JES
Journal of Engineering Sciences

Vol 13, Issue 05, May /2022
ISSN NO:0377-9254

RANKING MODEL ADAPTATION FOR DOMAIN-SPECIFIC SEARCH

M.SIVA, A. SIVA KUMAR
DEPARTMENT OF MCA
Sri Padmavathi College Of Computer Sciences & Technology

ABSTRACT
With the explosive emergence of vertical search domains, applying the broad-based ranking model directly to different domains is no longer desirable due to domain differences, while building a unique ranking model for each domain is both laborious for labeling data and time-consuming for training models. In this paper, we address these difficulties by proposing a regularization based algorithm called ranking adaptation SVM (RA-SVM), through which 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. Our algorithm only requires the Prediction from the existing ranking models, rather than their internal representations or the data from auxiliary domains. In addition, 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 quantitatively estimate if an existing ranking model can be adapted to a new domain. Experiments performed over Letor and two large scale datasets crawled from a commercial search engine demonstrate the applicabilities of the proposed ranking adaptation algorithms and the *ranking adaptability* measurement.

I. INTRODUCTION

LEARNING to rank is a kind of learning based information retrieval techniques, specialized in learning a ranking model with some documents labeled with their relevancies to some queries, where the model is hopefully capable of ranking the documents returned to an arbitrary new query automatically. Based on various machine learning methods, e.g., Ranking SVM the learning to rank algorithms have 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 broadbased search to specific verticals, for hunting information constraint to a certain domain. Different vertical search engines deal with different topicalities, document types or domain-specific features. For example, a medical search engine should clearly be specialized in terms of its topical focus, whereas a music, image or video search engine would concern only the documents in particular formats.

Since currently the broad-based and vertical search engines are mostly based on text search techniques, the ranking model learned for broad-based can be utilized directly to rank the documents for the verticals. For, example, most of current image search engines only utilize the text information accompanying images as the ranking features, such as the term frequency (TF)
of query word in image title, anchor text, alternative text, surrounding text, URL and so on. Therefore, 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 here directly. However, the broad-based ranking model is built upon the data from multiple domains, and therefore cannot generalize well for a particular domain with special search intentions. In addition, the broad-based ranking model can only utilize the vertical domain's ranking features that are same to the broadbased domain's for ranking, while the domain-specific features, such as the content features of images, videos or music can not 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.

II. EXISTING SYSTEM

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. From the probabilistic perspective, the broad-based ranking model provides a prior knowledge, so that only a small number of labeled 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 domain-specific ranking models.

III. PROPOSED SYSTEM

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. The first problem is solved by the proposed *rank-ing adaptability* measure, which quantitatively estimates whether an existing ranking model can be adapted to the new domain, and predicts the potential performance for the adaptation. We address the second problem from the regularization framework and a ranking adaptation SVM algorithm is proposed. Our algorithm is a black box ranking model adaptation, which needs only the predictions from the existing ranking model, rather than the internal representation of the model itself or the data from the auxiliary domains. With the black-box adaptation property, we achieved not only the flexibility but also the efficiency. To resolve the third problem, we assume that documents similar in their domain specific feature space should have consistent rankings.

Advantage:

1. Model adaptation.
2. Reducing the labeling cost.
3. Reducing the computational cost.

IV. IMPLEMENTATION

Implementation is the stage of the project when the theoretical design is turned out into a working system. Thus it can be considered to be the most critical stage in achieving a successful new system and in giving the user, confidence that the new system will work and be effective.

The implementation stage involves careful planning, investigation of the existing system and it's constraints on implementation, designing of methods to achieve changeover and evaluation of changeover methods.

Modules:

After careful analysis the system has been identified to have the following modules:

1. **Ranking Adaptation Module.**
2. **Explore ranking adaptability Module.**
3. **Ranking adaptation with domain specific search Module.**
4. **Ranking Support Vector Machine Module.**

1.Ranking adaptation Module:

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.

2.Explore Ranking adaptability Module:

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 fa and the ground-truth one labeled by human judges. Intuitively, if the two ranking lists have high positive correlation, the auxiliary ranking model fa 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.

3.Ranking adaptation with domain specific search Module:

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.

4.Ranking Support Vector Machines Module:

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 fa. 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.

V. CONCLUSION

As various vertical search engines emerge and the amount of verticals increases dramatically, a global ranking model, which is trained over a dataset sourced from multiple domains, cannot give a sound performance for each specific domain with special topicalities, document formats and domain-specific features. Building one model for each vertical domain is both laborious for labeling the data and time-consuming for learning the model. In this paper, we propose the ranking model adaptation, to adapt the well learned models from the broad-based search or any other auxiliary domains to a new target domain. By model adaptation, only a small number of samples need to be labeled, and the computational cost for the training process is greatly reduced. Based on the regularization framework, the Ranking Adaptation SVM (RA-SVM) algorithm is proposed, which performs adaptation in a black-box way, 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. We performed several experiments over Letor benchmark datasets and two large scale datasets obtained from a commercial internet search engine, and adapted the ranking models learned from TD2003 to TD2004 dataset, as well as from Web page search to image search domain. 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 Pro-cessing*, 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 Infor-mation 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.
[7] W. Dai, Q. Yang, G.-R. Xue, and Y. Yu. Boosting for transfer learning. In *ICML '07: Proceedings of the 24th international conference on Machine learning*, pages 193–200, 2007.
[8] H. Daume, III and D. Marcu. Domain adaptation for statistical classifiers. *Journal of Artficial Intelligence Research*, 26:101–126, 2006.
[9] Y. Freund, R. Iyer, R. E. Schapire, Y. Singer, and G. Dietterich. An efficient boosting algorithm for combining preferences. *Journal of Machine Learning Research*, 4:933–969, 2003.
[10] B. Geng, L. Yang, C. Xu, and X.-S. Hua. Ranking model adaptation for domain-specific search. In *CIKM '09: Proceeding of the 18th ACM conference on Information and knowledge management*, pages 197–206, 2009.