

述語項構造解析を用いた英語長文読解問題の自動解法

堯天貴之¹, 植田佳文¹, 東中竜一郎², 杉山弘晃², 平博順¹

¹ 大阪工業大学 情報科学部, ² NTT コミュニケーション科学基礎研究所

{e1c12024, m1m15a03}@st.oit.ac.jp

{higashinaka.ryuichiro, sugiyama.hiroaki}@lab.ntt.co.jp

hirotoshi.taira@oit.ac.jp

1 はじめに

我々は、国立情報学研究所が主導する「ロボットは東大に入れるか」プロジェクト [1] の英語問題に取り組んでおり、センター試験および東大二次試験において高得点を実現を目指している。2014年には我々が作成したシステムが「東ロボくん」として代ゼミセンター模試を受験し、全受験者の平均である 93.1 点 (200 点満点) を超える 95 点を達成した [2]。しかし長文問題に関しては 101 点中 32 点の低い得点に留まっており、長文問題における得点力向上が 1 次試験突破に向けて大きな課題となっている [3]。そこで本研究では、長文問題のうち現在の技術で得点力向上の余地があると思われる内容説明問題について、問題本文と選択肢文に対する述語項構造解析および単語類似度を利用した手法を提案し、簡単な評価結果を示す。

2 内容説明問題の概要

近年の大学入試センター試験における英語筆記試験は、以下のような問題構成となっている。

大問 1 発音・アクセント

大問 2 文法・語法・会話・語句整序

大問 3 未知語 (句) 語意推測・不要文除去・意見把握

大問 4 図表読解

大問 5 状況把握読解

大問 6 長文読解

さらに大問 6 は、最初に与えられた本文 (数百語からなる長文) に対し、本文と内容が合致するものを選択肢の内容説明問題 (6A) と本文の各段落に見出しをつける問題 (6B) から構成されている。

内容説明問題 6A の例として、図 1 に平成 26 年度 (2014 年度) 大学入試センター試験本試験における大

問 6 の問題本文の冒頭部分と問題 6A 冒頭の問題を示す。大問 6 の問題本文は通常 5~8 段落 (全体で 30 文~50 文) 程度の長さがあり、4 択問題が通常 5 問出題される。基本的に段落指定をされることが多いが本文全体を対象にした問題が出題されることもある。また、問題によっては「本文中で述べられていないこと」を解答させる場合もある。

この例の場合は、6 段落で構成された「オーディオ機器の音質」に関する本文 (647 語) に対し、問 1~5 でそれぞれ段落 (1), (3), (4), (5), (6) の内容について問われる問題があり、特に問 1 では、段落 (1) の内容について、ベル研究所の蓄音機がエジソンの蓄音機よりもどうであるかを問う問題になっており、正解は選択肢 4 の「より現実に近い音が再生できる」である。

この問題を正解するためには、1 文~数文の本文の該当箇所に対し、選択肢が妥当な説明となっているか否かを判定する必要があるが、現在のところ、意味的にはほぼ同じ内容であることの判定できれば説明の妥当性が判断できるという仮定のもとにシステムを作成している。

3 提案手法

提案手法における解析手順の概要を図 2 に示す。6A の内容説明問題では、まず問題選択肢と該当段落の内容が意味的にどのくらい近いかを測る必要がある。選択肢は 1 文程度と短く、正解の選択肢を導くための本文中の文章もほとんどの場合が 1~2 文程度と見られるため、今回は該当段落の各文と選択肢の文を総当りで類似度を算出し、最も高い類似度を持つ組み合わせがあった選択肢をシステムの回答とした。文同士の類似度は、述語項構造解析と word2vec を用いた単語類似度、否定表現の有無、推定表現の有無を用いて計算した。

【6】 次の文章を読み、下の問い (A・B) に答えよ。なお、文章の左にある (1)～(6) は段落の番号を表している。(配点 36)

(1) In 1877, Thomas Edison invented the phonograph, a new device that could record and play back sound. For the first time, people could enjoy the musical performance of a full orchestra in the convenience of their own homes. A few years later, Bell Laboratories developed a new phonograph that offered better sound quality; voices and instruments sounded clearer and more true-to-life. These early products represent two major (中略)

The advances over the years have been significant in both areas, but it is important not to let the music itself get lost in all the technology.

(2) Although the phonograph made listening to (以下略)

A 次の問い (問 1～5) の [47]～[51] に入れるのに最も適当なものを、それぞれ下の ①～④のうちから一つずつ選べ。

問 1 According to paragraph (1), Bell Laboratories' phonograph could [47] than Thomas Edison's.

