

半自動フィードバックを利用した英文誤り検出手法の性能向上

永田 亮† 河合 敦夫†† 森広浩一郎† 梶井 文人†† 井須 尚紀††

† 兵庫教育大学 †† 三重大学

E-mail: †{rnagata,mori}@info.hyogo-u.ac.jp, ††{kawai,masui,isu}@ai.info.mie-u.ac.jp

1. はじめに

英文の誤りを検出する手法は、ルールに基づいた手法 [7], 統計的な手法 [2], [5], 両者を組み合わせた手法 [6] など, 様々な手法が提案されているが, その性能はまだ充分でない. 特に, 学習者の書いた英文を対象とした場合, 技術的な課題が多く, 十分な性能が得られていない. 例えば, 手法 [2] の性能は, 学習者の書いた英文エッセイに対して, 検出率約 20%, 検出精度約 80% であると報告されている. 英文の誤りを自動的に検出できれば, 英文添削にかかる時間と労力を軽減できるが, 実用的な添削支援を実現するためには, 従来手法の性能を大幅に向上させる必要がある.

誤り検出システムの性能を向上させる有効な手段として, 図 1 に示されるようなフィードバックが考えられる. フィードバックを利用した検出性能の向上では, まず, 英語教師や言語学者が, システムで検出できなかった誤りや誤検出を訂正し, 学習者の書いたエッセイを正しい英文にする (図 1 (a)). 次に, 訂正した英文エッセイを基に, 検出に失敗した原因を分析し, 検出ルールの修正を行う (図 1 (b)). また, 新たな検出ルールの作成や不必要な検出ルールの削除も行う. これらの作業は, 英語教師と言語学者に加えて, システムの技術面に精通したエンジニアが行う. 最後に, エンジニアが, (b) の検出ルールをシステムに実装し, 必要があれば, システムのチューニングを行う (図 1 (c)). この手順を繰り返すことで, 検出ルールが改善され, 検出性能が向上する.

フィードバックを利用した検出性能の向上における最大の課題は, フィードバックに多くのコストと時間を要する点である. フィードバックは, 様々な専門家 (英語教師, 言語学者, エンジニア) の手によって行われるため, 多くのコストを必要とする. また, 図 1 の (a) ~ (c) は, 時間を要する作業である. よって, フィードバックを利用した検出性能の向上には, コストと時間の面で大きな課題がある.

そこで本論文では, 半自動フィードバック機能を備えた誤り検出手法を提案する. 提案手法では, 可算名詞/不可算名詞の判定に基づいた誤り検出手法 [6] を拡張し, 半自動フィードバック機能を備えた誤り検出手法を実現する. このフィードバックでは, 図 1 の (b) と (c) が自動的に行われ人手を要するのは (a) だけである. 言い換えると, 検出結果を手で訂正した英文 (以後, フィードバックコーパスと呼ぶ) が与えられると, 自動的にフィードバックを行い, 誤り検出

性能を向上させる. したがって, フィードバックにかかる作業の大部分が自動化されることになる.

以下, 2. で, 提案手法の基礎となる誤り検出手法 [6] の概要について述べる. 3. では, 提案手法の基本アイデアを説明する. 4. で, 半自動フィードバック機能を備えた誤り検出手法を提案する. 5. で, 実験を行い, 提案手法を評価する.

2. 誤り検出手法の概要

誤り検出手法 [6] は, 可算名詞/不可算名詞の判定に基づいて, 冠詞の誤りや単数/複数に関する誤りを検出する手法である. 以下, この手法の概要を説明する. なお, 詳細については, 文献 [6] を参照されたい.

手法 [6] の誤り検出は, 大きく次の 3 つの手順からなる: (i) コーパスから学習データを自動生成, (ii) 学習データから, 可算名詞/不可算名詞の判定のための決定リストを学習, (iii) 決定リストを用いて, 可算名詞/不可算名詞の判定を行い, 判定結果に文法規則を適用して誤りを検出, の 3 つの手順から成る.

学習データの生成は, 判定対象の名詞 (以後, ターゲット名詞と呼ぶ) ごとに行う. まず, 主名詞として使用されたターゲット名詞とその周辺の単語を英語新聞などの一般的なコーパス (以後, 一般コーパスと省略) から収集する. 収集は, 既存の構文解析器や句解析器を用いて行う. 次に, 表層情報を利用して, 収集したターゲット名詞を可算名詞と不可算名詞に分類する. 例えば, 通常, 不定冠詞は可算名詞のみを修飾するので, 不定冠詞に修飾された主名詞は可算名詞に分類できる. 同様に, 無冠詞単数の場合は, 不可算名詞に分類できる. 最後に, 分類した可算名詞/不可算名詞を保存し, 学習データとする.

