιμή τους με έναν απελή τρόπο, έτσι ώστε να επιβεβαιωθεί ότι μπορούν να λειτουργήσουν όπως αναμένεται. Παράλληλα, η δοκιμή αυτή παρέχει μεγάλο όγκο πληροφορίας, αφού καθιστά εμφανή τα σημεία και τις προϋποθέσεις που τα χαρακτηριστικά έχουν μεγάλη ακρίβεια, αλλά και εκεί που υστερούν.

Ο κώδικας της δοκιμής θεωρεί δεδομένη και σταθερή την ένταση κλοπής και την ημέρα που κάθε καταναλωτής ξεκινά να αλλοιώνει τις τιμές του. Ειδικότερα, το ποσοστό των καταναλωτών που αλλοιώνει τις μετρήσεις του είναι 50 τοις εκατό, η ένταση της κλοπής είναι της τάξης του 80 τοις εκατό και η μέρα κλοπής ορίζεται η 182η, δηλαδή μετά από 6 μήνες κανονικής κατανάλωσης ο χρήστης εισάγει σύστημα αλλοίωσης της μέτρησής του. Δοκιμάζοντας ξεχωριστά τα χαρακτηριστικά διαχωρισμού ελέγχουμε το όριο κάθε χαρακτηριστικού έτσι ώστε να δώσει μεγαλύτερη ακρίβεια στις επιθέσεις δεδομένων υπό τις παραπάνω συνθήκες. Αν ο παρατηρηθούν τέτοια χαρακτηριστικά ο καταναλωτής θεωρείται θετικός στην κλοπή. Αντίθετα αν ο καταναλωτής δεν έχει την αναμενόμενη συμπεριφορά το χαρακτηριστικό δεν καταγράφει κάποια τιμή και ο καταναλωτής θεωρείται αρνητικός στην κλοπή. Αναλυτικότερα για κάθε χαρακτηριστικό διαχωρισμού λήφθηκαν τα παρακάτω αποτελέσματα:

1. Κινούμενος μέσος όρος μηνιαίου μέσου όρου

Στην πρώτη δοκιμή δόθηκε έμφαση στη γενικότερη συμπεριφορά του χαρακτηριστικού ως προς το όριο που τίθεται κάθε φορά. Έτσι παρατηρείται εύκολα πως για μεγάλο όριο ο διαχωρισμός έχει χαμηλή ακρίβεια με εξαιρετικά μικρό ποσοστό αποτυχίας. Αντίθετα, αν το όριο χαμηλώσει αισθητά ο χάνεται η έννοια του διαχωρισμού και ο αλγόριθμος θεωρεί θετικούς σε κλοπές σχεδόν όλους τους καταναλωτές.

Δοκιμή 1ου χαρακτηριστικού					
Όριο	DR	FPR	BDR	Accuracy	F1
0,8	44,8	1,4	0,97	71,7	61,29
0,7	98,7	1,9	0,98	98,4	98,4
0,6	99,3	3,6	0,97	97,85	97,88
0,5	99,8	7,5	0,93	96,15	96,15
0	99,9	91,5	0,52	54,2	68,57

2. Κινούμενος μέσος όρος μηνιαίας τυπικής απόκλισης

Αντίστοιχα και εδώ για παρόμοιες τιμές του ορίου με το προηγούμενο χαρακτηριστικό ο διαχωρισμός είναι εξαιρετικά εύστοχος και δεν αφήνει περιθώρια για αμφισβήτηση.

Δοκιμή 2ου χαρακτηριστικού					
Όριο	DR	FPR	BDR	Accuracy	F1
0,7	98,2	2,3	0,98	98,3	98,31
0,6	99,8	4,1	0,96	97,85	97,89
0,5	99,5	8,2	0,92	95,65	95,81

3. Συμμετρική διαφορά καταναλώσεων

Το συγκεκριμένο χαρακτηριστικό δεν δίνει αξιόπιστα αποτελέσματα, καθώς η συμμετρία που προκύπτει από τον χρησιμοποιούμενο τύπο απάτης κάνει το συγκεκριμένο χαρακτηριστικό να αποτυγχάνει σε αυτή τη δοκιμή. Παρόλα αυτά, το χαμηλό ποσοστό αποτυχίας αφήνει δεύτερες σκέψεις, καθώς δεν επιβαρύνει αισθητά τα αποτελέσματα, αλλά βοηθά στη γενίκευση του τύπου κλοπής.

Δοκιμή 3ου χαρακτηριστικού					
Όριο	DR	FPR	BDR	Accuracy	F1
0,1	26,3	5,7	0,82	60,3	39,85

4. Συμμετρική διαφορά τυπικής απόκλισης

Αντίστοιχα συμπεράσματα ισχύουν και στη συμμετρική διαφορά τυπικής απόκλισης που οριακά ξεπερνά το 10 τοις εκατό στο FPR. Η γενίκευση που προσφέρει παρόλα αυτά το συγκεκριμένο χαρακτηριστικό είναι χρήσιμη, καθώς εν τέλει όλα τα χαρακτηριστικά θα ενώσουν τα δυνατά τους σημεία για να διαχωρίσουν απάτες με μεγαλύτερο τυχαίο παράγοντα.

