

オンライン開発環境を用いた文章自動採点システムの構築

Establishment of an automatic essay scoring system based on online development environment

高橋 光騎^{*1}, 関根 務^{*1}
Kohki TAKAHASHI^{*1}, Tsutomu SEKINE^{*2}
^{*1} 成蹊大学理工学部

^{*1} Faculty of Science and Technology, Seikei University
Email: us183067@cc.seikei.ac.jp

あらまし：情報通信技術を用いた教育環境を整備することは、新型コロナウイルス感染症の禍中にある近年では各教育機関において益々重要になってきている。本研究では、無償で公開されているオンラインリソースをいくつか組み合わせて、日本語の記述式解答に対する自動採点システムを開発した。また、そのシステムの基本的な特性について調べた。

キーワード：Automated essay scoring, Latent Semantic Analysis, Key sentence, Key word

1. はじめに

情報通信技術を用いた教育環境を整備することは、新型コロナウイルス感染症の禍中にある近年では各教育機関において益々重要になってきている。文章の自動採点に関する研究は、1960年代頃からさまざまな形で報告されてきている。英文の自動採点の代表的なものとしては、Project Essay Grader や Intelligent Essay Assessor 等があることが知られている。一方で、和文の自動採点に関する研究はまだ少ないようである。日本語の短答式の自動採点システムに関しては、いくつかの研究報告が存在しているが、小論文のような比較的長文を対象とした研究報告はさらに僅少である。

近年、自然言語処理技術が飛躍的に発展し、さまざまなオンラインリソースが無償で利用できるようになってきている。公開されているオンラインリソースをいくつか組み合わせることによって、比較的容易に文章の自動採点システムを構築することが可能になると考えられる。その中で、機械学習を活用した環境をつくることも可能ではあるが、前処理として学習モデルを構築するための日本語文章のリソースを入手することは極めて難しい。

そこで本研究では、オンラインリソースをいくつか組み合わせて、日本語の記述式解答に対する自動採点システムを開発した。また、そのシステムの特性について調べた。

2. 文章自動採点システムの開発

採点する対象として、著者の1人が過去の授業内で各学生に課したリアクションペーパー（以下、RP）を用いることにした。授業においては、オンラインシステム上に RP を書き込むように指示しており、受講者 72 名からの回答が記録されている。自動採点システムの開発環境としては、Google Docs Editors を利用した。特に、Google spreadsheet と Google Apps script を利用した。潜在意味分析（Latent Semantic Analysis）を用いて、重要語（以下、KW）、重要文

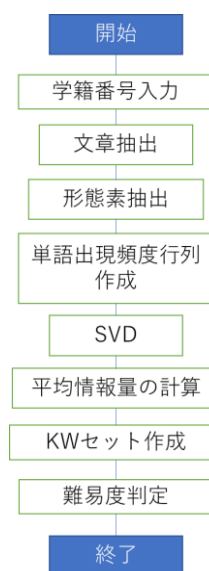


図 1 基準採点処理

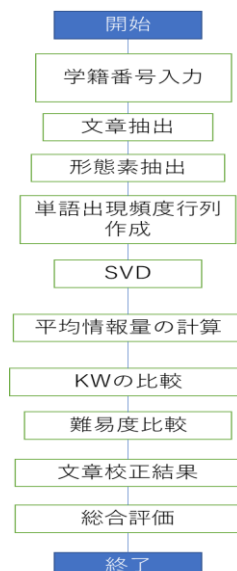


図 2 採点対象側処理

（以下、KS）を選出し、採点することにした。また、日本語の難易度の診断には、「やさしにチェック」 という外部サイトの結果を参照した。この外部サイトでは、文章を挿入することにより、語彙・漢字・硬さ・長さ・文法の 5 つを 5 段階で評価できる。ただし、本研究では実際にいくつか文章を採点したとき、模範解答の文章と書いている趣旨は一致しているのにも関わらず、文法の評価に差が生じることが確認できたため、文法を除いた 4 項目で評価することにした。

図 1 に基準採点処理の流れを示す。最初に優良な記述内容を含む RP を採点者が選定する。学籍番号を入力することにより、該当生徒の文章を抽出する。そして、その文章の形態素を抽出し、それを基に単語（行）×文（列）で構成された単語出現頻度行列を作成する。その行列に対して特異値分解（以下、SVD と呼ぶ）を行い、再構成された行列に対して平

均情報量を計算することで、優良な記述内容を含む RP の KS (以下、基準 KS) を選定し、さらにそこから KW (以下、基準 KW) を選定する。そこで選定された基準 KW の同義語を 10 個選出する[4]。最後にやさしにちチェッカーで文章の難易度を評価する。

