StackOverFlow - Finding the right expert

Divya Spoorthy
Indian Institute of Technology
Hyderabad, Telengana
es15btech11001@iith.ac.in

ABSTRACT

Internet, especially social media is one of the most important sources of information and knowledge in our everyday life activities. Schooling, working in a professional environment and many other day to day activities involve querying and trying to get the best relevant and most useful information. This has become a part of our routine in this modern society. Any query should be efficient, accurate and fast, and most importantly it should be reliable for an easy an dependable usage from the user point of view. Stackoverflow is in itself one of the biggest search engines and also a site for users to post questions. But there is an over all need for better query results and people usually do not know which expert to consult among the myriad of answers. Hence we develop a ranking system for stackoverflow sites based on query results of search engines in the stackoverflow websites.

KEYWORDS

QA sites, Ranking, Question modeling

1 INTRODUCTION

Everyone relies on social media in order to get their questions answered. There are many active and popular QA sites which help people resolve their queries. Quora, StackOverflow, Facebook, Twitter are popular examples of such sites where people post questions and hope to receive answers to them from other people's opinions and responses. While this is a brilliant platform for all of us to get our queries resolves more than often we find ourselves not having the exact answer whenever we try and query the questions. We might get slightly different results from what one expects and sometimes even get most irrelevant answers.

This is mainly because of high variability among the responses in the sites and QA forums. Because of this reason, it is very important to select the right platform to ask the questions first and then rank the responses based on the site. Various sites have variable characteristics that people want to have involved in their responses. So finding the right expert can be achieved through topic specialization. Previous work in this area of finding the right sight / expert has been primarily been directed towards behaviour of users[1][2]. Other source involved topic specialization[3] tackled the problem of finding the right social media site for asking a particular question from user. Here in this project we primarily focus on finding the right expert for stackoverflow websites.

2 PROBLEM STATEMENT

2.1 Input

In the general case scenario, we are given a specific Question, Q, and a set E, where each $e^i \in E$ is an Expert Profile.

2.2 Output

Output is a ranked version E_0 of the Expert Profile set, where each $e^i \in E_0$ is ranked according to its likelihood to give a response to the question.

3 QUESTION MODELING

It is extremely difficult to guess the most appropriate answer from just the one line questions that the user provides. We need to gather more information and perspective in order to map the questions to the most suitable and acceptable answer.

3.1 Extract Keywords

Separating the important words from the non - important ones helps in circumventing the topics that the user had in mind when he asks the question. Hence all the stop words are removed and nouns are extracted from each question. Then the words that are selected are ranked according to their importance.

3.2 Expand Keywords

Since just the key words do not give us any information regarding what the user might be asking we need to expand each word in the order of significance and try to collect more data as to what the user might have in his or her mind. So, we expand each key work using wikipedia websites and wikibooks. We scarpe all the data relevant to a particular word and store the urls from which data is scarped.

3.3 Vector Indexing

All the returned data from Wikibooks is indexed using tf.idf of the bi-grams obtained from the questions. These keywords are indexed as a frequency vector

$$W(q) = p(w_1|W_ik_i).....p(w_m|W_ik_i).$$

4 MODELING AN EXPERT PROFILE

We obtained the content data of the experts form the Stackoverflow Data - UCI.The stop words are removed and the data is processed to obtain the important words. Index each of the Expert data into an expert profile vector. Given an expert data

$$e_iE, T(e_i) = [p(w1|W_ik_i)....,p(w_m|W_ik_i)]$$

Thus for question q, the expert E can be ranked by

$$D(e_i, q) = T(e_i).W(q)/||T(e_i)||X||W(q)||$$

5 REFERENCES

[1] Elena Smirnova. 2011. A model for expert finding in social networks. In Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval (SIGIR '11). ACM, New York, NY, USA, 1191-1192.

[2] Z. Yang, I. Jones, X. Hu and H. Liu, "Finding the right social media site for questions," 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), Paris, 2015, pp. 639-644. doi: 10.1145/2808797.2809391