

# MIDAS: A Dialog Act Annotation Scheme for Open Domain Human Machine Spoken Conversations

**Dian Yu**

University of California, Davis  
dianyu@ucdavis.edu

**Zhou Yu**

Columbia University  
zhouyu@cs.columbia.edu

## Abstract

Dialog act prediction in open-domain conversations is an essential language comprehension task for both dialog system building and discourse analysis. Previous dialog act schemes, such as SWBD-DAMSL, are designed mainly for discourse analysis in human-human conversations. In this paper, we present a dialog act annotation scheme, MIDAS (Machine Interaction Dialog Act Scheme), targeted at open-domain human-machine conversations. MIDAS is designed to assist machines to improve their ability to understand human partners. MIDAS has a hierarchical structure and supports multi-label annotations. We collected and annotated a large open-domain human-machine spoken conversation dataset (consisting of 24K utterances). To validate our scheme, we leveraged transfer learning methods to train a multi-label dialog act prediction model and reached an F1 score of 0.79.<sup>1</sup>

## 1 Introduction

Human-machine conversations have different dynamics compared to human-human conversations due to the power imbalance between humans and machines in conversational settings. Such differences include content and voice quality (Hill et al., 2015). For instance, humans tend to use a more authoritative voice when they talk to a machine. We found that “commands” account for about 9% of utterances in a human-machine social conversation corpus we collected, while a similar dialog act, “Action-Directive”, only accounts for 0.4% in a human-human conversation corpus (Switchboard Dialog Act Corpus (SwDA) (Jurafsky et al., 1997)). Moreover, a human-machine dialog act scheme is not simply used to understand the dialog flow but also to help dialog systems plan their

<sup>1</sup>Code, data, and trained models are available at [https://github.com/DianDYu/MIDAS\\_dialog\\_act](https://github.com/DianDYu/MIDAS_dialog_act)

Scheme	#labels	context	multi-label	dataset
DAMSL	44	✓	✗	✓
TOPIC	14	✓	✗	✗
ISO	88	?	✓	✗
MIDAS	23	✓	✓	✓

Table 1: Comparison of different major dialog act schemes. TOPIC refers to the dialog act designed for topic modeling from Khatri et al. (2018). “?” indicates that guideline on context using is not clear (Mezza et al., 2018)

next steps. Therefore, a dialog act scheme for spoken dialog systems must capture semantic information necessary for fine-grained dialog planning. More importantly, a useful human-machine dialog act scheme should incorporate dynamics of realistic settings such as real-time automatic speech recognition (ASR) outputs. However, previously available dialog act schemes are all designed for manual dialog transcriptions, which are very different from ASR outputs. For instance, ASR outputs are noisy and may not contain punctuation. To test if a dialog act predictor trained on manual transcriptions would generalize to ASR outputs, we trained a model using the SwDA dataset annotated with the SWBD-DAMSL scheme (Jurafsky et al., 1997). We tested the model on the ASR output of our open-domain human-machine dialog system conversations and found that the model’s performance is only 47.38% in accuracy, mostly due to incompatible schemes and confusion in annotation in addition to domain shifts. This low accuracy score suggests that using existing schemes and datasets to train models for human-machine dialog systems is impractical.

In this paper, we propose a hierarchical multi-label dialog act annotation scheme, MIDAS, specifically designed for real-time open-domain human-

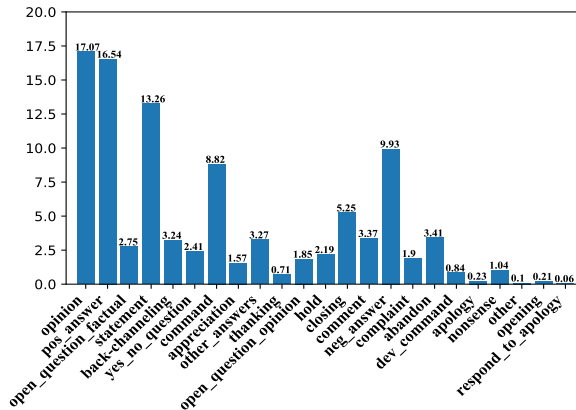


Figure 1: Distribution of dialog acts in the training set

machine spoken conversations. We show a comparison to major dialog act schemes in Table 1. We also annotated real-world human-machine social conversations using the MIDAS scheme and Figure 1 shows the distribution of dialog acts. The scheme is easy for humans to follow. Two annotators achieved an inter-annotator agreement of  $\kappa = 0.94$ . We trained a multi-label dialog act classifier using transfer learning methods and reached an F1 score of 0.79. Multiple Amazon Alexa Prize social chatbots have deployed our dialog act model and reported better conversational quality due to improved language understanding compared to using SWBD-DAMSL. We share our annotated data and trained models with the research community for easy adoption.

## 2 Related Work

Previous dialog act annotation schemes are mostly designed for task-oriented dialogs with a specific task, such as MapTask (Thompson et al., 1993) and DATE (Walker and Passonneau, 2001), or in a specific setting (Kl uwer, 2011). There are a few dialog act schemes designed for human-human task-independent conversations, such as the Discourse Annotation and Markup System of Labeling (DAMSL, Core and Allen, 1997) and SWBD-DAMSL (Jurafsky et al., 1997). SWBD-DAMSL is used to annotate the Switchboard (Godfrey et al., 1992) corpus, a task-independent telephone conversation corpus, with inter-annotator agreement of  $\kappa = 0.80$ . SWBD-DAMSL is also applied to annotate human-human meeting conversations (Shriberg et al., 2004). In this paper, we design a dialog annotation scheme specifically for human-machine social chitchat conversations without any

topic constraints.

Khatri et al. (2018) introduces a human-machine dialog act annotation scheme with 14 tags. However, the scheme is designed for modeling conversation topics instead of training dialog act predictors. The scheme has tags such as *Information Request*, *General Chat*, and *Multiple Goals*, and the annotation is performed on unsegmented user utterances. Even though the small number of tag categories makes annotation more reliable, it may not provide enough information for understanding user semantics. For example, tags such as *Multiple Goals* do not provide explicit information about user intent. In contrast, we propose a dialog act annotation scheme that focuses on improving open-domain dialog system understanding. We also build a dialog act predictor based on the annotated corpus.

