

【D-5】

Moodleを用いたLMS上の 自動採点システムの試作

山本 恵・名古屋外国語大学
梅村 信夫・名古屋学芸大学
河野 浩之・南山大学

2016年度 教育改革ICT戦略大会 東京, 2016/9/8(木)

研究概要

◆ 採点支援システムのアーキテクチャ

- (1) LMS (Moodle)
- (2) 形態素解析, 構文解析 (MeCab, CaboCha)
- (3) 分類器 (SVM, 決定木)

◆ 自動採点項目と採点方法

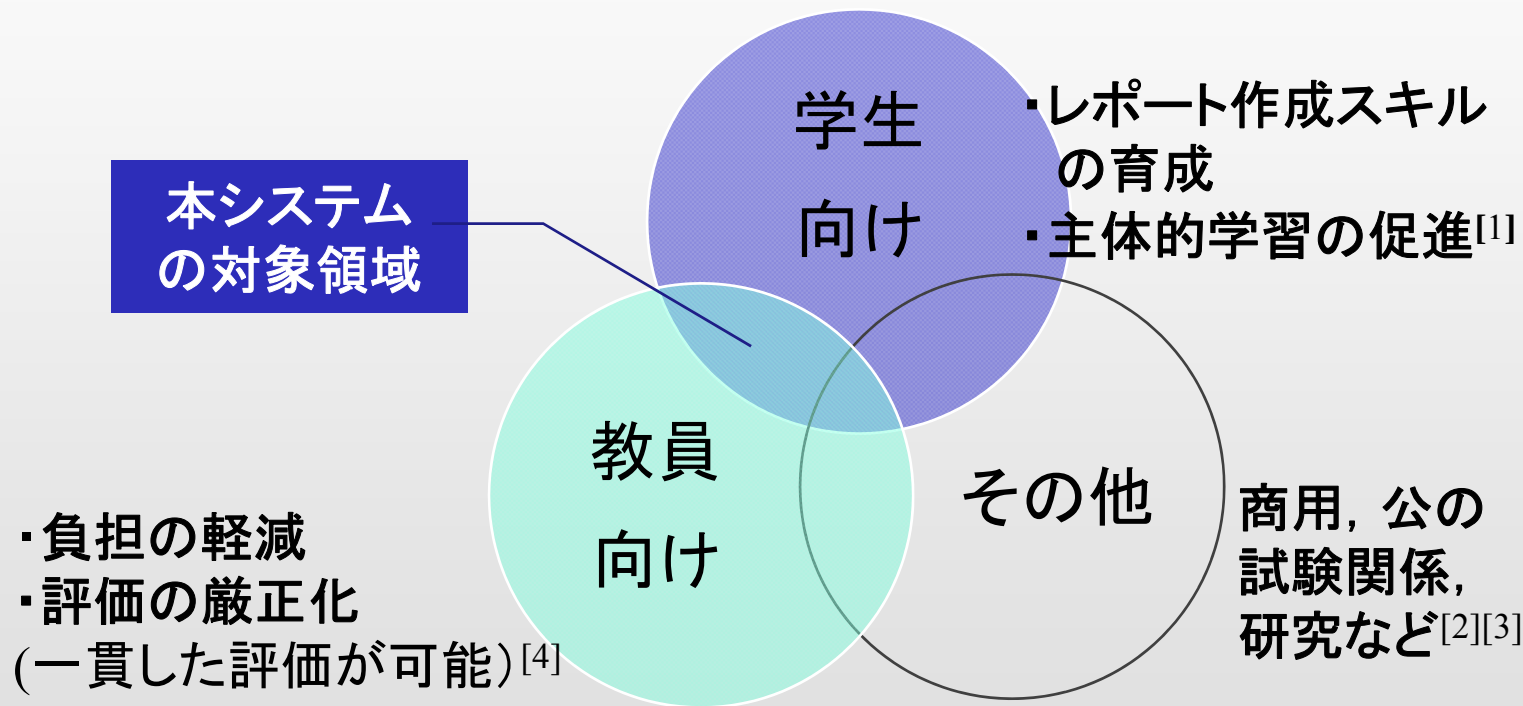
- (1) ルーブリックの提案
- (2) 採点アルゴリズムの検討

◆ 評価実験

- (1) 情報リテラシー関連レポートの採点
- (2) 教員採点と自動採点の相関
- (3) SVM等による成績レベルの分類

