A Survey of Multiclassifier Algorithms For Handling The Dynamics Of Web Data

Meghna Khatri, Chhavi Rana

Computer Science and Engineering Department, Maharishi Dayanand University
Rohtak, Haryana, India.

meghna823@gmail.com

chhavi.rana@gmail.com

***Abstract*— The data available on the web is highly dynamic in nature especially in the case of e-commerce systems. It needs to be handled efficiently in order to provide competitive and efficient applications such as recommender system that would be able to predict the ever changing tastes of users accurately. This aim can be fulfilled by building a model which is a combination of several classifier algorithms and also including a time-variable function. Over the last decade, several multiclassifier approaches were proposed and shown to be useful in many applications. Some approaches in temporal dynamics also emerged and have been proven fruitful in respective areas. In this paper an overview of such techniques of multiclassifier algorithms as well as temporal dynamics is presented.**

***Keywords*— Multiclassifier algorithms, temporal dynamics**

I. INTRODUCTION

The amount of information available on the web is increasing at a very fast rate. The mechanism of the e-commerce systems should be such that they are able to obtain new customers and retain existing ones. Also the customer preferences for products are drifting over time as new selections changes the product perception and popularity and customer inclinations are evolving, leading them to ever redefine their taste. this problem is known as concept drift [1] and to solve it a temporal model is required.

Data mining provides a number of algorithms to build profiles of users based on historical data, which is used to predict the preferences of new users. The predictive models induced by these algorithms are named (as) classifiers. The web data is divided into two sets, training set (used to train the model) and the test set (used to test the model). However, the data from the web used to build these models is heterogeneous and the behaviour of classifier [18] in an individual way sometimes fails with some training sets, when a wide variety of data exist. Therefore, multiclassifiers are a better option.

The multiclassifiers [2] are the result of combining several individual classifiers. The methods for building multiclassifiers are divided in two groups: ensemble and hybrid methods. The first method, such as Bagging [3] and Boosting [4], induce models that merge classifiers with the same learning algorithm but modifications in the training data set whereas the second method such as Stacking [16] creates new hybrid learning technology from different base learning algorithm.

The significance of the multiclassifiers may be enhanced if the temporal factors are included in the algorithm. This may be done by including a time variable function. This will trace the changes or the dynamics of the web data and the web data could be handled through it.

II. BACKGROUND

There are several reasons why multiclassifiers should be considered over a single classifier. In several applications, the volume of data to be analyzed is too large to be handled by a single classifier. Training a classifier with such a vast amount of data is usually not practical. A multiclassifier system will be an efficient approach, where data is partitioned into smaller subsets, trained with different classifiers for different subsets and the outputs are combined. Also a single classifier cannot perform well when the nature of features is different. Using multiple classifiers with a subset of features may provide a better performance. Another reason for combining classifiers is to improve the generalization performance. Finding a single classifier to work well for all test data is difficult. Instead multiple classifiers can be combined to give a better output than a single classifier.

Two or more classifiers can be combined together to create a multiclassifier. The ensemble methods induce models that merge classifiers with the same learning algorithm but introducing modifications in the training data set. A set of classifiers is constructed from the training data which will predict the class of previously unseen records by aggregating predictions made by multiple classifiers [6]. It can be performed in two ways Bagging and Boosting. Bagging [3] stands for Bootstrap Aggregating where equally weighted predictions from multiple models are combined. It exploits the instability in learning schemes. In Bagging approach several independent training sets are collected and a classifier is build from each training set e.g. learn decision tree [7] from each training set. The class of a test instance is the prediction that received most votes from all the classifiers. In Boosting [4], multiple data mining methods are combined using weighted voting. Boosting is an iterative model. Each new model is built to overcome the deficiencies in the earlier models. Whereas the hybrid methods such as Stacking [9] [16] is applied to models of different types and a metalearner is used to combine the different base learners.

The time factor can be included as a form of a function in any of these models which could be baseline predictors [1], Jaccard distance, Euclidean distance or by using the implicit feedback.

