Implicit Discourse Relation Identification for Open-domain Dialogues

Mingyu Derek Ma[1], Kevin K. Bowden[2], Jiaqi Wu[2], Wen Cui[2] and **Marilyn Walker[2]**

[1]Human-Computer Communications Laboratory &
Stanley Ho Big Data Decision Analytics Research Centre
The Chinese University of Hong Kong
`derekma@cuhk.edu.hk`
[2]Natural Language and Dialogue Systems Lab
University of California, Santa Cruz
{`kkbowden, jwu64, wcui7, mawalker`}`@ucsc.edu`

Abstract

Discourse relation identification has been an active area of research for many years, and the challenge of identifying implicit relations remains largely an unsolved task, especially in the context of an open-domain dialogue system. Previous work primarily relies on a corpora of formal text which is inherently non-dialogic, i.e., news and journals. This data however is not suitable to handle the nuances of informal dialogue nor is it capable of navigating the plethora of valid topics present in open-domain dialogue. In this paper, we designed a novel discourse relation identification pipeline specifically tuned for open-domain dialogue systems. We firstly propose a method to automatically extract the implicit discourse relation argument pairs and labels from a dataset of dialogic turns, resulting in a novel corpus of discourse relation pairs; the first of its kind to attempt to identify the discourse relations connecting the dialogic turns in open-domain discourse. Moreover, we have taken the first steps to leverage the dialogue features unique to our task to further improve the identification of such relations by performing feature ablation and incorporating dialogue features to enhance the state-of-the-art model.

1 Introduction

Discourse analysis considering relations between clauses has received increasing attention from the field, and implicit discourse relation identification is one of the most challenging problems in discourse parsing since it is purely based on textual features. Previous work has defined four widely accepted major classes of discourse relation - "Comparison", "Expansion", "Contingency" and "Temporal" (Miltsakaki et al., 2008; Prasad et al., 2008). These four relations can either be explicitly or implicitly realized. When explicitly realized, there are often clear connective words between clauses which result in an associated discourse relation, while implicit realizations are often much harder to detect. For example, people can imply there is a "Comparison" relation between the following two sentences by understanding the meaning. Without clear keywords like "but" however, it is hard for machines to recognize such implicit relations.

> Arg 1: *it's a great album.*
> Arg 2: *it's probably not their best.*

Since the development of the Penn Discourse Treebank (PDTB)[1], discourse relation identification has been treated as a supervised learning problem. For explicit discourse relation pairs, simple classification methods based on connective cues achieve more than 90% accuracy (Pitler et al., 2008). For implicit discourse relations however, where there is no discourse clue, relations needs to be inferred on the basis of textual features, making this a challenging problem in discourse parsing (Li and Nenkova, 2014; Lin et al., 2009).

While previous work has suggested that discourse relations may hold between dialogue turns, this idea is relatively unexplored (Stent, 2000; Tonelli et al., 2010). We posit that discourse relation identification could have wide application in dialogue systems, by cultivating a more aware state space in order to improve the continuity between an extended sequence of turns. The detected discourse relation could additionally serve as a query or ranking parameter for possible next turns, retrieved from a database of content, or generated by natural language generation. Adding this additional natural language understanding component might be especially useful when navigating open-domain dialogue where user input is unpredictable and the model must be topic-robust.

---

[1]More details about Penn Discourse Treebank can be found at `https://www.seas.upenn.edu/~pdtb/`

There are many fundamental challenges with identifying and utilizing discourse relations in an open-domain dialogue system. All existing datasets for discourse relation identification are based on monologic text such as news; these datasets are unlikely to provide good training material for dialogue. Moreover there is no previous work investigating the feasibility of applying a machine learning model developed on formal text to dialogic content, where turns in are normally short, informal text. Thus, the lack of labeled dialogue data for implicit discourse relation pairs in open-domain dialogue is the first challenge that must be addressed.

