**ΠΑΝΕΠΙΣΤΗΜΙΟ ΚΡΗΤΗΣ**

**ΤΜΗΜΑ ΕΠΙΣΤΗΜΗΣ ΥΠΟΛΟΓΙΣΤΩΝ**

**ΠΑΡΟΥΣΙΑΣΗ / ΕΞΕΤΑΣΗ ΜΕΤΑΠΤΥΧΙΑΚΗΣ ΕΡΓΑΣΙΑΣ**

**Μυρτάκης Νικόλαος**
**Μεταπτυχιακός Φοιτητής**

**Τμήμα Επιστήμης Υπολογιστών, Πανεπιστήμιο Κρήτης**

**Επόπτης Μεταπτυχιακής Εργασίας: Καθηγητής, Β. Χριστοφίδης**

**Δευτέρα , 19 Οκτωβρίου 2020 ,ώρα 12:30 π.μ.**

**Τηλεδιάσκεψη ( μέσω του συστήματος e:Presence), Τμήμα Επιστήμης Υπολογιστών, Πανεπιστήμιο Κρήτης**

**Διεύθυνση μετάδοσης (url):** http://video.ucnet.uoc.gr/live/show/321

**Κανάλι YouTube του Τμήματος**
https://www.youtube.com/channel/UC7uE3QiMTQjkrpByB_Gnt6Q/live

# " Ερμηνεύοντας Ανωμαλίες σε Δεδομένα: Από Περιγραφικές σε Προβλεπτικές Εξηγήσεις"

## Περίληψη

Σε πολλές εργασίες διερεύνησης δεδομένων, ακανόνιστα ή σπανίως εμφανιζόμενα μοτίβα που ονομάζονται ανωμαλίες (αποκλίνοντα ή πολύ διαφορετικά δεδομένα), είναι συχνά πιο ενδιαφέροντα από τα συνήθη μοτίβα. Για παράδειγμα, ακανόνιστα μοτίβα μπορεί να αναπαριστούν συστηματικά σφάλματα, απάτες σε τραπεζικές συναλλαγές,

παρεισφρήσεις δικτύων και συστημάτων ελέγχου ή άλλα ενδιαφέροντα φαινόμενα. Πολυάριθμοι αλγόριθμοι έχουν προταθεί για την ανίχνευση ανωμαλιών. Δυστυχώς, οι περισσότεροι ανιχνευτές χωρίς επίβλεψη δεν προσφέρουν κάποια εξήγηση σχετικά με το γιατί ένα δοσμένο δείγμα (καταγραφή) χαρακτηρίστηκε σαν ανωμαλία και ως εκ τούτου να διαγνωστούν οι αιτίες που προκλήθηκε.

Οι εξηγήσεις ανωμαλιών συχνά παίρνουν τη μορφή υποσυνόλων γνωρισμάτων, σημαντικά μειωμένης διάστασης σε σύγκριση με τον αρχικό χώρο γνωρισμάτων. Εξετάζοντας μόνο τα γνωρίσματα σε έναν επεξηγηματικό υπόχωρο, αρκεί ώστε να καθοριστεί εάν ένα δείγμα είναι ανωμαλία ή όχι σύμφωνα με έναν ανιχνευτή. Οι εξηγήσεις μπορούν να κατηγοριοποιηθούν στις εξής (i) περιγραφικές με την έννοια ότι εξηγούν μόνο τα δείγματα που εκπαιδεύτηκε ο ανιχνευτής και (ii) περιγραφικές οι οποίες γενικεύονται και σε απαρατήρητα δεδομένα. Σε αυτήν την εργασία, αποτιμούμε πειραματικά τους κύριες περιγραφικές μεθόδους εξήγησης που έχουν προταθεί στην βιβλιογραφία, καθώς επίσης εισάγουμε την πρώτη μέθοδο για προβλεπτική εξήγηση, εμπνευσμένη από πρόσφατες εξελίξεις στο πεδίο της Αυτοματοποιημένης Μηχανικής Μάθησης (AutoML).

