The Personalization Technique for Social Recommender Systems using Machine Learning

Huan Du[1] and Haiyan Chen[2*]

[1]The Third Research Institute of the Ministry of Public Security, Shanghai, China
[2]East China University of Political Science, Shanghai, China

Keywords: Personalization Technique, Social Recommender System, Machine Learning.

Abstract: Recent years have seen the explosive growth of information in the form of web services. Recommender systems suggest items of interest to users based on users' explicit and implicit feedback and also based on the preferences and interests of other similar users/items. As a small step towards extending the footprint of the applications of big data, this paper tries to depict the machine learning techniques to perform Social network analytics that may provide a 360 degree insight into the social network data. The term machine learning aptly denotes that, the system is made to learn by providing necessary inputs and carefully examining the obtained outputs. The applications of machine learning are as diverse as the applications of big data. Adaptive websites, Bio informatics, Computational advertising, Information retrieval, credit card fraud detection, medical diagnosis, Natural language processing, stock market analysis are some areas where machine learning has found its use.

1 INTRODUCTION

Recent years have seen the explosive growth of information in the form of web services. Recommender systems suggest items of interest to users based on users' explicit and implicit feedback and also based on the preferences and interests of other similar users/items. The two basic entities of any recommender system are items, which are the product/services and users, who procure those product/services. A user of a recommender system receives recommendations about items, makes use of those items and also provides opinion about various items. The history of recommender system dates back to early 1990 when certain experimental applications employed filtering mechanisms to provide the item of interest to the user (Allison, 2003). Initially the recommender systems were query based information system more like a search engine (Luo, 2011).With the advent of internet and the World Wide Web, there was endless possibilities of electronic data available to the end users. This paved the way for the recommender system which is a resource that helps to make a choice from infinite possibilities (Xu, 2015). The recommender system

---

*The corresponding author: Haiyan Chen

helps the consumer to narrow down his/her set of choices from the abundant list and also help in discovering new item of interest. The invasive presence of E-commerce (Liu, 2010) in today modern society and the aggressive consumers present three key challenges to the recommender system. The first and foremost is to produce high quality recommendations. Secondly, it is necessary to generate many recommendations per second for millions of customers and products, and the last is to achieve high coverage in the face of data sparsity. Now research is focused on improving the methods of recommending items to users.

As a small step towards extending the footprint of the applications of big data, this paper tries to depict the machine learning techniques to perform Social network analytics that may provide a 360 degree insight into the social network data. The term machine learning aptly denotes that, the system is made to learn by providing necessary inputs and carefully examining the obtained outputs. Machines can learn under different circumstances namely, Supervised, Unsupervised and Reinforcement. Machine learning is a subfield of computer science that evolved from the computational learning theory in Artificial Intelligence. Machine learning algorithms help us to make effective predictions

based on big data, upon which the operational convenience of a business can rely. Various machine learning tasks are categorized depending on the desired output of the machine learned system. The applications of machine learning are as diverse as the applications of big data. Adaptive websites, Bio informatics, Computational advertising, Information retrieval, credit card fraud detection, medical diagnosis, Natural language processing , stock market analysis are some areas where machine learning has found its use.

2 PERSONALIZATION FOR SOCIAL NETWORK

Social networks such as Face book, Twitter, LinkedIn etc pave way for generation of huge amount of diverse data in short period of time. Such social media data require the application of big data analytics to produce meaningful information to both information consumers and 3 data generators. The impact of different big data analytics tools and techniques over processing social network data will be discussed in detail in this section of this paper.

