Dialogue Act-based Breakdown Detection in Negotiation Dialogues

Atsuki Yamaguchi, Kosui Iwasa, Katsuhide Fujita

Tokyo University of Agriculture and Technology, Tokyo Japan.

8 Oct. 2021, SCAI'21, Online

Background

- Negotiation is an essential task.
- Studies of negotiation dialogues get attentions, recently.
  - Goal-oriented dialogue systems
    - Deal or No Deal (DN) (Lewis et al., 2017) – Item division
    - Craigslist Bargain (CB) (He et al., 2018) – Price negotiation
  - Support for human-human negotiation
    - Nash bargaining solution estimation (Iwasa and Fujita, 2018)
    - Real-time negotiation coaching (Zhou et al., 2019)
    - Negotiation breakdown detection (Yamaguchi and Fujita, 2020)

Problems and Contributions

- **Problems:**
  1. **Few negotiation dialogues corpus:**
     - The existing corpus was DealOrNoDeal(DN) and CraigslistBargain(CB).
     - Negotiation settings are simplified.
     - Some other corpora are too small to use.
  2. **Negotiation breakdown detection**: No effective method has been proposed.

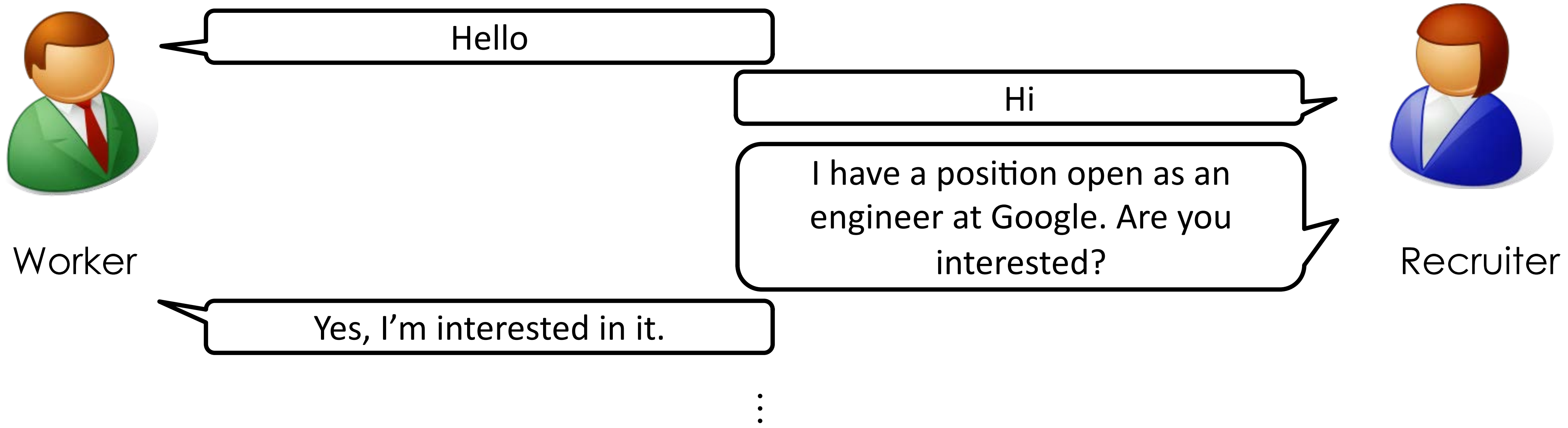
- **Contributions**:
  1. Develop a new negotiation dialogue corpus; **Job-interview dialogues Corpus**
  2. Propose **a dialogue act-based feature with a machine learning model** for negotiation breakdown detection.

Job Interview Negotiation Dataset (1 / 2) - Overview

- **Scenario**:
  Job interview between two negotiators in English.
- **Mathematical settings**:
  (Negotiation score) = (Linear additive utility function) + (Bias)
- **Data collection**: Amazon MTurk.