To tackle these two problems and utilize the unexplored benefits of features unique to dialogue systems, we carry out two steps. First, we construct a discourse relation pair dataset from a large corpus of open-domain dialogue, which to our knowledge is the first of its kind. Second, we investigated a feature-based model with different dialogue feature combinations and enhanced a deep learning model by incorporating dialogue features that utilize aspects unique to dialogue. The dataset and related code are publicly available.[2]

2 Related Work

The release of the Penn Discourse Treebank (PDTB) (Prasad et al., 2008) makes research on machine learning based implicit discourse relation recognition possible. Most previous work is based on linguistic and semantic features such as word pairs and brown cluster pair representation (Pitler et al., 2008; Lin et al., 2009) or rule-based systems (Wellner et al., 2006). Recent work has proposed neural network based models with attention or advanced representations, such as CNN (Qin et al., 2016), attention on neural tensor network (Guo et al., 2018), and memory networks (Jia et al., 2018). Advanced representations may help to achieve higher performance (Bai and Zhao, 2018). Some methods also consider context paragraphs and inter-paragraph dependency (Dai and Huang, 2018).

To utilize machine learning models for this task, larger datasets would provide a bigger optimization space (Li and Nenkova, 2014). Marcu and Echihabi (2002) is the first work to generate artificial samples to extend the dataset by using rules to

convert explicit discourse relation pairs into implicit pairs by dropping the connectives. This work is further extended by methods for selecting high-quality samples (Rutherford and Xue, 2015; Xu et al., 2018; Braud and Denis, 2014; Wang et al., 2012).

Most of the existing work discussed so far is based on the PDTB dataset, which targets formal texts like news, making it less suitable for our task which is centered around informal dialogue. Related work on discourse relation annotation in a dialogue corpus is limited (Stent, 2000; Tonelli et al., 2010). For example Tonelli et al. (2010) annotated the Luna corpus,[3] which does not include English annotations. To our knowledge there is no English dialogue-based corpus with implicit discourse relation labels, as such research specifically targeting a discourse relation identification model for social open-domain dialogue remains unexplored.

3 Dataset Construction

Previous work on discourse relation identification suggests that the most effective approach is supervised learning, but limited amounts of annotated data constrain the application of such algorithms. Previous work has additionally proven that weakly labeled data, which contains a small number of false labels and can be generated automatically, helps improve classifier performance with implicit relations (Rutherford and Xue, 2015).

We therefore constructed Edina-DR, the novel dataset of discourse relation pairs based on the publicly available self-dialogue Edina corpus which contains 24,165 multi-turn social conversations across 23 topics (Fainberg et al., 2018; Krause et al., 2017).[4] To the best of our knowledge, this is the first English discourse relation dataset based on open-domain dialogues. The Edina dataset initially contains no discourse relation labels. Inspired by the approaches taken to automatically extend PDTB, we designed a pipeline to extract discourse relation argument pairs through utilizing the connective words which are known as clear relation indicators. The pipeline automatically extracts argument pairs and assign discourse relation labels to each of the utterances. We

---

[2] https://github.com/derekmma/dialogue-discourse-relation

[3] EU FP6 contract No. 33549, http://www.ist-luna.eu/

[4] The Edina dataset is publicly available at https://github.com/jfainberg/self_dialogue_corpus

then have humans annotate a small sample of the data in order to validate the automated pipeline. Our pipeline targets the four level-1 discourse relations, i.e., "Comparison", "Expansion", "Contingency" and "Temporal".

We obtained this initial connectives pool according to statistical analysis of connective frequencies in PDTB conducted by Pitler et al. (2008), in which we only consider connectives which are strongly associated (probability $> 95\%$) with only one class of relation.[5] For example, we exclude the connective word "since" because it may often appear as an indicator of either a "Temporal" or "Contingency" relation.

