

به نام خدا



دانشگاه صنعتی شریف

دانشکده مهندسی برق

مقدمه‌ای بر یادگیری ماشین

پروژه پایانی

فاز اول

استاد درس: دکتر امینی

مهران مظاهری – محمدرضا صفوی

۹۸۱۰۶۷۰۱ – ۹۸۱۰۲۳۴۶

۲۸ دی

1. Explain Recommender Systems according to the references.

According to primary definition, recommender systems are systems which people provide recommendations as inputs, which the system then aggregates and directs to appropriate recipients, The term now has a broader connotation, describing any system that produces individualized recommendations as output or has the effect of guiding the user in a personalized way to interesting or useful objects in a large space of possible options. Such systems have an obvious appeal in an environment where the amount of on-line information vastly outstrips any individual's capability to survey it.

2. Write some of the applications of recommender systems in different areas.

Product recommendation – Movie recommendation (generally media recommendation) – News articles – Social media (recommend contents which is most likely to be liked by user) – online Ads (show users ads related to user search history)

3. There are various challenges faced by Recommendation Systems. Write these challenges.

1. Data sparsity: Since the pool of available items is often exceedingly large, overlap between two users is often very small or none. Further, even when the average number of evaluations per user/item are high, they are distributed among the users/items very unevenly and hence majority of users/items may have expressed/received only a few ratings.

2. Scalability: While the data is mostly sparse, for major sites it includes millions of users and items. It is therefore essential to consider the computational cost issues and search for recommender algorithms that are either little demanding or easy to parallelize.

3. Cold start: When new users enter the system, there is usually insufficient information to produce recommendation for them.

4. Diversity vs. accuracy: When the task is to recommend items which are likely to be appreciated by a particular user, it is usually most effective to recommend popular and highly rated items. Such recommendation, however, has very little value for the users because popular objects are easy to find (often they are even hard to avoid) without a recommender system.

5. Vulnerability to attacks: Due to their importance in e-commerce applications, recommender systems are likely targets of malicious attacks trying to unjustly promote or inhibit some items.

6. The value of time: While real users have interests with widely diverse time scales (for example, short term interests related to a planned trip and long term interests related to the place of living or political preferences), most recommendation algorithms neglect the time stamps of evaluations.

7. Evaluation of recommendations: While we have plenty of distinct metrics, how to choose the ones best corresponding to the given situation and task is still an open question. Comparisons of different recommender algorithms are also problematic because different algorithms may simply solve different tasks.

8. User interface: It has been shown that to facilitate users' acceptance of recommendations, the recommendations need to be transparent: users appreciate when it is clear why a particular item has been recommended to them. Another issue is that since the list of potentially interesting items may be very long, it needs to be presented in a simple way and it should be easy to navigate through it to browse different recommendations which are often obtained by distinct approaches.

4. Explain similarity-based methods for recommender systems.

This class of algorithms divided into methods employing user and item similarity, respectively.

The basic assumption of a method based on user similarity is that people who agree in their past evaluations tend to agree again in their future evaluations. Thus, for a target user, the potential evaluation of an object is estimated according to the ratings from users ("taste mates") who are similar to the target user.

Different from user similarity, an algorithm based on item similarity recommends a user the objects that are similar to what this user has collected before. Sometimes the opinions from dissimilar users or the negative ratings can play a significant (even positive) role in determining the recommendation, especially when the data set is very sparse and thus the information about relevance is more important than that about correlation. For additional information see the recent review articles, and is a nice survey that contains a number of similarity indices.

5. Explain classes of hybridization.

Hybrid RCs are categorized in 7 different classes based on the techniques which are used. They are as follows:

1. Weighted: it uses weighted linear function to aggregate output scores of each technique.

2. Feature combination: This type of hybrid RSs treats one recommender's output as additional feature data, and uses the other recommender (usually content-based which makes extensive use of item features) over the new extended data.

3. Cascade: These methods are example of staged recommendation process. The technique is employed to generate a coarse ranking of candidate items and then a second technique refines the list from the preliminary candidate set.

4. Switching: in this class, using technique changes as recommender system grows.

5. Feature augmentation: In this class of hybrids, using technique is obtained from combination of technique to produce an item prediction or classification which is then comprised in the operation of the other recommendation technique.

6. Meta level: This is example of order-sensitive hybrid RSs that use an entire model produced by the first technique as input for the second technique. It is typical to use content-based recommenders to build item representation models, and then employ these models in collaborative recommenders to match the items with user profiles.

7. Mixed: as it could be realized from the name “mixed”, this class puts together a high number of different recommenders simultaneously.

6. In this question, we learn the utility matrix.

A. What is the utility matrix?

A numerical representation, giving for each user-item pair for certain feature, a value that represents what is known about the degree of preference of that user for that item.

