

学習者の誤り傾向を反映した英語動詞選択誤りへの訂正候補推薦

澤井 悠 小町 守 松本 裕治
奈良先端科学技術大学院大学 情報科学研究科
{yu-s, komachi, matsu}@is.naist.jp

1 はじめに

英語学習者に対する作文支援としての、誤り検出および訂正についての研究が盛んにおこなわれている。これらの研究では主にスペル誤りや、前置詞・冠詞・時制の一致誤りといった文法誤りが対象となっている。一方、英語学習者コーパスにおいて頻出の誤りとして動詞選択の誤りがある。例えば、CLC-FCE コーパス [1]^{*1}では動詞の置換誤りを示す“RV”タグは、すべての誤り種類のうち、スペル誤りに次いで 2 番目に多い誤りである。また、Konan-JIEM コーパス^{*2}において動詞選択誤りは、18 の誤り種別中 6 番目に多い誤りである。

*They *connect / communicate with other businessmen and do their jobs with the help of computers.*

この例文^{*3}では、“connect”が学習者の選択した動詞であり、“communicate”は訂正された動詞である。どちらも自動詞としての用法をもつ動詞であり、文法誤りではない。しかし、意思疎通を図るという文意においては“communicate”のほうが適しているため、動詞選択誤りであると考えられる。

前置詞や冠詞の誤り検出・訂正において、訂正候補を **confusion set** (以下、候補集合) によって制限することが効果的であると示されている [2]。動詞選択誤り検出・訂正では、前置詞や冠詞と異なり訂正候補となる動詞数は膨大である。ゆえに、候補集合の重要性はそれらに対する研究に比べて高いといえる。

動詞選択誤りに関するこれまでの研究においては、類義語辞書に基づく候補集合生成が用いられることが多い [3]。しかし、学習者の誤りパターンは必ずしも類義語同士での混同だけではなく、既存手法にはカバレ

ジの問題がある。これに対し、日本語の動詞名詞コロケーション中の動詞選択支援タスクにおいて、カバレッジを向上する目的で、言語学習 SNS データを大規模学習者コーパスとして利用した研究がある [4]。

また既存研究では、複数の訂正候補が正解となる場合があるにも関わらず、自動訂正のみを主眼に置いたシステムが多い。一方で、システムが出力する訂正候補を学習者が無批判に受け入れる場合は少なく、提示された候補の質を見分けられると報告されている [5]。複数の訂正候補を提示し最終的な判断を学習者に委ねるというシステムには、実用性があると考えられる。

そこで本稿では、学習者作文中の動詞選択誤りに対し、訂正候補を推薦するシステムについて述べる。提案手法では、複数の訂正候補間の多クラス分類の結果から得た訂正候補のランキングを用いて、入力文中の動詞選択誤りの訂正候補推薦をおこなう。広範な学習者の誤りパターンに対応しシステムのカバレッジを向上するために、言語学習 SNS から収集した大規模な誤りタグ付きコーパスから動詞選択誤りの候補集合を作成し、学習者の誤り傾向を反映させる。訂正候補推薦の評価は、公開済み英語学習者コーパス上でおこなった。

本研究の主な貢献は、大規模な学習者コーパスから作成した候補集合を用いることにより、類義語辞書から候補集合を作成する既存手法に比べ、動詞選択誤りに対するカバレッジの向上を示した点である。

2 動詞選択支援の関連研究

Liu ら [3] は、中国人英語学習者作文の動詞選択誤りの自動訂正をおこなった。動詞の候補集合は、類義語辞書および英中対訳辞書から作成した。検出・訂正対象の動詞は、中国人英語学習者が用いる動詞のうち頻出の 50 語に絞っている。本稿では、候補集合を大規模な学習者コーパスから抽出している点や、対象とする動詞数が 500 語とより幅広いものである点、そして、自動訂正ではなく訂正候補の推薦が目的である点が異

