

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

Συνημμένα

Αναφορές

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