Usage-Based Classification and Ranking with Machine Learning Techniques for Recommendations

Praful Koturwar
Department of Information Technology
Maharashtra Institute of Technology
Pune, India
prafulkoturwar@gmail.com

Sheetal Girase
Department of Information Technology
Maharashtra Institute of Technology
Pune, India
girase.sheetal@gmail.com

Debajyoti Mukhopadhyay
Department of Information Technology
Maharashtra Institute of Technology
Pune, India
debajyoti.mukhopadhyay@gmail.com

***Abstract*— Recommendation systems aim at recommending relevant items to the users. These systems provide efficient recommendations based on algorithms used for classification and ranking. There exists various ways by which classification can be achieved in a supervised or unsupervised manner. Since the sample datasets that are used for experiments are large and also contain more number of feature sets, it is essential to understand dataset beforehand. Also when results are shown to the user, big challenge is how well data can be ranked so that user satisfaction is guaranteed.**

When data sets are large, some ranking algorithms perform poorly in terms of computation and storage. Thus, these kind of algorithms are quiet expensive. In this paper, we are using various Machine Learning Techniques which will reduce computational cost and dimensionality of data without affecting diversity of feature set. Support Vector Machine (SVM) classification technique is used for filtering the data set according to the format of the data. We are also using Area under the Curve (AUC) and Weighted Approximately Ranked Pairwise (WARP) ranking functions depending upon data set for ranking purpose. We introduced new functionality called Sigmoid Meta Decider which will be responsible for selecting one appropriate function to improve response time.

Keywords- Recommendations; Classification; Ranking; SVM; AUC; WARP.

I. INTRODUCTION

Recommender systems are now pervasive in users' lives. They aim to help users in finding items that they would like to buy or consider based on huge amounts of data collected. These items can be of any type like movies, music, books, websites, or news articles. Recommender systems or recommendation systems are a subclass of information filtering system that seek to predict the 'rating' or 'preference' that user would give to an item. The user's interest in an item is expressed through the rating/ranking the user gives the item. A recommendation system has to predict the ranking for items that the user has not yet seen. With these estimated ratings the system can recommend the ranking of items that have the highest estimated rating. Recommendation system also predicts users' preference on the basis of his/her demographic features like age, gender, location, etc.

Parsing a huge amount of data to predict a user's preference or his or her similarity with other group of users is the core of a recommender system. Amazon, Facebook, LinkedIn, and other commercial and social networking websites use these systems. Amazon.com (http://www.amazon.com) uses a recommendation system to predict users' purchase behaviors, recommends goods of potential interest, and improves Amazon's commercial profit. Some websites use recommendation systems to increase user satisfaction [15].

The knowledge learned from the users' behavior is the basis for the recommendations. Because online businesses have no real space constraint, they can offer much larger stocks, providing their customers with more choices. These large stocks become impossible to stack search, so e-commerce stores must provide personalized versions with reduced choices to the individual users.

Classification technique is used to classify the recommendations according to the format of the data that must be processed, the type of analysis to be applied, the processing techniques at work, and the data sources for the data that the target system is required to acquire, load, process, analyze and store [4]. Many classification techniques are used based on applications selected. Before actual classification begins, required information is extracted from recommendations and then classification is done. There are two main classification techniques, supervised and unsupervised. Supervised classification techniques are also known as predictive or directed classification. In this method set of possible class is known in advanced. Unsupervised classification techniques are also known as descriptive or undirected. In this method set of possible class is unknown, after classification we can assign name to that class.

Support Vector Machine is a supervised method that analyzes data and recognizes patterns which is used for classification. Given a training set and the data needs to be classified into two classes, a SVM classifier builds a model that assigns the data into one of the categories. Extraction of huge training set is modelled as a multi-dimensional classification problem with one class for each action and its aim is to assign a class label to a given action or activity.

As the volume of information is increasing day by day so there is a challenge for recommender systems (RS) to provide proper and relevant dataset to the user. Fig. 1, shows a working of a typical RS, which shows the flow graph for a searched query by a user.

