Data Science Tutor: Διαδικτυακή πλατφόρμα αυτόματης επίλυσης ασκήσεων στην Επιστήμη Δεδομένων

Χρήστος Τρυφωνόπουλος, Χάρης Κωτσιόπουλος, Παρασκευή Ραυτοπούλου
{trifon, kotsiopoulos, praftop}@uop.gr
Πανεπιστήμιο Πελοποννήσου, Τρίπολη, Ελλάδα

Περίληψη

Στην παρούσα εργασία παρουσιάζεται μία καινοτόμα, δωρεάν διαδικτυακή πλατφόρμα που δημιουργήθηκε για να υποστηρίξει τη μαθησιακή διαδικασία στο πεδίο της Επιστήμης Δεδομένων (Data Science) μέσω της αυτοματοποιημένης δημιουργίας και επίλυσης ασκήσεων. Η πλατφόρμα εστιάζει στην τριτοβάθμια εκπαίδευση και είναι ήδη σε χρήση εδώ και ένα χρόνο, με θερμή αποδοχή, από αρκετά πανεπιστημιακά τμήματα της Ελλάδας και του εξωτερικού. Στην εργασία αυτή παρουσιάζονται συνοπτικά η αρχιτεκτονική, η λειτουργικότητα και οι δυνατότητες της πλατφόρμας, καθώς και μία πρώτη αποτίμηση της χρήσης της.

Λέξεις κλειδιά: επιστήμη δεδομένων, επίλυση ασκήσεων, τριτοβάθμια εκπαίδευση

Εισαγωγή

Η έκρηξη στον όγκο των διαθέσιμων δεδομένων έδωσε το έναυσμα για τη δημιουργία ενός νέου διεπιστημονικού πεδίου, αυτού της Επιστήμης Δεδομένων (ΕΔ). Στο πλαίσιο αυτό, τα τελευταία χρόνια η βασική και εφαρμοσμένη έρευνα επικεντρώθηκε σε τρόπους διαχείρισης, ανάλυσης, εξόρυξης, και οπτικοποίησης της πληροφορίας που βρίσκεται πίσω από τα δεδομένα, με κύριο σκοπό την αξιοποίησή της στη λήψη αποφάσεων. Με γνώμονα τη ζήτηση που δημιουργήθηκε σε αυτό το νέο και ραγδαία αναπτυσσόμενο πεδίο, αρκετά πανεπιστημιακά τμήματα στην Ελλάδα και το εξωτερικό ενέταξαν στα (προπτυχιακά ή/και μεταπτυχιακά) προγράμματα σπουδών τους μαθήματα όπως μηχανική μάθηση, ανάκτηση πληροφοριών, ή εξόρυξη γνώσης, που σχετίζονται άμεσα με το πεδίο της ΕΔ. Υπάρχει επομένως ένας σημαντικός αριθμός εμπλεκόμενων στην εκπαιδευτική διαδικασία (καθηγητές, εργαστηριακοί βοηθοί, φοιτητές), οι οποίοι εκπαιδεύουν ή εκπαιδεύονται σε αλγορίθμους, μεθόδους και τεχνικές διαχείρισης δεδομένων. Την ίδια στιγμή όμως υπάρχει σημαντική έλλειψη σύγχρονων εκπαιδευτικών εργαλείων, τα οποία θα είναι σε θέση να επικουρήσουν τη διδακτική και μαθησιακή διαδικασία σε ένα τόσο σημαντικό και ραγδαία εξελισσόμενο πεδίο.

Σχετικές εργασίες. Παρά τις πρώτες ερευνητικές εργασίες για τον τρόπο διδασκαλίας της ΕΔ (Brunner & Kim, 2016; Hicks & Irizarry, 2016) δεν υπάρχουν προσεγγίσεις οι οποίες να εστιάζουν στη χρήση τεχνολογιών ηλεκτρονικής μάθησης. Επιπλέον, όσα συστήματα έχουν προταθεί για υποπεριοχές της ΕΔ, δίνουν έμφαση στη διδασκαλία μεθόδων αναζήτησης (Grivokostopoulou et al., 2017), στην παιγνιοποίηση αλγορίθμων (Halttunen & Sormunen, 2000), στην οργάνωση και διαχείριση της μελέτης (Nakayama et al., 2004; Willms, 2003), και στη συνεργασία των φοιτητών (Efthimiadis et al., 2011), αλλά όχι στην επίλυση ασκήσεων.