Secondly, some connectives cannot be removed without changing the original meaning (Sporleder and Lascarides, 2008). We follow the method proposed by Rutherford and Xue (2015) to identify the connectives which are freely omissible by measuring the Omissible Rate and Context Differential. Since we need some manually labeled connectives for this task, we implement the connective selection on the PDTB dataset and generalize the selection result to the dialogue dataset. The selected connectives include:

- **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

The third step is to select the conversations matching specific predefined patterns for different structures of the sentences with the selected connective words shown above. Inspired by (Braud and Denis, 2014; Marcu and Echihabi, 2002), we use two patterns: (Arg 1) (connective) (Arg 2) and (Arg 1). (Connective),(Arg 2). In other words, we have one pattern for when connectives appear in the middle of an utterance, and another pattern for when connectives link two arguments in adjacent utterances across separate turns. Finally,

---

[5]The list of connectives for each relation in detail can be found in (Pitler et al., 2008).

we defined several heuristic rules to filter out low-quality pairs which have been applied in previous work (Braud and Denis, 2014). The program only accepts full sentence arguments and we use certain POS tags for particular connectives to make sure the connective function as relation indicators. A segment window is defined so that our method only picks the closest phrases or sub-sentences if the whole conversation contains several sentences.

For example, in the sentence "*they had a $5 off the price, so i bought it.*", the connective "so" is identified in the list of connective words for "Contingency" relation and the sentence matches our pattern 1. Therefore we convert this sentence to a "Contingency" discourse relation pair and the two arguments are "they had a $5 off the price" and "i bought it".

	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

Table 1: Statistics of the extracted dataset Edina-DR

The statistics of the annotated dialogue discourse relation pairs dataset Edina-DR is shown in Table 1. The new dataset contains more than twice the pairs compared to PDTB, which should prove useful for machine learning. We note that the distribution of discourse relations in the Edina-DR dataset is different from PDTB. Most of the pairs belong to the "Comparison" relation, which is a natural way to structure dialogue. The number of "Temporal" pairs however is smaller, one possible explanation being that people do not use connectives words often in dialogues when talking about time-related events. These differences highlight the need for this work, as it's clear that human dialogue is in fact structured differently than more formal non-dialogic text.

We annotated discourse relations for 400 samples out of the extracted dataset by an expert annotator, 12% of the samples do not form a discourse relation which probably due to failures by the automatic extraction program to catch particular linguistic structures. 88% of the samples which do hold relations match the relation labels of the hu-

man annotations, which proves the reliability of our proposed extraction method.

4 Model

We propose the novel approach of applying the unique dialogue features encapsulated in the state-space of a real deployed dialogue systems to enhance discourse relation identification. Firstly, we use a feature-based classifier for feature selection and then we explore the feasibility of utilizing existing deep learning model in dialogue discourse relation identification task.

4.1 Feature-based Classifier

We extract dialogue features using the Natural Language Understanding (NLU) capabilities in SlugBot, a deployed open-domain dialogue system (Bowden et al., 2018a). These features are normally used for dialogue management and content retrieval. We input raw argument pairs into the NLU pipeline and get dialogue features which are then fed as one-hot vectors to a logistic regression classifier. A full dialogue feature vector contains 448 features. The dialogue features include:

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. The full list of dialogue acts is described in (Forsyth and Martell, 2007).

Sentiment: The sentiment of a dialogue utterance is obtained from the Stanford CoreNLP Toolkit (Manning et al., 2014) and there are five possible sentiment values: VERY POSITIVE, POSITIVE, NEUTRAL, NEGATIVE, and VERY NEGATIVE.

Intent: An utterance intent ontology consisting of 33 discrete intents is developed and recognized using heuristics and a trained model. It is designed to obtain utterance intent without conversational context, so only the input utterance is considered for intent detection. Some sample intents are REQUEST_OPINION, REQUEST_SERVICE, REQUEST_CHANGE_TOPIC. It is trained using a subset of Common Alexa Prize Chats (CAPC) dataset with roughly 50K utterances and 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) toolkit topic classification model (Khatri et al., 2018), which is a Deep Average network BiLSTM model. The model is trained on over 120,000 utterances and labeled across 22 topics. This includes commonly discussed topics such as POLITICS, FASHION, SPORTS, SCIENCE AND TECHNOLOGY, and MUSIC.

