

ΠΑΝΕΠΙΣΤΗΜΙΟ ΚΡΗΤΗΣ

ΤΜΗΜΑ ΕΠΙΣΤΗΜΗΣ ΥΠΟΛΟΓΙΣΤΩΝ

ΠΑΡΟΥΣΙΑΣΗ / ΕΞΕΤΑΣΗ ΜΕΤΑΠΤΥΧΙΑΚΗΣ ΕΡΓΑΣΙΑΣ

**Τρουλλινού Ειρήνη
Μεταπτυχιακή Φοιτήτρια**

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

Τετάρτη, 28/02/2018, 11:00

Αίθουσα Β108, Τμήμα Επιστήμης Υπολογιστών, Πανεπιστήμιο Κρήτης

**“Ανάλυση και Μοντελοποίηση της Αυθόρμητης Νευρωνικής Δραστηριότητας με χρήση
Συστημάτων Εκμάθησης Λεξικών”**

ΠΕΡΙΛΗΨΗ

Ένα κεντρικό δόγμα του κλάδου των νευροεπιστημών είναι ότι οι εξαιρετικές υπολογιστικές ικανότητες του εγκεφάλου οφείλονται στις διασυνδέσεις που υπάρχουν ανάμεσα σε πληθυσμούς διαφόρων νευρώνων. Βρισκόμαστε σε μια εξαιρετική περίοδο στην ιστορία του κλάδου της νευροεπιστήμης, καθώς το πεδίο βιώνει γρήγορη ανάπτυξη αναφορικά με την πολυπλοκότητα και τη ποσότητα της νευρωνικής δραστηριότητας που καταγράφεται. Αυτό οφείλεται στη πρόοδο που έχει γίνει στον πειραματικό σχεδιασμό, στις τεχνικές μέτρησης αλλά και στις υπολογιστικές δυνατότητες που υπάρχουν σήμερα, τα οποία μας επιτρέπουν πρωτοφανή πρόσβαση στην ανάλυση της νευρωνικής δραστηριότητας σε διάφορες περιοχές του εγκεφάλου. Ένας από τους στόχους της νευροεπιστήμης είναι η εύρεση ερμηνεύσιμων περιγραφών σχετικά με το τι αναπαριστά και υπολογίζει ο εγκέφαλος αλλά και η ερμηνεία πολύπλοκων φαινομένων σε απλή ορολογία. Θεωρώντας αυτό ως ένα πρόβλημα μείωσης διάστασης χώρου μας επιτρέπει να ανακαλύπτουμε αναπαραστάσεις απευθείας από τα πειραματικά δεδομένα, το οποίο είναι ένα βήμα κλειδί για την ανάπτυξη κατανοητών μοντέλων σχετικά με τη λειτουργία του εγκεφάλου. Οι μέθοδοι που χρησιμοποιούνται για τη μείωση διάστασης του χώρου παράγουν από δεδομένα υψηλής διάστασης αναπαραστάσεις με χαμηλότερη διάσταση, οι οποίες διατηρούν και δίνουν έμφαση σε ενδιαφέροντα χαρακτηριστικά που υπάρχουν στα δεδομένα μας.

Σε αυτή την εργασία χρησιμοποιούμε δύο διαφορετικά σύνολα δεδομένων με πραγματικές δυαδικές μετρήσεις, τα οποία αναφέρονται στην αυθόρμητη νευρωνική δραστηριότητα δύο εργαστηριακών ποντικών, την οποία εμείς στοχεύουμε να αναπαραστήσουμε αποδοτικά σε χαμηλότερη διάσταση. Τα πραγματικά δεδομένα, σε αντίθεση με τα συνθετικά, δεν είναι τόσο δομημένα και η ύπαρξη θορύβου είναι πιο έντονη. Ο θόρυβος αυτός μπορεί να οφείλεται είτε σε λάθη κατά την δημιουργία ή την επεξεργασία των δεδομένων, είτε σε τυχαία ενεργοποίηση κάποιων νευρώνων. Επομένως, για να μπορέσουμε να αποσπάσουμε χρήσιμες πληροφορίες για το πως συνδέονται οι νευρώνες μεταξύ τους, θα πρέπει να μπορούμε να διαχωρίζουμε τα πραγματικά από τα θορυβώδη πρότυπα ενεργοποίησης. Για να μπορέσουμε επομένως να αντιμετωπίσουμε και τις δύο αυτές προκλήσεις, δηλαδή την αποδοτική αναπαράσταση των δεδομένων σε χαμηλότερη διάσταση αλλά και το διαχωρισμό ανάμεσα σε πραγματικά και θορυβώδη πρότυπα, συνδυάζουμε τεχνικές εκμάθησης λεξικών και μηχανικής μάθησης με επίβλεψη. Πιο συγκεκριμένα προτείνουμε έναν αλγόριθμο για εκμάθηση λεξικών, ο οποίος διαδοχικά διαλέγει τα στοιχεία από το σύνολο των δεδομένων που διαθέτουμε, και γεμίζει το λεξικό, δηλαδή το νέο χώρο μειωμένης διάστασης, μόνο με εκείνα τα στοιχεία τα οποία συνεισφέρουν στη καλύτερη αναπαράσταση των πραγματικών προτύπων ενεργοποίησης από ότι εκείνων με θόρυβο. Η απόφαση για την είσοδο ενός στοιχείου στο λεξικό που βασίζεται στη χρήση πραγματικών ενθόρυβων προτύπων ενεργοποίησης δικαιολογεί το χαρακτηρισμό "ανταγωνιστικά" που έχουμε δώσει για να περιγράψουμε τον αλγόριθμο μας. Αυτή η μέθοδος αναζητά επαναλαμβανόμενα πρότυπα, για τα οποία υπάρχει μόνο απόλυτος συγχρονισμός διέγερσης ανάμεσα στους νευρώνες. Θεωρούμε και μια πιο χαλαρή προσέγγιση αυτού του αλγορίθμου, βάσει της οποίας μπορούμε να ανακαλύπτουμε πρότυπα που έχουν και χρονική συσχέτιση μεταξύ τους μέσα σε ένα μεγαλύτερο χρονικό παράθυρο. Τέλος, χρησιμοποιούμε ένα ταξινομητή, ο οποίος εκπαιδεύεται με τα ανακατασκευασμένα πραγματικά και ενθόρυβα πρότυπα ενεργοποίησης.

