ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 2, Issue 8, August 2013

SVM and AdaBoost Based Ranking Model Adaptation for Domain Specific Search

Greeshma.L[1], Srinivasa Rao.M[2], R.V.Krishnaiah[3]

Student, Department of CSE, DRKCET, Hyderabad, India[1]

Associate Professor, Department of CSE, DRKCET, Hyderabad, India [2]

Principal, Department of CSE, DRKCET, Hyderabad, India [3]

Abstract: The invent of Web 2.0 has enabled building of state of the art web applications including sophisticated search engines. The web applications are capable of providing domain specific search which enables end users to gain access to required data quickly. However, the results might be bulky and may not be relevant to the user-intended results. In other words users have to spend some time browsing results for finding required information. To overcome this problem many ranking algorithms came into existence. The ranking algorithms help users to find required results quickly. But the ranking models in the existing work were built based on broad-based ranking which is not useful for other domains. Recently Geng et al. proposed a ranking model adaption framework which can adapt to various domain specific searches. They used SVM for building ranking model. In this paper we built a prototype application that demonstrates ranking model adaption using a novel ranking model meant for ranking the search results besides adapting to new domains. The experimental results revealed that the proposed application is useful in searching data across the domains.

Index Terms: Ranking model, domain specific search, SVM, model adaptation

I. INTRODUCTION

Data mining has been around for performing various data mining operations. There are many techniques that can be used to discover knowledge from OLTP data. With the invent of Web 2.0 technologies, it is possible that the web applications are supporting search options that can be used to search domain specific data. Such search engines return huge number of records. All the records may not be relevant to end users. User has to spend his valuable time in order to search for required information. This kills the time of users. To overcome this drawback all search engines like Google, Yahoo, others employ a mechanism to present results based on ranking. However, the ranking model used by them is domain specific and can't be adapted to new domains. Especially the general search engines provide broad – based search that can't be directly used when user needs information from various domains. There are many approaches for ranking the results. For instance Ranking SVM [1], [2], RankBoost [3], LambdaRank [4], RankNet [5] and ListNet [6]. These ranking models are used in various real world applications for improving search results. However they are domain specific can can't be adapted to new domains easily.

Many information retrieval systems also use some mechanisms to present results to end users. Term frequency is used in case of ranking of documents based on the content of the documents. These models are also broad-based. These models cannot cope with the data of various domains. However, creation of different ranking models for different domains is not a suitable solution. Lexical analysis is also used by the information retrieval systems to fill the semantic gap between the search word and the results retrieved by them. For this reason model adaption is essential. There are many researches on this aspect. They include [7], [8], [9], and [10] that did not focus on the adaptation of domain specific search. In [11] concept drifting and classifier adaptation concepts are explored. These concepts also can't adapt to new domains easily. This is because the data of each domain is different along with metadata. Recently Geng et al. [12] built a ranking model which can adapt to new domains. It is something different from other aspects like classifier adaptation or data adaptation. They developed adaptive SVM for the purpose.

In this paper we proposed a novel algorithm for ranking model adaptation. This algorithm helps improve the results when compared with the technique proposed in [12]. The rest of the paper is structured as follows. Section 2 focuses on review of literature. Section 3 provides information about the new ranking adaptation model. Section 4 gives details

ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 2, Issue 8, August 2013

about the experiments, results and evaluation while section 5 provides conclusions.

II. PRIOR WORKS

In literature prior works on the ranking models are found. However, many of them have a common thread that is they are almost either domain specific or broad based. The domain specific approaches can't be adapted to new domains. The broad based search facilities also can't cope with new domains easily. In this context, this section explores various developments in the area of ranking models for search engines. All such search engines deal with information retrieval [13], [14]. These models have certain limitations. For this reason they can't adapt to new domains. This is because they need certain parameters to be changed and the underlying logic won't work with new datasets. With respect to ranking many ranking algorithms were found in the literature [4], [5], [6], [3], and [1].

Recently Geng et al. [12] presented a framework that has built in ranking model adaptation mechanism that helps to switch to different domains without modifying the basic ranking model. The new datasets can be used to work with that for ranking. It does mean that the users can switch between domains with the same framework for domain specific search. This kind of vertical search is very useful to end users as it can serve their information needs effectively. Some ranking model work is found in [15], [7], [16], and [9]. However, they are not close to the work done in [12]. Cross domain video detection algorithm was presented in [9]. The algorithm in [9] and [12] is known as adaptive SVM. The adaptive SVM is capable of supporting ranking model that can be adapted to various domains.

