A Click stream-Based Recommendation System with Machine Learning

Anurag Singh[1],
Research schooler
AKS University Satna(M.P) ,India
[1]*Email* singhanurag.jbp@gmail.com

Dr Subhadra Shaw
Assistant Professor of Computer Science
AKS University Satna(M.P) ,India

Abstract: *With the increment in the utilization of the Internet, the pace of increment of social networks is getting ubiquitous in recent years. Click-stream data is an important source to enhance user experience and pursue business objectives in e-commerce. The paper uses click-stream data to predict online shopping behavior and target marketing interventions. The paper represents a study of various existing recommendation systems and proposes a recommendation system that utilizes different Machine Learning procedures which results in showing that Random Forest Classifier (RFC) gives the most noteworthy expectation accuracy when contrasted with different procedures.*

Keywords: **Machine learning, recommendation system, naïve Bayes, random forest, cold start problem**

1. INTRODUCTION

Recommender systems were originally defined as ones in which ”people provide recommendations as inputs, which the systems then aggregates and directs to appropriate recipient” [1]. Now the concept has been extended from recommending products to recommending friends, movies, songs, and efforts are increasing to know more about an individual’s personal preferences more clearly and precisely. Leading companies, most notably Amazon, YouTube, and Netflix, have definitively demonstrated their value and have radically transformed what customers expect from any digital experience. Amazon, for example, directly attributes an estimated 35% of sales to its recommender system.[2] Netflix’s VP of Product Innovation Carlos Gomez-Uribe and Chief Product Officer mention in a research paper [3] that the combined effect of personalization and recommendations save more than $1B per year.
Basically, to keep things simple, a recommender system is able to provide suggestions (recommendations) to users, in multiple contexts such as when they are making a choice among a large catalog of items or whenever they want to receive suggestions. Identified features:

- Help to Decide: predicting a rating for a user for an item
- Help to Compare: rank a list of items in a personalized way for a user
- Help to Discover: provide a user with unknown items that will be appreciated
- Help to Explore: give items similar to a given target item

Most of the applications of recommender systems are on e-commerce websites. The site displays a list of recommended items to the end user. A recommender system is defined as a tool that helps users search for something which is related to their tastes or as a strategy of choice for users under complex information environments [4].
Recommender systems handle the problem of overload of information that users find, by providing them with personalized and targeted content. For building this system, distinct approaches have been developed. They can be collaborative filtering, content-based filtering, or hybrid filtering [4].Collaborative filtering recommends items by identifying other users with similar tastes. On the other hand, Content-based filtering recommends elements that are similar in content to products that users appreciated in the past or matched to user attributes. The hybrid method combines both collaborative and content-based techniques.

