2. ΑΝΙΧΝΕΥΣΗ ΣΥΜΒΑΝΤΩΝ ΣΕ ΡΟΕΣ ΔΕΔΟΜΕΝΩΝ ΑΙΣΘΗΤΗΡΩΝ ΠΟΛΛΑΠΛΩΝ ΜΕΤΑΒΛΗΤΩΝ

Σε αυτή την ενότητα μελετάται η διαδικασία ανίχνευσης συμβάντων και παραγωγής ροών συμβάντων πάνω σε ένα υπάρχον σύνολο ροών αισθητήρων. Το πρόβλημα ανίχνευσης συμβάντων πάνω σε πολλαπλές ροές αισθητήρων μπορεί να διαμορφωθεί όπως περιγράφεται στη συνέχεια. Αρχικά γίνεται παρατήρηση σε πραγματικό χρόνο με κάποια ενιαία συχνότητα χρονοσειρών πολλαπλών μεταβλητών των ποσοτικών παραμέτρων απόδοσης του συστήματος. Μία ροή αισθητήρων, η οποία αποτελείται από αριθμητικές τιμές αισθητήρων συμβολίζεται με s_i και με $s_i(t)$ συμβολίζεται η τιμή της ροής s_i σε χρόνο t, όπου ισχύει $t \epsilon [0, +\infty)$. Υποθέτοντας ότι n ροές αισθητήρων συγχρονίζονται για την αναφορά των τιμών τους περιοδικά, γίνεται αντιπροσώπευση του συνόλου πληροφορίας πλαισίου πολλαπλών μεταβλητών σε κάθε χρονική στιγμή t με ένα διάνυσμα πλαισίου $\Delta\Pi_t = (s_1(t), s_2(t), \dots, s_n(t)) \epsilon \Re^n$. Πρακτικά, κάθε ροή αισθητήρων διαμορφώνει μία μονοδιάστατη χρονοσειρά, ενώ η ροή διανυσμάτων πλαισίου αντιπροσωπεύει μία χρονοσειρά πολλαπλών μεταβλητών.

Υπάρχουν πολλά προβλήματα σε τομείς επιστημών, τα οποία απαιτούν την ακολουθιακή ανίχνευση μίας αλλαγής ή ενός συμβάντος σε μία διαδικασία. Στην πιο απλή μορφή, γίνεται προσπάθεια για ανίχνευση μίας αλλαγής στο μέσο όρο μίας ακολουθίας, όπου η αλλαγή είναι είτε απότομη ή σταδιακή. Μία ροή δεδομένων αποτελείται από μία δυνητικά άπειρη ακολουθία πλειάδων δεδομένων. Τα δεδομένα ροής έχουν δύο χαρακτηριστικά, τα οποία αποτελούν πρόκληση στην επεξεργασία τους, ο υψηλός ρυθμός άφιξης και το ενδεχόμενο μη προβλέψιμης συμπεριφοράς.

Η ανίχνευση συμβάντων πάνω σε ροές αισθητήρων έχει ως στόχο το προσδιορισμό των τιμών $s_i(t)$, οι οποίες αποτελούν απότομες μεταβολές μέσα σε μία ροή διανυσμάτων πλαισίου. Συγκεκριμένα, κάθε διάνυσμα πλαισίου μήκους n μετατρέπεται σε ένα δυαδικό διάνυσμα του ιδίου μήκους, με κάθε τιμή να αντιπροσωπεύει κάποια πιθανή μεταβολή στην αντίστοιχη ροή αισθητήρων. Τέτοιες αποκλίσεις από τη φυσιολογική συμπεριφορά ονομάζονται συμβάντα και τα δυαδικά διανύσματα ονομάζονται διανύσματα συμβάντων.

Ένα συμβάν μπορεί να είναι μία παρατήρηση η οποία δεν είναι σύμφωνη με ένα αναμενόμενο πρότυπο στο σύνολο δεδομένων. Τα συμβάντα μπορεί να έχουν προκληθεί από διάφορους λόγους, όπως για παράδειγμα βλάβη ή δυσλειτουργία στους αισθητήρες, τιμές απόκλισης ή ουσιαστικές αλλαγές οι οποίες μπορεί να επηρεάσουν τη συμπεριφορά του συστήματος. Ως εκ τούτου, ένα διάνυσμα συμβάντων σε χρόνο t αντιπροσωπεύεται από $\Delta\Sigma_t = (e_1^t, e_2^t, \dots, e_n^t) \epsilon \{0,1\}^n$ όπου $e_i^t = e_i(t)$ είναι η δυαδική τιμή η οποία αντιπροσωπεύει κατά πόσο εμφανίστηκε μια μη φυσιολογική συμπεριφορά στη ροή, η οποία αντιπροσωπεύεται με τιμή ίση με ένα, σε χρόνο t ή τιμή $s_i(t)$ συμπεριλαμβανόταν στο αναμενόμενο εύρος τιμών.

Η μετατροπή ενός διανύσματος πλαισίου σε ένα διάνυσμα συμβάντων βασίζεται σε αλγόριθμους ανίχνευσης μεταβολών, οι οποίοι έχουν ως στόχο τον εντοπισμό μη φυσιολογικών αποκλίσεων στις τρέχουσες τιμές σε σχέση με τις τιμές που προέκυψαν σε προηγούμενα βήματα. Οι αλγόριθμοι ανίχνευσης μεταβολών μπορούν να ταξινομηθούν σε δύο κατηγορίες, η ανίχνευση μεταβολών μίας μεταβλητής και η ανίχνευση μεταβολών πολλαπλών μεταβλητών.

Οι αλγόριθμοι οι οποίοι ανήκουν στην κατηγορία ανίχνευσης μεταβολών μίας μεταβλητής λαμβάνουν υπόψη κάθε ροή αισθητήρων ξεχωριστά και κάνουν ανίχνευση πιθανών ανωμαλιών μέσα από μία ακολουθιακή ανάλυση χρονοσειρών. Οι αλγόριθμοι οι οποίοι ανήκουν στην κατηγορία ανίχνευσης μεταβολών πολλαπλών μεταβλητών εκμεταλλεύονται αυτοπαλίνδρομα μοντέλα πολλαπλών μεταβλητών για την αναπαράσταση κάθε διανύσματος πλαισίου ως ένα γραμμικό άθροισμα της προηγούμενης συμπεριφοράς. Στη συνέχεια, ο στόχος απόκτησης μίας δυαδικής τιμής η οποία υποδεικνύει την μεταβολή ή μη μεταβολή για κάποια συγκεκριμένη μεταβλητή, δηλαδή μία ροή αισθητήρων, ανάγεται σε μία λειτουργία ελέγχου κατωφλίου μεταξύ του μελλοντικού εκτιμώμενου διανύσματος και του πραγματικού διανύσματος.

Οι μέθοδοι ανίχνευσης μεταβολών λαμβάνουν υπόψη τη χρονική σειρά των τιμών των μετρήσεων και κάνουν αναζήτηση για χρονικά σημεία στα οποία οι στατιστικές ιδιότητες των μετρήσεων αλλάζουν απότομα. Σύμφωνα με το [2] η λέξη απότομα συγκεκριμενοποιείται ως «σε άμεσο χρόνο ή τουλάχιστον πολύ γρήγορα εάν λάβουμε υπόψη την περίοδο δειγματοληψίας των μετρήσεων». Οι παρακολουθούμενες στατιστικές ιδιότητες θεωρείται ότι παρουσιάζουν καμία ή πολύ μικρή απόκλιση στις χρονικές στιγμές στις οποίες δεν παρατηρείται κάποια μεταβολή. Λαμβάνοντας υπόψη τις προαναφερθείσες συνθήκες, μπορεί να γίνει ανίχνευση ακόμα και μικρών μεταβολών με αρκετά μεγάλη πιθανότητα. Η πιθανότητα ανίχνευσης μπορεί να είναι ακόμα πιο μεγάλη εάν αυτές οι μεταβολές είναι επίμονες για κάποιο μεγάλο χρονικό διάστημα.

