

The Negochat Corpus of Human-agent Negotiation Dialogues

Vasily Konovalov¹, Ron Artstein², Oren Melamud¹, Ido Dagan¹

¹Computer Science Department, Bar-Ilan University, Israel

²USC Institute for Creative Technologies, USA

vaskonov@yahoo.com, artstein@ict.usc.edu, melamuo@cs.biu.ac.il, dagan@cs.biu.ac.il

Abstract

Annotated in-domain corpora are crucial to the successful development of dialogue systems of automated agents, and in particular for developing natural language understanding (NLU) components of such systems. Unfortunately, such important resources are scarce. In this work, we introduce an annotated natural language human-agent dialogue corpus in the negotiation domain. The corpus was collected using Amazon Mechanical Turk following the ‘Wizard-Of-Oz’ approach, where a ‘wizard’ human translates the participants’ natural language utterances in real time into a semantic language. Once dialogue collection was completed, utterances were annotated with intent labels by two independent annotators, achieving high inter-annotator agreement. Our initial experiments with an SVM classifier show that automatically inferring such labels from the utterances is far from trivial. We make our corpus publicly available to serve as an aid in the development of dialogue systems for negotiation agents, and suggest that analogous corpora can be created following our methodology and using our available source code. To the best of our knowledge this is the first publicly available negotiation dialogue corpus.

Keywords: dialogue systems, negotiation corpora, crowdsourcing

1. Introduction

Annotated in-domain corpora are crucial to the successful development of dialogue systems of automated agents, and in particular for developing natural language understanding (NLU) components of such systems (Lasecki et al., 2013). While NLU is considered to be a challenging task, previous studies in the Human Computer Interactions field have found that an important goal of any human-agent dialogue system should be an intuitive interface with interaction environments as real and natural as possible. This highlights the importance of achieving good natural language capabilities for automated agents, which go beyond traditional menu-based interactions (Coen, 1998).

Negotiation is an important task in our daily lives. It concerns anything from a mundane ‘time to go to lunch’, to our salary, and to issues that can have dramatic effects on the lives of millions (Musambachime and Hopmann, 2001). As such, various automated agents have been created to negotiate with people in different settings with varied factors, such as the number of parties, number of interactions and number of issues to be negotiated. (Zuckerman et al., 2013).

In particular, the popularity of the negotiation domain led to the establishment of the international Automated Negotiating Agents Competition (ANAC) Workshop. In this popular competition human-agent dialogues are currently restricted to a fixed set of menu-based interactions (Baarslag et al., 2012), while extension to natural language interactions is on the future roadmap. Since 2010, the competition has focused on agents that are based on the open Genius environment (Lin et al., 2012).

Although natural language annotated corpora are considered an important resource for training and tuning NLU components, to the best of our knowledge there are no such publicly available negotiation corpora.

In this work we address the above mentioned resource scarcity by collecting a corpus of natural language human-agent negotiations, which is compatible with the Genius framework and comprises 105 crowd-sourced dialogues (1484 utterances). Every natural language utterance in this corpus is manually annotated using a formal semantic language, based on both the utterance itself and the specific context in which it appears in the dialogue. Our contributions are as follows. First, we make our corpus publicly available to promote the development of natural language dialogue systems for negotiation agents.¹

Second, we share the methodology and dialogue system source code used to create our corpus and suggest that analogous corpora can be generated following the same approach.²

2. Dialogue System Architecture

Our dialogue system is based on the NCAgent system by (Rosenfeld et al., 2014) with a standard dialog system architecture (Martin and Jurafsky, 2000), which includes the following components: (1) Natural Language Understander (NLU) - translates natural language human utterances from the user side to a formal semantic representation. For example, the utterance *I offer you a salary of 20,000* is translated to the predicate-argument structure, *Offer(Salary=20,000)*. When collecting our corpus the NLU function was carried out by a human. (2) Natural Language Generator (NLG) - translates statements in formal semantic representation from the agent side to natural language utterances. (3) Dialogue Manager (DM) and Agent - the DM is responsible

¹The corpus is available at https://github.com/vaskonov/negochat_corpus

²The components of the dialogue system and the negotiation corpus are available at <https://github.com/vaskonov/>

for maintaining the state and flow of the dialogue, while the Agent in the back-end makes the negotiation decisions. NCAgent, the negotiation agent used for our data collection, was developed specifically to interact with humans, supporting partial agreements and issue-by-issue interactions. It was shown previously that NCAgent is able to reach significantly better agreements, in less time than the prior state-of-the-art KBAgent (Oshrat et al., 2009).

