



Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών
Υπολογιστών
Εργαστήριο Διοίκησης Πληροφοριακών Συστημάτων

Διοίκηση Ψηφιακής Επιχείρησης

Εργασία 4: Αναζήτηση και Recommender Systems

Καραβαγγέλης Αθανάσιος
ΑΜ: 03117022

1. ΑΝΑΖΗΤΗΣΗ – TFIDF

Υποθέστε ότι πραγματοποιείτε μια αναζήτηση σε μια μηχανή αναζήτησης με τους όρους “Sci-Fi, space, astro”. Θεωρείστε σαν μοναδικά δεδομένα της μηχανής τις παρακάτω περιγραφές ταινιών. α. Χρησιμοποιείστε τη μετρική TF-IDF (χωρίς TF normalization) για να βρείτε ποια θα είναι η κατάταξη των αποτελεσμάτων. Να χρησιμοποιηθεί Cosine Similarity και να ληφθούν υπόψη το κείμενο των τίτλων και οι παράγωγες/σύνθετες λέξεις (Stemming). Σχολιάστε τα αποτελέσματα.

Term Frequency (TF)

Term	2001: A Space Odyssey	Solaris	Gravity	Moon	Interstellar
Sci-fi	1	0	1	2	1
space	3	2	2	1	0
astro	3	0	1	2	0

Inverse Document Frequency (IDF)

Term	IDF	N	d
Sci-fi	0.223	5	4
space	0.223	5	4
astro	0.511	5	3

$$IDF = \log \frac{N}{|\{d \in D : t \in d\}|},$$

όπου N ο συνολικός αριθμός εγγράφων και d το πλήθος των εγγράφων όπου εμφανίζεται ο όρος t .

TF – IDF

Term	2001: A Space Odyssey	Solaris	Gravity	Moon	Interstellar
Sci-fi	0.223	0	0.223	0.446	0.223
space	0.669	0.446	0.446	0.223	0
astro	1.533	0	0.511	1.022	0

$$TF - IDF = TF \times IDF$$

Query

Term	TF	IDF	TFIDF
Sci-fi	0.333	0.223	0.0743
space	0.333	0.223	0.0743
astro	0.333	0.511	0.1702

Euclidean Distance: $\|p\| = \sqrt{p_1^2 + p_2^2 + \dots + p_n^2}$

Άρα:

$$\|movie\| = \sqrt{(sci - fi)_{TFIDF}^2 + space_{TFIDF}^2 + astro_{TFIDF}^2}$$

Οπότε αναλυτικά θα είναι:

$$\|2001: A Space Odyssey\| = \sqrt{0.223^2 + 0.669^2 + 1.533^2} = 1,6874$$

$$\|Solaris\| = \sqrt{0^2 + 0.446^2 + 0^2} = 0,446$$

$$\|Gravity\| = \sqrt{0.223^2 + 0.446^2 + 0.511^2} = 0,7139$$

$$\|Moon\| = \sqrt{0.446^2 + 0.223^2 + 1.022^2} = 1,1372$$

$$\|Interstellar\| = \sqrt{0.223^2 + 0^2 + 0^2} = 0,223$$

$$\|Query\| = \sqrt{0.0743^2 + 0.0743^2 + 0.1702^2} = 0,2$$

Cosine similarity:

$$sim(i, j) = cos(\vec{i}, \vec{j}) = \frac{\vec{i} \cdot \vec{j}}{\|\vec{i}\| \cdot \|\vec{j}\|}$$

Έγραψα αναλυτικά την πράξη για το 1^ο cosine similarity αλλά για τα υπόλοιπα για εξοικονόμηση χώρου και χρόνου θα παραθέσω κατευθείαν τα τελικά αποτελέσματα.