Οι μέθοδοι ανίχνευσης μεταβολών στις πλείστες των περιπτώσεων λειτουργούν χωρίς κάποια υπόθεση ότι οι παρακολουθούμενες μεταβλητές περιγράφονται από κάποια συγκεκριμένη κατανομή. Με άλλα λόγια, οι μέθοδοι ανίχνευσης μεταβολών συνήθως είναι μη παραμετρικές. Ακόμα ένα χαρακτηριστικό των μεθόδων ανίχνευσης μεταβολών αποτελεί η ανίχνευση των μεταβολών σε πολύ σύντομο χρονικό διάστημα ή ακόμα και άμεσα. Επίσης, η πληροφορία μεγέθους κάποιας μεταβολής στις περισσότερες περιπτώσεις δεν είναι κάτι μετρήσιμο ή απαραίτητο.

Ο σχεδιασμός διαδικασιών ανίχνευσης απότομων μεταβολών αποτελείται από δύο μεγάλες υποδιαδικασίες. Η πρώτη υποδιαδικασία είναι προαιρετική και περιλαμβάνει μία επεξεργασία των αρχικών δεδομένων έτσι ώστε οι τελικές τιμές του συνόλου δειγμάτων να μην αποκλίνουν κατά πολύ, από μία αρχική τιμή, από μετρικές όπως ο μέσος όρος, η απόκλιση και άλλα, όταν δεν παρατηρείται μεταβολή. Η αρχική τιμή μπορεί να είναι μηδενική ή κάποια άλλη κατάλληλη τιμή. Σε αυτή την υποδιαδικασία οι τελικές τιμές του συνόλου δειγμάτων αποκλίνουν σε σημαντικό βαθμό από την προαναφερθείσα τιμή αναφοράς όταν παρατηρείται κάποια μεταβολή. Η δεύτερη διαδικασία περιλαμβάνει την ανάπτυξη αλγορίθμων που ανήκουν στην κατηγορία των στατιστικών μεθόδων. Οι αλγόριθμοι αυτοί πρέπει να είναι ικανοί για ανίχνευση των απότομων μεταβολών στο σύνολο δειγμάτων και τις ακριβείς χρονικές στιγμές κατά τις οποίες εμφανίστηκαν.

Στις επόμενες παραγράφους γίνεται μία περιγραφή των αλγορίθμων ανίχνευσης μεταβολών που εξετάζονται σε αυτή την εργασία για τον εντοπισμό πραγματικού χρόνου των συμβάντων σε ροές αισθητήρων.

2.1 Μετρικές απόδοσης αλγορίθμων ανίχνευσης συμβάντων

Ο στόχος ενός αλγορίθμου ανίχνευσης απότομων μεταβολών είναι η ανίχνευση μίας απότομης μεταβολής στην κατανομή πιθανοτήτων μίας ακολουθίας από τυχαίες παρατηρήσεις. Ο ιδανικός αλγόριθμος ανίχνευσης απότομων μεταβολών δεν ανιχνεύει μία απότομη μεταβολή, μέχρι να συμβεί μία απότομη μεταβολή, και όταν η απότομη μεταβολή συμβεί, η ανίχνευση γίνεται σε άμεσο χρόνο. Ωστόσο, λόγω της στοχαστικής διακύμανσης, η ιδανική αυτή περίπτωση δεν είναι εφικτό να συμβεί.

Στην πράξη, υπάρχουν χρονικές στιγμές στις οποίες ο αλγόριθμος κάνει ανίχνευση απότομης μεταβολής, όταν αυτή δεν έχει συμβεί. Η περίπτωση αυτή ονομάζεται ψευδής συναγερμός (false alarm). Επίσης, στις περιπτώσεις απότομων μεταβολών, υπάρχει κάποια καθυστέρηση στην ανίχνευση της αλλαγής. Συνοπτικά, υπάρχουν δύο τυπικές μετρικές απόδοσης, $M\Delta E_0$ και $M\Delta E_1$, όπου $M\Delta E$ ορίζεται ως η μέση διάρκεια εκτέλεσης [32]. Η μετρική $M\Delta E_0$ ορίζεται ως ο μέσος χρόνος μεταξύ ψευδών συναγερμών, ενώ η μετρική $M\Delta E_1$ ορίζεται ως ο χρόνος μέσης καθυστέρησης μεταξύ χρόνου που συμβαίνει απότομη μεταβολή και του χρόνου που γίνεται η ανίχνευση από τον αλγόριθμο. Σε ιδανική περίπτωση, ένας αλγόριθμος απότομων μεταβολών έχει υψηλή τιμή $M\Delta E_0$ και χαμηλή τιμή $M\Delta E_1$. Ωστόσο, συντονίζοντας τις παραμέτρους ενός αλγορίθμου για επίτευξη επιθυμητής τιμής ενός από τα μεγέθη $M\Delta E_0$ και $M\Delta E_1$ έχει αρνητική επίπτωση στο άλλο μέγεθος.

2.2 Γενικός αλγόριθμος διαγραμμάτων ελέγχου

Μία συνήθης μέθοδος στατιστικής ανίχνευσης μεταβολών σε άμεσο χρόνο είναι η μέθοδος διαγραμμάτων ελέγχου [33]. Σε ένα διάγραμμα ελέγχου, η μέση τιμή και η μεταβλητότητα μίας παρακολουθούμενης μεταβλητής περιγράφονται με τα μεγέθη κεντρικού άξονα, ανώτατο όριο ελέγχου και κατώτατο όριο ελέγχου. Γίνεται ανίχνευση μίας μεταβολής εάν η τιμή της μεταβλητής σε κάποια χρονική στιγμή υπερβαίνει κάποιο από τα όρια ελέγχου, το ανώτατο όριο ελέγχου ή το κατώτατο όριο ελέγχου. Η ανίχνευση μεταβολής μπορεί να εκφραστεί με έναν έλεγχο υποθέσεων με μηδενική υπόθεση Υ_0 να περιγράφει καμία σημαντική μεταβολή και η συμπληρωματική υπόθεση Υ_1 να περιγράφει κάποια σημαντική μεταβολή.

Η ιδέα διαγραμμάτων ελέγχου περιγράφηκε για πρώτη φορά με αρχικό κίνητρο την ανίχνευση αλλαγής στις διαδικασίες παραγωγής, για σκοπούς ελέγχου ποιότητας. Ένα διάγραμμα ελέγχου αποτελείται από σημεία $y_1, y_2, \ldots$ τα οποία αντιπροσωπεύουν στατιστικά όρια και όρια ελέγχου α και δ, όπου ισχύει $\alpha < \delta$. Όταν ισχύει $y_\kappa \in (\alpha, \delta)$, η διαδικασία βρίσκεται σε κατάσταση ελέγχου, ενώ όταν ισχύει $y_\kappa \notin (\alpha, \delta)$, η διαδικασία βρίσκεται σε κατάσταση εκτός ελέγχου. Μία ακολουθία από στατιστικές συμβολίζεται με y_κ, συμβολισμός διαφορετικός από το συμβολισμό x_κ, ο οποίος αντιπροσωπεύει μία ακολουθία από παρατηρηθείσες τιμές. Το α ονομάζεται ανώτατο όριο ελέγχου και το δ ονομάζεται κατώτατο όριο ελέγχου. Το $\hat{\tau}$ ονομάζεται το σημείο αλλαγής της ροής δεδομένων, εάν ισχύει $y_{\hat{\tau}} \notin (\alpha, \delta)$, αλλά $y_t \in (\alpha, \delta)$ για όλα τα $t < \hat{\tau}$. Στον παραπάνω συμβολισμό το τ συμβολίζει ένα πραγματικό σημείο αλλαγής.

Ο αλγόριθμος διαγραμμάτων ελέγχου διατηρεί έναν κεντρικό άξονα, ο οποίος αντιπροσωπεύει τη μέση τιμή της παρακολουθούμενης μεταβλητής κάτω από κανονικές συνθήκες. Επίσης διατηρεί τιμές ανώτατου ορίου ελέγχου και κατώτατου ορίου ελέγχου, τα οποία όρια έχουν τιμές πάνω και κάτω από την τιμή κεντρικού άξονα αντίστοιχα, και καθορίζουν το εύρος της κατάστασης που χαρακτηρίζεται από κανονική μεταβλητότητα ή μεταβλητότητα εντός ελέγχου. Η συνάρτηση ελέγχου μεταβλητότητας κάνει μετάβαση στην κατάσταση μη κανονικής μεταβλητότητας ή μεταβλητότητας εκτός ελέγχου εάν η μετρηθείσα τιμή της μεταβλητής είναι εκτός των προκαθορισμένων ορίων.

