Design and Implementation of Ranking Model Adaptation for Domain-Specific Search

Ms. Wahida Banu S  
Assistant Prof, Dept Of CSE  
SVCE,  
Banglore, India  
syedawahida@gmail.com

Mr. Yerriswamy T  
Assistant Prof, Dept Of ISE  
SVCE,  
Banglore, India  
ys_2123@yahoo.com

Mr. Balaji K  
Assistant Prof, Dept Of CSE  
SVCE,  
Banglore, India  
kannappan.balaji@gmail.com

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 labelling data and time-consuming for training models. This paper addresses 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 labelled data and the training cost is reduced while the performance is still guaranteed. Ranking Adaptability Support Vector Machine algorithm only requires the prediction from the existing ranking models, rather than their internal representations or the data from auxiliary domains. Finally, ranking adaptability measurement is proposed to quantitatively estimate if an existing ranking model can be adapted to a new domain.

Keywords — Domain Specific engine; Support Vector Machine; Information retrieval

I. INTRODUCTION

Learning is a process of ranking a model based on information retrieval technique. This technique is specialized by some documents labelled with their relevancies to queries, where the model is capable of ranking the documents returned to an arbitrary new query automatically. Based on various machine learning methods, e.g., Ranking SVM, Rank Boost [6], Rank Net, List Net, Lambda Rank, etc., are model ranking 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 broad-based 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 the broad-based and vertical search engines are mostly based on text search techniques, the ranking model learned for broad-based [3] 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 webpage ranking, and text based ranking model can be applied here directly. However, the broad-based ranking model is built on the data from multiple domains and cannot be generalized well for a particular domain with special search intentions. In addition, the broad-based ranking model [1] can only utilize the vertical domain’s ranking features that are same to the broad-based domain's for ranking, where the domain-specific features, such as the content features of images, videos or music cannot be utilized directly. These 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.

Alternatively, each vertical can learn its own ranking model independently. However, it’s laborious to label sufficient training samples and time-consuming to train different models for various verticals, since the number of verticals are large and increasing drastically. Ranking model of the broad-based search can provide a reasonable, though not as perfect as the specifically trained, ranking model for vertical search applications. Thereafter, we can make a trade-off between the direct using of the broad-based model and the independent learning of a completely new ranking model, for each specific vertical [3]. Broad-based ranking model can be adapted, with the help of several labelled samples and their domain-specific features, for ranking the documents in new domains. Because the existing 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 [2], 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 domain-specific ranking models [1].

A. Ranking Model

Ranking model is a technique which is used to rank the data present in specific domain. Ranking model is adapted instead of generating new ranking model because that ranking model which is adapted from broad-based search by using RA-SVM algorithm [2] has the previous knowledge about domain and need not to incorporate the domain information into it. Domain-specific search means the search related to specific topic like videos, audio, image etc... Ranking is given based on number of searches made by the users to a data in specific domain.

B. Domain Specific Search Engines

Using a general search engine to search for a particular keyword, chances are you will be met with search results that are promotional in nature - sites that sell products related to the keyword. If don’t want to waste time wading through irrelevant search results and get straight to the heart of the matter, turn to domain specific search engines. Also known as vertical search engines, they focus on a specialized segment of the huge mass of online content. These search engines use focused crawlers to index only websites that are focused on topics that are predefined. They focus on content that is classified according to topicality, media type or different genres of content.

These search engines would limit themselves to searches relating to medical, travel, automobile, art or video content. A vertical search engine searches a specific industry, topic, type of content (e.g., travel, movies, images, blogs and live events), piece of data, geographical location, and so on. It may help to think of vertical search as a search for a particular niche. Some of this content cannot be found, or is difficult to find, on general search engines. For this reason, the topic of vertical search is closely related to that of the deep Web. Since the less prominent vertical search engine is relevant to consumers, the lower is the price- per-click that the general search engine can charge.

C. Problem Statement

The aim of the paper is to design and implement the ranking model for domain specific search. Ranking model will rank the data which is in the specific domain by using RA-SVM algorithm. Domain specific search performs the vertical search and provide the required data to the user without wasting time, money and system resources. Ranking model provides the rankings of the data to the user which helps to know which is the best data and number of views on that data. Rankings to data also helpful for the web page owners or product owners to view their rankings.

D. Scope of the Paper

- The Application is implemented to fulfill all the client requirements. The interfaces designed for the system is very user friendly and attractive. It has successfully implemented the operations of an organization like creating the various domains and uploading the images.

- The user can register and comment on the domains inserted by the admin by the search option. The comments of the users are viewed by the admin and they are differentiated based on the good and bad comments. Administrator can view the description of the comments received and they can perform the ways to increase their product efficiency.

This paper overcomes the disadvantages of broad based ranking model like:

- Keep on searching for required information to which user don’t know the correct URL leads to pay more money to information provider.
- There are no registration formalities for existing system hacker can misguide user to get wrong information.
- User, keep on trying for required data by entering the unknown URL or labeling the query which is unknown to him.