Core Entities Types: We use SlugNERDS to detect our named entities (Bowden et al., 2018b, 2017). SlugNERDS is specialized for open-domain dialogue interactions. It can sift through noisy user data and it uses the constantly updated Google Knowledge Graph[6] to remain aware of even the latest named entities. Both of these points are vital for understanding social chit-chat. We only consider the entity types of the entities as feature rather than entities themselves. We use standard `schema.org` types and there are totally 614 types. For example, if SlugNERDS detects "Cam Newton", which is an entity with type PERSON, then PERSON is used as feature.

4.2 Deep Learning Model with Dialogue Features

To investigate the adaptability of existing discourse relation identification models on dialogue data and our proposed features, we build on the Deep Enhanced Representation (DER) model of Bai and Zhao (2018)[7], which demonstrated its efficiency by achieving the current state-of-the-art performance on the PDTB dataset. It utilized different grained text representations including character, sub-word, word, sentence, and sentence pair levels, with embeddings obtained by ELMo (Peters et al., 2018). The model first generates representations for the argument pairs using an encoder and bi-attention module; these are then sent to the classifier, consisting of multiple layer perceptrons with softmax, to predict the discourse relation.

We take the DER design and architecture and train on Edina-DR dataset to evaluate the adaptability of existing model in dialogue environment. Then we explore a variation of this model by connecting dialogue feature vectors to the argument pairs representation vector to extend the representation. We use the same method to encode all dialogue features as the feature-based classifier. With the help of previous experiments, we use the best feature combination for the dialogue feature vec-

---

[6]https://developers.google.com/knowledge-graph/

[7]Original implementation of the authors can be found at https://github.com/hxbai/Deep_Enhanced_Repr_for_IDRR.

tors.

5 Evaluation and Analysis

For the following experiments, we randomly selected 400 samples to be used as test set with discourse relation labels annotated by an expert. We repeat the experiments five times and take the average score as the final report results.

5.1 Feature-based Classifier and Dialogue Feature Selection

We first analyze the performance of the feature-based model with different feature combinations shown in Table 2.

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

Table 2: Feature-based Model Evaluation

For single dialogue features, INTENT and ENTITIES TYPES provide the largest performance boost compared to other single dialogue features, and this demonstrates the effectiveness of using intent and types of entities for discourse relation identification. Other three features maintain the same level of performance, except a large drop in precision with respect to SENTIMENT. One possible explanation is that our sentiment classification results are obtained using the Sentiment Annotator from Stanford CoreNLP Toolkit, which is trained on movie reviews corpus (Manning et al., 2014; Socher et al., 2013). The nature of training data is not suitable for our dialogue corpus in this task. Using Table 2, we see that the best configuration includes all of our dialogue features except SENTIMENT.

5.2 Deep Learning Models

In Table 3, we see the results of our experiments, where DER represents our baseline model. We use the default parameter for DER models. We also show the result of the DER model trained and tested on the PDTB dataset for comparison marked as "DER (PDTB)". The first observation is that the DER model performs surprisingly well

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**

Table 3: Performance of Deep Learning Models (Dataset name is shown in parentheses)

with an F1 score of 0.76 on the new dialogue discourse relation dataset Edina-DR with p-value of 0.008, which demonstrates its strong adaptability to the task of discourse relation identification in dialogues. Comparing the same DER model on PDTB, the large drop in F1 score shows the difference between formal and informal data. We also find that the model with dialogue features enhance the performance by 1% on F1 score with p-value 0.006, which indicates the potential of using dialogue features to further enhance discourse relation identification models.

