論 文 の 内 容 の 要 旨

Abstract

論文題目　Learning Semantic Textual Relatedness using Natural Deduction Proof

（自然演繹に基づく論理推論を用いた文間関連性の評価）

氏　　名　谷中　瞳

1. Introduction

Capturing a semantic relation between two sentences is one of the most important core problems in natural language processing (NLP). This problem is closely related to many NLP applications such as question answering, information retrieval or text summarization. There are two tasks for capturing a semantic relation between two sentences: recognizing textual entailment (RTE) and semantic textual similarity (STS). RTE aims to judge whether one sentence logically entails or contradicts another, whereas STS aims to measure the graded similarity between two sentences. In other words, RTE is a three-way (entailment, contradiction, and neutral) classification task, whereas STS is a graded classification/regression task. To capture textual entailment and similarity accurately, we should consider how to represent and calculate two kinds of word meanings: the lexical meanings of content (lexical) words and the functional meanings of logical or functional (grammatical) words. To capture a semantic relation between two sentences, there are three main approaches: machine learning-based approaches, logic-based approaches and their hybrid approaches.

In machine learning-based approaches, vector-based representation models have been used to compare and rank words, phrases, and sentences using various relatedness scores. However, these models often use shallow information such as words and characters, and it remains unclear whether they are capable of accounting for functional meanings of sentences such as negations and connectives.

By contrast, logic-based approaches that use logical inference systems have been successful in representing the functional meanings of sentences as logical formulas; they have thus had a positive effect on the RTE task. However, purely logic-based approaches assess only entailment or contradiction relations between sentences and do not offer graded notions of semantic similarity. In addition, previous logic-based approaches have failed to account for lexical relations between phrases due to their strict logical inference.

To have advantages over both of logic-based and machine learning-based approaches, hybrid approaches are proposed for learning both textual entailment and similarity. Hybrid approaches use logic-based features derived from the proof results of first-order theorem proving combined with shallow features such as sentence lengths. While hybrid approaches open a door to a fusion of logic and machine learning to predict textual entailment and similarity, previous hybrid approaches use only proof results as logic-based features and thus a more effective way to combine logic-based approaches with machine learning-based approaches should be explored.

Considering the conception of proof-theoretic semantics, not only proof results but also the theorem proving processes represent semantic relations between sentences. The idea is that the careful treatment of proof processes obtained from logical inference system is useful for capturing semantic relations more precisely. The contributions of this thesis are summarized as follows. First, I propose a new hybrid method of learning semantic textual similarity from proof processes to achieve high prediction performance for the STS task. Second, I show that my hybrid approach is general and also can be applied to learning textual entailment, thereby obtaining the highest performance for the RTE task. Third, to improve our logical inference system, I propose the phrase abduction mechanism, which detects a lack of phrasal knowledge in logical inference during the proof processes.

2. Learning textual entailment and similarity using proof processes

In this research, I develop a new hybrid approach to learning textual similarity and entailment by using features based on proof processes of bidirectional entailment relations between sentences. I capture the semantic relation between two sentences as a function of the provability of bidirectional entailment relations for a sentence pair and combine it with shallow features. First, a sentence pair is mapped to logical formulas as semantic representations. To convert sentences into logical formulas, the sentences are parsed into syntactic trees based on Combinatory Categorial Grammar (CCG). The meanings of words are described using lambda terms and semantic representations are obtained by combining lambda terms in accordance with meaning composition rules specified in a CCG syntactic tree.

Then, our logical inference system attempts a natural deduction proof, aiming to prove bidirectional entailment relations. One formula is set to the premise and another formula is set to the conclusion. Then, the premise is decomposed into a pool of premises and the conclusion is decomposed into a set of formulas to be proved (called sub-goals). The proof is performed by searching for a premise whose predicate matches that of a sub-goal. If such a premise is found, then the sub-goal is removed. If all sub-goals are removed, we succeed in proving the entailment relation. If we fail to prove entailment relations, then we check whether one sentence contradicts another, which amounts to proving the negation of the original conclusion. Next, if we fail to prove entailment or contradiction, we then attempt to prove bidirectional entailment relations using word axiom injection (abduction), which generates the axiom that is necessary to fill the gap between the premise and the conclusion with lexical knowledge. In this process of word abduction, I combine two kinds of lexical knowledge: a lexical database and a word embedding model. The axiom candidates whose score is the highest and exceed the threshold are selected as axioms to inject into the proof. Again, if the proofs fail, we attempt to prove the negation of the conclusion using word abduction. If word abduction fails because of a lack of lexical knowledge, we can obtain partial proof information by simply accepting the unproved sub-goals and forcibly completing the proof.

After the proof is completed, details about proving processes such as generated axioms and skipped sub-goals are used as features. These features are used for learning semantic similarity and textual entailment with a random forest model. I combine logic-based features extracted from the proof with non-logic-based features. I propose 15 logic-based features from which 12 of those are derived from the bidirectional natural deduction proofs. That is, six features are extracted from the direct proof and another six from the reverse proof; the remaining three features are derived from semantic representations of a sentence pair. Regarding non-logic-based features, I propose nine features such as the string similarity between sentences.

To evaluate my hybrid approach, I used standard datasets provided for the RTE and STS tasks. The datasets were originally developed for evaluating compositional distributional semantics, and they contain logically challenging expressions such as quantifiers, negations, conjunctions, and disjunctions. My system achieved higher prediction performance than previous hybrid approaches on the STS task. Furthermore, my system achieved the state-of-the-art accuracy in the RTE task. The case analysis indicates that my approach has the capability of reflecting the syntactic structures or logical relations in sentences, which causes the state-of-the-art performance. Furthermore, the feature ablation study in both STS and RTE tasks indicates that logic-based features have more impact on improving the performance than non-logic-based features. In particular, the results indicate that proof processes are effective features for learning textual entailment and similarity. Lastly, I point out that the Pearson correlation of the “neutral” portion of the dataset on the STS task was 0.766, which suggests that my hybrid approach to learning textual similarity can be applied not only to sentence pairs with the entailment/contradiction label but also to sentence pairs with the neutral label.

3. Phrase Abduction

The error analysis in the previous section indicates that handling phrasal lexical knowledge in logical inference is a crucial problem within my approach. This problem is general in logic-based approaches, and there are three main difficulties that prevent effective identification and use of phrasal lexical knowledge in logical inference. The first difficulty is the presence of out-of-context phrase relations in previous lexical databases. The second difficulty is finding semantic phrase correspondences between the relevant text segments. The third difficulty is the intrinsic lack of coverage of databases for logical inference despite their large size.

To tackle these three problems, I propose an automatic phrase abduction mechanism to inject phrasal knowledge during the proof construction process. In addition, I consider multiple alignments by backtracking the decisions on variable and predicate unifications, which is a more flexible strategy. I represent logical formulas using directed acyclic graphs, since this is a general formalism that is easy to visualize and analyze. Then, I formalize a theorem proof routine and variable unification using graphs. I detect phrase-to-phrase semantic relations between the sentences by finding alignments between the subgraphs of their meaning representations during the proof process for proving the entailment relation or contradiction between the sentences.

I evaluate the utility of my phrase abduction mechanism in logical inference for the RTE task using the dataset. The evaluation results demonstrated that my method of detecting phrase alignments on semantic representations automatically detects various phrase correspondences, including antonym phrases and non-contiguous phrases, which is one of the main advantages over previous shallow or syntactic-based methods for phrase alignments. Furthermore, the evaluation results also demonstrated that my phrase abduction mechanism compensated for the lack of phrasal knowledge in logical inference and achieved the best performance for the RTE task among purely logic-based approaches.