Algorithms and Methods in Recommender Systems

Daniar Asanov
Berlin Institute of Technology
Berlin, Germany

Abstract—Today, there is a big variety of different approaches and algorithms of data filtering and recommendations giving. In this paper we describe traditional approaches and explain what kind of modern approaches have been developed lately. All the paper long we will try to explain approaches and their problems based on a movies recommendations. In the end we will show the main challenges recommender systems come across.

I. INTRODUCTION

With the beginning of the Web 2.0 era, the internet began growing up and developing with tremendous speed. Many opportunities, such as sharing knowledge, information, opinion with other users, came out. This did favor the development of social networks like Facebook. Nowadays, authors can share their creations with millions of readers around the globe. Amateur-musicians can get famous faster than ever before just with uploading their tracks. Business world have found more customers and profit in the internet. The variety of online shops, auctions or flea markets opened up in the internet. Today, every user of the World Wide Web can purchase almost any item being in any country of the world. As opposed to real shops, in the internet there are no place-limitations. In fact, there is almost endless place. Nevertheless people came across a new problem in the WWW. The amount of information and items got extremely huge, leading to an information overload. It became a big problem to find what the user is actually looking for. Search engines partially solved that problem, however personalization of information was not given. So developers found a solution in recommender systems. Recommender systems are tools for filtering and sorting items and information. They use opinions of a community of users to help individuals in that community to more effectively identify content of interest from a potentially overwhelming set of choices. [1] There is a huge diversity of algorithms and approaches that help creating personalized recommendations. Two of them became very popular: collaborative filtering and content-based filtering. They are used as a base of most modern recommender systems.

Appearance of mobile devices with new technologies, such as GPS ans 3G standards, on market issued new challenges. Recommender systems got involved in developing process of tourism, security and other areas. Modern recommender systems improving their recommendations accuracies by using context-aware, semantic and other approaches. Today, recommendations are more specific and personalized. Problems of combining different technologies and recommending approaches for better results will always exist and will be the reason of new researches. I wish you the best of success.

II. TRADITIONAL RECOMMENDER APPROACHES

A. Content-based filtering

Content-based recommender systems work with profiles of users that are created at the beginning. A profile has information about a user and his taste. Taste is based on how the user rated items. Generally, when creating a profile, recommender systems make a survey, to get initial information about a user in order to avoid the new-user problem. [2]

In the recommendation process, the engine compares the items that were already positively rated by the user with the items he didn’t rate and looks for similarities. Those items that are mostly similar to the positively rated ones, will be recommended to the user.

Figure 1 shows an example of a user profile with the movies he/she has watched and the ratings the user made. Figure 2 shows the list of movies and their attribute-values. A content-based recommender system would find out movies from the list (Figure 2) that the user has already watched and positively rated. Then, it would compare those movies with the rest of the movies from the list (Figure 2) and look for similarities. Similar movies would be recommended the user. In the current example we can see that there is a movie “Scary Movie” similar to the movie ”American Pie” that the user positively rated. The user hasn’t rated “Scary Movie” so it will be recommended him/her.

<table>
<thead>
<tr>
<th>Movies</th>
<th>Comedy</th>
<th>Violence</th>
<th>Horror</th>
<th>Explicit Content</th>
</tr>
</thead>
<tbody>
<tr>
<td>American Pie</td>
<td>10</td>
<td>3</td>
<td>1</td>
<td>9</td>
</tr>
<tr>
<td>Scary Movie</td>
<td>8</td>
<td>8</td>
<td>4</td>
<td>9</td>
</tr>
<tr>
<td>Saw</td>
<td>2</td>
<td>10</td>
<td>10</td>
<td>7</td>
</tr>
</tbody>
</table>

Figure 1: The movies the user has watched

<table>
<thead>
<tr>
<th>Movies</th>
<th>Green Lantern</th>
<th>Source Code</th>
<th>American Pie</th>
<th>Hangover 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ratings</td>
<td>8</td>
<td>7</td>
<td>9</td>
<td>10</td>
</tr>
</tbody>
</table>

Figure 2: The movies list

There are different algorithms of measuring similarities among items in database and those in users profile. [3] One of such approaches is cosine similarity. Representing items as vectors on a coordinate space it measures angles between vectors and gives out their cosine value. Vectors \( \vec{w}_c \) and \( \vec{w}_s \) of two items with attributes are compared in cosine similarity function as follows: [4]

\[
\cos \theta = \frac{\vec{w}_c \cdot \vec{w}_s}{\| \vec{w}_c \| \| \vec{w}_s \|}
\]
\[ u(c, s) = \cos(w^c_c, w^s_s) = \frac{\vec{w}^c_c \cdot \vec{w}^s_s}{||\vec{w}^c_c|| \times ||\vec{w}^s_s||} = \]
\[ = \frac{\sum_{i=1}^{K} w^c_i w^s_i}{\sqrt{\sum_{i=1}^{K} w^2_i c} \sqrt{\sum_{i=1}^{K} w^2_i s}} \]

The more similar two items are, smaller the angle between their vectors. [5]

Collaborative filtering is widely used in e-commerce. Customers can rate books, songs, movies and then get recommendations regarding those issues in future. Moreover collaborative filtering is utilized in browsing of certain documents (e.g. documents among scientific works, articles, and magazines). [7]

Going in details of methods of collaborative filtering we can distinguish most popular approaches: user-based, item-based and model-based approaches.