6 Conclusion and Future Work

In this paper, we proposed a novel pipeline specifically designed for implicit discourse relation identification in open-domain dialogue. We constructed a novel dataset of discourse relation pairs for dialogue conversations, and utilized unique dialogue features to enhance the performance of a state-of-the-art classifier. Our experiments show that dialogue intent and entities types play important roles and dialogue features can increase the performance of the discourse relation identification model.

Since implicit discourse relation identification is a key task for dialogue systems, there are still many approaches worth investigating in future work. More sophisticated dialogue features and classification algorithms are needed for the discourse relation identification task in addition to a larger more balanced corpus.

References

Hongxiao Bai and Hai Zhao. 2018. Deep enhanced representation for implicit discourse relation recognition. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 571–583.

Kevin K Bowden, Shereen Oraby, Jiaqi Wu, Amita Misra, and Marilyn Walker. 2017. Combining

search with structured data to create a more engaging user experience in open domain dialogue. *arXiv preprint arXiv:1709.05411.*

Kevin K Bowden, Jiaqi Wu, Wen Cui, Juraj Juraska, Vrindavan Harrison, Brian Schwarzmann, Nick Santer, and Marilyn Walker. Slugbot: Developing a computational model and framework of a novel dialogue genre.

Kevin K Bowden, Jiaqi Wu, Shereen Oraby, Amita Misra, and Marilyn Walker. 2018a. Slugbot: An application of a novel and scalable open domain socialbot framework. *arXiv preprint arXiv:1801.01531.*

Kevin K Bowden, Jiaqi Wu, Shereen Oraby, Amita Misra, and Marilyn Walker. 2018b. Slugnerds: A named entity recognition tool for open domain dialogue systems. *arXiv preprint arXiv:1805.03784.*

Chloé Braud and Pascal Denis. 2014. Combining natural and artificial examples to improve implicit discourse relation identification. In *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*, pages 1694–1705.

Zeyu Dai and Ruihong Huang. 2018. Improving implicit discourse relation classification by modeling inter-dependencies of discourse units in a paragraph. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, volume 1, pages 141–151.

Joachim Fainberg, Ben Krause, Mihai Dobre, Marco Damonte, Emmanuel Kahembwe, Daniel Duma, Bonnie Webber, and Federico Fancellu. 2018. Talking to myself: self-dialogues as data for conversational agents. *arXiv preprint arXiv:1809.06641.*

Eric N Forsyth and Craig H Martell. 2007. Lexical and discourse analysis of online chat dialog. In *International Conference on Semantic Computing (ICSC 2007)*, pages 19–26. IEEE.

Fengyu Guo, Ruifang He, Di Jin, Jianwu Dang, Longbiao Wang, and Xiangang Li. 2018. Implicit discourse relation recognition using neural tensor network with interactive attention and sparse learning. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 547–558.

Yanyan Jia, Yuan Ye, Yansong Feng, Yuxuan Lai, Rui Yan, and Dongyan Zhao. 2018. Modeling discourse cohesion for discourse parsing via memory network. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, volume 2, pages 438–443.

Chandra Khatri, Behnam Hedayatnia, Anu Venkatesh, Jeff Nunn, Yi Pan, Qing Liu, Han Song, Anna Gottardi, Sanjeev Kwatra, Sanju Pancholi, et al. 2018. Advancing the state of the art in open domain dialog systems through the alexa prize. *arXiv preprint arXiv:1812.10757.*

Ben Krause, Marco Damonte, Mihai Dobre, Daniel Duma, Joachim Fainberg, Federico Fancellu, Emmanuel Kahembwe, Jianpeng Cheng, and Bonnie Webber. 2017. Edina: Building an open domain socialbot with self-dialogues. *arXiv preprint arXiv:1709.09816.*

Junyi Jessy Li and Ani Nenkova. 2014. Addressing class imbalance for improved recognition of implicit discourse relations. In *Proceedings of the 15th Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGDIAL)*, pages 142–150.