Δοκιμή 4ου χαρακτηριστικού					
Όριο	DR	FPR	BDR	Accuracy	F1
0,1	38,9	10,2	0,79	64,35	52,18

5. Τμηματική διαφορά κατανάλωσης με όμοιους καταναλωτές

Σε αυτό το χαρακτηριστικό παρατηρείται σχετική αστοχία σε σχέση με τα πρώτα χαρακτηριστικά υποδεικνύοντας ανάγκη για καλύτερο ρύθμισμα του χαρακτηριστικού.

Δοκιμή 5ου χαρακτηριστικού					
Όριο	DR	FPR	BDR	Accuracy	F1
0,1	98,4	11,4	0,9	93,5	93,8
0,2	93,9	7	0,93	93,45	93,48
0,3	88,3	4,9	0,95	91,7	91,41

Δεδομένης της διαφοράς των καταναλώσεων με την γενικευμένη κατανάλωση μιας ομάδας δημιουργείται η ανάγκη για κανονικοποίηση σε κάθε διάνυσμα κατανάλωσης. Με αυτό τον τρόπο επιτυγχάνονται πολύ καλύτερα αποτελέσματα.

Δοκιμή 5ου χαρακτηριστικού με κανονικοποίηση					
Όριο	DR	FPR	BDR	Accuracy	F1
0,3	98,9	5,4	0,95	96,75	96,82

6. Τμηματική διαφορά τυπικής απόκλισης με όμοιους καταναλωτές

Αντίστοιχη μεθοδολογία εφαρμόστηκε και σε αυτό το χαρακτηριστικό. Τα αποτελέσματα ήταν ικανοποιητικά, αλλά όχι αρκετά. Έτσι χρησιμοποιήθηκε κανονικοποίηση για μπορέσει να μειωθεί το FPR, ενώ αυξάνεται το DR.

Δοκιμή 6ου χαρακτηριστικού					
Όριο	DR	FPR	BDR	Accuracy	F1
0,1	99,4	16,1	0,86	91,65	92,25
0,2	95,7	8,3	0,92	93,7	93,82
0,3	89,8	6,2	0,94	91,8	91,63

Δοκιμή 6ου χαρακτηριστικού με κανονικοποίηση					
Όριο	DR	FPR	BDR	Accuracy	F1
0,4	97	4,9	0,95	96,05	96,09
0,3	96,3	5,3	0,95	95,5	95,54

Τέλος έγινε μια ακόμη τελική δοκιμή στα χαρακτηριστικά, αυτή τη φορά με μεγαλύτερο τυχαίο παράγοντα. Η ένταση της κλοπής καθορίζεται από μια βήτα κατανομή με παραμέτρους 6 και 3, ενώ η ημέρα που ξεκινά η κλοπή επιλέγεται από μια κανονική κατανομή με παραμέτρους 182.5 και 56.1538. Σε κάθε καταναλωτή που έχει επιλεγεί για κλοπή επιβάλλονται διαφορετικές τιμές των παραπάνω κατανομών κρατώντας όμως το γενικότερο πλαίσιο του τύπου της κλοπής που είδαμε προηγουμένως. Αυτό που μας ενδιαφέρει να δούμε σε αυτό το σημείο είναι χαμηλό FPR καθώς αναμένεται να χαμηλώσει σημαντικά η ακρίβεια, λόγω της απλότητας του κριτηρίου μας.

Δοκιμή χαρακτηριστικών με τυχαίο παράγοντα						
Χαρακτ	Όριο	DR	FPR	BDR	Accuracy	F1
1	0,7	42,8	2,1	0,95	70,35	59,08
2	0,7	46,5	1,8	0,96	72,35	62,71
3	0,1	58	9,9	0,85	74,05	69,09
4	0,1	59,6	9,4	0,86	75,1	70,53
5	0,3	66,4	8,2	0,89	79,1	76,06
6	0,4	58,4	5,6	0,91	76,4	71,22

Καθίσταται λοιπόν σαφές πως τα χαρακτηριστικά σε γενικές γραμμές έχουν χαμηλότερη ακρίβεια, αλλά κρατούν χαμηλό FPR κάτι που μας ενδιαφέρει περισσότερο σε αυτό το σημείο. Παράλληλα, τα χαρακτηριστικά 3 και 4 που είχαν απογοητευτικά αποτελέσματα στην προηγούμενη δοκιμή, σε αυτήν δείχνουν να βελτιώνονται αισθητά σε σχέση με την επίδοση των υπόλοιπων. Αυτό μας πληροφορεί ότι η αρχική υπόθεση μας για το λόγο αστοχία τους ήταν αληθής.

Συνημμένα

Αναφορές

- [1] V. Chandola, A. Banerjee, V. Kumar, 2009. *Anomaly Detection: A Survey*, University of Minnesota. September, p. 6, 8.