Ο αλγόριθμος που προτείνουμε δημιουργεί ένα λεξικό, το οποίο χρησιμοποιείται για να παράγει τα πραγματικά, θορυβώδη ανακατασκευασμένα πρότυπα, τα οποία στη συνέχεια δίνονται ως είσοδος στο ταξινομητή. Πειραματικά αποτελέσματα δείχνουν ότι αυτός πετυχαίνει ακρίβεια ταξινόμησης 60% στην περίπτωση της απόλυτης συγχρονισμένης δραστηριότητας και 90% όταν αναζητάμε πρότυπα σε μεγαλύτερα χρονικά παράθυρα. Συγκριτικά, όταν ο ταξινομητής εκπαιδεύεται με τα αρχικά δεδομένα επιτυγχάνει ακρίβεια ταξινόμησης 51%. Επίσης, μέσω πειραματικών αποτελεσμάτων δείχνουμε τη καλύτερη επίδοση του αλγορίθμου μας τόσο ποσοτικά, όσο και ποιοτικά συγκρίνοντας τον με τον K-SVD, ένα καθιερωμένο αλγόριθμο για εκμάθηση λεξικών.

Troullinou Eirini

M.Sc. Thesis

Computer Science Department

University of Crete

Master's Thesis Supervisor: Professor Panagiotis Tsakalides

Wednesday, 28/02/2018, 11:00

Room B108, Computer Science Dept., University of Crete

“Analysis and Modeling of Spontaneous Neuronal Activity using Dictionary Learning Systems”

Abstract

A central tenet of neuroscience is the remarkable computational abilities of our brains that arise as a result of populations of interconnected neurons. Indeed, we find ourselves at an exciting moment in the history of neuroscience, as the field is experiencing rapid growth in the complexity and quantity of the recorded neural activity, as advances in experimental design, measurement techniques, and computational analysis allow us unprecedented access to the dynamics of neural activity in different brain areas. Thus, one of the goals of neuroscience is to find interpretable descriptions of what the brain represents and computes, and also to explain complex phenomena in simple terms. Considering this task from the perspective of dimensionality reduction provides an entry point into principled mathematical techniques that allows us to discover these representations directly from experimental data, a key step to developing rich yet comprehensible models for brain function. Dimensionality reduction methods produce low-dimensional representations of high-dimensional data, where the representation is chosen to preserve or highlight some feature of interest in the data.

In this master thesis, we employ two real binary datasets that refer to the spontaneous neuronal activity of two laboratory mice over time, and we aim to their efficient low-dimensional representation. Real datasets compared to synthetic ones are not so highly structured and background noise is more intense. Noise could be a result of mistakes during the creation or the processing of the datasets or it could also exist due to accidental firings produced by neurons. So, in order to get insights regarding how neurons are connected to each other, we also need to be able to discriminate the true from the noisy activation patterns. In order to address both challenges, namely the low-dimensional representation of the data and the discrimination between true and spurious activation patterns, we combine dimensionality reduction techniques with supervised machine learning. More specifically, we propose a Sequential Adversarial Dictionary Learning Algorithm, which selects sequentially the elements that are included in the dataset and fills the dictionary, namely the new reduced space, only with those elements that contribute to the better representation of the true, rather than the noisy activation patterns, which have been synthetically created. The entry of an element in the dictionary, which is based on the use of true and noisy activation patterns, justifies the name "Adversarial", which is used in the title of our algorithm. This method searches only for repeated patterns with total synchronous firing activity. Thus, we also consider the idea of a more relaxed approach, where we can discover patterns that have also some temporal correlation within a time window interval. Subsequently, a supervised classifier is used, which takes as input the reconstructed signals in order to discriminate the true from the noisy activation patterns.

Experimental results show that our algorithm creates a dictionary which, when used to produce the reconstructed patterns given to the classifier, it results to a classification accuracy of 60% in the case of synchronous firing activity and 90% when we search for patterns in bigger time window intervals. By comparison, the classifier achieves a classification rate of only 51% when raw data is used as input. We also demonstrate that our system achieves better results both quantitatively as well as qualitatively when compared with K-SVD, an established dictionary learning algorithm.