決定リストの学習は, 学習データを用いて, 可算名詞/不

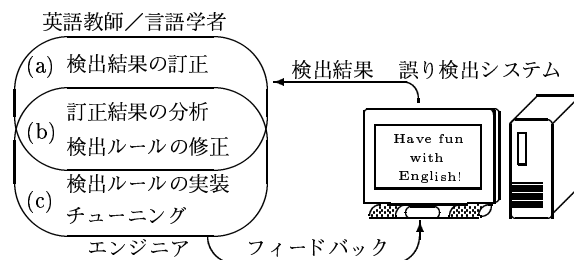


図 1: フィードバックを利用した誤り検出システムの性能向上

可算名詞の判定のための規則を生成し、規則に適用順位を付けることで行う。各規則は、次のようなテンプレートに従う：

$$w \rightarrow MC. \quad (1)$$

ここで、“ \rightarrow ”の左辺は規則を適用するための条件、右辺は判定結果を表す。また、 w は単語、 MC は可算名詞または不可算名詞を値にとる変数とする。式(1)は、「もしターゲット名詞の周辺^(注1)に単語 w が現れたら、ターゲット名詞を MC と判定」と解釈される。このテンプレートを満たすターゲット名詞周辺の単語を学習データから収集することで、規則を生成できる。規則の順位付けは、対数尤度比[9]に基づいて行う。すなわち、対数尤度比が高い規則ほど適用順位を高くする。対数尤度比は、式(1)と同じ記号を用いると、

$$\log \frac{p(MC|w)}{p(\overline{MC}|w)} \quad (2)$$

で定義される。ただし、 p と \overline{MC} は、それぞれ確率と MC の背反事象^(注2)を表すとする。確率 p は、学習データから得られる統計量を用いて、

$$p(MC|w) = \frac{f(w, MC) + \alpha}{f(w) + 2\alpha} \quad (3)$$

で推定する。ただし、 f は学習データ中の頻度、 α は平滑化パラメータとする。

誤りの検出は、可算名詞／不可算名詞の判定結果に基づいて行う。決定リスト中の規則を適用順位の高い順に試行し、適用可能な規則が見つかった時点で、その規則に従って判定を行う。最終的には、個々の名詞に依存しない基本的な文法規則を判定結果に適用して誤りの検出を行う。例えば、決定リストで不可算名詞と判定されたターゲット名詞が、不定冠詞や“many”に修飾されていたら、誤りとして検出する。

3. フィードバックの基本アイデア

提案手法の基本アイデアは、手法[6]を用いて、フィードバックコーパス^(注3)から決定リストを学習し、誤り検出性能を向上させるといものである。フィードバックコーパスを一般コーパスと見なすことで、一般コーパスの場合と同様に、決定リストを学習できる。一般コーパスと学習者の書いた英文とでは、単語の用法や出現傾向が異なる[2],[4]ことを考慮すると、一般コーパスに比べ、フィードバックコーパスは決定リストの学習に適した質の良いコーパスといえる。フィードバックコーパスから学習された決定リストによって、可算

名詞／不可算名詞の判定精度が向上すれば、可算名詞／不可算名詞の判定に基づいた誤り検出手法[6]の検出性能も向上する。

しかしながら、基本アイデアには、実用上大きな問題がある。一般に、決定リストの学習には、大規模なコーパスが必要となる。例えば、文献[6]では、約8000万語のコーパスを用いて決定リストを学習している。フィードバックコーパスは、学習者の書いた英文エッセイを基に作成されるので、大規模なものを収集することは困難である。更に、収集した英文エッセイの誤りを、誤り検出システムで検出し、その検出結果を手で訂正しなければならない。したがって、大規模なフィードバックコーパスを用意することは実用的でない。

この問題を解決するために、提案手法では、フィードバックコーパスと一般コーパスを組み合わせて決定リストを学習する。一般コーパスであれば、大規模なものを容易に手にいれることができる。そこで、大規模な一般コーパスとフィードバックコーパスを組み合わせることで、実用的なフィードバックを実現する。

4. フィードバック手法

既に述べたように、提案手法では、フィードバックコーパスと一般コーパスを組み合わせて決定リストの学習を行う。具体的には、対数尤度比を計算する際に、それぞれのコーパスから式(3)を用いて確率を推定し、フィードバックコーパスから推定された確率を用いて一般コーパスから推定された確率を補正することで両者を組み合わせる。

