

小論文採点支援システムにおける 文字誤り検出モジュールの構築

小畑 友也

岡山大学大学院自然科学研究科
pbgn8vxd@s.okayama-u.ac.jp

泉仁 宏太

岡山大学大学院自然科学研究科
pm9n6cei@s.okayama-u.ac.jp

飯塚 誠也

岡山大学全学教育・学生支援機構

竹内 孔一

岡山大学大学院自然科学研究科
koichi@cl.cs.okayama-u.ac.jp

田口 雅弘

岡山大学院社会文化科学研究科

阿保 達彦

岡山大学大学院自然科学研究科

大野 雅幸

岡山大学大学院自然科学研究科
pw2z9792@s.okayama-u.ac.jp

稲田 佳彦

岡山大学院教育学研究科

上田 均

岡山大学大学院自然科学研究科

1 はじめに

著者らは日本語で書かれた小論文に対して自動採点を行うシステムを開発中である [1, 2, 3, 4]。本稿ではそこで作成している誤字脱字などの文字誤りを検出するモジュールについて記述する。

小論文の自動採点システム構築のために、本研究ではまず、講義形式による模擬試験を行い、研究利用可能な小論文答案データを作成している [2]。筆記の小論文答案をみると図 1 に示すように、「侵食」の「食」の字を間違えて記述している。このように誤字などがあった場合小論文では減点する。そのため、自動採点では誤字脱字を検出することは必要である。

誤字脱字を検出する手法として言語モデルを利用した手法を提案する。また本手法による誤字脱字検出とリクルート社が提供する A3RT の Proof Reading API¹を比較する。その結果、提案する言語モデルによる手法が Proof Reading API を上回る結果を得たことを報告する。

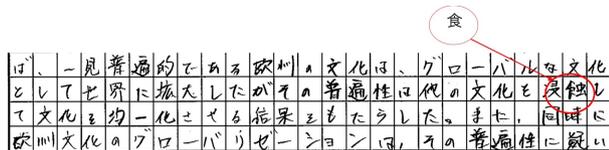


図 1: 実際の小論文文中での誤り例

2 関連研究

文字誤り検出の研究は訂正も含めて様々な言語モデルが提案されてきた。例えば文献 [5] では大量の新聞記事データをもとに隠れマルコフモデルを学習して、言語モデルを作成し、OCR が出力する誤った文字列を形態素単位で修正する手法を提案した。近年では、ニューラルネットワークを利用したモデルが提案されており、例えば [6] では seq2seq のニューラル翻訳モデルを利用して、文字列ベースの誤字訂正システムを提案している。また、言語学習者の作文を訂正する目的で、[7] は日本語の末尾表現を対象に、擬似的に作成した誤字データを作成し、誤字を正しい字に変換するニューラル翻訳モデルの手法を提案した。その中で、文字列ベースと単語ベースのモデルを比較し、単語ベースの方が精度が高いことを示した。

本研究では形態素ベースの誤字検出手法を提案しており、タスクが関連していることから参考になる手法である。

また、国内の小論文の自動採点の先行研究には石岡らの Jess [8, 9] がある。Jess は毎日新聞の社説およびコラムを学習し文章作法を評価する「修辞」、アイデアが理路整然と表現されていることを示す「論理構成」、トピックに関連した語彙が用いられているかを示す「内容」の 3 項目から小論文の評価を行うシステムである。

3 モデル

2.1 節では扱う文の前処理について記述し、2.2 節、2.3 節では二つの提案手法の概要について記述する。

¹A3RT <https://a3rt.recruit-tech.co.jp>

3.1 モデル準備

本研究では形態素区切りにした文章を扱う。これには主に2つの理由がある。1つは、人が扱う文字の最小単位が形態素であり、形態素単位で考えることは最も自然であると考えられるからである。他方では、文字誤りを含む文章を形態素分割したとき、異常な分割になりやすい性質があり、これを利用することで文字誤りを検出しやすいものと考えられることができるからである。

なお、形態素分割には形態素解析器 Mecab²でUnidicを用いた分割を行う。

3.2 統計モデル

統計モデルでは、連続する3形態素において、直前2形態素に対する3番目の形態素の確率値を新聞記事やwikipediaなどの膨大な文章群から取得し、得られた確率値について、人手で設定した閾値を下回るものを誤りとする。

確率値の式は(1)式のように表すことができる。

$$P_i = P(s_i | s_{i-1}, s_{i-2}) \quad (1)$$

このモデルを最も基本的なモデルとして比較対象とする。なお、実際の小論文解答中でのTrigram確率については表1のように算出される。

表 1: 実際の小論文解答中での Trigram 確率

	形態素	確率
	#	-
	#	-
1	グローバルゼーション	0.00000352
2	は	0.125
3	世界	0.0
4	全体	0.0113
5	の	0.484
6	所得	0.000347
7	格差	0.0182
8	を	0.169
9	縮小	0.0217
	...	-

