A Frustratingly Simple Conditional Likelihood-based Method for Dialogue Orientation Estimation

Yin Jou Huang, Sadao Kurohashi
Kyoto University
{huang, kuro}@nlp.ist.i.kyoto-u.ac.jp

Abstract

During a conversation, the participants constantly switch between strategies of different orientations. A participant can either adopt a forward strategy to advance the conversation toward a desirable direction, or a backward strategy to address previous conversation contents. In this work, we use dialogue orientation as a measure of an utterance's intention to change the conversational topic and propose a conditional likelihood-based method for dialogue orientation estimation. We performed experiments in both Japanese and English and verified the effectiveness of our proposed method.

1 Introduction

During a conversation, an interlocutor can choose to direct the conversation towards a new direction (**forward** strategy), or simply address what is previously said (**backward** strategy) [1]. Take Figure 1 as an example. In response to the utterance "I like to read in my spare time.", the utterance "What genre of books do you prefer?" is a forward utterance that prompts the subsequent discussion of book genres. On the other hand, the response "Reading is a good pastime!" is a backward utterance that makes comments on the previous utterance. **Dialogue orientation** is used to quantify the forward/backward strategies. In this work, we define dialogue orientation as the degree of an utterance's intent to direct the conversation toward a specific topic. By this definition, a typical forward utterance would have a high orientation score, and a backward utterance would have a low orientation score.

A good interlocutor controls the flow of the dialogue by adequately switching between strategies of different dialogue orientations. If the backward strategy is overly dominant, the conversation may be dull. On the other hand, excessive use of the forward strategy may progress the conversation too hastily and make it difficult to build rapport between the interlocutors. Thus, controlling dialogue orientation is critical for building engaging dialogue systems.

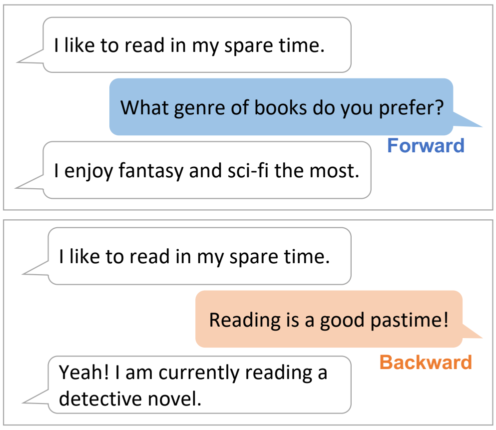


Figure 1: Forward and backward utterances.

Previous work of dialogue orientation focused on counseling dialogues and categorized counselors' utterances into backward utterances that show empathy and forward utterances that guide the patient to a calmer state [1, 2]. They use singular value decomposition to acquire utterance representations and calculate the orientation score based on the variance of preceding and following utterances. Inspired by their work, we further improved the dialogue orientation estimation method to be more robust, and also considered the semantic relations of the target utterance and the surrounding context.

In this paper, we propose a simple conditional likelihood-based method for dialogue orientation estimation. We observe that while a forward utterance is semantically related to both its preceding and following contexts, a backward utterance is semantically related only to its preceding context but not its following context. Based on the above characteristics, we calculate the conditional like-

This work is licensed by the author(s) under CC BY 4.0
(https://creativecommons.org/licenses/by/4.0/).

lihood of a target utterance under different contexts and calculate the orientation score of the utterance. After the orientation scores are calculated, we use a heuristic method to decide the threshold value between backward and forward strategies and get the discrete orientation labels.

We conduct experiments in both Japanese and English, using DailyDialog dataset [3] and Kyoto University Hobby Chat Dialogue Corpus (KUHCC) [4] respectively. Evaluations on manually labeled data and showed improved results over previous method. We also perform a thorough analysis of the characteristics of dialogue orientation and verified the effectiveness of our proposed method.

2 Dialogue Orientation Estimation

In this section, we introduce our proposed method for dialogue orientation estimation. We train a conditional utterance prediction model and use the model to estimate the dialogue orientation of a given utterance. Also, we propose a heuristic method to deduce discrete orientation labels.

2.1 Conditional Utterance Prediction Model

We train a conditional utterance prediction model to capture the semantic relations among utterances. Taking the history and future contexts as input, the model predicts the utterance that fits the given context.