Ziheng Lin, Min-Yen Kan, and Hwee Tou Ng. 2009. Recognizing implicit discourse relations in the penn discourse treebank. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 1-Volume 1*, pages 343–351. Association for Computational Linguistics.

Christopher Manning, Mihai Surdeanu, John Bauer, Jenny Finkel, Steven Bethard, and David McClosky. 2014. The stanford corenlp natural language processing toolkit. In *Proceedings of 52nd annual meeting of the association for computational linguistics: system demonstrations*, pages 55–60.

Daniel Marcu and Abdessamad Echihabi. 2002. An unsupervised approach to recognizing discourse relations. In *Proceedings of the 40th annual meeting of the association for computational linguistics.*

Eleni Miltsakaki, Livio Robaldo, Alan Lee, and Aravind Joshi. 2008. Sense annotation in the penn discourse treebank. In *International Conference on Intelligent Text Processing and Computational Linguistics*, pages 275–286. Springer.

Matthew E. Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, and Luke Zettlemoyer. 2018. Deep contextualized word representations. In *Proc. of NAACL.*

Emily Pitler, Mridhula Raghupathy, Hena Mehta, Ani Nenkova, Alan Lee, and Aravind K Joshi. 2008. Easily identifiable discourse relations. *Technical Reports (CIS)*, page 884.

Rashmi Prasad, Nikhil Dinesh, Alan Lee, Eleni Miltsakaki, Livio Robaldo, Aravind Joshi, and Bonnie Webber. 2008. The penn discourse treebank 2.0. In *Proceedings of the 6th International Conference on Language Resources and Evaluation.*

Lianhui Qin, Zhisong Zhang, and Hai Zhao. 2016. A stacking gated neural architecture for implicit discourse relation classification. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 2263–2270.

Ashwin Ram, Rohit Prasad, Chandra Khatri, Anu Venkatesh, Raefer Gabriel, Qing Liu, Jeff Nunn, Behnam Hedayatnia, Ming Cheng, Ashish Nagar, et al. 2018. Conversational ai: The science behind the alexa prize. *arXiv preprint arXiv:1801.03604.*

671

Attapol Rutherford and Nianwen Xue. 2015. Improving the inference of implicit discourse relations via classifying explicit discourse connectives. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 799–808.

Richard Socher, Alex Perelygin, Jean Wu, Jason Chuang, Christopher D Manning, Andrew Ng, and Christopher Potts. 2013. Recursive deep models for semantic compositionality over a sentiment treebank. In *Proceedings of the 2013 conference on empirical methods in natural language processing*, pages 1631–1642.

Caroline Sporleder and Alex Lascarides. 2008. Using automatically labelled examples to classify rhetorical relations: An assessment. *Natural Language Engineering*, 14(3):369–416.

Amanda Stent. 2000. Rhetorical structure in dialog. In *INLG'2000 Proceedings of the First International Conference on Natural Language Generation*.

Sara Tonelli, Giuseppe Riccardi, Rashmi Prasad, and Aravind K Joshi. 2010. Annotation of discourse relations for conversational spoken dialogs. In *LREC*.

Xun Wang, Sujian Li, Jiwei Li, and Wenjie Li. 2012. Implicit discourse relation recognition by selecting typical training examples. *Proceedings of COLING 2012*, pages 2757–2772.

Ben Wellner, James Pustejovsky, Catherine Havasi, Anna Rumshisky, and Roser Sauri. 2006. Classification of discourse coherence relations: An exploratory study using multiple knowledge sources. In *Proceedings of the 7th SIGdial Workshop on Discourse and Dialogue*, pages 117–125.

Yang Xu, Yu Hong, Huibin Ruan, Jianmin Yao, Min Zhang, and Guodong Zhou. 2018. Using active learning to expand training data for implicit discourse relation recognition. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 725–731.