Decalog 2007: Proceedings of the 11th Workshop on the Semantics and Pragmatics of Dialogue, pages 173–174.
Trento, Italy, 30 May – 1 June 2007. Edited by Ron Artstein and Laure Vieu.

An empirical study on detection and prediction of topic shifts in information seeking chats

Ichikawa Hiroshi **Tokunaga Takenobu**

Department of Computer Science
Tokyo Institute of Technology
Tokyo Meguro Ôokayama 2-12-1, 152-8552 Japan
take@cl.cs.titech.ac.jp

Introduction This paper describes an empirical study of detecting and predicting topic shifts in *information seeking chat* (Stede and Schlangen, 2004), which is characterised by its more exploratory and less task-oriented nature, where the user does not have a specific goal but obtains useful information of his interest through interaction with the system.

Unlike (Stede and Schlangen, 2004), we do not assume predefined domain specific knowledge to navigate the dialogue; instead we rely on more superficial clues to deal with a broad range of topics.

Differing from a series of topic segmentation research (Hearst, 1997; Ries, 2001; Galley et al., 2003; Olney and Cai, 2005; Arguello and Rosé, 2006), we deal with topic shifts in real-time interaction, aiming at using this technique in a dialogue system. In addition, we predict relevant topic shifts as well as topic shift detection.

Corpus analysis To find useful features indicating topic shifts, we manually analysed a part of the *Mister O corpus* (Ochiai et al., 2005), which is a cross-linguistic video corpus consisting of various types of conversations. Based on the predefined criteria, the transcribed texts of 6 Japanese conversations were divided into topic segments by one of the authors. Extracted features indicating *topic shift utterances (TSU)*[1], and utterances ahead of them are summarized in Table 1 and 2.

Table 1: Features of topic shift utterances

Feature	Occurrences
*clue expression**	11
*new words**	10
*initiative change**	9
prior topic	5
others	11
Total	45

Table 2: Features of preceding utterances ahead of topic shifts

Feature	Occurrences
*back-channel**	14
*silence**	13
repetition	6
generalisation	4
*impression**	4
others	13
Total	45

Automatic feature detection The features with the asterisk in Table 1 and 2 can be automatically extracted by using superficial clues as follows.

clue expressions: Since we do not have a thorough list of Japanese cue phrases as in (Hirschbeerg and Litman, 1993), we collected a set of cue expressions suggesting topic shifts based on the corpus analysis and introspection. These cue expressions imply this feature.

new words: We assume every content word in each utterance is accumulated in a word pool during interaction. A new content word in an utterance implies this feature.

initiative change: A heuristic algorithm of initiative change detection using predefined cue expressions and information of the speaker suggests this feature.

back-channel: When two consecutive utterances by different speakers include back-channel cue expressions, and they do not include any content words, reciprocal back-channel is implied.

silence: A silence longer than 1 second between utterances implies this feature.

[1]The first utterance of a topic segment.

impression: Cue words suggesting impression and sentiment in an utterance implies this feature.

Detecting and predicting topic shifts To detect topic shifts, we use *clue expression*, *new words*, *initiative change* and their combinations. When at least one of them is detected in an utterance, it is considered as a topic shift utterance.

To predict topic shifts, we use *back-channel*, *silence*, *impression* and their combinations. Since these features tend to be observed just before topic shifts, when at least one of them is detected in an utterance, we predict a topic shift after this utterance. We call this in-between point *topic shift relevance place (TSRP).*

Table 3: Results of TSU detection

Combination of features	Prec.	Recall	F
clue expression	0.43	0.10	0.17
new words	0.09	0.31	0.14
initiative change	0.39	0.10	0.16
clue expression+new words	0.10	0.38	0.16
clue expression+initiative change	0.41	0.21	**0.27**
initiative change+new words	0.10	0.35	0.15
clue exp.+init. chg.+new words	0.11	0.41	0.17

Table 4: Results of TSRP detection

Combination of features	Prec.	Recall	F
back-channel	0.25	0.59	**0.35**
silence	0.23	0.31	0.26
impression	0.24	0.35	0.28
back-channel+silence	0.24	0.76	**0.36**
back-channel+impression	0.25	0.69	**0.36**
silence+impression	0.23	0.48	0.31
back-channel+silence+impression	0.23	**0.79**	**0.36**

Evaluation The held-out data of 20 dialogues from *Mister O corpus* were manually divided into topic segments by three annotators. The average of pair-wise κ was 0.41. The outputs of the system with various combinations of features were compared with the manual annotation.

We used different gold standards in calculating precision and recall. All three annotators should agree for the gold standard in calculating recall, while a single annotator is enough in calculating precision. When a topic shift is found within two succeeding utterances after a TSRP, that TSRP is considered correct (Reynar, 1994).

Table 3 and 4 show the result of the evaluation using the *Mister O corpus*. F measure is calculated by $F = 2PR/(P + R)$.

Table 3 shows that the combination of *clue expression* and *initiative change* provides the best performance, although it is still worse than the past work as (Arguello and Rosé, 2006). What is interesting is that their recall is poor separately, but combining them doubles the value.

Table 4 shows that combinations including *back-channel* give rise to good F-measure values, suggesting that reciprocal back-channel is a good clue of an ending topic. Recall of detecting TSRPs is fairly good in contrast to precision.

Conclusion Based on the analysis of a free conversation corpus, we proposed a method of detecting topic shifts and topic shift relevance places in information seeking chats. The algorithm was evaluated in comparison with human performance. Although there is much room for improvement, we obtained initial clues for managing topic shifts in real-time information seeking chats. Future work includes sophisticated feature detection modules at the same time as devising a method to select an appropriate new topic at a topic shift.

Acknowledgement This research used the *Mister O corpus* which was built under the project on "Empirical and theoretical studies on culture, interaction, and language in Asia" (No. 15320054, directed by Prof. Ide Sachiko at Japan Women's University) supported by the *Grant-in-Aid for Scientific Research* program from the *Japanese Society for the Promotion of Science (JSPS).*

References

J. Arguello and C. Rosé. 2006. Museli: A multi-source evidence integration approach to topic segmentation of spontaneous dialogue. In *HLT/NAACL*, pages 9–12.

M. Galley, K. McKeown, E. Fosler-Lussier, and H. Jing. 2003. Discourse segmentation of multi-party conversation. In *ACL*.

M. A. Hearst. 1997. Texttiling: Segmenting text into multi-paragraph subtopic passages. *Computational Linguistics*, 23(1):33–64.

J. Hirschbeerg and D. Litman. 1993. Empirical studies on the disambiguation of cue phrases. *Computational Linguistics*, 19(3):501–530.

R. Ochiai, Y. Nomura, and K. Ueno. 2005. Overview of Mister O corpus. In *9th International Pragmatics Conference*. In Panel: Exploring the relationship among culture, language and interaction: Cross-linguistic perspectives.

A. Olney and Z. Cai. 2005. An orthonormal basis for topic segmentation in tutorial dialogue. In *HLT/EMNLP*, pages 971–978.

J. C. Reynar. 1994. An automatic method of finding topic boundaries. In *Proc. of ACL*, pages 331–333.

K. Ries. 2001. Segmenting conversations by topic, initiative, and style. In *SIGIR Workshop: Information Retrieval Techniques for Speech Applications*, pages 51–66.

M. Stede and D. Schlangen. 2004. Information-seeking chat: Dialogue management by topic structure. In *Proc. CATALOG '04*, pages 117–124.