補正された確率を定式化するために次の記号を導入する。いま、フィードバックコーパスから推定された確率を p_{fb} 、一般コーパスから推定された確率を p_g で表す。このとき、パラメータ λ を用いて、補正された確率 p_i を、

$$\begin{aligned} p_i &= \lambda p_{fb} + (1 - \lambda) p_g \\ &= p_g + \lambda(p_{fb} - p_g) \end{aligned} \quad (4)$$

を定義する。式(4)では、フィードバックコーパスから推定された確率 p_{fb} と一般コーパスから推定された確率 p_g との差にパラメータ λ を乗じて、 p_g を補正している。この p_{fb} と p_g の差は、フィードバックコーパスと一般コーパスにおける可算名詞／不可算名詞の用法の差と捉えることができる。したがって、式(4)は、フィードバックコーパスと一般コーパスにおける可算名詞／不可算名詞の用法の差に応じて p_g を補正していると解釈できる。

λ は、 p_{fb} の(推定の)信頼性と p_g の信頼性に基づいて決定する。もし、 p_{fb} の信頼性が十分に高ければ、決定リストの学習に一般コーパスを用いる必要はなく、フィードバックコーパスのみから決定リストを学習すれば良い。すなわち、 λ の値を大きくして、 p_i を p_{fb} に近づけるべきである。逆に、

(注1)：ターゲット名詞の周辺とは、ターゲット名詞の前後± k 単語やターゲット名詞の存在する名詞句内などの文脈を表す。

(注2)： MC = 可算名詞であれば、 \overline{MC} = 不可算名詞である。また、その逆も成り立つ。

(注3)：1. で述べたように、フィードバックコーパスとは、誤り検出システムの検出結果を手で訂正した英文のことである。

p_{fb} の信頼性が低ければ、 λ を小さくして、 p_{fb} が p_i に与える影響を小さくするべきである。

p_{fb} の信頼性は次のように与えられる。 p_{fb} は、ターゲット名詞周辺に単語 w が現れたときの、可算名詞（または不可算名詞）として使用されたターゲット名詞の割合と捉えることができる。割合の信頼性はその標準偏差の 2 乗の逆数で示される [8] ことから、 p_{fb} の信頼性 c_{fb} は、

$$c_{fb} = \frac{f_{fb}(w)}{p_{fb}(MC|w)(1 - p_{fb}(MC|w))} \quad (5)$$

で与えられる。ただし、 f_{fb} は、フィードバックコーパス中の頻度を表す。 p_g の信頼性も同様に与えられる。

式 (5) の信頼性を用いて、上記の性質を満たす λ を次のように定義する。いま、 p_{fb} の信頼性と p_g の信頼性をそれぞれ c_{fb} と c_g する。このとき、 λ を

$$\lambda = \frac{c_{fb}}{c_g} \quad (6)$$

で定義する。更に、式 (6) を用いて、 p_i を、

$$p_i = \begin{cases} p_g + \frac{c_{fb}}{c_g}(p_{fb} - p_g), & c_{fb} < c_g \\ p_{fb}, & c_{fb} \geq c_g \end{cases} \quad (7)$$

で定義する。式 (7) では、 p_{fb} の信頼性 c_{fb} が大きくなると、 p_i は p_{fb} に近づく。また、 p_{fb} の信頼性と p_g の信頼性が等しいとき、 $p_i = p_{fb}$ となる。これは、 p_{fb} と p_g が同じだけ信頼できる場合は、一般コーパスは利用せず、フィードバックコーパスのみを学習に利用することを意味する。

式 (7) では、非常に規模の大きい一般コーパスを使用した場合、問題が起こることがある。信頼性 c_g は、 $f_g(w)$ に比例して大きくなるため、大規模な一般コーパスを使用した場合、 c_g の値が非常に大きくなり、 p_{fb} の信頼性が十分に高い場合でも、 p_{fb} が p_i にほとんど影響を与えないことがある。例えば、 $f_{fb}(w) = 100$ 、 $p_{fb} = 0.7$ 、 $f_g(w) = 10000$ 、 $p_g = 0.3$ の場合、 $\lambda = 1/100$ 、 $p_i = 0.304$ となり p_{fb} が p_i に与える影響は非常に小さくなる。しかしながら、 $f_{fb}(w) = 100$ であれば、 p_{fb} を推定するサンプル数として少くないので、 λ を大きくすることが望ましい。そこで、 c_g の対数をとった、

$$p_i = \begin{cases} p_g + \frac{c_{fb}}{\log_2 c_g}(p_{fb} - p_g), & c_{fb} < \log_2 c_g \\ p_{fb}, & c_{fb} \geq \log_2 c_g \end{cases} \quad (8)$$