For each utterance u_t in the corpus, we take k preceding utterances as history context and j following utterances as future context. The model input is defined as follows:

$$u_{t-k} \text{ [SEP] } ...u_{t-1} \text{ [MASK] } u_{t+1} \text{ [SEP] } ... \ u_{t+j} \tag{1}$$

where special token [SEP] indicates the utterance boundaries, and [MASK] indicates the position of the target utterance. Given the above input, the model generates the target utterance u_t.

2.2 Conditional Likelihood-based Dialogue Orientation Score

Based on the conditional utterance prediction model (Section 2.1), we calculate the conditional likelihood-based dialogue orientation score.

The following is the intuition of our proposed method: Since a forward utterance directs the conversation topic toward a specific direction, it imposes semantic constraints on its subsequent utterances. Thus, the future context provides important information for predicting a forward utterance. Take the forward utterance in Figure 1 as an example. It is difficult to predict the forward utterance accurately only by looking at the history context. However, if we look at the future context ("I enjoy fantasy and sci-fi the most."), we can get a rough idea that the utterance before it is talking about genres of books.

In contrast, a backward utterance addresses previous conversational contents and does not impose strong additional constraints on its following context. Thus, the future context of a backward utterance often provides little information in the prediction of the backward utterance.

Based on the above observations, we consider the conditional likelihood of a target utterance u given different contexts. For typical forward utterance u^f and backward utterance u^b, we expect the conditional likelihood of u to exhibit the following property:

$$\begin{cases} p(u^f|history, future) > p(u^f|history) \\ p(u^b|history, future) \approx p(u^b|history) \end{cases} \tag{2}$$

with $history$ and $future$ being the history and future context of u.

We use perplexity to measure the conditional likelihood and calculate the dialogue orientation score. Given k preceding utterances and j future utterances of the target utterance u_t, we calculate the perplexity of u_t conditioned on both history and future context (ppl_{ALL}), and the perplexity of u_t conditioned on history context alone (ppl_{HIST}):

$$\begin{cases} ppl_{ALL} = ppl(u_t|u_{t-k}, ..., u_{t+1}; u_{t+1}, ..., u_{t+j}) \\ ppl_{HIST} = ppl(u_t|u_{t-k}, ..., u_{t-1}) \end{cases} \tag{3}$$

We define the orientation score under this condition as:

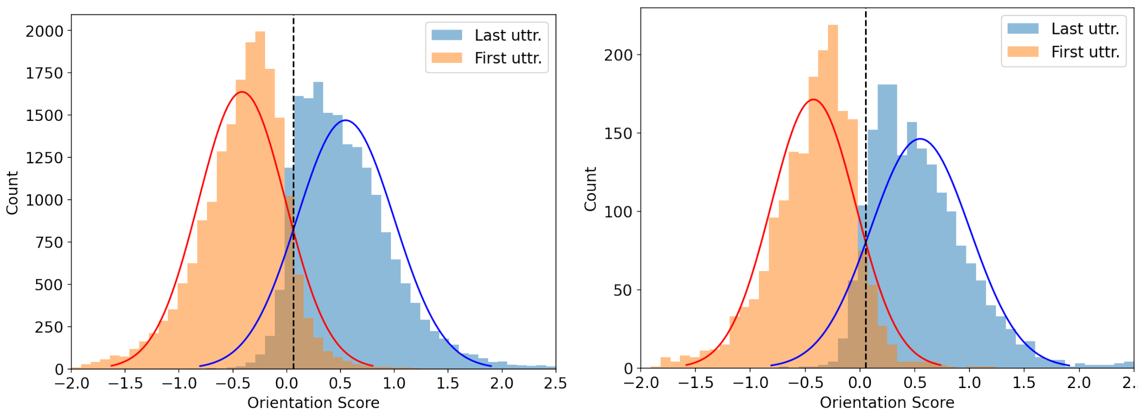
$$s_t(k, j) = -(ppl_{ALL} - ppl_{HIST}) \tag{4}$$

Note that the value of $s_t(k, j)$ depends on the length of history and future context (k and j). To calculate the orientation score of u_t, we fix the history context size to $k = 3$ and take the future context size j that gives the highest $s_t(k, j)$:

$$s_t = \max_{j \in [1,5]} s_t(k, j) \tag{5}$$

Multi-sentence utterances An utterance can contain more than one sentence, and different sentences could exhibit different orientation strategies. Thus, in addition to the utterance-level orientation scores (u_t), we also calculate the sentence-level orientation scores. Replace u_t with its constituent sentence in equation (3) gives the sentence-level orientation score.

