Κ23α - Ανάπτυξη Λογισμικού Για Πληροφοριακά Συστήματα

Χειμερινό Εξάμηνο 2020 – 2021

Καθηγητής Ι. Ιωαννίδης

Άσκηση 1 – Παράδοση: Δευτέρα 14 Δεκεμβρίου 2020

Υλοποίηση αρνητικής συσχέτισης προϊόντων

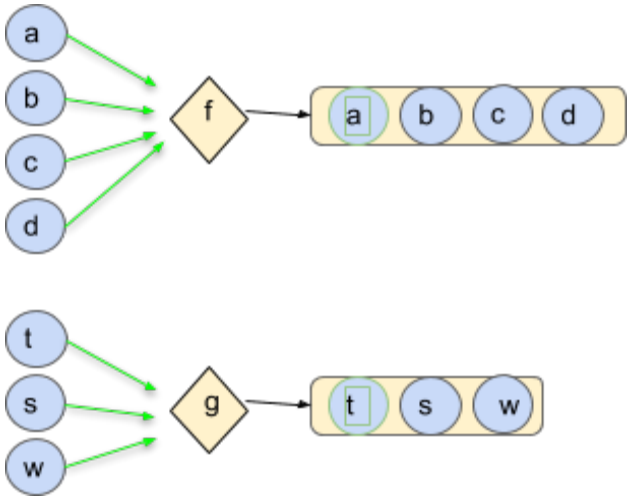
Έχοντας υλοποιήσει τις κλίκες για τα προϊόντα που ταιριάζουν μεταξύ τους, η υλοποίηση θα προχωρήσει στην επεξεργασία προϊόντων που δεν ταιριάζουν μεταξύ τους. Συνεχίζουμε να θεωρούμε ότι κάθε spec θα αντιστοιχιστεί με ένα id (έστω a, b, c,..). Κάθε id θα δείχνει κάθε στιγμή σε ένα σύνολο από άλλα id με τα οποία ταιριάζει, αλλά και πιθανώς σε άλλα σύνολα με τα οποία δεν ταιριάζει.

Η επεξεργασία γραμμών αρνητικής συσχέτισης, δηλαδή του τύπου (a, b, 0) θα πρέπει να λαμβάνει υπόψη της ότι δεν αρκεί να καταγραφεί η αρνητική συσχέτιση a, b, αλλά και το ότι θα πρέπει επαγωγικά να καταγραφεί και αρνητική συσχέτιση μεταξύ της κλίκας στην οποία ανήκει το a και της κλίκας στην οποία ανήκει το b.

Συγκεκριμένα θα ισχύσει ο παρακάτω πίνακας

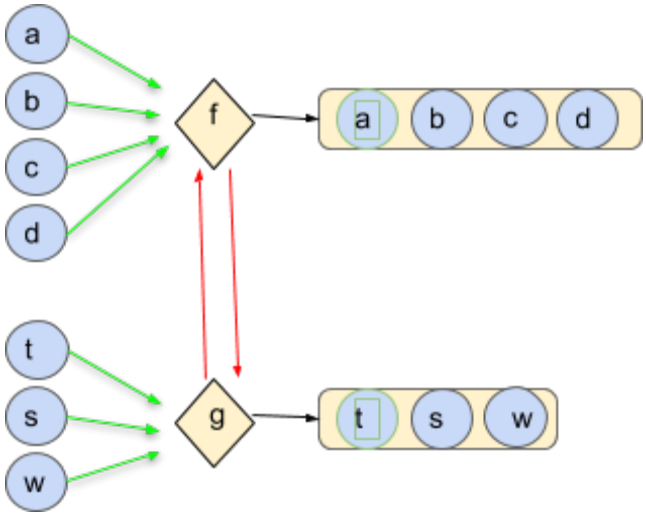
Υπάρχουσα κατάσταση	a, b, 1	a, b, 1	a, b, 0	a, b, 0
Νέα γραμμή	a, c, 1	a, b, 0	a, c, 0	a, c, 1
Επεξεργασία	b, c, 1	b, c, 0	?	b, c, 0

Ο πίνακας αυτός μπορεί να εφαρμοστεί και στη γενικότητα των κλικών. Έστω η εξής



κατάσταση:

Αν σε αυτή την κατάσταση, αναγνωστεί γραμμή π.χ (c, s, 0), τότε θα πρέπει να δημιουργηθεί αρνητική συσχέτιση μεταξύ όλων των μελών της μιας κλίκας με όλα τα μέλη της άλλης. Ένας απλός τρόπος να συμβολιστεί αυτό είναι μέσω της απευθείας αρνητικής συσχέτισης μεταξύ των κλικών, δηλαδή:



Όπως φαίνεται και στο σχήμα, θα πρέπει να υπάρχει αρνητική συσχέτιση τόσο από το σύνολο f προς το σύνολο g, όσο και για την αντίστροφη κατεύθυνση.