も定義する。式 (8) では、 c_g の対数をとることで、一般コーパスが非常に大きい場合でも、 p_{fb} が p_i に与える影響が小さくなることを防いでいる。また、式 (8) は、 c_{fb} に指数関数的に重みを付けて、 p_g の補正を行っているとも解釈できる。前述の通りフィードバックコーパスは決定リストの学習に適した質の良いコーパスである。式 (8) は、このフィードバックコーパスの質の良さを考慮したモデルと解釈できる。

以上をまとめると、提案手法のフィードバックの手順は次のようになる。まず、フィードバックコーパスと一般コーバ

ス、それぞれから学習データを生成する。次に、それぞれの学習データから、決定リストの規則を生成する。次に、式 (2) で定義される対数尤度比に基づいて、生成された規則に適用順位を付ける。ただし、対数尤度比を計算する際には、確率 p として式 (7)（または式 (8)）で定義される補正された確率 p_i を用いる。最後に、この学習された決定リストを、フィードバックの結果として出力する。

5. 実験と考察

5.1 実験条件

誤りの検出対象として、日本人英語学習者が書いた英文エッセイ [1] を用いた。このエッセイから、トピック “Gardening” と “Traveling” について書かれたものを、それぞれ 54 エッセイ（約 4000 語）、47 エッセイ（約 3000 語）用いた。このエッセイを、英文校正を職業とする英語ネイティブスピーカーがチェックしたところ、それぞれ 145 個と 105 個の誤りが確認された^(注 4)。これらの誤りを修正して、性能評価のための正解データとした。また、フィードバックコーパスとしても利用した。

一般コーパスとして、EDR 電子化辞書 [3] に収録されている英語コーパスと日英対訳辞書の英語語義文とを使用した。この一般コーパスの規模は約 300 万語となった。

検出性能の評価尺度には、検出率と検出精度を用いた。検出率は、

$$\frac{\text{正しく検出した誤りの数}}{\text{英文エッセイ中の誤りの数}} \quad (9)$$

で定義した。検出精度は、

$$\frac{\text{正しく検出した誤りの数}}{\text{検出した誤りの数}} \quad (10)$$

で定義した。

5.2 実験手順

まず、誤り検出手法 [6] を用いて、一般コーパスから、トピック “Gardening” のエッセイに出現した一般名詞に対して決定リストを学習した。平滑化パラメータは、 $\alpha = 1$ とした。名詞句の解析や主名詞の抽出には、英文解析ツール “OAK System^(注 5)” を用いた。その後、学習された決定リストの判定に基づいて、このエッセイ中の誤りを検出した。

次に、提案手法を用いて、フィードバックを行った。トピック “Gardening” のエッセイから、一つエッセイを選び出し検出対象とし、残りの 53 のエッセイをフィードバックコーパスとしてフィードバックを行った。この手順を各エッセイに対して行い、計 54 回のフィードバックと検出を行った。それぞれの検出結果を統合することで、最終的な検出結果とし

(注 4) : ここでの誤りの数は、誤り検出手法 [6] が検出対象としている誤りの数である。

(注 5) : <http://nlp.cs.nyu.edu/oak/>

た。フィードバックは、式 (7) または式 (8) を用いた 2 通りの手法で行った。また、比較のために、単純にフィードバックコーパスを一般コーパスに加えて、決定リストを学習し、誤りの検出を行った。

最後に、各手法の検出性能を評価した。前述の正解データと各手法での検出結果を比較し、検出率と検出精度を求めた。また、トピック “Traveling” のエッセイについても、同じ手順で実験を行った。

5.3 実験結果

表 1 に、実験結果を示す。表 1 中の “手法 [6] + FB corpus” は、一般コーパスにフィードバックを加えて決定リストを学習した手法を表す。また、“Feedback” は提案手法を表し、括弧の中は確率の補正に用いた式を表す。

表 1 から、提案手法のフィードバックによって、誤り検出手法 [6] の性能が向上したことがわかる。また、単純にフィードバックコーパスを一般コーパスに加えた場合に比べても、提案手法のフィードバックによって性能が向上していることがわかる。特に、式 (8) を用いた場合、検出率、検出精度ともに改善している。これは、式 (8) では、フィードバックコーパスから推定された確率の信頼性に指数関数的に重み付けをして、フィードバックを行ったからである。この結果から、フィードバックコーパスと一般コーパスを同等に扱うのではなく、フィードバックコーパスに重み付けをつけてフィードバックを行うべきであるといえる。