3. Corpus Design And Collection Process

3.1. Overview

In this work we collected a natural language human-agent negotiation corpus in the *job-candidate* domain by recording the interaction of users with the dialogue system described in Section 2. Given our formal semantic language, described below in Section 3.3., the implementation of an NLG component is pretty straightforward. The challenging part is the parsing of natural language utterances into a formal semantic language, which is the responsibility of the NLU component. While we hope that our corpus would help developing automated NLU components, in our data collection setting the function of the NLU was carried out by a human. This was done following the ‘Wizard-Of-Oz’ approach, which is described in more detail in section 3.4. Finally, once we collected the corpus, we employed human annotators to annotate the human utterances in the corpus with their respective formal semantic representations (Section 3.5.).

3.2. Job-Candidate Domain

The *job-candidate* negotiation domain includes bilateral multi-issue closed negotiations and was used in several previous works (Lin et al., 2008; Oshrat et al., 2009). This domain is compatible with Genius³, a general platform for negotiation agents, which was also adapted by the Automated Negotiating Agents Competition (ANAC) as the official competition platform (Lin et al., 2012; Baarslag et al., 2013). The negotiation takes place between an *employer* and a *candidate*. The goal of both sides is to reach a consensus on the hiring conditions or attributes of the agreement, while optimizing their own score objective. If no agreement is reached by the end of the allotted time (30 minutes), both sides receive some predefined score. Each side can also decide to opt-out of the negotiation if one feels that the prospect of reaching an agreement with the opponent is poor and it is useless to negotiate any further. Accordingly, the result of each negotiation was either reaching a full agreement, or failing to do so because of an opt-out or because the allotted time was over. Before starting a negotiation, participants take a tutorial⁴ describing the game interface and the objectives.

The attributes included in our semantic language with their predefined set of values appear in Table 1.

The negotiation dialogue comprises the exchange of one or more of the following dialogue acts in each utterance: *Offer*, *Accept*, *Reject*, *Query*, *Greet*, *Quit*.

3.3. Formal Semantic Language

We used a frame-based semantic representation, where one or more composite semantic labels are used to represent each utterance in a dialogue. Our representation is characterized by a common predicate-argument construction. Each composite label consists of up to three components, intent, attribute and value, where attribute and value are optional. For example, following is a natural language utterance with its corresponding composite label consisting of all three components:

NL utterance: *I offer you a pension of 10%*
Semantic label: *Offer(Pension Fund=10%)*

Alternatively, in this next example the value component is missing:

NL utterance: *I reject the job position*
Semantic label: *Reject(Job position)*

In many cases, such missing values can be implied based on the dialogue context. For example, the value ‘programmer’ can be implied for the above *Reject* utterance if it came as a response to *Offer(Job position = Programmer)*. Resolution of such implied values is done by the Dialogue Manager and not by the NLU component.

Finally, there are several intents that can be used even without attributes, for example:

NL utterance: *Hello, how are you?*
Semantic label: *Greet*

We note that we intentionally chose a simple semantic representation, as it is easier to train NLU models on, especially with relatively small training corpora. However, there is room for future extensions of this representation. Some possible extensions are adding the representation of conditioned offers and ‘OR’ relations. For example, the utterance *Giving you a company car would require you to either work 10 hours or drop your pension to 10%* cannot be expressed in our current semantic language representation.

Our semantic language is formally defined using Synchronous context-free grammar (SCFG) with corresponding natural language translations, which is mostly based on the representation that was used in (Rosenfeld et al., 2014)⁵.