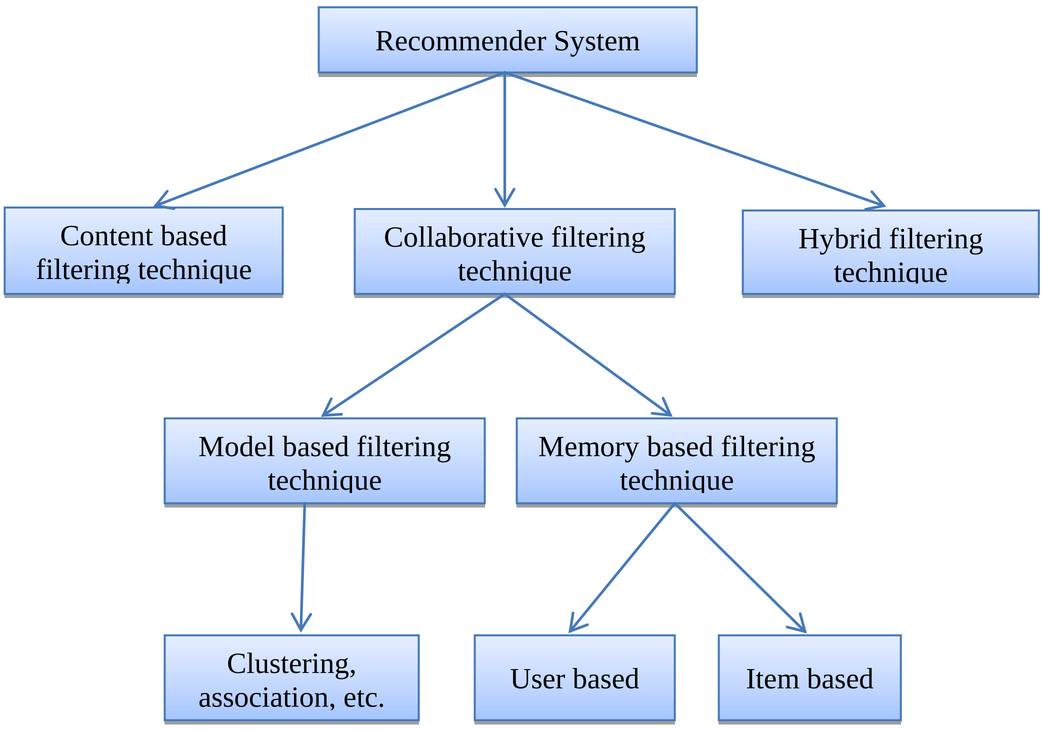


Figure 1: Recommendation systems [4]

This paper presents the existing mechanism and proposed click-stream based recommendation system.

2. LITERATURE REVIEW

The system helps users discover personalized items of their own interests; some platforms, such as Douban and Netflix, rely on the accuracy of the recommendation system to Maximize user stickiness and satisfaction. The successful application of the recommendation system is a powerful means to improve the profitability of commercial platforms [5][6]. Therefore, a recommender system that can provide users with accurate and reliable recommendations has great commercial value.

Typically, recommender systems are based on collaborative filtering, which calculates the similarity based on the rating profiles of two users and recommends Items that are highly rated by users [7]. Collaborative filtering is widely used in e-commerce represented by Amazon by virtue of its excellent efficiency system. However, collaborative filtering technology also has obvious disadvantages—it needs enough historical evaluations to calculate the similarity between users.

Most users will only rate a very small number of items that can reach millions, and the absence of common reviews may reduce the accuracy of the recommendation system [8]. With the increasing popularity of online social platforms, recommendation algorithms that incorporate social network factors have emerged as the times require. These methods are combined in the recommended.

Combining rating information and social relationship information, and assuming that the user's tendency will be influenced to a certain extent by other users who have established social relationships with the user affected by households. Previous studies have shown that recommender systems combined with social information can make more accurate recommendations, and can target cold start phenomenon is particularly effective [9].

Social recommender systems can be viewed as a combination of traditional score-based recommender systems and social networking services. They all depend heavily on the use of User profile. However, due to the openness of the scoring system and social networks, the recommendation represented by collaborative filtering is very vulnerable to malicious attacks. In the recommendation system, malicious attackers can inject false rating information and user relationship information to achieve attack push. This attack is called a shilling attack [10].

As an important web service, the social network has a wide range of applications, such as collaborative work, collaborative service quality rating, resource sharing, and discovery of new friends [11]. The concept of social networks has not only stayed in theory but has also entered the field of practice. As long as it is possible to describe and analyze the user's relationships, social networking applications and online services can be found and defined. Network theory is the study of the representation of relationships between nodes [12]. A social network is a network formed by a set of specific connections connecting participants. Social network theory suggests that the position of participants in a network will influence their access to resources, friends, and information [13].

In recent years, more and more studies have introduced social network analysis theory into business decision-making and business recommendation. Much has been written about social applications of network analysis. For example, DeMeo et al. [14] developed a framework based on social network analysis to recommend similar users and resources. Zhen et al. [15] applied the concept of a social network to develop a recommender system for end-to-end knowledge sharing. Business platforms increase sales by studying how to use social relationships to improve customers' purchasing decisions [16] and have a close relationship with other users. Users with social connections are believed to be able to exert greater influence over others [17]. In a previous study, Carchiolo et al. [18] found that social Relationships between friends and friends-of-friends in the network play a key role when it comes to trust and reliable information. Albert and Barabasi [19] pointed out that a social network is a complex network, with social entities as nodes and links to show the relationship. The nature of social networking focuses more on the relationships between the components that make up the network rather than on its own structure. Assessing the closeness of social relationships, it can be concluded that the social grade or score of a social node in a social network indicates the strength of influence or trust [20]. In practice, Wang and Chiu [21] combine social intimacy and social reputation to discover trusted online auction sellers. In the study of advertising targeting target groups, the research of Kempe et al. [22] shows that the information disseminated by people with higher intimacy will have a greater impact on other nodes in the network influences. András A. et. Al[24] In this paper, the Authors overview the main online learning methods for classification and regression, the most important machine learning tasks. They highlight the most important ideas, including linear models, gradient descent, and tree-based methods. Furthermore, due to the infinite nature of the data stream, online classifiers and regressors are best evaluated by the prequential method, which they also described in this paper. Marie Al-Ghossein, et al.[25] authors reviewed SBRS, underline their relation with time-aware RS and online adaptive learning, and present and categorize existing work that tackle the corresponding problem and its multiple facets.

