Volume 4, Issue 2, February 2014 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Recommendation Systems for Predictive Analytics -An Insight

Monika Singh, Monica Mehrotra
Department of Computer Science,
Jamia Millia Islamia,
New Delhi, India

Abstract-- Predictive analytics is an analytical technique that is forward thinking in nature. Recommending some items or services to the users or predicting about some event before it happens has become possible and is purely scientific in terms of the procedure employed. The ample user generated data has lead to the opportunity to analyze the domain of predictive analytics at a depth never explored before. Huge item sets can be recommended online as compared to the brick-and-mortar stores. This paper is an attempt to closely study the development of recommendation systems (RS), types of RS, application areas of RS, goals, techniques, evaluation of RS, user modeling and human computer interaction (HCI). It also presents the pointers to the emerging and ongoing research areas in this field.

Keywords: Predictive Analytics, Recommendation Systems, Information Overload, Content based Filtering, Collaborative Filtering, Hybrid approach for RS, Human Computer Interaction (HCI), User Modeling, Evaluating RS.

I. INTRODUCTION

The outburst of continuously generated gigantic data can transform many science and engineering disciplines as effective mining of the data can lead to development, which is beneficial, not only to the individuals but also the society and the nation as a whole. Utilizing the precious data along with the effective techniques of predictive analytics will be useful in obtaining fruitful results in different domains like market analysis, disease control, social network analysis, just to name a few.

With easy access of Internet to more than a billion people across the globe generates huge amount of data every day. Though some information can be implicitly drawn from the data, predictive analytics deals with the data for some explicit information. Getting the information explicitly requires analyzing data and interpreting thereof.

With many competitors in the market, attracting a large number of customers towards the company's products has always been a challenging task. Earlier the focus was to display the products at the E-commerce sites and customer had to search for the items/services she needed. As there is tremendous increase in services, products and brands, the customers' capability to search for an item from such huge set is vastly outstripped. Searching and surveying from a huge black box was a turn off for them. This was not an effective way to attract the customers.

Predictive analytics encompasses a variety of techniques from statistics, machine learning, and data mining that analyze current and historical facts to make predictions about future, or otherwise unknown events. Predictive analytics is used for unleashing the relevant information from huge datasets that are available. These predictions are useful to various government organizations, companies in finance or marketing, education and advertisement etc. in formulating decisions.

The use of predictive analytics in recommendation systems is stupendous, where users can be recommended the news articles, products, movies/shows based on their previous search history and interests. Similarly users can be recommended some communities/groups and friends or some applications. Users are recommended for buying specific products based on the analysis of their browsing actions. The concept of recommender system flourished in mid 1990's, which acts like a virtual friend for suggesting the type of product or service to the customers. Recommending a user accurately about what she needs is a challenge because of the dynamism over the web.

II. BACKGROUND AND RELATED WORK

Predictive analytics –how exactly is a prediction drawn?

A huge amount of literature is present in RS subject. The work on the applicative areas like travel based RS, Television RS (through Electronic Program Guide EPG), books RS, music RS and movie RS is present. There is a substantial research in formulating the techniques employed in framing recommendations, quality of recommendations and protecting the RS against malicious users attack. Work on User modeling and HCI are also primarily focused.

A. Development in RS and its Goals

RS is used to give personalized recommendation as an output or guide the users in a personalized way about the useful items from a huge set of possibilities [9]. RS helps individuals who lack required personal experience to evaluate grand number of alternatives. It has been observed that people do rely upon the recommendations in making decisions.

IR or search engines returns the results of querying on the basis of "matching" and the system returns all items that are matching to query. RS and Information retrieval systems differ by the semantics of user interactions.The results of RS can be interpreted as recommendation worthy of consideration while the result of an IR can be iterpreted as the result of user's query [9].

Searching an item from E-commerce sites was a turn off for the users, as the problem of information overload arose. Schwartz and Ward describe the freedom of excessive information as 'Misery inducing Tyranny' [18]. The need of filtered information from a huge set of available alternatives thus arose. So, E-commerce services introduced RS to make users take decisions based on what they want. It became tremendously useful for the companies to increase the sales and for customers to get narrowed but effective search results. Mass media companies are offering RS recommendations as a part of their service.

According to [12] the goals of RS are shown in Fig 1.

Fig 1: Goals of RS

The rage of deploying best RS was seen when Netflix, an online movie rental service has challenged to give one million dollar as a prize to the one who can improve their RS substantially.

