Semi-supervised Algorithm for Human-Computer Dialogue Mining

Calkin S. Montero and **Kenji Araki**
Graduate School of Information Science and Technology
Hokkaido University, Kita 14-jo Nishi 9-chome
Kita-ku, Sapporo 060-0814 Japan
{calkin,araki}@media.eng.hokudai.ac.jp

Abstract

This paper describes the analysis of weak local coherence utterances during human-computer conversation through the application of an emergent data mining technique, *data crystallization*. Results reveal that by adding utterances with weak local relevance the performance of a baseline conversational partner, in terms of user satisfaction, showed betterment.

1 Introduction

Data mining can be defined as the process of finding new and potentially useful knowledge from data. An enhanced trend of data mining is *chance discovery*, which in spite of being an emergent field of research has been applied to different branches of science. Recently, data crystallization (Ohsawa, 2005) has been proposed as a chance discovery extension devoted to find desired unobservable events within a given data. This method of data mining has been intended for revealing events that are significant but are *not* included in the analyzed data (sparseness).

The knowledge database of a conversational partner computer program is a good example of data sparseness, making very difficult to realize a human-like conversation between a user and a computer, since the database is missing relevant or coherent answers for a given user input, e.g., ELIZA system (Weizenbaum, 1966), corpus-based dialogue systems (Inui et al., 2003), and so forth. Herring (1999) noticed the weak sequential coherence during human-computer interaction that is caused by weak local relevance, pointing out the violation of the Gricean maxim of local relevance during the dialogue. However, we argue that conversational responses - often of very weak local relevance - generated by a conversational partner computer program (or "chatbot") could successfully simulate human-like coherent dialogue (see example given by (Herring, 1999), where topical coherence is stated in spite of violation of local relevance Gricean maxim). Moreover, we argue that weak local relevance utterances could be regarded as *topic-shifting utterances*. These utterances are vital in order to keep the flow of the dialogue and its overall coherence. Determining the moment when the topic-shifting utterances are needed is hence essential to the success of a chatbot.

In this paper KeyGraph and data crystallization techniques are applied to analyze human-computer trivial dialogue, *chat*, considering as chances to discover the *missing* topic-shifting utterances, used to improve the overall sequential coherence and the flow of the computer dialogue.

2 Human-Computer Conversation Analysis Approach

During the process of dialogue mining, in order to analyze the behavior of human-computer conversational interaction, the interrelationship of utterances of a dialogue section was obtained. The following is a fragment of an analyzed dialogue section between a user and ALICE chatbot (Wallace, 2005):

[1] User: hi
[2] Chatbot: hello there! what is your name?
[3] User: Joe
[4] Chatbot: Ok, I will call you Joe

[5] User: what should I call you?

[6] Chatbot: my name is Alice

[7] User: cool [....]

(Conversation continues)

[17] User: what do you mean by robot? there are lots of different robots

[18] Chatbot: It is a good question. Have you tried a web search? Would you like to know more?

[19] User: sure

[20] Chatbot: Ok let's do it. What would you like to know?

[21] User: what Alice thinks a robot means

[22] Chatbot: is that a rhetorical question? I would search for it

[23] User: when I hear robot, I think of something that moves

[24] Chatbot: you never know what to expect from life. Why? oh yeah, we are talking about something that moves [....]

For the analysis, the application of data crystallization is introduced in order to identify the missing topic-shifting utterances. The interrelationship of utterances (user-chatbot) is represented in a co-occurrence document created as described below.

2.1 The *KeyGraph*

The KeyGraph has been used as a data-mining tool for extracting patterns of the appearance of chance events (Ohsawa et al.(2003)). The KeyGraph identifies relationships between terms in a document particularly focusing on co-occurrence relationships of both high-probability and low-probability events. Montero et al. (2005) have applied this tool for analyzing the dynamic behavior of human-human chat, identifying criticality.

In this paper the KeyGraph is applied in combination with data crystallization in order to visualized utterances that do not appear during human-computer chat. The interrelationship of utterances (user-chatbot) is represented in a co-occurrence document created by the following algorithm: a) Each utterance (from both, the user and the chatbot) was considered as one sentence. b) Each sentence was segmented into words. c) High frequency words were eliminated, i.e., I, you, is, follow-ups and the like, as to avoid false co-occurrence. d) A vectorial representation of each sentence (at word level) was obtained and sentences co-occurrence relationship was determined as[1]:

$$D = w_1 :: S_1, S_2, S_4 .../ w_2 :: S_9, S_{25} .../$$
$$w_3 :: S_1, S_3, S_{10} .../ ... / w_n :: S_{24}, S_{25}, ... S_m$$

---

[1] Since follow-ups were eliminated, the number of sentences in D might be smaller than the actual number of sentences in the dialogue.

where: w_k $(k = 1, 2, 3, ..., n)$, represents a word in a sentence. S_l $(l = 1, 2, 3, ..., m)$, represents a sentence.