^{*1} <http://ilexir.co.uk/applications/clc-fce-dataset/>

^{*2} <http://www.gsk.or.jp/catalog/GSK2012-A/>

^{*3} CLC-FCE コーパス中の“RV”タグが付与されている箇所から引用した。

なる。

Wu ら [6] は、動詞名詞コロケーションにおける動詞推薦タスクを、最大エントロピー法による多クラス分類器を用いておこなった。対象ドメインは学術論文であり、評価は人工データでおこなった。ここで、クラスとなるのはコロケーション中の動詞であり、790 語の動詞からなる単一の候補集合から、コロケーションの動詞を選択する。これに対し、本稿では対象となる動詞それぞれに候補集合を作成し、多クラス分類に用いるクラス数（候補数）を削減している。また、より広範なトピックが含まれる学習者作文コーパス（CLC-FCE）に対して評価をおこなっている点で異なる。

3 多クラス分類に基づく動詞選択誤り検出と訂正候補推薦手法

動詞選択誤りの検出および訂正候補推薦には、各対象動詞ごとの候補集合と多クラス分類器の出力するスコアを用いる。まず、予め対象とする動詞 v_i に対して訂正候補集合を定めておく。対象とする動詞の候補集合ごとに、ネイティブコーパスからの事例を用いて、候補集合中の各動詞をクラスラベルとする多クラス分類モデル M_i を構築する。学習者文が入力となるとき、各チェックポイントに対する動詞選択誤りの検出および訂正候補推薦は、リスト 1 に示す手順でおこなわれる。各チェックポイントに対する検出は、リスト 1 に示す DETECTERROR によりおこなわれる。与えられた文脈の空欄に入りうる各候補のスコアは、動詞周辺文脈の素性ベクトル \mathbf{f} についてのモデル M_i の出力により求まる。スコアの上位 k 個中に、元の動詞 v_{org} が含まれていなければ、そのチェックポイントを誤りとして検出する。誤りを検出した場合には、SUGGESTALTCANDS によって、訂正候補のスコア上位 n 個を推薦する。

動詞の語彙は一般に膨大であるものの、本稿では CLC-FCE コーパスの全動詞語彙中^{*4}、頻出の 500 語を対象の動詞とした。この頻出語彙は、CLC-FCE コーパス中の一般動詞の 90% をカバーする。

3.1 各候補集合に対するモデルの学習

分類器に与える事例は穴埋め形式であり、動詞を空欄としてクラスラベルとし、周辺文脈から素性ベクトルを生成する。動詞周辺文脈の表層および POS の n-gram に加えて、動詞前後の名詞句の主辞、およびそれらに対する単語クラスタリング素性を用いる。具体的な素性

^{*4} 全動詞語彙は 1,292 語である。

リスト 1 各事例に対する検出および訂正候補推薦手順。

```

procedure DETECTERROR( $M_i, \mathbf{f}, k, v_{org}$ )
   $ranked\_cands \leftarrow get\_ranked\_candidates(M_i, \mathbf{f})$ 
   $rank\_v_{org} \leftarrow get\_rank(v_{org}, ranked\_cands)$ 
  if  $rank\_v_{org} > k$  then
    return True else return False
  end if
end procedure

procedure SUGGESTALTCANDS( $ranked\_cands, n$ )
   $alt\_cands \leftarrow ranked\_cands_{[1, n]}$ 
  return  $alt\_cands$ 
end procedure

```

表 1 穴埋め形式の事例に用いる素性の例

(例: *They communicate with other businessmen and do their jobs with the help of computers.*)

“<S>” は文頭を表し，“*V*” は空欄にした動詞の部分を表す。

素性名	例
N-grams (surface)	they-*V*-with <S>-they-*V* *V*-with-other
N-grams (POS)	PRP-*V*-IN <S>-PRP-*V* *V*-IN-JJ
NP head (Left, Right)	L_they, L_PRP R_businessmen, R_NNS
NP head cluster (Left, Right)	L_01110001, L_0111000, L_011100 R_11011001, R_1101100, R_110110