Big data analytics techniques and tool types include all of the following such as Predictive analytics, data mining, statistical analysis, complex SQL, data visualization, artificial intelligence and natural language processing. The analysis of structured and unstructured data from social networks leads to social network analytics (Balabanovic, 1997). Even blogs, micro blogs and wikis contribute to social network analytics data sets. Though there are various sources of information available in social media, we are largely concerned about the user generated contents such as sentiments, images, videos and bookmarks and interactive relationships between people, organizations and products. These two classes of information is utilized in various big data analytics tool such as Hadoop and Map Reduce Framework, Apache Pig, Apache Hive, Jaql, NoSQL etc. When user posted information is used in the analytics approach, it is called as content based analytics and when relationships between entities is used for analytics, it is known as structure based analytics. Social networks consist of millions of connected objects and analysis of data from those objects is computationally intensive and expensive. Hence there are two different approaches that shall be followed. They are parallelization approach and Graph databases approach (Adomavicius, 2005). In

parallelization approach, focus is towards dividing a huge data set into smaller sub sets and utilize the computational power through cloud computing to process the data in a parallel manner. Map Reduce and Pregel from google are pioneer in this approach. However, lots of open source initiatives in the form of Hadoop are gaining popularity in the social network analytics. Spark and Hama are also registering their market share in the research of social network data. (Burke, 2007)

Map reduce framework consists of Map phase and Reduce phase which uses Key/Value pairs and Key-Value List pairs respectively. Any mapreduce application contains various hotspots such as Input Reader, Map, Partition, Compare, Reduce, Output Writer. Application of Map reduce is considered to enable the scalability of social networks, for the determination of graph based metrics. This application is used to determine the betweenness centrality. The chaining of Mapreduce jobs in social network analytics is carried out for the estimation of shortest paths in a graph. Blocking mechanism is an important part of Map Reduce that deals with machine failures in the application of social network data.

The preprocessor cleans, integrates, selects and transforms the knowledge base of the users and items to relevant user and item data store. Then various types of filters are applied to data stored in these databases. The filtering algorithms can be broadly classified into memory based algorithms and model based algorithms. Recommender system using memory based algorithm learns at a particular instance of time considering all previous instances. After the recommendation, the system immediately knows the result of the prediction and hence uses the feedback for further recommendations. Memory based algorithms use similarity metrics to obtain the similarity distance between two users, or two items and aggregation measures that helps in generating the prediction. Model-based methods use user and item information to create a reference that generates the recommendations. The most widely used model based algorithms are based on Bayesian classifiers, neural networks, fuzzy logic based algorithms, genetic algorithms and singular value decomposition techniques.

3 MACHINE LEARNING FOR SOCIAL NETWORK

Machine learning techniques, as implied by the term, is the process of inculcating knowledge to any machine like, PC, laptop or mobile devices to learn about a system with a set of input /dependent variables and the desired output. Any machine can perform learning under three modes. They are Supervised, Unsupervised and Reinforcement learning techniques. Normally, machine learning techniques are employed in any system to carryout and produce results as part of predictive analytics and forecasting methods. Any machine learning techniques will be classified under the categories of Decision tree based, linear and logistic regression based and neural network based. Many organizations have kick started to utilize the impact of social media data in the decision making process. When social media data is utilized for such a critical decision making, it becomes necessary to process the huge datasets obtained from social networks using machine learning techniques. This will help organizations to foresee certain situations and decide based on the output of the social media analytics. The key aspect of any machine learning technique is iteration. This iterative aspect will make the system to independently adapt to new sets of input as they will be continuously subjected to variety of datasets. The advent of new computing technologies like big data have created a revolution in the machine learning domain, that complex mathematical calculation can be applied to heterogeneous huge datasets.

Machine learning algorithms that have played a major role in social media analysis include Decision tree learning, Naïve Bayes, Nearest Neighbor classifier, Maximum Entropy method, Support vector machine(SVM), Dynamic Language Model classifier, linear regression and logistic regression, Simple logistic classifier, Bayes Net and Multilayer Perceptron.

Upon carrying out literature research, it becomes quite evident that considerable amount of work has been carried out in the social network analytics field utilizing the decision tree learning mechanisms. Decision tree learning uses decision trees to predict the values of a target variable and relate the same to the observations of that variable. Two types of trees can be built using a decision tree learning mechanism namely, classification trees and regression trees. Classification trees provide finite set of values to the target variables and regression trees provide continuous values to the target variable. In social network analysis, decision tree learning has been utilized to profile users based on their relationship with other users, and depending upon the decision tree obtained, clustering of users can take place. Two important algorithms that employ top down, greedy search through the space of decision trees are ID3 and C4.5. The working principle of ID3 algorithm is that it learns decision trees by constructing them top down and starts at the top of the tree and then decides on the attribute to be tested. C4.5 is an extension of ID3 algorithm and it builds decision trees based on the concept of information entropy and a set of training data. Decision tree has been used to obtain the rules that govern the relationships among users in the online social network. These decision trees are also used to discover interesting patterns among the users.. Gradient Boosted Decision Trees (GBDT) is used in classification of users based on certain attributes in social networks. GBDT is proved to provide much smaller decision trees and reduced decoding compared to Support Vector Machines (SVMs).