Previously, most popular annotation schemes, such as DAMSL, use mutually-exclusive tags (Mezza et al., 2018) (0.01% of the labeled utterances have multiple labels so we do not consider DAMSL as multi-label) and each utterance is labeled with a single tag in SWBD-DAMSL (Stolcke et al., 2000). However, Bunt (2009) argues that conversation utterances are complex. Each functional segment can have four to five functions on average, so dialog act tags should serve multiple functions. Dynamic Interpretation Theory (DIT) (Bunt, 1997) and its extension, DIT++ (Bunt, 2009) try to solve this problem by supporting multi-dimension and multi-function tags. The 88 tags are organized in a hierarchical structure and separated into dimension-specific and general-purpose functions. The fifth version of DIT++, ISO 24617-2 (ISO standard, Bunt et al., 2010, 2017), is introduced to incorporate not only linguistic theory but also empirical discourse analysis on real domain-independent conversations. Although much effort has been expended in designing ISO, no large dataset was annotated with the scheme except DialogBank (Bunt et al., 2016). This is probably due to the complexity of the scheme and the lack of clear guidelines on how to use contextual information (Ribeiro et al., 2015; Mezza et al., 2018). Due to the huge complexity of open domain social conversations, we propose a hierarchical structure in the annotation scheme and allow one utterance to have multiple dialog acts. We limit the number of dialog acts to 23 to capture the fundamental acts necessary to the system while keeping the total number relatively small (two annotators reached 0.94 in Kappa). We

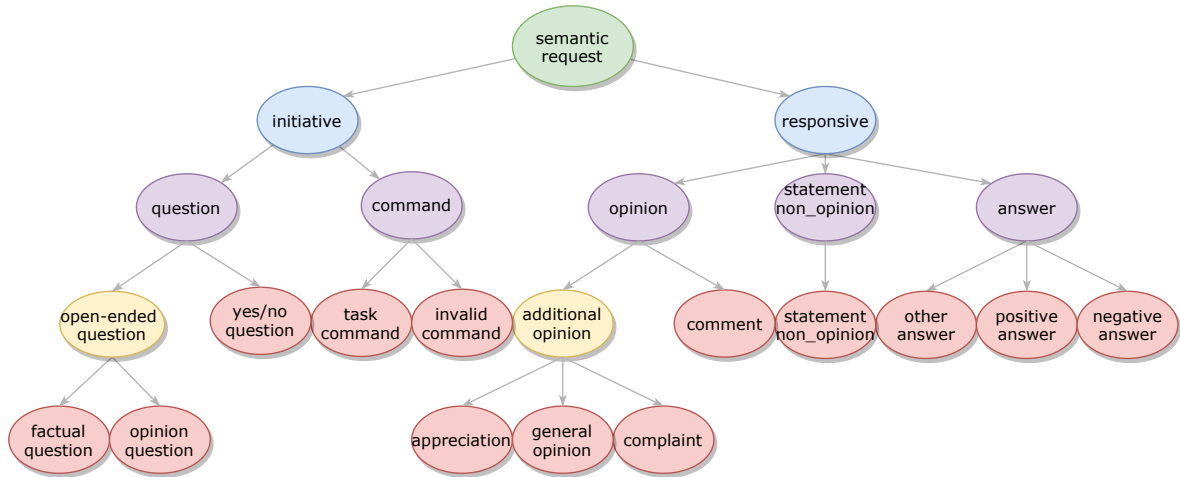


Figure 2: Semantic request tree. Scheme types, classes, categories, and sub-categories are in green, blue, purple, and yellow, respectively. Dialog act tags are leaf nodes in red. Tags can co-occur in one utterance, except tags under *opinion* and *statement non\_opinion*, *question* and *answer* categories due to semantic and syntactic conflicts. For example, “User1:Do you watch TV shows? User2: I prefer watching movies.” User2 is labeled both *general opinion* and *negative answer*.

also publish our annotated human-machine chatbot corpus through user studies containing 24,000 utterances.

### 3 MIDAS Annotation Scheme

We present MIDAS, a contextual hierarchical multi-label dialog act annotation scheme for human-machine conversations. MIDAS follows DIT++ and ISO (Bunt, 2009; Bunt et al., 2010) to ensure that the scheme facilitates both annotation and the training of automatic dialog act predictors. MIDAS focuses on helping dialog systems understand their human users, while previous schemes mainly focus on analyzing human-human dialog with not fully open-domain data (SWBD-DAMSL), mutually-exclusive tags (DAMSL), or lack of contextual information (DIT++ and ISO). Therefore, MIDAS provides a unique hierarchical structure and proposes a set of new labels for the human-machine setting while inheriting labels from previous schemes. A complete description of MIDAS is in Appendix A.2. Similar to Chowdhury et al. (2016); Mezza et al. (2018), we also show a potential mapping from SWBD, SWBD-DAMSL, and ISO to MIDAS in Appendix A.4. We discuss the three main features of MIDAS: hierarchical structure, multi-label format, and context consideration respectively.

#### 3.1 Hierarchical structure

Previous schemes such as ISO design their hierarchical structure as multiple dimensions and define

dialog acts as dimension-specific functions. Such a turn-by-turn specific taxonomy mixes in-depth analysis into the discourse level, thereby requiring a distinguished definition in each dimension and creating a complex hierarchy. For instance, “accept” and “decline” are defined individually for corresponding parent dimensions such as “address offer” and “address suggestion”. This detailed design is beneficial to study interlocutors in a conversations, but may not be necessary for language understanding in a dialog system. In comparison, MIDAS focuses on a concise hierarchical structure to mainly distinguish semantic and functional intents which are critical for dialog understanding and planning. This design facilitates both annotation and model prediction in human-machine conversations.

Specifically, we design MIDAS to have a tree structure. It has two sub-trees: *semantic request* type (Figure 2) and *functional request* type (Figure 3). Under each type, there are classes, categories, and tags arranged in a hierarchical tree structure. Please refer to Figure 2 and 3 for detailed organization. Utterances are labeled with dialog act tags, which are the leaf nodes. The non-leaf nodes are used to categorize different dialog acts, help annotators find the correct tags, and assist customized requirements. We explain the reasoning for the hierarchical design and provide justification for the dialog act definitions with use cases.

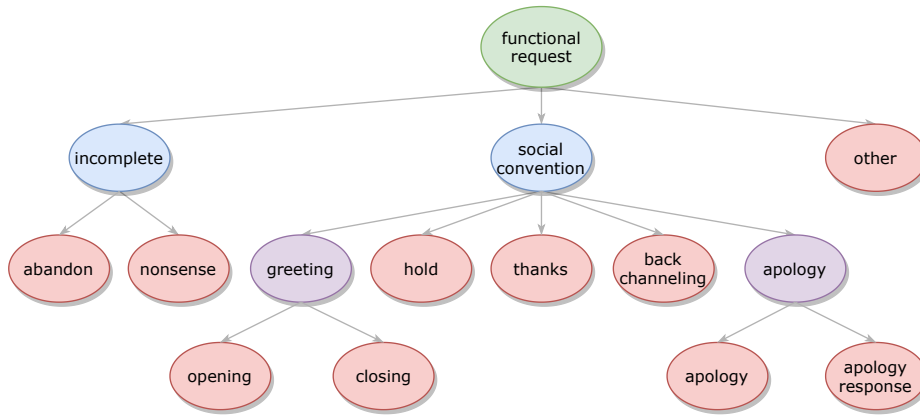


Figure 3: Functional request tree. We remove class and category nodes if there is only one label under them.