This work is licensed by the author(s) under CC BY 4.0
(https://creativecommons.org/licenses/by/4.0/).



(a) DailyDialog. (b) KUHCC.

Figure 2: Distribution of the forward and backward utterances.

2.3 Discrete Orientation Label

The scores calculated in Section 2.2 provide a measure of dialogue orientation, but how large/small does a score have to be to suggest a forward/backward strategy? In this section, we introduce a heuristic method to decide a threshold value between forward and backward utterances.

In dialogues between human interlocutors, a speaker tends to first address the previous conversational content (backward) and then advance to some new topic (forward), but not vice versa. We focus on utterances with multiple sentences, and calculate the orientation score of the first and the last sentence. If the last utterance exhibits a higher orientation score than the first utterance, then the utterance is likely to exhibit the 'backward-then-forward' pattern. Figure 2 shows the distribution of orientation score of the first and last sentences in multi-sentence utterances (DailyDialog dataset). Here, we view the first sentence as a backward utterance and the last sentence as a forward utterance.

Next, we fit Gaussian distributions to the orientation scores of the last sentences (forward) and first sentences (backward), respectively. We use the intersection of the two Gaussian curves as the threshold value δ between forward and backward orientation scores. Using this threshold value, we classify utterances with an orientation score larger than δ as forward utterances, and the ones with an orientation score smaller than δ as backward utterances.

3 Experiments

In this section, we conduct experiments to verify the effectiveness of our proposed orientation estimation method.

Method	Accuracy	
	DailyDialog	KUHCC
Baseline (Zhang+2020)	60.6	81.7
Proposed	**81.9**	**86.0**

Table 1: Orientation prediction accuracy of different orientation estimation methods.

3.1 Experimental Settings

In this work, we conduct experiments in English and Japanese. For English, we use the general-purpose DailyDialog dataset [3]. The dataset contains 13,118 dialogues of 10 different categories. For Japanese, we use the KUHCC [4] corpus. The corpus contains 4,911 dialogues collected with crowdsourcing. In each dialogue, the two participants talk about their hobbies.

We use the pretrained BART base model [5] for the conditional utterance prediction model. For each target utterance, we take the previous $k = 3$ utterances as the history context and randomly use $j = 5$ utterances as the future context. The threshold value obtained by the discrete label deduction method method in Section is $\delta_{en} = 0.209$ for English and $\delta_{jp} = 0.053$ for Japanese corpus.

3.2 Evaluation

We perform quantitative analysis with manually labeled samples. We manually labeled 158 utterances in the KUHCC corpus with orientation labels (forward or backward). Similarly, for the DailyDialog corpus, 160 labeled utterances are collected as the evaluation set.

Table 1 shows the accuracy score of our proposed method (**Proposed**) and the orientation estimation method

This work is licensed by the author(s) under CC BY 4.0 (https://creativecommons.org/licenses/by/4.0/).

	Proposed	Baseline
S_1: I love watching baseball. However, I haven't watched any game since last year. **S_2: You haven't been to even one game?** S_1: I really want to go, but the tickets are difficult to come by due to the restrictions.	**B**	F
S_1: I love watching soccer. **S_2: You go watch the game in person?** S_1: I got into the habit only several years ago, I haven't got the chance to watch in person.	**F**	**F**
S_1: I am the fan of the baseball player Martin. S_2: I don't know much about baseball, but I have heard of him! S_1: I just went to a game in Makuhari last weekend. **S_2: You go watch the game in person?** S_1: Yeah, it was really exciting!	**B**	F

Table 2: Case analysis. The target utterance is in bold font and the correct label is highlighted in bold. The correct orientation label is highlighted in bold. Examples from the KUHCC corpus are translated into English.

proposed by Zhang et al. [1] (**Baseline**). The proposed method outperformed the baseline method by 21.3 points for DailyDialog, and 4.3 points for KUHCC corpus.

3.3 Case Study

We perform a case study of the predicted orientation labels (Table 2).

