国語教育的評価項目を考慮した機械学習による日本語文章の 自動評価と評価モデルの構築

藤田 彬[†]・藤田 央[†]・田村 直良^{††}

本稿では、文章に対する評点と国語教育上扱われる言語的要素についての特徴量から、個々の評価者の文章評価モデルを学習する手法について述べる。また、学習した文章評価モデルにおける素性毎の配分を明示する手法について述べる。評価モデルの学習にはSVRを用いる。SVRの教師データには、「表層」「語」「文体」「係り受け」「文章のまとまり」「モダリティ」「内容」というカテゴリに分けられる様々な素性を用意する。これらには日本の国語科教育において扱われる作文の良悪基準に関わる素性が多く含まれる。なおかつ、全ての素性が評価対象文章に設定される論題のトピックに依存しない汎用的なものである。本手法により、文章の総合的な自動評価、個々の評価者が着目する言語的要素の明示、さらに評点決定に寄与する各要素の重みの定量化が実現された。

キーワード：テキスト評価、小論文、国語教育、機械学習、教育用システム

Automated Evaluation of Japanese Compositions based on Features along Japanese Education and Construction of the Individual Evaluation Model

AKIRA FUJITA[†], HIROSHI FUJITA[†] and NAOYOSHI TAMURA^{††}

We propose a method to learn an individual model, which is to evaluate Japanese Compositions via Support Vector Regression, based on features along Japanese education and scores, marked by human in advance. We also propose a method to represent a way of evaluation. Features in training data of SVR are categorized as 7 types according to what each features refer to. The features include some features regarding criterions of Japanese compositions in education. Besides, all the features do not depend on topic of a composition's prompt. Our methods implemented to score an integrated point of a composition automatically, and also to account elements considered by individual evaluator, to quantify weights of the each elements that contributes decision of scores.

Key Words: *Text Evaluation, Short Essay, Japanese Education, Machine Learning, Educational System*

[†] 横浜国立大学大学院環境情報学府, Graduate School of Environment and Information Sciences, Yokohama National University

^{††} 横浜国立大学大学院環境情報研究院, Graduate School of Environment and Information Sciences, Yokohama National University

1 はじめに

近年、作文技術の習熟度を評定する目的で文章を自動的に評価する技術に対して、需要が高まっている。大学入試や就職試験等の大規模な学力試験において課される小論文試験の採点や、e-learning等の電子的な学習システムにおいて学習者の作文技術についての能力を測るために出題される記述式テストの採点が、例として挙げられる。このような、多数の文章を同一の基準で迅速に評価する必要があるタスクにおいて、対象となる全ての文章を手で評価することは、多くの場合困難を伴う。

第一に、評価に要する時間と労力が問題となる。記述式回答の評価は、選択式回答の評価に比べて、評価者が捉えるべき情報と考慮すべき基準が多く、それらの情報や基準自体も複雑である。

第二に、評価基準の安定性が問題となる。文章の良悪を決定する基準は、評価者個々において完全に固定的なものではない。評価する順序による系列的効果や、ある要素についての評価が他要素の特徴に歪められるハロー効果 (Nisbett and Wilson 1977) の影響も考えられる。また、このような状況において他者による評価基準に自基準を合わせる場合、少なくとも他者との基準の差異についての定量的な情報が無い限り、基準の統合は困難といえる。

これらの問題の存在は、「個々の評価者が着目する言語的要素」や「評点決定に寄与する各要素の配分（重み）」に相違が生じる要因となり得る。結果的に、それらの相違が評価者間での評点の差異として表れることも考えられる。

これらに対し、文章評価の自動化は、評価の公平性を損なう要因となる問題の解消に役立つと考えられる。また、評価者が着目する言語的要素やその配分の定量的な提示を行うことで、正確かつ円滑な評価者間の基準統合が可能になると考えられる。