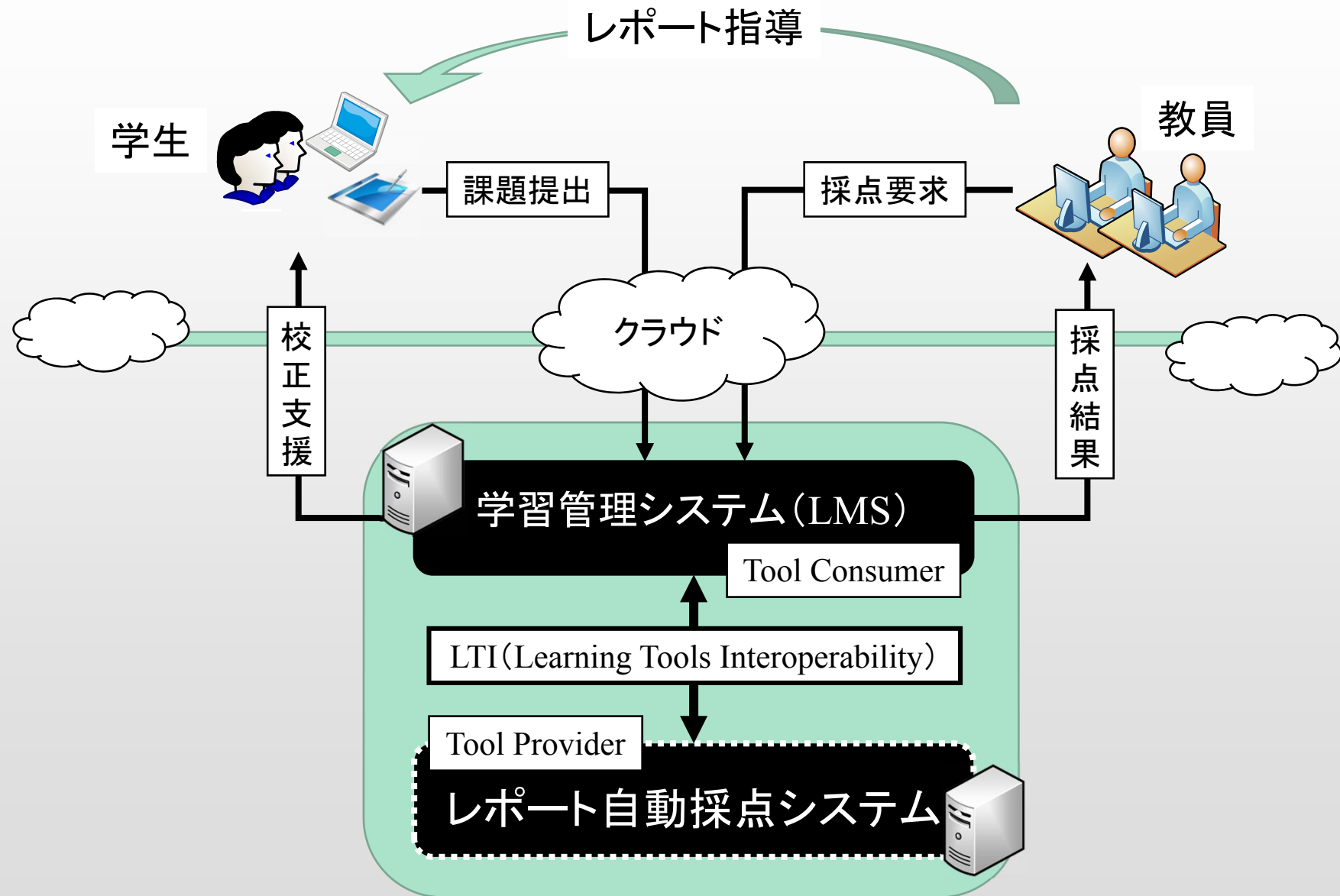
1. 研究の位置付け

自動採点システムの目的



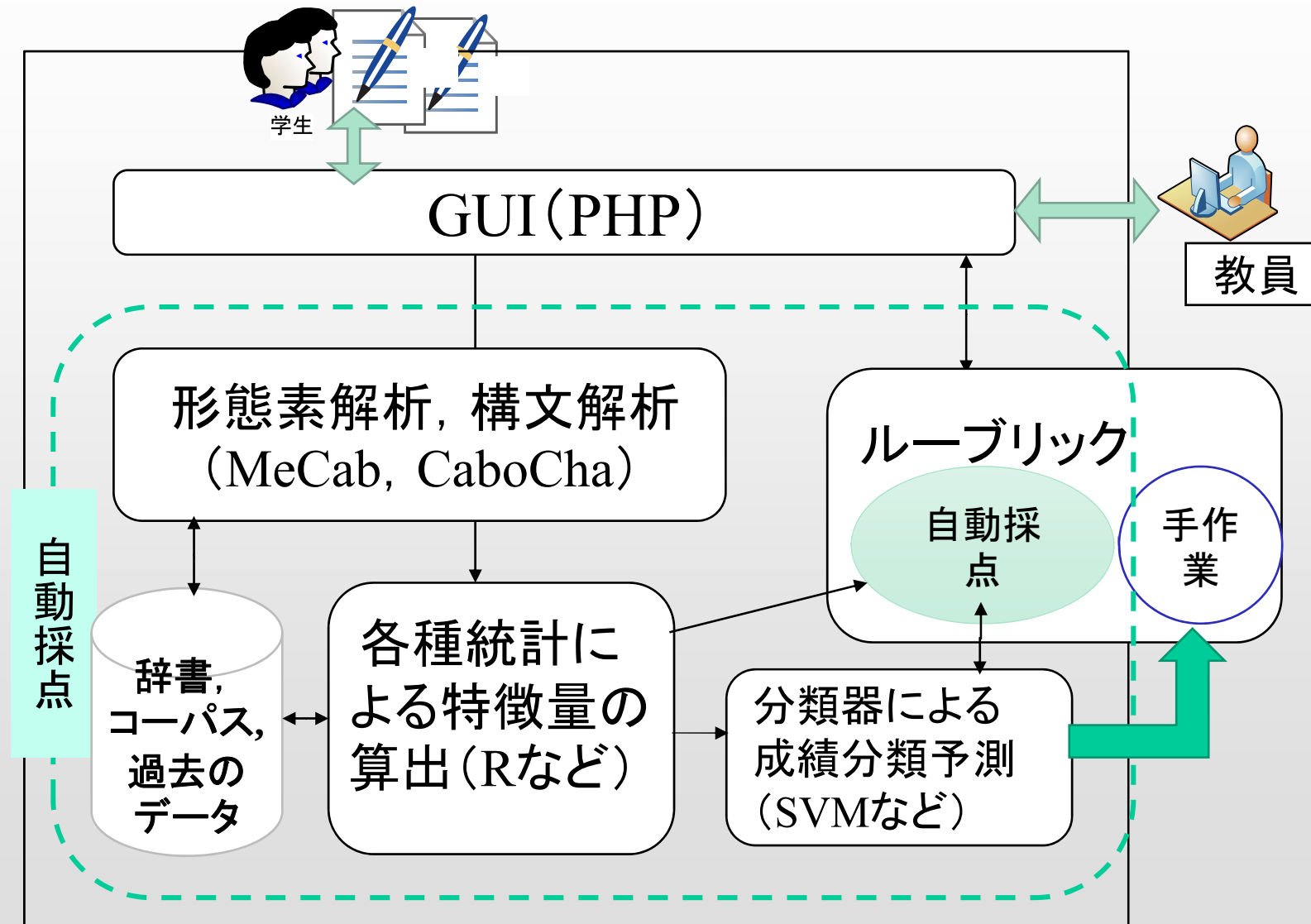
※自然言語処理, 語学教育, 機械学習, 知識情報など様々な分野で研究

2. 研究の目的



3. システムの概要

採点の流れ



4. 自動採点項目と採点方法(ルーズブリックの提案)