の例は表 1 に示す。動詞前後の名詞句の主辞の素性は、n-gram 素性では捉え切れない動詞の主語、目的語を扱うために用いる。また、名詞句主辞の単語表層素性のスパースネスを軽減するために、教師なし・階層型のクラスタリング手法 [7]^{*5}によるクラスタリング素性を用いる。これは後述する ukWaC コーパスから作成した次元数 256 のもので、表 1 中の異なる長さのビット列は、3 段階のクラスタ粒度に対応している。

3.2 言語学習 SNS から抽出した大規模学習者コーパスを用いた候補集合作成

候補集合は、言語学習 SNS 上の相互添削データからマイニングした誤りタグ付きコーパスから抽出された。使用したコーパスは相互添削 SNS Lang-8^{*6}から収集された 2011 年 10 月時点までの添削付きの英作文である。添削タグ^{*7}が付与されている文対数は 1,122,589 文対であった。これらの文対に Natural Language Toolkit

^{*5} <https://github.com/percyliang/brown-cluster>

^{*6} <http://lang-8.com/>

^{*7} この添削タグは定型化されていない。そのためアノテーションの粒度は一定ではないし、誤り種別情報も付与されていない。

2.10^{*8}により POS タグを付与したものから、DP マッチングにより訂正前後ともに VB* の POS タグを持つ語の置換操作箇所を抽出した。結果、224,994 個の動詞置換ペアが得られ、パターン数は 76,000 種類であった。これらの訂正前後の動詞ペアは原形に戻したうえで、前項で述べた対象動詞以外の語を除外した。誤り語を見出し語として、各見出し語に対する訂正候補の頻度順上位 50 個までを各対象語彙の候補集合とした。対象動詞によって訂正候補の数にばらつきがあるため、候補集合全体での候補数の平均は 20.4 個である。

4 英語学習者コーパスを用いた評価実験

大規模学習者コーパスを用いた候補集合の効果を評価するために、英語試験作文の公開コーパスである CLC-FCE コーパス上で、訂正候補の推薦性能を自動評価尺度を用いて評価した。

4.1 実験設定

システムのチェックポイントは CLC-FCE コーパス中の全 “VB*” の POS タグの箇所 (“be” を除く) であり、訂正候補推薦の対象となるものは、同コーパス中で動詞の置換誤りを表す “RV” 誤りタグの箇所である。本稿では、複数語にわたる訂正箇所や、入れ子になっている訂正タグ箇所はチェックポイントから除外している。誤り箇所は、全チェックポイント中の約 4% である 1,083 箇所となる^{*9}。

大規模学習者コーパスから作成した候補集合（以下、Lang8CS と呼称）の有効性を示すために、つぎに述べる 2 つの候補集合をベースラインとした。まず、これまでの研究で用いられることが多い類義語辞典からの候補集合として、WordNet から候補集合を作成した (WordNetCS)。これは、対象動詞の Synset と、同一の Hypernym を共有する語を収集したものである。この WordNetCS の平均候補数は 14.8 個である。また、複数の学習者コーパス^{*10}の動詞選択誤り箇所から作成した候補集合を作成した (LearnerCS)。この LearnerCS の平均候補数は 5.1 個である。

検出における元の動詞の許容順位 k は、 $1 \leq k \leq 50$ とした。 $k=1$ の場合、先行研究 [3] のように、最尤候補

が元の動詞でない場合に検出をおこなう。また、 k を増加させると、元の動詞を正しいとみなしやすくなる。

4.2 モデル構築

多クラス分類器の学習には、ukWaC コーパス [8] をネイティブコーパスとして用いた。これは “.uk” ドメインのウェブサイトから収集されたコーパスであり、多様なドメインの文章を含む。実験では、極端な短文や表組などをクリーニングして得られた約 44,000,000 文の中から、各対象動詞についてそれぞれ 20,000 文をランダムに抽出して利用した。各対象動詞についての多クラス分類器は、ロジスティック損失関数・L2 正則化項のオンライン線形分類器であり、Scikit-Learn 0.13^{*11}を実装として使用した^{*12}。システムが多数の分類器からなるため、それぞれの分類器のパラメータチューニングはおこなわず、デフォルト値を用いた。学習およびテストにおける構文解析には Fanseparser 0.22^{*13}を用いた。