本稿では、単独の評価者により対象文章に与えられる総合的な評点と、国語教育上扱われる言語的要素についての多種の特徴量から、任意の試験設定における個人の評価者の文章評価モデルを推定する手法について述べる。また、個人の評価者の評価モデルにおいて評点決定に寄与する要素毎の配分（重み）について、他の評価者の評価モデルとの間で定量的に比較可能な形で提示する手法について述べる。ただし、複数の評価者の評価モデルによる評価から最終的な評価判断を導き出すことについては扱わない。

提案手法は、文章を採点する行為を順序付き多クラス分類として捉え、Support Vector Regression (SVR) (Smola and Sch 1998) を用いた回帰手法により、評価者が付けうる評点を予測する。SVRの教師データには、表層や使用語彙、構文、文章構造などの特徴に関する様々な素性を用意する。これらの素性には、日本の国語科教育において扱われる作文の良悪基準に関わる素性が多く含まれる。なおかつ、全ての素性は、評価対象文章で議論されるトピック固有のものは含まない汎用的なものである。

本手法は、国語教育¹上扱われる言語的要素を SVR の素性に用いて文章評価をモデル化し、SVR の回帰係数の差として評価者間での評価基準の個人差を明示できるという点に、新規性を持つ。国語教育上扱われる要素に基づいて文章評価モデルを説明することができるため、教育指導を行う立場にある評価者が、普段の指導で参照する要素を介して容易に文章評価モデルを認識、比較することができる。

作文技術についてのあらゆる能力評価に対応可能であるよう、素性を網羅的に設定するが、「文章を意味面で適切に記述する能力」の評価に関しては扱わない。ここでいう意味面での適切さとは、文章中の文が示す個々の内容の正しさを指す。例えば、「月は西から昇る」のような文が示す内容が正しいか正しくないかについての判断は、本研究では扱わない。

2 関連研究

自動的に文章の評価を行うためのモデルを得る先行研究には様々なアプローチが存在する。一つは、評価者による文章のスコアをラベル、文章上の素性を事例として、教師付き学習により単一のスコア推定モデルを求めるものである (Burstein, Kukich, Wolff, Lu, Chadorow, Braden-Harder, and Harris 1998; Burstein and Wolska 2003; Elliot 2003; Ellis 1966, 1994; Landauer, Laham, and Foltz 2003b, 2003a; Foltz, Laham, and Landauer 1999; Attali and Powers 2008; Attali and Burstein 2006)。もう一方は、模範と考えられる文章上の素性値を基準として、その基準との距離を用いてスコア推定モデルを求めるものである (Ishioka and Kameda 2006; 石岡, 亀田 2003; 石岡 2008b)。

e-rater (Attali and Burstein 2006) は、12 の固定的な素性²を説明変数、評価者によるスコアを従属変数として重回帰分析を行い、得られた回帰式をスコア推定モデルとする。しかし、この手法では、説明変数として用いられる特徴量は何の変量であるかが抽象的であり、評価者の評価基準の違いを十分に表現できないと考えられる。例えば、e-rater では「総ワード数に対する語の使用法についてのエラーの割合」や「総ワード数に対する文法エラーの割合」といった特徴量が説明変数として扱われるが、「語の使用法」や「文法エラー」が具体的に示す言語現象が不明瞭である。したがって、評価基準の個人差をモデル式の回帰係数の差として示すことはできても、その差が何を意味するかについての具体性が乏しく、評価者にとって明確な差として捉えづらいつと考えられる。また、モデルが重回帰分析であることから多重共線性が問題となり、扱う特徴量を詳細化する上では、各特徴量の独立性が厳密に保たれる必要がある。この点で、提案手法は、多重共線性の影響が少ない回帰的手法 (SVR) を用いており、素性間の関連性