1) User-based approach: This approach was proposed in the end of 1990s by the professor of University of Minnesota Jonathan L. Herlocker.[8] In the user-based approach, the users perform the main role. If certain majority of the customers has the same taste then they join into one group. Recommendations are given to user based on evaluation of items by other users from the same group, with whom he/she shares common preferences. If the item was positively rated by the community, it will be recommended to the user. Thus in the user-based approach the items that were already rated by the user before play an important role in searching a group that shares appreciations with him. [9] [10] [11] (See Figure 5).

2) Item-based approach: This approach was proposed by the researchers of University of Minnesota in 2001 [12]. Referring to the fact that the taste of users remains constant or change very slightly similar items build neighborhoods based on appreciations of users. Afterwards the system generates recommendations with items in the neighborhood that a user would prefer [12] [8] (See Figure 6).
C. Hybrid recommendation approaches

For better results some recommender systems combine different techniques of collaborative approaches and content-based approaches. Using hybrid approaches we can avoid some limitations and problems of pure recommender systems, like the cold-start problem. The combination of approaches can proceed in different ways [13]:

1) Separate implementation of algorithms and joining the results.
2) Utilize some rules of content-based filtering in collaborative approach.
3) Utilize some rules of collaborative filtering in content-based approach.
4) Create a unified recommender system, that brings together both approaches.

Robin Burke worked out a taxonomy of hybrid recommender systems categorizing them. [14]

CinemScreen is an example of a recommendation agent that gives its users recommendations based on hybrid filtering about the movies that are shown in cinemas. On the website a user can create an account and evaluate all movies he/she has seen in cinemas. At first the system uses collaborative filtering. On the outcome of collaborative filtering it applies content-based filtering. The combination of two approaches gives CinemaScreen opportunity to make more recommendations [15] (See Figure 7).

III. Modern recommendation approaches

A. Context-aware approaches

Context is the information about the environment of a user and the details of situation he/she is in. Such details may play much more significant role in recommendations than ratings of items, as the ratings alone don’t have detailed information about under which circumstances they were given by users. Some recommendations can be more suitable to a user in evening and doesn’t match his preferences in the morning at all and he/she would like to do one thing when it’s cold and completely another when it’s hot outside. The recommender systems that pay attention and utilize such information in giving recommendations are called context-aware recommender systems.

Nowadays, when the mobile devices are getting popular and taking part in our lives, this kind of recommender systems are especially urgent. Using GPS, 3G access to internet and other technologies recommender systems can rapidly get any information about location of a user and the user himself. [16], [17]

As opposed to content information that is saved in profiles, context changes dynamically and often saved just permanently, as it is more likely to lose its currency after a certain period of time. For example most of the people would prefer to go to swimming pool, rather than to go to beach when it’s raining. So in such recommendations the weather would play a leading role. Thats why it is very important to periodically refresh the information. Context-aware recommender systems became much attention, as they noticeably increased the quality of recommendations and the approaches became more specific to use in certain areas.

Gediminas Adomavicius and Alexander Tuzhilin introduced three different algorithmic paradigms of using the contextual information in recommendation process: contextual pre-filtering, post-filtering and modeling. In contextual pre-filtering contextual information is used at selecting data at the beginning, like ratings, to be used for a further predictions, utilizing any approach. In the case of contextual post-filtering contextual information is used for filtering after predictions of ratings. In the last paradigm contextual information is used in prediction approach itself. [18]

Baltrunas and Ricci have purposed contextual item rating pre-filtering approach that is based on item splitting. At first their approach searches the items that have significant differences in their ratings. Each of such items is discovered for a contextual feature, up to which the ratings significantly change. If there is one, it creates two new items instead of the old one with definite contextual information. The main idea of algorithm is to find a contextual feature to split the item. For example, the movies about Santa Claus and his adventures might be higher rated in winter, at Christmas time. Adult movies might be more watched in the night than in the morning or afternoon. In the last example the future of item splitting is the daytime. The approach would split the adult movie ”Wild Orchid” into two: the one with high ratings and contextual information ”night” for the feature ”daytime” and the other one with lower ratings and the contextual information ”day”. A recommender system that uses this approach would then use the contextual information about time to make recommendations about movies. This approach is an extension of collaborative filtering.[19]