Many attempts for making better recommendations are found in the literature. Adomavicius [1] suggested incorporating contextual information to the recommendation process. This multidimensional recommendation model makes recommendations based on multiple dimensions, profiles and aggregation hierarchies.

B. Techniques employed in Predictive Analytics

The recommendations provided by the RS can be personalized or non-personalized. Non-personalized recommendations include recommendations like top 20 songs, or top 5 actors, which is based on overall liking and demand of people. Though RS do provide non-personalized recommendations, it typically focus on giving the personalized recommendations, which means the core recommendation techniques are customized to the individual.

Users are asked to rate the items. This serves as explicit mode of getting to know the preference of items. In implicit mode users' behavior is captured based on how she navigates. Even though she is not interested in buying the item, mere visiting shows her interest [12]. The responses are stored in the recommender databases and next recommendations are inferred by consulting the databases.

Recommendation systems can require utility matrix for its modeling. Users and items are entities of the matrix. Users are asked to rate products, movies or news articles and based on their ratings matrix is filled. In recommendation systems users and items can be seen as two classes of entities. The users and items can be placed as the rows and the column in a utility matrix. To fill the matrix, users are asked to rate the item. The users rate the items. The utility matrix thus obtained is a sparse one as most of the entries are "unknown" [16]. The goal of a recommendation system is to fill the blanks in the utility matrix based on predictions. Though it is not desirable to fill each blank. Similarity between the items can be obtained by checking the similarity in ratings of the items by the users who have rated both items. Similarity between users is obtained through the similarity in the choice of items.

RS designing uses results from various computer science disciplines like machine learning, information retrieval, data mining, HCI. Machine learning and data mining, subfields of artificial intelligence, allow the computer to learn how to optimally perform a certain task using examples, data or past experiences [16].

The RS uses knowledge deficient data or knowledge dependent data. The data obtained from ratings and evaluation of items is knowledge deficient whereas data obtained from ontological descriptions of the users or items is knowledge dependent. Harnessing knowledge from ontology is a recent trend in RS. Adomavicius [2] work focuses on using domain knowledge in giving recommendations.

RS are implemented through search engines indexing on primarily unstructured data. The architectures defined in section **D** elaborate the techniques employed in great depth.

C. Applicative Areas of Recommendation Systems

Recommender systems are applied in a variety of applications like, music, news, books, research articles, movies, social tags, and products in general. However, there are also recommender systems for restaurants, financial services, life insurances etc [15].

Recommendation system for a book can be designed in a manner, which can capture the details of a book like its name, its author, publishing house and genre. Based on these details books similar to users' choice can be recommended.

D. Recommendation System Architectures

Burke [9] provided the first overview of the six different approaches of RS: Content-based, Collaborative filtering, Demographic, Knowledge-based, Community-based, Hybrid recommender systems.

Each has its own shortcomings. Knowledge based approaches has knowledge engineering bottleneck [16]. Collaborative filtering techniques suffer from cold start problem, which means it won't be able to accurately recommend to a user for whom it hasn't gathered enough information. Not even a single recommendation for an item can be given until someone has rated it [19]. If the number of users who rated items is relatively small compared to the number of items in the database, the significant similarity between the users can't be seen. The current generation RS still requires further improvements to make recommendations more effective [2].

1) Content based Filtering: It primarily focuses on the properties of the items. Similarities in items can be obtained by looking at the similarities in the characteristics of items. A profile is created for each item which is the record(s) reflecting the characteristics of the item. As the user goes on rating the items, the content-based recommender system creates a user profile based on his interests. The 'Profile Learner' generalizes the users' preferences to create the user profile. While recommending a new item to a user, the content (properties) of the item is matched against the user profile. If any new item's profile matches up with the user's profile, the item is recommended to the user. Information filtering and Information retrieval are the ancestors of content based filtering.

It is relatively much simple to do item profiling in contrast to finding out characteristics of documents and images. Characteristics of a document or an image are not apparent. The 'Content Analyzer' [12] analyzes the content of items for feature extraction. The widely used algorithm is TF.IDF. To create a profile for a news article or blog, first stop words are eliminated and then TF.IDF score for each word is calculated. The words with highest TF.IDF score characterize the document [16].

All of the content-based approach sees the documents by the 'important words' in the document. Example being Fab, which recommends to the users the web page content with the 100 most important words [4].