3.3 ニューラル言語モデル

本モデルの概要を図2に示す。本モデルの基本構造は統計モデルと同じで、直前2形態素から次に出現す

²形態素解析機 MeCab <http://taku910.github.io/mecab/>

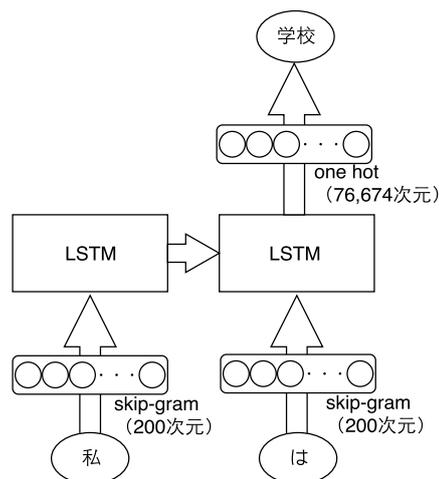


図 2: ニューラル言語モデル

る形態素の確率を予測する。

本モデルではRNNの一種であるLSTMを用いて2層ニューラルネットワークを構築する。入力を2形態素(200次元のSkip-gram)、出力を1形態素(one hot)とし、中間層を128次元、活性化関数には双曲線関数を用いる。Skip-gramは国立国語研究所の公開するfastText[10]で学習したものである。入力をone hotにする場合に比べ、使用する記憶域の縮小および計算時間の大幅な短縮が実現できる。なお、Skip-gramに登録のない単語についてはすべての要素が0である200次元のベクトルを入力とする。

4 実験

3.1節では実験で扱うデータについて、3.2節では評価を行うためのデータの作成方法について、3.3節では精度評価についてそれぞれ述べる。

4.1 実験データ

実験では、モデルごとに異なるデータを用いる。これは、統計的言語モデルに比べ、ニューラル言語モデルは時間・空間計算量の観点から学習に用いるデータを小さくせざるを得ないためである。今後の課題として、使用するデータを揃え、同一条件下での比較を行う必要がある。

実験に使用したデータは表2の通り。これらのデータを訓練データとして学習を進める(あるいは統計値を取得する)。

また、本研究で作成したモジュールの精度を調べるために実際に小論文模試を実施した。(表3) その一部

について人手による誤り検出データを作成し、それらの誤りが検出できるかどうかを評価指標とした。

表 2: 実験用データ

統計的言語モデル	毎日新聞 11 年分 日本経済新聞 8 年分 wikipedia 全記事 (4GB)
ニューラル言語モデル	毎日新聞 11 年分

表 3: 実施した試験の概要と受験者数

試験 1	グローバル化の光と影
実施年度	2016 年
受験者数	321 人
人手による誤り検出	83 人分
試験 2	自然科学の構成と科学教育
実施年度	2016 年
受験者数	321 人
人手による誤り検出	83 人分

4.2 人手による誤り検出データの作成

本研究で作成したモジュールの精度を検証するために、人手による誤り検出データを作成する。表 4 のような誤りを定義し、それらを人手により検出する。

表 4: 定義した文字誤りの種類

誤り種類	詳細
助詞誤り	助詞の使い方
漢字誤り	漢字の書き間違い
体裁誤り	「～たり～たり」などの用法
削除誤り	必要な文字が抜けている
挿入誤り	余分な文字が入っている
慣用句誤り	慣用句の誤用
論文らしさ	口語になっていないか

本システムのために用意した小論文の解答データは受験者の手書き文章であるが、本システムで扱うために人手により電子化している。このことから本システムで扱うデータ上では 2 種類の誤りが存在する。本モジュールでは、共に誤りとして検出することを目指す。

1. 小論文受験者が文字誤りをした。
2. 小論文受験者が正しい文字を書いたが、電子化する際に入力ミスがあった。

4.3 精度評価

精度の評価は Precision, Recall, F 値の各値をモデルごとに検証する。なお、各値はすべての小論文の文章に対して表 5 のすべての要素を数え上げたのち算出する。

表 5: 混同行列

		真の結果	
		正	負
予測結果	正	TP	FP
	負	FN	TN

5 実験結果

本章ではモデルごとの実験結果の比較と Recruit 社の Proof Reading API (PRA と略記) との比較を行う。図 6 の実験結果から、F 値の面で統計モデル (閾値 0.0) が最も精度が高く、次いでニューラルモデル、Proof Reading API の順で精度が高い。また、Precision においても統計モデル (閾値 0.0) が最も高い。Recall に関していえば、どのモデルでも閾値を上げることで 1 に近付くが、文字誤り検出の観点からは無意味である。

表 6: 各モデルの精度

モデル	Precision	Recall	F 値
統計 (閾値 0.0)	0.0357	0.287	0.0636
統計 (閾値 0.05)	0.0319	0.629	0.0607
統計 (閾値 0.3)	0.0307	0.846	0.0592
NN モデル	0.0284	0.367	0.0528
PRA	0.0324	0.134	0.0521

