

ΕΚΤΕΝΗΣ ΠΕΡΙΛΗΨΗ

Συνδυασμός μεθόδων Μηχανικής Μάθησης και Επέκτασης Δεδομένων για αναγνώριση απευθυγράμμισης σε έδρανα ολίσθησης

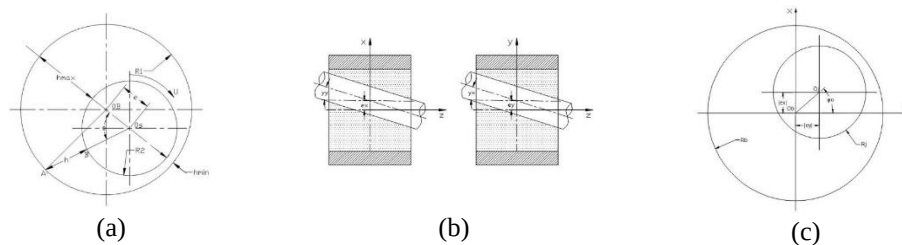
Κωνσταντίνος Αρβανίτης

Ένα από τα κυριότερα προβλήματα που αντιμετωπίζουν τα έδρανα ολίσθησης είναι η ατελής ευθυγράμμιση του άξονα από το έδρανο. Τα επακόλουθα της ατελούς ευθυγράμμισης είναι η μείωση της ικανότητας παραλαβής φορτίου, η φθορά του άξονα ή του εδράνου, η αύξηση των απωλειών ενέργειας, καθώς και η δημιουργία ταλαντώσεων που μπορεί να οδηγήσει σε διάδοση ρωγμών, φαινόμενα συντονισμού, εναλλασσόμενων φορτίων (κόπωσης) και θορύβου. Ως εκ τούτου, η μελέτη του προβλήματος αυτού είναι κρίσιμη, και είναι απαραίτητο να βρεθούν τρόποι για τον περιορισμό του. Η χρησιμοποίηση εργαλείων Τεχνητής Νοημοσύνης, όπως η Μηχανική Μάθηση σε προβλήματα που πραγματεύεται η επιστήμη του μηχανικού ολοένα και αυξάνεται καθώς παρουσιάζουν σημαντικά πλεονεκτήματα σε σχέση με τις κλασικές υπολογιστικές μεθόδους. Σκοπός της παρούσας εργασίας είναι η ανάδειξη των πλεονεκτημάτων αυτών στο πρόβλημα ενός ατελούς ευθυγραμμισμένου εδράνου, συγκεκριμένα στην περίπτωση περιορισμένου αριθμού πειραματικών δεδομένων. Η συλλογή δεδομένων από μια πειραματική διάταξη είναι μια χρονοβόρα και μεγάλου κόστους διαδικασία, οπότε η ανάπτυξη μιας αξιόπιστης μεθόδου αναγνώρισης της ατελούς ευθυγράμμισης σε περίπτωση περιορισμένων δεδομένων καθίσταται αναγκαία. Η ανάπτυξη μεθόδων Επέκτασης Δεδομένων αποτελεί έναν αναπτυσσόμενο τομέα έρευνας που παρέχει ικανοποιητικές λύσεις σε τέτοιου είδους προβλήματα. Στην παρούσα

εργασία, τα πειραματικά δεδομένα από ένα έδρανο σε ατελή ευθυγράμμιση παρουσιάζει δύο βασικά ζητήματα που οδηγούν σε κακή απόδοση ή και σε overfitting των αλγορίθμων Μηχανικής Μάθησης: μεγάλη ανισότητα μεταξύ των περιπτώσεων ευθυγράμμισης και ατελούς ευθυγράμμισης, και περιορισμένο αριθμό πειραματικών δεδομένων. Για το πρώτο πρόβλημα υιοθετείται μια μέθοδος υπερδειγματοληψίας που ονομάζεται SMOTE-LOF, ενώ λύση στο δεύτερο ζήτημα δίνει η χρήση ενός Παραγωγικού Ανταγωνιστικού Δικτύου (GAN). Στην συνέχεια, στο νέο σύνολο δεδομένων εφαρμόζονται πέντε (5) αλγόριθμοι Μηχανικής Μάθησης: Logistic Regression, K-Nearest Neighbors, Support Vector Machines, Decision Tree και Random Forest.

Ατελής Ευθυγράμμιση Εδράνου Ολίσθησης

Σε ένα έδρανο ολίσθησης σε υδροδυναμική λίπανση, η κύρια εξίσωση που χρησιμοποιείται είναι αυτή του Reynolds. Η γεωμετρία του συστήματος άξονα-εδράνου για την περίπτωση που δεν υπάρχει ατελής ευθυγράμμιση φαίνεται στην Εικόνα 1a, ενώ στις εικόνες 1b και 1c παρουσιάζονται οι τομές των επιπέδων στην περίπτωση ατελούς ευθυγράμμισης.