¹ 本稿では便宜上、小学校、中学校、高等学校における作文教育を国語教育と呼ぶこととする。

² 変数選択を行うことがなく、常に 12 の素性を説明変数とする。

を殆ど考慮することなく、多種の詳細な言語現象についての素性を用いて、評価基準をモデル化することができる。

Jess (Ishioka and Kameda 2006) は、あらかじめ三種の観点（修辞，論理構成，内容）に沿って模範となる文章（新聞の社説やコラム）における種々の素性値の分布を獲得し、理想的な分布とする。評価対象文章の各素性値が模範文章における素性値分布の四分位数範囲の 1.5 倍を超える場合、外れ値とみなしてそれぞれについて評点を減ずる。しかしこの手法では、模範文章の選択の妥当性について、評価が行われる背景（試験の目的等）毎に検証が必要である。その検証自体も、実用的に難しいと考えられる。また、評価基準をある基準に固定することが前提とされているため、評価者の基準の個人差を明示する提案手法とは目的が異なる手法であるといえる。

なお、これらの関連研究に関しては石岡によるサーベイ (石岡 2008a) が詳しい。

提案手法は、多種の詳細な言語現象についての変量を教師付き学習の素性として用いることで、「詳細な言語的要素に視座を置いた、評価基準の個人差の明示」を実現するスコア推定手法である。

3 評価基準の共通性に関する調査

図 1、表 1、表 2 に同一の文章群に対する複数人の手による評価結果の例を示す。これは宇佐美の小論文評価に関する研究 (宇佐美 2011) の中で示された、高校生による 584 編の小論文データとそれを 4 人の国語教育専門家が評価した結果のデータから、評点当たりの事例数とその分

図 1 各評価者の評点当たりの事例数 (x 軸は評点, y 軸は小論文の事例数)

表 1 評価者間での評点の κ 係数

	A-B 間	A-C 間	A-D 間	B-C 間	B-D 間	C-D 間
κ 係数	0.110	0.035	0.093	0.034	0.115	0.081

表 2 各評価者の評点分布の平均と標準偏差

評価者	A	B	C	D
平均	5.420	5.438	6.639	5.856
標準偏差	1.081	1.037	2.092	1.436

布や一致具合をまとめたものである。評価は特定の観点に沿ったものではなく総合的なもので、10段階の絶対評価により施される。評点は10点が最高点、1点が最低点である。

Krippendorff の考察 (Krippendorff 1980) によれば、あるデータ間の κ 係数が 0.7 未満の場合、両データの関連を示すことは困難であることが多いとされる。表 1 において評点の κ 係数は全評価者間で 0.7 を下回っている。また、表 2 と図 1 より、評点の分布にも異なりが認められる。このことから、本例においては評価者の評価基準が共通のものであるとはいえない。以上の結果から本研究では、全評価者の評点をまとめた点数ではなく、個々の評価者の評点を再現する評価モデルを推定する方針をとり、前述した 10 段階で評価する小論文試験の設定に沿って評価行為を模倣する。

4 評価モデルの学習

4.1 自動評価の定式化と学習の手法

文章の自動評価は、入力となる文章についての評点が出力となる。この評点は、順序関係を持った個別のクラスとみなすことができる。そこで本研究では、自動評価を、全順序関係にある多クラスに分類されたスコアを回帰的に推定する問題として定式化し、Support Vector Regression (SVR) を用いて評点の予測を行う。

SVR を用いて全順序関係にある多クラスに分類されたスコアを回帰的に推定する先行研究には、岡野原らの研究 (岡野原, 辻井 2007) がある。岡野原らは、評判分類タスクを順序付き多クラス分類問題として定式化した上で、レビューが評価対象に与える評価の度合を二極指標（実数値）で表す手法を提案している。また、SVR の回帰係数の絶対値がモデル上での重要性に相関することを用いて、レビュー記事の評価決定に重要と思われる表現を導出する手法を提案している。この中で岡野原らは、（順序関係を考慮しない）多クラス分類問題を解く分類器である pairwise Support Vector Machine (pSVM) (Kresel 1999) と SVR の間で、順序付き多クラス分類問題に対する適合性を比較している。その結果、SVR がより高い精度で分類を行うことが