2.3 Διαγράμματα ελέγχου αλγορίθμου συσσωρευτικού αθροίσματος

Η μέθοδος διαγραμμάτων αλγορίθμου συσσωρευτικού αθροίσματος είναι ένα από τα πρώτα εργαλεία που χρησιμοποιήθηκαν στην ανίχνευση απότομων μεταβολών. Προτάθηκε για πρώτη φορά το 1954 [11] και έχει μελετηθεί ευρέως στη βιβλιογραφία. Η μέθοδος είναι εύκολα διαχειρίσιμη και χρήσιμη για ανίχνευση των θέσεων των σημείων μεταβολής. Συγκεκριμένα, έχει χρησιμοποιηθεί για έλεγχο μεταβολής στις συναρτήσεις μέσης τιμής, απόκλισης και κατανομής.

Ένα πλεονέκτημα της μεθόδου έγκειται στο γεγονός ότι οι συναρτήσεις μέσης τιμής, απόκλισης και κατανομής εκφράζονται ως το άθροισμα ανεξάρτητων και πανομοιότυπα κατανεμημένων τυχαίων μεταβλητών. Έχει επίσης το πλεονέκτημα να λαμβάνει υπόψη το ιστορικό της υπό διερεύνηση σειράς και είναι ικανό να ανιχνεύσει αποτυχία μοντέλου πιο γρήγορα όταν το σφάλμα πρόβλεψης είναι σχετικά μικρό. Το διάγραμμα αλγορίθμου συσσωρευτικού αθροίσματος ενσωματώνει άμεσα όλη την πληροφορία στην ακολουθία του δείγματος με τη γραφική παράσταση των συσσωρευμένων αθροισμάτων των αποκλίσεων των τιμών του συνόλου δειγμάτων από την τιμή του κεντρικού άξονα. Το διάγραμμα μπορεί να κατασκευαστεί τόσο για μεμονωμένες παρατηρήσεις, όσο και για τους μέσους όρους των λογικών υποομάδων του συνόλου δειγμάτων.

2.3.1 Ο αλγόριθμος συσσωρευτικού αθροίσματος και θεωρία ελέγχου αποφάσεων

Ο αλγόριθμος συσσωρευτικού αθροίσματος [11] [13] βασίζεται στο γεγονός ότι το μέγεθος $\sigma_t = \sigma(y_1, \dots, y_t)$ έχει αρνητική τάση τιμών σε κανονικές συνθήκες και θετική τάση τιμών μετά από μία μεταβολή. Η συνάρτηση απόφασης α_t συγκρίνει την αύξηση του σ_t από την ελάχιστη τιμή του με ένα κατώφλι κ:

$$\alpha_t = \sigma_t - min_{1\le i\le t} s_i = \max(0, \sigma(y_t) + \ a_{t-1}) = \ [a_{t-1} + \sigma(y_t)]^+ \ge \kappa \qquad όπου\ \alpha_0 = 0$$

Γίνεται ανίχνευση ενός συμβάντος που περιγράφει μεταβολή εάν η συνάρτηση α_t ξεπερνά την τιμή κατωφλίου κ. Στην περίπτωση αυτή, εάν ο αλγόριθμος συνεχίζει και σε επόμενες χρονικές στιγμές, ο αλγόριθμος επανεκκινεί με τιμή μηδέν στη συνάρτηση α_t.

Ο αλγόριθμος συσσωρευτικού αθροίσματος μπορεί να περιγραφεί και με τη θεωρία ελέγχου υποθέσεων. Σε μια τέτοια περιγραφή, το διάγραμμα ελέγχου αλγορίθμου συσσωρευτικού αθροίσματος εκτελεί με επαναληπτική συμπεριφορά έναν Ακολουθιακό Έλεγχο Λόγου Πιθανοφανειών, στον οποίο κάθε απόφαση λαμβάνει υπόψη όσες διαδοχικές παρελθοντικές παρατηρήσεις είναι απαραίτητο για να γίνει η αποδοχή της κατάστασης Υ_0 ή Υ_1. Ο αλγόριθμος επανεκκινά τη διαδικασία απόφασης με τη μέθοδο Ακολουθιακού Ελέγχου Λόγου Πιθανοφανειών εάν έχει γίνει αποδεκτή η κατάσταση Υ_0. Σε αντίθετη περίπτωση, εάν γίνει αποδεκτή η κατάσταση Υ_1, γίνεται σηματοδότηση ανίχνευσης μεταβολής και ο αλγόριθμος σταματά. Η τιμή κατωφλίου κ προσφέρει μία εξισορρόπηση μεταξύ του μέσου χρόνου καθυστέρησης ανίχνευσης και του μέσου χρόνου μεταξύ λανθασμένων ανιχνεύσεων.

Μία τυπική στατιστική για ανίχνευση θετικών αποκλίσεων από την τιμή κεντρικού άξονα είναι $u^{+}(y) = y - (\mu_0 + K)$, όπου το K ονομάζεται τιμή αναφοράς. Για ανίχνευση και αρνητικών αποκλίσων, μία δεύτερη στατιστική είναι απαραίτητη $u^{-}(y) = (\mu_0 - K) - y$. Οι συναρτήσεις ανίχνευσης μεταβολών ορίζονται:

$$\alpha_t^{+} = [a_{t-1}^{+} + \ y_t - (\mu_0 + K)\ ]^{+} \ \geq \kappa$$

$$\alpha_t^{-} = [a_{t-1}^{-} \ + \ (\mu_0 - K) \ - \ y_t]^{+} \ \geq \kappa$$

Τυπικές τιμές των K και κ είναι $K = \frac{\sigma}{2}$ και $\kappa = 4\sigma$ ή $\kappa = 5\sigma$ όπου σ είναι η τυπική απόκλιση της Y_t.

2.3.2 Ο αλγόριθμος συσσωρευτικού αθροίσματος CUSUM

Ο αλγόριθμος συσσωρευτικού αθροίσματος [11] έχει ως στόχο την ανίχνευση, σε πραγματικό χρόνο, μίας μεταβολής στην κατανομή μίας χρονοσειράς σε σχέση με μία τιμή - στόχο. Συγκεκριμένα, αρχικά γίνεται η υπόθεση μίας μονοδιάστατης χρονοσειράς x_t η οποία αποτελείται από τιμές δεδομένων, οι οποίες συλλέχθηκαν με την πάροδο του χρόνου, και μίας τιμής - στόχος μ για αυτή τη ροή δεδομένων. Ο αλγόριθμος συσσωρευτικού αθροίσματος περιλαμβάνει τον υπολογισμό θετικών μεταβολών Θ και αρνητικών μεταβολών A στη χρονοσειρά x_t συσσωρευτικά με την πάροδο του χρόνου και συγκρίνει αυτές τις μεταβολές με ένα θετικό κατώφλι $κατώφλι^{+}$ και ένα αρνητικό κατώφλι $κατώφλι^{-}$. Κάθε φορά που οι τιμές κατωφλίων ξεπερνούνται, γίνεται μία αναφορά μεταβολής μέσω του σήματος άνω ανίχνευσης σ^{+} και του σήματος κάτω ανίχνευσης σ^{-}, ενώ τα συσσωρευτικά αθροίσματα παίρνουν μηδενική τιμή. Προκειμένου να αποφευχθεί η ανίχνευση μη απότομων μεταβολών ή αργών μετατοπίσεων, ο αλγόριθμος λαμβάνει υπόψη παραμέτρους ανεκτικότητας για θετικές μεταβολές α^{+} και για αρνητικές μεταβολές α^{-}.

Οι παράμετροι εισόδου για τον αλγόριθμο συσσωρευτικού αθροίσματος είναι η τιμή στόχος μ, η τιμή άνω ανεκτικότητας α^{+}, η τιμή κάτω ανεκτικότητας α^{-}, η τιμή άνω κατωφλίου $κατώφλι^{+}$ και η τιμή κάτω κατωφλίου $κατώφλι^{-}$. Οι παράμετροι εξόδου είναι το σήμα άνω ανίχνευσης σ^{+} και το σήμα κάτω ανίχνευσης σ^{-}. Στη συνέχεια γίνεται παρουσίαση του αλγορίθμου συσσωρευτικού αθροίσματος.

