

コーパスを利用したレポート自動採点品質の向上 Improving Automated Essay Scoring Quality with Corpus-Based Similarity

山本 恵[†] 梅村 信夫[‡] 河野 浩之^{†*}
Megumi Yamamoto Nobuo Umemura Hiroyuki Kawano

1. はじめに

筆者らは、レポート作成指導、教員の採点の厳正化、負担軽減を目的として、ループリックに基づく LMS 上の自動採点システムを構築している。具体的には 5 つの評価観点からなる教員の手動採点用ループリックを作成し、それらを細分化した評価項目を自動採点するものである。またこれらの自動採点結果からサポートベクターマシン (SVM) により総合成績レベルの分類を行うことができる。現段階では分類精度は 53.6% と低い。これは分類器作成のためのデータ件数が少ないこと、細分化した自動採点評価項目の中に精度が高くない採点方法があることが原因であると推測する。分類のための学習データは随時できるため、今後成績に十分なバラツキがあるデータを学習データとして増やすことで、ある程度の改善が見込める。しかしながら自動採点方法そのものが誤りであれば、データを増やしても精度の向上は期待できない。したがって計算方法や計算に用いる評価基準を見直すことが、精度向上に必要であると考える。そこで、本稿では、これらの取り組みのうち、特に「語彙の水準」の算出に着目し、精度を高める取り組みを行ったので、報告する。

2. 自動採点モデルの概要

採点システムの中核として、5 種の評価観点 (Contents: 課題の理解度と解答内容の妥当性, Structure: 論理的な展開, Evidence: 資料と根拠の妥当性, Style: 文章作法の遵守と適切な推敲, Skill: 読みやすさ・表現の巧みさ) を持つレポート採点用ループリックを提案した。

これらの評価観点を機械で処理するために、25 項目に細分化し自動採点用ループリックとも言うべき評価項目を設定している。このうち、Style, Skill の細分化した評価項目を表 2 に示す。すべての項目について自動採点し、その値を特徴量として SVM で総合成績レベル算出のための分類器を作成している。

表 1 の Style に従属する評価項目 (文体や誤字脱字など文章校正に関わる部分) は精度改善に向けた検討の余地があまりない。また Skill の文長や漢字使用率についても然りである。一方で語彙の水準や豊富さは様々な計算方法が選択肢としてあり、妥当性について議論の余地がある。特に語彙の水準は 0-6 点の範囲で各語彙のレベルを算入し採点しており、他の項目よりも評価の差が大きくなる可能性がある。そこで、レポートの語彙水準の採点方法を再検討することとした。

現在語彙水準表として利用しているのは、砂川らにより作成された日本語教育のための語彙表であり、18000 語について、6 段階 (初級前半, 初級後半, 中級前半, 中級後半, 上級前半, 上級後半) のレベルを付与している [2]。我々の自動採点システムでは、学生レポートの語彙の水準を計算するにあたり、レポートから抽出した形態素 (名詞・動詞・形容詞) について、索引語 (単語) 文書行列を作

成し、上記語彙表に存在する単語すべてのレベルを割り当て、平均を求める。しかしながら、この語彙水準表は日本語を教育するための基礎的な語に絞られており、小中学校の教科書レベルの単語は網羅されているが、より難易度の高い単語が含まれていない。本研究では大学生のレポートの評価を目的としており、ループリックの評価観点 Skill (文章の読みやすさや表現の巧みさなどの作文スキル) を細分化した「語彙の水準」を計算するための指標として語彙水準表が必要となるため、見直す必要がある。そこである程度専門的な単語も含まれたレポート評価のための語彙水準表を作成するために、まず広範な単語数を有し、各単語の語彙の水準が計算できるようなコーパスを作成することとした。

表 1 評価観点 Style, Skill の細分化した評価項目

評価観点	評価項目	評価基準値
[Style]	15 文体の統一性	0
	16 誤字・脱字の排除	0
	17 構文の妥当性	1.988
	18 主述関係の妥当性	1
	19 句読点の妥当性	15
	20 冗長さ・二重否定の排除	0
	21 表記ゆれ・曖昧さの排除	0
[Skill]	22 漢字の使用率	35%
	23 文長の妥当性	62.5 文字
	24 語彙の豊富さ	値が高い程、評価が高い
	25 語彙の水準	

3. 関連研究

語彙の水準は、各単語の難易度と言い換えることができる。語彙の難易度測定については、すでにシステムで利用している日本語教育語彙表をはじめとする言語教育のための研究が盛んである。日本語教育語彙表は、均衡コーパスを基に、複数ジャンルのテキストから語の使用を全般的に捉えて、一般的な日本語教育としての難易度を付与している。またコーパスの範囲を専門的な分野に絞り、専門性との関連から難易度を推測するものや [3][4]、均衡コーパスや人手を介する言語資源に頼らずに、生コーパスから直接難易度を推測するもの [5] などが提案されている。

4. レポート採点のためのコーパスの試作

4.1 コーパスの対象範囲

本研究では、情報リテラシー科目履修生の「表計算ソフトによるデータ分析」をテーマとするレポート 83 件の採点に限定した語彙の難易度示す表 (現段階では難易度表とする) を、採点のためのコーパスとして試作した。出現頻度が低いものほど、一般的に使われていない語として、難易度が高いとする。したがって採点用のコーパスでは各単語に出現頻度の情報を示し、難易度の高さを示す指標とす

[†]名古屋外国語大学, Nagoya University. of Foreign Studies

[‡]名古屋学芸大学, Nagoya University of Arts and Sciences

^{*}南山大学, Nanzan University

る。もとななる言語資源は、専門分野に特化しない一般的なレポート課題を想定し、出現する日本語をできるだけ広範囲にカバーする目的で、非均衡コーパスである Wikipedia 日本語版データ (以下 Wiki) を利用する。