Then it could be said that the obtained D document contains the co-occurrence relationship of the utterances during the analyzed dialogue section. In the graph, the most frequent items in D are shown as black nodes and the most strongly co-occurring item-pairs are linked by black lines according to the Jaccard coefficient:

$$J(S_x, S_y) = p(S_x \cap S_y)/p(S_x \cup S_y)$$

where $p(S_x \cap S_y)$ is the probability that both elements S_x and S_y co-occur in a line in D, and $p(S_x \cup S_y)$ is the probability that either S_x or S_y appears in a line. In the graph, nodes are interpreted as sentences (from D) and clusters of nodes as particular topics (Figure 1).

2.2 Data Crystallization

Data crystallization (Ohsawa, 2005), is dedicated to experts working in real domains where discoveries of events that are important but are not included in the analyzed data are desired. The process of data crystallization involves to insert *dummy items* in the given data in order to represent unobservable events. In this paper, each dummy item inserted in the D document (one in each vector of D) is named X_Y, where X represents the level of the insertion and Y represents the line where the dummy item was inserted. The KeyGraph is applied to the new D document and all of the dummy nodes that did not appear linking clusters in the graph are eliminated from the data, and then the cycle is iterated to higher levels. In the case of the D document of Sec.2.1, after the first level of insertion it becomes:

$$D' = w_1 :: S_1, S_2, S_4 ...1_1 / w_2 :: S_9, S_{25} ...1_2 /$$
$$w_3 :: S_1, S_3, S_{10} ...1_3 / ... / w_n :: S_{24}, S_{25}, ... S_m 1_n$$

where 1_o $(o = 1, 2, 3, ..., n)$, represents each dummy item inserted in each vector of D.

After feeding the KeyGraph with D', all the dummy items that did not appear linking clusters as bridges in the outputted graph are deleted. At this point new dummy items with higher hierarchy (2_x) are inserted in D', and the cycle iterates. Unobservable events and their relations with other events are to be visualized by the application of KeyGraph iteratively to the data that is been crystallized (Figure 2).
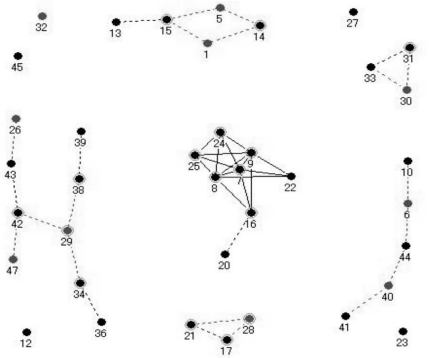
Figure 1: User-Computer Chat Graph

Figure 2: Crystallized Data Graph

3   Experiment and Visual Results

The performed experiment was carried out in three stages. In the first stage of the experiment, three different dialogue sections (including the one shown in Sec.2) between three native English speakers and a chatbot (Wallace, 2005) were analyzed in order to find co-occurrence between the users' utterance and the chatbot replies, i.e., D document. This D document was then examined by the KeyGraph (unsupervised process). Figure 1 shows the graphical view of the dialogue in Sec.2 (48 turns, user - chatbot, in total). A characteristic of the KeyGraph is the visualization of co-occurring events by means of clusters. In Figure 1, the nodes represent sentences from the D document, the clusters represent the relationship among those sentences, i.e., a specific topic, and the nodes that link the clusters represent the transition from one topic to the next. It can be observed that the main clusters are not interconnected, leading to the conclusion that the chatbot in many cases could not keep a smooth and natural flow of the dialogue.

In the second stage of the experiment, a *crystallized document* of utterance co-occurrence, i.e., D' document, was obtained for the same dialogue sections, following the process described in Sec.2.2. The graphical output of the dialogue in Sec.2, after crystallization, can be observed in Figure 2. It can be seen in this figure how the two main clusters appear to be interconnected by the dummy item *1_3*.
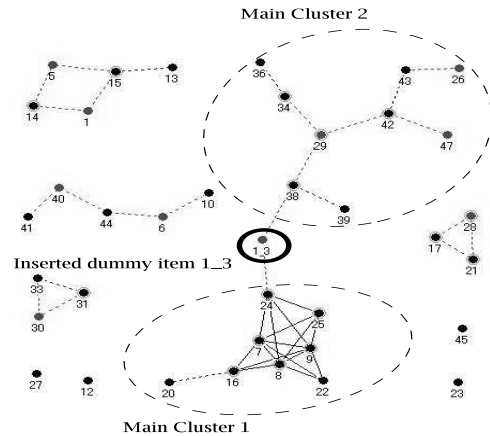
Although this dummy item was inserted in the third line of the D document, it appears in the graph connecting the two main clusters. The dummy item *1_3* branches from utterance [24]. This interconnecting point can be regarded as the system considering it appropriate to insert a topic-shifting utterance at this point of the conversation. In doing so, a well interconnected graph is obtained (Figure 2). This information is valuable for making the chatbot to ask "intelligent questions" as a mean of *conversational* responses to keep the interest from the user.