Issue	Option
Salary	\$20 to \$50 per hour (integer)
Weekly day off	2 days to 5 days (integer)
⋆Position	{Engineer, Designer, Manager, Sales}
⋆Company	{Google, Apple, Facebook, Amazon}
Workplace	{Tokyo, Seoul, Beijing, Sydney}

Table: List of issues. ★ denotes an interdependent relationship



Job Interview Negotiation Dataset (2 / 2) - Statistics

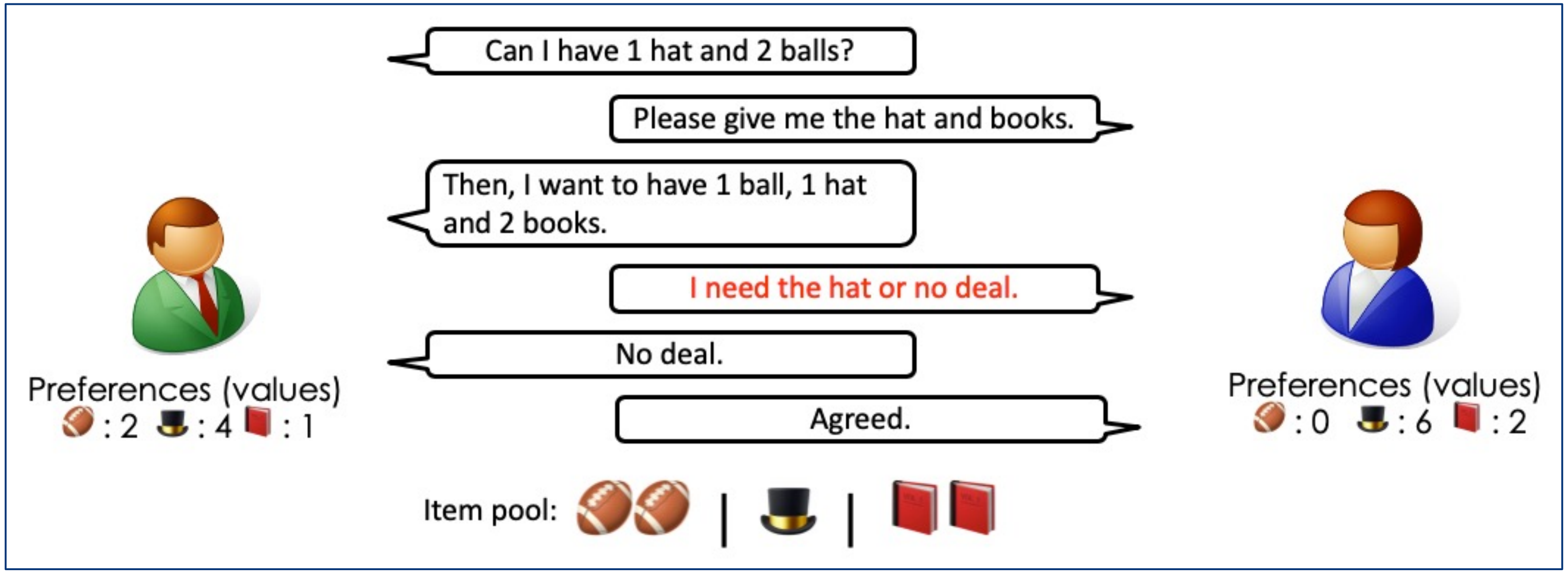
- Negotiators in the JI dataset likely had enough conversation.
- **Agreement ratio**: Huge gap between JI and others.
- **Complexity of Negotiation Scenarios**: JI has far fewer Pareto optimal solutions for agreements.
  1. Large number of possible solutions per dialogue
  2. Introduction of interdependent relationship

	JI	DN	CB
# of dialogues	2,639	6,251	6,682
Avg turns per dialogue	12.7	4.97	7.53
Avg words per turn	6.12	8.56	13.60
Vocab size	4,476	2,631	12,139
Agreed [%]	92.9	76.2	74.9
PO solutions [%]	13.4	75.0	
PO bids for all bids [%]	0.98	18.0	
# of all bids per dialogue	9,920	22.5	
Avg score	6.4 / 10	5.7 / 10	

Table: Statistics of the three negotiation corpora.

