

センター試験における英語問題の回答手法

東中竜一郎¹, 杉山弘晃¹, 磯崎秀樹², 菊井玄一郎², 堂坂浩二³, 平博順⁴, 南泰浩⁵

¹NTT コミュニケーション科学基礎研究所, ²岡山県立大学, ³秋田県立大学

⁴大阪工業大学, ⁵電気通信大学

1 はじめに

我々は、国立情報学研究所が主導するグランドチャレンジである「ロボットは東大に入れるか」[1]に英語チームとして参画している。本チャレンジの目的は東大に入学できるような人工知能を目指して研究を進めることにより、人工知能や自然言語処理の各領域の統合を推進することである。そして、英語チームの目的は、センター試験および東大二次試験の英語科目において高得点を実現することである。我々は今年の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。

例年のセンター試験と同様、今年の代ゼミセンター模試は全部で大問が6つある。前半の3つの大問が短文問題で、後半が長文問題である。今回、短文問題は99点の配点のうち63点(得点率は63.7%)で、長文問題は101点の配点のうち32点(得点率は31.7%)であった。以降、短文問題のそれぞれについてどのように回答したかについて詳細を述べる。なお、短文問題のうち「文脈に合わない文を推測する問題」については未着手であるため本稿では触れない。長文問題については、現状精度が低いため、現在の回答手法を簡単に説明するとともに、今回の受験で明らかになった問題点について述べるにとどめる。

2 発音・アクセント問題

ここ数年の代ゼミセンター模試でのアクセント問題は、四つの単語選択肢の中から、最もアクセントの強いシラブルの位置が他の選択肢と異なるものを選ぶものがほとんどである。選択肢の例を図1の左側に示す。この問題を解くため、カーネギーメロン大(CMU)で作成された音声認識・合成用の発音・アクセント辞書¹を引く。辞書引きの結果は図中の右側となる。アルファベットは発音記号を表し、数字は0がアクセントのない母音、1がアクセントのある母音を表す。この数字からアクセントのある母音のシラブル位置が分かり、MODERNが解答であることが分かる。

¹<http://www.speech.cs.cmu.edu/>

ADVICE	AEO D V AY1 S
CONTROL	K AHO N T R OW1 L
EVENT	IHO V EH1 N T
MODERN	M AA1 D ERO N

図1: CMUの発音・アクセント辞書を用いたアクセント位置の異なる単語選択

発音辞書に登録されている情報

FLIGHT	F L AY1 T
IDLE	AY1 D AHO L
IGNORANT	IH1 G N ERO AHO N T
LIVELY	L AY1 V L IYO

↓ アラインメント実行

F L I G H T	F L AY T
I D L E	AY D AH L
I G N O R A N T	IH G N ER AH N T
L I V E L Y	L AY V L IY

図2: CMUの発音・アクセント辞書を用いた対象文字の発音の異なる単語選択

発音に関する問題では、単語中の指定された文字の発音が他の単語中の文字の発音と異なるものを探す問題がほとんどである。選択肢の例を図2の上部左側に示す。ここでは、単語中の下線で示すIの発音が異なるものを探す。ここでもCMUの辞書を用いて、単語の発音・アクセントを調べる。しかし、この辞書の情報だけでは指定された文字がどの発音に対応するのかが分からない。そこで、全ての辞書項目と発音のデータからEMアルゴリズムにより、文字と発音の対応関係を統計的に学習し、その統計量を使って、文字と発音のアラインメントを行う[4]。これにより、指定された文字に対応する発音が明らかになる。この例の場合、IGNORANTの文字Iが異なる発音であることが分かる。以上の手法により過去5回の代ゼミのセンター模試で満点を達成できた。

3 文法・語法・語彙問題

文法・語法・語彙問題とは、以下のように、文中に開いている穴に最もふさわしいものを4つの候補の中から選ぶ穴埋め問題である。

I had a severe toothache, so I made 9 with the dentist.
 (1) a promise (2) a reservation
 (3) an appointment (4) an arrangement

この場合、穴 9 に最もふさわしいのは (3) an appointment である。このような穴埋め問題を解くため、統計的機械翻訳で流暢な訳を作るのに使われている言語モデルを利用した。言語モデルを計算するためのコーパスには、One Billion Word Language Modeling Benchmark のトレーニングデータ²を用い、SRILM³で5グラムまで計算した。この言語モデルを用いて、穴の開いた文に候補を入れて完成させた文を ngram コマンドで採点し、logprob が最大になる候補を回答した。

