Summarizing Online Forum Discussions – Can Dialog Acts of Individual Messages Help?

Sumit Bhatia[1], Prakhar Biyani[2] and Prasenjit Mitra[2]
[1]IBM Almaden Research Centre, 650 Harry Road, San Jose, CA 95123, USA
[2]Information Science and Technology, Pennsylvania State University, University Park, PA 16802
`sumit.bhatia@us.ibm.com`, {`pxb5080`, `pmitra`}`@ist.psu.edu`

Abstract

A typical discussion thread in an online forum spans multiple pages involving participation from multiple users and thus, may contain multiple view-points and solutions. A user interested in the topic of discussion or having a problem similar to being discussed in the thread may not want to read all the previous posts but only a few selected posts that provide her a concise summary of the ongoing discussion. This paper describes an extractive summarization technique that uses textual features and dialog act information of individual messages to select a subset of posts. Proposed approach is evaluated using two real life forum datasets.

1 Introduction

In recent times, online discussion boards (or *forums*) have become quite popular as they provide an easily accessible platform to users in different parts of the world to come together, share information and discuss issues of common interest. The archives of web forums contain millions of discussion threads and act as a valuable repository of human generated information that can be utilized for various applications. Oftentimes, the discussions in a thread span multiple pages involving participation from multiple users and thus, may contain multiple view-points and solutions. In such a case, the end-user may prefer a concise summary of the ongoing discussion to save time. Further, such a summary helps the user to understand the background of the whole discussion as well as provides an overview of different view-points in a time efficient manner. In addition to generic forums on the web, automatic forum summarization methods can prove to be useful for various domain specific applications, such as helping students and support-ing tutors in virtual learning environments (Carbonaro, 2010).

A typical discussion thread in a web forum consists of a number of individual *posts* or messages posted by different participating users. Often, the thread initiator posts a question to which other users reply, leading to an active discussion. As an example, consider the discussion thread shown in Figure 1 where the thread starter describes his problem about the missing headphone switch in his Linux installation. In the third post in the thread, some other user asks about some clarifying details and in the next post the topic starter provides the requested details that makes the problem clearer. On receiving additional details about the problem, some other user provides a possible solution to the problem (fifth post). The topic starter tries the suggested solution and reports his experience in the next post (sixth post). Thus, we see that each individual post in a discussion thread serves a different purpose in the discussion and we posit that *identifying the purpose of each such post is essential for creating effective summaries of the discussions.* Intuitively, the most important messages in a discussion are the ones that describe the problem being discussed and the solutions being proposed to solve the problem.

The *role* of an individual message in a discussion is typically specified in terms of *dialog acts*. There have been efforts to automatically assign dialog acts to messages in online forum discussions (Jeong et al., 2009; Joty et al., 2011; Bhatia et al., 2012) and also using dialog acts for linguistic analysis of forum data, such as in subjectivity analysis of forum threads (Biyani et al., 2012; Biyani et al., 2014). In this paper, we describe our initial efforts towards addressing the problem of *automatically creating summaries of such online discussion threads*. We frame forum summarization as a classification problem and identify messages that should be included in a summary of the

discussion. In addition to textual features, we employ dialog act labels of individual messages for summarization and show that incorporating dialog acts leads to substantial improvements in summarization performance.
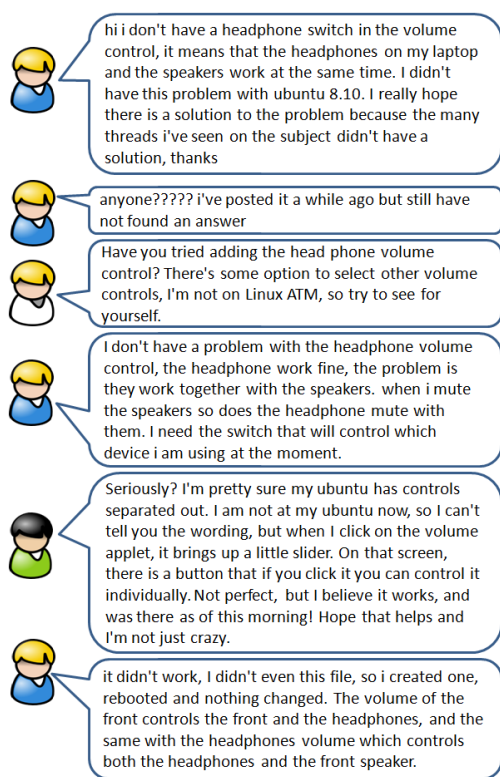

Figure 1: An example thread illustrating different role played by each post in the discussion. Different users are indicated by different colors.

