ISSN 2229-5518

Ranking Algorithm RA-SVM for Domain Adaptation: A Review

Paras V. Kothare, Yogesh K. Gedam, Ratnadeep R. Deshmukh

Abstract— Different new vertical domains are coming everyday so running a broad-based ranking model is no longer desirable as the domain are different and building a separate model for each domain is also not beneficial because there much time required for labeling the data and training the samples. In this paper we are handling the above problem by regularization based algorithm called as ranking adaptation SVM (RA-SVM), the algorithm is used to adapt existing ranking model of broad-based search engine to new domain. Here performance is still guaranteed and times taken to label the data training the samples are reduced. The algorithms only requires prediction from existing ranking model and do not require internal structure of it. Adapted ranking model concentrate on specific domain to achieve better results which are relevant to the search, further it reduces the searching cost also as the most appropriate search results are shown. Single ranking model is not good for training the search engine as the information retrieval is complicated, Domains are highly great, used in the global search engines and the data set is also large. So we can't generalize the information well for specific search intensions. That is why we are moving towards training the global ranking model for each specific domain for fetching the appropriate information from each respected domain for doing so we are using robust supervised classification algorithm. The parameters learned during the model adaptation and ranking SVM from global ranking model are capable retrieving the required information. Adapting the model is lot easier than building a unique ranking model for each domain.

Index Terms— Broad-based search, Regularization, Support vector machine (SVM), Adaption of model, Ranking Model, Supevise classification.

---

1 Introduction

Now a days people are more dependent on the internet for their day to day work, official work and academic work. As the result they want perfect result within short time. People perform their work Using search engines available on the net like Google, Bing, Yahoo etc. they insert search query in it. Search is an operation where the inserted key words are sent over the networks to the thousands web servers Consisting of corers of web pages. Keywords containing web pages and most relevant information are shown to the user by the search engines [1]. The search engine uses broad based ranking model for retrieving the information and the ranking model of broad-based engine search is build upon the data from multiple domain. Search performed by User with specific search intention couldn't get the specific information as it fail to generalize the information. Focus is now moving broad based ranking model to domain specific search for performing the Special search intentions, gives the best results free from anomalies [2].Learning to Rank is supervised learning technique where ranking model is to be learned through the repetitive machine learning process, Once the ranking model is learned it is hopefully capable of ranking the documents according to the query inserted by the user.

* *Paras V. Kothare*
  Student, Department of CS and IT, Dr Babasaheb Ambedkar Marathwada University, Aurangabad (MH), Email: paraskothare@gmail.com
* *Yogesh K. Gedam*
  Student, Department of CS and IT, Dr Babasaheb Ambedkar Marathwada University, Aurangabad (MH), Email: yogesh.gedam0103@gmail.com
* *Ratbadeep R. Deshmukh*
  Professor, Department of CS and IT, Dr Babasaheb Ambedkar Marathwada University, Aurangabad (MH), Email: rrdeshmukh.csit@bamu.ac.in

Based on the machine learning technique there are many ranking algorithms e.g. Lambda Rank [3], Rank Net [4], List Net [5], rank Boost [6] etc. Further domain specific features can be used to further boost the search result, like content features of the image, videos or music. While developing the algorithm for adaptation of ranking model majorly we are facing the three problems they are as fallow.

- Ranking model containing learned parameters over Broad based search engine.
- How to learn the parameters.
- Which parameters are to be used to fully acquire Domain specific features.

The algorithm handles the most of the problems as its using black box testing.

2 Related Works

Classifier adaptation which were mainly use in the broad based search engines are BM25 [7] and Language models For Information Retrieval [LMIR] [8] [9]. Which can rank the documents according to the query inserted by the user? By adjusting the few parameters they were best suited for broad based search engines, only problem was they were suffered by covariate shift and concept drifting. Mainly working on binary targets not the document as a whole. All were classification problem. Daume and Marcu proposed statistical learning method to adapt the domain which is mainly focusing on training and testing set. Similarly Boosting frame work and natural language were also proposed for the model adaptation. All above used learning to rank algorithms they are as fallows Ranking SVM, Rank Boost, List Net, Lambda Rank etc.