4.2 作成手順

はじめにコーパス作成のための単語の抽出を行う。学生のレポートデータをテキスト型データ分析ソフトである khCorder の茶釜を利用して、形態素解析し、名詞を抽出した。名詞は文章の内容を表現する意味語としての役割を持ち、語彙力を顕著に示すため、文章の重要度や難易度、類似度を測定する研究の多くで名詞を取り上げている[6]。本研究でも学生レポートの採点に向けた難易度表の作成を目的としているため、コーパス作成の対象を名詞とした。また労働問題”など、複合語を検出し、抽出リストに追加した。83名のレポートで636件が採点用コーパスの対象となる。

次に、抽出した語 (以下、単語) の出現頻度を調べるための Wiki コーパスを次の手順で作成する。

(1) Wiki コーパスの入手

Wiki は膨大な情報量を有している。ここでは Web 上で提供されている Wiki 日本語版データベースのうち、(6/24 時点の最新版である全記事の抽象データ) をまとめた XML ファイルを利用する。4つのフィールド (title, url, abstract, link) を持つ総記事数 1065396 件が認められた。処理の負荷を考慮し、任意の 171420 件を処理の対象とする。

(2) データベースシステムを作成

SQL Server Express Edition に Wiki の XML ファイルをインポートしてデータベースを作成する。得られたフィールドのうち、各記事の抽象データ本文が格納されている abstract のみ使用する。

(3) Wiki 内の単語の出現頻度を調べる

各記事毎に単語の出現回数を算出し、単語をキーとしたテーブルに格納する。すべての単語の出現頻度を調べる。

(4) レポート採点コーパスの作成

単語ごとに、Wiki 全体での単語の出現頻度総数、および出現した記事の件数を求める。1記事あたりの出現頻度を求め、難易度を示す指標とする。

4.3 考察

表 2・表 3 は、以上の手順により作成した採点用コーパスの一部である。表 2 は、コーパスの全単語 646 件を難易度順に並べ替えた結果の上位 10 の単語である。一方表 3 は難易度下位の単語 10 であり、出現頻度が高かったことを示す。単語を見ると、難易度に関する本研究での考え方は、現在水準表として用いている日本語教育語彙表と矛盾はみられない。上位の単語は Wiki に一度も出現していないが、90 万件近い残りの Wiki データを処理すれば、これらの単語を網羅した広範なコーパスを作成できる可能性がある。

5. まとめと今後

本研究では、構築した LMS 上の自動採点システムの採点精度の向上を目指した取り組みについて報告した。教員が評定ガイドとするべきレポート採点用ルーブリックを提案し、それに基づいたシステムを構築している。自動採点項目のうち、学生が記述したレポートの語彙レベルを求め

る際使用している語彙の水準表を改善する目的で、Wiki を利用した語彙の難易度を示すコーパスを試作した。今後は Wiki 本文のコーパスを利用するなど、精度を高め、実際にこれらを利用してレポートの採点精度を確認する予定である。

表 2 難易度上位 10 の単語*

名詞	レポート中の出現頻度	Wiki 記事内の出現頻度総数	出現した Wiki 記事数	1記事あたりの出現頻度	日本語教育語彙表で示される水準
一石二鳥	1	0	0	0.0	5.上級前半
気風	1	0	0	0.0	—
実体験	1	0	0	0.0	—
目先	1	0	0	0.0	5.上級前半
歴然	1	0	0	0.0	—
リラックス	1	0	0	0.0	4.中級後半
固執	1	0	0	0.0	5.上級前半
買い占め	1	0	0	0.0	—
密航	1	0	0	0.0	—
民泊	1	0	0	0.0	—

表 3 難易度下位 10 の単語*

名詞	全レポート中の出現頻度	Wiki 記事内の出現頻度総数	出現した Wiki 記事数	1記事あたりの出現頻度	日本語教育語彙水準表で示される水準
漫画	1	3106	2680	1.16	2.初級後半
会社	6	3205	2361	1.36	1.初級前半
人口	1	3554	3474	1.02	2.初級後半
アメリカ	16	4103	3511	1.17	—
鉄道	2	4238	3288	1.29	4.中級後半
大学	2	4615	3032	1.52	1.初級前半
学校	5	5123	2506	2.04	1.初級前半
京都	2	6082	5344	1.14	—
東京	11	7472	6245	1.20	—
日本	292	21811	18491	1.18	1.初級前半

*難易度表の名詞の総数 n=646

参考文献

- [1] 山本 恵, 梅村 信夫, 河野 浩之, “ルーブリックに基づくレポート自動採点システム”, 大学 ICT 推進協議会 2016 年度年次大会論文集 (2016).
- [2] 砂川有里子, “学習辞書編集支援データベース作成について”, 『学習辞書科研』プロジェクトの日本語教育連絡会議論文集 (2012).
- [3] 滝川 真弘, 山名 早人, “ノイズに頑健な分野別単語排他度の提案 Twitter ユーザの専門性推定への適用”, DEIM Forum (2017).
- [4] 滝川 真弘, 山名 早人, “特定分野を対象とした単語重要度計算手法の提案と Twitter における専門性推定への適応”, FIT2016, pp.1-7 (2016).
- [5] 江原 遥, “生コーパスからの単語難易度関連指標の予測”, 言語処理学会 第 23 回年次大会 発表論文集, pp.843-846 (2017).
- [6] 舟木 類佳, 黒田 久泰, “難易度及び類似度を用いたコンピューター関連書籍推薦システムの開発”, 情報処理学会研究報告, Vol.2014-NL-215, No.8, pp.1-6 (2014).