2 Definition of Dialog Acts Used

We use the same set of dialog acts as defined by Bhatia et al. (2012). Note that based on the application context and requirements new dialog acts can be defined and added.

1. Question: The poster asks a question which initiates discussion in the thread. This is usually the first post in the thread but not always. Often, the topic initiator or some other user may ask other related questions in the thread.

2. Repeat Question: Some user repeats a previously asked question (e.g. *Me too having the same problem.*).

3. Clarification: The poster asks clarifying questions in order to gather more details about the problem or question being asked. For example, *Could you provide more details about the issue you are facing*.

4. Further Details: The poster provides more details about the problem as asked by other fellow posters.

5. Solution: The poster suggests a solution to the problem being discussed in the thread.

6. Positive Feedback: Somebody tries the suggested solution and provides a positive feedback if the solution worked.

7. Negative Feedback: Somebody tries the suggested solution and provides a negative feedback if the solution did not work.

8. Junk: There is no useful information in the post. For example, someone justs posts a smiley or some comments that is not useful to topic being discussed. For example, *"bump", "sigh"*, etc., or messages posted by forum moderators such as *this thread is being closed for discussion*.

3 Proposed Approach for Thread Summarization

In general, text summarization techniques can be classified into two categories, namely extractive Summarization, and Abstractive Summarization (Hahn and Mani, 2000). Extractive summarization involves extracting salient units of text (e.g., sentences) from the document and then concatenating them to form a shorter version of the document. Abstractive summarization, on the other hand, involves generating new sentences by utilizing the information extracted from the document corpus (Carenini and Cheung, 2008), and often involves advanced natural language processing tools such as parsers, lexicons and grammars, and domain-specific knowledge bases (Hahn and Mani, 2000). Owing to their simplicity and good performance, extractive summarization techniques are often the preferred tools of choice for various summarization tasks (Liu and Liu, 2009) and we also adopt an extractive approach for discussion summarization in this work.

3.1 Summarization Unit – Individual Sentence vs Individual Message

Before we can perform extractive summarization on discussion threads, we need to define an appropriate text unit that will be used to construct the desired summaries. For typical summarization tasks, a sentence is usually treated as a unit of text and summaries are constructed by extracting most relevant sentences from a document. However, a typical discussion thread is different from

a generic document in that the text of a discussion thread is created by multiple authors (users participating in the thread). Further, the text of a discussion can be divided into individual user messages, each message serving a specific role in the whole discussion. In that sense, summarizing a discussion thread is similar to the task of multi-document summarization where content of multiple documents that are topically related is summarized simultaneously to construct an inclusive, coherent summary. However, we also note that an individual user message in a discussion is much smaller than a stand-alone document (compare $3 \sim 4$ sentences in a message to a few dozen sentences in a document). Thus, the sentences in a message are much more coherent and contextually related to each other than in a stand-alone document. Hence, selecting just a few sentences from a message may lead to loss of context and make the resulting summaries hard to comprehend. Therefore, in this work, we choose each individual message as a text unit and thus, the thread summaries are created by extracting most relevant posts from a discussion.

3.2 Framing Thread Summarization as Post Classification

We consider the problem of extracting relevant posts from a discussion thread as a binary classification problem where the task is to classify a given post as either belonging to the summary or not. We perform classification in a supervised fashion by employing following features.

1. Similarity with Title (TitleSim): This feature is computed as the cosine similarity score between the post and the title of the thread.

2. Length of Post (Length): The number of unique words in the post.

3. Post Position (Position): The normalized position of the post in the discussion thread. It is defined as follows:

$$\frac{\text{Position of the post in the thread}}{\text{Total \# of posts in the thread}} \quad (1)$$

4. Centroid Similarity (Centroid): This feature is obtained by computing the cosine similarity score between the post document vector and the vector obtained as the centroid of all the post vectors of the thread. Similarity with centroid measures the *relatedness* of each post with the underlying discussion topic. A post with a higher similarity score with the thread centroid vector indi-

cates that the post better represents the basic ideas of the thread.