示されている。pSVM は、予測クラスを間違えた際のペナルティに全クラス間で差が無いため、SVR に比べて分類モデルに大きな誤差を含む可能性が高いと思われる。

4.2 システムで採用した素性

評価者は、様々な言語的要素について文章を捉えて評価を行った上で、その判断結果を評点として数値化する。本研究ではこの過程を国語教育で参照される言語的要素を素性にとってモデル化する。

学校教育上考えられる様々な試験を想定して、それらで参照される言語的要素をマージし、素性として扱う。評価対象文章から自動的にそれぞれの素性値を算出する。得られる数値は、SVR の素性として評価モデルの訓練に用いられる。以下、提案手法において素性として用いる言語的要素を、説明の分かり易さのためカテゴリ毎に列挙した上で、一部の素性についてその詳細を述べる。本研究において独自に設定した素性については、素性名の末尾に (*) を付す。

4.2.1 カテゴリ「表層」の素性群

文字数や文数、字種などの表層的特徴に関する素性を表 3³に列挙し、特に FV2 について説明する。この群の素性は、主に、文章の形式面の妥当性についての評価に役立つ。国語教育上扱われる事項に関連する素性として、FV2 は「目的や意図に応じて簡単に書いたり詳しく書いたりする」(文部科学省 2008) という事項に、FV10 は「習得した漢字を使用して記述する」という事項にそれぞれ関連する。

FV2：文字数制限には、(I)「～字以上」、(II)「～字以内」、(III)「～字程度」の 3 種類が存在する。制限（指定）文字数を r 、評価対象文章の文字数を n としたとき、下記のように達成度 d を算出することにする。ただし、(I)(II) については制限が守られなかった場合、達成度

表 3 カテゴリ「表層」の素性群

ID	素性名	ID	素性名
FV1	文字数	FV9	カタカナ使用率
FV2	文字数制限の達成度*	FV10	漢字使用率
FV3	文数	FV11	記号使用率*
FV4	文の文字数の平均	FV12	英字使用率*
FV5	文の文字数の分散	FV13	算用数字使用率*
FV6	文の最小文字数	FV14	文当たりの読点数の平均*
FV7	文の最大文字数	FV15	形式段落当たりの文数の平均*
FV8	ひらがな使用率		

³ FV8-13 は、当該文字種の文字数を、総文字数で割った値とする。

を 0 とする.

$$(I) : d = 1 \quad (II) : d = n/r$$

$$(III) : d = \begin{cases} \frac{n}{r}, & n \leq r \\ \frac{2r-n}{r}, & n > r \end{cases} \quad (\text{バートレット窓関数の変形})$$

FW7: 文は長くなるほど, 内部の係り受け関係に曖昧さを生じやすいとされており (森岡 1963), これを用いる.

4.2.2 カテゴリ「語」の素性群

単語⁴ (特に自立語) の用法や品詞, 記法に関する素性を表 4⁵ に列挙する. 国語教育上扱われる事項に関連する素性として, FW3, FW4, FW5 は「表現の効果などについて確かめたり工夫したりすること」(文部科学省 2008) という事項に, FW10 は「文章内で数字の表記法を統一する」という事項にそれぞれ関連する.

FW2: 自立語の異なり数 (タイプ数) を延べ数 (トークン数) で割った値とする. ただし用言については活用形を計数の考慮に入れず, 異なり数, 延べ数ともにその用言の原形を数える.

FW4: オノマトペは副詞に分類される. しかし, 論説文におけるオノマトペの使用は特に着目されることが多いと考えられるため, 副詞とは別途に扱う.