### 3.1.1 Semantic request

Semantic request type captures dialog content, therefore it is essential for dialog topic planning. *Semantic request* separates into *initiative* class and *responsive* class based on whether the user is proposing or continuing a topic.

**Initiative class** is especially important in the human-machine setting, because in such an imbalanced power setting, the machine must follow the topic that its human partner proposes. Therefore, understanding whether the user is proposing a new topic with their specific intent is the first step for the system to be coherent. There are two categories, *question* and *command* in the *initiative* class, that are designed to distinguish information requests (*question*) from action requests (*command*).

MIDAS first separates *question* into *yes/no question* and *open-ended question* based on syntax. This separation helps the system to generate a coherent response. For example, it is more natural for system responses to start with “yes” or “no” when replying to a *yes/no question*. Then MIDAS further divides *open-ended question* into *factual question* and *opinion question* based on the different types of information that users seek. The system can thus leverage this information and prepare responses by searching different knowledge bases. For example, *factual questions* require factual information from knowledge graphs such as Wikipedia, while *opinion question* requires information from corpora with opinionated material such as Twitter.

Unlike *question*, *command* conveys orders and is particularly popular in human-machine dialogs. The system needs to follow users’ commands, both implicit and explicit, because the system has less power in the conversation than its human interlocutors. Therefore, unlike previous schemes, MIDAS

uses *task command* for task-oriented and device related requests, and introduces *invalid command*, which is specific to smart devices. Users sometimes produce commands that are beyond the system’s capability. For example, users may want to control device hardware to which the dialog system does not have access. The system needs to identify these utterances and handle them separately. Utterances labeled with *invalid command* are commands involving device functions which are specific to tasks and can be replied by templates and APIs. We present this tag in the leaf level to be consistent with other dialog acts in prediction.

**Responsive class** indicates that the utterance is a continuation of the previous topic. SWBD-DAMSL points out that opinions are often followed by general opinions, whereas statements are followed by back-channels (Jurafsky et al., 1997). This distinction may be subtle in human-human conversations (Jurafsky et al., 1997) as humans do not need to explicitly distinguish between the two tags to generate corresponding responses. However, knowing whether an utterance is a statement or an opinion is essential for a system to generate an appropriate response. Hence, MIDAS further breaks the responsive class into *opinion*, *statement non\_opinion*, and *answer*, based on the conversation history.

MIDAS separates the *opinion* category into the *additional opinion* subcategory and the *comment* tag because we observed examples such as “User1: my friend thinks we are living in matrix. User2: she’s probably right”. User2 comments on the previous utterance without contributing extra information. *additional opinion* indicates that utterances labeled as dialog acts under this subcategory may contribute extra information, whereas *comment* of-

ten indicates that an utterance is a simple reply without explicit feedback. MIDAS separates three types of opinions: *appreciation*, *complaint*, and *general opinion*. Understanding whether the user is complaining or praising the system is essential. Such information is critical to influence dialog policies and extract user feedback.

We also split *answer* into *positive answer*, *negative answer*, and *other answer*, based on utterance's sentiment. One caveat is that utterances, such as "why not", contain negative words but are actually a positive answer to questions such as "Can we talk about movies?", and thus should be labeled as *positive answer*. Such phenomena suggest that automatic dialog act prediction models require semantic understanding and incorporate context in feature representation.

### 3.1.2 Functional request

*Functional request* type helps dialog systems achieve discourse level coherence and control conversation functions. We define *incomplete*, *social convention*, and *other* classes under the *functional request* type.

**Incomplete class** describes utterances that are not complete. There are two *incomplete* types, *abandon* and *nonsense*. In real-world settings, human users can be cut off due to issues such as background noise and long pauses. These cases are labeled as *abandon*. By comparison, *nonsense* is used to label utterances that human annotators cannot understand. These utterances usually contain many ASR errors.

**Social convention class** is similar to the *social obligations management* and *discourse structure management* dimensions in ISO (Bunt et al., 2010). We define *opening*, *closing*, *thanks*, *apology*, *apology response*, *hold*, and *back channeling* to provide discourse level information.

Finally, utterances that cannot be assigned to any other tag in this hierarchical structure are labeled as an *other* tag.

## 3.2 Multi-label support

Compared to single-label schemes, multi-label schemes capture different dimensions and functions, which better support dialog system building and discourse analysis. For example,

User1: *what books have you read recently*

User2A: *i haven't read any*

User2B: *i don't want to talk about books*

User2C: *i prefer watching movies*

Users may use different strategies to express a *negative answer* intent. A single-label scheme cannot differentiate above three sentences. An extra label that captures additional semantic information besides *negative answer* benefits dialog system understanding. For instance, User2B has the additional *task command* intent that requests to end the current topic compared to User2A. User2C has the *general opinion* intent that initiates a different topic. Dialog system may not need to change a topic for User2A, but it needs to change the topic for User2B and User2C.

SWBD-DAMSL allows an utterance to be tagged as *double labels* and lists the preferred tag first (Jurafsky et al., 1997). However, *double labels* and ordering heuristics are not explicitly defined in the scheme and thus are infrequent in Switchboard. In MIDAS, except for two exclusive category pairs (*opinion* and *statement non\_opinion*, *question* and *answer*), dialog acts are designed to be compatible across hierarchies so that labels in each category can co-occur in one utterance.

Due to the power difference in human-machine conversations, the system needs to prioritize certain tags over the others to keep a coherent conversation. MIDAS has a priority list to focus, from high to low: *question*, *command*, *answer*, *opinion*, and *statement non\_opinion*. For example, "User1: what do you want to talk about? User2: how about the financial market". User2's utterance can be tagged as *task command*, *opinion question*, and *general opinion*. Among the three tags, *opinion question* and *task command* are more useful for the system to direct the conversation towards a specific topic.

## 3.3 Context consideration

Mezza et al. (2018) suggests that most dialog act schemes, including ISO, are not clear on how to leverage contextual information in the annotation process. So tags such as "Answers" are confused with "Inform", which leads to noisy annotations. In comparison, MIDAS scheme is designed to consider context utilizing the hierarchical structure. Dialog acts are thus explicitly distinguished by contextual information. For instance, *general opinion* and *comment* both refer to personal views, but they belong to different sub-categories capturing different semantics and expectations from interlocutors given a specific context. Correspondingly, during

annotation, annotators are instructed to consider the context from previous two turns and locate each level in the structure tree suggested by the context.