Είσοδος: τιμή στόχος μ, τιμή άνω ανεκτικότητας α^{+}, τιμή κάτω ανεκτικότητας α^{-}, τιμή άνω κατωφλίου $κατώφλι^{+}$, τιμή κάτω κατωφλίου $κατώφλι^{-}$

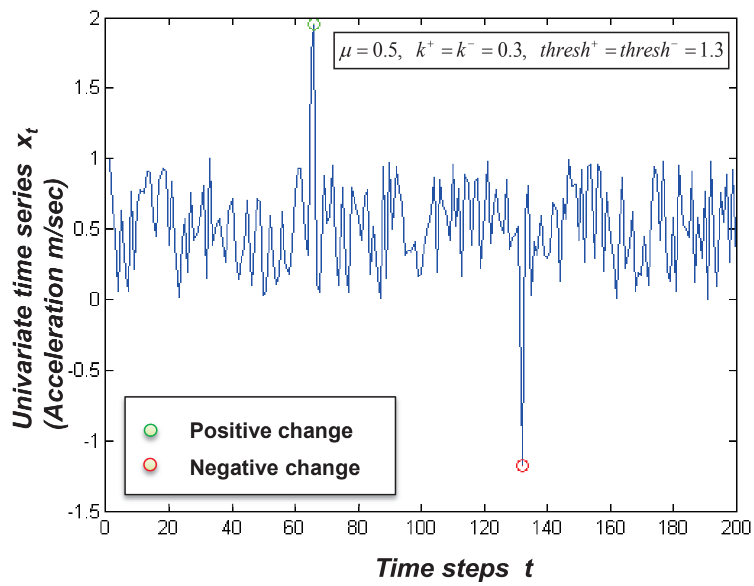
Έξοδος: σήμα άνω ανίχνευσης σ^{+}, σήμα κάτω ανίχνευσης σ^{-}

$\Theta \leftarrow 0$;
$A \leftarrow 0$;
$t \leftarrow 1$;
while ($true$)
    $\sigma^{+} \leftarrow 0$;
    $\sigma^{-} \leftarrow 0$;
    $\Theta \leftarrow \max(0, x_t - (\mu + \alpha^{+}) + \Theta)$;
    $A \leftarrow \min(0, x_t - (\mu - \alpha^{-}) + A)$;
    if $(\Theta > κατώφλι^{+})$ **then**
        $\sigma^{+} \leftarrow 1$;
        $\Theta \leftarrow 0$;
        $A \leftarrow 0$;
    end
    if $(A < κατώφλι^{-})$ **then**
        $\sigma^{+} \leftarrow 1$;
        $\Theta \leftarrow 0$;
        $A \leftarrow 0$;
    end
    $t \leftarrow t + 1$;
end

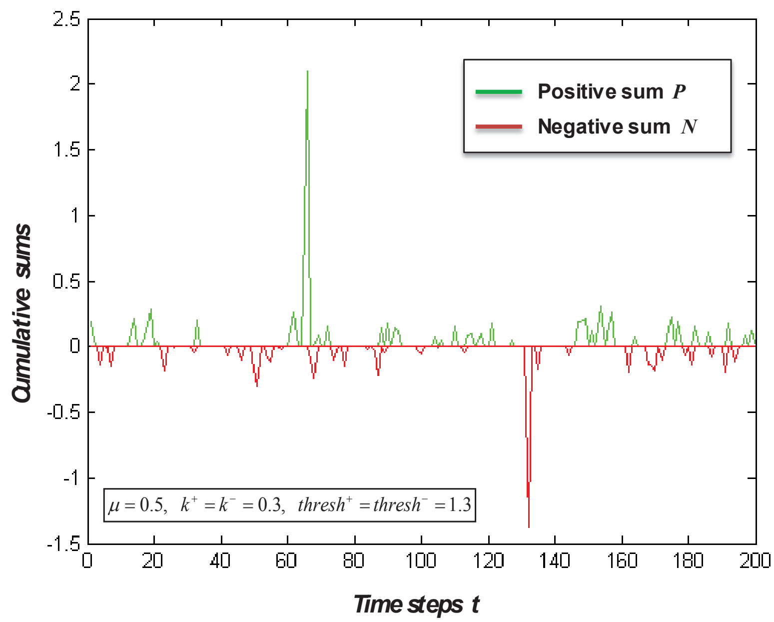
Αλγόριθμος 1: Αλγόριθμος Συσσωρευτικού Αθροίσματος

Ο αλγόριθμος υποθέτει ότι η χρονοσειρά που καταφθάνει ακολουθεί κανονική κατανομή. Προκειμένου ο αλγόριθμος να λειτουργήσει σωστά, πρέπει να γίνει συντονισμός των παραμέτρων ανεκτικότητας και κατωφλίου με τρόπο που να καθορίζει τι είναι μία πραγματική μεταβολή για κάποια συγκεκριμένη χρονοσειρά. Ο συντονισμός αυτός μπορεί να εκτελεστεί ακολουθώντας κάποια βήματα [34]. Αρχικά, γίνεται έναρξη της διαδικασίας με μεγάλες τιμές των $κατώφλι^{+}$ και $κατώφλι^{-}$. Στη συνέχεια, επιλέγονται οι παράμετροι άνω ανεκτικότητας α^{+} και κάτω ανεκτικότητας α^{-} στο μέσο της αναμενόμενης μεταβολής ή γίνεται προσαρμογή τους έτσι ώστε οι μετρικές Θ και A να έχουν μηδενική τιμή περισσότερες από το ένα δεύτερο φορές. Οι τιμές των $κατώφλι^{+}$ και $κατώφλι^{-}$ ρυθμίζονται έτσι ώστε να επιτυγχάνεται ο απαιτούμενος αριθμός λανθασμένων συναγερμών ή ο απαιτούμενος χρόνος καθυστέρησης. Εάν το απαιτούμενο είναι ταχύτερη ανίχνευση, γίνεται μείωση των τιμών άνω ανεκτικότητας α^{+} και κάτω ανεκτικότητας α^{-}. Εάν είναι επιθυμητός μικρότερος αριθμός από λανθασμένους συναγερμούς ή γίνεται ανίχνευση μεταβολών οι οποίες δεν βγάζουν νόημα, γίνεται αύξηση των τιμών άνω ανεκτικότητας α^{+} και κάτω ανεκτικότητας α^{-}.

Στις Εικόνες 3 και 4 απεικονίζεται ένα παράδειγμα του αλγορίθμου συσσωρευτικού αθροίσματος πάνω σε μία ροή αισθητήρων, όπου γίνεται ανίχνευση δύο μεταβολών, θετικής και αρνητικής. Η τιμή στόχος για x_i ορίζεται σε $\mu = 0,5,$ οι τιμές ανεκτικότητας ορίζονται σε $\alpha^{+} = \alpha^{-} = 0,3$ και οι τιμές κατωφλίου προσδιορίζονται σε $κατώφλι^{+} = κατώφλι^{-} = 1,3$. Συγκεκριμένα, η Εικόνα 3 παρουσιάζει τα αρχικά δεδομένα αισθητήρων και τα χρονικά βήματα στα οποία ανιχνεύεται μεταβολή από τον αλγόριθμο. Η Εικόνα 4 απεικονίζει τα συσσωρευτικά αθροίσματα των θετικών και αρνητικών μεταβολών σε συνάρτηση με το χρόνο για x_t. Στην περίπτωση δεδομένων αισθητήρων πολλαπλών μεταβλητών, ο αλγόριθμος συσσωρευτικού αθροίσματος πρέπει να εφαρμόζεται σε κάθε μεταβλητή ξεχωριστά.



Εικόνα 3: Αυθεντική ροή αισθητήρων καταμέτρησης επιτάχυνσης μέσω MPU και ανίχνευση μεταβολών με το αλγόριθμο συσσωρευτικού αθροίσματος



Εικόνα 4: Συσσωρευτικά αθροίσματα θετικών και αρνητικών μεταβολών

2.3.3 Μονόπλευρος και αμφίπλευρος αλγόριθμος συσσωρευτικού αθροίσματος

Οι αλγόριθμος συσσωρευτικού αθροίσματος μπορεί να διαχωριστεί σε δύο τύπους ή παραλλαγές [13] [35]. Ο πρώτος τύπος είναι ο μονόπλευρος αλγόριθμος συσσωρευτικού αθροίσματος, ο οποίος μπορεί να χρησιμοποιηθεί όταν οι τιμές του μέσου όρου πριν και μετά από απότομη μεταβολή είναι γνωστές εκ των προτέρων. Ο δεύτερος τύπος είναι ο αμφίπλευρος CUSUM, ο οποίος χρησιμοποιείται όταν το μέγεθος της μεταβολής δεν είναι γνωστό εκ των προτέρων.