4.2.3 カテゴリ「文体」の素性群

文末の形式や文内で用いられる文体等に関する素性を表 5 に列挙する. 国語教育上扱われる事項に関連する素性として, FF1, FF9 は「表現の効果などについて確かめたり工夫したりすること」(文部科学省 2008) という事項に, FF7 は「文章の敬体と常体の違いに注意しながら書くこと」(文部科学省 2008) という事項に, FF10 は「話し言葉と書き言葉との違いに気付くこ

表 4 カテゴリ「語」の素性

ID	素性名	ID	素性名
FW1	自立語の最大長	FW6	複合名詞の使用率*
FW2	自立語の多様性	FW7	カタカナ語の使用率*
FW3	副詞使用率*	FW8	自立語の文字数の平均
FW4	オノマトペ使用率*	FW9	自立語内のひらがな使用率*
FW5	感動詞, 間投詞の使用率*	FW10	数量表現における漢・算用数字の混用有無*

⁴ 本研究では, 形態素解析による出力単位を単語として扱う.

⁵ これらの素性の素性値は, 当該語の延べ数を全自立語の数で割った値とする. このとき複合名詞は一つの自立語として数える.

と」(文部科学省 2008) という事項にそれぞれ関連する.

FF4: 述語が“名詞句+断定の助動詞(「だ」「である」「です」等)”で構成される文の出現数を, 文の総数で割った値とする.

FF7: 文中の助動詞に着目し, 文体が一貫している場合は 0, 混用が認められる場合は 1 を素性値とする.

FF8: 式 $-\log(n/N)$ により算出する. ただし, n は文の最終文節の表記異なり数, N は文の総数とする.

4.2.4 カテゴリ「係り受け」の素性群

同一表層格の多用や文節間の修飾関係の複雑さを捉える素性を表 6 に列挙する. これらの素性は, 主に係り受けの適切さに関わる素性である. 国語教育上扱われる事項に関連する素性として, FD1~4, FD16 は「一文の意味が明確になるように語と語の続き方を考える」(文部科学省 2008) という事項, FD5~8 は「文の中における主語と述語との関係に注意すること」(文部科学省 2008) という事項にそれぞれ関連する.

FD6, FD7: 格助詞「ガ」, 係助詞「ハ」を付属語に持つ文節が文中に出現する回数の最大値を素性値とする. 文中で同じ助詞を繰り返し使用することで, 文の表す内容が不明瞭になる場合がある(岩淵 1979).

表 5 カテゴリ「文体」の素性

ID	素性名	ID	素性名
FF1	体言止め文の出現率*	FF6	括弧 () による補足を含む文の出現率*
FF2	体言が存在しない文の出現率*	FF7	常体・敬体の混用有無*
FF3	用言が存在しない文の出現率*	FF8	文末の単調さ*
FF4	名詞述語文の出現率*	FF9	感嘆符が用いられる文の出現率*
FF5	鉤括弧「」による引用を含む文の出現率*	FF10	口語的表現の使用有無*

表 6 カテゴリ「係り受け」の素性群

ID	素性名	ID	素性名
FD1	格助詞「ノ」で終わる文節の最大連続出現数*	FD9	文当たりの連体修飾回数の最大値
FD2	格助詞「ト」で終わる文節の最大連続出現数*	FD10	文当たりの連体修飾回数の最小値
FD3	格助詞「ニ」で終わる文節の最大連続出現数*	FD11	文当たりの連体修飾回数の平均
FD4	格助詞「ヲ」で終わる文節の最大連続出現数*	FD12	文当たりの連体修飾回数の分散
FD5	係助詞「モ」で終わる文節の最大連続出現数*	FD13	全文節中に占める名詞文節の割合*
FD6	格助詞「ガ」で終わる文節の文内最大出現数*	FD14	修飾の詳細さ (平均)*
FD7	係助詞「ハ」で終わる文節の文内最大出現数*	FD15	修飾の詳細さ (分散)*
FD8	一文中での「ハ・モ・ガ」の最大出現数*	FD16	一文中での中止法の最大使用回数*