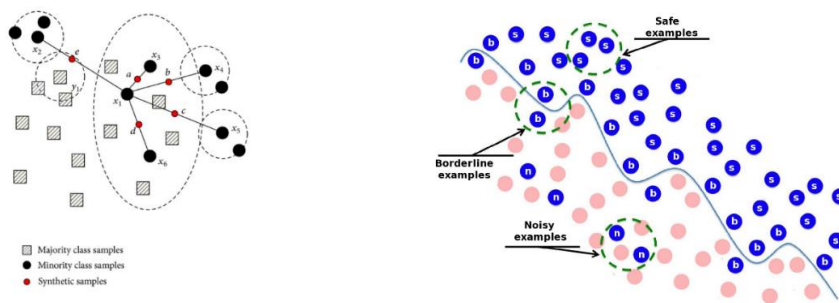


Εικόνα 1: Γεωμετρία εδράνου σε ατελή και μη ευθυγράμμιση.

Το πάχος του φιλμ του λιπαντικού αποτελεί μια καθοριστική παράμετρο των εδράνων σε μη ατελή ευθυγράμμιση. Το ελάχιστο πάχος του φιλμ του λιπαντικού, περιγράφεται από τις παρακάτω σχέσεις για τα δύο άκρα του εδράνου

SMOTE-LOF

Σε περιορισμένα σύνολα δεδομένων, είναι σύνηθες να παρουσιάζεται μια μεγάλη ανισότητα μεταξύ των τάξεων που ανήκουν σε αυτά. Λύση σε αυτό το πρόβλημα δίνουν οι μέθοδοι υπερδειγματοληψίας, οι οποίες παράγουν συνθετικά δεδομένα από την μειονοτική τάξη, ώστε να υπάρξει ισορροπία στο τελικό σύνολο δεδομένων. Η Τεχνική Υπερδειγματοληψίας Συνθετικής Μειονότητας (SMOTE) χρησιμοποιείται κατά κόρον σε τέτοιου είδους προβλήματα. Βασίζεται στην βασική ιδέα του αλγορίθμου Μηχανικής Μάθησης k Nearest Neighbor (kNN) και παράγει νέα δεδομένα της μειονοτικής τάξης πάνω στο τμήμα που συνδέει κάθε σημείο της τάξης αυτής με τους αντίστοιχους ‘γείτονές’ της.

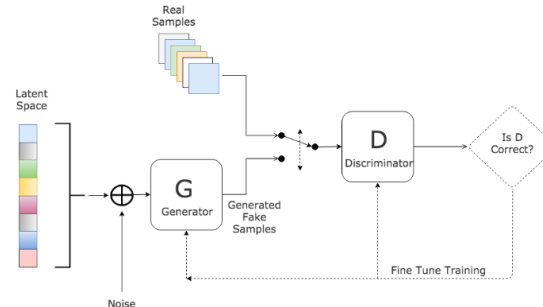


Εικόνα 2: Παραγωγή συνθετικών δεδομένων της μειονοτικής τάξης με την μέθοδο SMOTE και παραγόμενα δεδομένα της μειονεκτικής τάξης που χαρακτηρίζονται ως θόρυβος.

Ωστόσο, η μέθοδος SMOTE μπορεί να παράγει κάποια συνθετικά δεδομένα τα οποία να χαρακτηρίζονται ως θόρυβος και να μειώνουν σημαντικά την ποιότητα του τελικού συνόλου δεδομένων. Για τον λόγο αυτό, χρησιμοποιείται μια βελτιωμένη μορφή της που εμπεριέχει τον παράγοντα LOF (Local Outlier Factor), ο οποίος εξαλείφει τα δεδομένα που χαρακτηρίζονται ως θόρυβος, ή αλλιώς outliers. Σημειώνεται πως ως θόρυβος αναγνωρίζονται μόνο σημεία που παράγονται από την μέθοδο SMOTE, και όχι από το αρχικό σύνολο δεδομένων.

CTGAN

Τα Παραγωγικά Ανταγωνιστικά Δίκτυα (GANs) αποτελούν μια από τις βασικότερες μεθόδους επέκτασης δεδομένων. Ένα συγκεκριμένο είδος τέτοιων δικτύων αποτελούν τα Conditional Tabular GANs, τα οποία παράγουν με επιτυχία αληθοφανή δεδομένα σε μορφή πινάκων. Η βασική αρχή των GAN βασίζεται στον ανταγωνισμό δύο (2) νευρωνικών δικτύων: του Generator και του Discriminator. Ο Generator προσπαθεί να παράγει όσο πιο αληθοφανή συνθετικά δεδομένα γίνεται, ενώ ο Discriminator ξεχωρίζει τα πραγματικά από τα συνθετικά δεδομένα. Η εκτίμηση της απόδοσης ενός Παραγωγικού Ανταγωνιστικού Δικτύου αποτελεί μια περίπλοκη διαδικασία. Λύση στο πρόβλημα αυτό δίνει μια έννοια της θεωρίας των παιγνίων που ονομάζεται ισορροπία Nash, κατά την οποία δύο παίκτες ενός υποθετικού ανταγωνιστικού παιχνιδιού δεν μπορούν να βελτιώσουν περαιτέρω την στρατηγική τους.