ISSN 2229-5518

However this paper concentrates on domain adaptation with RA-SVM rather than building new ranking model.

3 RANKING ADAPTATION

We define the ranking adaptation problem formally as follows: for the target domain, a query set Q {q_1,q_2,....... qm} and a document set D {d_1;d_2; ... ,d_N) are given. For each query q_i 2 Σ Q, a list of documents d_i =(d_{i1},d_{i2}; ... ;$d_{i;n}$(q)}$_i$ are returned and labeled with the relevance degrees y_i ={y_{i1};y_{i2}; ... ;$y_{i;n}$(q_i) } by human annotators. The relevance degree is usually a real value, i.e., y_{ij} Σ IR, so that different returned documents can be compared for sorting an ordered list. For each query document pair <q_i,d_{ij}>, an s- dimensional query dependent feature vector ǿ (q_i,d_{ij})Σ IRs is extracted, e.g., the term frequency of the query keyword qi in the title, body, URL of the document d_{ij}. Some other hyperlink based static rank information is also considered, such as Pagerank ,HITS and so on. n(q_i) denotes the number of returned documents for query q_i. The target of learning to rank is to estimate a ranking function f Σ IRs IR so that the documents d can be ranked for a given query q according to the value of the prediction f(ǿ) (q,d).

In the setting of the proposed ranking adaptation, both the number of queries m and the number of the returned documents n(q_i) 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 f^a , which is well trained in another domain over the labeled data Q^a and D^a, is available. It is assumed that the auxiliary ranking model f^a 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. Before the introduction of our proposed ranking adaptation algorithm, it's important to review the formulation of 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.

3.1 Raking SVM

Similar to the conventional Support Vector Machines (SVM) for the classification problem, the motivation of Ranking SVM is to discover a one dimensional linear subspace, where the points can be ordered into the optimal ranking list under some criteria. Thus, the ranking function takes the form of the linear model *f* (ǿ(q,d))where the bias parameter is ignored, because the final ranking list sorted by the prediction *f* is invariant to the bias. The optimization problem for Ranking SVM is defined as follows:

where C is the tradeoff parameter for balancing the large-margin regularization ||f||^2 and the loss term Σ $_{i,j,k}$Σ$_{ijk}$ Because f is a linear model , we can derive that f (ǿ(q_i,d_{ij}))- *f* (ǿ (q_i,d_{ik})) *f* (ǿ(q_i,d_{ik})) with *f* (ǿ (q_i,d_{ik}))- ǿ(q_i,d_{ik}) denoting the

$$\begin{aligned}
&\min_{f,\xi_{ijk}} \frac{1}{2}\|f\|^2 + C\sum_{i,j,k}\xi_{ijk}\\
&\text{s.t. } f(\phi(q_i,d_{ij})) - f(\phi(q_i,d_{ik})) \geq 1 - \xi_{ijk}\\
&\qquad \xi_{ijk} \geq 0,\\
&\text{for} \quad \forall i \in \{1,2,\ldots,M\},\\
&\qquad \forall j \forall k \in \{1,2,\ldots,n(q_i)\} \text{ with } y_{ij} > y_{ik}.
\end{aligned}$$

diffe pair d_{ij} and d_{ik}. If we further introduce the binary label sign (y_{ij} - y_{ik}) for each pair of documents d_{ij} and d_{ik}, the above Ranking SVM problem can be viewed as a standard SVM for classifying document pairs into positive or negative, i.e., whether the document d_{ij} should be ranked above d_{ik} or not.

