Resolving Intent Ambiguities by Retrieving Discriminative Clarifying Questions

Kaustubh D. Dhole
Amelia Science
R&D, IPsoft
New York, NY 10004
kdhole@ipsoft.com

Abstract

Task oriented Dialogue Systems generally employ intent detection systems in order to map user queries to a set of pre-defined intents. However, user queries appearing in natural language can be easily ambiguous and hence such a direct mapping might not be straightforward harming intent detection and eventually the overall performance of a dialogue system. Moreover, acquiring domain-specific clarification questions is costly. In order to disambiguate queries which are ambiguous between two intents, we propose a novel method of generating discriminative questions using a simple rule based system which can take advantage of any question generation system without requiring annotated data of clarification questions. Our approach aims at discrimination between two intents but can be easily extended to clarification over multiple intents. Seeking clarification from the user to classify user intents not only helps understand the user intent effectively, but also reduces the roboticity of the conversation and makes the interaction considerably natural.

1 Introduction

Task oriented dialogue systems aim at extracting semantic information from natural language queries in order to decipher user's intents. Such systems play a vital role in commercial applications like personal assistants (e.g. Google Home, Alexa, Siri, etc.) for a variety of domain specific tasks like flight-booking, call routing, restaurant booking and so on which typically model a dedicated Natural Language Understanding (NLU) component that performs inference for downstream tasks like domain classification, intent detection and slot filling.

A major driving component of NLU is Intent Detection which operates over users' queries. However, users' queries are generally ambiguous and underspecified. Eg. in a banking domain, given two pre-defined intents, "opening_a_savings_account" and "opening_a_checking_account", even a simple user query like "I want to open an account" does not directly map to either of the two intents and requires disambiguation. Managing this would require creating a separate intent representing "opening an account" but that would mean creating the corresponding task workflows, acquiring extra training data to incorporate the new intent and retraining intent and possibly other subsequent classifiers.[1]

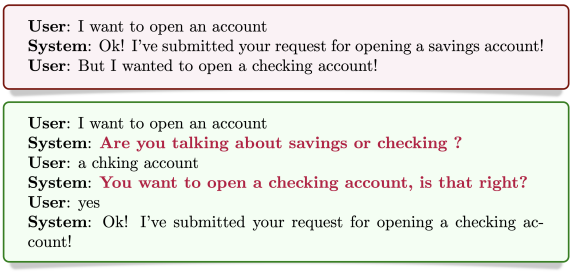

Figure 1: In the first conversation, the system suffers due to unavailability of a separate intent and hence misunderstands the user's intent. In the second conversation, the system generates two clarifying questions in order to disambiguate and clarify the user's intent successfully.

In this paper, we explore this problem specifically in the more pragmatic task oriented dialog setting to improve intent classification by incorporating a limited form of unsupervised interaction as shown in Figure 1. In order to disambiguate between two intents, given an ambiguous natural language query, we describe a simple rule-based system to generate discriminative questions using an existing question generator and a sentence similarity model. Generating discriminative questions has significant advantages over a one-to-one utterance-to-intent classification: (i) It improves the overall

[1]While this also depends on the design of the dialog system, we assume the pipelined approach of classifying the domain first followed by the intent.

accuracy of classifying the user's intent since it boils down the role of non-deterministic classifiers from a top-1 to an easier top-k classification problem permitting the classifiers a little slack in performance by acquiring clarification from the end-user herself. (ii) Rather than relying on a single user input, the communication with the dialog system becomes highly interactive.

2 Related Work

Clarification requests were studied in dialogue extensively by (Purver et al., 2003a,b; Purver, 2004; Healey et al., 2003) who also established a taxonomy of the various types of clarification. Coden et al. (2015) discussed challenges involved in disambiguating entities via clarification. With the rise of conversational systems, there has been enormous interest in generating clarifying questions and datasets recently. Xu et al. (2019) constructed a clarification dataset to address ambiguity arising in knowledge-based question answering. Alianne-jadi et al. (2019) proposed a clarification dataset to improve open-domain information-seeking conversations. Kumar and Black (2020) built a clarification dataset by sampling comments from StackExchange posts. Rao and III (2019); Cao et al. (2019); Zamani et al. (2020) have attempted to use neural models to train over (context, question) pairs to generate clarifying questions. Rao and III (2019) proposed an RL based model for generating a clarifying question in order to identify missing information in product descriptions. Cao et al. (2019) described an interesting approach feeding expected question specificity along-with the context to generate specific as well as generic clarifying questions. However, most of these models still require large amounts of training data with Wizard-of-Oz style dialog annotations. Yu et al. (2020) attempt to generate binary and multiple choice questions and show the benefit of incorporating interaction for determining user's intent.

Xu et al. (2020) use a graph neural network and a novel attention mechanism to capture the discriminative attributes of confusing law articles.Emphasizing that ambiguity is a function of the user query and the evidence provided by a very large text corpus, Min et al. (2020) introduce an interesting dataset and an associated task to generate disambiguating rewrites of an original open-domain question. Li et al. (2017) explore an effective method for generating discriminating questions to disambiguate pairs of images.