4.3 評価尺度

動詞選択誤り検出の評価尺度には、システムが検出をおこなった箇所に関する Precision と、誤りタグ付き箇所に関する Recall を用いる。また、すべての誤りタグ付き箇所に対する訂正候補推薦性能の評価には、次式に示す Mean Reciprocal Rank (平均逆順位) を用いる。

$$MRR = \frac{1}{N} \sum_{i=1}^N R R_i; \quad R R_i = \begin{cases} \frac{1}{r(gold_i)} & (gold_i \in S_i) \\ 0 & (\text{otherwise}) \end{cases}$$

ここで N は事例数で、実験においてはすべての誤りタグ付き箇所数である。事例 i についての逆順位 $R R_i$ は、推薦候補リスト S_i 中の正解 $gold_i$ の順位 $r(gold_i)$ を用いて定義される。推薦した候補中に正しい訂正が含まれていない場合や、システムの対象動詞に含まれない場合は逆順位は 0 となる。MRR 値の逆数は、正解が平均的に何番目に含まれているかを表すため、MRR 値が高いほど推薦性能が高いことになる。

4.4 結果

図 1 には各候補集合ごとに、ランキングに基づく検出時の許容順位 k を変化させ、横軸を Precision、縦軸を Recall としたプロットを示す。図 1 中では、提案手法である Lang8CS を用いたシステムが Recall, Precision

^{*8} <http://nltk.org>

^{*9} CLC-FCE コーパスの誤りタグの中では、動詞選択誤りはほぼ全て当該タグに含まれるものの、語彙選択以外の原因による誤りも含むことに注意したい。

^{*10} 本稿では、明示的に動詞選択誤りのタグが存在する Konan-JIEM コーパスと NICT-JLE コーパス、さらに NUS Corpus of Learner English (NUCLE) を用いた。

^{*11} <http://www.scikit-learn.org>

^{*12} 実装の都合上、分類器は One-Vs-Rest 形式であり、二値分類器をクラス数 (候補数) 分組み合わせて構築した。

^{*13} <http://www.isi.edu/publications/licensed-sw/fanseparser>

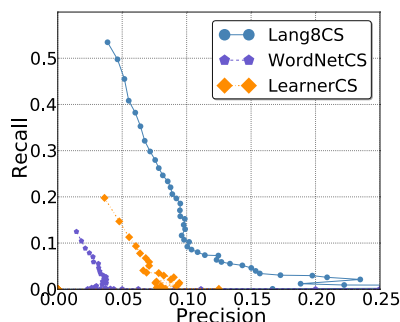


図1 許容順位 k による検出性能の変化. ($1 \leq k \leq 50$)

表2 誤り箇所に対する訂正候補推薦性能 (MRR 値)

WordNetCS	LearnerCS	Lang8CS
0.069	0.125	0.220

ともに他の候補集合を用いた場合よりも検出時の性能が高いことがわかる。また、候補集合の大きさは最も小さいに関わらず、LearnerCS は WordNetCS より性能が良かった。全候補集合についての傾向として、元の動詞が常に全候補中で最尤でなければならないという検出の条件 ($k=1$) のときに、最も Recall が高い。

表2に、全誤りタグ付き箇所に対する訂正候補推薦における各実験条件の MRR 値を示す。提案手法である Lang8CS の MRR 値は 0.220 であり、全候補集合中最も高い。この値は、推薦した候補中の 4.5 位までに正しい訂正候補が含まれていることを意味する。