#### 4 Dataset and Annotation Process

We collected 380K human-machine conversations ASR outputs using Gunrock, the 2018 Alexa Prize winning social bot (Chen et al., 2018). The average number of turns in a conversation and the average tokens per user utterance is 21.76 and 2.85, respectively. Table 2 shows a sample conversation. Two annotators read the descriptions and examples of each dialog act illustrated in Appendix A.2. The annotators reached an inter-annotator agreement of  $\kappa = 0.94$  on 1,185 segmented utterances verified by the scheme designers (compared to  $\kappa = 0.80$  using DAMSL reported by Stolcke et al. (2000)). Then they annotated the rest of the data separately. In total, they annotated 468 conversations, including 24K segmented sentences on both user and system, among which 12.9K segments are from users. Note that if one dialog act spans multiple segments, we annotate each segment with the same dialog act tag. *general opinion* and *statement non\_opinion* are the most frequent tags. For multiple labels, (*positive\_answer*, *command*) and (*negative\_answer*, *command*) are the most frequent co-occurring tags. Detailed annotation results, including dialog act tag distribution, are shown in Appendix A.2 and Appendix A.3.

**USER1:** what do you want to talk about  
**USER2:** what can you tell me what the top books are right now  
**USER1:** i am so excited to talk to you about books. i'm actually a pretty big bookworm, and i love to read when i'm not chatting  
**USER2:** oh [SEG] what are some titles of the books you've read  
**USER1:** recently, i'm reading the great gastby. it's really thought provoking, and i can see why some people call it the great american novel. how about you? what book do you like?  
**USER2:** i have {not} read a book in a while [SEG] do you have recommendations in the sci fi

Table 2: An example conversation between a machine (USER1) and a human (USER2). The word “not” is dropped in the last sentence due to ASR errors.

We note that there may be some limitations in the corpus we annotated in terms of both ASR results, and the current interactions between humans and machines. Our motivation is that compared to carefully corrected transcriptions such as SwDA, our scheme and annotated corpus can be more ro-

bust against ASR errors and noisy inputs. Since the proportion of utterances with ASR errors are relatively small in our corpus, models trained with the corpus should still perform well on utterances without such errors from more advanced ASR systems. In addition, we believe that our hierarchical multi-label schemes with a manageable size and structure can be easily modified to customized models, as well as more advanced human-machine conversational models. Moreover, our scheme is designed to bridge the gap in human and machine interactions, and are not restricted by the annotation corpus. More importantly, this scheme and annotated corpus can be extended outside of predictions and understanding tasks to language generation tasks. We leave detailed comparison and application to additional tasks such as language generation to future work.

For dialog act prediction, user utterances in human-machine dialogs are ASR outputs and have no punctuation. Therefore, we train a model to segment utterances into semantic units for pre-processing. Note that utterance segmentation is a different line of research and is out of the scope of this paper. We focus on dialog act prediction on each segmented unit following Stolcke et al. (2000). Previous research detects sentence boundaries by predicting the exact punctuation in the training dataset (Cho et al., 2015). However, correct punctuation also relies on deep semantic understanding beyond the sentence surface forms. For instance, a misused question mark can lead the dialog act model to predict a sentence as a question. So following Favre et al. (2008), we only predict the boundary of a sentence instead of predicting punctuation to avoid introducing errors.

It is expensive to annotate sentence boundaries, so we use the Cornell Movie-Quotes Corpus (Danescu-Niculescu-Mizil and Lee, 2011) to train a sentence segmentation model. The Cornell dataset contains 300K utterances from movie transcripts. We reformat the transcripts by replacing punctuation to sentence breaker tokens (denoted as [SEG]). We then train a sequence-to-sequence (seq2seq) model (Sutskever et al., 2014) to predict sentence breaker similarly to Klejch et al. (2017) and Peitz et al. (2011). The input to the model is a reformatted sentence, and the output is the same sentence with added sentence breaker tokens. An example can be seen in the last USER2 utterance in Table 2. Word embeddings are pre-trained with fastText

(Mikolov et al., 2018) using Common Crawl. We evaluate the segmentation model on 2k manually labeled human utterances of the collected data. The segmentation model achieves 84.43% in micro F1 score, 84.97% in precision, and 84.57% in recall. We apply the trained segmentation model on the entire collected dataset to obtain segmented sentences. All the dialog act annotation and predictions are done on the automatic segmentation results.

## 5 Dialog Act Prediction

We formulate the dialog act prediction problem as a multi-label classification problem. Following Katakis et al. (2008), we evaluate the proposed methods using F1 score. For evaluation simplicity, we predict one or two labels in the following experiments. We leverage both unlabeled data and annotated data to improve classification performance.

### 5.1 Baseline model

RNN models have shown promising results in text classification (Rojas-Barahona et al., 2016). Our baseline model uses a 2-layer Bi-LSTM to encode the context representation and a multi-layer perceptron (MLP) to decode the output. For multi-label prediction, we use a binary cross-entropy objective function to learn co-occurring tags independently. This training objective also helps with transfer learning from other single label dialog act datasets. During testing, we choose the labels with the highest values predicted from the MLP as the potential output and filter them with an empirical threshold (0.5).

### 5.2 Context representation

Contextual information plays an important role in dialog act prediction (Liu et al., 2017; Khatri et al., 2018). We consider two methods to represent previous turns: the actual utterance (text), and the dialog act of the utterance (DA). For each method, the most recent segmented sentence unit from each speaking party is considered as the history suggested by the length of context from Khatri et al. (2018). We append the last segmented system unit (sys\_unit), the previous segmented user unit (user\_prev), and the current segmented user unit (user\_cur) as *sys\_unit*  $\langle u_p \rangle$  *user\_prev*  $\langle u_c \rangle$  *user\_cur* where  $\langle u_p \rangle$  and  $\langle u_c \rangle$  are special tokens to separate utterances. For instance, to predict the dialog act for the segment “do you have recommendations in the sci fi” in the last

*USER2* utterance in Table 2, the context representation is formed as *what book do you like*  $\langle u_p \rangle$  *i haven’t read a book in a while*  $\langle u_c \rangle$  *do you have recommendations in the sci fi*. However, if the current utterance is the first one in the current turn, i.e. there is no contextual information, we use an empty token for *usr\_prev* instead.

Another method to incorporate history is to replace the actual previous segment unit with its dialog act labels (if there are two labels for one segment, we combine both labels). The results for these two methods are shown in Table 3.