① be build more quickly and cheaply
 ② be operated with less difficulty
 ③ play more musical instruments
 ④ reproduce sound more realistically

図 1: 内容説明問題の例

3.1 述語項構造解析と単語類似度による意味同一性推定

図 3 に示すように、本文中の 1 文と選択肢の 1 文の文間類似度を計算する際には、一旦各文を述語項構造解析し、(Arg1, Verb, Arg2, Arg3) の組を抽出したあと、項タイプ (Arg1～3) および動詞 (Verb) ごとに分けて単語類似度を計算し、その合計値を文間類似度とした。述語項構造の解析には、英語の HPSG パーザである Enju [4, 5] を使用した。また単語類似度の計算には word2vec [6] を用いた。

各ラベル名で格納した本文と選択肢の単語の一致度をラベル名ごとに計算して合計値を求め、最も値が大きな選択肢を解答として出力する。

この際、段落指定がある問題については該当段落の文と選択肢との一致度を計算し合計値を比較する。段

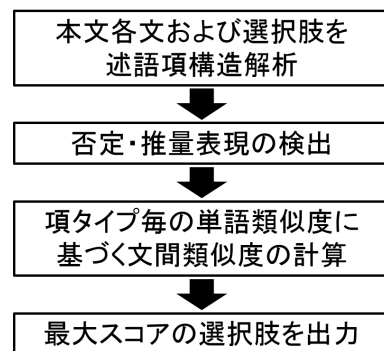


図 2: 解析手順の概要

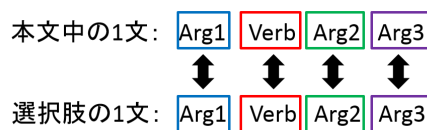


図 3: 項タイプ毎の単語類似度の計算

落指定がない問題は全段落の文と選択肢との一致度を計算し合計値の比較を行う。

3.2 否定表現と推量表現の考慮

否定表現を含む文と含まない文とでは、述語項構造レベルで類似度が高くても異なる内容であると考えられる。そのため、与えられた文が否定表現を含むか否かを特定し、否定表現の有無が一致しない場合は、文間類似度にペナルティを与えた。ペナルティの値は学習データを用いて実験的に求めた。

推量表現については、推量表現を含む文では曖昧性が増し、類似度に影響を与えられられる。そのため、与えられた文が推量表現を含むか否かを特定し、推量表現の一致度に応じたスコアを文間類似度に加味した。スコアの値は否定表現の場合と同様実験的に求めた。

4 評価実験

4.1 実験設定

今回の実験では、大学入試センター試験の過去問 11 回分 (1991～2013 年度の奇数年度 (2009 年度, 2011 年度を除く) および 2014 年度の本試験), および代ゼミセンター模試 5 回分 (2013 年第 1～4 回および 2014 年第 1 回), ベネッセ模試 3 回分 (2014 年 6 月, 9 月, 11 月) を評価対象とした。各試験問題は人手で XML

フォーマットで入力された電子化データで与えられる。なお、6Aの問題で段落が指定されるようになったのは2009年度以降であり、それ以前の問題は、段落指定がなく本文全体に対して内容説明を選択する問題である。本データでは、本試験11回のうち古い方の9回分がもともと段落指定が無い問題である。

解析システムにおいては、英語の述語項構造解析器としてEnju (Ver. 2.4.2)を使用した。また、word2vecの辞書データとしてはGoogle Newsデータから学習された300万単語に対応したものを用了。

また、我々の提案法では基本的に問題文で指定された段落の本文のみ選択肢との文間類似度を計算して解答を導いている。段落の情報がどの程度精度に影響しているかを調べるため、問題文で段落が指定されていてもその情報を無視して本文全体について文間類似度を計算する方法についても実験を行った。

4.2 実験結果

表 1: 評価結果

試験	段落指定あり	段落指定なし
本試験 (11 回分)	28% (16/57)	32% (18/57)
代ゼミ (5 回分)	40% (10/25)	36% (9/25)
ベネッセ (3 回分)	60% (9/15)	60% (9/15)
合計	36% (35/97)	37% (36/97)

今回の実験結果を表1に示す。全体的にはランダムに選択肢を選んだ場合のベースライン (25%) よりは高い解析精度が得られている。また、段落指定の有無による影響はほとんどなかった。なお、試験により精度に差が出ているが、今回実験に使用した問題数が異なるため、この結果が直接、試験の難易度を表すものではない。

4.3 解析誤りの分析

今回の実験で正しく解答できなかった問題について分析を行った。その結果、以下のような場合があることが分かった。

4.3.1 単語レベルを超える言い換えがある場合

提案手法は、同じ内容を表す文は類似の単語を用いて構成されているという仮定に基づいた手法であるが、本文該当箇所の表現と選択肢の表現がかなり異なる

ベネッセ模試 2014 年度 11 月 6A 問 1
本文該当箇所 : Spending time on social networking websites, one can see how much people love sharing stories of their lives.
問題文 : According to (1), it is common to 47 through social networking services.
正解選択肢 : share one's experiences with others

る場合があった。上記の例では、本文の“one can see how much people love sharing stories of their lives.”と選択肢の“it is common to share one's experiences with”の内容がほぼ同じであることを認識する必要があるが、表層的な単語だけをみると一致している単語がほとんどないことが分かる。単語レベルでの言い換えを超えたこのような言い換えがされている場合は、今回の手法では正解を得ることは難しい傾向にあった。

4.3.2 名詞句の解析が必要な場合

センター本試験 2007 年度, 6A 問 4
本文該当箇所 : Because of his rare color he had become quite famous.
問題文 : Why was Snowflake a popular exhibit at the zoo? 49
正解の選択肢 : He was an unusual color.

