Volume 5, Issue 1, January 2015 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Domain Specific Search using Ranking Model Adaptation

[1]Priyanka Chandragiri, [2]A. Soumya, [3]Ch. Srinivasa Rao
[1]Associate Professor in CSE, [2]Assistant Professor in CSE, [3]Professor in CSE
Christu Jyothi Institute of Technology and Science, Telangana, India

Abstract: Applying the broad-based ranking model directly to different domains is not desirable due to domain differences it is time-consuming for training models while building a unique ranking model for each domain. We address these difficulties by proposing a regularization based algorithm called ranking adaptation SVM (RA-SVM), An adaptation process is described to adapt a ranking model constructed for a broad-based search engine for use with a domain-specific ranking model . We can adapt an existing ranking model to a new domain, so that the amount of labeled data and the training cost is reduced while the performance is still guaranteed. We only require the Prediction from the existing ranking models. We assume that documents similar in the domain-specific feature space should have consistent rankings, and add some constraints to control the margin and slack variables of RA-SVM adaptively. Finally, ranking adaptability measurement is proposed to be assumed that documents similar in their domain specific feature space should have consistent rankings.

Index terms: Ranking, Ranking Adaptation, Domain.

I. INTRODUCTION

The most basic assumption used in statistical learning theory is that training data and test data are drawn from the same underlying distribution. Unfortunately, in many applications, the "in domain" test data is drawn from a distribution that is related, but not identical, to the "out-of-domain" distribution of the training data[1]

We study the problem of learning to accurately rank a set of objects by combining a given collection of ranking or preference functions. This problem of combining preferences arises in several applications, such as that of combining the results of different search engines, or the "collaborative filtering" problem of ranking movies for a user based on the movie rankings provided by other users[2] . Intuitively, a good information retrieval system should present relevant documents high in the ranking, with less relevant documents following below. While previous approaches to learning retrieval functions from examples exist, they typically require training data generated from relevance judgments by experts. The goal of this paper is to develop a method that utilizes click through data for training, namely the query-log of the search engine in connection with the log of links the users clicked on in the presented ranking[3]

The existing broad-based ranking model provides a lot of common information in ranking documents only few training samples are needed to be labeled in the new domain. 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.

An adaptation process is described to adapt a ranking model constructed for a broad-based search engine for use with a domain-specific ranking model. An example process identifies a ranking model for use with a broad-based search engine and modifies that ranking model for use with a new (or "target") domain containing information pertaining to a specific topic. Based on various machine learning methods, e.g., Ranking SVM the learning to rank algorithms has already shown their promising performances in information retrieval, especially Web search. However, as the emergence of domain-specific search engines, more attentions have moved from the broad based search to specific verticals. Web images are actually treated as text-based documents that share similar ranking features as the document or Web page ranking, and text based ranking model can be applied directly. In addition, 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. Those features are generally important for the semantic representation of the documents and should be utilized to build a more robust ranking model for the particular vertical.

Proposed System focus whether we can adapt ranking models learned for the existing broad-based search or some verticals, to a new domain, so that the amount of labeled data in the target domain is reduced while the performance requirement is still guaranteed, how to adapt the ranking model effectively and efficiently and how to utilize domain-specific features to further boost the model adaptation.

II. IMPLEMENTATION AND ANALYSIS

The objective is solved by proposing a ranking adaptability measure, that quantitatively calculates whether an existing ranking model can be adapted to the new target domain, and determines the level of performance for the adaptation. The analysis of the system has been identified with Ranking Adaptation ,Explore ranking adaptability ,Ranking adaptation with domain specific search ,Ranking Support Vector Machine

Ranking adaptation is closely related to classifier adaptation, which has shown its effectiveness for many learning problems. Ranking adaptation is comparatively more challenging. Unlike 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 domains. In ranking the relevance levels between different domains are sometimes different and need to be aligned. we can adapt ranking models learned for the existing broad-based search or some verticals, to a new domain, so that the amount of labeled data in the target domain is reduced while the performance requirement is still guaranteed and how to adapt the ranking model effectively and efficiently .Then how to utilize domain-specific features to further boost the model adaptation.

Ranking adaptability measurement by investigating the correlation between two ranking lists of a labeled query in the target domain, i.e., the one predicted by f^a and the ground-truth one labeled by human judges. Intuitively, if the two ranking lists have high positive correlation, the auxiliary ranking model f^a is coincided with the distribution of the corresponding labeled data, therefore we can believe that it possesses high ranking adaptability towards the target domain, and vice versa. This is because the labeled queries are actually randomly sampled from the target domain for the model adaptation, and can reflect the distribution of the data in the target domain.

Data from different domains are also characterized by some domain-specific features, e.g., when we adopt the ranking model learned from the Web page search domain to the image search domain, the image content can provide additional information to facilitate the text based ranking model adaptation. In this section, we discuss how to utilize these domain-specific features, which are usually difficult to translate to textual representations directly, to further boost the performance of the proposed RA-SVM. The basic idea of our method is to assume that documents with similar domain-specific features should be assigned with similar ranking predictions. We name the above assumption as the consistency assumption, which implies that a robust textual ranking function should perform relevance prediction that is consistent to the domain-specific features.