実際、式 (8) のフィードバックで得られた規則には、フィードバックの効果が見られた。例えば、フィードバックによって、ターゲット名詞 “car” に対する規則 “by → 不可算名詞” の適用順位が上がり、正しく使用されている “by car” を誤りとしてしまう誤検出を減らした。日本人英語学習者の書いた英文では、“by car” の “car” は手段を表し、無冠詞単数になる正しい用法が頻出するため、人手によるフィードバックでも同様に、“by → 不可算名詞” の適用順位が上げられると予想できる。このように、人手によるフィードバックでも予想される規則の修正が、提案手法のフィードバックによって行われることが確認できた。

以上の結果から、提案手法は、誤り検出性能を向上させる有効な手法であるといえる。性能向上に加えて、提案手法のフィードバックは、半自動で行われることも重要な有効性である。すなわち、提案手法では、図 1 の (b) と (c) が自動

的に行われる。フィードバックにかかるコストと時間を考慮すると、この点も提案手法の大きな利点である。

6. おわりに

本論文では、フィードバックを半自動的に性能を向上させる英文誤り検出手法を提案した。実験により、提案手法のフィードバックによって、検出率、検出精度ともに向上することを確認した。また、性能向上だけでなく、提案手法によって、フィードバックにかかる作業の大部分が自動化されるという有効性も確認した。提案手法は、フィードバックコーパスから推定された確率で一般コーパスから推定された確率を補正することでフィードバックを行うので、他の確率に基づいた誤り検出手法にも応用できる可能性が高い。例えば、そのような手法として、Chodorow らの手法 [2] が挙げられる。この応用可能性も提案手法の利点である。

今後は、提案したフィードバック手法を拡張して行く予定である。例えば、フィードバックコーパスと訂正前の英文エッセイを比較すれば、誤りの例からなる誤りコーパスが構築できる。この誤りコーパスも有用な情報源としてフィードバックに利用できる。また、誤り検出結果とフィードバックコーパスを比較することで、誤り検出に使われた検出ルールの性能評価も行える。これらの情報を利用することで、更なるフィードバックが可能になると予想される。

参考文献

- [1] 朝尾幸次郎, 第二言語習得研究のための英語学習者コーパスの構築とその利用, 科学研究費補助金研究成果報告書, 2000.
- [2] M. Chodorow and C. Leacock, “An unsupervised method for detecting grammatical errors,” Proc. 1st Meeting of the NAACL, pp.140–147, Oct. 2000.
- [3] Japan electronic dictionary research institute ltd, EDR electronic dictionary specifications guide, Japan electronic dictionary research institute ltd, Tokyo, 1993.
- [4] S. Granger, “Prefabricated patterns in advanced EFL writing: collocations and formulae,” Phraseology: Theory, Analysis and Applications, Clarendon Press, pp.145–160, 1998.
- [5] E. Izumi, K. Uchimoto, T. Saiga, T. Supnithi, and H. Isahara, “Automatic error detection in the Japanese learners’ English spoken data,” Proc. of 41st Annual Meeting of ACL, pp. 145–148, July 2003.
- [6] R. Nagata, T. Wakana, F. Masui, A. Kawai, and N. Isu, “Detecting article errors based on the mass count distinction,” Proc. of 2nd International Joint Conf. on Natural Language Processing, pp. 815–826, Oct. 2005.
- [7] D. Schneider and K.F. McCoy, “Recognizing syntactic errors in the writing of second language learners,” Proc. 17th International Conf. on Computational Linguistics, pp.1198–1205, Aug. 1998.
- [8] 田中良久, 心理学的測定法, 東京大学出版, 東京, 1977.
- [9] D. Yarowsky, “Unsupervised word sense disambiguation rivaling supervised methods,” Proc. 33rd Annual Meeting of ACL, pp. 189–196, June 1995.

表 1: Experimental results

Method	Gardening		Traveling	
	検出率	検出精度	検出率	検出精度
手法 [6]	0.59	0.78	0.70	0.68
手法 [6]+FB corpus	0.60	0.78	0.71	0.69
Feedback (式 (7))	0.60	0.78	0.71	0.70
Feedback (式 (8))	0.61	0.79	0.71	0.72