The same principle is used in micro-profiling technique...
of Baltrunas and Amatriain. This time-dependant approach splits user profiles based on change of preferences at different periods of time.[20]

There are many ways of representation of contextual information and its relationship with items and users. One of such representations is a contextual graph. Other than representation, utilizing graph algorithms, it can also help to improve predictions. In a contextual graph items, users and contextual objects are represented as nodes of the graph and the edges would be ratings of items and similarity indicators between users. Toine Bogers used contextual graphs in his movie recommendation algorithm ContextWalk that is based on Markovs random walks.[21] Context-aware recommender systems became a part of tourism domain.

One of the biggest problems of context-aware recommender systems is obtaining context information. The information can be obtained explicitly by directly interacting with user asking him/her to fill out a form and making a survey. Although it is mostly desirable to obtain context information without making the whole rating and reviewing process complicated. Another way is gathering information implicitly using the sources like GPS, to get location, or a timestamp on transaction.[18] The last way of information extraction is analyzing users observing their behavior or using data mining techniques. For example: obtaining information from reviews. People usually like to write their reviews in free text form. The problem is to get important contextual information from such reviews. There are many text mining algorithms including artificial intelligence techniques.[22]

Context-aware recommender systems that work with positioning systems became a part of tourism domain. Such systems are called location-based recommender systems.[23] One of them is COMPASS - an application for mobile devices that utilizes GPS. Based on the location of a tourist and his/her interests the system gives recommendations. If the tourist is interested in antique buildings, the software will show up such building around him on the map. In connection with mobile internet technologies the system can give more detailed recommendations like opening times of museums or special offers of restaurants. Knowing the location the application can also require a taxi to that place [17] (See Figure 8).

B. Semantic based approaches

Most of the descriptions of items, users in recommender systems and the rest of the web are presented in the web in a textual form. Using tags and keywords without any semantic meanings doesn't improve the accuracy of recommendations in all cases, as some keywords may be homonyms. That's why understanding and structuring of text is a very significant part recommendation. Traditional text mining approaches that base on lexical and syntactical analysis show descriptions that can be understood by a user but not a computer or a recommender system. That was a reason of creating new text mining techniques that were based on semantic analysis. Recommender systems with such techniques are called semantic-based recommender systems.[24]

The performance of semantic recommender systems are based on knowledge base usually defined as a concept diagram (like taxonomy) or ontology.[25]

On Wikipedia Taxonomy is defined as a practice and science of classification.[26] Taxonomy plays an important role in semantic analysis. Classification of items and users concerning their domains and groups brings much efficiency in recommendation system. In movies recommending, for example, the classification of movies up to their genre and building a tree would lead to a better understanding of users tastes by a system (See Figure 9).

However, in some cases taxonomies are not the best way of sorting and classifying items, as implementation of taxonomies may be very complicated and expensive, or even not possible. Web services for photo sharing are good example of difficult taxonomy creation. One of the best solutions of this problem would be folksonomies (folks + taxonomy). Folksonomy is items sorting approach, whereby users add tags to items and the most frequent and popular tags would build categories of items. [27] [28] [29]

Tags in folksonomies and taxonomies are related with each other using different ontologies. Swartout defined Ontology as a hierarchically structured set of terms for describing a domain that can be used as a skeletal foundation for a knowledge base.[30] Gruber said that an ontology defines the basic terms and relations comprising the vocabulary of a topic area, as well as the rules for combining terms and relations to define extensions to the vocabulary.[30] So with the help of ontology recommender systems can understand how some terms (items, users, tags, etc) are related to each other (See Figure 10).
As an ontology description language, developers use Web Ontology Language (OWL).[31]

Despite an existence of commonly used ontologies, like WordNet, OpenCyc and Gene Ontology, there are still some challenges[24]. One of them is their currency. Relations between items may change and that would bring us to not up to date ontologies.

Ahmed Elgohary purposed an approach that uses Wikipedia as an ontology for the semantic analysis of text. Articles in Wikipedia are constantly updated and some of them are given in many languages. They are also connected with each other via hyper-links. Using a Wikipedia-based annotator the recommender system creates an annotations repository, which is further used for semantic descriptions of terms (items).[24]

Semantic analysis can be used in different type of recommender systems for a better understanding relation between items and users, like creating a semantic neighborhood among ontological profiles. Taxonomy of items is one of successful solutions for "New Item" problem.[31][32]

TripFromTV is an example of a recommender agent that works with Digital Television based on a semantic information. Application collects the information about what kind of channels, programs and films are favorite ones while the user watches TV and creates their classification. Using this information the agent creates a profile of the user. Based on his history and profile the agent can give a recommendation about the places he can go, travel or just spend time (See Figure 11). If the user watches programs about Japan or sushi the system can recommend him a restaurant he would like near his place. With the help of internet the application can also show the opening times and special offers of that restaurant. [33]

C. Cross-domain based approaches

Finding similar users and building an accurate neighborhood is an important part of recommending process of collaborative recommender systems. Similarities of two users are discovered based on their appreciations of items. But similar appreciations in one domain dont surely mean that in another domain valuations are similar as well. Users sharing preferences in comedies are not by all means like the same type of horrors (See Figure 12).

