

複数文からなる文章読解タスクへのテキスト含意認識の適用

笠原 要¹, 平 博順¹, 永田 昌明¹, 柴田 知秀², 黒橋 禎夫²¹NTT コミュニケーション科学基礎研究所 {kasahara.kaname, taira.hiroshi, nagata.masaaki} @lab.ntt.co.jp²京都大学大学院情報学研究所 {shibata, kuro} @i.kyoto-u.ac.jp

1. はじめに

テキストの汎用的な意味処理方法の1つとして Recognizing Textual Entailment (テキスト含意認識、以下、RTE と略記) が提唱された[1]。与えられた文章(テキスト T) に対して別の文(仮説 H) の意味を含んでいる/いないかをシステムで判定させるものであり、幅広い応用が期待されている。本稿では、Reading Comprehension (文章読解、以下、RC と略記) タスクに RTE を応用可能か検討するために段階的な読解問題の作成を試みて評価した結果について報告する。

2. 文章読解タスク

RC タスクとは、国語や外国語学習者の文章読解力を計るためのテストをシステムで自動解答するものである。テストには文章(問題文章)があり、学習者はこれを読んで関わる質問に対して選択肢、単語、文章等の指定された形式で解答する。第一回の RTE チャレンジ[1]では、情報検索や質問応答タスクとともに RC 応用を考慮したテストデータが作成されている。新聞記事のテキストに対して、高校生レベルの読解力で包含関係を判定できる文が作成され、仮説(H)とされている。2回目以降のチャレンジでは、テストデータ作成の応用先として RC 応用は想定されていない。一方、チャレンジで扱うテキストは1文程度と短いものであったが、RTE-3[2]からはテストデータの一部に段落レベルのテキストが含まれるようになってきている。テキスト中の複数文を考慮することが必要となっており、文章読解へ応用も可能と考えられる。

そこで、高校生を想定した4択式の読解問題を作成し、RTE方式のベースラインとして考えられる bag-of-words モデル[3]で検討を行った[4]。その結果、文単位に分割して比較した結果を総合するよりも、複数の文を1つにまとめて比較した場合のほうが、良い正答率となった。この原因として、個々の文のみの解析では誤った選択を導きやすい高度な“ひっかけ問題”を全ての問題の選択肢について含めたことが影響していると思われる。問題文及び、質問の選択肢の1例をあげる。

[京都観光]紅葉真っ盛りです

大学構内もすっかり紅くなって、きれいですね。市内の名所に見に行きたいものですが、NFやら授業やらでなかなか時間がとれそうにありません。時間がとれたころには散っていた、なんてことにならないといいのですが。個人的にはここからが本題。大学に入学してからずっと疑問に思っていたのですが、高校生や中年女性の団体がバスで大学にやってきて時計台の前で記念撮影などをしているのは、京

都大学が京都の観光名所の一部になっているからなんですか。どうも観光、特に京都観光となると寺や神社なんかを思い浮かべるので、大学が観光地であるとすれば、それはとても意外で新鮮な感じがします。考えてみれば京都大学はそれなりに歴史のある大学で、見所もあるものだけでも、それが遠くからはるばるやってきた人たちの巡回ルートの中に入るというのは自分の頭の中にはない発想で、それはつまり京都大学と金閣銀閣清水寺その他が肩を並べているとも言えられるのではないかと思います。おそらく大学が自由でオープンな雰囲気を出しているのもこの現象の理由のひとつなのではないかと考えますが、これをどうとらえていいのか自分にはよくわかりません。今日の構内の紅葉はきれいだったので、たくさんの人が見て、きれいだなあという感覚を共有したい気持ちはありますが……。

表1 選択肢の例

| 選択肢 | 仮説 | 含意判定 |
|-----|--------------------------|-------|
| 1 | 観光バスが京都大学に来ることがある。 | TRUE |
| 2 | 大学内の紅葉は終わっている。 | FALSE |
| 3 | 学生として京都大学に来る高校生や中年女性が多い。 | FALSE |
| 4 | 著者は市内の名所をよく見にいっている。 | FALSE |

問題文を含意する選択肢1については、「団体がバスで大学にやってくる」箇所と「大学が京都の観光名所の一部になっている」を総合して判定する必要がある。一方、包含しない選択肢4については、「市内の名所に見に行ってみよう」という表現が関連性が高いが、続く「なかなか時間がとれそうにありません。」から時間がとれないのでよく見にいっていないという因果関係まで推論する必要がある。このような複雑な問題では、テキスト含意認識の方法が適切であるかがなかなかわかりづらいので、難易度が低い問題から適宜評価、改善することが必要と考えた。

そこで本稿では、解析の観点では小学生レベルの問題を新たに作成し、bag-of-wordsのRTE方式とともに述語項構造解析に基づく方法を比較評価した。

3. 実験

3-1 読解問題の作成

文献[4]での検討と同じ、NTTと京大で共同作成したKNBコーパス(Kyoto-University and NTT Blog コーパス)[5]の京都観光に関する記事41件を用い、初級の読解問題を作成した。

小学校1-2年の国語指導要綱で、「文の中における主語と述語との関係に注意すること。」とある[6]。文法能力として初級レベルで求められるものと考え、複数文の中で1文に着目し、その中に含まれる述語と名詞を組み合わせさせて4つの選択肢中で1つのみが含意する質問文を作成した。例を表2にあげる。