4 EVALUATIONS

Evaluation of recommender system quality implies measurement of the quality attributes that a recommender system is preferred to have, for instance its functionality, maintainability, usability and so on. Various recommender algorithms, their advantages and limitations are summarized. Evaluation of recommender systems depends on values of the measurement carried out. The main objective of the recommender system is to improve customer experience through personalized recommendations and also achieve the sellers' interest in promoting the product.

In empirical research methods, data is collected to answer a particular research question. Empirical research methods can be divided into two categories, quantitative research methods and qualitative research methods. In quantitative research methods, data collected are in the form of numbers (numerical

data) and patterns and relationship in the data are identified and analyzed using statistical methods. In qualitative research methods, data collected are qualitative data such as text, images, sounds drawn from observations, interviews and documentary evidence, and the data is analyzed using qualitative data analysis methods. An offline experiment of recommender system is performed using historical dataset. Using this dataset the behavior of the user is simulated. Offline experiments help to understand the behaviour of various algorithms at a low cost. The scalability of the algorithm can be measured by increasing the size of the dataset. Certain experimental constrain can be embedded in the dataset. The main advantage of offline algorithm is that it is cheaper and it does not require the interaction of the real users. The major disadvantage of offline algorithm is the recommender's influence on users' behaviour cannot be determined and also recommender's characteristics like serendipity and diversity cannot be determined. Online experiments are deployed large scale application where the users are unaware about the experiment being conducted. Online experiments are designed to learn about user behaviour characteristics. The performance of the recommender system varies on many user dependent factors such as users' intent, users' context and various characteristics of the graphical user interface of the recommender system. Online Experiments help to test multiple algorithms by submitting the user request to different alternative recommendation engine.

5 CONCLUSIONS

As a small step towards extending the footprint of the applications of big data, this paper tries to depict the machine learning techniques to perform Social network analytics that may provide a 360 degree insight into the social network data. The term machine learning aptly denotes that, the system is made to learn by providing necessary inputs and carefully examining the obtained outputs. The applications of machine learning are as diverse as the applications of big data. Adaptive websites, Bio informatics, Computational advertising, Information retrieval, credit card fraud detection, medical diagnosis, Natural language processing , stock market analysis are some areas where machine learning has found its use.

REFERENCES

Lloyd Allison, "Types and classes of machine learning and data mining", *Proceedings of the 26th Australasian Computer Science Conference*, Vol.16, Page:207-215, 2003.

Luo, X., Xu, Z., Yu, J., and Chen, X. 2011. Building Association Link Network for Semantic Link on Web Resources. *IEEE transactions on automation science and engineering*, 8(3):482-494.

Xu, Z. et al. 2015. Knowle: a Semantic Link Network based System for Organizing Large Scale Online News Events. *Future Generation Computer Systems*, 43-44:40-50.

Guojin Liu, Ming zhang, Fei Yan, " Large Scale social network analysis based on MapReduce", *International conference on computational aspects of social networks*, Page:487-490, 2010

Balabanovic,M., Y.Shoham (1997) Fab: Content-based, Collaborative Recommendation, *Communications of the ACM*, vol. 40, pp. 66-72.

Adomavicius, G., Sankaranarayanan, R., Sen, S., Tuzhilin, A. (2005) Incorporating Contextual Information In Recommender Systems Using A Multidimensional Approach, *ACM Transactions on Information Systems*, vol. 23, pp.103–145.

Robin Burke (2007) Hybrid Web Recommender Systems, In: *The Adaptive Web-Lecture notes in Computer Science*, Springer Verlag, pp. 377-408.