Negotiation Breakdown Detection Task (1 / 2)

- **Task:** Given a dialogue composed of many turns' utterances between two negotiators, a model needs to classify whether its outcome is successful or not.



Negotiation is

Breakdown

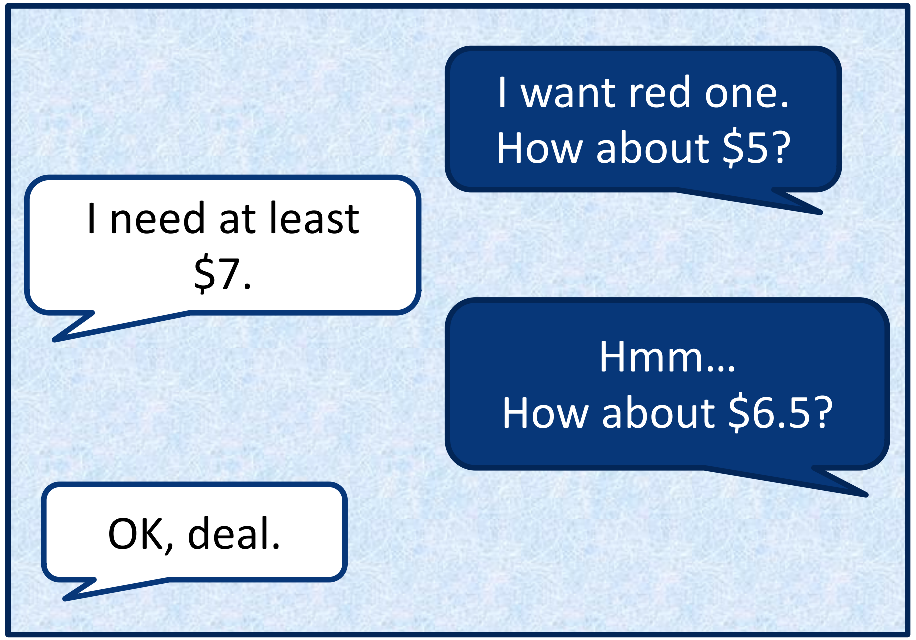
or

Successful?

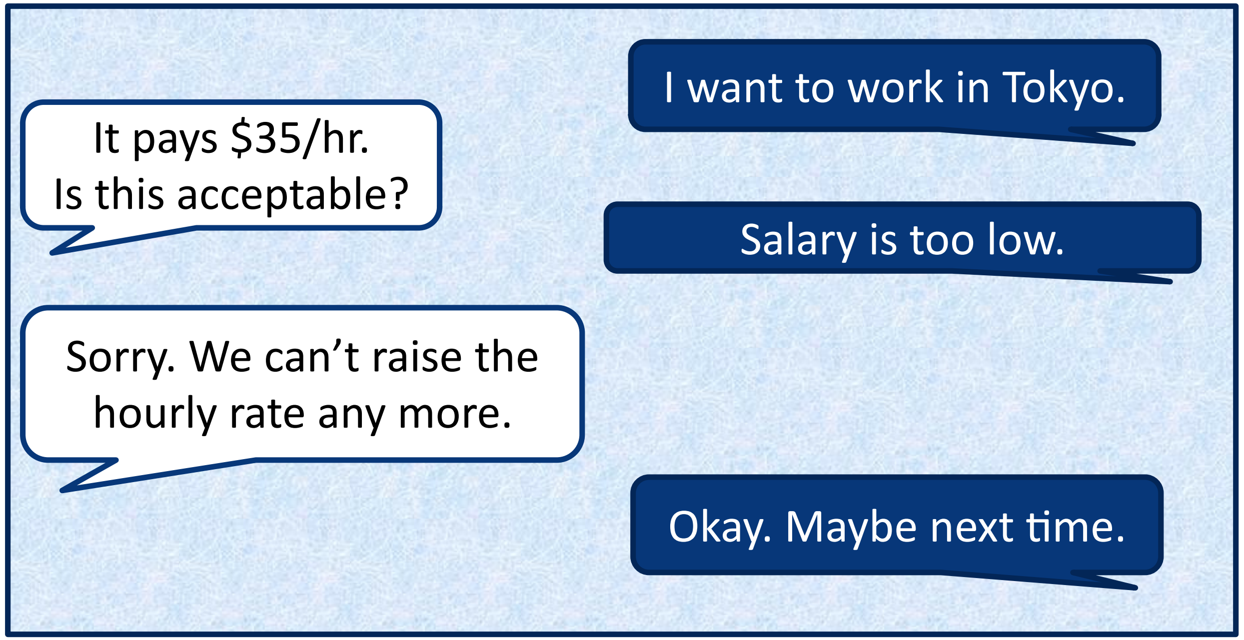
Figure: Example of a breakdown negotiation dialogue in an item division bargaining problem.

