

ルーブリックに基づくレポート自動採点システムの構築

Implementation of Automated Report Scoring System Based on Rubric

山本 恵† 梅村 信夫‡ 河野 浩之†‡

名古屋外国語大学† 名古屋学芸大学‡ 南山大学†‡

1. はじめに

学生のレポート作成能力育成および教員による厳正な採点支援を行うために、レポート採点ルーブリックを提案し、採点の自動化を目指している。開発途上の本システムは、ルーブリックに基づく自動採点と手動採点を組み合わせるハイブリッド型である。

以下、構築中のシステムのアーキテクチャと、ルーブリックを取り入れた採点方針および採点方法について報告する。

2. 先行研究

実用段階にある自動採点システムに E-rater[1] (TOEFL 試験等) や IntelliMetric[2] (米医学部入学資格試験等) がある。前者は、構造／組織／内容に関する 12 変量の重回帰モデルで採点する。後者は、一貫性／内容／構成などを、プロの評価者が採点した模範解答から、機械学習により採点ルールを求める。

センター試験利用を目的とする Jess[3] は E-rater を参考に、修辞／論理構成／内容を採点する。プロの評価者ではなくプロのライターの記述からメトリクスなどの統計量の理想的な分布を求め、外れ値となる文書を減点法で採点する。

上記の先行研究では、人間の採点結果に近い評価を行う自動採点システムを開発している。

3. 提案システムのアーキテクチャ

本研究は、大学の教育現場での活用を目指し、自動採点と手動採点を統合する。自動採点領域では E-rater に倣い、重回帰モデルで算出する。

システムアーキテクチャを図 1 に示す。学生のレポート提出や教員による採点では、LMS (Moodle) を操作する。操作と処理の流れを、図 1-I 内の番号により説明する。

- (a) 学生は、Moodle 上の授業用学習コースの提出場所 (課題提出モジュール: オンラインテキスト) に学生レポートテキストデータ (以下、文書と表記) を入力
- (b) 文書は Moodle 内データベースに格納される

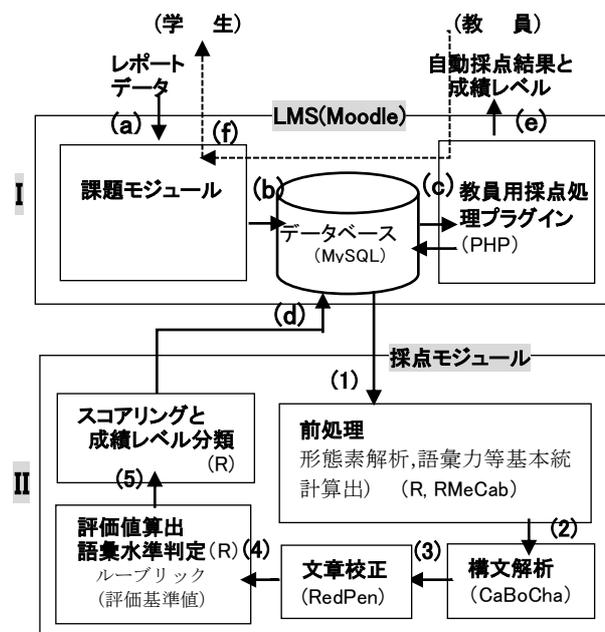


図 1 システムアーキテクチャと処理の流れ

- (c) 教員は、採点システムプラグインから採点対象文書をクラス単位で指定し、ブラウザして確認し採点を開始
- (d) 採点モジュールから、自動採点対象である評価観点の点数と全体の成績レベルを採点結果として受け取る
- (e) (d)を参考に手動採点部分を採点
- (f) 成績を学生にフィードバック

4. 採点モジュール

本章では採点の方針と実装中の R を用いた採点モジュールの詳細を述べる。

4.1 採点の方針

大学等組織レベルで作成・活用されているルーブリック等を参考に、レポート採点ルーブリックを提案する (表 1)。5 つの評価観点のうち、Content/Structure/Evidence を手動で、Style/Skill を自動採点する。自動採点結果をもとに機械学習アルゴリズムを用いて分類し、手動採点部を含めた成績レベル A+~D を提示し、教員が主導採点する。現段階では、サポートベクターマシン等のアルゴリズムの精度を比較・検討している。