ベネッセ模試 2014 年度 9 月 6A 問 2
本文該当箇所 : Interestingly, growing numbers of foreign students have accompanied changes in U.S. general education.
問題文 : In paragraph (3), the author suggests changes in general education at U.S. universities are happening along with changes in 48 .
正解の選択肢 : the ratio of foreign students

上記のように本文、選択肢いずれか、または両方の内容が名詞句で表現されている場合、今回の (Arg1, Verb, Arg2, Arg3) の組で照合を行う方法では、中心となる動詞が取り出せず解析ができなかった。また、選択肢が文を構成していない場合は述語項構造解析自体がうまくいかず正答できないことがあった。

4.3.3 単語の表層形が異なる場合

代ゼミ模試 2013 年度第 2 回, 6A 問 1
本文該当箇所 : When they find good flowers and come back to their hive and meet their peers, they “dance.”
問題文 : According to paragraph (2), honeybees dance in order to <input type="text" value="46"/> .
正解の選択肢 : tell their fellows about flowers they have found

上記の動詞の過去分詞と現在形のように表層形が異なることで、現在使用している word2vec の設定では類似度スコアがそれほど高い値が得られない場合がある。例えば、‘find’ と ‘found’ について word2vec の単語類似度を計算したところ、0.60 の値しか得られなかった。今後、正規化した単語における類似度なども考慮して単語類似度を計算することも検討したい。

4.3.4 該当箇所が複数の文に分散して記述されている場合

センター本試験 2007 年度, 6A 問 3
本文該当箇所 : “His name was Snowflake,” he continued, “and he was a gorilla , a very special albino gorilla , with white fur and pink skin. When only three years old, Snowflake was captured in the forests of Africa and then brought to the zoo. (中略) Grandpa smiled at Valerie and said, “Anyway, maybe it’s best not to plan everything. All kinds of wonderful, unexpected encounters may be waiting for you on your trip.
問題文 : What unexpected experience did Grandpa describe from his first trip to Barcelona? <input type="text" value="48"/>
正解選択肢 : Encountering an unusual gorilla .

上記のように選択肢に合致する内容を表現した単語が本文中では複数の文に分散して存在する場合があった。この場合、いずれかの項タイプでの単語一致度が高くてもその他の項タイプでの単語一致度が低くなり、文全体の一致度が低くなってしまふ。その結果、一致度の合計スコアが他の選択肢より低くなり正解が得られないことが多かった。

5 おわりに

本研究では英語の長文読解における内容説明問題において、本文各文と各選択肢について、項タイプ別の単語類似度を用いることにより正解の選択肢を推定する手法について検討を行った。また、否定、推量表現の有無が考慮できるような工夫も行った。センター試験問題および模擬試験問題で評価したところ、一定の精度向上はあったが、名詞句の扱いや複数文に分散して内容が記述される場合があるなど、改善の余地があることが分かった。今後は名詞句の詳細な解析や複数文と一文との類似度計算などの検討を行っていきたいと考えている。

謝辞

本研究を推進するにあたって、大学入試センター試験問題のデータをご提供下さった独立行政法人大学入試センターおよび株式会社ジェイシー教育研究所に感謝いたします。また、模擬試験データをご提供下さった学校法人高宮学園、株式会社ベネッセコーポレーションに感謝いたします。

参考文献

- [1] 新井紀子. ロボットは東大に入れるか. イースト・プレス, 2014.
- [2] 東中竜一郎, 杉山弘晃, 磯崎秀樹, 菊井玄一郎, 堂坂浩二, 平博順, 南泰浩. センター試験における英語問題の回答手法. 言語処理学会第 21 回年次大会 (NLP2015), 2015.
- [3] 松崎拓也, 横野光, 宮尾祐介, 川添愛, 狩野芳伸, 加納単人, 佐藤理史, 東中竜一郎, 杉山弘晃, 磯崎秀樹, 菊井玄一郎, 堂坂浩二, 平博順, 南泰浩, 新井紀子. 「ロボットは東大に入れるか」プロジェクト: 代ゼミセンター模試タスクにおけるエラーの分析. 自然言語処理, Vol. 23, No. 1, 2016.
- [4] Miyao Yusuke and Tsujii Jun’ichi. Maximum entropy estimation for feature forests. In *Proceedings of the Second International Conference on Human Language Technology Research (HLT)*, pp. 292–297, 2002.
- [5] Takuya Matsuzaki, Yusuke Miyao, and Jun’ichi Tsujii. Efficient HPSG parsing with supertagging and CFG-filtering. In *Proceedings of the Twentieth International Joint Conference on Artificial Intelligence (IJCAI)*, pp. 1671–1676, 2007.
- [6] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. Distributed representations of words and phrases and their compositionality. In *Advances in Neural Information Processing Systems 26*, pp. 3111–3119, 2013.