An efficient ranking of query words has a major role in efficient system for query words. There are various challenges associated with the ranking of datasets such that some dataset are made only for navigation purpose and some data of the dataset do not possess the quality of self-descriptiveness. In this paper, the different classification and ranking recommendation techniques being experimented, analyzing them in terms of the data that

supports the recommendations, the algorithms that operate on that data and examines the best recommendation techniques that have been proposed for recommending datasets.

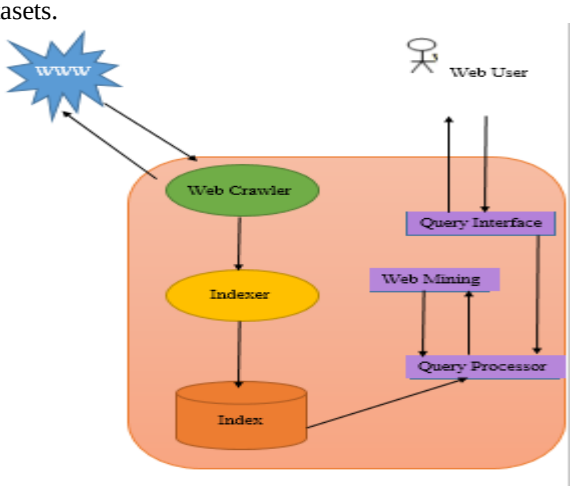

Figure 1. Process of Ranking

Ranking is a central part of many information retrieval problems, such as document retrieval, collaborative filtering, sentiment analysis, computational advertising (online ad placement). A possible architecture of a machine-learned search engine is shown in the Fig. 1, to the right Training data consists of queries and documents matching them together with relevance degree of each match. It may be prepared manually by human assessors, who check results for some queries and determine relevance of each result. It is not feasible to check relevance of all documents, and so typically a technique called pooling is used only the top few documents, retrieved by some existing ranking models are checked. Alternatively, training data may be derived automatically by analyzing click through logs (i.e. search results which got clicks from users), query chains, or such search engines' features as Google's Search Wiki.

We propose a two-stage optimization strategy for learning ranking functions that is robust to outliers and applicable to any method that learns a linear ranking function. This meta-ranker is computationally economical and, although developed for document ranking, this method generalizes across many ranking tasks. With Sigmoid global loss to decide which loss function suitable to order user interest rank page using WRAP and AUC with the Help of SVM [14].

The rest of this paper is organized as follows. In Section II, related work done by various researchers in this field is presented. In Section III, proposed methodology is discussed. The conclusion is given in last section.

II. RELATED WORK

Xindong Wu et al. [1], have studied Big Data concern large-volume, complex, growing data sets with multiple, autonomous sources for the fast development of networking, data storage, and the data collection capacity. In this paper the author has presented a HACE theorem that characterizes the features of the Big Data revolution and he proposed a Big Data processing model in the data mining perspective involves demand-driven aggregation of information sources, mining and analysis, user interest modeling, and security and privacy considerations.

Dingxian Wang et al. [2], has presented a stock futures prediction strategy by using a hybrid method to forecast the price trends of the futures which is essential for investment decisions. In order to deal with huge amounts of futures data, the author have studied strategy which consists of two parts, such as, I. Raw Data Treatment and Features Extraction, and II. DT-SVM Hybrid Model Training. In this paper, they employed real world transaction data of stock futures contracts for their study that data are first stored in a distributed database then the data are distributed to a group of computing nodes to extract statistical features and finally, a hybrid method combing DT (Decision Tree) and SVM (Support Vector Machine) algorithms is applied. This method can filter most noisy data with the DT algorithm in the first phase, and then using the SVM algorithm to process the big training data in the second phase.

G. Kesavaraj et al. [3], has presented the basic classification techniques and different kinds of classification method which are decision tree induction, bayesian networks, k-nearest neighbor classifier.

Shan Suthaharan [4], focuses on the specific problem of Big Data classification of network intrusion traffic. In this paper, He discussed the system challenges presented by the Big Data problems associated with network intrusion prediction as well as problems and challenges in handling Big Data classification using geometric representation-learning techniques and the modern Big Data networking technologies. In this paper, He also discussed the issues related to combining supervised learning techniques, representation learning techniques, machine lifelong learning techniques and Big Data technologies (e.g. Hadoop, Hive and Cloud) for solving network traffic classification problems.