$$sim(Query, 2001: A Space Odyssey) = cos(\overrightarrow{Query}, \overrightarrow{A Space Odyssey})$$

$$= \frac{\overrightarrow{Query} \cdot \overrightarrow{A Space Odyssey}}{\|\overrightarrow{Query}\| \cdot \|\overrightarrow{A Space Odyssey}\|} =$$

$$= (0.0743 * 0.223 + 0.0743 * 0.669 + 0.1702 * 1.533) / 0.2 * 1.6874$$

$$= \mathbf{0,9695}$$

$$sim(Query, 2001: A Space Odyssey) = \mathbf{0,9695}$$

$$sim(Query, Solaris) = \mathbf{0,3715}$$

$$sim(Query, Gravity) = \mathbf{0,9573}$$

$$sim(Query, Moon) = \mathbf{0,9833}$$

$$sim(Query, Interstellar) = \mathbf{0,3715}$$

Άρα η τελική κατάταξη θα είναι η εξής:

1. Moon
2. 2001: A Space Odyssey
3. Gravity
4. Interstellar – Solaris

Σχολιασμός

Οι ταινίες Interstellar και Solaris κατατάσσονται στην τελευταία θέση αφού και οι 2 περιέχουν μόνο μία από τις 3 λέξεις. Οι 3 πρώτες ταινίες είναι αρκετά κοντά αλλά η ταινία Moon υπερिशύει περιέχοντας την λέξη sci-fi 2 φορές σε αντίθεση με τις 2001: A Space Odyssey και Gravity που την περιέχουν από 1 φορά. Η ταινία Moon είναι πρώτη επειδή δεν ψάχνουμε την ταινία με τις περισσότερες εμφανίσεις των λέξεων (η οποία είναι η 2001: A Space Odyssey με 7) αλλά την ταινία με διάνυσμα πιο κοντά σε αυτό του Query.

2. ΑΝΑΖΗΤΗΣΗ – PRECISION / RECALL

Εχετε | Google | e.com | αναζητή | Αγορά | ΒΙΒΛΙΑ | Βιβλία | πατάκι | PATAK | ΕΚΔΟΣ |


← → ↻ 🏠 google.com/search?q=gravity&xsrf=APq-WBsM0S6_LkWKz4TfhAsYOjrNVWe22Q:1647092208385&ei=8


Google gravity X 🔍


🔍 All 🖼 Images 📺 Videos 📰 News 📍 Maps ⋮ More Tools


About 378,000,000 results (0.65 seconds)


On Earth all bodies have a weight, or downward force of gravity, proportional to their mass, which Earth's mass exerts on them. Gravity is **measured by the acceleration that it gives to freely falling objects**. At Earth's surface the acceleration of gravity is about 9.8 metres (32 feet) per second per second.





 <https://www.britannica.com › Science › Astronomy> ⋮
gravity | Definition, Physics, & Facts - Encyclopedia Britannica


 <https://www.imdb.com › title> ⋮
Gravity (2013) - IMDb
A medical engineer and an astronaut work together to survive after an accident leaves them adrift ... Two astronauts work together to survive after an accident ...
Oct 6, 2013 · ★★★★★ Rating: 7.7/10 · 807,182 votes
[Gravity](#) · [Awards](#) · [Trivia](#) · [Sandra Bullock as Ryan Stone](#)


 <https://en.wikipedia.org › wiki › Gravity> ⋮
Gravity - Wikipedia
Gravity or gravitation, is a natural phenomenon by which all things with mass or energy, including planets, stars, galaxies and even light, are attracted to ...
[Anti-gravity](#) · [Linearized gravity](#) · [Massive gravity](#) · [Entropic gravity](#)


 [https://en.wikipedia.org › wiki › Gravity_\(2013_film\)](https://en.wikipedia.org › wiki › Gravity_(2013_film)) ⋮
Gravity (2013 film) - Wikipedia
Gravity is a 2013 science fiction thriller film directed by Alfonso Cuarón, who also co-wrote, co-edited, and produced the film. It stars Sandra Bullock and ...
Budget: \$80–130 million Release dates: [August 28, 2013 \(Venice\)](#); ...
Box office: \$723.2 million Production companies: [Heyday Films](#); [Es...](#)

 <https://spaceplace.nasa.gov › what-is-gravity> ⋮
What Is Gravity? | NASA Space Place – NASA Science for Kids
Gravity is the force by which a planet or other body draws objects toward its center. The force of gravity keeps all of the planets in orbit around the sun.