III. PROPOSED FRAMEWORK RANKING MODEL ADAPTATION

In this paper we design and implement a new ranking model adaptation algorithm which is based on SVM and AdaBoost. However this work is influenced by [12]. We have mixed the functioning of SVM presented in [12] and the AdaBoost algorithm for better performance. SVM (Support Vector Machine) is a machine learning algorithm that is used for ranking the results of search. This algorithm is widely used for classification of objects. In this paper is it used to rank search results. AdaBoost is an algorithm for boosting any other learning algorithms. SVM can be used along with AdaBoost. In our framework we combined both AdaBoost and SVM for ranking model adaptation. The results are compared with the ranking model adaption framework proposed in [12].

Prototype Application

A web based prototype application is built in Java platform. The technologies used include Servlets and JSP. The environment used for developing the prototype is a PC with 4GB RAM, Core 2 Dual processor running in Windows 7 operating system. NetBeans IDE is used for rapid application development.

IV. EXPERIMEENTAL RESULTS

Experiments are done on datasets collection from Internet sources. The datasets include TD2003 and TD2004. These datasets are also used for demonstrating proof concept for ranking adaption to different domains. The model tested with one data set will also work with other data set as the proposed model can adapt to different domains. With AdaBoost algorithm the SVM's performance is improved further. Therefore the results revealed that the proposed framework shows better performance when compared with [12]. More information about ranking model adaption concepts and empirical work can be found in [12]. The results of proposed framework are compared with that of [12].

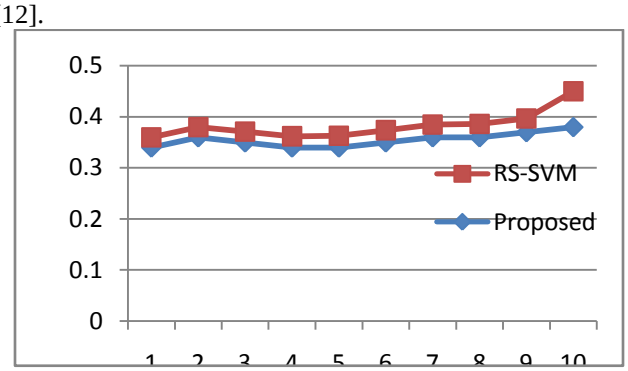

Fig. 1 - TD2003 to TD2004 adaptation with five queries

As can be seen in fig. 1, adaptation comparison is made between datasets TD2003 and TD2004. The proposed framework is compared with that of RA-SVM. The results reveal that the proposed approach outperforms RA-SVM.

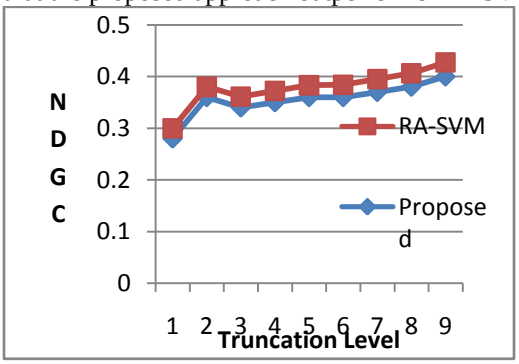

Fig. 2 - TD2003 to TD2004 adaptation with ten queries

As can be seen in fig. 2, adaptation comparison is made between datasets TD2003 and TD2004. The proposed

ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 2, Issue 8, August 2013

framework is compared with that of RA-SVM. The results reveal that the proposed approach outperforms RA-SVM.