### 5.3 Transfer learning

We use an unsupervised domain adaptation task and a supervised dialog act predictor trained on SwDA (Jurafsky et al., 1997) to improve model performance. We used a BERT model trained on Wikipedia (Devlin et al., 2019) to leverage its contextual word embeddings. However, one potential drawback of using pre-trained BERT is that it has a different domain from our conversational data. Inspired by Siddhant et al. (2019), we use 50 million unlabeled segmented utterances from 380K conversations from Gunrock to fine-tune the BERT language model to resolve this issue. We also leverage an annotated dataset from a similar task. We automatically map 42 tags from SWBD-DMSL to our 23 tags. The detailed mapping can be found in Appendix A.4. We remove all the punctuation (except apostrophes) and non-verbal information such as “ $\langle \text{laugh} \rangle$ ” from the carefully annotated dataset. We also remove sentences with dialog acts that are not applicable to ours such as *3rd-party-talk*. After pre-processing, we extract a total of 200K annotated utterances using the context representation in Section 5.2. We train a single-label prediction model based on BERT before fine-tuning it on our multi-label prediction task.

## 6 Experiments

**Setting** The main purpose of a dialog act predictor in human-machine dialogs is to improve the dialog system’s understanding of user intent and preparing corresponding responses. Therefore, we build a dialog act prediction model on user utterances only. After pre-processing (refer to Section 5.2), there are 12.9K user segments, 13.78% of which have multiple labels. We use 10.3K for training and 2.6K for testing. All the testing data are from in-lab user studies. In order to protect Amazon Alexa

	Pre(%)	Rec(%)	F1(%)
LSTM-text	75.94	75.91	75.51
LSTM-DA	75.83	73.48	73.77
BERT-text	79.57	79.31	79.11
BERT-DA	79.29	76.12	76.87
BERT-no_context	73.88	70.43	71.30
BERT-DA+text	79.79	79.47	79.28
BERT_F-text	79.83	79.64	79.40
BERT_F-DA	79.30	76.15	76.89
BERT_F-DA+text	<b>79.93</b>	79.61	<b>79.44</b>
BERT-SwDA	79.26	76.43	78.98
BERT-SwDA_F	79.58	<b>79.76</b>	79.28

Table 3: BERT\_F-DA+text achieves the best precision and F1 score. Results reported are an average score of six different random seed runs.

user’s privacy, we only make the 2.6K annotated testing utterances public as they were collected with consent for releasing the data to the public.

**Models.** We implemented 11 models. We use **LSTM** to represent the baseline model trained with LSTMs. **BERT** represents Transformer models with a pre-trained BERT language model. Based on different transfer learning methods described in Section 5.3, **BERT\_F** is a pre-trained BERT language model fine-tuned on unlabeled in-domain data, whereas **BERT\_SwDA** is a pre-trained BERT language model fine-tuned on labeled SwDA task. Combining these two methods, **BERT-SwDA\_F** is fine-tuned on both the unlabeled and labeled tasks. After fine-tuning, the models are trained on our data annotated with the MIDAS scheme. To evaluate the impact of context representation for the above models, we use **-text** and **-DA** to represent using text and dialog act as the context, respectively. We denote **-no\_context** when predicting the current utterance without any context.

See Appendix A.1 for implementation details.

## 7 Results and Analysis

Table 3 describes the experimental results on all 11 models. Transformer models (Vaswani et al., 2017) using BERT embeddings (BERT-text) outperform Bi-LSTM models with pre-trained word embeddings (LSTM-text) by a large margin (from 75.51% to 79.11% in F1). If we further fine tune the BERT language model on an unsupervised training task with similar data distribution (BERT\_F-text), the classification result further improves from 79.11%

to 79.40% in F1. This is consistent with previous research on in-domain pre-training (Siddhant et al., 2019). However, the performance improvement is not statistically significant. One possible reason is that models pre-trained on a very large text dataset, such as Wikipedia, already encode sufficient semantics for dialog act prediction. Therefore, fine-tuning the model on a more domain-aligned data set does not significantly improve the performance.

We found that incorporating context improves the model performance. Adding text information as context improves the BERT model from 71.30% to 79.11% in F1. We also compare the impact of different context embedding methods on dialog act classification performance. The results show that replacing text with dialog act achieves a high precision, but suffers from a low recall. This is because an utterance can have multiple intents while dialog act itself does not provide enough context information to achieve accurate prediction. For example, when “i don’t think so” is a response to a simple *yes/no question* such as “have you read the book”, it is a *negative\_answer*. But if it is a response to a more complex *yes/no question*, such as “do you want to talk about books”, then it has two tags, *command* and *negative\_answer*. The latter conveys user’s implicit request on changing the topic. Therefore, only using dialog act as context could lead to high recall but low F1. We found combining both the previous segment’s dialog act label and its surface text achieves the best performance in F1 (79.44%). However the performance improvement over including text only is not statistically significant. This suggests that dialog act and text may contain more overlapping information than complementary information.

We also found that fine-tuning the model using the supervised dialog act prediction task on the SwDA data did not improve performance in F1 but improved recall slightly. The reduced performance may be due to difference in the data. Even though both datasets are open domain conversational data, the SwDA task uses pre-processed Switchboard data that does not have ASR errors. Moreover, SwDA is human-human conversations, and they are more coherent and consistent compared to human-machine conversations. Another reason is that SwDA dataset has exactly one label for each utterance. When fine-tuning on our multi-label task, the pre-trained single-label model may tend to predict more labels to quickly reduce loss



but fail to learn better representations.

We further looked into the errors made by our best model (*BERT\_F-DA+text*) and found that the model confuses *statement non\_opinion* and *general opinion*. This is most likely caused by only including one turn context. Sometimes, users have questions that break the conversation flow, such as “can you say it again clearly”. The model needs to consider not only this utterance but also the prior turns to perform dialog act prediction. We plan to incorporate longer context in future work. In addition, some of the *nonsense* sentences are misclassified as *statement non\_opinion* such as “it doesn’t outside break a car”. It is also worth noting that some incorrectly segmented units resulted in inaccurate dialog act prediction. For instance, the utterance “we are they love each other” is incorrectly segmented into “we” and “are they love each other” given the context “you must be great friends with your dogs”. The second segment is incorrectly predicted as a *yes/no question*.

## 8 Conclusion

We propose MIDAS, a dialog act scheme designed for open-domain human-machine conversational systems. MIDAS is a hierarchical annotation scheme that supports multiple labels. We annotated 24K sentences from a human-machine social conversation dataset using MIDAS. We also trained dialog act classification models based on the annotated dataset. We tested different transfer learning techniques to improve model performance. We found that fine-tuning a pre-trained BERT model on unannotated target human-machine conversations improved model performance. However, fine-tuning the model on a supervised dialog act task with human-human conversations did not improve model performance. MIDAS has been deployed in real world dialog systems, demonstrating its efficacy in the open-domain dialog setting.