 <https://www.warnerbros.com › movies › gravity> ⋮
Gravity | Movies - WarnerBros.com
Academy Award® winners Sandra Bullock (The Blind Side) and George Clooney (Up in the Air, Syriana) star in this heart-pounding thriller written,

 <https://www.gravityforms.com> ⋮
Gravity Forms | The Best WordPress Form Plugin | Form Builder
Create custom forms on WordPress with Gravity Forms, the easiest-to-use, secure, and reliable WordPress form builder plugin for your website.

 <https://www.merriam-webster.com › dictionary › gravity> ⋮
Gravity Definition & Meaning - Merriam-Webster
The meaning of GRAVITY is dignity or sobriety of bearing. How to use gravity in a sentence.

 <https://www.skroutz.gr › Baseus-B&B...> ⋮ [Translate this page](#)
Baseus Βάση Κινητού Αυτοκινήτου Osculum Type Gravity με ...
Δες τιμή, χαρακτηριστικά & πραγματικές κριτικές χρηστών για το προϊόν Baseus Βάση Κινητού Αυτοκινήτου Osculum Type Gravity με Ρυθμιζόμενο Άγκιστρο SUYL-XP0S ...
★★★★★ Rating: 3.7 · 79 reviews · Starting from €6.40 · In stock



<https://gravitymobility.com>

Gravity Electric Vehicles NYC

Experience the next generation of urban sustainable mobility and electric vehicle infrastructure.
Book a luxury ride with us today.



<https://www.dictionary.com/browse/gravity>

Gravity Definition & Meaning | Dictionary.com

Gravity definition, the force of attraction by which terrestrial bodies tend to fall toward the center of the earth. See more.



<https://www.0-gravity.ae>

Zero Gravity Dubai: The Best pool and beach bar night club in ...

Zero Gravity is the best pool and beach bar night club in Dubai. We are offering the Best Ladies night and DJs party in Dubai, UAE.



<https://www.gravityrenewables.com>

Home - Gravity Renewables, Inc.

Advancing Small Hydroelectric Projects. Gravity Renewables is an owner, operator, and developer of small hydroelectric power plants in the United States.



<https://dictionary.cambridge.org/dictionary/gravity>

GRAVITY | meaning in the Cambridge English Dictionary

3 days ago — gravity definition: 1. the force that attracts objects towards one another, especially the force that makes things fall.... Learn more.



<https://gravityvault.com>

The Gravity Vault | Indoor Rock Climbing Gyms | NJ, NY & PA

The best indoor rock climbing gyms! Located in Chatham, Flemington, Hoboken, Middletown, Upper Saddle River, & Voorhees NJ | Melville & Poughkeepsie, ...



<https://www.gravitysketch.com>

Gravity Sketch | 3D design and modelling software

Gravity Sketch is an intuitive 3D design platform for cross-disciplinary teams to create, collaborate, and review in an entirely new way.



<https://www.gravity-uk.com>

Gravity: Welcome

Welcome to Gravity. The number one trampoline park in the UK. Gravity Trampoline Parks provide fun for the whole family with a range of activities for you ...



<https://www.camirafabrics.com/.../Contract-Fabrics>

Gravity - Contract Upholstery Fabric Supplier | Camira

We design, manufacture and supply fabrics for commercial interiors, offices, schools, universities, bars, restaurants, hotels & healthcare.



<https://www.gravitymedia.com>

Gravity Media | Global Leader in Broadcast Solutions

Gravity Media is a global leader in broadcast solutions formed by established broadcast and production facilities: Gearhouse Broadcast, ...