Pandora uses the power of Music Genome project that predicts what the user is going to like next based on what they listened earlier. It does so by content based filtering.

Content based RS suffers from serendipity problem which means that content based RS can recommend with limited degree of novelty. This RS is not able to recommend items to a new user as his profile is not created yet [12].

2) Collaborative Filtering: It recommends items on the basis of relationship between users or items. The process of filtering for information or patterns using techniques involving collaboration among multiple agents, viewpoints, data sources, etc. It can use concepts from clustering and similarity search.

Last.fm recommends songs by keeping record of the bands and tracks the user listens to. It plays the songs present in his library and in addition plays songs often played by users of similar interests. It does so by collaborative filtering technique.

Collaborative filtering is considered to be the most popular and widely implemented technique in RS. Most commonly used algorithms for measuring similarities in users or items in collaborative filtering include k-nearest neighbor approach and Pearson Correlation.

Although collaborative filtering is mostly of the times used to find correlation among users, it may also be used to find correlation among the objects rated [17]. A collaborative RS algorithm uses matrix factorization.

3) Demographic: It provides recommendations on the basis of demographic profile of the user [12]. Products are recommended based on combining the ratings of users based in different demographic groups. This recommender system is based on finding the relationship between an item and types of people who liked that item [17].

Example includes recommending web pages, new articles or products to the people based on demographics like their country, language, religion, gender or age.

4) Knowledge-based: Knowledge based RS builds knowledge domain and gives appropriate recommendations to users based on user queries by learning from the knowledge domain. A knowledge-based recommender suggests products based on inferences about a user's needs and preferences.[1]

Experiments by Burke [9] shows the knowledge component combined with other RS can be used to build some best

[1] [1] http://en.wikipedia.org/wiki/Recommender_system

hybrid RS.

5) Community based RS: The recommender system is based on the observation that recommendations from friends are more reliable and trusted than some random recommendation by an unknown. Community based RS captures the preferences of the user's friends and based on that, it recommends.

6) Hybrid Approach: A hybrid recommender system combines collaborative based filtering and content based filtering for effectiveness. It helps avoiding the limitations of collaborative based filtering and content based filtering [2]. The recommendations given by hybrid RS are empirically proved to be more accurate than the pure approaches[1].

Balabanovic and Shoham [4], discuss the hybrid approach of RS for the web called Fab, which combines collaborative and content based filtering systems to incorporate the advantages of both methods and eliminating many weaknesses found in each approach. Claypool [10] presented a filtering approach combining collaborative and content based filters in an online newspaper. Netflix, an online movie rental RS is also based on hybrid approach.

E. User Modeling and Human Computer Interaction

Recommender Systems uses data of items, users or transactions. It has become the need of the hour to do user modeling to be used by RS. A user should be recommended the items according to his taste. This can be done when his personalized information is stored and inferences can be made using that. A User profile can be learnt based upon collaborative, content based and demographic filtering.

Fisher [11] defines user models as models residing in a computational environment. Berkovsky [5] describes RS as a tool that generates recommendation by building and exploiting user models. Adomavicius [2] states including the improved modeling of users and items as an important extension to current RS. To maintain a user's profile system needs relevant information about users interests [15]. Fisher [11] states, "The ultimate objective of user modeling is that it has to be done for the benefit of users". User modeling adaptive to change are also a new research avenue. Adaptive hypermedia (AH) systems build a model individual user and adapt it to change during the course of interaction [8].

Computer graphics are more accessible than command-line based displays. Information visualization is a growing area to visually present very large information space [3]. Well-designed graphics are more appealing to users.

F. Evaluating RS

It is very important to evaluate the RS before deploying it. Though majority of literature shows that attempts are on improving the accuracy of RS, work by Wen Wu [20] takes into consideration some other criteria like coverage, serendipity, risk, novelty, adaptability etc.

For many reasons, evaluating RS and its algorithms is difficult [14]. Herlocker [14] describes various factors that are to be considered while evaluating a RS.

III. DISCUSSION AND CONCLUSION

This paper is aimed at exploring RS, as a tool for predictive analytics. The essence of utilizing the tremendous data for predictive analytics is discussed. The evolution of RS along with its major goals is being discussed. It illustrates the need of RS over e-commerce. It touches the boundaries of the RS types. It discusses the applicative areas of RS and some work done in the field of RS applications.

Much research work is in on-going phase about the type of RS and their relative effectiveness and optimal use. It presents the approaches on which recommender systems are built. The goals of RS are discussed along with the problem of 'information overload'.

