Recommendation System based on Prediction of User Behaviour with Hybrid Approach: A Survey

Dharti Akkawar[1] Dhanshree Kaurase[2] Sufiyan Dagli[3] Ketan Gadwale[4] Prof. A. R. Mune Sir[5]
[1,2,3,4]Student [5]Professor
[1,2,3,4,5]Department of Computer Science & Engineering
[1,2,3,4,5]Prof. Ram Meghe Institute of Technology & Research, Badnera, India

Abstract— Nowadays, the usage of e-commerce is growing day by day, so online users are also rising. Every user spends their most of the time on e-commerce websites and their behavior is different from one another. E-commerce has become very competitive so, knowing user's behavior has become prior concern. Most of the e-commerce platforms are using either content based or collaborative approach to predict user behavior. Collaborative approach has become obsolete. Many e-commerce websites mainly rely on the content based approach. But for better recommendation, it is important to make use of the user data as well. We propose design and implementation of hybrid system by K-mean algorithm and Hidden Markov Model. Machine learning regression algorithms are used to fetch user's priorities and clustering of data through K-mean algorithm.

Key words: Content Based, Collaborative Approach, K-Mean Algorithm, Hidden Markov Model, Clustering

I. Introduction

A. *Recommendation System*

Recommender systems typically produce a list of recommendations in one of two ways – through collaborative filtering or through content-based filtering (also known as the personality-based approach).[1]Collaborative filtering approaches build a model from a user's past behaviour (items previously purchased or selected and/or numerical ratings given to those items) as well as similar decisions made by other users. This model is then used to predict items (or ratings for items) that the user may have an interest in. Content-based filtering approaches utilize a series of discrete characteristics of an item in order to recommend additional items with similar properties.

Recommender systems are a useful alternative to search algorithms since they help users discover items they might not have found otherwise. Of note, recommender systems are often implemented using search engines indexing non-traditional data.

II. Approaches

A. *Content Based Approach*

Content-based filtering methods are based on a description of the item and a profile of the user's preferences. In a content-based recommender system, keywords are used to describe the items and a user profile is built to indicate the type of item this user likes. In other words, these algorithms try to recommend items that are similar to those that a user liked in the past (or is examining in the present). In particular, various candidate items are compared with items previously rated by the user and the best-matching items are recommended. This approach has its roots in information retrieval and information filtering research.

To create a user profile, the system mostly focuses on two types of information:

1) A model of the user's preference.
2) A history of the user's interaction with the recommender system.

Basically, these methods use an item profile (i.e., a set of discrete attributes and features) characterizing the item within the system. The system creates a content-based profile of users based on a weighted vector of item features. The weights denote the importance of each feature to the user and can be computed from individually rated content vectors using a variety of techniques. Simple approaches use the average values of the rated item vector while other sophisticated methods use machine learning techniques in order to estimate the probability that the user is going to like the item.[2]

Direct feedback from a user, usually in the form of a like or dislike button, can be used to assign higher or lower weights on the importance of certain attributes.

A key issue with content-based filtering is whether the system is able to learn user preferences from users' actions regarding one content source and use them across other content types. When the system is limited to recommending content of the same type as the user is already using, the value from the recommendation system is significantly less than when other content types from other services can be recommended.

B. *Collaborative Approach*

One approach to the design of recommender systems that has wide use is collaborative filtering.[3] Collaborative filtering methods are based on collecting and analyzing a large amount of information on users' behaviors, activities or preferences and predicting what users will like based on their similarity to other users. A key advantage of the collaborative filtering approach is that it does not rely on machine analyzable content and therefore it is capable of accurately recommending complex items such as movies without requiring an "understanding" of the item itself. Many algorithms have been used in measuring user similarity or item similarity in recommender systems.

Collaborative filtering is based on the assumption that people who agreed in the past will agree in the future, and that they will like similar kinds of items as they liked in the past.

When building a model from a user's behavior, a distinction is often made between explicit and implicit forms of data collection.

Examples of explicit data collection include the following:

– Asking a user to rate an item on a sliding scale.
– Asking a user to search.
– Asking a user to rank a collection of items from favorite to least favorite.