Beyond prediction in open-domain dialogs for human request understanding, our proposed scheme can also be applied to task-oriented dialogs together with chatbots. Furthermore, MIDAS can be used in end-to-end dialog systems for cases such as probing and model interpretation. In addition, our scheme and method can be extended to controlled text generation. We will explore broader applications in future work.

## 9 Acknowledgement

We thank Kenji Sagae for valuable discussions and suggestions. We also thank Sam Davidson and anonymous reviewers for their helpful comments. Dian Yu is supported by the National Science Foundation under Grant No. 1840191. Any opinions, findings, and conclusions or recommendations expressed are those of the authors and do not necessarily reflect the views of the NSF.

## References

- Harry Bunt. 1997. Dynamic interpretation and dialogue theory.
- Harry Bunt. 2009. The dit++ taxonomy for functional dialogue markup. In *AAMAS 2009 Workshop, Towards a Standard Markup Language for Embodied Dialogue Acts*, pages 13–24.
- Harry Bunt, Jan Alexandersson, Jean Carletta, Jae-Woong Choe, Alex Chengyu Fang, Koiti Hasida, Kiyong Lee, Volha Petukhova, Andrei Popescu-Belis, Laurent Romary, Claudia Soria, and David Traum. 2010. [Towards an ISO standard for dialogue act annotation](#). In *Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC’10)*, Valletta, Malta. European Languages Resources Association (ELRA).
- Harry Bunt, Volha Petukhova, and Alex Fang. 2017. Revisiting the iso standard for dialogue act annotation. In *Proceedings 13th Joint ISO - ACL Workshop on Interoperable Semantic Annotation (isa-11)*, pages 37–50. Research Unit(s) information for this publication is provided by the author(s) concerned.
- Harry Bunt, Volha Petukhova, Andrei Malchanau, Kars Wijnhoven, and Alex Fang. 2016. [The DialogBank](#). In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*, pages 3151–3158, Portorož, Slovenia. European Language Resources Association (ELRA).
- Chun-Yen Chen, Dian Yu, Weiming Wen, Yi Mang Yang, Jiaping Zhang, Mingyang Zhou, Kevin Jesse, Austin Chau, Antara Bhowmick, Shreenath Iyer, Girithaja Sreenivasulu, Runxiang Cheng, Ashwin Bhandare, and Zhou Yu. 2018. Gunrock: Building a human-like social bot by leveraging large scale real user data. In *2nd Proceedings of Alexa Prize*.
- Eunah Cho, Kevin Kilgour, Jan Niehues, and Alexander H. Waibel. 2015. Combination of nn and crf models for joint detection of punctuation and disfluencies. In *INTERSPEECH*.
- Shammur Absar Chowdhury, Evgeny Stepanov, and Giuseppe Riccardi. 2016. [Transfer of corpus-specific dialogue act annotation to ISO standard: Is it worth it?](#) In *Proceedings of the Tenth International Conference on Language Resources and*

- Evaluation (LREC 2016)*, pages 132–135, Portorož, Slovenia. European Language Resources Association (ELRA).
- 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*, pages 28–35.
- Cristian Danescu-Niculescu-Mizil and Lillian Lee. 2011. Chameleons in imagined conversations: A new approach to understanding coordination of linguistic style in dialogs. In *Proceedings of the Workshop on Cognitive Modeling and Computational Linguistics, ACL 2011*.
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. [BERT: Pre-training of deep bidirectional transformers for language understanding](#). In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.
- B. Favre, D. Hakkani-Tur, S. Petrov, and D. Klein. 2008. Efficient sentence segmentation using syntactic features. In *2008 IEEE Spoken Language Technology Workshop*, pages 77–80.
- John J. Godfrey, Edward C. Holliman, and Jane McDaniel. 1992. Switchboard: telephone speech corpus for research and development. In *[Proceedings] ICASSP-92: 1992 IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 1, pages 517–520 vol.1.
- Jennifer Hill, W. Randolph Ford, and Ingrid G. Farreras. 2015. [Real conversations with artificial intelligence: A comparison between human–human online conversations and human–chatbot conversations](#). *Computers in Human Behavior*, 49:245 – 250.
- Dan Jurafsky, Liz Shriberg, and Debra Biasca. 1997. Switchboard SWBD-DAMSL shallow-discourse-function annotation coders manual. Technical Report Draft 13, University of Colorado, Institute of Cognitive Science.
- Ioannis Katakis, Grigorios Tsoumakas, and Ioannis Vlahavas. 2008. Multilabel text classification for automated tag suggestion. In *Proceedings of the ECML/PKDD 2008 Discovery Challenge*.
- C. Khatri, R. Goel, B. Hedayatnia, A. Metanillou, A. Venkatesh, R. Gabriel, and A. Mandal. 2018. [Contextual topic modeling for dialog systems](#). In *2018 IEEE Spoken Language Technology Workshop (SLT)*, pages 892–899.
- Diederik P. Kingma and Jimmy Ba. 2014. [Adam: A method for stochastic optimization](#). *CoRR*, abs/1412.6980.
- O. Klejch, P. Bell, and S. Renals. 2017. Sequence-to-sequence models for punctuated transcription combining lexical and acoustic features. In *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 5700–5704.
- Tina Klüwer. 2011. “i like your shirt” - dialogue acts for enabling social talk in conversational agents. In *Intelligent Virtual Agents*, pages 14–27, Berlin, Heidelberg. Springer Berlin Heidelberg.
- Yang Liu, Kun Han, Zhao Tan, and Yun Lei. 2017. Using context information for dialog act classification in dnn framework. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 2170–2178. Association for Computational Linguistics.
- Thang Luong, Hieu Pham, and Christopher D. Manning. 2015. [Effective approaches to attention-based neural machine translation](#). In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 1412–1421, Lisbon, Portugal. Association for Computational Linguistics.
- Stefano Mezza, Alessandra Cervone, Evgeny Stepanov, Giuliano Tortoreto, and Giuseppe Riccardi. 2018. Iso-standard domain-independent dialogue act tagging for conversational agents. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 3539–3551. Association for Computational Linguistics.
- Tomas Mikolov, Edouard Grave, Piotr Bojanowski, Christian Puhersch, and Armand Joulin. 2018. Advances in pre-training distributed word representations. In *Proceedings of the International Conference on Language Resources and Evaluation (LREC 2018)*.
- Stephan Peitz, Markus Freitag, Arne Mauser, and Hermann Ney. 2011. Modeling punctuation prediction as machine translation. In *IWSLT*.
- Eugénio Ribeiro, Ricardo Ribeiro, and David Martins de Matos. 2015. [The influence of context on dialogue act recognition](#). *CoRR*, abs/1506.00839.
- Lina M. Rojas-Barahona, Milica Gašić, Nikola Mrkšić, Pei-Hao Su, Stefan Ultes, Tsung-Hsien Wen, and Steve Young. 2016. [Exploiting sentence and context representations in deep neural models for spoken language understanding](#). In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, pages 258–267, Osaka, Japan. The COLING 2016 Organizing Committee.
- Elizabeth Shriberg, Raj Dhillon, Sonali Bhagat, Jeremy Ang, and Hannah Carvey. 2004. [The ICSI meeting recorder dialog act \(MRDA\) corpus](#). In *Proceedings of the 5th SIGdial Workshop on Discourse and Dialogue at HLT-NAACL 2004*, pages 97–100, Cambridge, Massachusetts, USA. Association for Computational Linguistics.