The multiclassifiers can be used to handle the dynamics of web data and has many uses in web usage mining, text mining, personalisation [11] of the recommender system and web page mining

III. LITERATURE REVIEW

A. Multiclassifier Algorithms

The multiclassifiers are the result of combining several individual classifiers. The methods for building multiclassifiers [17] are divided in two groups: ensemble and hybrid methods. The first methods, such as Bagging [3] and Boosting [4], induce models that merge classifiers with the same learning algorithm, but introducing modifications in the training data set. The second type methods, such as Stacking [16], create new hybrid learning techniques from different base learning algorithms. The architectures and main methods of multiclassifiers were described in an another study.[11]

Another work by Mangai UG, et al [12] gives a layout of the various techniques that are applied for the combination of the classifier algorithm to form the multiclasssifer. The classifier combinations are the results of two parallel lines of study (i) decision optimization method in which an attempt is made to find an optimal combination of classifiersis made to find an optimal combination of classifiers, decision [19], given a fixed set of carefully designed and highly specialized classifiers. (ii) coverage optimization method – generate a set of mutually complementary, generic classifiers that can be combined to achieve optimal accuracy assuming a fixed combination rule.

The multiple classifier system can be achieved in one of the following ways :-

- Variation of initial parameters of the classifiers: a set of classifiers can be created by verifying the initial parameters, using which each classifiers is trained with the same training data.
- Variations of the training dataset of the classifiers: multiclass systems can be built by training the same classifiers with different training datasets.
- Variations in the number of individual classifiers used: training different types of classifiers like SVM, ANN, etc., with the same training dataset.

When combining various classifiers to obtain a multiclassifier, a new approach was proposed in [14] as an algorithm was developed which firstly applies bagging method to create some feature subsets. Secondly using principal component analysis of feature extraction method on each feature subsets and select classifiers with the classification accuracy. Then apply the classifier diversity measure to select diversity classifiers for their combination.

Work by Kittler [15] develops a common theoretical framework for combining classifiers which use distinct pattern representations and show that many existing schemes can be considered as special cases of compound classification where all the pattern representations are used jointly to make a decision. An experimental comparison of various classifier combination schemes is also given that demonstrate the combination rule developed under the most restrictive assumptions, the sum rule, outperforms other classifier combinations schemes.

Associative classification and the combination strategy for multiclass classification: Classification is an important management task. Many methods such as the agent-based approach, decision tree, and data envelopment analysis have been proposed to solve the decision analysis problems in various fields. Associative classification is a relatively new classification method whose aim is to apply the Apriori algorithm to mine association rules and construct associative classifiers. Rule mining will find the associations between attributes (rule preconditions) and ratings (results). In associative classification, the support degree is defined as the ratio of the number of objects satisfying a specific rule precondition and having a specific rating result over the total number of objects in the database.[10]

B. Time factor

Work by Lathia, N. et al [8] gives an outline of a method of how to depict user-similarity over time. In order to incorporate time factor, the user rating is sorted according to when they were input and then simulate a system that iteratively updates (every μ days). Beginning at time (t=ε) all ratings before ε are used to train the algorithm and test on all ratings input before the next update, at time (ε+μ). This process is repeated for each time t, incrementing by μ at each step. At each step, what was previously tested on becomes incorporated into the training set and simulated on the system.

Time-changing baseline predictors was given by Koren, Y [7] in which it is proposed to include the temporal variability within the baseline predictors through two major temporal effects. First is addressing the fact that an item's popularity is changing over time. For example, movies can go in and out of popularity as triggered by external events such as the appearance of an actor in a new movie. This is manifested in our models by the fact that item bias bi will not be a constant but a function that changes over time. The second major temporal effect is related to user biases - users change their baseline ratings over time. For example, a user who tended to rate an average movie "4 stars", may now rate such a movie "3 stars", for various reasons explained earlier. Hence, in our models we would like to take the parameter b_u as a function of time. This induces the following template for a time sensitive baseline predictor:

$$b_u i(t) = \mu + b_u(t) + b_i(t)$$

The function $b_{ui}(t)$ represents the baseline estimate for u's rating of i at day t. Here, $b_u(t)$ and $b_i(t)$ are real valued functions that change over time.

More work by Koren, Y., Bell, R., [13] which gives a more detailed approach of capturing temporal dynamics with the baseline predictors and also states more prediction rules.

In another work by Lathia, N., [9] a new approach is proposed. They say that by minimizing the mean error produced when predicting hidden user ratings and also if we adopt an approach of adaptive neighbourhoods [20] then root mean square error is considered to be a criterion for including temporal factor.

Another approach of implicit feedback is given by Lee, T.Q., Park, T., [6] which proposes to give the pseudo ratings matrix an entry '1' as a rating value when a user u purchases. A Time-based Pseudo Rating Matrix is generated where two kinds of temporal information are incorporated - the time when the item was launched and the time when the user purchased an item - into the simple pseudo rating matrix. Two observations are taken:

- More recent purchases better reflect a user's current preference.

