Automatic Extraction of Discussion based on Sentence Type Estimation

Yuto Hoshikawa
Graduate School of Library, Information and Media Studies, University of Tsukuba
1-2 Kasuga, Tsukuba, Ibaraki, Japan
s1521643@u.tsukuba.ac.jp

Kei Wakabayashi
Faculty of Library, Information and Media Science, University of Tsukuba
1-2 Kasuga, Tsukuba, Ibaraki, Japan
kwakaba@slis.tsukuba.ac.jp

CSCW '17 Companion, February 25 - March 01, 2017, Portland, OR, USA
ACM 978-1-4503-4688-7/17/02.
http://dx.doi.org/10.1145/3022198.3026336

Abstract

Automatic analysis of conversational data in computer-supported communication is useful to support us to understand a network of people. We propose a method that detects *discussion* from conversations automatically. The proposed method estimates types of utterance on the basis of speech act theory and applies the insight from conversation analysis studies to measure how likely is this conversation a discussion. We confirm the effectiveness of the proposed method by an experiment that examines correlation with human annotated purposefulness scores.

Author Keywords

conversation; discussion; social media

ACM Classification Keywords

I.2.7 [Natural Language Processing]

Introduction

Computer-supported communication is pervasive in our activities including business messaging, social chat and discussions. These communications are usually recorded as data, which is possibly useful for understanding a network of people. For example, a manager who organizes teams of staffs should be concerned about the human relationships between members for making a productive pairing that yields a constructive discussion. [4] An assistance to

Sentimental
Fact
Supplement
Value judgment
Knowledge gain
Knowledge provide
Knowledge teach
Being taught
Request
Confirm
Request action
Request saying
Propose
Thank
Apology
Agreement
Opposite
Accept action request
Accept saying request
Satisfaction
Say reason
Hold
Switch topic

Table 1: Types of sentence in a modified version of DAMSL annotation tag set [1] that is reorganized to adapt for Japanese conversation analysis by Nishihara et al. [5] They also made a semantic dictionary that relates Japanese auxiliary verbs to each sentence type.

such manager using a computer system is helpful especially when there are a lot of staffs to be managed. For this purpose, we should detect people who often makes discussions, not a daily conversation, because the discussion is an important factor in the cooperative work.

In this study, we classify conversations into two classes: aimless conversation, *chat*, and purposeful conversation, *discussion*. We propose an algorithm that automatically extracts discussions among conversations with the definition of *discussion score* that indicates a degree of the purposefulness of conversation. The discussion score is designed to reflect the insight from conversation analysis studies [3] by combining the speech act theory that is extensively implemented by DAMSL annotation method. [2] We confirmed the effectiveness of the proposed method by examining correlation between the discussion score and human annotated purposefulness score. In this paper, we report a work that focuses on Japanese conversation, but we believe the method can be extended to another language.

Related Work

It is demonstrated that an automatic analysis of conversation is useful to support human communication to date. Azaria et al. [1] proposed an automated agent that can help people detect deception in a discussion environment by using machine learning technique. Nishihara et al. [5] proposed a method that estimates how much intimate persons are by analyzing conversation data. Nishihara et al. refer to DAMSL annotation tag set [2], which is an approach to implement speech act theory in a computational way, to make an accurate conversation analysis possible in the algorithm (Table 1). Our proposed method is based on their method that calculates an intimacy indicator, *friendship score*. However, a pair of people that have high friendship score can be one that just makes a long conversation, which is possibly a contentless chat that is not appropriate to make in a working group.

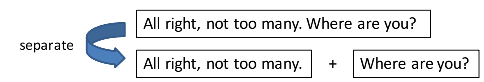

Figure 1: Separating tweet

We extend the Nishihara's friendship scoring method by applying the insight from conversation analysis, which is an approach of social study that focuses on conversational interaction. Koiso et al. [3] conducted an analytical survey of human conversations and suggested that conversation can be categorized by focusing on its purposefulness. Ojima et al. [6] categorized an utterance in a purposeful discussion into nine types that is listed in Table 1. We apply these research results in our proposed method that extracts discussions.