このベースライン手法により、2005 年以降の奇数年のセンター試験の問題 51 問を解いたところ、34 問 (67%) を解けた。今回は、動詞の活用形を表す VBG などの品詞名を、その動詞の直前に疑似語として挿入するという方法 (以下では VPOS 法, Verb Part-of-Speech) を試し、36 問 (71%) 解けたが、統計的有意差はなかった。

今回の模試では、10 問中、ベースライン法で 5 問、VPOS 法により 6 問正解した。VPOS 法で解けなかった 4 問のうち 2 問は、穴が 2 つあいている新傾向の問題で、プログラムの修正により、現在では両方とも解けており、ベースライン法で 7 問、VPOS 法で 8 問正解している。

まだ正解できていない 2 問は、離れた単語の関係を考慮する必要がある問題で、具体的には、問 4 の can't even ... much less と問 6 の never ... without である。現在、このような離れたパターンに対応するため、Dependency Language Model [3] などの利用を検討している。

4 会話文完成問題

会話文完成問題は、以下のように、二人の話者の会話の空所に相応しい文を 4 つの選択肢から選び、会話文を完成させる問題である。この問題では正解は選択肢 (4) となる。

Parker: I hear your father is in hospital.
 Brown: Yes, and he has to have an operation next week.
 Parker: 19 Let me know if I can do anything.
 Brown: Thanks a lot.

選択肢: (1) Exactly, yes. (2) No problem.
 (3) That's a relief. (4) That's too bad.

²<http://www.statmt.org/lm-benchmark/1-billion-word-language-modeling-benchmark-r13output.tar.gz>

³<http://www.speech.sri.com/projects/srilm/>

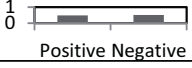
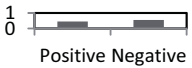
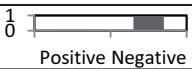
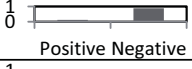
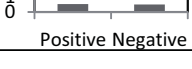
話者	発話	発話意図	感情極性
Parker	I hear your father is in hospital.	表明	 Positive Negative
Brown	Yes, and he has to have an operation next week.	表明	 Positive Negative
Parker	That's too bad.	評価	 Positive Negative
Parker	Let me know if I can do anything.	申し出	 Positive Negative
Brown	Thanks a lot.	感謝	 Positive Negative

図 3: 発話意図と感情極性の流れ

この問題を解くため、4 つの選択肢の各場合について会話文の流れの自然さを推定し、最も自然な流れとなる選択肢を選ぶという方法を用いた。会話文の流れの自然さのスコア Conv は、コーパスに照らして、発話意図 (表明, 評価など) の流れの自然さのスコア Int と感情極性 (ポジティブかネガティブか) の流れの自然さのスコア Sent の重み付き和として算出した。

$$\text{Conv} = (1 - w_1)\text{Int} + w_1\text{Sent}$$

図 3 に先ほどの問題の選択肢 (4) の場合における発話意図と感情極性の流れを示す。発話意図としては、話者の父親の入院と手術についての表明に始まり、選択肢 (4) ではそれに対する評価、続いて援助の申し出と感謝という自然な流れが推定されている。感情極性に関しても、入院や手術に関する発言の感情極性と選択肢 (4) の発言の感情極性がともにネガティブであるという自然な流れが推定されている。

発話意図の流れの自然さのスコア Int を算出するため、CRF を用いて Switchboard Dialog Act Corpus[5] から会話中の発話意図列の識別モデルを学習した。素性は発言中の単語 1-gram, 2-gram を用い、クラスは 43 種類の発話意図を用いた。CRF の実装は CRFsuite[7] を利用した。このモデルにより発話意図列の生起確率を推定し、選択肢ごとの発話意図列の生起確率の和が 1 になるように正規化したものをスコアとした。

感情極性の流れの自然さのスコア Sent を算出するため、SVM を用いて感情極性コーパス [8] から発言の感情極性の識別モデルを学習した。素性は文に含まれる単語 1-gram, 2-gram を用いた。SVM の実装は LIBLINEAR[2] を用い、Log-linear モデルを学習した。このモデルにより会話中の各発言の感情極性がポジティブあるいはネガティブである確率を推定した。感情極性の流れの自然さ Sent は、話者独立な感情極性の流れの自然さ Sent_{indep} と話者内の感情極性の流