Ranking Support Vector Machines (Ranking SVM), which is one of the most effective learning to rank algorithms, and is here employed as the basis of our proposed algorithm. The proposed RA-SVM does not need the labeled training samples from the auxiliary domain, but only its ranking model f^a. Such a method is more advantageous than data based adaptation, because the training data from auxiliary domain may be missing or unavailable, for the copyright protection or privacy issue, but the ranking model is comparatively easier to obtain and access.

2.1 Design and Implementation Constraints

All modules are coded thoroughly based on requirements from software organization.
The software is designed in such a way that the user can easily interact with the screen. Software is designed in such a way that it can be extended to the real time business.

2.2 User Interfaces

This application include GUI standards or product family style guides that are to be followed, screen layout constraints, buttons and functions that will appear on every screen, error message display standards, and so on.

2.3 User Classes and Characteristics

End user of the application is the customer or anyone who uses a certain product. The user can provide comments on any product they have used. They can search the products of any domain which are available. The admin can view the comments received and he can enhance the methods to improve the quality of the products that are falling in the corresponding domains. The admin can view the graph in terms of good comments received so far for the products of the particular domain.

III. EXPERIMENTS

We perform several experiments under two different settings, to demonstrate the effectiveness of the proposed RA-SVM based algorithms and the ranking adaptability measurement.

Table 1 ranking adaptation dataset information

Dataset	#Query	#Query-Document	Relevance Degree	Feature Dimension
TD203	50	39171	2	44
TD204	85	8750	2	44
Web Page Search	1675	112875	5	354
Image Search	1121	51046	3	354

IV. CONCLUSION

The ranking model adaptation is proposed to adapt the well learned models from the broad-based search or any other auxiliary domains to a new target domain, i.e., only the relevance predication of the auxiliary ranking models is needed for the adaptation. Based on RASVM, two variations called RA-SVM margin rescaling (RA-SVM-MR) and RA-

SVM slack rescaling (RA-SVMSR) are proposed to utilize the domain specific features to further facilitate the adaptation, by assuming that similar documents should have consistent rankings, and constraining the margin and loss of RA-SVM adaptively according to their similarities in the domain-specific feature space. Furthermore, we propose *ranking adaptability*, to quantitatively measure whether an auxiliary model can be adapted to a specific target domain and how much assistance it can provide. Based on the results, we can derive the following conclusions: The proposed RA-SVM can better utilize both the auxiliary models and target domain labeled queries to learn a more robust ranking model for the target domain data. The utilization of domain-specific features can steadily further boost the model adaptation, and RA-SVM-SR is comparatively more robust than RASVM- MR. The adaptability measurement is consistent to the utility of the auxiliary model, and it can be deemed as an effective criterion for the auxiliary model selection. The proposed RA-SVM is as efficient as directly learning a model in a target domain, while the incorporation of domain-specific features doesn't brings much learning complexity for algorithms RASVM-SR and RA-SVM-MR.

REFERENCES

[1] M. Belkin, P. Niyogi, and V. Sindhwani. Manifold regularization:A geometric framework for learning from labeled and unlabeled examples. *Journal of Machine Learning Research*, 7(Nov):2399–2434, 2006.

[2] J. Blitzer, R. Mcdonald, and F. Pereira. Domain adaptation with structural correspondence learning. In *EMNLP '06: Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing*, pages 120–128. Association for Computational Linguistics, July 2006.

[3] C. J. C. Burges, R. Ragno, and Q. V. Le. Learning to rank with nonsmooth cost functions. In *NIPS '06: Advances in Neural Information Processing Systems*, pages 193–200. MIT Press, Cambridge, MA, 2006.

[4] C. J. C. Burges, T. Shaked, E. Renshaw, A. Lazier, M. Deeds, N. Hamilton, and G. Hullender. Learning to rank using gradient descent. In *ICML '05: Proceedings of the 22th International Conference on Machine Learning*, 2005.

[5] Z. Cao and T. yan Liu. Learning to rank: From pairwise approach to listwise approach. In *ICML '07: Proceedings of the 24th Interna-tional Conference on Machine Learning*, pages 129–136, 2007.

[6] J. Cui, F. Wen, and X. Tang. Real time google and live image search re-ranking. In *ACM Multimedia*, pages 729–732, 2008.

ABOUT AUTHOR

Ms Priyanka Chandragiri received her Bachelor's and Master's degrees in Computer Science & Engineering in the years 2006 and 2011 resp. from Jawaharlal Nehru Technological University, Hyderabad. Currently she is working as Associate Professor in Computer Science & Engineering at Christu Jyothi Institute of Technology & Sciences, Jangaon, Warangal. She has published 5 research papers in various National/ International journals. Her areas of research are wireless and Sensor Networks and Network security.

Mrs A. Soumya received her Bachelor's and Master's degrees in Information Technology in the years 2007 and 2011 resp. from Jawaharlal Nehru Technological University, Hyderabad. Currently she is working as Sr. Assistant Professor in Computer Science & Engineering at Christu Jyothi Institute of Technology & Sciences, Jangaon, Warangal. She has published 2 research papers in various National/ International journals. Her areas of research are Data Mining and Computer Networks.

Mr.CH. Srinivasa Rao working as Professor and Head in Computer Science & Engineering at Christu Jyothi Institute of Technology & Sciences, Jangaon, Warangal.

