

Ανάκτηση Πληροφορίας και Μηχανές Αναζήτησης

Εργασία 05 – Collaborative Filtering

Ονοματεπώνυμο: Παναγιώτης Γιαννουτάκης

AM: 12/38

Email: it1238@uom.edu.gr

Γενικές οδηγίες για το πρόγραμμα

Το πρόγραμμα όταν σχεδιαζόταν έτρεχε για 10 χρήστες και 10 αντικείμενα και αυτό για να εμφανίζονται εύκολα τα αποτελέσματα του πίνακα στο τερματικό. Αν αλλαχθούν οι μεταβλητές M και N και είναι διαφορετικές μεταξύ τους θα έχουμε αρκετές NaN (Not a Number) τιμές οπότε δεν συστήνεται. Αν αλλαχθούν οι χρήστες (N) θα εμφανίζει error τιμές οπότε δεν συστήνεται. Το πρόγραμμα τρέχει σωστά όταν M και N έχουν ίδιες τιμές.

Υπάρχουν σχόλια σε κάθε κλάση εκτός από τις Cosine.java και AdjustedCosine.java και αυτό γιατί ο κώδικας πέρα από τον υπολογισμό του τύπου που διαφέρει, είναι ίδιος. Ότι ισχύει στις κλάσεις Jaccard.java και Dice.java ισχύει και στις άλλες δύο.

Οι K-Κοντινότεροι Γείτονες που επιλέχθηκαν είναι 3. Μπορεί να αλλαχθεί η τιμή μέσα από τον κώδικα.

Το ποσοστό των συνολικών βαθμολογιών που είναι άγνωστες είναι 10%. Κάθε χρήστης δηλαδή έχει ένα αντικείμενο που δεν έχει βαθμολογήσει. Το ποιο είναι αυτό το αντικείμενο είναι στην τύχη. Ο κώδικας για να αποφασίζει στην τύχη το πρόγραμμα από μόνο του ποιο θα είναι το ποσοστό των άγνωστων βαθμολογιών υπάρχει σε μορφή σχολίων. Αυτό διότι σε περιπτώσεις που έβγαине πολύ μεγάλο (50% και πάνω) δεν γινόταν να υπολογισθούν οι τύποι επειδή δεν υπήρχαν επαρκή δεδομένα. Έτσι επέλεξα να την απενεργοποιήσω αυτήν την δυνατότητα και να το θέτω στο 10% εκ των προτέρων.

Πειράματα και Αποτελέσματα

Στα παρακάτω πειράματα οι μεταβλητές M και N έχουν τιμή 100 και ο αριθμός των επαναλήψεων για κάθε αλγόριθμο είναι 10. Τα αποτελέσματα φαίνονται παρακάτω.

Αλγόριθμος: Jaccard, KNN: 5

```
Poion algorithmo theleis na xrisimopoiiseis? <1: Jaccard> <2: Dice> <3: Cosine> <4: AdjustedCosine>
1
Me poses epanalipseis theleis na ektelestei o algorithmos?
10
To geniko Meso Apoluto Lathos olon ton epanalipseon einai: 0.03305454545454545
time to compute = 0.172000 seconds
```

Αλγόριθμος: Dice, KNN: 5

```
Poion algorithmo theleis na xrisimopoiiseis? <1: Jaccard> <2: Dice> <3: Cosine> <4: AdjustedCosine>
2
Me poses epanalipseis theleis na ektelestei o algorithmos?
10
To geniko Meso Apoluto Lathos olon ton epanalipseon einai: 0.034557575757575884
time to compute = 0.125000 seconds
```

Αλγόριθμος: Cosine, KNN: 5

```
Poion algorithmo theleis na xrisimopoiiseis? <1: Jaccard> <2: Dice> <3: Cosine> <4: AdjustedCosine>
3
Me poses epanalipseis theleis na ektelestei o algorithmos?
10
To geniko Meso Apoluto Lathos olon ton epanalipseon einai: 0.033385858585858685
time to compute = 0.127000 seconds
```