れの自然さ $Sent_{within}$ の重み付き和として算出した。

$$Sent = (1 - w_2)Sent_{indep} + w_2Sent_{within}$$

話者独立な感情極性の流れの自然さとは、会話中の隣接発言間で感情極性の差異が小さいことを意味し、話者内の感情極性の流れの自然さとは、特定話者の隣接発言間での感情極性の差異が小さいことを意味する。話者独立な感情極性の流れの自然さ $Sent_{indep}$ は、選択肢ごとに次の式で算出した。ここで、会話には N 個の発言が含まれ、会話中の i 番目の発言の感情極性がポジティブ、ネガティブである確率を P_i, N_i とする。

$$Sent_{indep} = 2.0 - Diff$$

$$ただし, Diff = \frac{\sum_{i=2}^N (|P_{i-1} - P_i| + |N_{i-1} - N_i|)}{N - 1}$$

話者内の感情極性の流れの自然さは、同様の式を使って、話者ごとに隣接発言間の流れの自然さのスコアを算出し、そのスコアの平均値とした。

過去 5 回の代ゼミセンター模試での評価をもとに、重み w_1, w_2 はそれぞれ 0.4, 0.6 とした。本方法により今回の模試では 3 問中 1 問に正答でき、過去 5 回の模試では 15 問中 7 問に正答できた。

5 語句整序完成問題

語句整序完成問題とは、空所を含む文に対して、与えられた数個の単語列を適切に並べ替えて、文法・意味的に正しい文を完成させる問題である。以下に例題を示す。

This problem is too _____ 21 _____
22 _____ ease.

- (1) complex (2) for (3) me
 (4) solve (5) to (6) with

この場合、正しい並びは complex for me to solve with である。このような語句整序完成問題を解くにあたり、我々は、語句の並べ方を全通り作成して元の文にあてはめ、最も流暢な文ができた並び方を回答とするアプローチを採用した。文の流暢さの評価には、文法・語法・語彙問題でも用いられている、言語モデルを用いる。言語モデルを計算するためのコーパスは UMBC webbase corpus⁴ を用い、SRILM⁵ で 5 グラムまで計算した。この言語モデルを用いて、語句を並べ替えてあてはめた文を ngram コマンドで採点し、logprob が最大になる候補を回答とした。

この手法によって、1997 年以降の奇数年度のセンター試験本試を回答したところ、27 題中 20 題 (74%)

⁴<http://ebiquity.umbc.edu/blogger/2013/05/01/umbc-webbase-corpus-of-3b-english-words/>

⁵<http://www.speech.sri.com/projects/srilm/>

George: I must get this paper finished by next Monday.
 Paul: So you can't go dinner with me this weekend.
 George: No, but I'll take a rain check.
 Paul: Sure, how about next weekend?
 未知フレーズ: take a rain check
 選択肢 (1) accept your offer later
 (2) change my mind
 (3) go with you, weather permitting
 (4) refuse the invitation

図 4: 未知語 (句) 語彙推測問題の例

を正答した。また、2012 年、2013 年の代ゼミ模試については、15 題中 13 題 (87%) を正答した。今回の模試では、3 問中 2 問を正答した。不正解であった 1 問に正答するためには、語の抽象化や、not only ... but also のような離れた語句間の関係性の理解、自然な英文ではあまり用いられない受験英語表現への対策が必要と考えられる。

6 未知語 (句) 語彙推測問題

この問題は、出現頻度が低く一般にはあまり知られていないような文章中の単語またはフレーズについて語義を推定し、与えられた選択肢の中から最も意味の近い語義を選択する問題である。今回の模試での問題例を図 4 に示す。

この問題の場合、take a rain check は「延期する」の意味で用いられており正解の選択肢は (1) となる。我々はこの語義推定問題を解くために word2vec [6] およびイディオム辞書を利用した。Google News に出現した 1 千億単語のデータより作られた 300 万の異なり語 (またはフレーズ) に対する 300 次元のベクトルを使用して、未知語またはフレーズと選択肢との類似度を測り、最も高いスコアが得られた選択肢を正解とみなして出力した。また、対象の語句がイディオム辞書に記載されている場合には、一旦イディオム辞書に記載の語義文に変換した後、ベクトル化を行うことで、より高い精度が得られるようになった。今回の模試では 2 問中 2 問正解し、それ以外の過去 5 回の模試も含めた評価では、計 12 問中、9 問 (75%) について正解することができた。

7 意見要旨把握問題

意見要旨把握問題は、以下のように、複数人の議論において、各発言者の意見の要旨を最もよく表す文を 4 つの選択肢から選ぶ問題である。この問題では正解は選択肢 (1) となる。

Saburo: I can't agree with Maki at all. It's cyclists who

should change their minds. Some of them are very reckless and careless ... (省略) ...

Teacher: OK. You've made a very good point, too, Saburo. What you want to say is that cyclists should

33, isn't it?