評価観点	達成レベルと配点		
	4	～	0
[Content] 課題の理解度と解答 (記述)内容の妥当性	課題を完全に理解し、 的確に解答している。 関連用語を適切に用 いている。改善の必要 がない。	:	解答内容が、課題とは 無関係である。
[Structure] 論理的な展開	順序立てて理論を展 開している。意見・主 張があり、説得力があ る。	:	記述内容にまとまりが ない。
[Evidence] 資料と根拠(エビデ ンス)の妥当性	当該の学問分野にふ さわしく、信頼でき関 連性のある資料を、う まく使いこなしている。	:	資料を全く参照してい ない。根拠を示してい ない。
[Style] 文章作法の遵守と適 切な推敲	よく推敲している。全 く誤りがない。	:	複数にわたってルール を守っていない。文章 が全く推敲されていな い。
[Skill] 読みやすさ・表現の巧 みさ	読み手に明確に意味 を伝えることができ、 読みやすい。語彙が 豊富である。	:	文章が読み辛い。明ら かに文章スキルがない。

具体的な
採点評価
項目



評価項目	
1	論題と記述の合致度
2	主要な関連語の存在
3	出題意図の理解度
4	内容の総合評価
5	学修内容の理解度
6	論理性の水準
7	意見・主張の妥当性
8	事実と意見の区分け
9	説得力
10	参照資料の質水準
11	参照資料の関連性
12	論拠資料の妥当性
13	図表への説明付加
14	引用量の妥当性
15	文体の統一性
16	誤字・脱字の排除
17	構文の妥当性
18	主述関係の妥当性
19	句読点の妥当性
20	冗長さの排除
21	表記ゆれの排除
22	漢字の使用率
23	文長の妥当性
24	語彙の豊富さ
25	語彙の水準

<提案の基となったルーブリック>

「文章コミュニケーション VALUEルーブリック」松下(2013)[5] から引用

	キャップストーン 4	マイルストーン 3 2		ベンチマーク 1
文章作成の文脈と目的	文脈・読者・目的について完璧な理解を示し、それによって、与えられた課題に対応し、作業(ワーク)のあらゆる要素に焦点をあてることができる。	文脈・読者・目的について適切な理解を示し、与えられた課題(例えば、読者・目的・文脈を結びつけること)に明確に焦点をあてることができる。	文脈・読者・目的や与えられた課題(例えば、読者の認知や了解事項への気づきを見せ始めること)への自覚を示す。	文脈・読者・目的や与えられた課題(例えば、読者としての授業者や自己の期待)に対し最低限の注意を示す。
内容の展開	適切で関連性があり説得力に富む内容を用いることによって、科目の習得ぶりを示すとともに、書き手の理解したことを伝え、作業(ワーク)全体を形づくることができる。	適切で関連性があり説得力に富む内容を用いることによって、学問分野の文脈の中で自分の考えを探究し、作業(ワーク)全体を形づくることができる。	適切で関連性のある内容を用いることによって、作業(ワーク)の大半を通じて、自分の考えを展開・探究することができる。	適切で関連性のある内容を用いることによって、作業(ワーク)の何れ所かで、シンプルな考えを展開することができる。
ジャンルと学問分野の約束事	特定の学問分野や文章作成課題に関連する広範な約束事(構成、内容、提示、書式、文体選択を含む)に対し、細かい注意を向けうまく遂行することができる。	特定の学問分野や文章作成課題に関連する重要な約束事(構成、内容、提示、文体選択を含む)を一貫性をもって使用することができる。	特定の学問分野や文章作成課題にふさわしいものとして、期待されることがら(基本的構成、内容、提示など)に従う。	基本的構成や提示について一貫した体系を使おうとしている。
資料(ソース)と根拠(エビデンス)	当該の学問分野やジャンルにふさわしい考えを展開するために、質が高く、信頼でき、関連性のある資料をうまく使いこなしている。	当該の学問分野やジャンルの中に位置づける考えを支持するために、信頼でき、関連性のある資料を一貫して使っている。	当該の学問分野やジャンルにふさわしい考えを支持するために、信頼できるか関連性がある資料を使おうとしている。	考えを支持するために、資料を使おうとしている。
構文と技法を操ること	読み手に明確かつ流暢に意味を伝えることができる格調ある言葉遣いをする。ほとんど全く誤りがない。	読み手に意味を伝える直截的な言葉遣いをする。減多に誤りがない。	文章に数か所誤りを含むが、明確に意味を伝える言葉遣いをする。	用語法に誤りがあるために、意味の伝達が妨げられるような言葉遣いをする。