†Megumi Yamamoto · Nagoya University, of Foreign Studies

‡Nobuo Umemura · Nagoya University of Arts and Sciences

†‡Hiroyuki Kawano · Nanzan University

表1 レポート評価ルーブリックと採点の提案

評価観点(レベルと配点) D(0-1),C(2-3),B(4-5),A(6-7),A+(8-9)	採点方法
[Content] 課題の理解度と解答(記述)内容の妥当性	← 成績資料 (成績 レベル 分類結果)
[Structure] 論理的な展開	
[Evidence] 資料と根拠(エビデンス)の妥当性	
[Style] 文章作法の遵守と適切な推敲	自動
[Skill] 読みやすさ・表現の巧みさ	

4.2 自動採点モジュールの詳細

自動採点モジュールでは、表2のように評価観点を11の評価項目に細分化して採点する。

表2 自動採点項目

評価観点	評価項目	評価基準値	重み付け
[Style]	15 文体の統一性	0	0.20
	16 誤字・脱字の排除	0	0.23
	17 構文の妥当性	1.988	0.35
	18 主述関係の妥当性	1	-0.09
	19 句読点の妥当性	15	0.37
	20 冗長さ・二重否定の排除	0	0.06
	21 表記ゆれ・曖昧さの排除	0	0.14
[Skill]	22 漢字の使用率	35%	-0.14
	23 文長の妥当性	62.5 文字	-0.12
	24 語彙の豊富さ	値が高い程、 評価が高い	-0.04
	25 語彙の水準		0.04

自動採点モジュールの処理の詳細は次のとおりである(図1-IIの番号に対応)。

- (1) RMeCabにより文書を形態素解析し索引語を抽出し、品詞情報を付加。索引語の出現頻度を示す文書行列の作成、文書毎に、文字数、文節数、トークン比を算出
- (2) CaBoChaによる構文解析の実行。全索引語の係り先との距離の平均を算出
- (3) 文章校正ツール (RedPen サーバ[4])により、誤りの数を文書毎、評価項目毎に算出
- (4) (1)~(3)から各文書のStyle項目の誤り率を算出し、自動採点ルーブリックパラメータとの差異を各文書の評価値として算出。また、語彙水準表[5]を用いて各学生の索引語の平均水準を求め、Skill25の評価値として算出
- (5) (4)の評価値を0-9点に補正し、重み付け(表2)により自動採点結果算出。また、補正前の結果を用いてサポートベクターマシンによりA+~Dに分類
- (d) 教員に採点結果として提示

以上が自動採点の流れである。各評価項目の計算式詳細は文献[6]を参照されたい。

5. 採点実験

情報リテラシー科目(表計算ソフトを使ったデータ解析をテーマとする)を受講する大学1年次生(2クラス)が提出した小レポートの採点を行ったところ、表3の結果を得た。現段階の実験結果において、Skillの評価観点の自動採点結果が、手動採点と相関が低く、課題を残している。

表3 提案システムと手動採点との相関

レポート	文書数	平均文字数	標準偏差	手動採点との相関		
				Style	Skill	成績レベルの分類
A	43	427.6	216.0	0.602	0.210	0.581
B	40	325.5	171.7	0.463	0.089	0.550

6. おわりに

本研究では、ルーブリックに基づく自動採点システムのアーキテクチャを提案し、構築中のシステムの概要と処理の流れを中心に報告した。提案システムでは、評価項目を手動と自動に区分して採点し、自動採点結果から手動採点部分を予測して成績レベルを提示することで教員の採点支援を目指している。今後、採点結果表示インターフェースの作成、テストデータによる処理速度の評価、自動採点評価項目による採点アルゴリズムの改善と、機械学習アルゴリズムによる採点結果の分類精度の向上をめざす。

参考文献

- [1] Yigal Attali, Jill Burstein, Automated Essay Scoring With e-rater® V.2, The Journal of Technology Learning and Assessment, Vol. 4, No. 3, p. 22, 2006.
- [2] 石岡恒憲, 小論文およびエッセイの自動評価採点における研究動向, 人工知能学会誌, Vol. 23, pp. 17-24, 2008.
- [3] 石岡恒憲, コンピュータ上で実施する記述式試験について, 情報処理学会研究報告, Vol. 2016-CE-133, No. 19, 2016.
- [4] RedPen, <http://redpen.cc/>, (参照 2016-12-1)
- [5] 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.
- [6] 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.