Proposed Method

In this section, we explain the proposed method to extract *discussion* on the basis of Nishihara's method. [5] The input is text sentences of utterance. Generally, a text message includes multiple sentences. It is desirable to explicitly separate them because each sentence has different meaning. We divide a message into sentences with punctuations and particular symbols. Figure 1 shows an example of sentence separation.

We process each sentence to extract Japanese auxiliary verbs, which express speaker's attitude. We apply Nishihara's semantic dictionary that relates Japanese auxiliary verbs to each DAMSL-based sentence type in Table 1 (for detail, see [5].) Then, we calculate a vector of sentence types using rules for identifying of Table 1. In

Sentence type for discussion
Proposal: Proposals of the problem, topic, presentation of the new topic
Propose, Switch topic
Confirmation: Questions to be answered by yes/no
Confirm
Explanation: Presentation of the additional information and detailed information about the current topic
Knowledge provide
Agreement: Position expressed of agreement
Agreement
Disagreement: Position expressed of disagreement
Opposite
Question: Question for others
Request saying
Supplement: Additional information to the remarks just before
Supplement
Answer: Answer to a question
Knowledge teach, Knowledge provide
Other: An impossible classification remarks in any of the above
Not applicable

creating vectors of sentence types, we exclude sentence types that don't appear in discussion. In this paper, we use a set of sentence types that are used in discussions defined by Ojima et al. [6], which is shown in the side bar. We denote vector of sentence types for sentence d by $V_d = (x_1, x_2, ..., x_R)$, where R is the number of sentence type in the side bar, i.e. $R = 9$. The value of vector V_d is given by the number of meanings of auxiliary verbs with the sentence and the weight of sentence types corresponding to meanings. Set of auxiliary verbs with sentence d is P_d, meanings of auxiliary verbs in P_d is $m^j_{p_i}$ and the element value denoted $x_{l,d}$ of vector of sentence types V_d is calculated by (1).

$$x_{l,d} = \left(\sum_{p_i \in P_d} h(x_l, m^j_{p_i}) \right) \times w(x_l) \tag{1}$$

In (1), $h(x, m)$ is related sentence types x and meaning m in Table 1, is a function that returns 1 if meaning is on end of its sentence and 0.5 otherwise. $w(x)$ is a function that is a weight of the sentence type x.

The *Discussion score* is computed on the basis of the strength of sentence type that is used in discussion. We propose two versions of method to estimate discussion score. The first method estimates discussion score regarding the number of sentence types. A sentence d is regarded to have a sentence type l when $x_{l,d}$ exceeds a threshold $T = 1.0$, which is same to [5]. The number of sentence types $k(s)$ in all sentences of the speaker s is calculated by (2).

$$k(s) = \sum_{l=1}^{R} \begin{cases} 1 & \texttt{if} \sum_{d \in D_s} x_{l,d} \geq T \\ 0 & \texttt{otherwise} \end{cases} \tag{2}$$

where D_s is a set of sentences by speaker s. Discussion Score $Score1(s_a, s_b)$ of conversation c is calculated by (3) using $k(s)$ and the number of messages in conversation t.

$$Score1(s_a, s_b) = \frac{k(s_a) \times k(s_b)}{t} \tag{3}$$

The second version of the proposed method reflects the strength of vector elements, which might be important to estimate the discussion score. The total strength of sentence type $x_{l,d}$ is simply calculated by (4).

$$x_{sum}(s) = \sum_{l=1}^{R} \sum_{d \in D_s} x_{l,d} \tag{4}$$

Discussion Score $Score2(s_a, s_b)$ of conversation c is calculated by (5) with $x_{sum}(s)$ and the number of messages of the conversation t.

$$Score2(s_a, s_b) = \frac{x_{sum}(s_a) \times x_{sum}(s_b)}{t} \tag{5}$$

Experiment

We confirm the effectiveness of the proposed methods by an experiment that examines the correlation with the human annotated purposefulness score of conversation by using Twitter dataset. The dataset is composed of 100 conversations sampled from conversations between April 1, 2013 and April 30, 2013 extracted from Twitter. Each conversation is evaluated by one human annotator to decide its purposefulness score in scale of 1 to 5. The number of tweets, the number of characters per tweet and Friendship Score are comparative methods to verify the effectiveness of proposed methods.