Σκοπός και στόχοι. Η παρούσα εργασία διαπιστώνει την έλλειψη ενός εκπαιδευτικού εργαλείου το οποίο θα μπορεί να επικουρήσει την εκπαιδευτική διαδικασία στο πεδίο της ΕΔ και προτείνει μία καινοτόμα, δωρεάν, διαδικτυακή πλατφόρμα (Data Science Tutor - DST) η οποία επικεντρώνεται στο κομμάτι της επίλυσης ασκήσεων που σχετίζονται με τη διδαχθείσα ύλη. Έτσι, επιτρέπει σε εκπαιδευτές/εκπαιδευόμενους να δημιουργήσουν δικές τους ασκήσεις

για μία πλειάδα αλγορίθμων και τεχνικών διαχείρισης δεδομένων και κατόπιν να δουν άμεσα την (με αυτόματο τρόπο παραγόμενη) λύση τους. Οι λύσεις των ασκήσεων παρουσιάζονται βήμα-βήμα, με επεξηγήσεις που παραπέμπουν στη λειτουργία του εκάστοτε αλγορίθμου, και είναι συμβατές με την ορολογία και τη μεθοδολογία επίλυσης ασκήσεων των κύριων διδακτικών συγγραμμάτων -ενδεικτικά (Aggarwal, 2015; Manning et al., 2008; Witten et al., 2011)- που χρησιμοποιούνται στα αντίστοιχα πανεπιστημιακά μαθήματα. Η προτεινόμενη πλατφόρμα απευθύνεται σε δύο κατηγορίες χρηστών, με διαφορετικούς στόχους για κάθε μία. Οι προπτυχιακοί/μεταπτυχιακοί φοιτητές στο πεδίο της ΕΔ, ή σε συναφή αντικείμενα όπως η μηχανική μάθηση, η ανάκτηση πληροφοριών, η διαχείριση δεδομένων, και η εξόρυξη γνώσης, μέσω της δημιουργίας δικών τους ασκήσεων, αλλά και της μελέτης της συνοδευόμενης λύσης έχουν τη δυνατότητα να εντρυφήσουν στις ιδιαιτερότητες κάθε αλγορίθμου, να εντοπίσουν περιπτώσεις που δεν καλύπτονται από το υπόλοιπο διδακτικό υλικό, αλλά και να μάθουν με αλληλεπιδραστικό τρόπο τη συλλογιστική και τη διαδικασία επίλυσης που ακολουθείται για διαφορετικούς τύπους ασκήσεων. Αντίστοιχα το διδακτικό προσωπικό ενός μαθήματος (εισηγητής ή επικουρικό προσωπικό) μπορεί μέσα από το προτεινόμενο σύστημα να παράξει με ευκολία μια μεγάλη ποικιλία ασκήσεων με τις συνοδευόμενες λύσεις, έτοιμες για χρήση στην τάξη ή ως επικουρικό υλικό για τη διόρθωση γραπτών δοκιμασιών.

Συνεισφορά. Συμπερασματικά, η προτεινόμενη πλατφόρμα DST είναι η μοναδική παγκοσμίως διαθέσιμη λύση για την αυτόματη διαδικτυακή επίλυση ασκήσεων σε μία ευρεία γκάμα θεματικών που σχετίζονται με την ΕΔ. Έχει ήδη τύχει θερμής υποδοχής από σημαντικά ιδρύματα της Ελλάδας και του εξωτερικού τα οποία τη χρησιμοποιούν για το εργαστηριακό/φροντιστηριακό μέρος των σχετικών μαθημάτων. Η πλατφόρμα DST είναι στην Αγγλική, και βρίσκεται σε χρήση εδώ και έναν περίπου χρόνο, ενώ εμπλουτίζεται συνεχώς με νέα λειτουργικότητα και αλγορίθμους.