Indirect utterance The first row of Table 2 shows an example of *indirect utterance* whose surface form differs from its intent. The target utterance is in the interrogative form, but its intent is confirmation rather than asking for information. Since typical questions are followed by corresponding answers to fulfill the turn-taking rules of communication [6, 7], they lean towards the forward end of the orientation spectrum. However, a question could have intents other such as confirmation or irony [8], which indicates a backward strategy. Our proposed method classified this utterance as backward correctly, while the baseline method classified it as forward. We speculate that the baseline method is influenced by the textual indicators of questions[1)] and has the tendency to classify the utterance with question form as forward. This example shows that our proposed method predicts the orientation not only based on the surface form but also on the semantic relations between the target utterance and the surrounding context.

Context dependency The second and third rows in Table 2 illustrate that an utterance could have a different orientation depending on the context. In the first sample, the utterance "You go watch a game in person?" is a forward utterance and leads to further discussion about different ways of watching soccer games. However, in the second sample, the same utterance is a backward utterance that aims to express the emotion of surprise. The proposed orientation estimation method correctly predicts both samples. The above examples illustrate that our proposed method can adequately reflect the difference in context when predicting the orientation.

4 Conclusion

In this work, we studied dialogue orientation, which is an important dialogue characteristic indicating the forward strategy and backward strategy of conversation. We proposed a simple conditional likelihood-based method for estimating dialogue orientation. We applied our proposed dialogue orientation estimation method to Japanese and English corpora and conducted a thorough analysis. The analysis illustrated the robustness of our model in handling cases of misleading surface forms. Also, the proposed model adequately captures the context-dependent nature of dialogue orientation.

For future work, we consider downstream applications of dialogue orientation. We expect the dialogue orientation measure to be helpful in dialogue analysis, such as dialogue breakdown detection, quality control of crowd-sourced dialogue data, etc. Also, incorporating and controlling dialogue orientation to a dialogue agent can lead to more engaging conversations.

1) Many languages have textual indicators of questions, such as question marks, interrogative words such as the 5W1H in English, the *ka* in Japanese, to name a few.

This work is licensed by the author(s) under CC BY 4.0 (https://creativecommons.org/licenses/by/4.0/).

Acknowledgement

This work was supported by project JPNP20006, commissioned by New Energy and Industrial Technology Development Organization (NEDO). Also, this work was supported by JST, CREST Grant Number JPMJCR20D2, Japan.

References

[1] Justine Zhang and Cristian Danescu-Niculescu-Mizil. Balancing objectives in counseling conversations: Advancing forwards or looking backwards. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, pp. 5276–5289, Online, July 2020. Association for Computational Linguistics.

[2] Brian L Mishara, François Chagnon, Marc Daigle, Bogdan Balan, Sylvaine Raymond, Isabelle Marcoux, Cécile Bardon, Julie K Campbell, and Alan Berman. Which helper behaviors and intervention styles are related to better short-term outcomes in telephone crisis intervention? results from a silent monitoring study of calls to the us 1-800-suicide network. Suicide and Life-Threatening Behavior, Vol. 37, No. 3, pp. 308–321, 2007.

[3] Yanran Li, Hui Su, Xiaoyu Shen, Wenjie Li, Ziqiang Cao, and Shuzi Niu. DailyDialog: A manually labelled multi-turn dialogue dataset. In Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 1: Long Papers), pp. 986–995, Taipei, Taiwan, November 2017. Asian Federation of Natural Language Processing.

[4] Takashi Kodama, Frederic Bergeron, Hayato Atarashi, Ribeka Tanaka, Wataru Sakata, and Sadao Kurohashi. Japanese Dialogue Collection Platform Available in Crowdsourcing. In Proceedings of the 27th Annual Conference of the Association for Natural Language Processing (In Japanese), pp. 859–864, 2021.

[5] Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Veselin Stoyanov, and Luke Zettlemoyer. BART: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, pp. 7871–7880, Online, July 2020. Association for Computational Linguistics.

[6] Emanuel A Schegloff and Harvey Sacks. Opening up closings. 1973.

[7] Silvia Pareti and Tatiana Lando. Dialog intent structure: A hierarchical schema of linked dialog acts. In Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018), Miyazaki, Japan, May 2018. European Language Resources Association (ELRA).

[8] John R Searle. Indirect speech acts. In Speech acts, pp. 59–82. Brill, 1975.

This work is licensed by the author(s) under CC BY 4.0
(https://creativecommons.org/licenses/by/4.0/).