Baskin Engineering UC SANTA CRUZ

NATURAL LANGUAGE AND DIALOGUE SYSTEMS UC SANTA CRUZ

香港中文大學 The Chinese University of Hong Kong

Implicit Discourse Relation Identification for Open-domain Dialogues

Mingyu Derek Ma, Kevin K. Bowden, Jiaqi Wu, Wen Cui and Marilyn Walker

mingyuma@usc.edu; {kkbowden, jwu64, wcui7, mawalker}@ucsc.edu

Code and Data

Introduction

- Four major classes of discourse relation

  Comparison, Expansion, Contingency, Temporal
- **Implicit** VS **Explicit**: whether they have clear connective cues

  *it's a great album **(but)** it's probably not their best.*
- Implicit discourse relations are hard to detect

  - No discourse cue, only inferred on the basis of textual features
- Discourse relations are relatively unexplored in dialogue systems, they can

  - Cultivate a more aware state space to improve continuity
  - Serve as ranking parameters for possible next turns, refine database queries, or generate content with NLG

Challenges

- Existing datasets with discourse relation labels are based on monologic text such as news (e.g., Penn Discourse Tree Bank), which is unlikely to provide good training material for dialogue
- No previous work investigating the feasibility of applying a ML model developed on formal text to dialogic content, where turns are normally comprised of short, informal text
- **Lack of labeled data for pairs of implicit discourse relations in open-domain dialogue**

Dataset Construction

- Based on the Edina self-dialogue corpus - contains 24,165 multi-turn social conversations across 23 topics with no discourse relation labels
- Core idea: Converting explicit discourse relation pairs into implicit pairs by dropping the connectives

1. Initial connectives pool identified through statistical analysis of connective frequencies in PDTB (Pitler et al., 2008); we only select connectives strongly associated with each class of relation

2. Remove some connectives from the selection if they are not freely omissible (the meaning will change if the connectives are removed) by calculating Omissible Rate and Context Differential (Rutherford and Xue, 2015)

   Selected connective words for each relation:
   - Comparison: *but, however, although, by contrast*
   - Contingency: *because, so, thus, as a result, consequently, therefore*
   - Expansion: *also, for example, in addition, instead, indeed, moreover, for instance, in fact, furthermore, or, and*
   - Temporal: *then, previously, earlier, later, after, before*

3. Select the conversations matching specific predefined patterns (Arg 1) (connective) (Arg 2) and (Arg 1). (Connective),(Arg 2)

4. Heuristic rules for final filtering: only full sentence, POS tags for particular connectives

Edina-DR dataset statistics:

	Edina-DR	PDTB
# pairs of all relations	27998	11734
avg # words of arg 1	7.1	18.8
avg # words of arg 2	7.3	19.4
# pairs of 'Comparison'	20823	1799
# pairs of 'Contingency'	5080	2243
# pairs of 'Expansion'	1580	6933
# pairs of 'Temporal'	452	759

- Twice as many pairs as PDTB
- Distribution of discourse relations is different:
  - Most pairs belong to "Comparison"
  - Small number of "Temporal" pairs

Model

Feature-based Classifier

- Dialogue features extracted using the NLU in SlugBot, an open-domain Alexa Prize system
- Here we feed dialogue features as one-hot vectors to logistic regression

Dialogue Features

Dialogue Act: The act of a dialogue utterance is obtained using the NPS dialogue act classifier (Forsyth and Martell, 2007). There are 15 different dialogue acts, including GREET, CLARIFY, and STATEMENT.

Sentiment: The sentiment of a dialogue utterance is obtained from the Stanford CoreNLP Toolkit; there are five possible sentiment values.

Intent: An utterance intent ontology consisting of 33 discrete intents is developed and recognized using heuristics and a trained model. Some sample intents are REQUEST OPINION, REQUEST SERVICE, and REQUEST CHANGE TOPIC. It is trained using roughly 50K utterances from the Common Alexa Prize Chats (CAPC) dataset; the model ensembles both a Recurrent Neural Network and Convolutional Neural Network (Ram et al., 2018).

Topic: The topic of the utterance is obtained using the CoBot (Conversational Bot) topic classification model, which is a Deep Average network BiLSTM model.

Core Entities Types: We use SlugNERDS to detect our named entities (Bowden et al., 2018b, 2017). Here we use the constantly updated Google Knowledge Graph. Specifically, we use entity types rather than the entities themselves.

Deep Learning Model with Dialogue Features

- Built on Deep Enhanced Representation (DER) model (Bai and Zhao, 2018); best performer on discourse relation identification for PDTB
- Train DER model on the new Edina-DR dataset to evaluate the adaptability of the existing model to dialogic data
- Extend the argument pair representation vector by connecting the dialogue feature vectors

Experiments and Evaluation

- 400 expert annotated gold labels
  - 12% of the samples do not form a discourse relation due to grammar issues

Feature-based Classifier and Dialogue Feature Selection

Features	Precision	Recall	F1
DIALOGUE ACT	0.64	0.69	0.66
INTENT	0.63	0.74	0.68
TOPICS	0.62	0.71	0.66
SENTIMENT	0.56	0.74	0.64
ENTITIES TYPES	0.63	0.74	0.68
All	0.63	0.65	0.64
All - SENTIMENT	0.64	0.73	0.68

- The INTENT and ENTITIES TYPES show the best performance individually
- The SENTIMENT feature leads to a large drop in precision
- Best configuration: all features except SENTIMENT

Deep Learning Models

Model	Acc.	F1
DER (PDTB)	0.61	0.51
Logistic Reg. (Edina-DR)	0.64	0.68
DER (Edina-DR)	0.80	0.76
DER+Dialogue (Edina-DR)	**0.81**	**0.77**

- DER performs surprisingly well
- Strong adaptability for DER model to the task of discourse relation identification in dialogues
- The model with dialogue features improved performance