Design and Implementation of Ranking Adaptation Algorithm for Domain Specific Search

B.Maheswari [1], K. Chandra Shekhar Reddy[2], Prof.S.V.Achutha Rao[3]

[1]*Pursuing M.Tech(CSE), Vikas College of Engineering and Technology, Nunna, Vijayawada.Affliated to JNTUK, Kakinada, A.P, India*

[2]*K.Chandra Shekhar Reddy is working as an Associate Professor with Department of CSE at Vikas College of Engineering and Technology, Nunna, Vijayawada, India.*

[3] *S.V. Achuta Rao is working as a Professor & Head of the Department of CSE at Vikas College of Engineering and Technology , Nunna, Vijayawada, India.*

***Abstract*- In the market, various domain-specific search engines emerged, which are restricted to specific topicalities or document formats, and vertical to the broad-based search. Simply applying the ranking model trained for the broad-based search to the verticals cannot achieve a sound performance due to the domain differences, while building different ranking models for each domain is both laborious for labelling sufficient training samples and time-consuming or the training process. In this paper, to address the above difficulties, we investigate two problems: (1) whether we can adapt the ranking model learned for existing Web page search or verticals, to the new domain, so that the amount of labelled data and the training cost is reduced, while the performance requirement is still satisfied; and (2) how to adapt the ranking model from auxiliary domains to a new target domain. We address the second problem from the regularization framework and an algorithm called ranking adaptation SVM is proposed. The results demonstrate the applicability's of the proposed ranking model adaptation algorithm and the ranking adaptability measurement.**

***Keywords*— Search engine, Ranking Models, Framework.**

INTRODUCTION

A search engine is a program that can search the Web on a specific topic. By typing in a word or phrase (known as a keyword), the search engine will produce pages of links on that topic. The more relevant links are at the top of the list, but that is not always true. The information contains in the search engines is may be specialist in web pages, images and other types of files. The learning to rank is a kind of learning based information restoration techniques, specialized in the learning a ranking model with some documents labeled with their relevancies to some queries. Then the model is hopefully capable of ranking the documents returned to new queries groups automatically. The performances of the learning to rank algorithm have already showed in the information retrieval. It is especially in the web search services. Page Rank is a link analysis algorithm, named after Larry Page and used by the Google web search engine, that assigns a numerical weighting to each element of a hyperlinked set of documents, such as the World Wide Web (WWW), with the purpose of "measuring" its relative importance within the set. The algorithm may be applied to any collection of entities with reciprocal quotations and references. Page Rank is not Google's only algorithm that determines rankings in search results, but merely one of many factors used to determine ranking websites in search results pages for any given query.

Nowadays, this domain specific search engines are very powerful because of their accuracy and it consists of additional features on the different web search engines like Delta search, Google search and Bing search engines are mostly used search engines nowadays. In the emerging search engines domains, to implement a broad-based ranking model directly to different domain is no longer desirable due to the domain differences, while building a unique ranking model for each domain is both laborious for labeling data and time consuming for training models. There are many vertical search engines are introduced in recent years, they contains different domain-specific features, document types and topicalities. Let us take an example on image or music or videos, these are just focusing on the documents in particular formats, that is why these broad based and vertical search engines are becoming most important for the text based search formats recently.

For example web images are considered as text-based documents that is share with similar ranking features as a document or web page ranking, and text –based ranking model can be applied here directly. Nevertheless, the broad based ranking models are built upon the data from the multiple domains; it would not be generalize for particular domain. The broad based ranking model can only utilize the vertical domain's ranking features that are same to the broad

based domains for ranking, while the domain-specific features, such as the content features of images, videos, or music cannot be utilized directly. And those features are generally important for the same representation of the documents. And also it should used to built a more efficient and powerful model for the vertical i.e., known as ranking adaptation model.

Lets us discuss about domain specific search engines in olden days, it gives the lots of common information in the searched documents, only few new documents will be given as correct documents without giving common information so that the broad based ranking model can completely adapted to the new domains with many features, so it gives the result very efficiently because using this vertical search applications we can get the efficient documents without giving any repeated data to the user.

There are many services provided by the ranking models are effectively some issues are risen in these model. The general difficulties faced by the classifier adaption namely: covariate shift and concept drifting and it have more challenging compared to the ranking models. The classifier adaptation, which mainly deals with binary targets, ranking adaptation desires to adapt the model which is used to predict the rankings for a collection of documents. Though the documents are normally labeled with several relevance levels, which seem to be able to be handled by a multiclass classification or regression, it is still difficult to directly use classifier adaption for ranking. The reason lies in two ways: 1) in ranking, the mainly concentrated is about the preference of two documents or the ranking of a collection of documents, which is difficult to be modeled by classification or regression; 2) the relevance levels between different domains are sometimes different and need to be aligned.

In this mainly we can discuss about issues of an existing ranking model, to overcome these issues a new ranking model or technique in proposing system known as Ranking Model Adaptation for the Domain Specific search (RA-SVM). In this model it will take the auxiliary domains directly instead of using the labeled data. And also we overcome the problems of an existing ranking model those we can identify in below 3 statements.