2.3.3.1 Μονόπλευρος αλγόριθμος συσσωρευτικού αθροίσματος

Η συνάρτηση $\lambda_n = \sum_{i=1}^{n} z_i$ όπου

$$z_i = \ln\frac{P_{\theta_1}(y_i)}{P_{\theta_0}(y_i)} = \frac{\mu_1 - \mu_o}{\sigma^2}\left(y_i - \frac{\mu_0 + \mu_1}{2}\right) = \frac{\beta}{\sigma}\left(y_i - \frac{\mu_0 + \mu_1}{2}\right) = \frac{\beta}{\sigma}(y_i - \mu_0 - \frac{v}{2})$$

είναι ο λογαριθμικός λόγος πιθανοφανειών για τις παρατηρηθείσες τιμές από y_1 έως y_n, και το μέγεθος μεταβολής ορίζεται από $v = \mu_1 - \mu_0$ και ισχύει $\beta = \frac{\mu_1 - \mu_0}{\sigma}$. Στο n σημείο δειγματοληψίας, η συνάρτηση απόφασης έχει συσσωρευτικό χαρακτήρα και εκφράζεται ως:

$$\alpha_n = \frac{\beta}{\sigma}\sum_{i=1}^{n}(y_i - \mu_0 - \frac{v}{2}).$$

Στην περίπτωση θετικής μεταβολής στο μέσο όρο, η τυπική συμπεριφορά της συνάρτησης απόφασης παρουσιάζει μία αρνητική μετατόπιση πριν την μεταβολή και μία θετική μετατόπιση μετά την μεταβολή. Στην περίπτωση αρνητικής μεταβολής στο μέσο όρο, η τυπική συμπεριφορά της συνάρτησης απόφασης παρουσιάζει μία θετική μετατόπιση πριν τη μεταβολή και μία αρνητική μετατόπιση μετά τη μεταβολή.

Στην περίπτωση του μονόπλευρου αλγορίθμου συσσωρευτικού αθροίσματος, η βασική υπόθεση είναι ότι έχει παρουσιαστεί μία μεταβολή γνωστού μεγέθους στο μέσο όρο, μεταβολή που μπορεί να είναι θετική ή αρνητική. Στην περίπτωση θετικής μεταβολής στο μέσο όρο, η σχετική πληροφορία μεταβολής προκύπτει από τη διαφορά μεταξύ του λογαριθμικού λόγου πιθανοφανειών και του μέχρι τώρα ελαχίστου των τιμών δειγμάτων του συνόλου. Ο κανόνας απόφασης στην περίπτωση αυτή είναι η σύγκριση, σε κάθε χρονική στιγμή, της διαφοράς που προαναφέρθηκε με ένα προκαθορισμένο κατώφλι. Συνεπώς, εμφανίζεται συναγερμός απότομης μεταβολής εάν ισχύει

$$\sigma_n = \alpha_n - min_{0 \le t \le n} \alpha_t \ \geq \lambda \text{ όπου } \alpha_t = (\alpha_{t-1} + y_t - \mu_0 - \frac{v}{2})^{+}, \alpha_0 = 0$$

όπου σ_n είναι η συνάρτηση απόφασης αλγορίθμου συσσωρευτικού αθροίσματος.

2.3.3.2 Αμφίπλευρος αλγόριθμος συσσωρευτικού αθροίσματος

Μία άλλη περίπτωση ανίχνευσης μεταβολής χρησιμοποιώντας τον αλγόριθμο συσσωρευτικού αθροίσματος είναι η περίπτωση στην οποία το μέγεθος της μεταβολής δεν είναι γνωστό. Στην περίπτωση αυτή, το μέγεθος της μεταβολής εκφράζεται ως ένα προκαθορισμένο ελάχιστο μέγεθος μεταβολής, το οποίο μπορεί να είναι $\mu_1^+ = \mu_0 + v$ στην περίπτωση της θετικής μετατόπισης στο μέσο όρο και $\mu_1^- = \mu_0 - v$ στην περίπτωση της αρνητικής μετατόπισης στο μέσο όρο.

Στον αμφίπλευρο έλεγχο αλγορίθμου συσσωρευτικού αθροίσματος, είναι χρήσιμο να γίνεται εκτέλεση δύο παράλληλων ελέγχων, όπου ο πρώτος εμφανίζει συναγερμό απότομης μεταβολής στην περίπτωση αύξησης στο μέσο όρο και ο δεύτερος στην περίπτωση μείωσης στο μέσο όρο. Συνεπώς, εμφανίζεται συναγερμός απότομης μεταβολής εάν ισχύει

$$\sigma_n^+ = a_n^+ - min_{0\le t\le n} a_t^+ \ \ge \lambda$$

$$\sigma_n^- = max_{0\le t\le n} a_t^- - a_n^- \ \ge \lambda$$

όπου $\alpha_t^- = (a_{t-1}^- - y_t + \mu_0 - \frac{v}{2})^+ , a_0^+ = a_0^- = 0$

Η προκύπτουσα χρονική στιγμή εμφάνισης συναγερμού απότομης μεταβολής δίνεται από:

$$t_{\sigma\upsilon\nu\alpha\gamma\varepsilon\rho\mu\acute{o}\varsigma} = \min\{t \ge 1 : (a_t^+ \ \ge \lambda) \cup (a_t^- \ \ge \lambda)\}$$

Η παράμετρος μεγέθους μεταβολής, εφόσον δεν είναι εκ των προτέρων γνωστή, μπορεί να έχει διάφορες μορφές, ανάλογα με τα χαρακτηριστικά της εφαρμογής. Στις περισσότερες εφαρμογές η διαθέσιμη πληροφορία όσον αφορά το μέγεθος αυτό είναι πολύ μικρή. Το μέγεθος μεταβολής μπορεί να καθοριστεί ως το μικρότερο δυνατό μέγεθος μεταβολής, το πιο πιθανό μέγεθος μεταβολής ή το μέγεθος για το οποίο δεν εμφανίζεται συναγερμός για ένα ανεκτό αριθμό περιπτώσεων, ενώ υπήρχε απότομη μεταβολή.

2.3.4 Προσέγγιση τιμών μέσου όρου και τυπικής απόκλισης

Ο λανθασμένος καθορισμός τιμών μέσου όρου και τυπικής απόκλισης των προς διερεύνηση δεδομένων είναι πιθανό να οδηγήσει σε ένα μη επιθυμητό αριθμό από λανθασμένους συναγερμούς. Σε κάποιες εφαρμογές, τα δεδομένα ελέγχου μπορούν να χρησιμοποιηθούν για τον υπολογισμό της μέσης τιμής και της τυπικής απόκλισης. Στη συνέχεια, οι αρχικές προσεγγίσεις των παραμέτρων μπορούν να χρησιμοποιηθούν για τη βαθμονόμηση του αλγορίθμου συσσωρευτικού αθροίσματος.

Στην περίπτωση που δεν υπάρχουν διαθέσιμα δεδομένα ελέγχου στο σύνολο δεδομένων, χρησιμοποιείται ένας αλγόριθμος συσσωρευτικού αθροίσματος ο οποίος ξεκινάει χωρίς κάποια εκ των προτέρων πληροφορία, όσον αφορά τις τιμές μέσου όρου και τυπικής απόκλισης. Ο αλγόριθμος αυτός υπολογίζει τις τιμές μέσου όρου και τυπικής απόκλισης σε πραγματικό χρόνο, χρησιμοποιώντας τα δεδομένα που έχουν συγκεντρωθεί μέχρι την τρέχουσα χρονική στιγμή. Σε ένα τέτοιο σύστημα, οι τιμές μέσου όρου και τυπικής απόκλισης ενημερώνονται ακολουθιακά και σε πραγματικό χρόνο.