CSE



SS

評価観点	評価項目	採点支援
[Content] 課題の理解度と解答(記述)内容の妥当性	1 論題と記述の合致度	○
	2 主	
	3 出	
	4 内容の総合評価	×
	5 学修内容の理解度	×
[Structure] 論理的な展開	6 論理性の水準	×
	7 意見・主張の妥当性	
	8 事実と意見の区分け	
	9 説得力	
	10 参照資料の質水準	
[Evidence] 資料と根拠(エビデンス)の妥当性	11 参照資料の関連性	
	12 論拠資料の妥当性	×
	13 図表への説明付加	×
	14 引用量の妥当性	△
	15 文体の統一性	○
[Style] 文章作法の遵守と適切な推敲	16 誤字・脱字の排除	○
	17 構文の妥当性	○
	18 主述関係の妥当性	○
	19 句読点の妥当性	○
	20 冗長さの排除	○
[Skill] 読みやすさ・表現の巧みさ	21 表記ゆれの排除	○
	22 漢字の使用率	○
	23 文法の妥当性	○
	24 語彙の豊富さ	○
	25 語彙の水準	○

<採点アルゴリズムの検討>

自動採点結果が、
SSやCSEを説明で
きるかどうかを検証

＜採点アルゴリズムの検討(Style項目群)＞

No	評価項目	評価内容	計算式	使用ソフト
15	文体の統一性	「である調」または「ですます調」での統一の可否	非統一の割合=非統一の数※1／文節総数※2	Word 2016文章 校正ツール
16	誤字・脱字の排除	誤字・脱字の有無	誤字・脱字出現率=誤字・脱字の数※1／文節総数※2	
17	構文の妥当性	係り受けに関与する文節間の平均距離(適正基準1.988 ^[6])	係り受けの平均距離=文書(レポート)の全係り受け数／文節総数※2	CaboCha MeCab
18	主述関係の妥当性	1文における係助詞の数(適正基準1.0)	係助詞の平均数=係助詞の数／文の数	
19	句読点の妥当性	句読点間平均文字数(適正範囲13～17文字, 適正基準35)	句読点間平均文字数=全文字数／句読点の数	MeCab R
20	冗長さの排除	ことばの重複や繰り返しの有無	冗長さ出現率=(重複語の数※1＋語の繰り返しの数)／文節総数※2	Word2016 文章校正 ツール TeMP※3 MeCab
21	表記ゆれの排除	表記ゆれの有無	表記ゆれ出現率=表記ゆれ出現数※1／文節総数※2	

※1 Microsoft® Office Word 2016 の文章校正ツールを利用してカウント

※2 1文書(レポート)の全文節数(CaboChaにより算出)

※3 TeMP 日本語テキストマイニング支援ツール^[7]