- Recently launched items appeal more to users.

Based on these observations, they define a rating function *w* that computes rating values (rather than simply assigning 1) as follows:

w(pi, lj) = The rating value when an item with launch time lj was purchased at time pi.

IV. CONCLUSION

Combination of the individual classifiers to form an efficient multiclassifier algorithm is an important issue. Including a temporal variable in multiclassifier algorithm will be a requirement so as to ensure the accuracy and efficiency of the algorithm.

Improving time and space complexities of algorithm is a problem that would continue to attract attention.

The proper combination of the time functions and the classifier algorithms is an important issue and will provide a platform for enhancing the performance while handling the dynamics of web data.

REFERENCES

[1] Freund, Y., Schapire, R.E., Experiments with a new boosting algorithm. Proceedings of the 13th International Conference on Machine Learning, 1996, pp. 148-156.

[2] Breiman, L., Bagging predictors, Machine Learning, Vol. 24, No.2, 1996, pp. 123-140.

[3] Yang, Y., An evaluation of statistical approaches to text categorization, Journal of Information Retrieval, Vol. 1, 1999, pp. 69-90.

[4] Mor, E., Minguillón, J., An empirical evaluation of classifier combination schemes for predicting user navigational behavior. Proceedings of the International Conference on Information Technology: Computers and Communications, ITCC'03, pp. 467- 471, 2003

[5] Segerera, S. and Moreno, Maria N., An experimental comparative study of web mining methods for recommender system, WSEAS International Conference on Distance Learning and Web Engineering, 2006.

[6] Lee, Tong Q., Park, Young, Park Yong T., A Time-based Recommender System using Implicit Feedback, Expert Syst. Appl., Vol. 34, No. 4. (May 2008), pp. 3055-3062.

[7] Koren, Y., Collaborative filtering with temporal dynamics, Yahoo! Research, 2009.

[8] Lathia N,Hailes S,Capra L, Temporal Collaborative Filtering With Adaptive Neighbourhoods, International ACM SIGIR conference on Research and development in information retrieval SIGIR 09.,pp.796-797

[9] Lathia N,Hailes S,Capra L, Evaluating Collaborative Filtering Over Time, ACM SIGIR Workshop on the Future of IR Evaluation, JUL-2009

[10] Yuanchun Jiang a,b,, Jennifer Shang b, Yezheng Liu a,c, Maximizing customer satisfaction through an online recommendation system: A novel associative classification model, Decision Support Systems 48 pp 470–479, 2010.

[11] L. Xu, A. Krzyzak, and C. Y. Suen, "Methods of combining a multiple classifiers and their applications to handwriting recognition," IEEE Trans. Systems, Man, and Cybernetics, vol. 22, pp. 418-435, March 1992.

[12] Mangai UG, Samanta S, Das S, Chowdhury PR. A Survey of Decision Fusion and Feature Fusion Strategies for Pattern Classification. IETE Tech Rev 2010;27:293-307

[13] Koren, Y.; Bell, R. , Advances in Collaborative Filtering Recommender Systems Handbook, chapter 5, 2011.

[14] Lathia N., Hailes S., Capra L., Amatriain X., Temporal Diversity in Recommender Systems, SIGIR May, 2010.

[15] Kittler,J.V.,.Hatef, L., Duin, R. P. W. and Matas, J.,on combining classifiers,ieee trans. pattern analysis and machine Intelligence, vol.20, pp 226-239, march 1998.

[16] Wolpert, D.H., Stacked Generalization, Neural Networks, Vol. 5, 1992, pp. 241-259.

[17] Segrera, S., Moreno, M.N., Multiclasificadores: Métodos y Arquitecturas. Informe Técnico- Technical Report, DPTOIA-IT-2006-001, Departamento de Informática y Automática, Universidad de Salamanca, Mayo, 2006.

[18] Hansen, L.K., Salamon, P., Neural network ensembles, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 12, No.10, 1990, pp. 993-1001.

[19] Ho, T.K., Hull, J. And Srihari, S.N., "Decision in multiple classifier systems", IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 16, pp. 66-75, January 1994.

[20] Adomavicius, G. And Tuzhilin, A.. Towards the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions. IEEE TKDE, 17(6), June 2005.