3.2 Raking Adaptation SVM

It can be assumed that, if the auxiliary domain and the target domain are related, their respective ranking functions fa and f should have similar shapes in the function space IRs ! IR. Under such an assumption, *f*^{a} actually provides a prior knowledge for the distribution of *f* in its parameter space. The conventional regularization framework, such as L_p-norm regularization, manifold regularization designed for SVM, regularized neural network, and so on, shows that the solution of an ill-posed problem can be approximated from variational principle, which contains both the data and the prior assumption. Consequently, we can adapt the regularization framework which utilizes the f^a as the prior information, so that the ill-posed problem in the target domain, where only few query document pairs are labeled, can be solved elegantly. By modeling our assumption into the regularization term, the learning problem of Ranking Adaptation SVM can be formulated as

$$\begin{aligned}
&\min_{f,\xi_{ijk}} \frac{1-\delta}{2}\|f\|^2 + \frac{\delta}{2}\|f - f^a\|^2 + C\sum_{i,j,k}\xi_{ijk}\\
&\text{s.t. } f(\phi(q_i,d_{ij})) - f(\phi(q_i,d_{ik})) \geq 1 - \xi_{ijk}\\
&\qquad \xi_{ijk} \geq 0,\\
&\text{for} \quad \forall i \in \{1,2,\ldots,M\},\\
&\qquad \forall j \forall k \in \{1,2,\ldots,n(q_i)\} \quad \text{with } y_{ij} > y_{ik}.
\end{aligned}$$

The objective function consists of the adaptation regularization term ||*f* -*f*^{a}||^2, which minimizes the distance between the target ranking function and the auxiliary one in the function space or the parameter space, to make them close; the large-margin regularization ||*f*||^2 and the loss term Σ$_{i;j;k}$ ξ$_{ijk}$. The parameter ơξ (0,1) is a tradeoff term to balance the contributions of large-margin regular- ization ||*f*||^{2}which makes the learned model numerically stable, and adaptation regularization ||*f* -*f*^{a}||^2 which makes the learned model similar to the auxiliary one.

3.3 Raking SVM

To optimize Problem, we briefly denote x_{ijk}=ǿ(q_i;d_{ij}) ǿ(q_i;d_{ik}) and introduce the Lagrange multipliers to integrate the constraints into the objective function, which results in the primal problem

$$\begin{aligned}
L_P = &\frac{1-\delta}{2}\|f\|^2 + \frac{\delta}{2}\|f - f^a\|^2 + C\sum_{i,j,k}\xi_{ijk}\\
&- \sum_{i,j,k}\mu_{ijk}\xi_{ijk} - \sum_{i,j,k}\alpha_{ijk}(f(\mathbf{x}_{ijk}) - 1 + \xi_{ijk}).
\end{aligned}$$

Taking the derivatives of LP w.r.t. f, and setting it to zero, we can obtain the solution as

$$f(\mathbf{x}) = \delta f^a(\mathbf{x}) + \sum_{i,j,k}\alpha_{ijk}\mathbf{x}_{ijk}^T\mathbf{x}.$$

The proposed RA-SVM has several advantages, which makes

ISSN 2229-5518
our algorithm highly applicable and flexible when applied to the practical applications. We'll give more discussions of the characteristics of RA-SVM in the following.

- Model adaptation: 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.

Black-box adaptation: The internal representation of the model *f*a is not needed, but only the prediction of the auxiliary model to the training samples from the target domain $f^a(x)$ is used. It brings a lot of flexibilities in some situations where even the auxiliary model itself may be unavailable. Also, in some cases, we would like to use a more advanced algorithm for learning the ranking model for the new target domain, than the one used in the old auxiliary domain, or in other cases, the algorithm used in the old domain is even unknown to us. By the black-box adaptation property, we don't need to have any idea on the model used in the auxiliary domain, but only the model predictions are required.

- Reducing the labeling cost: by adapting the auxiliary ranking model to the target domain, only a small number of samples need to be labeled, while the insufficient training sample problem will be ad- dressed by the regularization term $||f - f^a||^2$, which actually assigns a prior to the target ranking model.

- Reducing the computational cost: It has been shown that our ranking adaptation algorithm can be transformed into a Quadratic Programming pro-blem, with the learning complexity directly related to the number of labeled samples in the target domain. Platt proposed the sequential minimal optimization (SMO) algorithm which can decom- pose a large QP problem into a series of subpro- blems and optimize them iteratively. The time complexity is around O (n2.3) for general kernels . cutting-plane method is adopted to solve SVM for the linear kernel, which further reduces the time complexity to O (n). Here, n is the number of labeled document pairs in the target domain. According to the above discussion, the size of the labeled training set is greatly reduced. Thus, n is substantially small, which in turn leads to the efficiency of our algorithm.