E. Objectives

- Easy access: This computing paradigm brings great convenience for information access. This application can be accessed through mobile, laptops, desktops.
- Low cost: Maintenance cost of the application will be less as it uses minimal system resources ex. hardware and software.
- Privacy: In proposed system, we overcome the problem of location privacy by implementing authentication. New users should register for first time in order to access the application.

II. RANKING ADAPTATION WITH DOMAIN SPECIFIC FEATURE

Data from different domains are characterized by certain domain specific features, e.g. When we adopt the ranking model that is used in a webpage search domain to an image search domain, the information surrounding the image can also provide additional information to support text based ranking model.

Here the domain specific features are utilized. These domain specific features are difficult to translate into textual forms. This boosts the performance of RA-SVM. The rule is that documents with similar domain specific features have to be ranked with similar rankings in the domain. This assumption is called as consistency assumption. The RA-SVM system has been identified to have the following modules:

- Ranking Model Adaptation (RMA) Module.
- Upload and Explore Module.
- Ranking Model with Domain Specific Search Module.
III. SYSTEM DESIGN

Figure 1 shows that the computing device may be a laptop, personal computer, mobile etc., the computing device enables the user to operate a browser to interact with a search engine. When the user is going to enter the query term or query terms, the query term passed through the network to the search engine server. The server may have the number of links to the web pages which are located in different machines. After getting the required web page into the server, the web page is ranked that by checking how many times the web page has been accessed by using the Ranking-Model Module and display the required web page and its rank to the user.

Figure 2 shows the Architecture for domain specific search. When the user query entered into the Domain-Specific search Engine, the indexer will fetch all the related documents and stored into the Focused index. From the Focused Index top-k results are fetched. When top-k results are given to the Domain-Specific Ranking Model, the web page which will be accurately matched with the user query is displayed to the user.

Figure 3 shows the Flow Chart Diagram for Ranking Model Adaptation for Domain Specific Search on the administrator side. Here the operations provided to administrator by application are

- **Domain upload**: Administrator can able to upload or create a new domain by entering details like domain name, sub domain, title etc. For example, if administrator wants to create new domain called product, he/she can enter domain name as product, sub domain as electronics, title as laptop, and type as dell and some description about that product and then submit these details to create new domain.

- **View user rankings**: Administrator can able to view the rankings of user which is given to the content of the particular domain by entering domain name, sub domain, title and type details to which domain he/she like to know the rankings.

- **View user comment**: Administrator can view the comment of the user by entering the type of particular domain.

- **View user registration**: Administrator can view the details of users who are registered to use the application.

- **View graph representation**: Administrator can view the graphical representation of the ranking of data in the particular domain by entering domain name and title.

Figure 3 shows the Flow Chart Diagram for Ranking Model Adaptation for Domain Specific Search on the user side. The following operations are going to provide by the application to the use

- **Domain specific search**: The user can search the required data by entering the specific domain into the domain specific search engine.

- **Enter domain name**: By entering the domain name of required data user can easily get the required dat
- **View domain details**: User can view user details by entering the domain name in a domain specific search engine. For example if the domain name is medical, there will be different types of sub domains like cancer and fever and there may be different types of cancer and fever in those two sub domains like thyroid cancer, lung cancer, kidney cancer and fevers like viral, malaria, typhoid etc.,

- **Add user Ratings**: User can upload ratings to the data after viewing the data in particular domain.

- **Add user comment**: The user can also add comment to the data of specific domain.

Figure 4 shows the Use case diagram on user side.

**Fig. 4 Flow chart diagram on user side.**

Figure 5 shows the Use case diagram for Ranking Model Adaptation for Domain Specific Search. Here the operations provided to administrator by application are:

- Domain upload.
- View user rankings.
- View user comment.
- View user registration.
- View graph representation.

Operations provided to user by application are:

- Domain specific search.
- Enter domain name.
- View domain details.
- Add user Ratings.
- Add user comment.

**IV. ALGORITHM FOR RA-SVM**

The RA-SVM algorithm will take input as the domain name and performs domain specific search. It has 4 functionalities as follows:

1) Ranking model is adapted to domain specific search from broad-based by using RA- SVM algorithm.

2) RA-SVM will categorize the data into specific domain by using binary technique which will divide data based on their property.

3) The adapted ranking model will not perform efficiently because of domain differences RA-SVM is used to make suitable so that ranking model will perform efficiently.

4) RA-SVM will provide the most relevant data that is suitable to user query.

The algorithm for RA-SVM is as shown below.

**Algorithm: RA-SVM**

**Input:** Domain Name

**Output:** Performs Domain-Specific Search

**Method:** It can be categorized into 4 functionalities

**Step-1:** Adapt ranking model from broad-based to domain-specific

**Step-2:** Categorize the data into specific domain based on binary technique

**Step-3:** Make adapted ranking model suitable for target domain

**Step-4:** Provide the most relevant information to the user query

**V. RESULTS**

The results of ranking model adaptation for domain specific search are as follows.
VI. 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 labelling 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. The system implemented, fulfil all the client requirements like easy to access, low cost and privacy. The interfaces designed for the system is very user friendly and attractive. It has successfully implemented the operations of an organization like creating the various domains and uploading the images.

REFERENCES