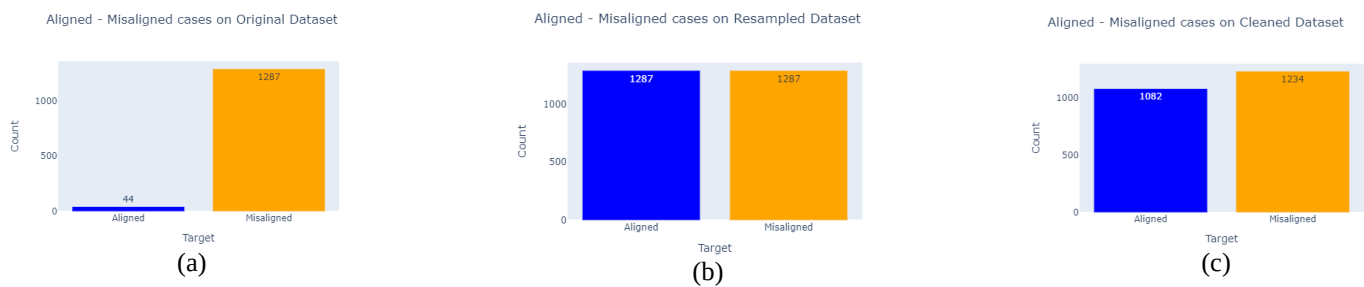
Εικόνα 3: Δομή ενός Παραγωγικού Ανταγωνιστικού Δικτύου.

Η βασική διαφορά ενός CTGAN είναι η ειδική κανονικοποίηση που εφαρμόζει στα συνεχή δεδομένα τύπου πίνακα, καθώς και ο Conditional Generator που λαμβάνει υπόψη τα δεδομένα που ανήκουν σε κατηγορίες (ηλικία, φύλο, κ.λπ.).

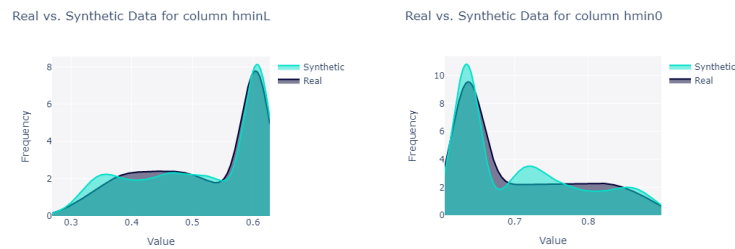
Αποτελέσματα

Η εργασία αυτή οδήγησε σε αρκετά σημαντικά συμπεράσματα σχετικά με την εφαρμογή μεθόδων επέκτασης δεδομένων στο πρόβλημα της ατελούς ευθυγράμμισης σε έδρανα ολίσθησης.

Αρχικά, μέσω της μεθόδου SMOTE εξαλείφεται η ανισότητα μεταξύ των δύο τάξεων του αρχικού συνόλου δεδομένων, ενώ ο παράγοντας LOF ‘καθαρίζει’ το νέο σύνολο δεδομένων από 258 σημεία που θεωρούνται θόρυβος. Στην συνέχεια, ένα CTGAN παράγει 3000, 4000, 5000 και 8000 νέα συνθετικά δεδομένα με στόχο την εύρεση του ποιοτικότερου μεταξύ των νέων ‘επεκταμένων’ συνόλων δεδομένων. Τέλος, σε αυτά τα σύνολα εφαρμόζονται οι αλγόριθμοι Μηχανικής Μάθησης, ώστε να βρεθεί ποιο έχει την βέλτιστη απόδοση.



Εικόνα 4: Εφαρμογή SMOTE-LOF για εξάλειψη της ανισότητας στο αρχικό σύνολο δεδομένων



Εικόνα 5: Πραγματικά και συνθετικά δεδομένα για παραγωγή 5000 συνθετικών σημείων.

ΠΙΝΑΚΑΣ ΜΕ ΕΠΕΞΗΓΗΣΕΙΣ ΤΗΣ ΟΡΟΛΟΓΙΑΣ

Αγγλικά	Ελληνικά
Journal Bearing	Έδρανο Ολίσθησης
Misalignment	Ατελής Ευθυγράμμιση
Shaft	Άξονας
Artificial Intelligence	Τεχνητή Νοημοσύνη
Backpropagation	Οπισθοδιάδοση
Machine Learning	Μηχανική Μάθηση
Oversampling	Υπερδειγματοληψία
Synthetic Minority Oversampling Technique	Τεχνική Υπερδειγματοληψίας Συνθετικής Μειονότητας
Data Augmentation	Επέκταση Δεδομένων
Generative Adversarial Network	Παραγωγικό Ανταγωνιστικό Δίκτυο