Domain Adaptive Answer Extraction for Discussion Boards

Ankur Gandhe
IBM Research India
ankugand@in.ibm.com

Dinesh Raghu
IBM Research India
dinraghu@in.ibm.com

Rose Catherine
IBM Research India
rosecatherinek@in.ibm.com

ABSTRACT

Answer extraction from discussion boards is an extensively studied problem. Most of the existing work is focused on supervised methods for extracting answers using similarity features and forum-specific features. Although this works well for the domain or forum data that it has been trained on, it is difficult to use the same models for a domain where the vocabulary is different and some forum specific features may not be available. In this poster, we report initial results of a domain adaptive answer extractor that performs the extraction in two steps: a) an answer recognizer identifies the sentences in a post which are likely to be answers, and b) a domain relevance module determines the domain significance of the identified answer. We use domain independent methodology that can be easily adapted to any given domain with minimum effort.

Categories and Subject Descriptors

H.3.3 [**Information Search and Retrieval**]: Retrieval Models

Keywords

question-answering, sequential patterns

1. INTRODUCTION

With the advent of Web 2.0, forums and discussion boards have become a source of rich unstructured information and it is becoming increasingly important to extract the useful information from the threads in these forums. Since the content is created by community, it is plagued by a lot of junk information and the answer, if present, is almost always hidden among the other posts. In a thread of large number of posts, it becomes difficult for a user to find the answer post quickly.

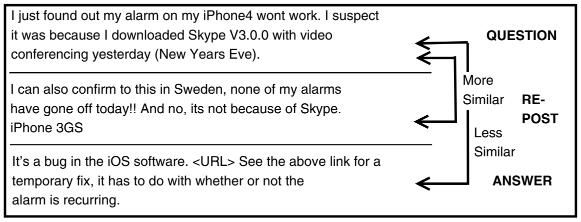
In a given thread, the first post is usually the question and the subsequent posts can be a reply to this question or a reply to a different post which was posted earlier. Reply posts can be further divided into three types: a) *re-posts* that express the same or similar problem, asking for an answer b) *answer posts* which contain the potential answer to the question and c) *junk posts* that contain information not related to the discussion.

Some research has been done on extracting questions and answers from documents, forums and email conversations. In the past, Cong et al. [1] used graph-based propagation method to extract answers from cQA sites. They use similarity to determine question-answer pairs which does not give satisfactory results for discussion boards, especially technical forums, due to the presence of *re-posts*. Other research has focused on question detection and question retrieval in cQA websites ([3]) based on lexical and syntactic features. This method works well for question detection but similar techniques have not been applied for answer detection.

Using similarity match for answer retrieval in forums does not work very well since answer posts often do not contain the same words as the question and also due to the presence of *re-posts*. Figure 1 gives an example of such a thread where the similarity between question and answer post is lower than the similarity with a re-post. On the other hand, technical forums tend to be highly domain specific and presence of domain specific words are a good indicator of it being an answer.

Our proposed method is different from the existing works in the following way:

1. Extraction of syntactical or lexical patterns in answers that work across domains
2. Domain relevance of the sentence is used to guage the significance of the answer in the domain



I just found out my alarm on my iPhone4 wont work. I suspect it was because I downloaded Skype V3.0.0 with video conferencing yesterday (New Years Eve).

I can also confirm to this in Sweden, none of my alarms have gone off today!! And no, its not because of Skype. iPhone 3GS

It's a bug in the iOS software. <URL> See the above link for a temporary fix, it has to do with whether or not the alarm is recurring.

Figure 1: Example of a thread post with low similarity of Question-Answer pair

2. PROPOSED METHOD

Our proposed method consists of two steps and we combine three features to model the likelihood of a post being an answer: a) domain relevance, b) Part-of-Speech(POS) tag patterns, and c) hybrid POS-lexical patterns.

Domain-Relevance: If a sentence does not have any domain words, it is more likely to be an irrelevant or junk post than be an answer post in a given thread. From the corpus, we extract domain-specific words by comparing the probability of occurrence of a noun or a verb in the corpus to the probability of it occurring in generic english text[1] using

[1] In our experiments, we crawled data from http://www.bbc.co.uk/, http://www.voanews.com/

WWW 2012 Companion, April 16–20, 2012, Lyon, France.
ACM 978-1-4503-1230-1/12/04.

the following inequality:

$$\frac{P_{domain(w)}}{P_{generic(w)}} \geq \tau \qquad (1)$$

We use this as a measure of how relevant a sentence in that domain. Given the sentences in a thread, we determine their domain relevance (ϕ_D) detecting the domain-specific words in the sentence, normalized by the number of words in the sentence. The more the number of domain words in the sentence, the greater is the relevance of the sentence in the concerned thread or forum.