- Presenting two items to a user and asking him/her to choose the better one of them.
- Asking a user to create a list of items that he/she likes.
- Examples of implicit data collection include the following:
- Observing the items that a user views in an online store.
- Analyzing item/user viewing times. [4]
- Keeping a record of the items that a user purchases online.
- Obtaining a list of items that a user has listened to or watched on his/her computer.
- Analyzing the user's social network and discovering similar likes and dislikes.

The recommender system compares the collected data to similar and dissimilar data collected from others and calculates a list of recommended items for the user.

Collaborative filtering approaches often suffer from three problems: cold start, scalability, and sparsity. [5]

Collaborative filtering methods are classified as memory-based and model based collaborative filtering. A well-known example of memory-based approaches is user-based algorithm[3] and that of model-based approaches is Kernel-Mapping Recommender.[6]

III. Hidden Markov Model

Hidden Markov Model (HMM) is a statistical Markov model in which the system being modeled is assumed to be a Markov process with unobserved (i.e. hidden) states.

The hidden Markov model can be represented as the simplest dynamic Bayesian network. The mathematics behind the HMM were developed by L. E. Baum and coworkers.[7] HMM is closely related to earlier work on the optimal nonlinear filtering problem by Ruslan L. Stratonovich,[8] who was the first to describe the forward-backward procedure.

In simpler Markov models (like a Markov chain), the state is directly visible to the observer, and therefore the state transition probabilities are the only parameters, while in the hidden Markov model, the state is not directly visible, but the output (in the form of data or "token" in the following), dependent on the state, is visible. Each state has a probability distribution over the possible output tokens.

The adjective hidden refers to the state sequence through which the model passes, not to the parameters of the model; the model is still referred to as a hidden Markov model even if these parameters are known exactly.

A hidden Markov model can be considered a generalization of a mixture model where the hidden variables (or latent variables), which control the mixture component to be selected for each observation, are related through a Markov process rather than independent of each other. Recently, hidden Markov models have been generalized to pairwise Markov models and triplet Markov models which allow consideration of more complex data structures[9] and the modeling of nonstationary data.[10]

IV. Algorithms

A. K-Mean Algorithm

k-means clustering is a method of vector quantization, originally from signal processing, that is popular for cluster analysis in data mining. k-means clustering aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean, serving as a prototype of the cluster. This results in a partitioning of the data space into Voronoi cells.

The problem is computationally difficult (NP-hard); however, there are efficient heuristic algorithms that are commonly employed and converge quickly to a local optimum. These are usually similar to the expectation-maximization algorithm for mixtures of Gaussian distributions via an iterative refinement approach employed by both k-means and Gaussian mixture modeling. Additionally, they both use cluster centers to model the data; however, k-means clustering tends to find clusters of comparable spatial extent, while the expectation-maximization mechanism allows clusters to have different shapes.

The algorithm has a loose relationship to the k-nearest neighbor classifier, a popular machine learning technique for classification that is often confused with k-means due to the k in the name. One can apply the 1-nearest neighbor classifier on the cluster centers obtained by k-means to classify new data into the existing clusters. This is known as nearest centroid classifier or Rocchio algorithm.

B. Linear Regression

Linear regression has many practical uses. Most applications fall into one of the following two broad categories:

- If the goal is prediction, or forecasting, or error reduction, linear regression can be used to fit a predictive model to an observed data set of values of the response and explanatory variables. After developing such a model, if additional values of the explanatory variables are collected without an accompanying response value, the fitted model can be used to make a prediction of the response.
- If the goal is to explain variation in the response variable that can be attributed to variation in the explanatory variables, linear regression analysis can be applied to quantify the strength of the relationship between the response and the explanatory variables, and in particular to determine whether some explanatory variables may have no linear relationship with the response at all, or to identify which subsets of explanatory variables may contain redundant information about the response.