S. B. Kotsiantis [5], has studied various supervised machine learning classification techniques.

Hwanjo Yu et al. [6], presented a new method called Clustering-Based SVM (CBSVM) which is specifically designed for handling very large data sets. Their experiments on synthetic and real data sets show that CB-SVM is highly scalable for very large data sets while also generating high classification accuracy.

Yongjun Piao et al. [7], had proposed an ensemble method for classification of high-dimensional data, with each classifier constructed from a different set of features determined by partition of redundant features. Their method for the redundancy of features was considered to divide the original feature space then, each generated feature subset was trained by support vector machine and the results of each classifier were combined by the majority voting method. The efficiency and effectiveness of this method were demonstrated through comparisons with other ensemble techniques, and the results showed that this method outperformed other methods.

The Vitthal Yenkar, Prof.Mahip Bartere [8], had discussed a characterizes applications of Big Data processing model and Big Data revolution, from the data mining outlook which is made for a very important issue, like the heterogeneous mixture property of data. In this paper, they introduced heterogeneous mixture learning and

the study of various issues in the Big Data revolution and also in the data-driven model.

The main goal of Mohammed GH. AL Zamil [9], is to provide a technique that facilitates extracting ontological patterns, which enhance the semantic interpretation of such pool of knowledge. Furthermore, the proposed framework facilitates integrating different heterogeneous sources of knowledge into a single one.

Wei Dai and Wei Ji, [10], have proposed a typical decision tree algorithm, C4.5, using MapReduce programming model. They used proposed algorithm to transform the traditional algorithm into a series of Map and Reduce procedures. In addition, the authors design some data structures to minimize the communication cost and also conduct extensive experiments on a massive dataset. The results indicate that their algorithm exhibits both time efficiency and scalability.

Jason Weston, et al. [11], proposed a strongly performing method that scales to huge datasets by simultaneously learning to optimize precision at k of the ranked list of annotations for a given image and learning a low-dimensional joint embedding space for both images and annotations. The proposed method had outperformed several baseline methods and, in comparison to them, is faster and consumes less memory.

Nicolas Usunier et al. [12], proposed to optimize a larger class of loss functions for ranking which is based on an Ordered Weighted Average (OWA) (Yager, 1988) of the classification losses. When aggregating hinge losses, the optimization problem is similar to the SVM for interdependent output spaces. Moreover, they showed OWA aggregates of margin based classification losses have good generalization properties. Experiments on the Letor 3.0 benchmark dataset for information retrieval validate our approach.

Jason Weston et al. [13], proposed a strongly performing method that scales to such datasets by simultaneously learning to optimize precision at the top of the ranked list of annotations for a given image and learning a low dimensional joint embedding space for both images and annotations. The proposed method is called WSABIE which outperformed several Classification with sigmoid function decides the algorithm for ranking until learning get threshold greater than 0 consumes less memory. In this paper [14], they presented a family of loss functions, the k-order statistic loss which includes stochastic gradient descent scale i.e. good for large collaborative data scale.

Alexandros Karatzoglou et al. [15], tutorial focuses on the cutting-edge algorithmic developed in the area of recommender systems that provide a depth picture of the progress of ranking models in the field, summarized the strengths and weaknesses of existing methods, and discussed open issues that could be promised for future research in the community.

Krisztian Balog et al. [16], performed an experimental comparison of these two strategies using supervised learning with a rich feature set. There were main finding is that ranking outperforms classification on all evaluation settings and metrics and analysis reveals that a ranking-based approach has more potential for future improvements.

III. PROPOSED METHODOLOGY

A. *Problem Statement*

To develop classification and ranking algorithm which will reduce computational cost and dimensionality of data without affecting the diversity of the feature set.

B. *K-ORDER STATISTIC LOSS using Meta Strategy*

Process 1: Rating by K-order

We consider the general recommendation task of ranking a set of items D for a given user; the returned list should have the most relevant items at the top. To solve this task, we are given a training set of users U each with a set of known ratings [11-14].