<https://www.framestore.com/work/gravity>

Gravity | Framestore

Gravity. Warner Bros. Pictures. Alfonso Cuarón's remarkable blockbuster Gravity enjoyed fantastic critical success, collecting enough stars from film reviewers ...

<https://www.washingtonpost.com/work/2013/10/21/>

Here's what 'Gravity' gets right and wrong about space - The ...

Oct 21, 2013 — Kevin Grazier, the science adviser for Alfonso Cuarón's new film "Gravity," has a bone to pick with Twitter-famous astrophysicist Neil ...



<https://science.howstuffworks.com/.../Geophysics>

How does gravity work? - Science | HowStuffWorks

Gravity is the force that causes two particles to pull towards each other. Learn about gravity and find out how Newton and Einstein explain gravity.

www.warnerbros.com > movies > gravity ▼

Gravity | Movies - WarnerBros.com

Academy Award® winners Sandra Bullock (The Blind Side) and George Clooney (Up in the Air, Syriana) star in this heart-pounding thriller written, directed and ...

www.imdb.com > title ▼

Gravity (2013) - IMDb

★★★★★ Rating: 7.7/10 - 738,784 votes

Two astronauts work together to survive after an accident leaves them stranded in space. ...
Ryan Stone (Sandra Bullock) is a brilliant medical engineer on her first shuttle mission, with veteran astronaut Matt Kowalski (George Clooney) in command of his last flight before retiring.



www.livescience.com > 37115-what-is-gravity ▼

What Is Gravity? | Live Science

Jun 3, 2013 - **Gravity** is the force that attracts two bodies toward each other, the force that causes apples to fall toward the ground and the planets to orbit the ...



www.space.com > classical-gravity ▼

What Is Gravity? | Space

Aug 21, 2019 - Einstein suggested that the shape of spacetime is what gives rise to the force we experience as **gravity**. A concentration of mass (or energy), such ...



www.britannica.com > science > gravity-physics ▼

gravity | Definition, Physics, & Facts | Britannica

Gravity, also called gravitation, in mechanics, the universal force of attraction acting between all matter. It is by far the weakest known force in nature and thus ...

www.rottentomatoes.com > gravity_2013 ▼

Gravity (2013) - Rotten Tomatoes

★★★★★ Rating: 96% - 347 reviews

Feb 25, 2014 - **Gravity** focuses on two astronauts, played by Sandra Bullock and George Clooney, as their Space Shuttle mission is torn apart (literally) by an ...

Rating: PG-13 (for intense perilous sequences, ... **Genre:** Drama, Science Fiction & Fantasy

α) Αρχικά, για το συγκεκριμένο ερώτημα ας εξηγήσουμε εν τάχει τις παρακάτω έννοιες:

- True positives: Τα αποτελέσματα που θα έπρεπε να εμφανίζει η μηχανή αναζήτησης και πράγματι τα εμφανίζει.
- False positive: Τα αποτελέσματα που δεν ταιριάζουν στην αναζήτηση αλλά η μηχανή τα εμφανίζει
- True negatives: Τα αποτελέσματα που δεν ταιριάζουν στην αναζήτηση και η μηχανή δεν τα εμφανίζει
- False negatives: Τα αποτελέσματα που ενώ ταιριάζουν στην αναζήτηση, η μηχανή αναζήτησης δεν τα εμφανίζει.

Στις παραπάνω εικόνες έχω σημειώσει με πράσινο χρώμα τα αποτελέσματα της αναζήτησης τα οποία μας παρέχουν πληροφορίες σχετικά με την ταινία Gravity, δηλαδή είναι σωστά – true positive με πλήθος 8 ενώ τα υπόλοιπα με κόκκινη ένδειξη είναι λάθος – false positive, με πλήθος 20.