Answer Recognition: Given a post from discussion forums, it can be classified into three categories as defined in Section 1. The word order and other lexical properties such as tense of the verb help us distinguish the reposting of the question from an answer. For instance, the word sequence "Try using the application ... " is a good indicator that the sentence might be an answer. To identify the frequently occurring subsequences of words that are indicative of an answer, we use answers of *frequently asked questions* (FAQs) to extract sequential patterns that occur in them. We learn the sequential patterns on two transactions:

1. Part-of-speech tag patterns (ϕ_P): To capture the grammatical patterns of answers, we replace all words by their POS tags. A POS tag pattern for the above example would be "VB VBG DT NN ... ".
2. Hybrid POS-lexical patterns (ϕ_G): Since nouns and verbs differ greatly when we switch domains, they are replaced by their POS tags before extracting the word patterns. A hybrid version of previous example would be "VB VBG the NN ... ".

We use PrefixSpan [2] to extract the sequential patterns and empirically set the minimum support (a user specified threshold of frequency of subsequences) to 4. Given the sequential patterns, we calculate the confidence of a sentence i being an answer as:

$$\phi = \frac{2|S_i|}{L_i(L_i+1)} \qquad (2)$$

where S_i is the set of sequential patters matching with sentence i and L_i is the number of words in the sentence. The term $L_i(L_i + 1)$ is used to normalize the scores, since the cardinality of pattern matches increases exponentially with number of words.

Ranking: Given the three sentence-wise feature scores from the previous steps (two for sequential patterns and one for domain relevance), we combine the individual features linearly:

$$Score = \lambda_P\phi_P + \lambda_G\phi_G + \lambda_D\phi_D \qquad (3)$$

The total score of a given post is calculated by averaging the per-sentence scores. The posts score in a thread are normalized to obtain the likelihood of post being an answer.

3. EXPERIMENTAL EVALUATION

Dataset: To build an answer recognizer, we extracted 91 answers from WolframAlpha FAQs[2] with a total of 201 sentences and extracted the sequential patterns as explained in Section 2. These were pre-processed and POS tagged using the Stanford Parser[3]. For evaluating our proposed method on a different domain, we crawled about 140K iPhone Forum[4] threads and used them to obtain the domain relevance probabilities. We randomly picked and annotated 549 threads, out of which 60 threads were used for training and 489 threads for evaluation. The training set was used to perform a grid search over Precision@1 to learn the weights of the features. The average number of posts per thread was 4.49 and the average number of answers per thread was 1.6.

	P@1	**MAP**	**MRR**
Cosine-Sim	0.52	0.70	0.70
Graph-Prop	0.67	0.81	0.79
SeqPat	0.70	0.83	0.81
SeqPat+Domain	**0.72**	**0.84**	**0.82**

Table 1: Evaluation metrics for different settings

Metrics: We evaluate the performance of our method using three metrics: Precision at 1 (P@1), Mean Reciprocal Rank (MRR) and Mean Average Precision (MAP). We compared our approach against 2 baselines:

1. A simple cosine similarity match between the question and the posts, treating each post as bag of words (Cosine-Sim)
2. A graph based propagation method as explained in [1] (Graph-Prop)

Results: Table 1 shows the P@1, MAP and MRR for different approaches. It can be seen that our approach performs the best on the test set for all the metrics. The Domain relevance scores improves the precision by about 2 %. Also note that though the Answer Recognizer was trained on a different domain, a significant improvement is achieved, suggesting the domain-adaptive nature of our approach.

4. CONCLUSION

In this paper, we proposed a domain-adaptive approach for answer extraction from forums that works across domains, and performs significantly better than previous work in this area. To overcome the short-comings of similarity match, we proposed a method to build an answer recognizer and then use easily obtainable domain knowledge to rank the answer posts. Future work involves comparing our method to supervised approaches for extracting answers, and utilizing additional attributes associated with a post in a thread. Adding similarity as a feature for deciding the ranking of answers also might lead to further improvement.

5. REFERENCES

[1] G. Cong, L. Wang, C.-Y. Lin, Y.-I. Song, and Y. Sun. Finding question-answer pairs from online forums. *Proceedings of 31st annual international ACM SIGIR*, pages 469–474, July 2008.

[2] J. Pei, J. Han, B. Mortazavi-As, and H. Pinto. Prefixspan: Mining sequential patterns efficiently by prefix-projected pattern growth. *ICDE'01*, pages 215–224, April 2001.

[3] K. Wang and T.-S. Chua. Exploiting salient patterns for question detection and question retrieval in community-based question answering. *Proceedings of 23rd COLING 2010*, pages 1155–1163, August 2010.

[2] http://www.wolframalpha.com/faqs.html

[3] http://nlp.stanford.edu/software/lex-parser.shtml

[4] https://discussions.apple.com/community/iphone