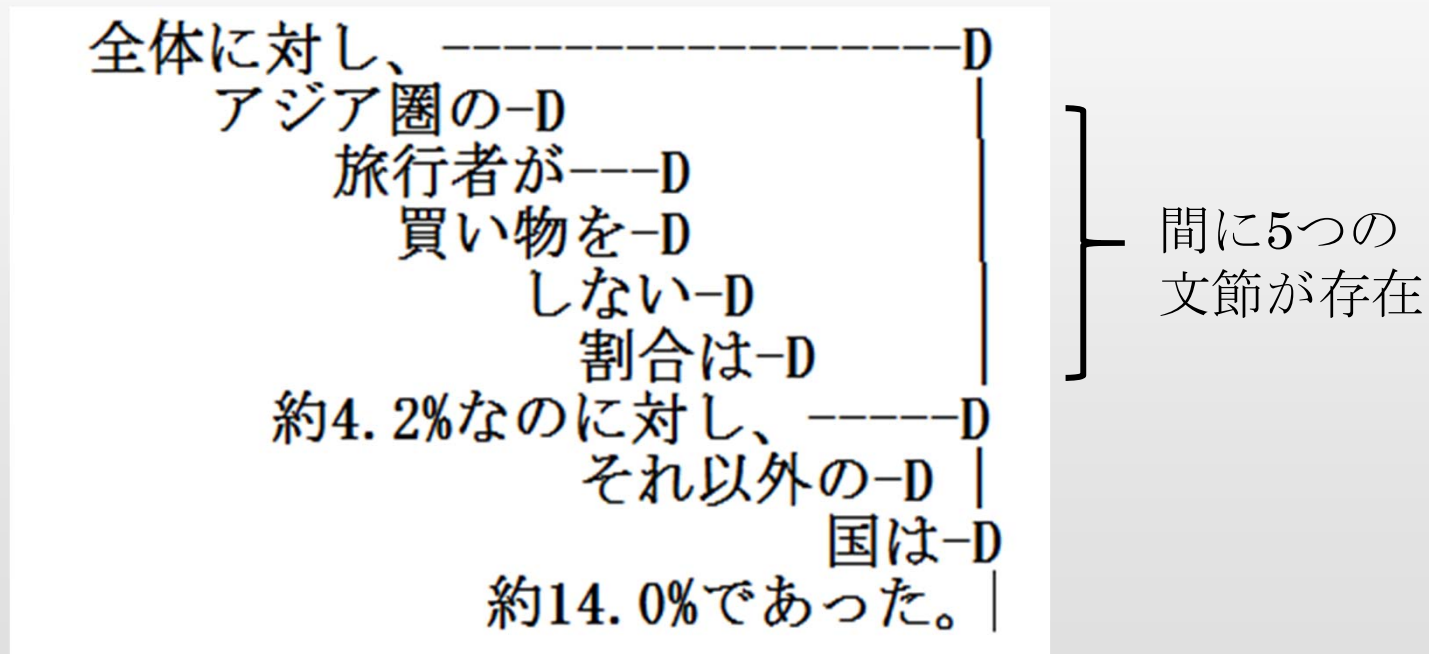
<採点アルゴリズムの検討(Skill項目群)>

No	評価項目	評価内容	計算式	使用ソフト等
22	漢字の使用率	漢字の使用率(適正範囲25~45%, 適正基準35%)	漢字の使用率 =漢字の数/全文字数	Word 2016文章 校正ツール
23	文長の妥当性	文の平均の長さ(適正範囲30~40文字, 適正基準35文字)	文の平均文字数 =全文字数/文の数	R
24	語彙の豊富さ	トークン比(値が大きいほど語彙が豊富)	トークン比 =異なり語数/述べ語数	TeMP
25	語彙の水準	主要語彙の平均水準(値が大きいほど高レベル)	主要語彙の平均水準 =(全名詞, 形容詞, 動詞の各々の語彙水準*出現頻度)の平均	日本語教育語彙表 ※1 TeMP

※1 日本語教育語彙表 ver1.0(<http://jhlee.sakura.ne.jp/JEL.html>)^[8]を利用

＜採点アルゴリズムの例(係り受け)＞

17	構文の妥当性	係り受けに関与する文節間の平均距離(適正基準1.988)	係り受けの平均距離 = 文書(レポート)の全係り受け数 ／ 文節総数
----	--------	------------------------------	--



係り受けの平均距離: $(5+1+1+1+1+1+2+1+1+0)/10=1.4$

5. 評価実験

＜実験対象とする学生レポートの特性＞

- 授業
情報リテラシー科目
- レポートのテーマ
表計算ソフトを使ってデータ解析し、見解を200文字以上で述べる
- 受講者数
43名
- 提出レポート
平均文字数 427.6 標準偏差216.0