3.4. Crowd-Sourced Data Collection

We recruited 105 English speakers on Amazon Mechanical Turk (AMT) to participate in the negotiations. AMT allows to apply specific filtering criteria for the participants. The only filtering criteria was physical location in the United States, United Kingdom, or Canada.

In our corpus the agent always assumes the role of the candidate while the human is the employer. The dialogues were collected by applying the Wizard-Of-Oz (WOZ) approach. Under this approach, the user believes that she is

³<http://ii.tudelft.nl/genius/>

⁴<https://github.com/vaskonov/tutorial>

⁵The SCFG grammar is available at <https://github.com/vaskonov/scfg/blob/master/grammars/NegotiationCandidateConcise.txt>

Salary	60000 USD, 90000 USD, 120000 USD
Working hours	8 hours, 9 hours, 10 hours
Leased car	With leased car, Without leased car, No agreement
Pension fund	0%, 10%, 20%, No agreement
Promotion possibilities	Fast promotion track, Slow promotion track
Job position	QA, Programmer, Team manager, Project manager

Table 1: The agreement attributes with the corresponding possible values

directly interacting with an agent, but behind the scene, a human who is not a turker, is performing the NLU function, translating the user utterances from natural language to semantic language (Dahlbäck et al., 1993). While this approach requires real-time manual translation of natural language utterances, its merit is in facilitating reliable interactive user-agent dialogues.

The NLG component was entirely automated. First, we generated offline a mapping from all the semantic representations that can potentially be used by the agent into natural language utterances. This was done using Synchronous Context-Free Grammar (SCFG) (Aho and Ullman, 1972), where we defined the SCFG rules for both the natural language and the semantic language.⁶ Then, in real-time, the NLG used a quick look-up in this mapping to translate the agent’s outputs.

3.5. Annotation

Once dialogue collection was completed, two independent annotators annotated the corpus following the annotation guidelines⁷. Disagreements were resolved by the WOZ person as the arbitrator. As a metric of inter-annotator agreement we used Krippendorff’s α with MASI distance that supports multi-label annotation (Passonneau, 2006), as implemented in the DKPro Agreement package (Meyer et al., 2014).⁸

The inter-annotator agreement before reconciliation was 0.89 and after reconciliation became 0.95.

The annotators were instructed to take the context of the dialogue into account while annotating each utterance. For example, the annotation of the following utterance could depend on the context of the dialogue:

OK, let’s move to pension fund

If this utterance is a reply to some *Offer*, then it would most likely be annotated as *Accept*, referring to the previous *Offer*, along with *Query(Offer=Pension Fund)*. However, if

⁶The script used to generate this mapping is available at <https://github.com/vaskonov/scfg/blob/master/scfg.js>. The generated set of all possible utterances, each in both natural and semantic language, is reachable via <https://github.com/vaskonov/scfg/blob/master/NLG.json>

⁷The guidelines are available at https://github.com/vaskonov/negochat_guidelines/blob/master/guidelines.pdf

⁸Two labels are considered to be in agreement only if all of their corresponding components are identical. For example, *Reject(Salary)* and *Reject(Salary=20000)* are considered different labels.

it comes after a *Reject* then *Query(Offer=Pension Fund)* alone would be the more appropriate annotation.

A sample of an annotated dialogue appears in Appendix.

4. Analysis

4.1. Corpus analysis

In total 105 dialogues with 1484 human utterances and 2140 generated agent utterances were collected. These include 1264 single-label human utterances, 110 multi-label ones and 110 non-labeled utterances. The total annotation time was 323 minutes. In 69 out of the 105 dialogues an agreement was reached. Total dialogues duration was 1291 minutes. The maximal dialogue duration was 30 minutes and the minimum dialogue duration was 3 minutes. The participants reported their demographic information in a pre-questionnaire, resulting in 55 male participants and 50 female participants. The participant age distribution appears in Table 2.

Age	Number
18-30	46
31-40	42
41-50	8
51-up	9

Table 2: Turker age distribution

The total expense on crowd-sourcing was a little under \$250, including AMT fees and web hosting of the game application. The post annotation of the utterances cost approximately \$200.