Στο πρώτο κομμάτι αυτής της εργασίας, παρουσιάζουμε ένα διεξοδικό πλαίσιο αποτίμησης αλγορίθμων εξήγησης ανωμαλιών χωρίς επίβλεψη, τόσο για μεμονωμένες όσο και για ομάδες ανωμαλιών με στόχο την αποσαφήνιση διαφόρων αναπάντητων ερωτημάτων από την τρέχουσα βιβλιογραφία όπως: (α) Πόσο αποτελεσματικός είναι ο συνδυασμός οποιουδήποτε αλγόριθμου εξήγησης με έναν οποιονδήποτε ανιχνευτή? (β) Πώς επηρεάζεται η συμπεριφορά μιας αλληλουχίας ανίχνευσης και εξήγησης ανωμαλιών από τον αριθμό ή την συσχέτιση των γνωρισμάτων στα δεδομένα? (γ) Ποια είναι η ποιότητα μιας σύνοψης στην περίπτωση που οι ανωμαλίες εξηγούνται από υποχώρους διαφορετικών διαστάσεων? Ένα μεγάλο ελάττωμα των περιγραφικών μεθόδων εξήγησης, πηγάζει από το γεγονός ότι πρέπει να ξανα υπολογιστούν για κάθε νέα παρτίδα δεδομένων.

Για να καταπολεμήσουμε αυτόν τον περιορισμό, στο δεύτερο κομμάτι αυτής της εργασίας, παρουσιάζουμε τη σχεδίαση και την πειραματική αποτίμηση του PROTEUS (Πρωτέας), ενός συστήματος αυτοματοποιημένης μηχανικής μάθησης. Ο PROTEUS παράγει καθολικές, προβλεπτικές εξηγήσεις χρησιμοποιώντας ένα υποκατάστατο μοντέλο, ειδικά σχεδιασμένο για επιλογή γνωρισμάτων σε μη ισορροπημένα δεδομένα ώστε να προσεγγίσει με τον καλύτερο δυνατό τρόπο την επιφάνεια επιλογής οποιουδήποτε ανιχνευτή χωρίς επίβλεψη. Υπολογιστικά πειράματα επιβεβαιώνουν την αποτελεσματικότητα και συνέπεια του PROTEUS στην παραγωγή προβλεπτικών εξηγήσεων για διαφορετικές οικογένειες ανιχνευτών ανωμαλιών καθώς και την αξιοπιστία του στην εκτίμηση της προβλεπτικής επίδοσης σε απαρατήρητα δεδομένα.

**University of Crete**

**Computer Science Department**

**M.Sc. Thesis presentation / examination**

**Myrtakis Nikolaos**

**Master's Thesis Supervisor: Professor V. Christophides**

**Monday, 19 October 2020, 12:30 p.m**

**Teleconference (will use the e: Presence system), Computer Science Department, University of Crete**

**(url) :** http://video.ucnet.uoc.gr/live/show/321

**YouTube channel :**

https://www.youtube.com/channel/UC7uE3QiMTQjkrpByB_Gnt6Q/live

# "Interpreting Data Anomalies: From Descriptive to Predictive Explanations"

## Abstract

In many data exploratory tasks, abnormal and rarely occurring patterns called anomalies (outliers, novelties) are more interesting than the prevalent ones. For instance, they could represent systematic errors, frauds in bank transactions, intrusions in network and system monitoring or other interesting phenomena. Numerous algorithms have been proposed for detecting anomalies. Unfortunately, unsupervised detectors in general, do not explain why a given sample (record) was labelled as an anomaly and thus diagnose its root causes.

Anomaly explanations often take the form of feature subsets of significantly lower dimensionality compared to the original feature space. By examining only the features of an explaining subspace suffices to determine whether a sample is an anomaly or not according to a detector. Explanations can be categorized as (i) descriptive in the sense that they explain the samples used to train the detector and (ii) predictive that generalize to unseen data. In this thesis we experimentally evaluate the main descriptive explanation methods proposed in the literature, as well as, introduce the first predictive explanation method that is inspired by recent advances in Automated Machine Learning systems (AutoML).

In the first part of our thesis, we present a thorough evaluation framework of unsupervised explanation algorithms for individual and groups of anomalies aiming to uncover several missing insights from the literature such as: (a) Is it effective to combine any explanation algorithm with any off-the-shelf outlier detector? (b) How is the behavior of an outlier detection and explanation pipeline affected by the number or the correlation of features in a dataset? and (c) What is the quality of summaries in the presence of outliers explained by subspaces of different dimensionality? A major drawback of the descriptive explanation methods stems from the fact that they should be recomputed for every new batch of data.

To address this limitation, in the second part of our thesis, we present the design and experimental evaluation of the PROTEUS AutoML pipeline. PROTEUS produces global, predictive explanations using a surrogate model, specifically designed for feature selection on imbalanced datasets in order to best approximate the decision surface of any unsupervised detector. Computational experiments confirm the efficacy and robustness of PROTEUS to produce predictive explanations for different families of anomaly detectors as well as its reliability to estimate their predictive performance in unseen data.