1. Whether the amount of labeled data in the target domain is reduced while the performance requirement is still guaranteed?

2. How to adapt the ranking model effectively and efficiently?

3. How to utilize domain-specific features to further boost the model adaptation?

RELATED WORK

In this, we described some works that are closely related to ranking model they rank the documents according to its relevant query. Some of the models have been proposed, which also have been successfully applied to web search engines. Classical BM and Language models for information retrieval work quite stable for the broad-based search adjusted. Broad-based ranking model provides a lot of common information in ranking documents only few training samples are needed to be labelled in the new domain. From the probabilistic perspective, the broad-based ranking model provides a prior knowledge, so that only a small number of labelled samples are sufficient for the target domain ranking model to achieve the same confidence. Hence, to reduce the cost for new verticals, how to adapt the auxiliary ranking models to the new target domain and make full use of their domain-specific features, turns into a pivotal problem for building effective. In this paper, instead of designing a new learning algorithm, we focus on the adaptation of ranking models across different domains based on the existing learning to rank algorithms. A lot of domain adaptation methods have been proposed to adapt auxiliary data to a new domain Adaptive SVM algorithm for the cross-domain video concept detection problem. However, these works are mainly designed for classification problems, while we focused on the domain adaptation problem for ranking in this paper.

Satoshi Oyama, Takashi Kokubo,(Both Satoshi Oyama and Takashi Kokubo we can find in the references of this domain specific search model) and Toru Ishida [this name too we can find in references], — Domain-specific Web search engines are effective tools for reducing the difficulty experienced when acquiring information from the Web. Existing methods for building domain-specific Web search engines require human expertise or specific facilities. However, we can build a domain-specific search engine simply by adding domain-specific keywords, called “keyword spices,” to the user’s input query and forwarding it to a general-purpose Web search engine. Keyword spices can be effectively discovered from Web documents using machine learning technologies. This paper will describe domain-specific Web search engines that use keyword like different domains i.e. Java, .Net, cars and so on.

And also we present some of the concepts of ranking model adaptation here. To create a ranking that can rank the particular documents according to their relevance to a given query, in this paper we have various types of ranking models have been implemented successfully applied to the web search engines. Classical BM25 and language models getting information is quite stable for the broad-based search with some of the parameters needed to be adjusted. However, with the development of statistical learning models, and more labelled information with complicated features being are available for achieving better ranking performance.

PROPOSED WORK

In this paper, we mainly focus on the issues of an existing model. In order to overcome the issues new technique in proposing in this paper namely Ranking Model Adaptation for the domain-specific search (RA-SVM). In the ranking models, instead of utilizing the labeled data from auxiliary domains directly, it should be inaccessible due to the privacy issues or data missing. Model adaptation is more desirable than data adaptation, because the learning complexity is now only correlated with the size of the target domain training set, which should be much smaller than the size of auxiliary data set. There are three issues raised in an existing ranking model.

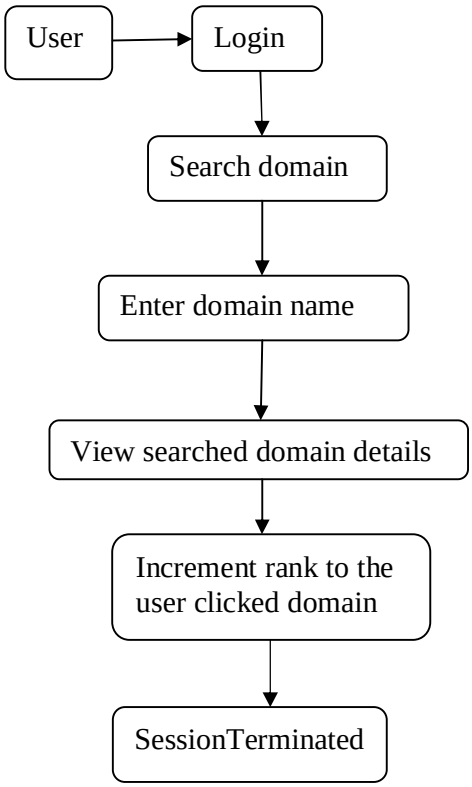
To solve the first problem the proposed ranking adaptability measure, this quantitatively estimates and predicts to the potential performance for the adaptation. And to solve the second issue we proposed the algorithm as ranking adaptation SVM (RA-SVM). Our algorithm is a black box ranking model adaptation. The black-box adaptation property not only achieved flexibility and also the efficiency. To resolve the third problem, we assume that documents similar in their domain-specific feature space should have consistent rankings, e.g., images that are similar in their visual feature space should be ranked into similar positions and vice versa. We implement this idea by constraining the margin and slack variables of RA-SVM adaptively, so that similar documents are assigned with less ranking loss if they are ranked in a wrong order.