4 Ranking Adaptation With Domain Specific Features

Conventionally, data from different domains are also characterized by some domain-specific features, e.g., when we adopt the ranking model learned from the webpage 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. To implement the consistency assumption, we are inspired by the work and recall that for RA-SVM , the ranking loss is directly correlated with the slack variable, which stands for the ranking loss for pairwise documents, and is nonzero as long as the ranking function predicts a wrong order for the two documents. In addition, as a large margin machine, the ranking loss of RA-SVM is also correlated with the large margin specified to the learned ranker. Therefore, to incorporate the consistency constraint, we rescale the ranking loss based on two strategies, namely margin rescaling and slack rescaling. The rescaling degree is controlled by the similarity between the documents in the domain-specific feature space, so that similar documents bring about less ranking loss if they are ranked in a wrong order.

4.1 Margin Rescaling

Margin rescaling denotes that we rescale the margin violation adaptively according to their similarities in the domain-specific feature space. Specifically, the Ranking Adaptation SVM with Margin Rescaling (RA-SVM-MR) can be defined as the following optimization problem:

$$
\begin{aligned}
\min_{f,\xi_{ijk}} \quad & \frac{1-\delta}{2}||f||^2 + \frac{\delta}{2}||f - f^a||^2 + C\sum_{i,j,k}\xi_{ijk} \\
\text{s.t.} \quad & f(\phi(q_i, d_{ij})) - f(\phi(q_i, d_{ik})) \geq 1 - \xi_{ijk} - \sigma_{ijk} \\
& \xi_{ijk} \geq 0, \\
\text{for} \quad & \forall i \in \{1, 2, \ldots, M\}, \\
& \forall j \forall k \in \{1, 2, \ldots, n(q_i)\} \quad \text{with } y_{ij} > y_{ik},
\end{aligned}
$$

where $0 \leq \delta_{ijk} \leq ijk$ denotes the similarities between document d_{ij} and d_{ik} returned for query q_i in the domain-specific feature space. The above optimization problem differs from in the first linear inequality constraint, which varies the margin adaptively. Compared to a pair of dissimilar documents, similar ones with larger δ_{ijk} will result in a smaller margin to satisfy the linear constraint, which produces less ranking loss in terms of a smaller slack variable ξ_{ijk} if the document pair d_{ij} and d_{ik} (namely d_{ijk})is ranked in a wrong order by the function f. The dual problem is

$$
\begin{aligned}
\max_{\alpha_{ijk}} \quad & -\frac{1}{2}\sum_{i,j,k}\sum_{l,m,n}\alpha_{ijk}\alpha_{lmn}\mathbf{x}_{ijk}^T\mathbf{x}_{lmn} \\
& + \sum_{i,j,k}(1 - \delta f^a(\mathbf{x}_{ijk}) - \sigma_{ijk})\alpha_{ijk} \\
\text{s.t.} \quad & 0 \leq \alpha_{ijk} \leq C, \\
\text{for} \quad & \forall i \in \{1, 2, \ldots, M\}, \\
& \forall j \forall k \in \{1, 2, \ldots, n(q_i)\} \quad \text{with } y_{ij} > y_{ik},
\end{aligned}
$$

ar

as shown above.