図 2 に、文章採点処理の流れを示す。図 1 と同様に平均情報量の計算までの作業を行い、採点対象となっている文章の KS (以下、採点対象 KS) を選出して、そこからさらに KW (以下、採点対象 KW) を選出する。そこから基準 KW (同義語含む) と採点対象 KW の比較を行い、やさしにちチェッカーを用いた基準採点の結果と採点対象となっている文章で得られている結果の比較を行い、最後に文章校正の結果も確認している。

ここで、形態素解析と文章校正に関しては、Yahoo! デベロッパーネットワークから公開されている日本語形態素解析と校正支援を API 機能を通して利用した。総合評価は、KW 比較得点と難易度比較得点の平均を算出し、文章校正結果を踏まえて減点したものを Total score として数値化した。

各文章の評価結果を基にして、自動採点システムの特徴を調べるために、Total score、文字数、KW 比較得点、KS 数、難易度、文章校正、文章数、形態素数を抽出するプログラムを作成し、それぞれの関係性について分析した。

3. 結果と考察

採点結果として出てきているいくつかの採点対象 KS を確認したところ、講義で重要だったものと一致していることが確認できた。このことから、構築した自動採点システムは正しく KS を抽出でき、一定の基準で公平に評価できていると考えられる。

次に形態素数と文字数の関係を図 3 に示す。図 3 から形態素数と文字数には相関関係があることがわかる。次に、文字数と KW 比較得点の関係を図 4 に示す。図 4 から文字数と KW 比較得点には相関関係がないことがわかった。文字数を多くした RP を無条件に高く評価してしまうことは RP の採点において適切とはいえない。構築した自動採点システムでは、文字数に比例して KW 比較得点が高くなることはないことが確認できた。この結果は、2 つの要因から影響していると考えられる。1 つ目は平均情報量の高い文を KS として抽出して、そこから KW を抽出している点である。すなわち文字数の多い RP であっても平均情報量の低い文は採点の対象にならず、得点に影響を及ぼさなくなっている。2 つ目は、比較 KS に対する得点の平均を採点対象 KW 得点としているためだと考えられる。これによって 1 つの基準 KS に対して複数の比較 KS が一致したとしても得点が高まることはない。言い換えれば、KS の数が多くなったとしても、無条件に得点が高くなることを避けられる。

これを踏まえて、次に採点対象 KS 数と採点対象

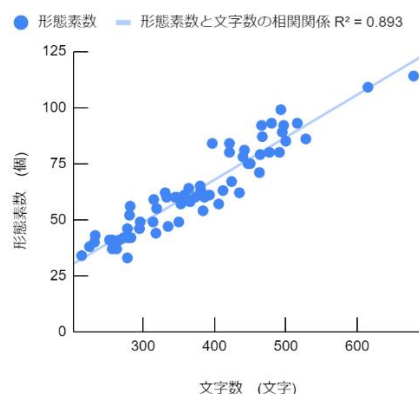


図 2 文字数と形態素数の関係

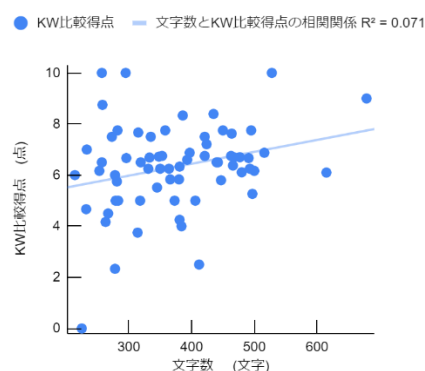


図 3 文字数と KW 比較得点の関係

KW 得点について調べたところ、KW 比較得点と採点対象 KS 数に相関関係がないことがわかった。このことから採点対象 KS 数が KW 比較得点に影響を与えていないことを確認できた。一方で、採点対象 KS 数が 1 でありながら、満点の 10 点をとっている RP があるとわかった。その RP を確認したところ 1 つの KS に基準 KW より多くの採点対象 KW があり、基準 KW と一致している状況にあった。多くの採点対象 KW があればあるほど、必然的に基準 KW と一致する可能性は高くなるが、基準 KW より多くの採点対象 KW があるということは関係のない KW も多いということを意味している。したがって、関係のない KW に対する減点を行うなどの改善点が考えられる。そうすることにより、基準 KS の一致度に基づいて、より適切な評価が可能になると考えられる。

4. まとめ

本研究では、オンラインリソースをいくつか組み合わせ、日本語の記述式解答に対する自動採点システムを開発した。また、そのシステムの特徴について調べた。その結果、開発した自動採点システムの基本的な特性を明らかにすることができた。

参考文献は省略