5. Inter Post Similarity: This feature is computed by taking the mean of the post's cosine similarity scores with all the other posts in the thread.

6. Dialog Act Label (Class): This is a set of binary features indicating the dialog act class label of the post. We have one binary feature corresponding to each dialog act.

4 Experimental Evaluation

4.1 Data Description

We used the dataset used by Bhatia et al. (2012) that consists of randomly sampled 100 threads from two different online discussion forums – `ubuntuforums.org` and `tripadvisor.com`. There are a total of 556 posts in the 100 threads from Ubuntu dataset and 916 posts in 100 threads from NYC dataset. The associated dialog act labels of individual messages in each of the threads are also available.

Next, for creating data for the summarization task, two independent human evaluators (H1 and H2) were recruited to create summaries of the discussion threads in the two datasets. For each thread, the evaluators were asked to read the whole discussion and write a summary of the discussion in their own words. The annotators were requested to keep the length of summaries roughly between 10% and 25% of the original text length. Thus for each thread, we obtain two human written summaries.

These hand-written summaries were then used to identify most relevant posts in a discussion thread in a manner similar to one used by Rambow et al. (2004). We compute cosine similarity scores for each post in the thread with the corresponding thread summary and the top k ranked posts are then selected to be part of the summary of the thread. The number k is determined by the compression factor used for creating summaries. We choose a compression factor of 20%. The top k ranked posts, thus constitute the gold summary of each thread. Note that we obtain two gold summaries for each thread – one corresponding to each evaluator. This summarization data can be downloaded for research purposes from `http://sumitbhatia.net/source/datasets.html`.

Evaluator	Method	Ubuntu		NYC	
		Precision	F-1	Precision	F-1
H1	Baseline	0.39	0.53	0.32	0.46
	Without Dialog Acts	0.578	0.536	0.739	0.607
	With Dialog Acts	0.620	0.608	0.760	0.655
	Gain	+7.27%	+13.43%	+2.84%	+7.91%
H2	Baseline	0.38	0.52	0.31	0.45
	Without Dialog Acts	0.739	0.607	0.588	0.561
	With Dialog Acts	0.760	0.655	0.652	0.588
	Gain	+14.94%	+20.53%	+10.88%	+4.81%

Table 1: Results of post classification for summarization task. H1 and H2 correspond to the two human evaluators. Percentage improvements obtained by addition of post class label information is also reported.

4.2 Baseline

As a baseline method, we use a rule based classifier that classifies all the *Question* and *Solution* posts in a thread as belonging to the summary and discards the remaining posts.

4.3 Results and Discussions

We used Naive Bayes classifier as implemented in the Weka machine learning toolkit (Hall et al., 2009) for classification experiments. We trained the classifier on 75% of the data and used the remaining 25% for testing. Table 1 reports the classification results using *(i)* the baseline method,*(ii)* features 1–5 only, and *(iii)* using all the features (dialog act labels, in addition to the five features). For both the datasets, we observe that incorporating dialog act information along with textual features results in performance gain across all reported metrics. The strong performance improvements achieved for the two datasets corroborate the proposed hypothesis that knowing the role of each individual message in an online discussion can help create better summaries of discussion threads. Further, we observe that the precision values are very low for the baseline algorithm (from 0.31 to 0.39) with moderate F-1 values (0.45 to 0.53), indicating a higher recall. This means that even though many of the posts in the gold summaries belong to question and solution categories, not all the posts belonging to these two categories are useful for summarization. Using textual features and dialog act labels in a supervised machine learning framework captures the distinguishing characteristics of *in-summary* and *out of summary* posts and thus, yields a much better classification performance.

5 Related Work

Among various applications of text summarization, work on E-Mail thread summarization (Rambow et al., 2004; Cohen et al., 2004) can be considered as closely related to the problem discussed in this paper. An E-Mail thread is similar to a forum discussion thread in that it involves back and forth communication with the participants, however, the problem of discussion thread summarization is very different (and difficult) due to a relatively larger number of participants, highly informal and noisy language, and frequent topic drifts in discussions. Zhou and Hovy (2005) identify clusters in internet relay chats (irc) and then employ lexical and structural features to summarize each cluster. Ren et al. (2011) have proposed a forum summarization algorithm that models the reply structures in a discussion thread.