4.2 Slack Rescaling

Compared to margin rescaling, slack rescaling is intended to rescale the slack variables according to their similarities in the domain specific feature space. We define the corresponding Ranking Adaptation SVM with Slack Rescal- ing (RA-SVM-SR) as the following optimization problem:

$$\begin{aligned}
&\min_{f,\xi_{ijk}} \frac{1-\delta}{2}\|f\|^2 + \frac{\delta}{2}\|f - f^a\|^2 + C\sum_{i,j,k}\xi_{ijk}\\
&\text{s.t.} \quad f(\phi(q_i, d_{ij})) - f(\phi(q_i, d_{ik})) \geq 1 - \frac{\xi_{ijk}}{1-\sigma_{ijk}}\\
&\qquad \xi_{ijk} \geq 0,\\
&\text{for} \quad \forall i \in \{1,2,\ldots,M\},\\
&\qquad \forall j \forall k \in \{1,2,\ldots,n(q_i)\} \quad \text{with } y_{ij} > y_{ik}.
\end{aligned}$$

Different from margin rescaling, slack rescaling varies the amplitude of slack variables adaptively. If a pair of documents are dissimilar in the domain-specific feature space, by dividing 1 $\sigma_{ijk,}$ the slack variables that control the ranking loss of the two documents are correspondingly amplified in order to satisfy the first linear equality, and vice versa. The dual problem of is

$$\begin{aligned}
&\max_{\alpha_{ijk}} -\frac{1}{2}\sum_{i,j,k}\sum_{l,m,n}\alpha_{ijk}\alpha_{lmn}\mathbf{x}_{ijk}^T\mathbf{x}_{lmn}\\
&\qquad + \sum_{i,j,k}\left(1 - \delta f^a(\mathbf{x}_{ijk})\right)\alpha_{ijk}\\
&\text{s.t.} \quad 0 \leq \alpha_{ijk} \leq (1-\sigma_{ijk})C,\\
&\text{for} \quad \forall i \in \{1,2,\ldots,M\},\\
&\qquad \forall j \forall k \in \{1,2,\ldots,n(q_i)\} \quad \text{with } y_{ij} > y_{ik},
\end{aligned}$$

and the solution of the ranking function, as for RA-SVM-MR, is same to, as shown in . It can be observed from the dual format of that, slack rescaling is equivalent to rescaling the tradeoff parameters C for each pairwise documents, based on their similarities. The optimizations of RA-SVM-MR and RA-SVM- SR have the exactly same time complexity as for the RA-SVM , i.e., O(n2:3) by using SMO algorithm and OðnÞ by means of cutting plane algorithm for the linear kernel. Therefore, although domain-specific features are incorporated for the model adaptation, we didn't bring about any additional efficiency problems.

5 Conclusions

As various vertical search engines emerge and the amount of verticals increases dramatically, a global ranking model, which is trained over a data set 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 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. We proposed suitable and efficient method for rank model adaptation for domain specific search. The model adapted for the construction uses the number of hits on link made by the user and automatically that search comes up. This model can be build for the Medical, Engineering, Agriculture, Pharmacy etc. Ranking model is adapted for small number of instances and parameters are needed to be adjusted resulting in good performance and search result. The searches performed by the user are usually dependent on the previous search they performed or specific search intension. One should make the search engine domain adapted to do easier work for the user.

References

[1] M. Belkin, P. Niyogi, and V. Sindhwani, "Manifold Regular- ization: A Geometric Framework for Learning from Labeled and Unlabeled Examples," J. Machine Learning Research, vol. 7,pp. 2399-2434, Nov. 2007.

[2] J. Blitzer, R. Mcdonald, and F. Pereira, "Domain Adaptation with Structural Correspondence Learning," Proc. Conf. Empirical Meth- ods in Natural Language Processing (EMNLP '06), pp. 120-128, July 2006.

[3] 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.

[4] 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.

[5] 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.

[6] 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. vol. 30, nos. 1/2, pp. 81-93, June 2007.

[7] 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.

[8] 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.

[9] 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, 2004.

[10] L. Page, S. Brin, R. Motwani, and T. Winograd, "The Pagerank Citation Ranking: Bringing Order to the Web," technical report, Stanford Univ., 2005

[11] J.M. Kleinberg, S.R. Kumar, P. Raghavan, S. Rajagopalan,

ISSN 2229-5518
and A.Tomkins, "The Web as a Graph: Measurements, Models and Methods," Proc. Int'l Conf. Combinatorics and Computing, pp. 1-18, 2008.