5. 評価実験

<教員採点と自動採点との相関>

Style項目群		Skill項目群	
No.	相関係数	No.	相関係数
15	0.371※	22	-0.117
16	0.406※	23	-0.099
17	0.080	24	0.0002
18	0.060	25	0.027
19	0.384※	予測値	0.233
20	0.045	$p < 0.05$, $p < 0.001$. (表内の係数値は、スピアマンの 順位相関係数に基づく)	
21	0.270		
予測値	0.608※※		

5. 評価実験

<SVMによる成績レベルの分類>

=== Confusion Matrix ===

```
a  b  c  d  <-- classified as
3  0 13  0 | a = C
3  1  1  0 | b = D
0  0 21  0 | c = B
0  0  1  0 | d = A
```



	教員 採点	正解	不正解	
A	1	0	1	Bに分類
B	21	21	0	
C	16	3	13	Bに分類
D	5	1	4	BとCに分類
計	43	23	20	

Correctly Classified Instances 58.1395 %

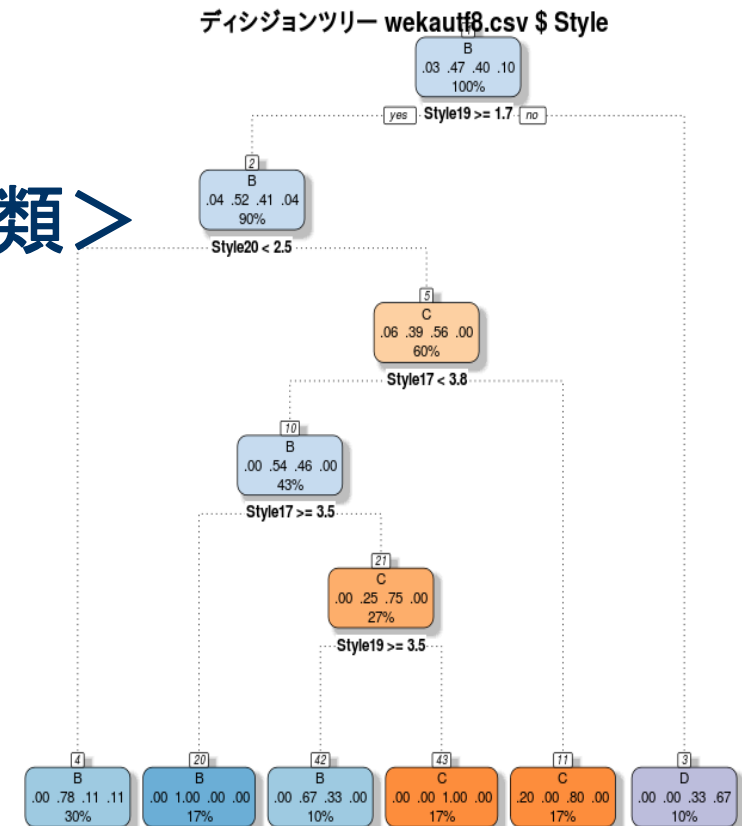
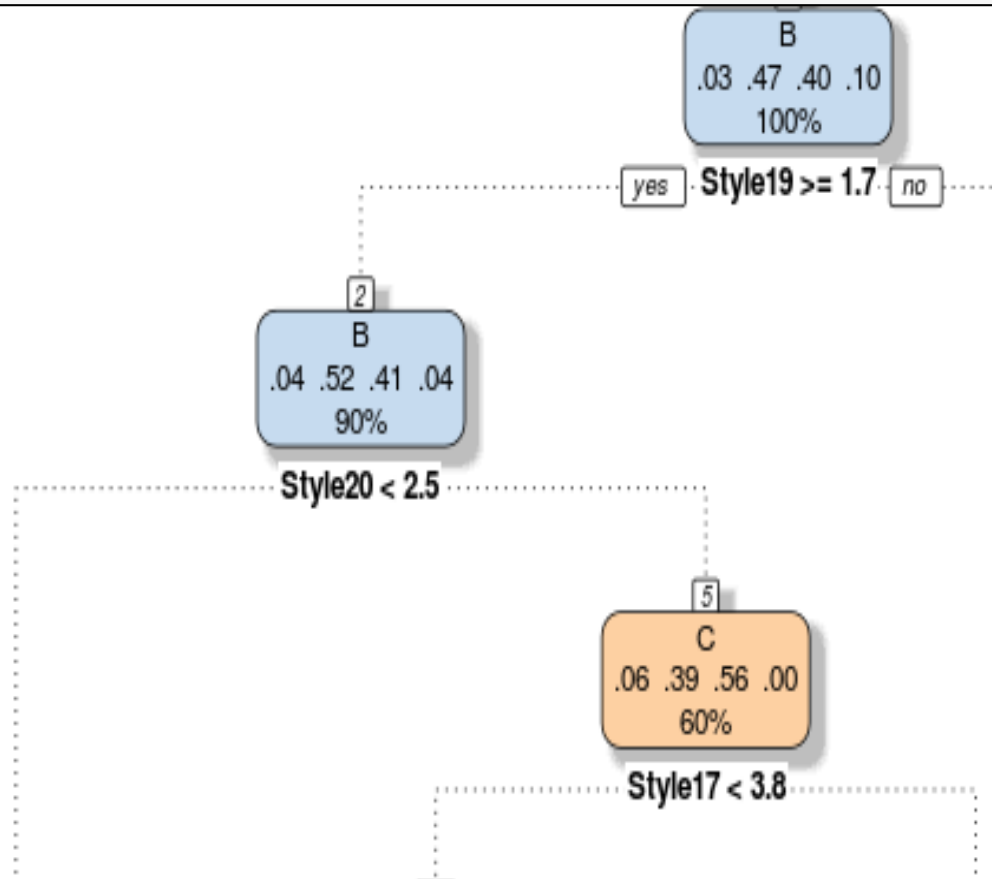
Incorrectly Classified Instances 41.8605 %

