

学習指導要領に基づく小学生作文支援システムの構築

後藤田 結† 田村 直良‡

横浜国立大学環境情報学府†

横浜国立大学環境情報研究院‡

E-mail: gotoda-yui-sx@ynu.jp † tam@ynu.ac.jp ‡

1 はじめに

本研究では、小学生の書く作文を対象に、学習指導要領に基づいた観点で作文の本文を解析し、より適切な文章に修正するための指摘事項を提示する手法について検討する。

近年、情報化社会の進展に伴い、学校における教育の情報化が進められている。しかし、国語教育で行われる作文のような分野においては、その内容を計算機で扱うことが難しく、現在、情報技術を活用した学習者の支援が充分に行われているとはいえない。このような分野で情報技術を活用するにあたっては、学習者が入力する文章を解析するための自然言語処理技術が必須である。

作文教育においては、段落分けや、筆者の伝えたいことが効果的に伝わるような書き方といった指導が行われる。段落分けはテキストセグメンテーションの問題として捉えることができ、筆者の伝えたいことが効果的に伝わるように書くことは重要文抽出の問題の応用として捉えることができる。テキストセグメンテーションに関しては語彙の出現頻度から話題の境界を検知する Hearst による手法[1]、重要文抽出に関しては複数の手がかりから機械学習手法によって重要文を抽出する平尾らによる手法[2]など、自然言語処理の分野の中で数多くの研究が行われてきている。だが、従来の研究はいずれも新聞記事・論文など、大人の書いた文章を対象としており、小学生の作文のよう

に、対象となる文章そのものが未熟である場合、そのまま適用するだけでは十分な結果を得ることができない。そこで本論文では、小学生の作文のように未熟な文章を対象とし、学習指導要領に基づいて文章の解析を行い、入力された文章の中から修正すべき点を指摘することで小学生の国語教育を支援するシステムを構築する方法を提案する。

2 学習指導要領における「書くこと」の指導事項

学習指導要領において、国語教育には「話すこと・聞くことの指導」「書くことの指導」「読むことの指導」の3つの観点がある。この中の「書くこと」の目標は、学年によって以下のように定められている[3]。

第1学年及び第2学年

経験したことや想像したことなどについて、順序を整理し、簡単な構成を考えて文や文章を書く能力を身に付けさせるとともに、進んで書こうとする態度を育てる。

第3学年及び第4学年

相手や目的に応じ、調べたことなどが伝わるように、段落相互の関係などに注意して文章を書く能力を身に付けさせるとともに、工夫をしながら書こうとする態度を育てる。

第5学年及び第6学年

目的や意図に応じ、考えたことなどを文章全体の構成の効果を考えて文章に書く能力を身に付けさせるとともに、適切に書こうとする態度を育てる。

また、指導事項についてより詳細に述べた学習指導要領解説[4]において、事実と感想・意見などを区別する、「始め—中—終わり」などの構成を意識できるようにする、といった指導事項が挙げられている。本研究では、これらの指導事項に基づいて、入力された作文の中から不適切と思われる文を抽出し、修正すべき点を出力するシステムを構築した。

3 構成タグつき小学生作文コーパスの構築

3.1 作文の収集

神奈川県内9校の小学校から藤田らが収集した672編の作文事例[5]をもとに、教科書の文章の紹介文、体験学習の感想文といった作文課題ごとに10～20編の作文を無作為に抽出し、合計100編をアノテーションの対象とした。

3.2 構成タグの付与

2章で述べた指導目標および学習指導要領解説の内容を踏まえ、小学生作文の構成を分析するため、以下に述べる4つのアノテーション項目を作成した。

- (ア) 「始め・中・終わり」それぞれの先頭である文
- (イ) 最も強く感じたことを表す文
- (ウ) 自分の感想・意見などを述べた文
- (エ) 段落の先頭である文