The distributions across the corpus of the intents in the utterances of both humans (in the employer role) and agent (in the candidate role) appears in Table 3. These distributions are naturally sensitive to the strategic policy, configured for our agent. We used the default agent configuration. We note that this is a mixed initiative negotiation and therefore both parties can suggest job conditions. Accordingly, both employer and candidate suggestions were annotated as Offers.

There were 110 human utterances for which the annotators could not assign any of the defined intents. For example, the following utterances were not annotated: “*Welcome to the company!*”, “*Here’s an offer, take a look.*”, “*hold on a minute*”.

4.2. Baseline NLU classification results

An intended goal of the corpus is to train the NLU component of a negotiation dialogue system. As a baseline for

Intent	Human (employer)	Agent (candidate)
Offer	1292	1704
Accept	249	298
Reject	248	542
Query	79	34
Greet	30	105
Quit	18	0
Total	1916	2683

Table 3: The distribution of the intents over the utterances of the human and the agent

future work we followed (Tur et al., 2010), which presents a competitive discriminative classifier method for dialogue systems. To this end, we used a multi-label SVM classifier with unigrams and bigrams as features and tf-idf feature weights. The input to the classifier was the features of the natural language utterance and the output was one or more labels in our formal semantic representation. The first 70 dialogues in the corpus were used as train set and the remaining 35 dialogues as test set. Due to the fact that the labels’ distribution is highly unbalanced we calculated both micro-average and macro-average metrics. Micro-average metrics are calculated globally based on the total true positives, false negatives and false positives for all labels. Macro-average metrics are calculated by first evaluating the individual metrics for each label, and then taking a simple average. The important difference is that the macro-average metrics are not biased towards the most frequent labels. The micro-average results are F1-score of 0.78, Precision of 0.83, and Recall of 0.74. The macro-average results are F1-score of 0.45, Precision of 0.54, and Recall of 0.43.

5. Conclusions and Future Work

In this work we collected a corpus of natural language human-agent negotiations, which is compatible with the Genius framework and comprises 105 crowd-sourced dialogues, including 1484 human natural language utterances. The utterances were annotated with formal semantic language labels, achieving a high inter-annotator agreement of 0.95 Krippendorff’s α .

We make our corpus publicly available to promote the development of natural language dialogue systems for negotiation agents. The rather mediocre baseline results, achieved when training a standard NLU classifier on our corpus, confirm that our corpus presents a challenging task to the research community. Furthermore, we suggest that additional corpora, analogous to ours, can be created easily and quite cheaply following our methodology and using the dialogue system components that are all made publicly available.

6. Acknowledgments

We thank Sarit Kraus, Avi Rosenfeld, Erel Segal-Halevi, Osnat Drein and Inon Zuckerman for their assistance and contribution. This work was partly supported by ERC Grant #267523.

Appendix: A Sample of an annotated dialogue

```
[{ "role": "Candidate",
  "data": "Hello. I would like to discuss the
        issues of my contract.",
  "output": [{ "Greet": true }]
},
{ "role": "Candidate",
  "input": "I would like a position of
        project manager",
  "output": [
    { "Offer": { "Job Description":
                "Project Manager" } }
  ]
},
{ "role": "Employer",
  "input": "Ok! I can offer you a programmer
        position with a 120,000 salary.",
  "output": [
    { "Offer": {"Job Description": "Programmer"}},
    { "Offer": {"Salary": "120,000 USD"}}
  ]
},
{ "role": "Candidate",
  "input": "I agree. Now let's talk about
        how we can agree on other issues.
        I would like a company car",
  "output": [
    { "Accept": true },
    { "Offer": {"Leased Car": "With leased car"}}
  ]
},
{ "role": "Employer",
  "input": "I can give you a company car,
        but your working hours will be 9.",
  "output": [
    { "Offer": {"Leased Car": "With leased car"}},
    { "Offer": {"Working Hours": "9 hours"}}
  ]
},
{ "role": "Candidate",
  "input": "I agree. Let's proceed to another
        issue. I am expecting a fast promotion track",
  "output": [
    { "Accept": true },
    { "Offer": { "Promotion Possibilities":
                "Fast promotion track" } }
  ]
},
{ "role": "Employer",
  "input": "Yes, that sounds good- but is a
        10% pension fun okay?",
  "output": [
    { "Accept": true },
    { "Offer": { "Pension Fund": "10%" } }
  ]
},
{ "role": "Candidate",
  "input": "I reject a pension of 10%.
        I want a pension of 20%",
  "output": [
    { "Reject": { "Pension Fund": "10%" } },
    { "Offer": { "Pension Fund": "20%" } }
  ]
},
{ "role": "Employer",
```

```



