Τίτλος Διδακτορικής Διατριβής

Ανάπτυξη πρωτότυπων αλγορίθμων Μηχανικής Μάθησης για χρήση σε εκπαιδευτικά δεδομένα και σε Συστήματα Διαχείρισης Εκπαιδευτικού Περιεχομένου.

Ονοματεπώνυμο

Γεώργιος Κωστόπουλος

Τριμελής Συμβουλευτική Επιτροπή

Ράγγος Όμηρος (Επίκουρος Καθηγητής, Επιβλέπων)

Κωτσιαντής Σωτήρης (Επίκουρος καθηγητής)

Τσέλιος Νικόλαος (Αναπληρωτής Καθηγητής)

Περίληψη

Η αξιοποίηση των τεχνολογιών της πληροφορίας και των επικοινωνιών στον χώρο της εκπαίδευσης συμβάλλει καθημερινά στην παραγωγή και αποθήκευση μεγάλων ποσοτήτων δεδομένων. Η ανάγκη αποτελεσματικής ανάλυσης αυτών των δεδομένων για την ανακάλυψη πολύτιμων πληροφοριών και τη μετατροπή τους σε συστηματική γνώση συντέλεσε στην ανάπτυξη των πεδίων της Εξόρυξης Γνώσης από Εκπαιδευτικά Δεδομένα και της Μαθησιακής Αναλυτικής. Η Εξόρυξη Γνώσης από Εκπαιδευτικά Δεδομένα επικεντρώνεται στην ανάπτυξη και εφαρμογή μεθόδων Εξόρυξης Γνώσης σε εκπαιδευτικά δεδομένα για την επίλυση σημαινόντων εκπαιδευτικών προβλημάτων, ενώ η Μαθησιακή Αναλυτική εστιάζει περισσότερο στη διαδικασία της μάθησης, αξιοποιώντας την ανάλυση δεδομένων για την ενίσχυση των διαδικασιών λήψης αποφάσεων. Ωστόσο, ανεξάρτητα από τον τρόπο προσέγγισης ενός εκπαιδευτικού προβλήματος, και τα δύο επιστημονικά πεδία έχουν κοινούς στόχους: τη βελτίωση της μάθησης και την αναβάθμιση της ποιότητας της προσφερόμενης εκπαίδευσης.

Η πρόβλεψη των μαθησιακών αποτελεσμάτων των εκπαιδευομένων συνιστά ένα από τα σημαντικότερα προβλήματα των πεδίων της Εξόρυξης Γνώσης από Εκπαιδευτικά Δεδομένα και της Μαθησιακής Αναλυτικής. Η αντιμετώπιση του συγκεκριμένου προβλήματος αφορά στη δημιουργία ενός μοντέλου κατηγοριοποίησης ή παλινδρόμησης, ανάλογα με τη φύση του χαρακτηριστικού πρόβλεψης, εφαρμόζοντας έναν κατάλληλα επιλεγμένο αλγόριθμο Επιβλεπόμενης Μηχανικής Μάθησης, όπως είναι, για παράδειγμα, ένα δέντρο απόφασης. Ωστόσο, η δημιουργία ενός προβλεπτικού μοντέλου προϋποθέτει την εκπαίδευση του αλγορίθμου σε ένα σύνολο ετικετοποιημένων δεδομένων, δηλαδή ένα σύνολο για το οποίο είναι γνωστές τόσο οι τιμές των ανεξάρτητων μεταβλητών, όσο και η τιμή της μεταβλητής απόφασης. Η δυσκολία συλλογής ετικετοποιημένων δεδομένων έχει οδηγήσει στην ανάπτυξη νέων μεθόδων Μηχανικής Μάθησης οι οποίες χαρακτηρίζονται, γενικά, με τον όρο «Μηχανική Μάθηση με Ελλιπή Επίβλεψη». Η Ημι-Επιβλεπόμενη Μηχανική Μάθηση και η Ενεργή Μηχανική Μάθηση αποτελούν τις κύριες συνιστώσες αυτού του ραγδαία αναπτυσσόμενου πεδίου, στοχεύοντας στη βέλτιστη αξιοποίηση ετικετοποιημένων και μη ετικετοποιημένων δεδομένων για τη δημιουργία αποτελεσματικών και εύρωστων μοντέλων Μηχανικής Μάθησης.

Τα τελευταία χρόνια, αρκετοί αλγόριθμοι Μηχανικής Μάθησης με Ελλιπή Επίβλεψη έχουν αναπτυχθεί και εφαρμοστεί με μεγάλη επιτυχία για την επίλυση διάφορων προβλημάτων σε πολλά επιστημονικά πεδία. Ωστόσο, η αποτελεσματικότητα αυτών των μεθόδων δεν έχει μελετηθεί στο πεδίο της εκπαίδευσης, όπως συνάγεται από την ανασκόπηση της σχετικής βιβλιογραφίας, δημιουργώντας νέες προκλήσεις για επιστήμονες και ερευνητές του χώρου. Οι προκλήσεις αυτές δεν αφορούν μόνο στην εφαρμογή υφιστάμενων μεθόδων Μηχανικής Μάθησης με Ελλιπή Επίβλεψη στα πεδία της Εξόρυξης Γνώσης από Εκπαιδευτικά Δεδομένα και της Μαθησιακής Αναλυτικής, αλλά και στην ανάπτυξη νέων αλγορίθμων για την διύλιση πολύτιμης γνώσης από τεράστιες ποσότητες εκπαιδευτικών δεδομένων.

