

ΠΑΝΕΠΙΣΤΗΜΙΟ ΚΡΗΤΗΣ

ΤΜΗΜΑ ΕΠΙΣΤΗΜΗΣ ΥΠΟΛΟΓΙΣΤΩΝ

ΠΑΡΟΥΣΙΑΣΗ / ΕΞΕΤΑΣΗ ΜΕΤΑΠΤΥΧΙΑΚΗΣ ΕΡΓΑΣΙΑΣ

**Μεταπτυχιακός Φοιτητής
Μπουντρογιάννης Κωνσταντίνος**

**Τμήμα Επιστήμης Υπολογιστών, Πανεπιστήμιο Κρήτης
Επόπτης Μεταπτ. Εργασίας: Καθηγητής, Π. Τσακαλίδης**

Πέμπτη, 09/07/2020, 10:00 μ.μ.

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

Διεύθυνση μετάδοσης (url): <http://video.ucnet.uoc.gr/live/show/291>

Κανάλι YouTube του Τμήματος

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

**“Συμβολικές αναπαραστάσεις βάσει δεδομένων για ανάλυση
χρονοσειρών σε υψηλό επίπεδο”**

Περίληψη

Η συστηματική συλλογή δεδομένων είναι πλέον μια εγγενής διαδικασία όλων των πτυχών της σύγχρονης ζωής. Από βιομηχανικά μηχανήματα έως μηχανήματα υγειονομικής περίθαλψης και φορητούς αισθητήρες, μια άνευ προηγουμένου ποσότητα δεδομένων διατίθεται για εξόρυξη και ανάκτηση πληροφοριών. Ο συνεχώς αυξανόμενος όγκος και η πολυπλοκότητα των δεδομένων χρονοσειρών απαιτούν αποτελεσματική μείωση των διαστάσεων των δεδομένων για τη διευκόλυνση των εργασιών εξόρυξης δεδομένων. Οι συμβολικές αναπαραστάσεις, ειδικότερα η

οικογένεια των συμβολικών συναθροιστικών προσεγγίσεων (SAX), έχουν αποδειχθεί πολύ αποτελεσματικές για τη συμπίεση της πληροφορίας που περιέχεται στις χρονοσειρές, ενώ εκμεταλλεύονται τον πλούτο των αλγορίθμων αναζήτησης που χρησιμοποιούνται στις κοινότητες της βιοπληροφορικής και της εξόρυξης κειμένου. Ωστόσο, οι τυπικές τεχνικές που βασίζονται στην SAX υποθέτουν ότι τα υποκείμενα στατιστικά χαρακτηριστικά των δεδομένων είναι Γκαουσιανά, με αποτέλεσμα συχνά να επιδεινώνεται η απόδοσή τους σε πρακτικές εφαρμογές. Για να ξεπεραστεί αυτός ο περιορισμός, η διατριβή αυτή παρουσιάζει μια μέθοδο που αναιρεί οποιαδήποτε υπόθεση σχετικά με την κατανομή πιθανότητας των χρονοσειρών, μέσω εκτίμησης πυκνότητας με πυρήνα (KDE) και Lloyd-Max κβάντισης. Η πειραματική αξιολόγηση σε πραγματικά δεδομένα καταδεικνύει την ανωτερότητα της προτεινόμενης μεθόδου, σε σύγκριση με τη συμβατική SAX και μια εναλλακτική μέθοδο βασιζόμενη στη SAX, που λειτουργεί με απευθείας προσαρμογή στα δεδομένα. Τέλος, στην παρούσα διατριβή, η προτεινόμενη μέθοδος μείωσης διαστάσεων αξιοποιείται για να παρέχει συμπαγείς αναπαραστάσεις χρονοσειρών με στόχο την ανίχνευση ανωμαλιών. Για το σκοπό αυτό, αναπτύσσεται ένα υπολογιστικά αποτελεσματικό, αλλά πολύ ακριβές, πλαίσιο για ανίχνευση ανωμαλιών σε ροές δεδομένων σε χώρους λιγότερων διαστάσεων, ενώ εναλλακτικά σχήματα κβαντισμού διερευνώνται και χρησιμοποιούνται για πιο ακριβή άντληση στατιστικών συμπερασμάτων.

Bountrogiannis Kostantinos

M.Sc. Thesis

Computer Science Department

University of Crete

Master's Thesis Supervisor: Professor, P.Tsakalides

Thursday, 09/07/2020, 10:00 p.m.

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

(url) : <http://video.ucnet.uoc.gr/live/show/291>

YouTube channel του Τμήματος

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

“Data-driven symbolic representations for high-level time series analysis”

Abstract

The systematic collection of data has become an intrinsic process of all aspects in modern life. From industrial to healthcare machines and wearable sensors, an unprecedented amount of data is becoming available for mining and information retrieval. The ever-increasing volume and complexity of time series data necessitate efficient dimensionality reduction for facilitating data mining tasks. Symbolic representations, especially the family of symbolic aggregate approximations (SAX), have proven very effective in compacting the information content of time series while exploiting the wealth of search algorithms used in bioinformatics and text mining communities. However, typical SAX-based techniques rely on a Gaussian assumption for the underlying data statistics, which often deteriorates their performance in practical scenarios. To overcome this limitation, this thesis introduces a method that negates any assumption on the probability distribution of time series, by means of kernel density estimation (KDE) and Lloyd-Max quantization. Experimental evaluation on real-world datasets demonstrates the superiority of the proposed method, when compared against the conventional SAX and an alternative data-adaptive SAX-based method. Finally, in the present thesis, the proposed dimensionality reduction method is utilized to provide compact representations of time series for the purposes of anomaly detection. To this end, a computationally efficient, yet highly accurate, framework for anomaly detection of streaming data in lower-dimensional spaces is developed, whereas alternative quantization schemes are explored and utilized for more accurate statistical inference.