In the third stage of the experiment, the information yielded by the previous analysis, i.e., regarding the timing where a topic-shifting utterance might be needed, was used to feed the chatbot database. *Topic-shifting responses* were inserted by hand (supervised process) as general patterns (around one hundred patterns) for smoothly change the topic when there is not a pattern that matches a given utterance. In this way a bridge, represented in Figure 2 by the dummy item, is created giving to the dialogue the desired smoothness. Seven users (four native English speakers, three non native speakers) were requested to perform a chat with the plain chatbot and with the enhanced chatbot (the users did not know which chatbot was plain or which was enhanced). The time set up was maximum 30 minutes-chat with each program, the user was free to stop at any time before the time limit. The evaluation of the chatbots performances was made through a ques-

	Turns	% V.R.*	Accuracy/relevance
Baseline	74	21.11%	fair
Enhanced	128	7.37%	good

V.R.* = Vague Reply (This table contains the average of Turns and VR)

Table 1: Chatbots Overall Performance

tionnaire given to the users. In the questionnaire the degree of satisfaction of the user was surveyed based on the accuracy/relevance of the chatbot responses and its ability to keep the dialogue flowing. The accuracy/relevance had a ranking scale: 1. excellent, 2. good, 3. average, 4. fair, and 5. poor, while the user satisfaction ranked : 1. very satisfied, 2. satisfied, 3. not sure, 4. dissatisfied, and 5. very dissatisfied.

The four native English speaker users globally agreed ranking enhanced chatbot as having "good" accurate/relevant responses during the overall dialogue, giving remarks like "interesting to talk with", "better at keeping the conversation going", "easier to talk to", and "more impression of consciousness". In the case of the plain chatbot, the native speakers showed "dissatisfied" at its performance, giving remarks like "loop responses", "slow in keeping the conversation going", "no so nice responses", "no constant memory". Table 1 shows a resume of the average performance of the chatbot for all of the users. An example of a vague reply is given in the following fragment:

Chatbot: What is like living in Sapporo?

User: sweet as

Chatbot: In the context of living in Sapporo I don't understand "what is like living in Sapporo?" **[vague reply]**

Two non native English speaker users ranked the enhanced chatbot as having "fair" and "average" accurate/relevant responses while the plain chatbot was ranked as having "poor" and "fair" accurate/relevant responses. The third non native English speaker user ranked both chatbots as "poor" due to "the chatbots lack of understanding deixis, and anaphor".

As a mean of discussion, in Figure 2 it could be expected that the dummy item *1_3* would branch from utterance [25] {*User: no, you asked me who is the best robot*}, which is in the same cluster with utterance [24]. However, under closer examination it becomes clear that utterance [24] has stronger co-occurrence with utterance [38] {*Chatbot: I know you are but what am I*} than utterance [25]. Hence, the algorithm suggests

to link the clusters via utterance [24].

In other aspect, based on the feedback given by the seven users of the experiment, the overall performance of the enhanced chatbot can be considered better than the plain chatbot. It is worth noticing that the evaluation of the non native English speaker users tended to emphasize the grammatical aspect of the chatbots responses. On the other hand, the evaluation of the native English speaker users tended to emphasize the smoothness of the dialogue. Although there is still plenty of room for improvement and research a betterment in the chatbot performance could be seen through this approach.

4 Conclusion

In this paper the application of a novel data mining method, data crystallization, for visualizing missing topic-shifting utterances during human-computer chat has been described. Based on this information, during the experiment, the use of weak local relevance utterances, i.e., topic-shifting responses, despite of violation of Grecian maxim of local relevance, showed to meliorate the overall dialogue flow. Future research will be oriented to the extended implementation of the obtained results for enhancing the chat flow modeling of a conversational partner program.

References

Susan Herring. 1999. Interactional coherence in cmc. *Journal of Computer Mediated Communication*, 4(1).

Nobuo Inui, Takuya Koiso, Junpei Nakamura, and Yoshiyuki Kotani. 2003. Fully corpus-based natural language dialogue system. In *Natural Language Generation in Spoken and Written Dialogue, AAAI Spring Symposium*.

Yukio Ohsawa and Peter McBurney, editors. 2003. *Chance Discovery*. Springer, Berlin Heidelberg New York.

Yukio Ohsawa. 2005. Data crystallization: Chance discovery extended for dealing with unobservable events. *New Mathematics and Natural Computation*, 1(3):373–392.

Calkin S Montero and Kenji Araki. 2005. Human chat and self-organized criticality: A chance discovery application. *New Mathematics and Natural Computation*, 1(3):407–420.

Richard Wallace. 2005. A.l.i.c.e. artificial intelligence foundation. http://www.alicebot.org.

Joseph Weizenbaum. 1966. Eliza a computer program for the study of natural language communication between man and machine. *Commun. ACM*, 9(1):36–45.