Linear regression models are often fitted using the least squares approach, but they may also be fitted in other ways, such as by minimizing the "lack of fit" in some other norm (as with least absolute deviations regression), or by minimizing a penalized version of the least squares cost function as in ridge regression (L^2-norm penalty) and lasso (L^1-norm penalty). Conversely, the least squares approach can be used to fit models that are not linear models. Thus, although the terms "least squares" and "linear model" are closely linked, they are not synonymous

V. RELATED WORK

Recommendation Systems have been widely used in ecommerce. Many domains such as book, electronic products, music, movies etc. uses recommendations. In last few years, rapid development of information technology, an increasing number of books are available on ecommerce websites. Recommendation system helps user to find relevant books. Recently many scholars have made significant progress in recommendation systems. Lu et al. proposed content based filtering and collaborative filtering recommendation methods. Antonio Hermando et al proposed a proposed a prediction method of collaborative filtering recommendation based on collaborative filtering for rating of users based on Bayesian probabilistic model. Kouki et al. designed hybrid probabilistic extensible hybrid recommendation method, which could automatically learn and make predictions. Typical model based collaborative filtering includes the clustering techniques based collaborative filtering, probabilistic method based collaborative filtering and matrix based decomposition collaborative filtering. The ratings of related items usually calculated in process of recommendation, but related time sequence behavior information easily recommendation, but related time sequence behavior information is easily ignored. Many authors and proposed the recommendation algorithm based on time sequence information for this problem.

Time sequence based recommendation algorithm adds time sequence into the existing recommendation model. This algorithm enables the model to learn the data changing over time. Hence the accuracy of recommendation results would be improved. Time sequence algorithm is newly introduced and hence if it is combined to the existing algorithm, result achieved will be accurate and efficient. Hence, current scholars use hybrid approaches like same. Gao.et.al developed an improved collaborative filtering recommendation also with time adjusting. A real time stream based recommendation algorithm was proposed based on collaborative filtering. Some authors also add time sequence information into feature vectors of product.

VI. CONCLUSION

We have proposed a hybrid method for product recommendation. The greatest advantage of this method is that it combines user review data and contents features for the accurate results. As both the approaches have their disadvantages, it's better to use the hybrid approach so that accurate results will be generated.

ACKNOWLEDGEMENT

We are thankful to our mentor, coach and guide Prof. Ankit Mune Sir. We would like to thank you for this opportunity to present a paper on Recommender System with Hybrid approach. The references used for this are given below.

REFERENCES

[1] Hosein Jafarkarimi; A.T.H. Sim and R. Saadatdoost A Naïve Recommendation Model for Large Databases, International Journal of Information and Education Technology, June 2012.

[2] Blanda, Stephanie (May 25, 2015). "Online Recommender Systems – How Does a Website Know What I Want?". American Mathematical Society. Retrieved October 31, 2016.

[3] John S. Breese; David Heckerman & Carl Kadie (1998). Empirical analysis of predictive algorithms for collaborative filtering. In Proceedings of the Fourteenth conference on Uncertainty in artificial intelligence (UAI'98).

[4] Parsons, J.; Ralph, P.; Gallagher, K. (July 2004). "Using viewing time to infer user preference in recommender systems". AAAI Workshop in Semantic Web Personalization, San Jose, California

[5] Sanghack Lee and Jihoon Yang and Sung-Yong Park, Discovery of Hidden Similarity on Collaborative Filtering to Overcome Sparsity Problem, Discovery Science, 2007.

[6] "Kernel-Mapping Recommender system algorithms". Information Sciences. 208: 81–104. doi:10.1016/j.ins.2012.04.012. Retrieved 1 June 2015.

[7] Baum, L. E.; Petrie, T. (1966). "Statistical Inference for Probabilistic Functions of Finite State Markov Chains". The Annals of Mathematical Statistics. 37 (6): 1554–1563. doi:10.1214/aoms/1177699147. Retrieved 28 November 2011.

[8] Stratonovich, R.L. (1960). "Conditional Markov Processes". Theory of Probability and its Applications. 5 (2): 156–178. doi:10.1137/1105015.

[9] Pieczynski, Wojciech (2007). "Multisensor triplet Markov chains and theory of evidence". International Journal of Approximate Reasoning.

[10] Lanchantin et al., P. Lanchantin and W. Pieczynski, Unsupervised restoration of hidden non stationary Markov chain using evidential priors, IEEE Trans. on Signal Processing, Vol. 53, No. 8, pp. 3091-3098, 2005.