B. For real world data, is it an sparse matrix?

Yes, because there isn't measured preference for all user-item pair in dataset and most of them are empty.

C. Is this sentence true? “The goal of a recommendation system is to predict all the blanks in the utility matrix”

No, it isn't necessary to predict every blank entry in a utility matrix. Rather, it is only necessary to discover some entries in each row that are likely to be high. In most applications, the recommendation system does not offer users a ranking of all items, but rather suggests a few that the user should value highly.

D. Write the two general approaches to discovering the value users place on items to build the utility matrix.

1. We can ask users to rate items: Movie ratings are generally obtained this way, and some on-line stores try to obtain ratings from their purchasers. Sites providing content, such as some news sites or YouTube also ask users to rate items. This approach is limited in its effectiveness, since generally users are unwilling to provide responses, and the information from those who do may be biased by the very fact that it comes from people willing to provide ratings.

2. We can make inferences from users' behavior. Most obviously, if a user buys a product at Amazon, watches a movie on YouTube, or reads a news article, then the user can be said to “like” this item.

7. content-based recommendations. And 8.different approach to recommendation

شماره 7

(a) در این گونه ای، تمرکز ما بر ویژگی های محصولات است. شایسته محصولات برای ما شایسته ویژگی های این محصولات نبوده باشد. برای و محصول باید مجموعه از خواص داشته باشیم و آن ها را در یک بردار قرار دهیم. این ویژگی ها اصولاً ویژگی های اصلی هستند.

(b) 1- مجموعه بازیگران 2- کارگردان 3- نویسنده 4- سال تولید 5- ژانر 6- نمره IMDb 7- بودجه تولید 8- کشور تولید کننده 9- میزان فروش در هند آدلین 10- لیکن تولید کننده

(c) 1- به کارخانه داریم

1 1 0 1 0 1 0 1 3
1 1 1 0 1 0 1 0 4

$$\cos \alpha = \frac{1+1+12}{\sqrt{5+9} \sqrt{5+16}} = \frac{14}{\sqrt{204}}$$

(d) مانند این که این فیلم در سال 2010 میلادی در هند تولید شده است پس در این مناطق با این

بازار برای این فیلم 0.2 باشد.

(e) $\frac{0+1+2}{3} = 1$ \therefore 1, 3-3, 4-3, 5-3

مسعود و پناه ۷ م یانلی ۲-۹، ۳-۹، ۵-۹، ۱-۹، ۴-۹
 $\frac{-2-1+1}{3} = \frac{-2}{3}$

سے شروع

$$\cos \alpha_{12} = \frac{2.98 \times 7.55 + 400 \times 250 \alpha^2 + 5 \times 3 \beta^2}{\sqrt{2.98^2 + 400^2 \alpha^2 + 5^2 \beta^2} \sqrt{2.55^2 + 250^2 \alpha^2 + 3^2 \beta^2}}$$

$$\cos \alpha_{13} = \frac{2.98 \times 2.87 + 400 \times 600 \alpha^2 + 5 \times 5 \beta^2}{\sqrt{2.98^2 + 400^2 \alpha^2 + 5^2 \beta^2} \sqrt{2.87^2 + 600^2 \alpha^2 + 5^2 \beta^2}}$$

$$\cos \alpha_{23} = \frac{2.55 \times 2.87 + 250 \times 600 \alpha^2 + 3 \times 5 \beta^2}{\sqrt{2.55^2 + 250^2 \alpha^2 + 3^2 \beta^2} \sqrt{2.87^2 + 600^2 \alpha^2 + 5^2 \beta^2}}$$

$$\cos \alpha_{12} = \frac{2.98 \times 2.55 + 400 \times 260 + 5 \times 3}{\sqrt{2.98^2 + 400^2 + 5^2} \sqrt{2.55^2 + 260^2 + 3^2}} = 0.9999$$

میں کو رسا ہے

$$GSA_{1,3} = 0.9999$$

$$\cos \theta_{23} = 0.9999$$

(h) نتائج قبل الحاف $G_{s\alpha_{12}} = 0.988$ $G_{s\alpha_{13}} = 0.978$ $G_{s\alpha_{23}} = 0.947$

جواب: $\beta = \frac{1}{\frac{5+3+5}{3}}$, $\alpha = \frac{1}{\frac{900+750+600}{3}}$ در این صورت (i)

$$\alpha = 2.4 \times 10^{-3}, \beta = 0.23$$

$\cos \alpha_{12} = 0.993$ $\cos \alpha_{13} = 0.990$ $\cos \alpha_{23} = 0.971$ دایره قدرت ۱

$$\Rightarrow \alpha_{12} \approx 6.78^\circ, \alpha_{13} \approx 8.10^\circ, \alpha_{23} \approx 13.83^\circ$$

(د) باید سوار شد، 2.8، باید سوار شد، 416.7، باید سوار شد، 4.3 است، باید سوار شد، 1.3

$$\begin{bmatrix} 0.18 & -0.25 & 0.07 \\ -16.7 & -16.7 & 183.3 \\ 0.7 & -1.3 & 0.7 \end{bmatrix}$$

(ک) راه حل آن است که روابط را به زبان کمترین طولانی Max-min و روابط استوار

بین هر یک در روابط اتیم به Max-min در آن می‌کنیم. به این صورت می‌توان به مقصود احوال رسید.