In the setting of the proposed ranking adaptation, both the number of queries m and the number of the returned documents n (qi) in the training set are assumed to be small. They are insufficient to learn an effective ranking model for the target domain. However, an auxiliary ranking model fa, which is well trained in another domain over the labeled data Qa and Da, is available. It is assumed that the auxiliary ranking model fa contains a lot of prior knowledge to rank documents, so it can be used to act as the base model to be adapted to the new domain. Few training samples can be sufficient to adapt the ranking model since the prior knowledge is available.

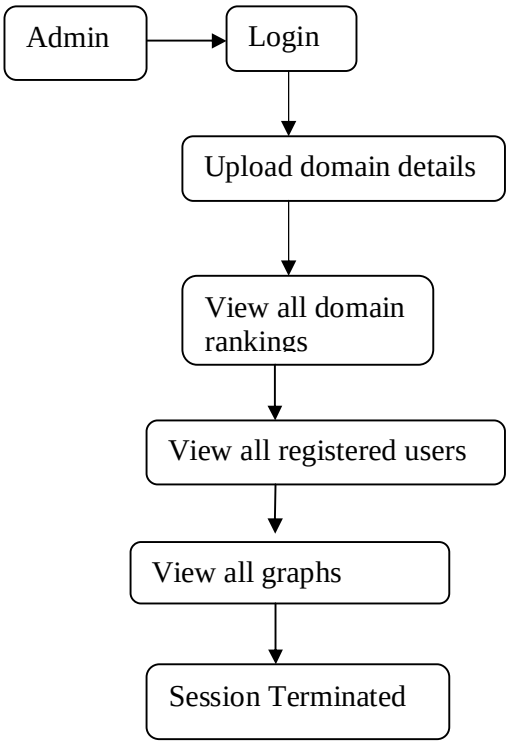
Where m-represents is the no of queries and n (qi) - represents no of returned documents, fa-represents auxiliary ranking model and Qa and Da represents trained data from another domain. Queries play a major role along with the implementation of the RA-SVM model, with the algorithm we can show the data as per the required analysis but with queries we can format depending upon the need of the user. So, SQL queries play an important role.

Below block diagrams show the working of the designed model with two modules considered in this paper,

At user:



At admin:



CONCLUSION

We have implemented a Ranking Adaptation Algorithm which will give the appropriate result to the user by searching in it the domains of various companies. The link which comes on top of the search is purely based on the number of users consuming that data for their reference. With this approach many users will get benefited and they will get the expected results in less time as the algorithm implemented will apply the ranking function and will give the highest rank allotted data for the user specific search. Based on the work Ranking Adaptation SVM (RA-SVM) algorithm is implemented and it has two variations in the implementation like slack rescaling and margin rescaling. Margin Rescaling is more preferred than Slack Rescaling.

REFERENCES

M.G. Kendall, "A New Measure of Rank Correlation," Biometrika, vol. 30, nos. 1/2, pp. 81-93, June 1938.

J.M. Kleinberg, S.R. Kumar, P. Raghavan, S. Rajagopalan, and A. Tomkins, "The Web as a Graph: Measurements, Models and Methods," Proc. Int'l Conf. Combinatorics and Computing, pp. 1-18, 1999.

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.

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.

Z. Cao and T. Yan Liu, "Learning to Rank: From Pairwise Approach to Listwise Approach," Proc. 24th Int'l Conf. Machine Learning (ICML '07), pp. 129-136, 2007.

M. Belkin, P. Niyogi, and V. Sindhwani, "Manifold Regularization: A Geometric Framework for Learning from Labeled and Unlabeled Examples," J. Machine Learning Research, vol. 7, pp. 2399-2434, Nov. 2006.

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.

K. Ja¨rvelin and J. Keka¨la¨inen, "Ir Evaluation Methods for Retrieving Highly Relevant Documents," Proc. 23rd Ann. Int'l ACM SIGIR Conf. Research and Development in Information Retrieval (SIGIR '00), pp. 41-48, 2000.

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, July2006.

J. Cui, F. Wen, and X. Tang, "Real Time Google and Live Image Search Re-Ranking," Proc. 16th ACM Int'l Conf. Multimedia, pp. 729- 732, 2008.

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.

H. Daume III and D. Marcu, "Domain Adaptation for Statistical Classifiers," J. Artificial Intelligence Research, vol. 26, pp. 101-126, 2006.

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.

AUTHORS PROFILE

B.Maheswari, Pursuing M.Tech(CSE) Vikas College of Engineering and Technology , Nunna, Vijayawada. Affiliated to JNTU-Kakinada, A.P., India

K. Chandra Shekhar Reddy is working as an Associate professor at Vikas College of Engineering and Technology, Nunna, Vijayawada, Affiliated to JNTU-Kakinada, A.P., India

Prof S.V.Achutha Rao, is working as an HOD of CSE at Vikas College of Engineering and Technology, Nunna, Vijayawada, Affiliated to JNTU-Kakinada, A.P., India