Table 4 shows correlation coefficients of each method with human annotated score. The result shows that the proposed methods are effective because their values are higher than other comparative methods. $Score1$ is the best and its

Everyone, Good morning. Today is going to be a busy day. ...
@A Hi, Pitch. Good morning. Congratulations on the entrance ceremony...
@B Good morning. Thank you. Today is going to be a good anniversary at ...

Table 2: Conversation like *chat*.

Speaking of memories of Satsuki Award, I think Mihono Burubon. ...
@C I think he was strong, too. ...
@D Because there was not the foreign jockey, there was the drama of ...

Table 3: Conversation like *discussion*.

	Correlation
$Score1$	0.573
$Score2$	0.555
Number of tweets	0.185
Number of characters per tweet	0.498
Friendship Score	0.447

Table 4: Correlation coefficient.

correlation is 0.573. We consider that there are two reasons why proposed methods show the higher correlation coefficients than comparative methods. One factor is to choose sentence types for discussion because the correlation coefficients are higher than that of Friendship Score. Also, because it is divided by the number of messages, the scores are not affected by the number of messages and can reflect how much conversation is like discussion. In $Score1$, the threshold is $T = 1.0$ on the experiment, but this is the threshold that Nishihara et al. defined to calculate the number of sentence types. Figure 2 shows the correlation coefficient varying the threshold. For Figure 2, the highest correlation coefficient is 0.593 when $T = 0.6$ and $T = 0.7$, which is lower than $T = 1.0$ of the experiment.

Conclusion

We proposed a method extracting discussion from conversations and confirmed its effectiveness by an experiment. We define an indicator that denotes the degree of purposefulness as discussion score. The proposed method computes vectors of sentence types from Japanese auxiliary verbs. The calculated vectors reflect the insight from study of conversational discussion analysis. Although we focus on Japanese conversation in this work, we believe the approach on the basis of speech act theory is applicable for another language.

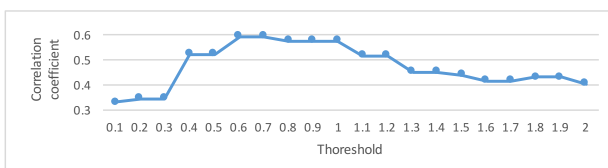

Figure 2: Threshold

Acknowledgement

This work was supported by JSPS KAKENHI 16H02904 and NII ’ s strategic open-type collaborative research.

REFERENCES

1. Amos Azaria, Ariella Richardson, and Sarit Kraus. 2015. An Agent for Deception Detection in Discussion Based Environments. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing (CSCW ’15)*. 218–227.
2. Mark G Core and James F Allen. 1997. Coding Dialogs with the DAMSL Annotation Scheme. In *Working Notes of the AAAI Fall Symposium on Communicative Action in Humans and Machines*.
3. Hanae Koiso, Tomoyuki Tsuchiya, Ryoko Watanabe, Daisuke Yokomori, Masao Aizawa, and Yasuharu Den. 2016. Survey of conversational behavior, corpus design, corpus of conversational Japanese. *NINJAL Research Papers* 10 (2016), 85–106.
4. Ioanna Lykourentzou, Angeliki Antoniou, Yannick Naudet, and Steven P. Dow. 2016. Personality Matters: Balancing for Personality Types Leads to Better Outcomes for Crowd Teams. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing (CSCW ’16)*. 260–273.
5. Yoko Nishihara and Wataru Sunayama. 2009. Estimation of friendship and hierarchy from conversation records. *Information Sciences* 179, 11 (2009), 1592 – 1598.
6. Keiichi Ojima and Toshio Okamoto. 1996. A Study on Presumption of Utterance Intention among the Dialogue over the CSCW(3). *Technical Report of IEICE. AI* 95, 460 (1996), 87–94.