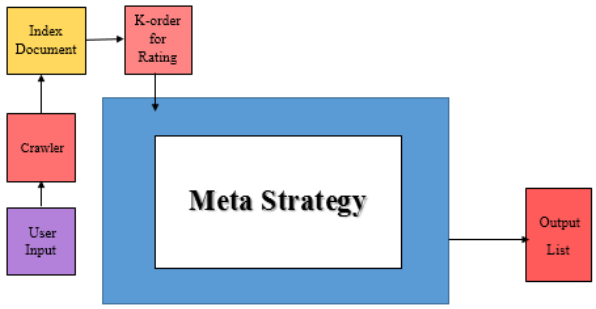

Figure 2. Proposed Block Diagram

We consider the case where each user has purchased / watched / liked a set of items, which are considered as positive ratings. No negative ratings are given. All non-positive rated items are thus considered as having an unknown rating1. We define the set Du to be the positive items for user u. We consider factorized models of the form

$$f_d(u) = \frac{1}{|D_u|} \sum_{i \in D_u} V_i^T V_d$$

We are given a probability distribution P of drawing the ith position in a list of size K. This defines the choice of loss function.

- Pick a user u at random from the training set.
- Pick i = 1. . . K positive items $d_i \in D_u$.
- Compute f_{di} (u) for each i.
- Sort the scores by descending order,
- Let o (j) be the index into d that is in position j in the list.
- Pick a position k 2 1. . . K using the distribution P.
- Perform a learning step using the positive item $d_o(k)$.

Process 2: Ranking by SVM with sigmoid loss

One of the most successful algorithms for classification, Support Vector Machine, has recently been successfully adapted to ranking using the pairwise framework. Given a binary paired dataset S0 from a set of queries, an SVM classifier can be naturally adapted to model this problem. The SVM model will attempt to solve the following quadratic optimization problem:

1. For All User Matrix **Ui** and Item Matrix **Ij** belongs To Document **D**
2. Min1/2$[w]^2$ +SUM(w, di – dj)

Where, non-negative slack variables _ql were introduced, and the trade-off between margin size and training error is controlled by the parameter C.

3. The sigmoid loss function is not convex, thus the learning procedure is only guaranteed to reach a local maximum. To avoid learning poor locally optimal solutions, the sigmoid ranker is used as a second optimization step as Global Loss work for a Local Loss i.e. WARP or AUC,

$$\min_{w} L(w) = \lambda \|w\|^2 + \sum_{q,l} [1 - sigmoid(\sigma, z_{ql} \langle w, d_{qi} - d_{qj} \rangle)].$$

Ranke[w] = WARP {If (sigmoid(x) =1/1+e^{Xx}) >0.5}

Ranke[w] = AUC {If (sigmoid(x) =1/1+e^{Xx}) <0.5}

Process 3: WARP

Weighted Approximate Ranked Pairwise loss function that attempts to focus on the top of list by comparing the positive and negative items from data sets. In the pairwise approach, the position of a given relevant element in the sorted list can be computed by counting the number of irrelevant elements that have a higher score. This count can be carried out by building the pair of Relevant (Positive) and Irrelevant (Negative) elements and checking the sign of different scores. Consider an example: In e-commerce site, we see that Amazon gets more hits than other e-commerce sites globally hence it is on the top of the ranked positive items and remaining are considered as negative ratings. Compare to Amazon, Flipkart has low rating. Though world-wide, we did not get the straight forward information from data, some data might be missing or crashed. At this condition we used WARP loss function otherwise we used AUC loss function.

Process 4: AUC

Area under the Curve is a well-known loss function, sometimes known as Margin Ranking Loss. With respective above WARP algorithm which only focus on top ranked elements and ignore the low ranked elements. These low ranked elements consider as a positive items by AUC loss function and then obtained well rank list. As per above example of e-commerce site, Flipkart becomes positively ranked item according to AUC loss function and will be consider for ranking purpose.

Process 5: LEARING

We utilized a gradient descent technique to learn the final ranking model w. specifically, we can differentiate the sigmoid-based loss function with respect to the parameter vector w. To obtain parameter vector w, we have to compute sigmoid of frequent relevant items. The gradient descent algorithm is used to compute the threshold value η_k

where, the index k defines the number of iterations.