この中で、(イ)最も強く感じたことを表す文に関しては、さらに次の2種類に分類した。

- (イ)-i 作文全体のまとめとなる、筆者の意見、感想などを述べている文
- (イ)-ii 作文全体の内容を受けて、今後の自分や周囲の展望について述べている文

これらの項目について、3名のアノテータがそれぞれ作文75編ずつのアノテーションを行った。表1に、アノテータごとの構成タグの付与事例数の合計を示す。

表1 構成アノテーションの付与事例数

項目	アノテータ		
	A	B	C
(ア)	132	143	97
(イ)-i	70	66	58
(イ)-ii	35	41	20
(ウ)	505	460	454
(エ)	213	323	167

3.3 構成タグ付きコーパス

前節のアノテーション結果により、コーパスの各文にタグを付与した。タグには各アノテータのアノテーションの一致度に応じて0～1の値が付与されている。これを構成タグ付き小学生作文コーパスと呼ぶ。

また模範的な文章として、毎日新聞社説10編を使用して同様なアノテーションを行い、構成タグ付き社説コーパスとした。

4 作文に対する指摘事項の提示手法

4.1 素性の抽出

入力された作文中の各文が3.2節で述べた項目に該当するかどうか判定するため、文章中の各文について、文の位置、語彙連鎖の有無、文中の名詞のTF-IDF値の合計を正規化したもの、該当文と前後文に出現する接続詞から算出する各文の統括力、文の主題が筆者かどうか、文の時勢、文末表現などといった素性を抽出した。

4.2 構成の解析

前節の素性を用いて、3.2節で述べた4つの項目ごとにSupport Vector Machine (SVM)による判定器を訓練した。判定器の訓練には、構成タグ付き小学生作文コーパスのうちアノテーションの一致度が高い作文と、構成タグ付き社説コーパスを合わせたものを用いた。

4.3 文・文章構成の適切さの判定

4.2 節の判定結果と、3.2 節で述べた模範的な作文のコーパスから得られる統計情報に基づき、入力された作文について、各文と文章構成のそれぞれの観点から適切さの判定を行う。

各文については、文の表層から得られる情報と、SVM で判定を行った際の分離平面からの距離の値を用いて、以下の観点で適切さの判定を行う。

- a-1 段落の先頭として不適切な文ではないか
- a-2 段落の先頭とすべき文ではないか
- a-3 自分の感想・意見などと事実が区別して書かれているか
- a-4 一文の長さが適切か
- a-5 文の長さに対して適切な数の読点が打たれているか

文章構成については、SVM で各文を分類した結果に基づいて、以下の観点で適切さの判定を行う。

- b-1 始め・中・終わりが分かれているか
- b-2 始め・中・終わりそれぞれの長さが適切か
- b-3 文章の長さに対する段落の数が適切か
- b-4 作文中での重要な文が作文中の先頭か末尾に近い位置に配置されているか
- b-5 作文中で自分の感想・意見と事実の両方が述べられているか

4.4 指摘事項の作成・提示

4.3 節で述べたそれぞれの観点について、不適切であると判定された点について、指摘事項を表示する。

図 1 に、不適切な点と対応して表示される指摘事項の例を述べる。ただし、i, j はそれぞれ、該当する文の文番号である。

a-2 の観点で不適切と判定された文

「i 文目は、まえの文と違う内容なので、新しい段落にしたほうがいいでしょう」

b-4 の観点で不適切と判定された文章

「この作文で一番大切な文は、j 文目ではありませんか。大切な文は、作文のはじめの方が、おわりの方に持ってきたほうが、読んだ人の心にのこります」

図 1: 表示する指摘事項の例

5 実験と考察

5.1 実験内容

提案手法に基づき、小学生作文における構成の解析・指摘事項の提示実験を行った。実験には、3 章で述べた小学生作文コーパスを用いた。構成の解析には SVM-Light¹を利用した。訓練データとして、構成タグつき小学生作文コーパスのうち、アノテーションの一致度が高いほうから上位 20 編の作文と、構成タグつき社説コーパスを合わせた、合計 30 編の文章を用いた。