2.3.5 Προσέγγιση τιμών αναμενόμενου μεγέθους μετατόπισης και κατωφλίου

Για την επίτευξη του βέλτιστου αλγορίθμου συσσωρευτικού αθροίσματος, πρέπει να γίνει προσέγγιση της αναμενόμενης τιμής μετατόπισης με τέτοιο τρόπο ώστε να ελαχιστοποιούνται οι αριθμοί χαμένων απότομων μεταβολών και λανθασμένων συναγερμών. Το κριτήριο βελτιστοποίησης είναι η επιλογή τιμής μετατόπισης, η οποία προσφέρει ανίχνευση απότομης μεταβολής με μικρό χρόνο καθυστέρησης και μικρό αριθμό λανθασμένων συναγερμών.

Τυπικά, το μέγεθος αναμενόμενης μετατόπισης μπορεί να οριστεί της τάξης ενός ή δύο τυπικών αποκλίσεων των δεδομένων ελέγχου. Εκτός από το μέγεθος αναμενόμενης μετατόπισης, η επιλογή τιμής κατωφλίου μπορεί επίσης να επηρεάσει το χρόνο καθυστέρησης ανίχνευσης απότομης μεταβολής. Τυπικά, αρχικά γίνεται μία προσέγγιση της τιμής μεγέθους αναμενόμενης μεταβολής και στη συνέχεια καθορισμός της τιμής κατωφλίου.

2.3.6 Επιλογή κριτηρίων

Το τυπικό μέγεθος απόδοσης για ανίχνευση απότομης μεταβολής σε πραγματικό χρόνο είναι ο χρόνος καθυστέρησης ανίχνευσης, ο οποίος πρέπει να ελαχιστοποιείται για ένα σταθερό ποσοστό λανθασμένων συναγερμών. Μπορεί επίσης να γίνει ορισμός των κριτηρίων, τα οποία μπορούν να χρησιμοποιηθούν για προσέγγιση βέλτιστων χρόνων διακοπής στην ανίχνευση απότομων μεταβολών. Τα κριτήρια πρέπει να ευνοούν ανίχνευση απότομων μεταβολών με ελάχιστο χρόνο καθυστέρησης και μικρό αριθμό λανθασμένων συναγερμών. Ο χρόνος καθυστέρησης ανίχνευσης απότομων μεταβολών είναι επιθυμητός για το λόγο ότι, η διαφορά μεταξύ των χρόνων πραγματικής απότομης μεταβολής και εμφάνισης συναγερμού αυτής μπορεί να οδηγήσει σε εσφαλμένη ερμηνεία των χαρακτηριστικών του προς διερεύνηση μεγέθους.

Από τα προαναφερθέντα εμφανίζεται η ανάγκη για εξισορρόπηση μεταξύ των μεγεθών μέσου χρόνου μεταξύ λανθασμένων συναγερμών και του χρόνου καθυστέρησης ανίχνευσης απότομης μεταβολής. Τα δύο προαναφερθέντα μεγέθη αυξάνονται, όσο ο αλγόριθμος είναι λιγότερο ευαίσθητος σε υψηλές συχνότητες. Μία σχετική εξισορρόπηση μεγεθών, η οποία είναι επίσης επιθυμητή, είναι η εξισορρόπηση μεταξύ των μεγεθών αποδοτικότητας και πολυπλοκότητας. Όταν το σχεδιασμένο σύστημα ανίχνευσης περιλαμβάνει σε κάθε συναγερμό απότομης μεταβολής μία επεξεργασία, η οποία είναι χρονοβόρα ή και ακριβή όσον αφορά τις πράξεις υπολογισμού, ή και αναδιαμόρφωση ελέγχου, ο αριθμός λανθασμένων συναγερμών είναι πιο μεγάλος. Σε κάποια συστήματα, είναι χρήσιμο να υπάρχει μείωση της πολυπλοκότητας υπολογισμού, χωρίς επηρεασμό της αποδοτικότητας του αλγορίθμου. Ένα παράδειγμα για την επίτευξη του στόχου αυτού είναι η χρήση πιθανού πλεονασμού δεδομένων.

2.4 Διαγράμματα ελέγχου αλγορίθμου Shewhart

Τα διαγράμματα ελέγχου αλγορίθμου Shewhart [25] καθορίζουν ένα μέγεθος κεντρικού άξονα, όπως επίσης και τα μεγέθη ανώτατου ορίου ελέγχου AOE και κατώτατου ορίου ελέγχου KOE. Τα μεγέθη αυτά καθορίζονται με βάση κάποια μετρική, η οποία προσδιορίζεται από N παρατηρηθείσες τιμές της μεταβλητής $y_{(j-1)N+1}, \dots, y_{jN}$. Ένα παράδειγμα τέτοιας μετρικής είναι η μέση τιμή $\overline{y_J}$, η οποία είναι κατάλληλη για ανίχνευση μεταβολών στο μέσο όρο:

$$\overline{y_J} = \frac{1}{N}\sum_{t=(j-1)N+1}^{jN} y_t, \text{ όπου } j = 1,2, \dots$$

Εάν οι παρατηρηθείσες τιμές είναι ανεξάρτητες και ομοιόμορφα κατανεμημένες με μέσο μ_0 και τυπική απόκλιση σ^2, η μετρική $\overline{y_J}$ είναι μία μετρική εκτίμησης του μ_0 με απόκλιση $\frac{\sigma^2}{N}$. Ως εκ τούτου, το ανώτατο όριο ελέγχου και το κατώτατο όριο ελέγχου ορίζονται ως εξής:

$$AOE = \mu_0 + \frac{L\sigma}{\sqrt{N}}$$

$$KOE = \mu_0 - \frac{L\sigma}{\sqrt{N}}$$

όπου L είναι η παράμετρος συντονισμού. Ενεργοποιείται ένας συναγερμός ανίχνευσης μεταβολής, εάν η τιμή $\overline{y_J}$ ξεπερνάει την τιμή ανώτατου όριου ελέγχου ή την τιμή κατώτατου όριου ελέγχου.

2.4.1 Ο αλγόριθμος διαγραμμάτων ελέγχου Shewhart και θεωρία ελέγχου υποθέσεων

Οι μέθοδοι ανίχνευσης οριακών τιμών αρχικά παρουσιάστηκαν στο πεδίο του ποιοτικού ελέγχου, και συγκεκριμένα στα συστήματα συνεχούς ελέγχου. Στη συνέχεια περιγράφονται οι μαθηματικές έννοιες, οι οποίες χαρακτηρίζουν τα συστήματα αυτά [13] [36]. Γίνεται υπόθεση K τιμών δειγμάτων ενός συνόλου δεδομένων με σταθερό μέγεθος N. Στο τέλος του κάθε δείγματος, γίνεται υπολογισμός ενός κανόνα απόφασης, ο οποίος αποτελεί είσοδο σε μία διαδικασία ελέγχου μεταξύ των υποθέσεων:

$$\Upsilon_0 : \theta = \theta_0$$

$$\Upsilon_1 : \theta = \theta_1$$

όπου το θ είναι η τιμή κεντρικού άξονα των δεδομένων, που στη συγκεκριμένη περίπτωση είναι η μετρική του μέσου όρου των δειγμάτων.

Εάν η διαδικασία απόφασης έχει ως αποτέλεσμα την υπόθεση Υ_0, συνεχίζεται η δειγματοληψία και ο έλεγχος. Εάν η διαδικασία απόφασης έχει ως αποτέλεσμα την υπόθεση Υ_1, η διαδικασία δειγματοληψίας διακόπτεται. Στην περίπτωση που το σύνολο δειγμάτων έχει σταθερό μέγεθος και είναι προκαθορισμένο, ο κανόνας απόφασης α_K δίδεται από:

$$\alpha_K = 0 \; \varepsilon\acute{\alpha}\nu \; A_1^N < \lambda, \varepsilon\pi\iota\lambda o\gamma\acute{\eta} \; \Upsilon_0$$

$$\alpha_K = 1 \; \varepsilon\acute{\alpha}\nu \; A_1^N \geq \lambda, \varepsilon\pi\iota\lambda o\gamma\acute{\eta} \; \Upsilon_1$$

όπου $K = 1, 2, 3, \dots, N$ και A_1^N είναι μία συνάρτηση απόφασης και λ είναι ένα κατώφλι, το οποίο έχει τιμή που διαφέρει και προσαρμόζεται ανάλογα με την εφαρμογή. Η απόφαση λαμβάνεται με τη βοήθεια ενός κανόνα διακοπής, ο οποίος στην περίπτωση αυτή ορίζεται ως:

$$t_{συναγερμός} = N.\min\{K: \alpha_K = 1\}$$

όπου α_K είναι ο κανόνας απόφασης για τον αριθμό των δειγμάτων από το σύνολο δεδομένων K ή το προκαθορισμένο αριθμό δειγμάτων N και $t_{συναγερμός}$ είναι ο τρέχων αριθμός εκτέλεσης της διαδικασίας, στον οποίο παρουσιάστηκε συναγερμός απότομης μεταβολής.