6 Conclusions and Future Work

We proposed that dialog act labels of individual messages in an online forums can be helpful in summarizing discussion threads. We framed discussion thread summarization as a binary classification problem and tested our hypothesis on two different datasets. We found that for both the datasets, incorporating dialog act information as features improves classification performance as measured in terms of precision and F-1 measure. As future work, we plan to explore various other forum specific features such as user reputation and quality of content to improve summarization performance.

References

Sumit Bhatia, Prakhar Biyani, and Prasenjit Mitra. 2012. Classifying user messages for managing web forum data. In *Proceedings of the 15th International Workshop on the Web and Databases 2012, WebDB 2012, Scottsdale, AZ, USA, May 20, 2012*, pages 13–18.

Prakhar Biyani, Sumit Bhatia, Cornelia Caragea, and Prasenjit Mitra. 2012. Thread specific features are helpful for identifying subjectivity orientation of online forum threads. In *COLING 2012, 24th International Conference on Computational Linguistics, Proceedings of the Conference: Technical Papers, 8-15 December 2012, Mumbai, India*, pages 295–310.

Prakhar Biyani, Sumit Bhatia, Cornelia Caragea, and Prasenjit Mitra. 2014. Using non-lexical features for identifying factual and opinionative threads in online forums. *Knowledge-Based Systems, In Press, doi = http://dx.doi.org/10.1016/j.knosys.2014.04.048.*

Antonella Carbonaro. 2010. Towards an automatic forum summarization to support tutoring. In MiltiadisD. Lytras, Patricia Ordonez De Pablos, David Avison, Janice Sipior, Qun Jin, Walter Leal, Lorna Uden, Michael Thomas, Sara Cervai, and David Horner, editors, *Technology Enhanced Learning. Quality of Teaching and Educational Reform*, volume 73 of *Communications in Computer and Information Science*, pages 141–147. Springer Berlin Heidelberg.

Giuseppe Carenini and Jackie Chi Kit Cheung. 2008. Extractive vs. nlg-based abstractive summarization of evaluative text: The effect of corpus controversiality. In *Proceedings of the Fifth International Natural Language Generation Conference*, pages 33–41. Association for Computational Linguistics.

William W. Cohen, Vitor R. Carvalho, and Tom M. Mitchell. 2004. Learning to Classify Email into "Speech Acts". In *EMNLP*, pages 309–316. ACL.

Udo Hahn and Inderjeet Mani. 2000. The challenges of automatic summarization. *Computer*, 33(11):29–36.

Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, and Ian H. Witten. 2009. The WEKA Data Mining Software: An Update. *SIGKDD Explorations*, 11(1).

Minwoo Jeong, Chin-Yew Lin, and Gary Geunbae Lee. 2009. Semi-supervised speech act recognition in emails and forums. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 3 - Volume 3*, EMNLP '09, pages 1250–1259.

Shafiq R. Joty, Giuseppe Carenini, and Chin-Yew Lin. 2011. Unsupervised modeling of dialog acts in asynchronous conversations. In *IJCAI*, pages 1807–1813. IJCAI/AAAI.

Fei Liu and Yang Liu. 2009. From extractive to abstractive meeting summaries: can it be done by sentence compression? In *Proceedings of the ACL-IJCNLP 2009 Conference Short Papers*, pages 261–264. Association for Computational Linguistics.

O. Rambow, L. Shrestha, J. Chen, and C. Laurdisen. 2004. Summarizing email threads. *Proceedings of HLT-NAACL 2004: Short Papers.*

Zhaochun Ren, Jun Ma, Shuaiqiang Wang, and Yang Liu. 2011. Summarizing web forum threads based on a latent topic propagation process. In *Proceedings of the 20th ACM International Conference on Information and Knowledge Management*, CIKM '11, pages 879–884, New York, NY, USA. ACM.

Liang Zhou and Eduard Hovy. 2005. Digesting virtual "geek" culture: The summarization of technical internet relay chats. In *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics*, ACL '05, pages 298–305, Stroudsburg, PA, USA. Association for Computational Linguistics.