Τα σωστά αποτελέσματα για εμάς είναι αυτά που περιέχου όρους όπως ‘movie’, ‘film’ και εταιρείες – ιστοσελίδες που σχετίζονται με ταινίες π.χ. IMDb, WarnerBros, Rotten Tomatoes κ.α. Από την άλλη τα περισσότερα λανθασμένα αποτελέσματα με βάση τη δική μας άσκηση αναφέρονται στη λέξη «gravity» ως τη βαρύτητα που όλοι γνωρίζουμε.

β) Αν γνωρίζουμε ότι υπάρχουν ακόμη 450 αποτελέσματα που σχετίζονται με την ταινία και δεν βρέθηκαν – false negative – υπολογίστε τα παρακάτω

i. Precision

ii. Recall

iii. F-Measure

Σχολιάστε τα αποτελέσματα.

Θεωρούμε:

- true positives: $t_p = 8$
- false positives: $f_p = 20$
- false negatives: $f_n = 450$

και έχουμε:

i. $Precision = \frac{t_p}{t_p + f_p} = \frac{8}{28} = 0,2857$

ii. $Recall = \frac{t_p}{t_p + f_n} = \frac{8}{458} = 0,0175$

iii. $F - measure = \frac{t_p}{t_p + \frac{1}{2}(f_p + f_n)} = \frac{8}{8 + 0.5(470)} = \frac{8}{242} = 0,0331$

Σχολιασμός

Δεδομένου ότι το precision δείχνει τι ποσοστό των αποτελεσμάτων ανταποκρίνεται σε αυτό που εμείς καθορίζουμε ως “positive” βλέπουμε ότι το 28% είναι και το λογικό αφού από όλα τα αποτελέσματα μόνο 8 ήταν «σωστά». Είναι ένα αρκετά χαμηλό precision.

Το recall μας δείχνει την πιθανότητα ή το αντίστοιχο ποσοστό ένα αποτέλεσμα που συσχετίζεται με την ταινία να επιστραφεί από την αναζήτηση. Δεδομένου ότι έχουμε 450 false negatives δηλαδή σχετικά αποτελέσματα που όμως δεν κατατάχθηκαν ως σωστά ενώ θα έπρεπε, το recall είναι πολύ χαμηλό, μόλις στο 1.75%.

Τέλος, η μετρική f-measure είναι ο αρμονικός μέσος του precision και του recall και είναι λογικό να έχει τιμή ενδιάμεση στα 2 και πιο κοντά στο recall λόγω του υψηλού αριθμού false negatives.

Γενικά, τα αποτελέσματα ήταν χαμηλά, κάτι αναμενόμενο λόγω του χαμηλού αριθμού true positives.

3. Recommender Systems

Έξι άνθρωποι αξιολόγησαν τις ταινίες του ερωτήματος 1 με βαθμολογία από 1 (καθόλου καλή) έως 10 (εξαιρετική) και τα αποτελέσματα φαίνονται στον παρακάτω πίνακα.

Αντικαταστήστε όπου Χ τον τελευταίο ψηφίο του Αριθμού Μητρώου σας, αφού προσθέσετε τον αριθμό 1. (ΑΜ: 03117022 άρα το ψηφίο 2+1 = 3)

Ο τελικός πίνακας είναι ο παρακάτω:

	2001: A Space Odyssey	Solaris	Gravity	Moon	Interstellar
Χρήστης 1	3	4		6	7
Χρήστης 2	3	2		5	4
Χρήστης 3	5	1	4	3	3
Χρήστης 4	7	6	7	9	6
Χρήστης 5		2	3	3	4
Χρήστης 6	9	7	8	7	10

- a. Υπολογίστε την ομοιότητα (similarity) μεταξύ των 6 χρηστών χρησιμοποιώντας δυο μεθόδους: Ευκλείδεια απόσταση και Pearson Correlation.

Αρχικά, θα εξηγήσουμε πως προκύπτει η ομοιότητα χρησιμοποιώντας κάθε μία μέθοδο και έπειτα θα παρουσιάσουμε τα αποτελέσματα.

