

Εξόρυξη Δεδομένων

Άσκηση 3η

User User Collaborative Filtering

Ευτυχία Κιάφα AM:3003

Χειμερινό εξάμηνο 2020-2021

Υλοποίηση

Η υλοποίηση της άσκησης έγινε σε notebook με χρήση ενός csv αρχείου: *user-user-collaborative-filtering-matrix.csv* για την φόρτωση των δεδομένων. Για τα κενά κελιά χρησιμοποιήθηκε η τιμή -1.

Για την πρώτη περίπτωση υπολογίστηκε αρχικά για κάθε χρήστη σε σχέση με τον χρήστη X, το cosine similarity. Οι τιμές που προέκυψαν ήταν:

Cosine Similarity			
	UserY	UserW	UserZ
UserX	0.7947194142390264	0.39901193756050635	0.23693955110363696

Οι πιο όμοιοι χρήστες είναι οι: UserY και UserW, καθώς η τιμή του similarity τους, είναι μεγαλύτερη.

Στην συνέχεια υπολογίστηκε ο σταθμισμένος μέσος όρος για την ταινία 6, με βάρος το cosine similarity κάθε χρήστη. Ο τύπος που χρησιμοποιήθηκε ήταν:

$$\frac{\text{cosineXW_similarity} * \text{UserW_Movie6['Rating']} + \text{cosineXY_similarity} * \text{UserY_Movie6['Rating']}}{\text{cosineXW_similarity} + \text{cosineXY_similarity}}$$

Η τιμή του σταθμισμένου μέσου όρου και τιμή του κελιού (X,6) είναι: **2.002768178013424** και με στρογγυλοποίηση προς τα κάτω είναι : 2

Για την δεύτερη περίπτωση για κάθε μη αρνητικό κελί και κάθε γραμμή, υπολογίστηκε ο μέσος όρος των βαθμολογιών:

Mean Value of no negative ratings				
	UserX	UserY	UserW	UserZ
Mean	3.0	2.0	3.333333	2.5

Αφαιρέθηκε ο μέσος όρος από τα μη αρνητικά κελιά και αντικαταστάθηκαν οι τιμές στις βαθμολογίες των χρηστών με τον πίνακα πια να διαμορφώνεται ως εξής:

	Movie1	Movie2	Movie3	Movie4	Movie5	Movie6
UserX	2	0	-2	1	-1	-1
UserY	2	0	-1	-1	-1	1
UserZ	0.5	-1.5	0.5	0.5	3.0	3.0
UserW	-1.333333	1.666667	-2.333333	1.666667	-0.333333	0.666667

Κατόπιν, υπολογίστηκε ξανά το cosine similarity για κάθε χρήστη με τον χρήστη X.

Cosine Similarity			
	UserY	UserW	UserZ
UserX	0.7947194142390264	0.39901193756050635	0.23693955110363696

Οι πιο όμοιοι χρήστες ήταν και πάλι οι UserY και UserW.

Για τον υπολογισμό του σταθμισμένου μέσου όρου απόκλισης χρησιμοποιήθηκε ο εξής τύπος:

$$\frac{\text{cosineXW_similarity} * \text{UserW_Movie6['Rating']} + \text{cosineXY_similarity} * \text{UserY_Movie6['Rating']}}{\text{cosineXW_similarity} + \text{cosineXY_similarity}}$$

Με αποτέλεσμα: **2.002768178013424**

Για την τελική τιμή του (X,6),προστέθηκε και ο μέσος όρος του X,οπότε προκύπτει:**5.002768178013424**

Με στρογγυλοποίηση της τιμής προς τα κάτω : **5**

Παρατηρήσεις

- 1)Το cosine similarity παρέμεινε ίδιο παρόλο που τροποποιήθηκαν οι τιμές του πίνακα,αφαιρώντας τον μέσο όρο κάθε γραμμής.
- 2)Πιο όμοιοι χρήστες παρέμειναν οι UserY και UserW.
- 3)Τα κελιά με την μικρότερη σημασία είναι τα κενά κελιά,καθότι μειώνουν σταθερά,σύμφωνα με το πλήθος τους τις τιμές.