D. Peer-to-Peer approaches

The recommender systems with P2P approaches are decentralized. Each peer can relate itself to a group of other peers with same interests and get recommendations from the users of that group. Recommendations can also be given based on the history of a peer. Decentralization of recommender system can solve the scalability problem. [35] [36]
E. Cross-lingual approaches

The recommender system based on cross-lingual approach lets the users receive recommendations to the items that have descriptions in languages they don’t speak and understand.

Yang, Chen and Wu proposed an approach for a cross-lingual news group recommendations. The main idea is to map both text and keywords in different languages into a single feature space, that is to say a probability distribution over latent topics. From the descriptions of items the system parses keywords than translates them in one defined language using dictionaries. After that, using collaborative or other filtering, the system gives recommendations to users.[37]

With the help of semantic analysis it’s possible to make a language-independent representation of text. Pascal Lops purposed an approach with a recommender system called MARS. Based on Word Sense Disambiguation algorithm that exploits online multilingual lexical database as sense repository the engine assigns right meaning to the words avoiding synonyme problems.[38]

Similar approach was purposed by Murad Magableh and Antonio Cau. They utilize online WordNet lexical ontology with EuroWordNet multilingual lexical resource to avoid synonym problem and build more accurate relations between tags in folksonomies. Every time a user adds a tag to an item synonyms of the tag-word are discovered by the collaborative tagging system and added to the tag using the WordNet ontology. EuroWordNet builds relations between tags in different European languages, which let recommender systems compare and classify items with tags in different languages. [27]

Cross-lingual recommender systems break the language barrier and gives opportunities to look for items, information, papers or books in other languages.

IV. CHALLENGES AND ISSUES

A. Cold-start

It’s difficult to give recommendations to new users as his profile is almost empty and he hasnt rated any items yet so his taste is unknown to the system. This is called the cold-start problem. In some recommender systems this problem is solved with survey when creating a profile. Items can also have a cold-start when they are new in the system and havent been rated before. Both of these problems can be also solved with hybrid approaches.

B. Trust

The voices of people with a short history may not be that relevant as the voices of those who have rich history in their profiles. The issue of trust arises towards evaluations of a certain customer. The problem could be solved by distribution of priorities to the users.

C. Scalability

With the growth of numbers of users and items, the system needs more resources for processing information and forming recommendations. Majority of resources is consumed with the purpose of determining users with similar tastes, and goods with similar descriptions. This problem is also solved by the combination of various types of filters and physical improvement of systems. Parts of numerous computations may also be implemented offline in order to accelerate issuance of recommendations online.

D. Sparsity

In online shops that have a huge amount of users and items there are almost always users that have rated just a few items. Using collaborative and other approaches recommender systems generally create neighborhoods of users using their profiles. If a user has evaluated just few items then its pretty difficult to determine his taste and he/she could be related to the wrong neighborhood. Sparsity is the problem of lack of information. [12]

E. Privacy

Privacy has been the most important problem. In order to receive the most accurate and correct recommendation, the system must acquire the most amount of information possible about the user, including demographic data, and data about the location of a particular user. Naturally, the question of reliability, security and confidentiality of the given information arises. Many online shops offer effective protection of privacy of the users by utilizing specialized algorithms and programs.

V. CONCLUSION

Recommendation systems have definitely opened new options of searching and filtering information. Internet stores have accelerated profits, music lovers have discovered new artists unknown to them before, and tourists might take a look to new interesting places. Having all these options available, the customers save their time in multiple numbers. And this is the minor part of the beneficial influence of recommendation system on the clients. At the same time, there are some shortcomings, limits, and defects. Some of them were discussed above. Numerous improvements are required in the sphere of development of user’s model, of dapper semantic analysis of information, and of acceleration and polishing of recommendations. Recommendation systems are not limited by only computers and mobile devices, but they can also open new security capabilities while embedded into automobile industry, and overall, into devices of everyday use. This, in turn, would require development of more specified recommendation systems. All these facts make us sure that these systems will be promising and topical for long time. And we are just in the initial stage of their development.

REFERENCES