Negotiation Breakdown Detection Task (2 / 2)

- **Task:** Given a dialogue composed of n turns' utterances between two negotiators, a model needs to classify whether its outcome is successful or not.



Successful



Negotiation Breakdown

Methodology (1 / 4)

- **Approach**: Dialogue act-based feature with a machine learning model.
  - **Intuition**: a breakdown dialogue should have distinct flow (e.g., many disagreements).
- **Pipeline**:
  1. Extract dialogue acts in an utterance based on the matching patterns.
  2. Align and filter the extracted dialogue acts according to the constrained negotiation flow.
     - To reduce noise due to the rule-based matching.
  3. Concatenate all filtered dialogue acts.

Methodology (2 / 4) - Pattern Matching

- Mostly based on He et al. (2018).
- Match words/phrases in an utterance by using regular expressions.

Dialogue act	Matching Pattern
<greet>	*hi, hello, yo, hey, hiya, howdy, how are you, good day, good afternoon, good morning*
<disagree>	Generic – *isn't, worse, bad, sorry, no, not, nothing, don't, can't, cannot, afraid, a lot lower/higher, too much/high/low* JI – An intermediate offer is rejected.
<agree>	*ok, okay, no problem, yes, great, perfect, thanks, gracias, thx, thank you, pleasure, fine, deal, cool, that works, that will work, that works, it will work, sounds good, very good, looks good, i can do*
<inquire>	*what, where, when, which, how's, how about, how does, do you, did you, will you, would you, could you, are you, do we, did we, could we, do i, let me know, ?*
<propose>	Generic – Any digits, *come down, highest, lowest, go higher/lower, i would like* DN – *ball(s), hat(s), book(s)* JI – A new intermediate offer is proposed.
<inform>	A previous utterance ends with <inquire> and its reply does not contain any other tags.

Table 4: Matching patterns for dialogue acts.

He He, Derek Chen, Anusha Balakrishnan, Percy Liang, "Decoupling Strategy and Generation in Negotiation Dialogues," Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, pp.2333–2343, 2018

Methodology (3 / 4) - Filtering and Alignment

- Align and filter extracted dialogue acts according to the Figure.
  - ➢ To reduce detection noise caused by the rule-based matching.
  - ➢ Inspired by automated negotiation.