C. System Architecture

The architecture will divide a given software application into two parts, so as to separate internal representations of information from the ways that information will present to or accept from the user. First part is the Web, from where user can send his request to the system and other one is Meta strategy (i.e., Meta Learning) that is the heart of this architecture which classified into classification and ranking.

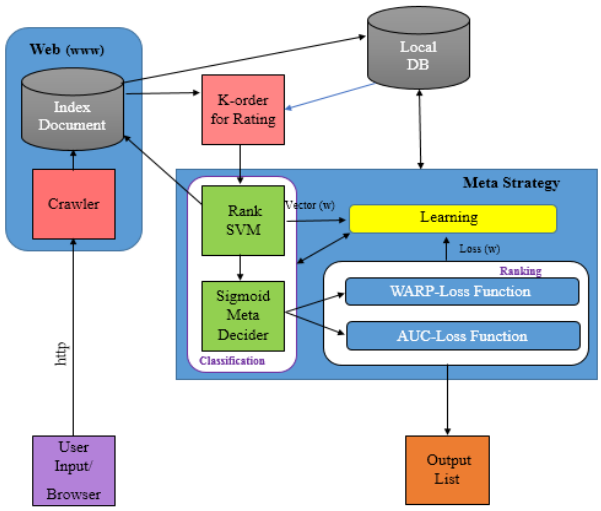

Figure 3. Proposed System Architecture

Classification will filter the recommendations that take k-order as an input. For this purpose, Classification uses the SVM supervised techniques which is the most effective supervised method than other classification techniques. Learning is the component take a filtered data from SVM for learning purpose. Sigmoid is the SVM kernel or function which is nothing but decider. Sigmoid perform decision making about the data it gains from SVM that either it send to AUC or WARP based on threshold value. If threshold value is greater than 0.5 then WARP algorithm compute the loss functions else AUC algorithm. Finally, the learning will compare optimized loss function items with the filtered data and then obtained item is the top ranked item.

IV. EXPERIMENTAL RESULTS

Consider an example shown in Table 1, In which 5 users i.e U1,U2,U3,U4,U5 and items I1,I2,I3 have shown. Similarity of users with respective U1 is calculated by using similarity evaluation matrix. After similarity evaluation classification algorithm is applied to classify the list into different categories. Suppose classification threshold value (η) is set to 0.5 then it classifies similarity values into two categories that is Category I(0 to 0.5) and Category II(0.5 to 1) then AUC ranking algorithm is applied to Category I and WARP ranking algorithm is applied to Category II to generate ranked recommendations for user.

TABLE 1. RANKING ALGORITHM BASED ON SIMILARITY WITH U1

	I1	I2	I3	Similarity with U1	η	Ranking Algorithms (WARP/AUC)
U1	5.0	3.0	2.5	1.000	>0.5	WARP
U2	2.0	2.5	5.0	0.203	<0.5	AUC
U3	2.5	-	-	0.286	<0.5	AUC
U4	5.0	-	3.0	0.667	>0.5	WARP
U5	4.0	3.0	2.0	0.472	<0.5	AUC

V. CONCLUSION

In this paper, K-Order Stochastic gives positive and negative data items, then Classification technique with sigmoid function decides the algorithm for ranking until learning get threshold. Based upon threshold value, ranking algorithm provides a definite rank to resultant data items. The existing techniques have limitations particularly in terms of response time, accuracy of results, importance of the results and relevancy of results.

A general class of ranking loss function is used for training large-scale factorized recommendation models based on threshold value. This class generalizes several well-known loss functions which are Area under the Curve and Weighted Approximately Ranked Pairwise and also provides new choices of objective function. In particular, by focusing on the training of more highly ranked items, one can obtain better precision and recall metrics compared to those existing approaches. Alternatively, by focusing the training on lower ranked items one can obtain better mean or maximum rank metrics. Depending on the overall goal, both of these approach may be useful.