Αρχιτεκτονική, δυνατότητες και αποδοχή

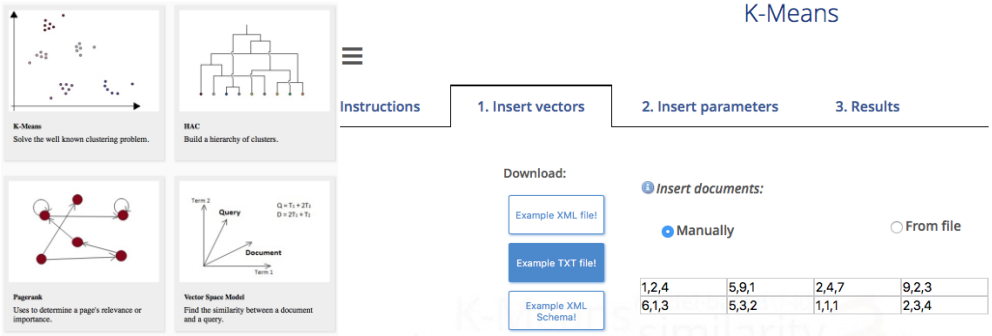
Στην ενότητα αυτή περιγράφουμε συνοπτικά την αρχιτεκτονική της πλατφόρμας DST, τις δυνατότητες και την παρεχόμενη λειτουργικότητα, καθώς και τα πρώτα συμπεράσματα που έχουν εξαχθεί από τη μέχρι τώρα χρήση της.

Αρχιτεκτονική και τεχνολογίες. Η πλατφόρμα DST έχει υλοποιηθεί με χρήση διαδεδομένων γλωσσών και εργαλείων, ενώ η αρχιτεκτονική της είναι διαστρωματωμένη με στόχο την εύκολη προσθήκη τεχνικών και αλγορίθμων. Είναι εξολοκλήρου βασισμένη σε php και javascript, πλήρως συμβατή με τα τελευταία πρότυπα των HTML και CSS, ενώ για την εμφάνιση γραφικών στοιχείων (π.χ., διαγράμματα ή γράφοι) χρησιμοποιείται το Chart API της Google. Στην παρούσα φάση η πλατφόρμα φιλοξενείται σε έναν τυπικό διακομιστή 64bit/Dual Core 2GHz/4GB RAM με Ubuntu Linux, έκδοση 16.04.3.

Δυνατότητες και λειτουργικότητα. Η πρόσβαση στην πλατφόρμα γίνεται μέσω οποιουδήποτε σύγχρονου φυλλομετρητή και είναι δωρεάν, ενώ για τη χρήση της δεν απαιτείται εγγραφή ή παροχή προσωπικών στοιχείων. Έτσι, αρκεί κάποιος να πληκτρολογήσει την ηλεκτρονική της διεύθυνση http://db.uop.gr/~ds-tutor και μεταφέρεται στο γραφικό περιβάλλον της εφαρμογής. Εκεί καλείται να επιλέξει μεταξύ διαφορετικών αλγορίθμων/μεθόδων, αυτόν για τον οποίο ενδιαφέρεται να υποβάλλει και να επιλύσει μία άσκηση (Εικόνα 1, αριστερά).

Στην παρούσα έκδοσή της, η πλατφόρμα DST παρέχει τη δυνατότητα επίλυσης διάφορων τύπων ασκήσεων για 12 αλγορίθμους με τις παραλλαγές τους, μεταξύ των οποίων αλγόριθμοι ανεπίβλεπτης μάθησης όπως ο K-means, δημοφιλή μοντέλα ανάκτησης πληροφορίας όπως το

Boolean και το Vector Space, αλγόριθμοι ιεραρχικής ομαδοποίησης όπως ο HAC, και μέθοδοι

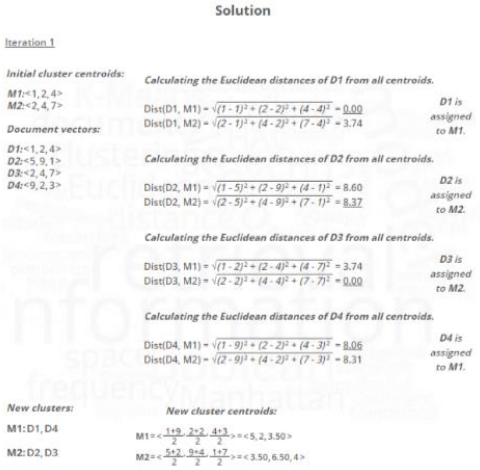


Εικόνα 1. Επιλογή αλγορίθμου στην αρχική οθόνη (αριστερά) και εισαγωγή παραμέτρων και δεδομένων(δεξιά)