表2 初級の質問文

| 選択肢 | 仮説 | 含意判定 |
|-----|--------------------|-------|
| 1 | 中年女性の団体は記念撮影などしている | TRUE |
| 2 | 中年女性の団体は疑問に思っていた。 | FALSE |
| 3 | 大学は疑問に思っていた。 | FALSE |
| 4 | 大学は記念撮影などしている。 | FALSE |

これは、下記の問題文中の1文から生成したものである。この含意を **bag-of-words** モデルで判定した場合、どれも近いスコアを与える可能性がある反面、構文解析や述語項構造解析の結果を考慮する高度な RTE システムでは適切に判定される可能性がある評価セットとなっている。

大学に入学してからずっと疑問に思っていたのですが、高校生や中年女性の団体がバスで大学にやってきて時計台の前で記念撮影などをしているのは、京都大学が京都の観光名所の一部になっているからなんですか。

3-2 RTE システム

適用するテキスト含意認識方法を比較するため、小谷らの研究[7]に基づいて作成された RTE システムを用いた。テキスト・仮説それぞれに対して構文・格・省略解析を行ない、解析結果に基づき述語項構造に正規化し、正規化された述語項構造を単位としてマッチングをとることにより、テキスト間の含意関係を認識する方法である。出力結果は含意する/しないの2値のみである。

3-3 実験結果

まず、過去に作成した高校生レベルの問題についての評価結果を表4に示す

表4 読解問題回答評価(高校生レベル)

| RTE 方式 | 正答率 (選択肢) | 正答率 (問題) |
|-----------------|--------------|-------------|
| 述語項比較[7] | 0.799 | 0.024 |
| Bag-of-words[3] | 0.665 | 0.293 |

Bag-of-words の方式では、訓練データを用いて含意する/しないを与える BLEU スコアの閾値を決定するが、その値は、問題の長さで変動するため、本稿では、相対値として利用し、最も高い値の選択肢のみを含意すると出力している。そのような条件を活用しても、選択肢レベルの判定では、述語項を比較する方法[7]が評価値が高まっている。但し、問題単位の正答率では正解の条件が大きく影響するので、それを利用した **Bag-of-words** の方式のほうが評価値が高くなっている。述語項の比較の方法でも、複数の選択肢が含意する場合には1つを選ぶ等行えば、正答率が高まると予想される。

次に小学生レベルの問題での評価結果を表5に示す。与えるテキストとして、問題の複数の全文全てを与えた場合と、選択肢を作成した1文を取り出して与えた場合双

方を比較した。選択肢単位の正答率では、述語項の比較に基づく RTE 方式が高い値となった。また、テキストとして全文を与えた場合も1文のみの場合も同一の判定結果を与えているため、テキストの長さに関して安定な処理ができる方法であると予想される。

表5 読解問題回答評価(小学生レベル)

| RTE 方式 | テキスト | 正答率 (選択肢) | 正答率 (問題) |
|---------------------|------|--------------|-------------|
| 述語項比較 [7] | 全文 | 0.726 | 0.171 |
| | 1文 | 0.726 | 0.171 |
| Bag-of-words [3] | 全文 | 0.632 | 0.293 |
| | 1文 | 0.622 | 0.293 |

4. おわりに

本稿では、複数文からなる読解問題へのテキスト含意認識の適用方法を検討する一環として、複数のレベルの問題を作成し、**bag-of-words** での含意認識方法と述語項に基づく方法を比較評価し、問題によって評価結果が変わる場合があることを確認した。今後は、複数段階の問題の作成を行いながら、テキスト含意認識結果を分析し、拡張方法の検討を行う予定である。

参考文献

- [1] Ido Dagan, Oren Glickman, and Bernardo Magnini. The PASCAL Recognising Textual Entailment Challenge. In Quiñero-Candela et al., editor, MLCW 2005, LNAI Vol. 3944, pp. 177-190, Springer-Verlag, 2006.
- [2] Danilo Giampiccolo, Bernardo Magnini, Ido Dagan, and Bill Dolan. The Third PASCAL Recognizing Textual Entailment Challenge, Proc. of the ACL-PASCAL Workshop on Textual Entailment and Paraphrasing, pp. 1-9, 2007.
- [3] Diana Pérez and Enrique Alfonseca. Using Bleu-like Algorithms for the Automatic Recognition of Entailment. In Quiñero-Candela et al., editor, MLCW 2005, LNAI Vol. 3944, pp. 191-204, Springer-Verlag, 2006.
- [4] 笠原 要, 平 博順, 永田 昌明: 複数文からなる文章読解タスクへのテキスト含意認識の適用可能性の検討, 言語処理学会第16回年次大会, 2010.
- [5] 橋本 力, 黒橋 禎夫, 河原 大輔, 新里 圭司, 永田 昌明. 構文・照応・評判情報つきプロコーパスの構築. 言語処理学会第15回年次大会, pp.614-617, 2009.
- [6] 文部科学省: 小学校学習指導要領 第1節 国語
http://www.mext.go.jp/b_menu/shuppan/sonota/990301/03122601/002.htm, 平成10年12月告示、15年12月一部改正.
- [7] 小谷 通隆, 柴田 知秀, 黒橋 禎夫: 言い換え表現の述語項構造への正規化とテキスト含意関係認識での利用, 言語処理学会第15回年次大会, pp.260-263, 2009.