3. PROPOSED SYSTEM

Very little information was provided beyond item and session ids, hence a supervised feature-based classification method is used instead of collaborative filtering directly. The system is modeled with classification algorithms:

1. Logistic Regression
2. Random Forest
3. Gradient Boosted Trees
4. Neural Networks

All the models used characteristic features extracted from each session that is being classified.
There are 3 types of features used in model building :

1. Session Statistics.
2. Global Features.
3. Time Features.

The process flow of the system is shown below:

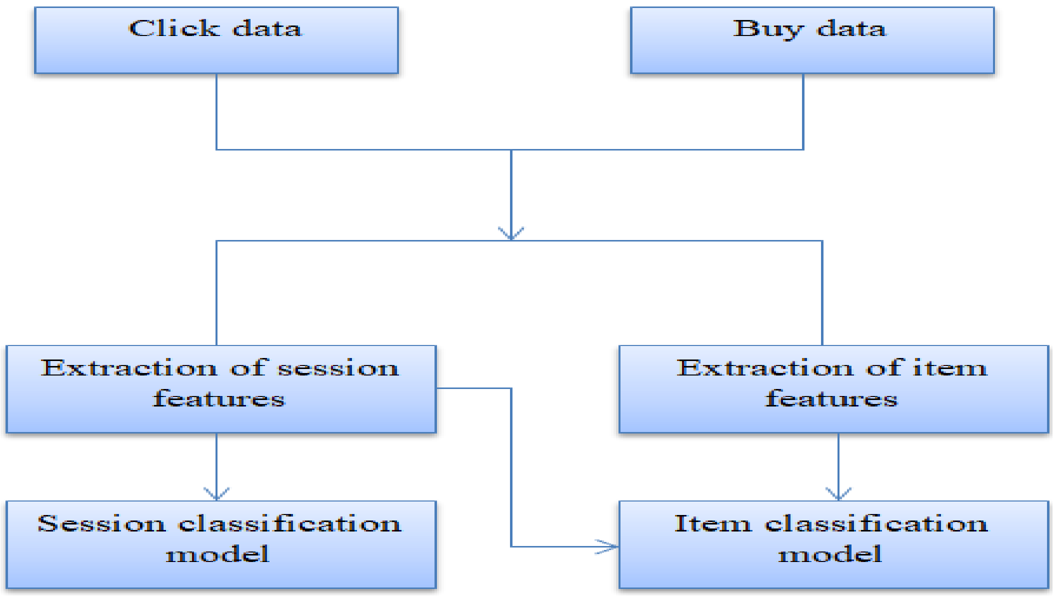


Figure 2: Proposed flow

General model for machine learning process is as follows:

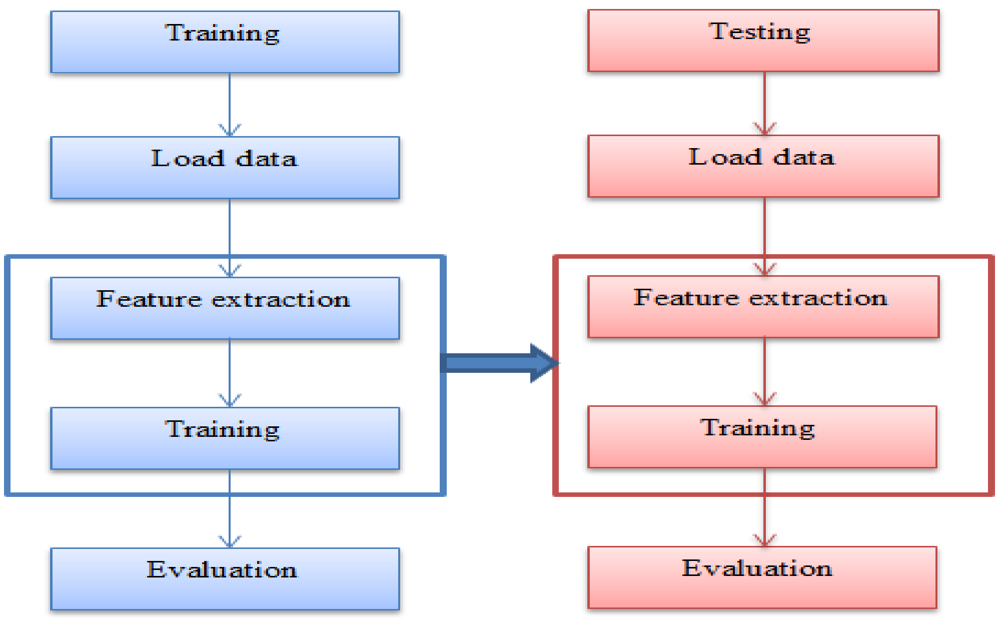


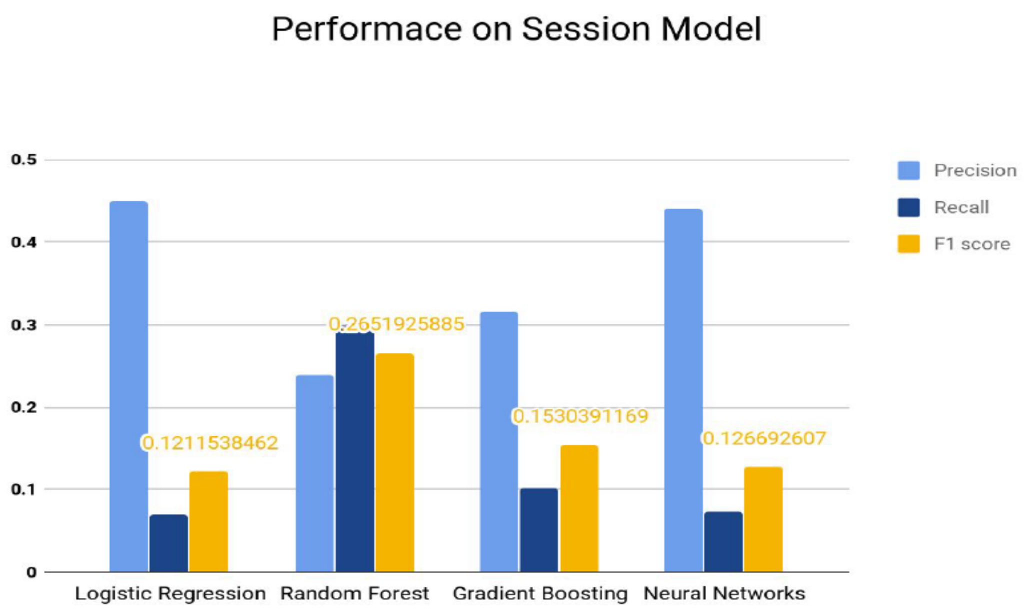
Figure 3: Proposed ML Model

Result

This dataset was constructed by yoochoose. The dataset contains a collection of sessions from a retailer, where each session is encapsulating the click events that the user performed in the session. The work is based on a large scaled dataset of over 9.2million user-item click sessions. There are two files in the training dataset: Dataset is logs of e-commerce website

1. Clicks Data
2. Buy Data

A comparative chart of results is shown below:



Best Results were given by Random Forest.

4. CONCLUSION

Cold start problem is very important in the case of a recommendation system. This can be solved by the proposed system by using various feature extraction methods with machine learning. The dataset we use is yoochoose. It is found that the proposed model performs well. The work can also be extended to solve shilling attacks and enhance the machine learning model.

Future Work

Further research can be performed to enhance the results, such as creating more than one training and testing dataset to be evaluated and gathering the average of the results. The execution and results assessment show the adequacy of the proposed model consequently the following future extension for the work is planned in near future. The proposed system is currently being developed in an experimental environment soon real-world data and applications are used.