Our approach is close to Yu et al. (2020); Zamani et al. (2020). In contrast to Yu et al. (2020) where questions and answer choices are manually generated, we seek to automate this by using a simple TF-IDF approach to generate potential answer choices to disambiguate to a particular intent. Additionally, instead of collecting domain specific discriminative questions which are harder to obtain, we show how we can generate discriminative questions by using only a sentence-level question generator and a discriminative similarity measure. Besides, our approach is simpler to incorporate in production systems with small amounts of training data for intents. Keeping this interaction component partitioned from the one-to-one intent classifier also eases its incorporation into dialogue systems with pre-deployed intent classifiers.

3 Model

Given an ambiguous utterance, our goal is to generate a discriminative question to obtain clarification between two highly probable intents.

- First, we train an intent classifier and classify the incoming utterance into one of several intents
- Using a pre-trained question generation system, we generate question answer pairs and select the question with the highest potential to discriminate
- If the question does not have high *discriminative similarity*, we generate a template based question from the intent's pre-computed discriminative attributes
- Finally, we classify the user's subsequent response to the discriminative question into either of the two intents or none of them. We describe each of the steps in detail in the following subsections.

3.1 Intent Classification and Ambiguity Detection

Given a set of utterances in the form of user sentences $x_1, x_2...x_n$ with their annotated intents $y_1, y_2...y_n$ where $y_i \in 1...m$, we train a sentence classifier in order to create an intent classifier. A softmax layer is used to assign probabilities to each intent $p_1, p_2...p_n$.

At runtime for a given user query q, we execute the intent classifier to get the probability scores.

If the softmax scores of the highest intent j namely p_j is lesser than a pre-determined confidence threshold t_1, we consider the query as ambiguous in itself.

If the softmax scores of the two highest intents j, k namely p_j, p_k is within a pre-determined threshold, we consider the query as ambiguous between intents j and k i.e. if $p_j - p_k < t_2$ where $t_2 \in [0, 1]$ is the two-intent ambiguity threshold.

3.2 Discriminative Question Selection

In order to generate a discriminative question, we use an existing question generation system and the annotated utterances used for training the intent classifier itself. We collect all the training utterances of the top two ambiguous intents:

$$J = \{x_i \forall i | y_i = j\}$$
$$K = \{x_i \forall i | y_i = k\}$$

For all the utterances in J and K, we generate question answer pairs using SynQG (Dhole and Manning, 2020) and accumulate them in the following two sets of question answer pairs respectively from which we select one question-answer pair from each set in order to further compute our representative discriminative question.

$$Q_J, Q_K$$

We are interested to select a question-answer pair from each of the above two sets whose question can serve as a potential discriminative question. We attempt to identify one question-answer pair from each set $(q_J^*, a_J^*) \in Q_J$ and $(q_K^*, a_K^*) \in Q_K$ using the following discriminatory conditions.

$$\forall j \in |Q_J|, \forall k \in |Q_K| :$$
$$s_{j,k} = score(q, q_j, a_j, q_k, a_k)$$
$$j^*, k^* = max(s_{j,k} \forall (j, k))$$

where the discriminative score is defined as follows:

$$score(q, q_j, a_j, q_k, a_k) = sim(q_j, q_k)$$
$$-sim(a_j, a_k)$$
$$+0.5(sim(q, q_j) + sim(q, q_k))$$

We hypothesize that an ideal discriminative question would be such that its corresponding answers for each of the intents would have to be not only different and discriminative, but the answers should exclusively be present in each of the two intents. Hence we would expect a_J^* and a_K^* to be highly dissimilar: $-sim(a_j, a_k)$[2]

Additionally, both q_J^* and q_K^* should be neutral to both the intents and highly similar to each other since we want a question to be identical enough to trigger both the intents: $sim(q_j, q_k)$

We draw a parallel here with Li et al. (2017)'s task of generating a discriminative question from a pair of ambiguous images by seeking to identify discriminative regions, choosing pairs with high contrast, high visual dissimilarity and high question similarity.

However, two intents might have multiple sources of ambiguity than provided by their definitions. We try to figure out the specific source by looking at the user utterances used within the two intents and the user query. Consider the following extreme case wherein the first question in both the given pairs belongs to a common intent and the second question also belongs to another common intent : The following pair qualifies as being rightly disambiguating:

(What is the type of account?, savings)

(What is the account type?, checking)

as well as this pair:

(What would you like to do?, open a savings account)

(What do you want to do?, open a checking account)

For a user query $q =$ "I want to open an account", questions belonging to the first pair can serve as discriminatory questions but not from the second pair. However, for a user query like $q =$ "I would like do do this", only questions from the second pair would be useful. Hence, we attempt to further re-rank the questions by the similarity with the user query: $+0.5(sim(q, q_j) + sim(q, q_k))$.

For each of the above similarity computations, we use encodings from Cer et al. (2018) and perform a cosine similarity.

Moreover, since user queries' grammatical style is meant to be in a form suitable to communicate facing the agent, these questions can't be presented back to the user directly. And hence, we perform a simple set of substitutions to perform the conver-

[2]This also depends on the choice of the similarity function like retro-fitting vectors (Faruqui et al., 2015) might be a better choice than GloVe (Pennington et al., 2014) or Word2Vec (Mikolov et al., 2013) when answers are common nouns or adjectives.