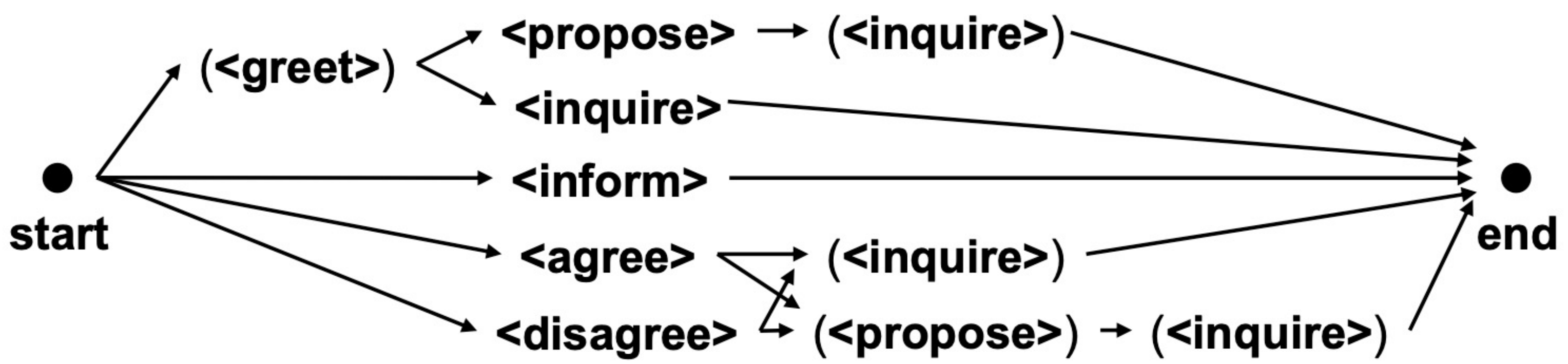


Fig: Constrained flow of each dialogue turn. Dialogue acts in parenthesis represents they do not always appear in the utterance.

Methodology (4 / 4) - Using Dialogue Act-based Features as Inputs for ML-based models

- Concatenate extracted features and create a single sequence.
  - Insert a special separator token:
    - ✓ <sep> into the start of each negotiator’s utterance
    - ✓ <end> into the end of the concatenated sequence.
  - **For Linear Models**: Count # of each dialogue act and weight it by TF-IDF.
  - **For NN-based Models**: Convert each dialogue act into a one-hot representation.

Experimental Settings

- **Data**
  - Use DealOrNoDeal (DN), CraigslistBargain (CB) and JobInterview (JI) datasets.
- **Classification Models**
  - Compare with text-based features.
  - Logistic regression with Bag-of-words (LR-BOW), GRU, GRU with a self-attention (GRU-Att), $BERT_{BASE}$, $BERT_{LARGE}$.
- **Data Pre-processing**
  - Remove short dialogues in the CB and JI datasets as most of them are labelled as a "breakdown". Breakdown ratio: 25%, 18%, 5%
- **Model training and testing**
  - Train & test a model using stratified five-fold cross-validation.
  - Further split training data into training and validation subsets with the ratio of 8:2.

Quantitative Results (Brief Summary)

- **Evaluation Metrics:** ROC-AUC, confusion matrix, Average Precision.
  - Mostly follows our previous work (Yamaguchi and Fujita, 2020).

- GRU-based model trained with dialogue act-based features perform well.
  - **Existing Datasets**:
    - Show similar results over models with textual features.
  - **Proposed Dataset**:
    - Achieve better results.
    - Small agreement ratio in the JI dataset (4.9%).

Ablation Study

- **Importance of Each Dialogue Act**
  - <agree> & <propose> are important
  - <disagree> & <inquire> are less important

- **Roles of Agree and Disagree**
  - <agree> → <diagree>: Rise in TP / Sharp drop in TN.
  - <disagree> → <agree>: Rise in TN / Sharp drop in TP.
  - ✓ The model properly took into account the roles of <agree> and <disagree> to some extent.