ανάλυσης γράφων/υπερσυνδέσμων όπως οι PageRank και HITS. Το σύνολο της αλληλεπίδρασης με την πλατφόρμα καθώς και η επίλυση των ασκήσεων είναι στην Αγγλική.

Για όλους τους αλγόριθμους παρέχεται μια συνοπτική παρουσίαση της θεωρίας καθώς και βοήθεια για το είδος της εισόδου που απαιτείται για τη δημιουργία μίας άσκησης. Η διαδικασία εισόδου των απαραίτητων δεδομένων για κάθε αλγόριθμο γίνεται με χρήση ενός απλού και φιλικού βοηθού (wizard) ο οποίος καθοδηγεί το χρήστη στη συμπλήρωση των απαιτούμενων πεδίων. Αξίζει να σημειωθεί πως η πλατφόρμα DST απευθύνεται στην τριτοβάθμια εκπαίδευση και εστιάζει σε ένα ευρύ αλλά απαιτητικό αντικείμενο με αλγορίθμους που αρκετές φορές απαιτούν εξειδικευμένη γνώση ακόμη και για την είσοδο των δεδομένων. Με βάση τα παραπάνω, η χρήση ενός τέτοιου οδηγού κρίθηκε απαραίτητη έπειτα από την ανάλυση απαιτήσεων και τις προσωπικές συνεντεύξεις χρηστών.

Η είσοδος των δεδομένων γίνεται με δύο τρόπους: είτε με χρήση του online γραφικού περιβάλλοντος, είτε με μεταφόρτωση στο διακομιστή κατάλληλα μορφοποιημένου XML αρχείου το οποίο περιέχει όλες τις απαραίτητες παραμέτρους και εισόδους που θα επιτρέψουν τη δημιουργία και επίλυση της άσκησης. Ο διττός τρόπος εισαγωγής διευκολύνει τόσο τους αρχάριους ή περιστασιακούς χρήστες, όσο και τους πιο έμπειρους οι οποίοι συχνά απαιτούν μία εύκολη, γρήγορη, και αυτόματοποιημένη διαδικασία εισόδου δεδομένων. Για τους έμπειρους χρήστες είναι διαθέσιμο και το XML schema για την υποστήριξη της διαλειτουργικότητα με άλλες εφαρμογές. Στην Εικόνα 1 (δεξιά) φαίνεται το περιβάλλον εισαγωγής δεδομένων και παραμέτρων (Βήμα 1) για έναν από τους διαθέσιμους αλγόριθμους. Στην εικόνα αυτή είναι εμφανή τα εξής στοιχεία: ο βοηθός (wizard) με την αλληλουχία βημάτων, οι δυνατότητες εισαγωγής δεδομένων απευθείας στο γραφικό περιβάλλον ή από αρχείο, καθώς και τα υποδείγματα αρχείων .txt (μόνο για είσοδο δεδομένων), .xml (για είσοδο



Εικόνα 2. Υπόδειγμα επίλυσης άσκησης

παραμέτρων και δεδομένων), και .xsd (για υποστήριξη διαλειτουργικότητας). Στο Βήμα 2 (δεν παρουσιάζεται λόγω έλλειψης χώρου), πραγματοποιείται η εισαγωγή των παραμέτρων αν αυτές δεν έχουν εισαχθεί μέσω του αρχείου XML, και στο Βήμα 3 (Εικόνα 2) παρουσιάζεται η λύση και η επεξήγηση κάθε σταδίου επίλυσης της άσκησης (το παράδειγμα εδώ είναι σύντομο λόγω έλλειψης χώρου). Τέλος, παρέχεται η δυνατότητα εκτύπωσης ή εξαγωγής σε αρχείο PDF της άσκησης με τη λύση της.