5.2 実験結果

作文の構成解析結果の精度を表 2 に、実際に入力として用いた作文本文からの抜粋と、それに対して出力されたアドバイスの例を図 2 にそれぞれ示す。

表 2 作文の構成解析結果

項目	適合率	再現率
(ア)	56.79	82.00
(イ)-i	17.88	80.80
(イ)-ii	30.06	81.92
(ウ)	27.78	39.94
(エ)	19.51	75.65

¹ Support Vector Machine SVM-Light:
<http://svmlight.joachims.org/>

小学生の作文（抜粋）

文1 ぼくはかんばんさがしをすることになりました。
文2 分からないことはすべて、おばさんにききました。
文3 はじめに見つけたのは①です。

出力された指摘事項

・3文目は、あたらしい段落からはじめてみましょう。

図2 出力された指摘事項の例

5.3 構成の解析・指摘事項に関する考察

表2において、作文の構成解析結果の数値は充分とはいえない。その原因として、対象となる文章そのものが未熟であり、人間が見た場合でもしばしば判断が一致しないことが挙げられる。そのような作文にも対応するため、指摘事項の作成に際して、SVMで判定を行ったときの分離平面からの距離の値を適切さの判定に用いる、一文の長さや読点の数といった構成の基本となる項目においては、構成タグつき社説コーパスや、構成タグ付き小学生作文コーパスのうちアノテーションの一致度が高い作文から得られる統計情報を用いた適切さの判定を行う、といった方法で対処している。

(ウ) 自分の感想・意見などを述べた文とそうでない文に関しては、文の意味を考慮しなければ判別できない事例が多いため、今回用いた素性だけでは判別が難しかった。また、模範的な文章として訓練データの一部に用いた新聞社説と小学生の作文では、意見の述べ方が大きく異なっていたことも、分類誤りの原因として考えられる。

実験に使用した小学生作文100編中、指摘事項を出力した作文は90編だった。文ごとにとみると、小学生作文中の文合計2921文中の204文(6.9%)に対して指摘事項を出力した。しかし、(ウ)自分の感想・意見などを述べた文とそうでない文の判定が不十分なために、作文中に一度も自分の意見や感想が述べられていない作文(4.3節で述べたb-5の観点に照らして不適切)において指摘事項が出力されないなどの事例があり、実用化のため

には作文構成の解析精度を高める必要があると考えられる。

6 まとめと今後の課題

本論文では、小学生の作文を対象に、学習指導要領に基づいて作成した項目によって作文の構成を解析し、その結果をもとに、不適切な文や文章構成を指摘することで、作文教育を支援する手法を提案した。評価実験の結果、実験に用いた小学生作文の90%において、修正すべき点を指摘事項として提示することができた。しかし、自分の感想・意見などを述べた文とそうでない文の判別などの項目において、構成の解析が不十分であるために、出力すべき指摘事項が出力されない事例がみられた。

参考文献

- [1] Marti A. Hearst, “TextTiling: Segmenting Text into Multi-paragraph Subtopic Passages”, *Computational Linguistics*, Vol.23 No.1 pp.33-64, 1997
- [2] 平尾努, 磯崎秀樹, 前田英作, 松本裕治, “Support Vector Machine を用いた重要文抽出法”, *情報処理学会論文誌*, Vol.44 No.8, pp.2230-2243, 2003
- [3] 文部科学省, “小学校学習指導要領”, 現行学習指導要領・生きる力 第2章 各教科 第1節 国語, 2008
- [4] 文部科学省, “小学校学習指導要領解説 国語編”, 現行学習指導要領・生きる力, 2008
- [5] 藤田彬, 田村直良, “作文事例に基づいた児童の「書くこと」に関する学習傾向についての分析”, *言語処理学会第18回年次大会発表論文集*, pp.987-990, 2012