選択肢: (1) be more sensible and considerate to other road users at all times, (2) be trained in order to control bicycles more rapidly at all times, (3) make artificial sounds at all times to be noticed by other road users, (4) stop using bicycles at once because they are dangerous

議論も会話の一種であり、会話文完成問題と同様の方法で問題を解いた。ただし、意見要旨把握問題では発話意図の流れの自然さは選択肢によって変わらないと想定されるので、感情極性の流れのみを使った。この方法により今回の模試では3問2問正解でき、過去5回の模試では15問中6問に正解できた。

8 長文問題とその難しさ

長文問題は大問で三つある。そのうち二つはイラストやグラフ・表を読み解く問題であり、もう一つは論説文に関する問題である。イラストやグラフ・表と本文の文章を統合して理解し、問題を解くことは現状の技術では難しい。そこで、今回の模試においては、選択肢と本文のみに着目し、選択肢の中で本文との間に最も強く含意関係が認められるものを選ぶことにした。選択肢がイラストの場合は、イラストに補助として付与されていた文章を用いて含意関係認識を行った。論説文についても同様に長文との間で含意関係の強さを評価し、最も強い選択肢を選ぶ方法を採用した。なお、含意認識のためのテキスト・仮説の作成および認識処理には次の処理を行った。1) 問題文において“According to paragraph (3),”のようにパラグラフが指定されている場合には長文全体ではなく当該パラグラフのみをテキストとして用いる。2) 選択肢を問題文の空欄に埋め込んだものを含意認識の仮説にする⁶。3) 最も適切でないものを答えさせる問題(含意の逆の問題)かどうかを判別し、その情報を出力処理に渡す。

含意認識エンジンとして今回は TIFMO[9] を用いた。TIFMO の出力するスコアを用いて、「含意の逆の問題」の場合はスコア最小の選択肢、それ以外はスコア最大の選択肢を選ぶ。ただし、スコア最大の選択肢が複数存在する場合は出力自体を行わない(信頼性が低いと考える)。以上の方法により、論説文では過去の模試も含めた6回分 合計30問中15問について出力が得られ、そのうち正解したのは3問のみであった。出力したものだけを見てもチャンスレベル以下であり

⁶問題文中の“The author claims that”のようなメタ表現は削除する。

抜本的な改善が必要である。

9 まとめと今後の課題

本稿では「ロボットは東大に入れるか」における英語チームが、昨年の代ゼミセンター模試でどのような回答手法を用いたかについて述べた。この受験を通して、短文問題については比較的正答できることが分かったが、一方で長文の難しさも分かった。英語チームとしては、今後長文問題に重点的を置いて取り組んでいく予定である。また、センター試験は4択であるが、東大の二次試験は記述式もあり、より問題が複雑だと考えられる。これらの問題についても取り組んでいきたい。

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に感謝いたします。

参考文献

- [1] Noriko H. Arai and Takuya Matsuzaki. The impact of A.I. on education—Can a robot get into the University of Tokyo? In *Proc. ICCE*, pages 1034–1042, 2014.
- [2] Rong-En Fan, Kai-Wei Chang, Cho-Jui Hsieh, Xiang-Rui Wang, and Chih-Jen Lin. LIBLINEAR: A library for large linear classification. *Journal of Machine Learning Research*, 9:1871–1874, 2008.
- [3] Joseph Gubbins and Andreas Vlachos. Dependency language models for sentence completion. In *Proc. EMNLP*, pages 1405–1410, 2013.
- [4] Sittichai Jiampojarn, Grzegorz Kondrak, and Tarek Sherif. Applying many-to-many alignments and hidden Markov models to letter-to-phoneme conversion. In *Proc. HLT-NAACL*, pages 372–379, 2007.
- [5] Daniel Jurafsky, Elizabeth Shriberg, and Debra Biasca. Switchboard SWBD-DAMSL shallow-discourse-function annotation coders manual, draft 13. Technical Report Technical Report 97-02, University of Colorado, Boulder. Institute of Cognitive Science, 1997.
- [6] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg Corrado, and Jeffrey Dean. Distributed representations of words and phrases and their compositionality. In *Proc. NIPS*, pages 3111–3119, 2013.
- [7] Naoaki Okazaki. CRFsuite: a fast implementation of conditional random fields (CRFs), 2007.
- [8] Bo Pang and Lillian Lee. Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales. In *Proc. ACL*, pages 115–124, 2005.
- [9] Ran Tian, Yusuke Miyao, and Takuya Matsuzaki. Logical inference on dependency-based compositional semantics. In *Proc. ACL*, pages 79–89, 2014.