- Aditya Siddhant, Anuj Goyal, and Angeliki Metallinou. 2019. [Unsupervised transfer learning for spoken language understanding in intelligent agents](#). *Proceedings of the AAAI Conference on Artificial Intelligence*, 33(01):4959–4966.
- Andreas Stolcke, Noah Coccaro, Rebecca Bates, Paul Taylor, Carol Van Ess-Dykema, Klaus Ries, Elizabeth Shriberg, Daniel Jurafsky, Rachel Martin, and Marie Meteer. 2000. [Dialogue act modeling for automatic tagging and recognition of conversational speech](#). *Comput. Linguist.*, 26(3):339–373.
- Ilya Sutskever, Oriol Vinyals, and Quoc V Le. 2014. [Sequence to sequence learning with neural networks](#). In Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 27*, pages 3104–3112. Curran Associates, Inc.
- Henry S. Thompson, Anne Anderson, Ellen Gurman Bard, Gwyneth Doherty-Sneddon, Alison Newlands, and Cathy Sotillo. 1993. The hrc map task corpus: Natural dialogue for speech recognition. In *Proceedings of the Workshop on Human Language Technology, HLT '93*, pages 25–30, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. [Attention is all you need](#). In *Advances in Neural Information Processing Systems*, volume 30, pages 5998–6008. Curran Associates, Inc.
- Marilyn Walker and Rebecca Passonneau. 2001. [DATE: A dialogue act tagging scheme for evaluation of spoken dialogue systems](#). In *Proceedings of the First International Conference on Human Language Technology Research*.
- Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, R’emi Louf, Morgan Funtowicz, and Jamie Brew. 2019. [Huggingface’s transformers: State-of-the-art natural language processing](#). *ArXiv*, abs/1910.03771.

## A Appendices

### A.1 Implementation Details

The baseline dialog act prediction model uses a 2-layer Bi-LSTM with a hidden size of 500. The LSTM layers use a dropout rate of 0.3. We optimize the model with Adam optimizer (Kingma and Ba, 2014). For the Transformer models, we use 12 layers with 12 attention heads and a hidden size of 768. All the fully connected layers use a dropout rate of 0.1. Because one utterance may have multiple labels, following Katakis et al. (2008), we calculate precision, recall, and F1 for multilabel classification) on each sample and then average them across all samples (micro F1). We use the efficient Transformer implementation (Wolf et al., 2019) for our experiments.

For sentence breaker seq2seq model, both the encoder and the decoder are 2-layer 500-dimension bi-LSTMs. In addition, the decoder uses global attention and input feed (Luong et al., 2015) with beam search.

## A.2 Dialog Act Scheme

Dialog Act - Semantic request		
Dialog Act Tag	Description	Example
<i>factual question</i>	factual questions	How old is Tom Cruise; How's the weather today
<i>opinion question</i>	opinionated questions	What's your favorite book; what do you think of disney movies
<i>yes/no question</i>	yes or no questions	Do you like pizza; did you watch the game last night
<i>task command</i>	commands/requests (can be in a question format) for some actions that may be different from the ongoing conversation	can i ask you a question; let's talk about the immigration policy; repeat
<i>invalid command</i>	general device/system commands that cannot be handled by the social bot	show me a picture; cook food for me
<i>appreciation</i>	appreciation towards the previous utterance	that's cool; that's really awesome
<i>general opinion</i>	personal view with polarized sentiment	dogs are adorable; (A: How do you like Tom) B: i think he is great
<i>complaint</i>	complaint about the response from another party	I can't hear you; what are you talking about; you didn't answer my question
<i>comment</i>	comments on the response from another conversation party	(A: my friend thinks we live in the matrix) B1: she is probably right; B2: you are joking, right; B3: i agree; (A: ... we can learn a lot from movies ...) B: there is a lot to learn; (A: He is the best dancer after michael jackson. What do you think) B: michael jackson
<i>statement non_opinion</i>	factual information	I have a dog named Max; I am 10 years old; (A: what movie have you seen recently) B: the avengers
<i>other answer</i>	answers that are neither positive or negative	I don't know; i don't have a favorite; (A: do you like listening to music) B: occasionally
<i>positive answer</i>	positive_answers	yes; sure; i think so; why not
<i>negative answer</i>	negative response to a previous question	no; not really; nothing right now

Table 4: Dialog Act Scheme for Semantic request

<b>Dialog Act - Functional request</b>		
<b>Dialog Act Tag</b>	<b>Description</b>	<b>Example</b>
<i>abandon</i>	not a complete sentence	So uh; I think; can we
<i>nonsense</i>	utterances that do not make sense to humans	he all out
<i>hold</i>	a pause before saying something	let me see; well
<i>opening</i>	opening of a conversation	hello; hi
<i>closing</i>	closing of a conversation	nice talking to you; goodbye
<i>thanks</i>	expression of thankfulness	thank you
<i>back-channeling</i>	acknowledgement to the previous utterance	Uh-huh; (A: i learned that ...) B: okay/yeah/right/really?
<i>apology</i>	apology	I'm sorry
<i>apology response</i>	response to apologies	That's all right
<i>other</i>	utterances that cannot be assigned to other tags	

Table 5: Dialog Act Scheme for Functional request

### A.3 Multi-functionality schemes

Multi-label tags		
Dialog Act Tags	Example	Count in User Utterances
<i>positive answer, task command</i>	(A: wanna know something interesting about it?) B: sure; (A: do you want to talk about some games) B: minecraft	698
<i>negative answer, task command</i>	(A: would you like to know more about it) B: I don't want to hear more	328
<i>task command, general opinion</i>	(A: what do you want to talk about) B: harry potter stuff	192
<i>task command, statement non_opinion</i>	let's talk about mario kart	141
<i>positive answer, statement non_opinion</i>	(A: have you read any books recently?) B: I'm reading the great gatsby	133
<i>task command, yes/no question</i>	do you know tom brady; (A: what do you want to talk about?) B: how about movies	116
<i>negative answer, statement non_opinion</i>	(A: do you have pets) B: I don't have any	66
<i>positive answer, general opinion</i>	(A: do you like animals) B: My favorite animals is panda	35
<i>invalid command, yes/no question</i>	can you speak louder	15
<i>task command, factual question</i>	what do you know about dodgers	12
<i>negative answer, general opinion</i>	(A: do you watch sports) B: I'm not into sports	10
<i>task command, opinion question</i>	(A: what did you find interesting recently) B: what do you think of the new movie	9
<i>task command, complaint</i>	I don't want to hear you talk about anything; would you stop asking me that question	5
<i>other answer, general opinion</i>	(A: what's your favorite movie) B: there are so many to choose from	5
<i>positive answer, comment</i>	(A: don't you think so) B: it's true	4
<i>general opinion, yes/no question</i>	(A: what would you imagine doing in such situation) B: can we just sleep all day	3
<i>negative answer, comment</i>	(A: isn't that interesting) B: that's ridiculous	3
<i>general opinion, opinion question</i>	(A: what book would you recommend me to read) B: how about antifrangible	3