Στην περίπτωση που η κατανομή των τιμών είναι Γκαουσιανή με μέση τιμή $\theta = \mu$ και σταθερή τιμή τυπικής απόκλισης σ^2, η συνάρτηση πυκνότητας πιθανότητας ορίζεται ως:

$$P_\theta(y) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(y-\mu)^2}{2\sigma^2}}$$

Γίνεται δημιουργία ενός διαγράμματος ελέγχου αλγορίθμου Shewhart, σύμφωνα με το οποίο όταν ισχύει $\mu_1 > \mu_0$, ο συναγερμός απότομης μεταβολής εμφανίζεται τη χρονική στιγμή που περιγράφεται από:

$$\bar{y}(K) \geq \mu_0 + \delta \frac{\sigma}{\sqrt{N}}\ , K = 1, 2, 3, \dots, N$$

όπου η τιμή του μέσου όρου των τιμών δειγμάτων μέχρι τη χρονική στιγμή κ είναι:

$$\bar{y}(K) = \frac{1}{N} \sum_{i=N(K-1)+1}^{NK} y_\iota$$

όπου δ και N είναι οι παραμέτροι ρύθμισης του διαγράμματος ελέγχου αλγορίθμου Shewhart, δηλαδή οι παράμετροι οι οποίοι ρυθμίζονται σύμφωνα με τα χαρακτηριστικά της εφαρμογής. Οι παράμετροι αυτοί μπορεί να οριστούν από την αρχή, αλλά συνήθως είναι πιο χρήσιμο να γίνει μία προσέγγιση τους με τη βοήθεια κάποιας διαδικασίας προεπεξεργασίας των δεδομένων.

Το μέγεθος $\mu_0 + \delta \frac{\sigma}{\sqrt{N}}$ είναι το ανώτατο όριο ελέγχου, το οποίο χρησιμοποιείται για ανίχνευση απότομης μεταβολής, η οποία προκύπτει από την αύξηση των τιμών δειγμάτων. Παρόλο που σε κάποιες εφαρμογές γίνεται χρήση μόνο του ανώτατου ορίου ελέγχου, στις περισσότερες των εφαρμογών είναι χρήσιμο να γίνεται ανίχνευση απότομης μεταβολής, η οποία προκύπτει από την αύξηση αλλά και τη μείωση των τιμών δειγμάτων. Στην περίπτωση αυτή, η τιμή του μέσου όρου μετά την απότομη μεταβολή μπορεί να είναι $\mu_1^+ = \mu_0 + \nu$ ή $\mu_1^- = \mu_0 - \nu$. Σε αυτή την περίπτωση, το κατώτατο όριο ελέγχου ορίζεται ως $\mu_0 - \delta \frac{\sigma}{\sqrt{N}}$ και ο συναγερμός απότομης μεταβολής εμφανίζεται τη χρονική στιγμή που περιγράφεται από:

$$|\bar{y}(K) - \mu_0| \geq \delta \frac{\sigma}{\sqrt{N}}, K = 1, 2, 3, \ldots, N$$

2.4.2 Ο αλγόριθμος διαγραμμάτων ελέγχου Shewhart

Τα διαγράμματα ελέγχου αλγορίθμου Shewhart [25] παρέχουν ένα στατιστικό μέτρο για την ανίχνευση απότομων μετατοπίσεων μίας χρονοσειράς. Στο διάγραμμα ελέγχου αλγορίθμου Shewhart, μία μεταβλητή x_t ανιχνεύεται να αποκλίνει σε χρόνο t από την κανονικότητα της, κάθε φορά που ξεπερνάει ένα από τα όρια ελέγχου, τα οποία καθορίζονται από τον αλγόριθμο: το ανώτατο όριο ελέγχου και το κατώτατο όριο ελέγχου.

Τα όρια ελέγχου ορίζονται ως η μετρικές απόστασης από την τρέχουσα τιμή μέσου όρου της στατιστικής διαδικασίας x_t. Συγκεκριμένα, το ανώτατο όριο ελέγχου και κατώτατο όριο ελέγχου ορίζονται ως:

$$AOE = \overline{x_t} + \kappa . \sigma_t$$

$$KOE = \overline{x_t} - \kappa . \sigma_t$$

όπου $\overline{x_t}$ αντιπροσωπεύει το μέσο όρο της χρονοσειράς σε χρόνο t και σ_t είναι η τυπική απόκλιση στο ίδιο χρονικό βήμα. Η παράμετρος κ αντιπροσωπεύει τη στεγανότητα της διαδικασίας ανίχνευσης μεταβολών. Χαμηλές τιμές κ οδηγούν σε στεγανό έλεγχο της διαδικασίας μετρήσεων, ενώ μεγάλες τιμές σηματοδοτούν μόνο τις μετρήσεις, οι οποίες είναι σε μεγάλο βαθμό εκτός ελέγχου.

Σε χρονικό βήμα t η ροή δεδομένων x_t ανιχνεύεται να ενεργοποιεί ένα συναγερμό, εάν ισχύει $x_t > AOE$ ή $x_t < KOE$. Ο αλγόριθμος επιστρέφει ένα σήμα ανίχνευσης εξόδου σ σε κάθε βήμα t όπου $\sigma = 1$ εάν υπάρχει ανίχνευση μεταβολής. Στην περίπτωση φυσιολογικής συμπεριφοράς, δηλαδή $x_i \in [AOE, KOE]$, ο αλγόριθμος ορίζει $\sigma = 1$. Στη συνέχεια παρουσιάζεται ο αλγόριθμος διαγραμμάτων ελέγχου Shewhart.

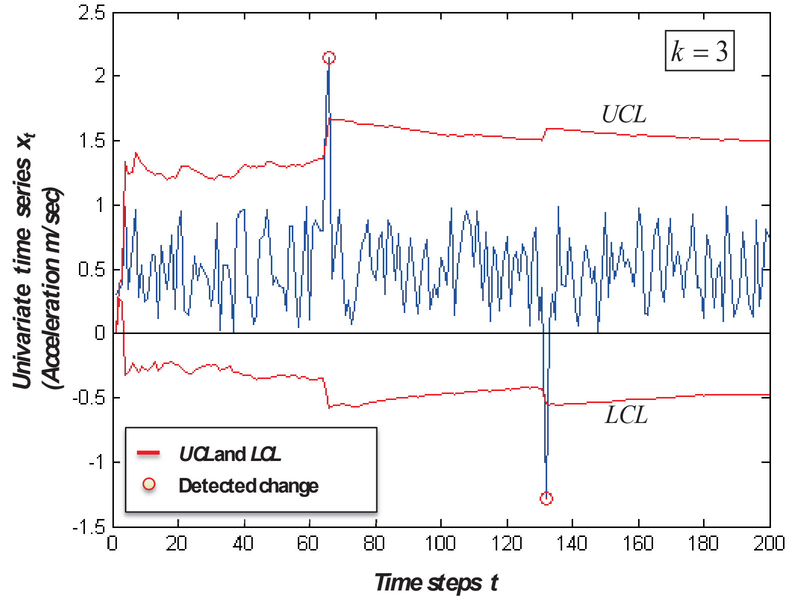
Είσοδος: μονοδιάστατη χρονοσειρά x_t, στεγανότητα κ