Herlocker [13] presented an algorithmic framework that breaks collaboration prediction components into components and provided empirical results regarding each component variants. Attempts on building RS based on hybrid approaches can also be seen. To optimize RS results, we saw that contextual information is included in recommendation model. Not much literature is present in the areas of demographic and community based RS.

Lastly HCI an important research constituent is discussed. User modeling, a crucial input in designing a RS is a vector on which research communities are focusing. Along with these issues on evaluating RS and its effectiveness is also a rage among research communities. Recommendations should speak for themselves. If recommendations are correct, it will gain users' trust and acceptance. For building the trust among users evaluating the RS for its correctness is also an important research area. This is an attempt to explore the scientific approaches and techniques employed to get into the depth of 'how exactly is a prediction drawn'.

REFERENCES

[1] Adomavicius, G., Sankaranarayanan, R., Sen, S., Tuzhilin, A.: Incorporating contextual information in recommender systems using a multidimensional approach. ACM Trans. Inf. Syst. 23(1), 103–145 (2005)

[2] Adomavicius, G., Tuzhilin, A.: Toward the next generation of recommender systems: A surveyof the state-of-the-art and possible extensions. IEEE Transactions on Knowledge and Data Engineering 17(6), 734–749 (2005)

[3] Baeza-Yates, R., & Riberio-Neto, B. (1999). Modern information retrieval (p. 30). Reading, MA:Addison-Wesley Publishing Co.

[4] Balabanovic, M. and Shoham, Y. (1997). Fab: Content-Based, Collaborative Recommendation. *Communications of the ACM* **40**(3), March 97.

[5] Berkovsky,S.,Kuflik,T.,Ricci,F.:Mediationofusermodelsforenhancedpersonalizationin recommender systems. User Modeling and User-Adapted Interaction 18(3), 245–286 (2008)

[6] Billsus, D. & Pazzani, M. J. (1998). Learning Collaborative Information Filters. In *Proceedings of the International Conference on Machine Learning*. Madison, WI: Morgan Kaufmann Publishers.

[7] Billsus, D. & Pazzani, M. J. (1999). A Hybrid User Model for News Classification. In *Proceedings of UM'99*, 99–108. Wien, New York: Springer-Verlag.

[8] Brusilovsky, P.: Methods and techniques of adaptive hypermedia. User Modeling and User- Adapted Interaction 6(2-3), 87–129 (1996)

[9] Burke, R.: Hybrid web recommender systems. In: The AdaptiveWeb, pp. 377–408. Springer

[10] Claypool, M., Gokhale, A., Miranda, T., Murnikov, P., Netes, D., & Sartin, M. (1999). Combining content-based and collaborative filters in an online newspaper. In ACM SIGIR _99 Workshop on Recommender Systems

[11] Fisher, G.: User modeling in human-computer interaction. User Modeling and User-AdaptedInteraction 11, 65–86 (2001)

[12] Francesco Ricci and Lior Rokach and Bracha Shapira,Introduction to Recommender Systems Handbook.

[13] Herlocker, J., Konstan, J., Borchers, A. & Riedl, J. (1999). An Algorithmic Framework for Performing Collaborative Filtering. In *Proceedings of the 1999 Conference on Research and Development in Information Retrieval*.

[14] Herlocker, J.L., Konstan, J.A., Terveen, L.G., Riedl, J.T.: Evaluating collaborative filtering recommender systems. ACM Transaction on Information Systems 22(1), 5–53 (2004)

[15] Montaner,M.,Lo ́pez,B.,delaRosa,J.L.:A taxonomy of recommender agents on the internet. Artificial Intelligence Review 19(4), 285–330 (2003).

[16] Pang-Ning Tan, Vipin Kumar and Michael Steinbach, “Introduction to data mining”.

[17] Pazzani, M. J.: A Framework for Collaborative, Content-Based and Demographic Filter- ing. AI Review 13 (5/6), 393-408. (1999)

[18] Schwartz, B., Ward., A.,: Doing Better but feeling worse

[19] Terveen, L. G. & Hill, W. (2001). Beyond Recommender Systems: Helping People Help Eachother. In Carroll, J. (ed.) *HCI in the New Millennium*. Addison Wesley.

[20] Wen, W., He, L., Yang, J.:Evaluating Recommender Systems. IEEE, 6-61(2012)