### A.4 Dialog act tag mapping

<b>SWBD-DAMSL</b>	<b>SWBD</b>	<b>MIDAS</b>
<i>statement_non_opinion</i>	<i>sd</i>	<i>statement non_opinion</i>
<i>Acknowledge (Backchannel)</i>	<i>b</i>	<i>back-channeling</i>
<i>Statement-opinion</i>	<i>sv</i>	<i>general opinion</i>
<i>Agree/Accept</i>	<i>aa</i>	<i>positive answer</i>
<i>Abandoned or Turn-Exit</i>	<i>% -</i>	<i>abandon</i>
<i>Appreciation</i>	<i>ba</i>	<i>appreciation</i>
<i>Yes-No-Question</i>	<i>qy</i>	<i>yes-no question</i>
<i>Non-verbal</i>	<i>x</i>	
<i>Yes answers</i>	<i>ny</i>	<i>positive answer</i>
<i>Conventional-closing</i>	<i>fc</i>	<i>closing</i>
<i>Uninterpretable</i>	<i>%</i>	<i>abandon</i>
<i>Wh-Question</i>	<i>qw</i>	
<i>No answers</i>	<i>nn</i>	<i>negative answer</i>
<i>Response Acknowledgement</i>	<i>bk</i>	<i>back-channeling</i>
<i>Hedge</i>	<i>h</i>	<i>other answers</i>
<i>Declarative Yes-No-Question</i>	<i>qy^d</i>	<i>yes-no question</i>
<i>Other</i>	<i>o,fo,bc,by,fw</i>	<i>other</i>
<i>Backchannel in question form</i>	<i>bh</i>	<i>back-channeling</i>
<i>Quotation</i>	<i>q̂</i>	<i>general opinion</i>
<i>Summarize/reformulate</i>	<i>bf</i>	<i>general opinion</i>
<i>Affirmative non-yes answers</i>	<i>na, ny^e</i>	<i>positive answer</i>
<i>Action-directive</i>	<i>ad</i>	<i>task command</i>
<i>Collaborative Completion</i>	<i>^2</i>	<i>general opinion</i>
<i>Repeat-phrase</i>	<i>b^m</i>	<i>general opinion</i>
<i>Open-Question</i>	<i>qo</i>	
<i>Rhetorical-Questions</i>	<i>qh</i>	
<i>Hold before answer/agreement</i>	<i>^h</i>	<i>hold</i>
<i>Reject</i>	<i>ar</i>	<i>negative answer</i>
<i>Negative non-no answers</i>	<i>ng,nn^e</i>	<i>negative answer</i>
<i>Signal-non-understanding</i>	<i>br</i>	<i>complaint</i>
<i>other_answers</i>	<i>no</i>	<i>other answer</i>
<i>Conventional-opening</i>	<i>fp</i>	<i>opening</i>
<i>Or-Clause</i>	<i>qrr</i>	<i>other</i>
<i>Dispreferred answers</i>	<i>arp,nd</i>	<i>negative answer</i>
<i>3rd-party-talk</i>	<i>t3</i>	
<i>Offers, Options Commits</i>	<i>oo,cc,co</i>	<i>other</i>
<i>Self-talk</i>	<i>t1</i>	<i>other</i>



<b>SWBD-DAMSL</b>	<b>SWBD</b>	<b>MIDAS</b>
<i>Downplayer</i>	<i>bd</i>	<i>apology response</i>
<i>Maybe/Accept-part</i>	<i>aap/am</i>	<i>positive answer</i>
<i>Tag-Question</i>	<i>^g</i>	<i>other</i>
<i>Declarative Wh-Question</i>	<i>qw^d</i>	
<i>Apology</i>	<i>fa</i>	<i>apology</i>
<i>Thanking</i>	<i>ft</i>	<i>thanking</i>

Table 6: Dialog act tag mapping among SWBD-DAMSL, SWBD, and MIDAS

<b>MIDAS</b>	<b>ISO</b>
<i>factual question</i>	<i>Question, Set-Question, Choice-Question</i>
<i>opinion question</i>	<i>Question, Set-Question, Choice-Question</i>
<i>yes/no question</i>	<i>Propositional-Question, Check-Question</i>
<i>task command</i>	<i>Instruct, Suggest, Request, Offer, Promise</i>
<i>invalid command</i>	<i>Instruct, Suggest, Request, Offer, Promise</i>
<i>appreciation</i>	<i>AutoPositive, AlloPositive</i>
<i>general opinion</i>	<i>Inform</i>
<i>complaint</i>	<i>AutoNegative, AlloNegative</i>
<i>comment</i>	<i>AutoPositive, AutoNegative, AlloPositive, AlloNegative</i>
<i>statement non_opinion</i>	<i>Inform, Init-Self-Introduction, Return Self-Introduction</i>
<i>other answer</i>	<i>Address-Offer, Address-Request, Address-Suggest, Answer, Correction, Disagreement, Agreement</i>
<i>postive answer</i>	<i>Confirm, Accept-Offer, Accept-Request, Accept-Suggest, Agreement</i>
<i>negative answer</i>	<i>Disconfirm, Decline-Offer, Decline-Suggest, Decline-Request. Disagreement</i>
<i>abandon</i>	
<i>nonsense</i>	
<i>opening</i>	<i>Init-Greeting, Return Greeting</i>
<i>closing</i>	<i>Init-Goodbye, Return Goodbye</i>
<i>hold</i>	<i>Staling, Pausing</i>
<i>thanks</i>	<i>Thanking, Accept Thanking</i>
<i>back-channeling</i>	<i>Propositional-Question</i>
<i>apology</i>	<i>Apology</i>
<i>apology response</i>	<i>Accept Apology</i>
<i>other</i>	<i>Other</i>

Table 7: Dialog act tag mapping between MIDAS and ISO communicative functions