Αλγόριθμος: AdjustedCosine, KNN: 5

```
Poion algorithmo theleis na xrisimopoiiseis? <1: Jaccard> <2: Dice> <3: Cosine> <4: AdjustedCosine>
4
Me poses epanalipseis theleis na ektelestei o algorithmos?
10
To geniko Meso Apoluto Lathos olon ton epanalipseon einai: 0.03160202020202031
time to compute = 0.128000 seconds
```

Αλγόριθμος: Jaccard, KNN: 3

```
Poion algorithmo theleis na xrisimopoiiseis? <1: Jaccard> <2: Dice> <3: Cosine> <4: AdjustedCosine>
1
Me poses epanalipseis theleis na ektelestei o algorithmos?
10
To geniko Meso Apoluto Lathos olon ton epanalipseon einai: 0.033505050505050545
time to compute = 0.128000 seconds
```

Αλγόριθμος: Dice, KNN: 3

```
Poion algorithmo theleis na xrisimopoiiseis? <1: Jaccard> <2: Dice> <3: Cosine> <4: AdjustedCosine>
2
Me poses epanalipseis theleis na ektelestei o algorithmos?
10
To geniko Meso Apoluto Lathos olon ton epanalipseon einai: 0.0346734006734007
time to compute = 0.098000 seconds
```

Αλγόριθμος: Cosine, KNN: 3

```
Poion algorithmo theleis na xrisimopoiiseis? <1: Jaccard> <2: Dice> <3: Cosine> <4: AdjustedCosine>
3
Me poses epanalipseis theleis na ektelestei o algorithmos?
10
To geniko Meso Apoluto Lathos olon ton epanalipseon einai: 0.03240740740740741
time to compute = 0.101000 seconds
```

Αλγόριθμος: AdjustedCosine, KNN: 3

```
Poion algorithmo theleis na xrisimopoiiseis? <1: Jaccard> <2: Dice> <3: Cosine> <4: AdjustedCosine>
4
Me poses epanalipseis theleis na ektelestei o algorithmos?
10
To geniko Meso Apoluto Lathos olon ton epanalipseon einai: 0.03421548821548825
time to compute = 0.108000 seconds
```

Ερωτήματα της Άσκησης

- 1) Πώς επηρεάζει το ποσοστό των γνωστών βαθμολογιών την απόδοση (ακρίβεια) του συστήματος;

Όπως αναφέρθηκε και προηγουμένως μεγάλο ποσοστό άγνωστων βαθμολογιών επηρεάζει το αποτέλεσμα. Συγκεκριμένα δεν γίνεται να υπολογισθούν οι τύποι. Σε περίπτωση που γίνεται, όσο μεγαλύτερο είναι το ποσοστό, τόσο πιο χρονοβόρα είναι η διαδικασία.

2) Ποια είναι η καλύτερη συνάρτηση ομοιότητας;

Σύμφωνα με τα πειράματα που φαίνονται παραπάνω για $KNN = 3$ η καλύτερη συνάρτηση είναι η Cosine. Για $KNN = 5$ είναι η AdjustedCosine. Άλλα οι διαφορές είναι αμελητέες όσον αφορά το Μέσο Απόλυτο Λάθος. Η μόνη αξιοσημείωτη διαφορά βρίσκεται στους χρόνους εκτέλεσης των συναρτήσεων και εκεί η Jaccard συνάρτηση είναι η πιο χρονοβόρα.

3) Πώς επηρεάζει το K την απόδοση (ακρίβεια) του συστήματος;

Από τα πειράματα βλέπουμε ότι το k δεν παίζει ιδιαίτερα σημαντικό ρόλο όσον αφορά το Μέσο Απόλυτο Λάθος. Αν το αυξήσουμε σημαντικά (π.χ. το κάνουμε 20) τότε βλέπουμε αύξηση στον χρόνο εκτέλεσης,

4) Ποια είναι η καλύτερη μέθοδος πρόβλεψης;

Ίδιο με το ερώτημα 2).