Modeling Message Roles and Influence in Q&A Forums

Jeonhyung Kang and Jihie Kim
University of Southern California
Information Sciences Institute
4676 Admiralty Was, Marina del Rey, CA, U.S.A
{jeonhyuk, jihie}@isi.edu

Abstract

We are modeling roles of individual messages and participants in Q&A discussion forums. In this paper, we present a mixed network model that represents message exchanges and message influences within a discussion thread. We first model individual message roles and thread-level user roles using discussion content features. We then combine the resulting message roles and the user roles to generate the overall influence network. Message influences and their aggregation over the network are analyzed using B-centrality measures. We use the results in identifying the most influential message in answering the initial question of the thread.

1 Introduction

Online discussion boards play an important role in various fields, including science, politics, and education. Understanding patterns of group interactions can be important in many applications. For example, in Q&A forums, some messages contain more useful or influential answers than others. Identifying useful content can help future discussions with similar issues. That is, useful information on a topic can be sent to related discussions (Kim et al., 2009).

There has been some work on analyzing dialogue patterns in online discussion boards (Feng et al, 2006). Some of these model message roles using dialogue acts such as question act or answer act. Most of these focus on modeling individual messages, often using surface forms. There has been limited study on thread-level modeling of the true roles of the messages, whether they provide information (*source*) or seek information (*sink*), or which message in the thread is most useful or influential as the source.

In this paper, we present a novel model of message influence within a discussion thread. In modeling the thread-level influence of the messages, besides the sink/source role of the individual messages, we take into account the roles of the message posters within the thread. In Q&A discussion threads, since the roles of the posters as an information provider or an information seeker often do not change within the same thread, such information can help us identify the true roles or influence of the messages.

We combine the message roles and the user roles with a network model. Message influences and their aggregation over the network are analyzed using B-centrality measures. We use the resulting influence scores in identifying the most influential message in answering the initial question of the thread.

2 Modeling Message Influence

We use discussion data from an Operating Systems course in the Computer Science department at the University of Southern California. Students use a discussion board, most commonly, to seek help on the project assignments. For this study, we use data from the Fall 2007 semester, with 177 discussion threads (randomly choose 133 for training and 44 for test) with 580 messages (randomly choose 451 for training and 129 for test).

2.1 Sink and source roles of a message

For each pair of messages where one is a reply to the other, we model the roles of the latter, as a *sink* or a *source* with respect to the former message or the message author. Some messages, especially long ones, can have both roles. Figure 1 shows an influence model of sink and source. A node represents either a user or a message. An edge is either a reply-to relation between two messages or an ownership of a message by a user. The direction of each edge indicates the direction of influence. A source is a message that provides information and it generates influence. In the top graph, B responds

Proceedings of the NAACL HLT 2010 Workshop on Computational Linguistics in a World of Social Media, pages 11–12,
Los Angeles, California, June 2010.

to A's message as an information source. A sink message requests information from others so the edge direction goes towards it. Note that sinks and sources are different from questions and answers since some of sources can take a form of a question (e.g. have you checked the manual?).

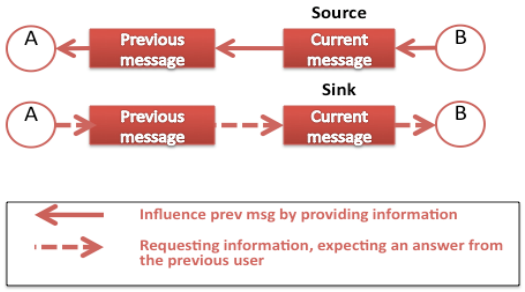

Figure 1: Sink and source role of a message

2.2 Information seeker and information provider: Thread level roles

Figure 2 shows an influence model of an information seeker and an information provider when they exchange messages. The information flows from the information provider to the information seeker, as indicated by the direction of the edges. In general, the initial poster seeks information and his or her role does not usually change within the thread. Without loss of generality, we assume that within a discussion thread, a user's role doesn't change for a certain amount of time, although he or she can post both Sink and Source messages, as shown in Figure 2. Using this model, we can capture the intention of the message based on who posted the message.

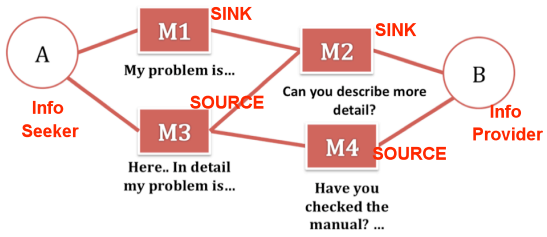

Figure 2: Thread-level user intention and message roles

2.3 Message and Participant Role Classifiers

Individual messages were annotated with sink and source information. The same message can have both sink and source roles with respect to the prior message or its poster.

The features used include cue phases (n-grams) and their positions, message position in the thread, author change information, the relative position of the message in the thread, a user's participation frequency (normalized), n-gram of previous messages with their positions and the message length. The details of these features are described in Kim et al (2009).

We used a Support Vector Machine (Chang and Lin 2001) to create two binary classifiers (since one message can have both roles) that identify message roles: source and sink, and one binary classifier for user roles: *information seeker and information provider*. The precisions/recalls range from 0.89 to 0.96. F-scores are within 0.92-0.93.

3 Profiling Influence of Message with Centrality Measures

We generate a message-role graph and a user-role graph using the above model, and generate influence network combining user and message roles. Message influences and their aggregation over the network are analyzed using B-centrality measures. To evaluate our source score accuracy (Ghosh and Lerman 2009), we annotated the most influential source message for the initial question (sink) in each thread. We use Mean Reciprocal Rank Score (MRR) to evaluate our results. The combined model provides better results with an MRR score of 0.90.

Ranking strategy	Information used	MRR
Influence Network Model score	**User role + msg role**	**0.90**
Earlier source msg	**msg role** + msg location	0.74
Earlier msg from info providers	**User role** + msg location	0.68

Table 2: MRR scores for different strategies

Acknowledgement

This work was supported by National Science Foundation, CCLI Phase II (#0618859).

References

Feng, D. Shaw, E. Kim, J. and Hovy, E. An Intelligent Discussion-Bot for Answering Student Queries in Threaded Discussions, Proc. of *IUI-2006*.

Ghosh, R. and Lerman, K. The structure of heterogeneous networks. Proc. of *IEEE Social Comp. Conf, 2009*.

Kim, J., Li, J., and Kim, T. Identifying student online discussions with unanswered questions,Proc. K-CAP 2009.