REFERENCES

[1] *P. Resnick and H. R. Varian, Recommender Systesm. Communications of the ACM, (1997).*

[2] *I. MacKenzie, C. Meyer, and S. Noble, How retailers can keep up with consumers. McKinsey and Company.*

[3] *C. A. Gomez-Uribe and N. Hunt, The Netflix Recommender System: Algorithms, Business Value, and Innovation. Netflix.Inc, (2015).*

[4] *F. Isinkaye, Y. Folajimi, and B. Ojokoh, "Recommendation systems: Principles, methods and evaluation," Egyptian Informatics Journal, vol. 16, no. 3, pp. 261–273, (2015).[Online].Available:https: //www.sciencedirect.com /science/ article/pii/S1110 86 6515000341*

[5] *Recommendation systems: Principles, methods and evaluation-(2011).*

[6] *Pup, Chen L, Hu R. A user-centric evaluation framework for recommender systems. In: Proceedings of the fifth ACM conference on Recommender Systems (RecSys'11), ACM, New York, NY, USA; (2011). p. 57–164.*

[7] *X. Su, T.M. Khoshgoftaar, A survey of collaborative filtering techniques, Adv. Artif. Intell. (2009) 4.*

[8] *J.B. Schafer, D. Frankowski, J. Herlocker, S. Sen, Collaborative filtering recommender systems, in: The Adaptive Web, Springer, (2007), pp. 291–324.*

[9] *H. Gao, J. Tang, H. Liu, gscorr: modeling geo-social correlations for new check-ins on*

location-based social networks, in: Proceedings of the 21st ACM International Conference on Information and Knowledge Management, ACM,(2012), pp. 1582–1586.

[10] *S.K. Lam, J. Riedl, Shilling recommender systems for fun and profit, in: Proceedings of the 13th International Conference on World Wide Web, ACM, (2004), pp. 393–402.*

[11] *J. Domingo-Ferrer, A. Viejo, F. Sebe, U. Gonzalez-Nicolas, Privacy homomorphisms for social networks with private relationships, Computer Networks 52 (2008) 3007–3016.*

[12] *C. Kiss, M. Bichler, Identification of influencers--measuring influence in customer networks, Decision Support Systems 46 (1) (2008) 233–253.*

[13] *P. Van Baalen, J. Bloemhof-Ruwaard, E. van Heck, Knowledge sharing in an emerging network of practice, European Management Journal 23 (2005) 300–314.*

[14] *P. DeMeo, A. Nocera, G. Terracina, D. Ursino, Recommendation of similar users, resources and social networks in a social internetworking scenario, Information Sciences 181 (7) (2011) 1285–1305.*

[15] *L. Zhen, Z. Jiang, H. Song, Distributed recommender for peer-to-peer knowledge sharing, Information Sciences 180 (18) (2010) 3546–3561.*

[16] *Y.A. Kim, J. Srivastava, Impact of social influence in e-commerce decision making, Proceedings of the ninth international conference on Electronic commerce, ACM, New York, NY, USA, (2007), pp. 293–302.*

[17] *K.O. Lee, N. Shi, M.K. Cheung, H. Lim, C.L. Sia, Consumer's decision to shop online: the moderating role of positive informational social influence, Information Management 48 (6) (2011) 185–191.*

[18] *V. Carchiolo, A. Longheu, M. Malgeri, Reliable peers and useful resources: searching for the best personalised learning path in a trust- and recommendation-aware environment, Information Sciences 180 (10) (2010) 1893–1907.*

[19] *R. Albert, A. Barabasi, Statistical mechanics of complex networks, Reviews of Modern Physics 74 (47) (2002) 47–97.*

[20] *J. Srivastava, N. Pathak, S. Mane, M.A. Ahmad, Data mining for social network analysis, IEEE International Conference on Data Mining, Hong Kong,(2006), pp. 18–22.*

[21] *J.C. Wang, C.C. Chiu, Recommending trusted online auction sellers using social network analysis, Expert Systems with Applications 34 (3) (2008) 1666–1679.*

[22] *D. Kempe, J. Kleinberg, E. Tardos, Maximizing the spread of influence through a social network, Proceedings of ACM SIGKDD'03, (2003), pp. 137–146.*

[23] *Marie Al-Ghossein, Talel Abdessalem, and Anthony Barré., Dynamic local models for online recommendation. In Companion Proc. of TheWebConf. 1419–1423., (2018)*

[24] *András A. Benczúr, Levente Kocsis, and Róbert Pálovics.. Recommender systems over data streams. In Encyclopedia of Big Data Technologies, Sherif Sakr and Albert Zomaya (Eds.). Springer, 1–9. (2019).*

[25] *Marie Al-Ghossein, Talel Abdessalem, and Anthony Barré , A Survey on Stream-Based Recommender Systems, ACM Computing Surveys, Vol. 54, No. 5, Article 104. Publication date: May(2021)*