REFERENCES

[1] Xindong Wu, Xingquan Zhu, Gong-Qing Wu, and Wei Ding, "Data Mining with Big Data," Transactions On Knowledge And Data Engineering, Vol. 26, No. 1. 1041-4347/14 January 2014, IEEE.

[2] Dingxian Wang, Xiao Liu, Mengdi Wang, "A DT-SVM Strategy for Stock Futures Prediction with Big Data,"16th International Conference on Computational Science and Engineering 978-0-7695-5096-1/13, 2013, IEEE.

[3] G. Kesavaraj, Dr. S. Sukumaran, "A Study on Classification Techniques in Data Mining," 4th ICCCNT – Tiruchengode, India, 31661, July 4 - 6, 2013, IEEE.

[4] Shan Suthaharan, "Big Data Classification: Problems and Challenges in Network Intrusion Prediction with Machine Learning," Department of Computer Science, University of North Carolina at Greensboro, Greensboro, NC 27402, USA, 2012.

[5] S. B. Kotsiantis, "Supervised machine learning: A review of classification techniques," Informatica 31, 249-268, 2007.

[6] Hwanjo Yu, Jiong Yang, Jiawei Han, "Classifying Large Data Sets Using SVMs with Hierarchical Clusters," SIGKDD '03 Washington, DC, USA, 1581137370/ 03/0008, 2003, ACM.

[7] Yongjun Piao, Hyun Woo Park, Cheng Hao Jin, Keun Ho Ryu, "Ensemble Method for Classification of High-Dimensional Data," 978-1-4799-3919-0/14, 2014, IEEE.

[8] Vitthal Yenkar, Prof.Mahip Bartere, "Review on Data Mining with Big Data," International Journal of Computer Science and Mobile Computing, Vol.3 Issue.4, pg. 97-102, April- 2014.

[9] Mohammed GH. AL Zamil, "The Application of Semantic-based Classification on Big Data," International Conference on Information and Communication Systems (ICICS) 978-1-4799-3023-4/14, 2014, IEEE.

[10] Wei Dai, Wei Ji, "A MapReduce Implementation of C4.5 Decision Tree Algorithm," International Journal of Database Theory and Application, SERSC, 2014.

[11] Jason Weston1, Samy Bengio1, and Nicolas Usunier, "Large Scale Image Annotation: Learning to Rank with Joint Word-Image Embeddings," Journal Machine Learning, Vol.81 Issue.1, pg. 21-35, October- 2010.

[12] Nicolas Usunier, David Buffoni, Patrick Gallinari, "Ranking with Ordered Weighted Pairwise Classi_cation," ICML '09 Proceedings of the 26th Annual International Conference on Machine Learning, Pages 1057-1064, ISBN: 978-1-60558-516-1, New York, NY, USA, 2009, ACM.

[13] Jason Weston, Samy Bengio, Nicolas Usunier, "WSABIE: Scaling Up To Large Vocabulary Image Annotation," IJCAI'11 Proceedings of the 21nd international joint conference on Artificial Intelligence, Vol.3,pg. 2764-2770, ISBN: 978-1-57735-515-1, 2011, AAAI.

[14] Jason Weston, Hector Yee, Ron J. Weiss, "Learning to Rank Recommendations with the k-Order Statistic Loss," RecSys'13, Hong Kong, China, 978-1-4503-2409-0/13/10 October 12–16, 2013, ACM.

[15] Alexandros Karatzoglou, Linas Baltrunas, Yue Shi, "Learning to Rank for Recommender Systems," RecSys'13, Hong Kong, China, 978-1-4503-2409-0/13/10 October 12–16, 2013, ACM.

[16] Krisztian Balog, Heri Ramampiaro, "Cumulative Citation Recommendation: Classification vs. Ranking," SIGIR'13, Dublin, Ireland, 978-1-4503-2034-4/13/07 July 28–August 1, 2013, ACM.

[17] Praful Koturwar, Sheetal Girase, Debajyoti Mukhopadhyay, "A Survey of Classification Techniques In The Area of Big Data," International Journal of Advance Foundation and Research in Computer (IJAFRC), ISSN 2348 – 4853, Volume 1, Issue 11, November 2014.