Σε αυτό το πλαίσιο, ο πρωταρχικός σκοπός της παρούσας διατριβής είναι η ανάπτυξη νέων μεθόδων Μηχανικής Μάθησης με Ελλιπή Επίβλεψη και η εφαρμογή τους στην πρόβλεψη των μαθησιακών αποτελεσμάτων των σπουδαστών σε διάφορες βαθμίδες της εκπαίδευσης. Πιο συγκεκριμένα, αναπτύσσουμε έναν αλγόριθμο Ημι-Επιβλεπόμενης Κατηγοριοποίησης με Συνεκπαίδευση και έναν αλγόριθμο Ημι-Επιβλεπόμενης Παλινδρόμησης για την πρόβλεψη της απόδοσης και του βαθμού, αντίστοιχα, προπτυχιακών φοιτητών στην εξ αποστάσεως τριτοβάθμια εκπαίδευση. Οι προτεινόμενες μέθοδοι κρίνονται κατάλληλες τόσο για την αποτελεσματική, όσο και την έγκαιρη πρόβλεψη των μαθησιακών αποτελεσμάτων των σπουδαστών, όπως τεκμαίρεται από τα πειραματικά αποτελέσματα των δημοσιευμένων εργασιών μας.

Ολοκληρώνοντας, θεωρούμε ότι η παρούσα διατριβή αποτελεί την πρώτη συστηματική και ολοκληρωμένη προσπάθεια για την αξιοποίηση της Μηχανικής Μάθησης με Ελλιπή Επίβλεψη στον χώρο της εκπαίδευσης, προσδοκώντας σαφώς καλύτερα αποτελέσματα από τις παραδοσιακές μεθόδους Επιβλεπόμενης Μάθησης.

Title of Thesis

Development of new Machine Learning algorithms for use in educational data and Educational Content Management Systems.

Full Name

Georgios Kostopoulos

Three-member Advisor Committee

Omiros Ragos (Assistant Professor, Supervisor)

Sotiris Kotsiantis (Assistant Professor, Member of the Advisor Committee)

Tselios Nikolaos (Associate Professor, Member of the Advisor Committee)

Abstract

The use of information and communication technologies in the educational field contributes daily to the production and storage of large amounts of data. The need for an effective analysis of these data for retrieving valuable information and their converting into systematic knowledge resulted in the genesis of the fields of Educational Data Mining and Learning Analytics. Educational Data Mining focuses primarily on the development and implementation of data mining methods on

educational data for solving important educational problems, while Learning Analytics is centered mainly around the learning process, exploiting data analysis in order to reinforce the decision-making procedures. However, irrespectively of the method used to approach an educational problem, both scientific fields have common goals; the improvement of learning and enhancing the quality of education offered.

The prognosis of students' learning outcomes constitutes one of the most significant problems in the fields of Educational Data Mining and Learning Analytics. The handling of this specific problem relates to the creation of a classification or regression model, depending on the nature of the outcome variable, implementing a supervised learning algorithm, such as a decision tree. However, the building of a predictive learning model presupposes the training of the algorithm in a set of labeled data. The difficulty of acquiring labeled data has resulted to the development of new machine learning methods which are generally referred to as "Weakly Supervised Learning". Semi-Supervised learning and Active Learning constitute the main components of Weakly Supervised Learning with a view to exploiting a small number of labeled examples together with a large number of unlabeled ones in the best possible manner, for building accurate and robust learning models.

In recent years, a plethora of Semi-Supervised Learning algorithms have been developed and implemented with great success for solving a variety of problems in many scientific fields. However, the effectiveness of these methods has not been thoroughly studied in the educational field, as it is evident from the pertinent literature, generating, thus, new challenges for scientists and researchers. These challenges relate not only to the implementation of existing Semi-Supervised Learning algorithms but also to the development of innovative algorithms for extracting valuable knowledge from various sources of educational data.

Within this frame, the main purpose of the present thesis is the development of new Semi-Supervised Learning methods and their implementation for the prognosis of students' learning outcomes at different levels of education. More specifically, we develop a co-training style algorithm and a semi-supervised regression algorithm for predicting the academic performance and the grades of undergraduate students in distance higher education, respectively. The proposed methods are deemed to be very effective for the accurate and early prediction of students' learning outcomes, as confirmed by the experimental results of our published studies.

Rounding up, we consider that the present thesis constitutes the first systematic and comprehensive attempt towards exploiting the Semi-Supervised Learning framework in the educational field, anticipating better results than those generated from traditional supervised methods.

Link ανάρτησης της διατριβής

https://thalis.math.upatras.gr/~sotos/Phd%20Thesis-Georgios%20Kostopoulos.pdf