به این صورت داریم.

$$\begin{bmatrix} 0.419 & -0.581 & 0.163 \\ -0.048 & -0.476 & 0.524 \\ 0.35 & -0.65 & 0.35 \end{bmatrix}$$

همه داریم، اما به دنبال اعداد جدید

$$\alpha_{12} = 145.4^\circ, \alpha_{13} = 65.46^\circ, \alpha_{23} = 152.1^\circ$$

$$0.33, -1.66, 1.33$$

(د) داریم، باید سوار شد

$$0.11, -0.55, 0.44$$

به داده‌های دیگر

$$0.11 \quad -0.55 \quad 0.44$$

(m)

$$2.28 \quad 2.55 \quad 2.87$$

$$400 \quad 250 \quad 600$$

$$5 \quad 3 \quad 5$$

(18) 1/2

$$J_{12} = \frac{4}{8} = \frac{1}{2}, \quad J_{13} = \frac{4}{8} = \frac{1}{2}, \quad J_{23} = \frac{4}{8} = \frac{1}{2}$$

(a)

$$\cos \alpha_{12} = 0.48, \quad \cos \alpha_{13} = 0.64, \quad \cos \alpha_{23} = 0.57$$

(b)

$$1 \ 1 \ 0 \ 1 \ 0 \ 0 \ 1 \ 0$$

(c) 1/2

$$0 \ 0 \ 1 \ 1 \ 0 \ 0 \ 0 \ 0$$

$$0 \ 0 \ 0 \ 1 \ 0 \ 1 \ 1 \ 1$$

$$J_{12} = \frac{1}{5}, \quad J_{13} = \frac{2}{6}, \quad J_{23} = \frac{1}{5}$$

(d)

$$\cos \alpha_{12} = 0.35, \quad \cos \alpha_{13} = 0.3, \quad \cos \alpha_{23} = 0.35$$

(e)

$$\begin{bmatrix} 1.84 & 0.84 & 0 & 0.84 & -2.16 & 0 & -0.16 & -1.16 \\ 0 & -0.16 & 1.84 & 0.84 & -1.16 & -0.16 & -1.16 & 0 \\ -0.83 & 0 & -1.83 & 0.17 & 0 & 1.17 & 1.17 & 0.17 \end{bmatrix}$$

(f)

$$\cos \alpha_{12} = 0.379, \cos \alpha_{13} = -0.21, \cos \alpha_{23} = -0.70 \quad (P)$$

Scanned with CamScanner

9. Explain Dimensionality Reduction (with examples).

In dimensionality reduction, we estimate blank entries of utility matrix by assuming that utility matrix would be equal to product of two long, thin matrices. This view makes sense if there are a relatively small set of features of items and users that determine the reaction of most users to most items. we sketch one approach to discovering two such matrices.

For example, in movie preference: Most users respond to a small number of features; they like certain genres, they may have certain famous actors or actresses that they like, and perhaps there are a few directors with a significant following. If we start with the utility matrix M , with n rows and m columns (i.e., there are n users and m items), then we might be able to find a matrix U with n rows and d columns and a matrix V with d rows and m columns, such that UV closely approximates M in those entries where M is nonblank.

$$\begin{bmatrix} 5 & 2 & 4 & 4 & 3 \\ 3 & 1 & 2 & 4 & 1 \\ 2 & & 3 & 1 & 4 \\ 2 & 5 & 4 & 3 & 5 \\ 4 & 4 & 5 & 4 & \end{bmatrix} = \begin{bmatrix} u_{11} & u_{12} \\ u_{21} & u_{22} \\ u_{31} & u_{32} \\ u_{41} & u_{42} \\ u_{51} & u_{52} \end{bmatrix} \times \begin{bmatrix} v_{11} & v_{12} & v_{13} & v_{14} & v_{15} \\ v_{21} & v_{22} & v_{23} & v_{24} & v_{25} \end{bmatrix}$$

If so, then we have established that there are d dimensions that allow us to characterize both users and items closely. We can then use the entry in the product UV to estimate the corresponding blank entry in utility matrix M . This process is called UV-decomposition of M .