Μέθοδος της Ευκλείδειας Απόστασης

Για τον υπολογισμό της ομοιότητας χρησιμοποιώντας ευκλείδεια απόσταση ακολουθούμε τα εξής βήματα:

- Έστω X_i, Y_i οι βαθμολογίες 2 χρηστών X , ή Y αντίστοιχα, στο i -οστό αντικείμενο.
- Η ομοιότητα προκύπτει ως:

$$\frac{1}{1 + \sqrt{(X_1 - Y_1)^2 + (X_2 - Y_2)^2 + \dots + (X_i - Y_i)^2}}$$

, όπου X_i, Y_i βαθμολογίες για ταινίες που έχουν ΚΑΙ οι 2 χρήστες

- Τα αποτελέσματα φαίνονται παρακάτω. Για να αποφύγω τις πράξεις έγραψα ένα python script με κατάλληλες τροποποιήσεις ώστε να μη συμπεριλαμβάνονται οι ταινίες που δεν έχουν βαθμολογηθεί και από τους 2 χρήστες.

Χρήστης 1	Χρήστης 2	Απόσταση	Ομοιότητα
1	2	3.742	0.211
1	3	6.164	0.14
1	4	5.477	0.154
1	5	4.69	0.176
1	6	7.416	0.119
2	3	3.162	0.24
2	4	7.211	0.122
2	5	2.0	0.333
2	6	10.05	0.09
3	4	9.11	0.099
3	5	1.732	0.366

3	6	11.533	0.08
4	5	8.485	0.105
4	6	5.099	0.164
5	6	10.1	0.09

Μέθοδος Pearson Correlation

Για αυτή τη μέθοδο ακολουθούμε τα εξής βήματα:

- Βρίσκουμε το άθροισμα των αξιολογήσεων για κάθε χρήστη, έστω S_X και S_Y και τα αθροίσματα των τετραγώνων των αξιολογήσεων για κάθε χρήστη, έστω $S_{X_squared}$ και $S_{Y_squared}$
- Έπειτα, ξεκινώντας με άθροισμα $s = 0$ και για κάθε ταινία που έχουν αξιολογήσει και οι 2 χρήστες, προσθέτουμε στο s την τιμή $X_i * Y_i$
- Επίσης χρησιμοποιούμε την παράμετρο n που ισούται με τον αριθμό των ταινιών που έχουν αξιολογήσει και οι 2 χρήστες
- Πριν βρούμε την ομοιότητα υπολογίζουμε τα εξής:

$$num = s - \frac{S_X S_Y}{n}, \quad den = \sqrt{\left(S_{X_squared} - \frac{S_X^2}{n}\right) \left(S_{Y_squared} - \frac{S_Y^2}{n}\right)}$$

$$pearson_correlation = \frac{num}{den}$$

- Τελικά: $similarity = \frac{1 + pearson_correlation}{2}$
- Τα αποτελέσματα φαίνονται παρακάτω, και πάλι χρησιμοποιήσα ένα python script για τις πράξεις:

Χρήστης 1	Χρήστης 2	Pearson correlation	Ομοιότητα
1	2	0.902	0.951
1	3	-0.51	0.245
1	4	0.075	0.538
1	5	0.361	0.68
1	6	0.21	0.605
2	3	-0.378	0.311
2	4	0.424	0.712
2	5	0.12	0.56
2	6	0.12	0.56
3	4	-0.228	0.386
3	5	-0.424	0.288
3	6	0.665	0.832
4	5	0.0	0.5
4	6	-0.47	0.265
5	6	0.076	0.538

b. Χρησιμοποιώντας K-Nearest Neighbors με $k=2$ και weighted average και με τις δύο μετρικές του ερωτήματος α. υπολογίστε πως περιμένουμε να αξιολογήσει την ταινία Gravity ο χρήστης 2. Σχολιάστε τα αποτελέσματα.