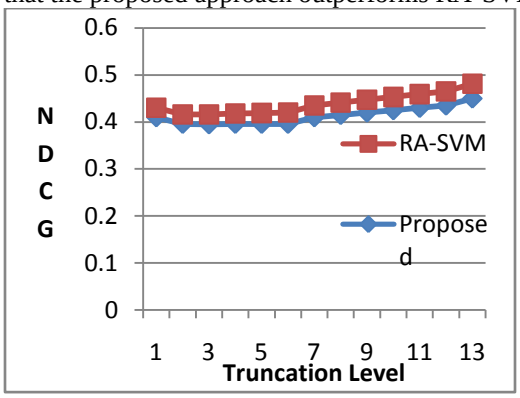

Fig. 3 – NDCG Results of web page search to image search adaptation with five labeled queries

As can be seen in fig. 3, adaptation comparison is made between datasets TD2003 and TD2004. The proposed framework is compared with that of RA-SVM. The results reveal that the proposed approach outperforms RA-SVM.

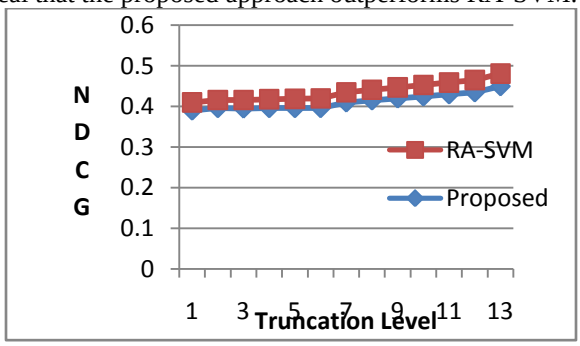

Fig. 4 – NDCG Results of web page search to image search adaptation with ten labeled queries

As can be seen in fig. 4, adaptation comparison is made between datasets TD2003 and TD2004. The proposed framework is compared with that of RA-SVM. The results reveal that the proposed approach outperforms RA-SVM.

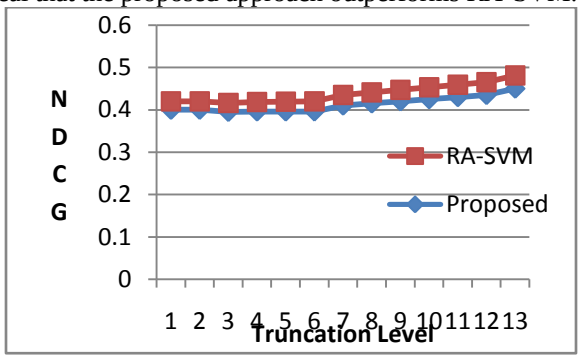

Fig. 5 – NDCG Results of web page search to image search adaptation with twenty labeled queries

As can be seen in fig. 5, adaptation comparison is made between datasets TD2003 and TD2004. The proposed framework is compared with that of RA-SVM. The results reveal that the proposed approach outperforms RA-SVM.

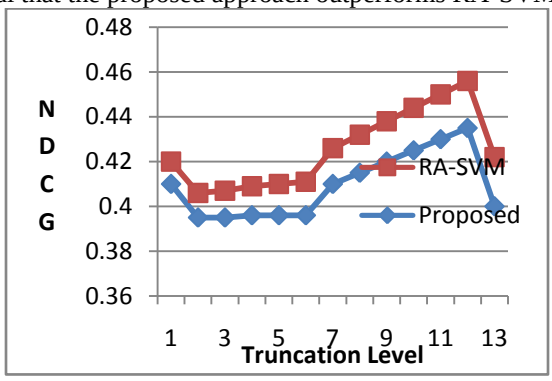

Fig. 6 – NDCG Results of web page search to image search adaptation with thirty labeled queries

As can be seen in fig. 6, adaptation comparison is made between datasets TD2003 and TD2004. The proposed framework is compared with that of RA-SVM. The results reveal that the proposed approach outperforms RA-SVM.

V. CONCLUSION

In this paper we implemented a new ranking model adaptation framework which combines the SVM and AdaBoost algorithms. The framework is built in such a way that it can be adapted to different domains. The domain specific search is possible without building different ranking models for different domains. This can't be achieved with domain specific search engines and broad based searching algorithms. Our decision to use AdaBoost along with SVM for ranking has shown positive results. We also built a prototype application with web based interface which demonstrates the proof of concept. The application is tested with data sets such as TD2003 and TD2004. The empirical results revealed that the proposed framework is adaptive to different domains.