Η υλοποίησή σας, θα πρέπει να μπορεί να διαχειριστεί περισσότερες της μίας αρνητικές συσχετίσεις που να εμπλέκουν οποιαδήποτε κλίκα.

Μηχανική Μάθηση

Ένας δυαδικός ταξινομητής είναι μία συνάρτηση που ταξινομεί στιγμιότυπα σε δύο κλάσεις (π.χ. 0 και 1). Η μηχανική μάθηση μας δίνει μεθόδους εκμάθησης ενός τέτοιου ταξινομητή με τρόπο αυτόματο, από ένα σύνολο παραδειγμάτων της μορφής (x,y), δηλαδή ένα σύνολο από ζευγάρια εισόδου-εξόδου. Είναι σύνηθες ο ταξινομητής αυτός να έχει κάποια παραμετρική μορφή φ(.,w) στην οποία περίπτωση το πρόβλημα εκμάθησης γίνεται ένα πρόβλημα βελτιστοποίησης ως προς τις παραμέτρους w κάποιας συνάρτησης σφάλματος L. Η συνάρτηση L που χρησιμοποιείται για δυαδική ταξινόμηση είναι συνήθως η cross-entropy (ή binary cross-entropy). Ένας καλός ταξινομητής πρέπει να μπορεί να ταξινομήσει σωστά νέα δεδομένα που δεν έχει επεξεργαστεί κατά τη διάρκεια εκμάθησης. Αυτό σημαίνει πως πολλές φορές ένας ταξινομητής που μαθαίνει να ταξινομεί καλά το σύνολο δεδομένων εκμάθησης μπορεί να μην είναι καλός ταξινομητής (καθώς μπορεί να μην γενικεύει καλά). Για αυτό το λόγο συνηθίζεται να κρατάμε ένα μικρό μέρος του συνόλου εκμάθησης ξεχωριστά, το οποίο δεν χρησιμοποιείται για την εκπαίδευση του ταξινομητή, άλλα χρησιμοποιείται για την αξιολόγηση ταξινομητών που εκπαιδεύουμε στα υπόλοιπα δεδομένα, με σκοπό να επιλέξουμε τον ταξινομητή που γενικεύει καλύτερα.

Ζητούμενο στην άσκηση είναι να βρεθεί για κάθε πιθανό ζεύγος από προϊόντα το εάν σχετίζονται (1) ή όχι (0). Για ένα μικρό υποσύνολο των ζευγών αυτή ή πληροφορία δίνεται στο αρχείο W. Θα πρέπει να χρησιμοποιήσετε τα δεδομένα αυτά για να φτιάξετε έναν ταξινομητή με τον οποίο θα πάρετε την ζητούμενη πληροφορία για όλα τα υπόλοιπα ζεύγη. Φυσικά, θα πρέπει να χρησιμοποιήσετε όλη τη διαθέσιμη πληροφορία που δίνεται για κάθε προϊόν και είναι επιτρεπτό να κάνετε κάποια προεπεξεργασία στα δεδομένα έτσι ώστε να τα φέρετε σε μία κατάλληλη αναπαράσταση για είσοδο στον ταξινομητή σας.

Για την εκμάθηση και υλοποίηση του ταξινομητή σας θα χρησιμοποιήσετε τη βιβλιοθήκη Tensorflow η οποία διαθέτει C/C++ API (https://www.tensorflow.org/api_docs/cc). Είναι αναγκαίο να χρησιμοποιήσετε το συγκεκριμένο API και δεν επιτρέπεται η χρήση άλλης γλώσσας προγραμματισμού.

Παράδοση εργασίας

Προθεσμία παράδοσης: 14/12/2020
Γλώσσα υλοποίησης: C / C++ χωρίς χρήση stl.
Περιβάλλον υλοποίησης: Linux (gcc > 5.4+).

Παραδοτέα: Η παράδοση της εργασίας θα γίνει με βάση το τελευταιο commit πριν την προθεσμία υποβολής στο git repository σας. **Η χρήση git είναι υποχρεωτική**.

Επιπλέον, εκτός από τον πηγαίο κώδικα, θα παραδώσετε μια σύντομη αναφορά, με τις σχεδιαστικές σας επιλογές καθώς και να εφαρμόσετε ελέγχους ως προς την ορθότητα του λογισμικού με τη χρήση ανάλογων βιβλιοθηκών (Software testing).