5 考察

検出の Recall が低かった動詞の例として、“get” や “take” のように、頻出かつ様々な語義の用例をもつ語がある。今回用いた素性で捉えられている場合には、熟語表現として用いられるものは検出および推薦に成功した。候補集合およびモデルをそれぞれの語義ごとに構築することで、このような語についても検出が可能になると期待できる。

モデルの傾向として、誤用も一般的に用いられてしまっている語には同程度のスコアが与えられる。「O に影響をおよぼす」という語義の “affect” の誤用である “effect” がこの例として挙げられる。これは、学習に用いたコーパスがくだけた表現も含むウェブコーパスであるため、誤用を含む文脈も含まれていることが考えられる。ゆえに、同様の周辺文脈での混同が学習データにある場合、現状の素性ではこれらの事例を識別で

きなかったと考えられる。述語項構造などのより深い意味的素性の追加によりモデルの改善をおこなう必要がある。

候補集合に関わる問題としては、言語学習 SNS 由来の添削データを用いたため、見出し語と意味的な類似度が低い語が入っていることが挙げられる。例えば、“hear” の候補集合に “tire” が含まれるような場合である。このような語は、推薦候補に含まれてしまうと学習者を混乱させる原因にもなるので、クリーニングする必要がある。候補集合生成時に類似度尺度による添削ペアの足切りをおこなうことで、訂正候補の質の向上が期待できる。

6 おわりに

本稿では、英語学習者の英作文中の動詞選択誤りへの訂正候補推薦タスクに対し、学習者の誤り傾向に関する知識を用いた多クラス分類に基づく手法を提案した。言語学習 SNS から得た大規模な英語学習者コーパスから候補集合を作成し、対象動詞ごとに制約を与えることで学習者の誤り傾向を反映させた。評価実験の結果、既存手法で用いられる類義語辞書や、既存の学習者コーパスから候補集合を作成する場合に比べて、誤りパターンに対するカバレッジの改善がみられた。

参考文献

- [1] Helen Yannakoudakis, Ted Briscoe, and Ben Medlock. A New Dataset and Method for Automatically Grading ESOE Texts. In *Proc. of ACL*, pages 180–189, 2011.
- [2] Alla Rozovskaya and Dan Roth. Generating Confusion Sets for Context-Sensitive Error Correction. In *Proc. of EMNLP*, pages 961–970, 2010.
- [3] Xiaohua Liu, Bo Han, Kuan Li, Stephan Hyeonjun Stiller, and Ming Zhou. SRL-based Verb Selection for ESL. In *Proc. of EMNLP*, pages 1068–1076, 2010.
- [4] Lis W.K. Pereira, Erlyn Manguilimotan, and Yuji Matsumoto. Collocation Suggestion for Japanese Second Language Learners. In *IPSJ SIG Technical Report NL-210*, number 3, pages 3–7, 2013.
- [5] Claudia Leacock, Michael Gamon, and Chris Brockett. User Input and Interactions on Microsoft Research ESL Assistant. In *Proc. of NAACL Workshop*, pages 73–81, 2009.
- [6] Jian-cheng Wu, Yu-Chia Chang, Teruko Mitamura, and Jason S Chang. Automatic Collocation Suggestion in Academic Writing. In *Proc. of ACL*, pages 115–119, 2010.
- [7] Peter F Brown, Vincent J Della Pietra, Peter V DeSouza, Jennifer C Lai, Robert L Mercer, and Vincent J Della Pietra. Class-Based n-gram Models of Natural Language. *Computational Linguistics*, 18(4):467–479, December 1992.
- [8] Adriano Ferraresi, Eros Zanchetta, and Marco Baroni. Introducing and evaluating ukWaC, a very large web-derived corpus of English. In *Proc. of the 4th Web as Corpus Workshop*, pages 45–54, 2008.