i Title:

Gao Cong, Long Wang Chin-Yew Lin, Young-In Song, Yueheng Sun,Finding Question-Answer Pairs from Online Forums Published in 2008

ii Keywords

Question Answering Service: Forums or services like Yahoo Answers, where people can ask questions and answer them.
Graph Based Ranking: It refers to Ranking generated from a graph based propagation method used in the paper
Labeled Sequential Patterns: Sequentially matching patterns that are used to identify comparative sentences and erroneous sentences.
Information Extraction: Extraction of valid information from forums

iii Brief Summaries

iii1 Motivation:

Mining question-answer pairs from forums has applications in the are of QA services including instant answers provided by search engines, QA search systems and so on. In spite of it being highly valuable to obtain these question answer pairs from forums which are widely unstructured very less work in the past deals with this problem. Thus the application of mining QA for chatbots , search engines is the major motivation for the study of this area.

iii2 Data

Data from three forums of different scales was taken
1. 1,212,153 threads from TripAdvisor forum
2. 86,772 threads from LonelyPlanet forum 7 and
3) 25,298 threads from BootsnAll Network 8 as well as
300,000 question-answer pairs from Yahoo! Answers as training set.

iii3 Related Work

There has been some related work where the input-reply pairs are extracted for chatbox applications to automatically answer users' queries.
Feng et al. [5] implemented a discussion-bot to automatically answer students' queries by matching the reply posts from an annotated corpus of archived threaded

discussions with students' query using cosine similarity. This was a little bit different from the work that this paper talks about

iii4 Statistical tests:

To generate graph based rankings for finding question-answer pairs the authors use graph based propagation method. They first present how to build graphs for candidate answers, and then how to compute ranking scores of candidate answers using the graph.

To compute propagation score they use two approaches, Computation propagation without initial score where the product of the authority value and the initial ranking score between candidate answer and question will be returned as the final ranking score for and computing with initial score, where the propagation is computed recursively using the following formula where the parameter ∏ is a trade-off between the score of *a* and the scores of *a's* offsprings in the equation, and is determined empirically.

$$Pr(\mathbf{q}|\mathbf{a}) = \lambda \frac{Pr(\mathbf{q}|\mathbf{a})}{\sum_{\mathbf{t}\in C_\mathbf{q}} Pr(\mathbf{q}|\mathbf{t})} + (1-\lambda) \sum_{\mathbf{v}\in C_\mathbf{q}} nw(\mathbf{v}\rightarrow\mathbf{a}) \times Pr(\mathbf{q}|\mathbf{v}) \quad (10)$$

.

iv Scope Of Improvement

iv1: To detect questions without answers

iv2: To revisit more TREC QA techniques to see if they can help answer detection in forums

iv3: To investigate the effectiveness of the techniques in passage retrieval on TREC QA data