Έξοδος: σήμα ανίχνευσης σ

$\bar{x}_0 \leftarrow 0$;
$\sigma_0 \leftarrow 0$;
$t \leftarrow 1$;
while ($true$)
  $\bar{x}_t \leftarrow \bar{x}_{t-1} + \frac{x_t - \bar{x}_{t-1}}{t}$;
  $\sigma_t \leftarrow \sqrt{\frac{1}{t}\left((t-1)\cdot\sigma_{t-1}^2 + (x_t - \bar{x}_t)(x_t - \bar{x}_{t-1})\right)}$;
  $AOE_i \leftarrow \bar{x}_t + \kappa.\sigma_t$;
  $KOE_i \leftarrow \bar{x}_t - \kappa.\sigma_t$;
  if $((x_t > AOE)$ **or** $(x_t < KOE))$ **then**
    $\sigma \leftarrow 1$;
  else
    $\sigma \leftarrow 0$;
  end
  $t \leftarrow t+1$;
end

Αλγόριθμος 2: Αλγόριθμος διαγραμμάτων ελέγχου Shewhart

Η Εικόνα 5 απεικονίζει ένα παράδειγμα του αλγορίθμου διαγραμμάτων ελέγχου Shewhart πάνω σε μία ροή αισθητήρων επιταχυνσιόμετρου Arduino. Με παρόμοιο τρόπο, όπως στον αλγόριθμο συσσωρευτικού αθροίσματος, στην περίπτωση χρονοσειρών πολλαπλών μεταβλητών, το διάγραμμα ελέγχου αλγορίθμου Shewhart πρέπει να εφαρμοστεί σε κάθε μεταβλητή ξεχωριστά. Ωστόσο, ο αλγόριθμος υποθέτει κανονική κατανομή της μεταβλητής x_t. Το γεγονός αυτό κάνει τον αλγόριθμο αρκετά εύρωστο σε σύνολα δεδομένων πραγματικού χρόνου, όπου στις περισσότερες των περιπτώσεων δεν υπάρχει κάποια διαθέσιμη πληροφορία για την κατανομή πιθανοτήτων την οποία ακολουθεί μία ροή δεδομένων. Από την άλλη πλευρά, ο αλγόριθμος είναι λιγότερο προσαρμοστικός σε σύγκριση με τον αλγόριθμο συσσωρευτικού αθροίσματος, αφού τα όρια ελέγχου μπορούν να τροποποιηθούν ελάχιστα στην περίπτωση μεγάλων χρονοσειρών.



Εικόνα 5: Αυθεντική ροή αισθητήρων καταμέτρησης επιτάχυνσης μέσω MPU και ανίχνευση μεταβολών με τον αλγόριθμο διαγραμμάτων ελέγχου Shewhart

2.5 Σύγκριση αλγορίθμου συσσωρευτικού αθροίσματος και αλγορίθμου διαγραμμάτων ελέγχου Shewhart

Λόγω της ποικιλομορφίας μαθηματικών πολυπλοκοτήτων των αλγορίθμων ανίχνευσης μεταβολών, αλλά και του μεγάλου αριθμού πιθανών εφαρμογών, η επιλογή των κατάλληλων αλγορίθμων γίνεται ανάλογα με τα χαρακτηριστικά της κάθε εφαρμογής. Ο αλγόριθμος διαγραμμάτων ελέγχου Shewhart μπορεί να ανιχνεύσει αποκλίσεις από την τιμή κεντρικού άξονα όταν αυτές είναι αρκετά μεγάλες, και σε χρονικές στιγμές οι οποίες είναι μεταγενέστερες των αποκλίσεων. Ο αλγόριθμος διαγραμμάτων ελέγχου Shewhart, στις περισσότερες των περιπτώσεων, δεν μπορεί να ανιχνεύσει αποκλίσεις από την τιμή κεντρικού άξονα όταν αυτές είναι μικρού μεγέθους. Μάλιστα, έχει παρατηρηθεί ότι η μη ανίχνευση των αποκλίσεων αυτών συμβαίνει ακόμα και στις περιπτώσεις που οι αποκλίσεις αυτές είναι επίμονες [37].

Ο αλγόριθμος συσσωρευτικού αθροίσματος είναι ευαίσθητος σε αποκλίσεις μετρίου ή μικρού μεγέθους από την τιμή κεντρικού άξονα, οι οποίες είναι επίμονες. Η τιμή κεντρικού άξονα μπορεί να βασίζεται στο μέσο όρο, στην τυπική απόκλιση και άλλα. Επίσης, η ανίχνευση των αποκλίσεων αυτών γίνεται σε χρονικές στιγμές οι οποίες δεν είναι πολύ μεταγενέστερες των αποκλίσεων. Με άλλα λόγια, η καθυστέρηση ανίχνευσης αποκλίσεων είναι σχετικά μικρή.

Σε σύγκριση με τον αλγόριθμο διαγραμμάτων ελέγχου Shewhart, ο αλγόριθμος συσσωρευτικού αθροίσματος ανιχνεύει μεταβολές με υψηλότερες πιθανότητες, όταν αυτές έχουν μικρή μεταβλητότητα, αλλά επιμονή στο χρόνο. Αυτό δικαιολογείται από το γεγονός ότι ο αλγόριθμος συσσωρευτικού αθροίσματος που περιεγράφηκε προηγουμένως συσσωρεύει λίγες επιδράσεις σε συνάρτηση με το χρόνο. Στο άρθρο [22] έγινε μία μελέτη του αλγόριθμου συσσωρευτικού αθροίσματος μη παραμετρικού χαρακτήρα, για μία συγκεκριμένη οικογένεια εκθετικών κατανομών. Υπό αυτές τις προϋποθέσεις, το χρονικό διάστημα καθυστέρησης ανίχνευσης μεταβολής προσεγγίζει το θεωρητικό ελάχιστο, εάν η τιμή μέσου χρόνου μεταξύ ψευδών συναγερμών ανίχνευσης τείνει στο άπειρο.

2.6 Αυτοπαλίνδρομο μοντέλο πολλαπλών μεταβλητών

Ένας περιορισμός, ο οποίος εμφανίζεται στους αλγορίθμους που έχουν παρουσιαστεί μέχρι τώρα, είναι η επιβολή μίας μονόδρομης σχέσης, ότι η προς διερεύνηση μέτρηση ή μεταβλητή επηρεάζεται από τις εσωτερικές μεταβλητές του αλγορίθμου, αλλά όχι από τις μεταβλητές των αλγορίθμων άλλων μετρήσεων. Ωστόσο, σε ένα μεγάλο αριθμό εφαρμογών, είναι χρήσιμο να υπάρχει μία συσχέτιση ή επηρεασμός μεταξύ των διαφόρων μετρήσεων. Για παράδειγμα, στην περίπτωση δύο μετρήσεων, μία αμφίδρομη σχέση επηρεασμού μεταξύ τους ίσως να είναι κατάλληλη, όπως για παράδειγμα η μέτρηση θερμοκρασίας και η μέτρηση υγρασίας στην ατμόσφαιρα. Τέτοιες σχέσεις επηρεασμού είναι επιτρεπτές στο πλαίσιο της παλινδρόμησης πολλαπλών μεταβλητών. Σε ένα τέτοιο πλαίσιο, όλες οι μεταβλητές αντιμετωπίζονται συμμετρικά. Η μοντελοποίηση των μεταβλητών είναι τέτοια, ώστε να υπάρχει η υπόθεση ότι κάθε μεταβλητή επηρεάζει κάθε άλλη μεταβλητή ισοδύναμα. Με μαθηματική ορολογία, οι μεταβλητές θεωρούνται ενδογενείς. Σε αυτό το πλαίσιο, οι μετρήσεις ή μεταβλητές ορίζονται ως ys: η μεταβλητή $y_{1,t}$ είναι η τιμή της μεταβλητής y_1 τη χρονική στιγμή t, η μεταβλητή $y_{2,t}$ είναι η τιμή της μεταβλητής y_2 τη χρονική στιγμή t.

Ένα αυτοπαλίνδρομο μοντέλο πολλαπλών μεταβλητών [26] είναι μία γενίκευση του μονομεταβλητού αυτοπαλίνδρομου μοντέλου για την πρόβλεψη ενός συνόλου μετρήσεων. Το σύνολο μετρήσεων μπορεί επίσης να εκφραστεί και ως ένα διάνυσμα στο χρόνο. Αποτελείται από μία εξίσωση για κάθε μεταβλητή του συστήματος, όπου κάθε εξίσωση αποτελείται από μία σταθερά και από προηγούμενα βήματα όλων των μεταβλητών του συστήματος. Εάν λάβουμε υπόψη ένα σύστημα παλινδρόμησης