```

7. Bibliographical References

- Aho, A. and Ullman, J. (1972). *The theory of parsing, translation, and compiling*. Prentice-Hall, Inc.
- Baarslag, T., Hindriks, K., Jonker, C., Kraus, S., and Lin, R. (2012). The first automated negotiating agents competition (anac 2010). In *New Trends in Agent-Based Complex Automated Negotiations*, pages 113–135. Springer.
- Baarslag, T., Fujita, K., Gerding, E. H., Hindriks, K., Ito, T., Jennings, N. R., Jonker, C., Kraus, S., Lin, R., Robu, V., et al. (2013). Evaluating practical negotiating agents: Results and analysis of the 2011 international competition. volume 198, pages 73–103. Elsevier.
- Coen, M. H. (1998). Design principles for intelligent environments. In *AAAI/IAAI*, pages 547–554.
- Dahlbäck, N., Jönsson, A., and Ahrenberg, L. (1993). Wizard of oz studies: why and how. In *Proceedings of the 1st international conference on Intelligent user interfaces*, pages 193–200. ACM.
- Lasecki, W. S., Kamar, E., and Bohus, D. (2013). Conversations in the crowd: Collecting data for task-oriented dialog learning. In *First AAAI Conference on Human Computation and Crowdsourcing*.
- Lin, R., Kraus, S., Wilkenfeld, J., and Barry, J. (2008). Negotiating with bounded rational agents in environments with incomplete information using an automated agent. *Artificial Intelligence*, 172(6):823–851.
- Lin, R., Kraus, S., Baarslag, T., Tykhonov, D., Hindriks, K., and Jonker, C. M. (2012). Genius: An integrated environment for supporting the design of generic automated negotiators. In *Proceedings of ANAC*. Wiley Online Library.
- Martin, J. H. and Jurafsky, D. (2000). Speech and language processing. *International Edition*.
- Meyer, C. M., Mieskes, M., Stab, C., and Gurevych, I. (2014). Dkpro agreement: An open-source java library for measuring inter-rater agreement. *COLING 2014*, page 105.
- Musambachime, M. and Hopmann, P. T. (2001). The negotiation process and the resolution of international conflicts.
- Oshrat, Y., Lin, R., and Kraus, S. (2009). Facing the challenge of human-agent negotiations via effective general opponent modeling. In *Proceedings of The 8th International Conference on Autonomous Agents and Multi-agent Systems-Volume 1*, pages 377–384. International Foundation for Autonomous Agents and Multiagent Systems.
- Passonneau, R. (2006). Measuring agreement on set-valued items (masi) for semantic and pragmatic annotation.
- Rosenfeld, A., Zuckerman, I., Segal-Halevi, E., Drein, O., and Kraus, S. (2014). NegoChat a chat-based negotiation agent. In *Proceedings of the 2014 international conference on Autonomous agents and multi-agent systems*, pages 525–532. International Foundation for Autonomous Agents and Multiagent Systems.
- Tur, G., Hakkani-Tur, D., and Heck, L. (2010). What is left to be understood in atis? In *Spoken Language Technology Workshop (SLT), 2010 IEEE*, pages 19–24. IEEE.
- Zuckerman, I., Rosenfeld, A., Kraus, S., and Segal-Halevi, E. (2013). Towards automated negotiation agents that use chat interfaces. ANAC.