6 おわりに

本稿では統計的言語モデルおよび LSTM を利用した形態素ベースの誤字検出モデルを提案した。

実験結果より、統計モデルにおける文字誤り検出精度が最も高く、ニューラルモデル、Proof Reading API に対して高い精度を出したことを報告する。

しかしながら、すべてのモデルで共通して Precision が非常に低く、精度を上げるためにはこれを改善する必要がある。Precision が低くなる原因としては、文

字誤りの該当箇所を直接特定できていないこと、および文字誤りとは関係のない箇所を多く指摘してしまっていることにある。

図7に実際に小論文中にあった文字誤りを提案手法によって指摘する。この場合、「柔軟」の「軟」の字が誤っている。両モデルは共に誤りを正しく指摘しているが、誤り箇所以外の助詞も誤りとして指摘している。このように、文字誤りの前後を含めて多数の文字誤りを指摘している。この例では、誤り箇所を適切に指摘しているが文字誤り箇所を直接特定できずに周辺のみを指摘する場合もある。

これらの課題を解決するためにニューラルモデルで学習データを増やす必要がある。また、双方向のモデルを構築することで現状の単方向のモデルより精度の向上が望めると考える。

表 7: 文字誤り検出

入力	... 各国の文化やニーズに柔軟に...
統計	... 各国の文化やニーズに柔軟に ...
NN	... 各国の文化やニーズに柔軟に ...

参考文献

- [1] Masayuki Ohno, Koichi Takeuchi, Kota Motojin, Masahiro Taguchi, Yoshihiko Inada, Masaya Iizuka, Tatsuhiko Abo, and Hitoshi Ueda. *Construction of Open Basic Data for Automatic Scoring of Essay and Evaluation of Automatic Scoring Method at Current Stage*. PACLING-2017, 2017.
- [2] 大野雅之, 泉仁宏太, 竹内孔一, 小畑友也, 田口雅弘, 稲田佳彦, 飯塚誠也, 阿保達彦, 上田均. 参照データとidfを利用した事前採点不要な小論文評価手法. 言語理解とコミュニケーション研究会, pp. 368–371, 2018.
- [3] 泉仁宏太, 竹内孔一, 大野雅幸, 田口雅弘, 稲田佳彦, 飯塚誠也, 阿保達彦, 上田均. 小論文採点支援のための関連文書取得法の考察. 電子情報通信学会技術研究報告. NLC, 言語理解とコミュニケーション, pp. 47–51, 2017.
- [4] 竹内孔一, 大野雅幸, 泉仁宏太, 田口雅弘, 稲田佳彦, 飯塚誠也, 阿保達彦, 上田均. 小論文の自動採点に向けたオープンな基本データの構築および現段階での自動採点手法の評価. 言語処理学会第23回年次大会発表論文集, pp. 839–842, 2017.
- [5] 竹内孔一, 松本裕治. 統計的言語モデルを用いたocr誤り訂正システムの構築. 情報処理学会論文誌, Vol. 40, No. 6, pp. 2679–2689, 1999.
- [6] Chiao-Wen Li, Jhih-Jie Chen, and Jason S. Chang. Chinese spelling check based on neural machine translation. *Proceedings of PACLIC 2018.*, 2018.
- [7] Jun Liu, Fei Cheng, Yiran Wang, Hiroyuki Shindo, and Yuji Matsumoto. Automatic error correction on japanese functional expressions using character-based neural machine translation. *Proceedings of PACLIC 2018*, 2018.
- [8] 石岡恒憲. コンピュータ上で実施する記述式試験—エッセイタイプ, 短答式, マルチメディア利用について—. 電子情報通信学会誌, Vol. 99, No. 10, pp. 1005–1011, 2016.
- [9] 石岡恒憲, 亀田雅之, 劉東岳. 人工知能を利用した短答式記述採点支援システムの開発. 電子情報通信学会技術研究報告. NLC, 言語理解とコミュニケーション, pp. 87–92, 2016.
- [10] 浅原正幸, 岡照晃. nwjc2vec: 『国語研日本語ウェブコーパス』に基づく単語の分散表現データ. 言語処理学会第23回年次大会, pp. 94–97, 2017.
- [11] 石川巧. 「いい文章」ってなんだ?—入試作文・小論文の歴史. 筑摩書房, 2010.
- [12] Tshuneori Ishioka, Kohei Yamaguchi, and Thuneori Mine. Rubric-based Automated Japanese Short-answer Scoring and Support System Applied to QALab-3. In *Proceedings of the 13th NTCIR Conference on Evaluation of Information Access Technologies*, pp. 152–158, 2017.
- [13] 石岡恒憲. 日本語小論文の自動採点および作文支援システムの開発. 科学研究費補助金研究成果報告書, 2007.