Η έκταση και η μορφή των εξηγήσεων επίλυσης των ασκήσεων που δημιουργούνται και υποβάλλονται διαφέρει ανάλογα με το είδος του αλγορίθμου και της άσκησης. Επιπλέον, για όλες τις παρεχόμενες πληροφορίες ακολουθούνται οι μεθοδολογίες επίλυσης που προτείνονται από τη σχετική βιβλιογραφία (Aggarwal, 2015; Manning et al., 2008; Witten et al., 2011).

Αποδοχή. Στον ένα χρόνο λειτουργίας της πλατφόρμας DST είχαμε την ευκαιρία να κάνουμε μία πρώτη αποτίμηση της αποδοχής της από την κοινότητα της τριτοβάθμιας εκπαίδευσης. Η πλατφόρμα έχει μέχρι στιγμής χρησιμοποιηθεί στα πλαίσια μαθημάτων που σχετίζονται με την ΕΔ και παρέχονται σε προγράμματα σπουδών πρώτου ή δεύτερου κύκλου σε πανεπιστήμια της Ελλάδας και του εξωτερικού. Η αποδοχή της και από τους φοιτητές του οικείου τμήματος ήταν ενθαρρυντική· στα ερωτηματολόγια αξιολόγησης των σχετικών μαθημάτων που διανέμονται από τη Μονάδα Διασφάλισης Ποιότητας (ΜΟΔΙΠ) απάντησαν με πολύ θετικά σχόλια για την πλατφόρμα DST και για τη συμβολή της (i) στην κατανόηση και εμβάθυνση της ύλης, (ii) στην προετοιμασία για τις εξετάσεις, και (iii) στην αξιολόγηση των αποτελεσμάτων των γραπτών δοκιμασιών. Επομένως, παρότι δεν έχει μέχρι στιγμής γίνει συστηματική μελέτη χρήσης (user study) όσον αφορά στα μαθησιακά αποτελέσματα, η αποδοχή και τα πρωτόλεια σχόλια των χρηστών είναι ενθαρρυντικά.

Μελλοντικές ερευνητικές κατευθύνσεις

Μελλοντικές ερευνητικές κατευθύνσεις και επεκτάσεις περιλαμβάνουν τη διεξαγωγή μιας διευρυμένης μελέτης χρήσης για την εξαγωγή μακροσκοπικών συμπερασμάτων, η ενσωμάτωσή της πλατφόρμας σε συστήματα εικονικής μάθησης (π.χ., Open eClass, Moodle), και η αυτοματοποιημένη δημιουργία ασκήσεων με διαβαθμισμένη δυσκολία.

Αναφορές

Aggarwal, C.C. (2015). Data Mining: The Textbook, *Springer*.

Brunner, R.J. & Kim, E.J. (2016). Teaching Data Science, *Procedia Computer Science*, Vol. 80.

Efthimiadis, E.N., Fernandez-Luna, J.M., Huete, J.F., & MacFarlane, A. (2011). Teaching and Learning in Information Retrieval, *The Information Retrieval Series, Vol. 31, Springer*.

Grivokostopoulou, F., Perikos, I., & Hatzilygeroudis, I. (2017), An Educational System for Learning Search Algorithms and Automatically Assessing Student Performance. *I. J. Artificial Intelligence in Education* 27(1): 207-240.

Halttunen, K. & Sormunen, E. (2000). Learning Information Retrieval through an Educational Game. Is Gaming sufficient for learning?, *Education for Information* 18(4): 289-311.

Hicks, S.C. & Irizarry, R.A. (2016). A Guide to Teaching Data Science, *eprint arXiv:1612.07140.*

Manning, C.D., Raghavan P., & Schütze, H. (2008). Introduction to Information Retrieval, *Cambridge University Press*, 2008.

Nakayama, L., Nóbile de Almeida, V., & Vicari, R. (2004). A Personalized Information Retrieval Service for an Educational Environment. *Intelligent Tutoring Systems*, pp. 842-844.

Willms S. (2003). Visualizing a User Model for Educational Adaptive Information Retrieval. *User Modeling,* pp. 432-434.

Witten, I.H., Frank, E., & Hall, M.A. (2011). Data Mining: Practical Machine Learning Tools and Techniques, *Morgan Kaufmann*.