SMO -C 1.0 -L 0.001 -P 1.0E-12 -N 0 -V -1 -W 1 -K "

weka.classifiers.functions.supportVector.PolyKernel -E 1.0 -C 250007"

5. 評価実験

<決定木による成績レベルの分類>



Rattle 2016- 9月-07 19:07:19 root

6. まとめ

- レポート採点用ルーブリックの提案と自動採点可能な評価観点の提示
- Style項目群の半数は採点時の指標となり得る
- 構文の妥当性やSkill項目群は、アルゴリズムの再検討が必要

課題

- 各評価観点をより正確に説明できる自動採点項目の追加
- 自動採点が困難な評価観点への対応
- LMS(Moodle)への実装

参考文献

- [1]大野博之,稲積宏誠, ”技術文章作成能力の育成を目指した教育支援ツールの開発”,電子情報通信学会 第18回データ工学ワークショップ論文集,(2007),D9-4
- [2]石岡恒憲, 亀田雅之, ”コンピュータによる小論文の自動採点システムJessの試作”, 計算機統計学, Vol.16,(2003), pp.3-19
- [3]寺田凜太郎,久保顕大, 柴田知秀, 黒橋禎夫, 大久保智哉, ”ニューラルネットワークを用いた記述式問題の自動採点”,言語処理学会第22回年次大会発表論文集,(2016),pp.370-373
- [4]勝又大介, 藤田彬, 田村直良, ”文章構造解析に基づく小論文の論理構成における整然さの自動評価”,言語処理学会第19回年次大会発表論文集,(2013),pp.190-193
- [5]松下佳代, <新しい能力>と学習評価の枠組み
http://www.mext.go.jp/b_menu/shingi/chousa/shotou/095/shiryo/__icsFiles/afieldfile/2013/01/29/1330122_01.pdf
- [6]朝日新聞デジタル 7月23日(土)16時0分配信の記事,「東京五輪, 14知事『恩恵ある』 観光客誘導, 求める声」
<http://digital.asahi.com/articles/ASJ7Q5K0NJ7QUTIL02L.html>より算出
- [7]山本恵,梅村信夫, ”「語彙の豊富さ」からみた学生レポートの分析・評価”, 日本ムードル協会全国大会発表論文集,(2016),pp.6-8
- [8]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』プロジェクトの紹介—”, 日本語教育連絡会議論文集, (2012),pp. 164-169