Εδώ μας ζητείται μία πρόβλεψη για την βαθμολογία του χρήστη 2 για την ταινία Gravity χρησιμοποιώντας τις μεθόδους του α ερωτήματος. Αναμένουμε λίγο διαφορετικά αποτελέσματα καθώς βλέπουμε διαφορές στα similarities ανάμεσα στις 2 μεθόδους.

Ο τύπος που έχουμε για weighted average K-means με $K=2$ είναι ο εξής:

$$pred(2 \rightarrow Gravity) = \frac{similarity(X, 2) \cdot rating_X + similarity(Y, 2) \cdot rating_Y}{similarity(X, 2) + similarity(Y, 2)}$$

, όπου $similarity(y, h)$ η ομοιότητα μεταξύ των χρηστών z και h και X, Y οι κοντινότεροι γείτονες του χρήστη 2

- **Με Euclidean:**

Οι 2 κοντινότεροι γείτονες του χρήστη 2 είναι ο χρήστης 5 και ο χρήστης 3, με ομοιότητες 0.333 και 0.24 αντίστοιχα.

Άρα θα είναι:

$$pred(2 \rightarrow Gravity) = \frac{0.333 \cdot rating_5 + 0.24 \cdot rating_3}{0.333 + 0.24} = 3,4188$$

- **Με Pearson:**

Οι 2 κοντινότεροι γείτονες του χρήστη 2 είναι ο χρήστης 4 και ο χρήστης 5, με ομοιότητες 0.865 και 0.828 αντίστοιχα.

Άρα θα είναι:

$$pred(2 \rightarrow Gravity) = \frac{0.712 \cdot rating_4 + 0.56 \cdot rating_5}{0.712 + 0.56} = 5,23$$

Παρατηρήσεις

Παρατηρούμε πως τα αποτελέσματα απέχουν περίπου 2 μονάδες που δεν είναι μεγάλη απόσταση αλλά όχι και μικρή, γενικά το pearson correlation σα μέθοδος παρουσιάζει μεγαλύτερα similarities από ότι η Euclidean και αυτό επηρεάζει το αποτέλεσμα μας.

c. Αν υποθέσουμε ότι χρησιμοποιούμε τις προτιμήσεις των χρηστών στις ταινίες για να προτείνουμε φίλους, τότε ποιες σχέσεις σας φαίνονται πιο πιθανές; Εξηγήστε

Για να προτείνουμε φίλους με βάση τις προτιμήσεις των χρηστών θα βρίσκαμε ποιοι χρήστες έχουν μεγαλύτερο similarity μεταξύ τους. Αυτό θα μπορούσε να γίνει με κάθε έναν από τους 2 τρόπους που χρησιμοποιήσαμε παραπάνω. Για παράδειγμα, με βάση το Pearson Correlation τα ζευγάρια χρηστών που θα προτείναμε, με φθίνουσα τιμή similarity είναι:

1. Χρήστες 1 και 2 -> 0.951
2. Χρήστες 3 και 6 -> 0.832
3. Χρήστες 2 και 4 -> 0.712
4. Χρήστες 1 και 5 -> 0.68

Με βάση το Euclidean similarity αντίστοιχα θα προτείνουμε:

1. Χρήστες 3 και 5 $\rightarrow 0.366$
2. Χρήστες 2 και 5 $\rightarrow 0.333$
3. Χρήστες 2 και 3 $\rightarrow 0.24$
4. Χρήστες 1 και 2 $\rightarrow 0.211$
5. Χρήστες 1 και 5 $\rightarrow 0.176$

Ίσως θα είχε νόημα να υπολογίζαμε και ένα συνδυασμό των 2 μεθόδων, βάσει και ενός παράγοντα εμπιστοσύνης στο διάστημα $[0,1]$ ανάλογα πόσο εμπιστευόμαστε κάθε τρόπο ώστε να λάβουμε υπόψη και τις 2 μεθόδους.