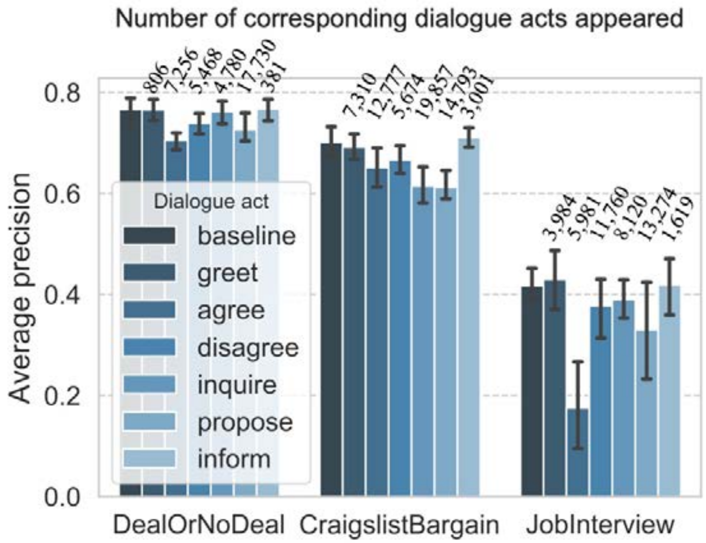


Figure 2: Classification performance comparison on five test folds when replacing a specific dialogue act with an unknown <unk> tag. Error bars denote the 95% confidence interval.

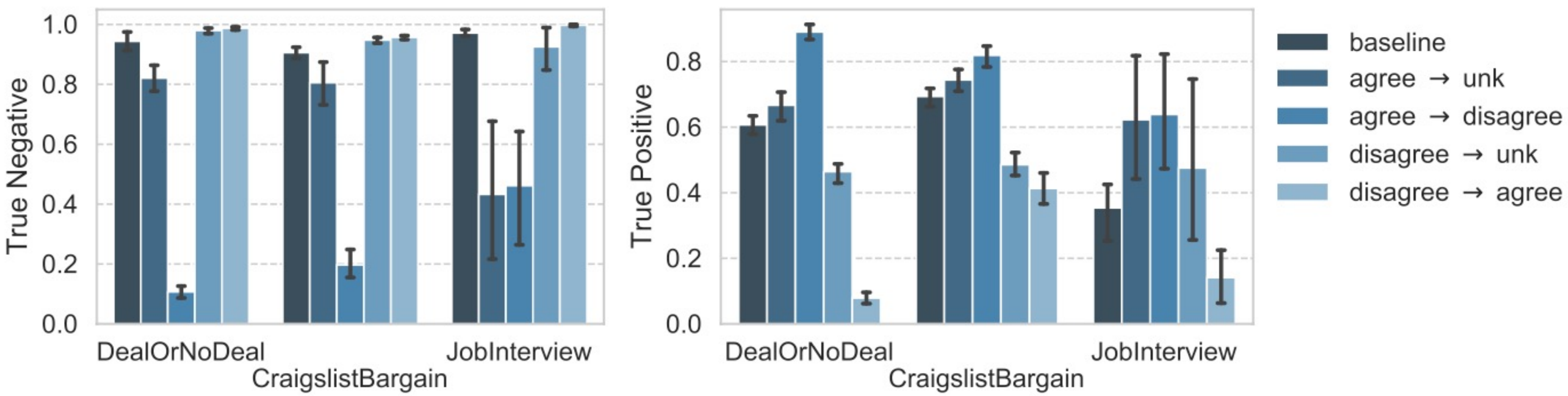


Figure 3: Performance comparison on five test folds when replacing <agree> and <disagree> tags with their counterpart or an <unk> tag. Error bars denote the 95% confidence interval.

Summary

- A job interview dataset with 2639 dialogues and increased complexities.
- A dialogue-act based breakdown detection model.
- GRU-based model with dialogue act-based features achieved comparable results.
- Analysed how dialogue act-based features worked.

Future work

- Explore applications of dialogue act-based features.
- Utilise the proposed corpus in related tasks.