REFERENCES

[1] R. Herbrich, T. Graepel, and K. Obermayer, "Large Margin Rank Boundaries for Ordinal Regression," Advances in Large Margin Classifiers, pp. 115-132, MIT Press, 2000.
[2] T. Joachims, "Optimizing Search Engines Using Clickthrough Data," Proc. Eighth ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining (KDD '02), pp. 133-142, 2002.
[3] Y. Freund, R. Iyer, R.E. Schapire, Y. Singer, and G. Dietterich, "An Efficient Boosting Algorithm for Combining Preferences," J. Machine Learning Research, vol. 4, pp. 933-969, 2003.
[4] C.J.C. Burges, R. Ragno, and Q.V. Le, "Learning to Rank with Nonsmooth Cost Functions," Proc. Advances in Neural Information Processing Systems (NIPS '06), pp. 193-200, 2006.
[5] C.J.C. Burges, T. Shaked, E. Renshaw, A. Lazier, M. Deeds, N. Hamilton, and G. Hullender, "Learning to Rank Using Gradient Descent," Proc. 22th Int'l Conf. Machine Learning (ICML '05), 2005.

ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 2, Issue 8, August 2013

[6] Z. Cao and T. Yan Liu, "Learning to Rank: From PairwiseApproach to Listwise Approach," Proc. 24th Int'l Conf. Machine Learning (ICML '07), pp. 129-136, 2007.
[7] W. Dai, Q. Yang, G.-R. Xue, and Y. Yu, "Boosting for Transfer Learning," Proc. 24th Int'l Conf. Machine Learning (ICML '07), pp. 193-200, 2007.
[8] H. Shimodaira, "Improving Predictive Inference Under Covariate Shift by Weighting the Log-Likelihood Function," J. Statistical Planning and Inference, vol. 90, no. 18, pp. 227-244, 2000.
[9] J. Yang, R. Yan, and A.G. Hauptmann, "Cross-Domain Video Concept Detection Using Adaptive Svms," Proc. 15th Int'l Conf. Multimedia, pp. 188-197, 2007.
[10] B. Zadrozny, "Learning and Evaluating Classifiers Under Sample Selection Bias," Proc. 21st Int'l Conf. Machine Learning (ICML '04), p. 114, 2004.
[11] R. Klinkenberg and T. Joachims, "Detecting Concept Drift with Support Vector Machines," Proc. 17th Int'l Conf. Machine Learning (ICML '00), pp. 487-494, 2000.
[12] Bo Geng, Linjun Yang, Chao Xu, and Xian-Sheng Hua, "Ranking Model Adaption for Domain Specific Search", IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 24, NO. 4.
[13] J. Lafferty and C. Zhai, "Document Language Models, Query Models, and Risk Minimization for Information Retrieval," Proc. 24th Ann. Int'l ACM SIGIR Conf. Research and Development in Information Retrieval (SIGIR '01), pp. 111-119, 2001.
[14] J.M. Ponte and W.B. Croft, "A Language Modeling Approach to Information Retrieval," Proc. 21st Ann. Int'l ACM SIGIR Conf. Research and Development in Information Retrieval, pp. 275-281, 1998.
[15] J. Blitzer, R. Mcdonald, and F. Pereira, "Domain Adaptation with Structural Correspondence Learning," Proc. Conf. Empirical Methods in Natural Language Processing (EMNLP '06), pp. 120-128, July 2006.
[16] H. Daume III and D. Marcu, "Domain Adaptation for Statistical Classifiers," J. Artificial Intelligence Research, vol. 26, pp. 101-126, 2006.

BIOGRAPHIES

Greeshma is student of DRK College of Engineering and Technology, Hyderabad, AP, INDIA. She has received B.Tech Degree computer science and engineering, M.Tech Degree in computer science and engineering. Her main research interest includes data mining, Databases and DWH.

M.Srinivasa Rao is working as an Associate Professor in DRK College of Engineering and Technology, JNTUH, Hyderabad, Andhra Pradesh, India. He is pursuing Ph.D in Information Security. He has completed M.Tech (C.S.E) from JNTUH. His main research interest includes Information Security and Computer Ad-Hoc Networks.

Dr.R.V.Krishnaiah (Ph.D) is working as Principal at DRK INSTITUTE OF SCINCE & TECHNOLOGY, Hyderabad, AP, INDIA. He has received M.Tech Degree EIE and CSE. His main research interest includes Data Mining, Software Engineering.