例：英語の早期教育がもたらす効果が測れないうちに実行に移すことが問題であることは、言うまでもない。

FD14, FD15：一文中における名詞文節の出現回数を用言の出現回数で割った値を全文について平均した値、またその分散を素性値とする。

FD16：一文中で、用言の連用形や接続助詞等によって文が途中で中止される回数の最大値を素性値とする。中止法の多用は、係り受け関係を曖昧にする原因となりうる (岩淵 1979)。

4.2.5 カテゴリ「文章のまとまり」の素性群

文章のまとまりに関する性質として、「テキスト一貫性」(田窪, 西山, 三藤, 亀山, 片桐 2004) と「テキスト結束性」(Halliday and Hasan 1976) が挙げられる。テキスト一貫性は、概念や事象の間の意味的なつながりの良さを指す。テキスト一貫性は、隣り合う二文間における一貫性を示す「局所的な一貫性」と、文章全体での話題遷移の一貫性を示す「大域的な一貫性」に区別できる。一方テキスト結束性は、意味的なつながりではなく、文法的なつながりの良さを指す。Barzilay ら (Barzilay and Lapata 2008) は、局所的な一貫性のモデルとして「entity grid モデル」⁶を提案している。横野ら (横野, 奥村 2010) は、結束性に寄与する要素⁷を entity grid モデルに組み込むことで、結束性と局所的な一貫性を同時に捉えるモデルを提案している。

横野らは、結束性を考慮する目的で既存手法に下記の手法を組み込んでいる。

- i. 語句要素の文間遷移確率の計算に文間の接続関係の考慮を加え、接続関係の種類別に遷移確率を計算する。
- ii. 意味的な類似性に基づいて語句要素のクラスタリングを行う。
- iii. 参照表現が正しく機能している割合を素性として導入する。

表 7 カテゴリ「文章のまとまり」の素性群

ID	素性名
FC1～FC36	論理的結合関係で接続する文間の構文役割遷移確率
FC37～FC72	多角的連続関係で接続する文間の構文役割遷移確率
FC73～FC108	拡充的合成関係で接続する文間の構文役割遷移確率
FC109～FC114	文章始端と冒頭文の間の構文役割遷移確率
FC115～FC120	最終文と文章終端の間の構文役割遷移確率
FC121	参照表現が正しく機能している割合
FC122	並列接続の有無*

⁶ 文を行、文章中の語句要素を列、文における語句要素の構文役割を成分とする行列を用いて、語句要素の分布パターンを表現するモデル。行列から構文役割の遷移確率と構文役割の出現確率を成分とするベクトルを導出し、局所的な一貫性の評価等に用いる。表 7 の FC1～FC120 は、構文役割についての 2-gram 生起確率に対応する。

⁷ Halliday ら (Halliday and Hasan 1976) は、参照、接続、語彙的結束性、省略を挙げている。

また, Barzilay らの entity grid において扱われる構文役割は 4 種類 (S: 主語, O: 目的語, X: その他, —: 出現せず) であるが, 横野らは構文役割の体系を日本語に特化する目的で, 2 種類の構文役割 (H: 主題, R: 述部要素) を加えている.

本研究では, 文章のまとまりについての特徴を捉える目的に, 横野らのモデルに基づいた素性を用いる (表 7). ただし, FC122 のみ筆者らが独自に設定する素性である. 接続関係の分類・同定方法, 語句要素が持つ構文役割の同定方法, 参照表現が機能している割合の導出方法は, それぞれ横野らの方法に従う. 類似性に基づいた語句要素のクラスタリングについては, EDR 電子化辞書 (日本電子化辞書研究所 2010) の日本語単語辞書内で同一の概念識別子を持つ語句要素を同じクラスタとして扱う. クラスタの持つ構文役割は, $H > S > O > R > X$ という優先順位 (cf., (Walker, Iida, and Cote 1994)) に基づいて, クラスタ毎に一つの構文役割を決定する. また, 考慮する遷移確率は文 2-gram のみとする. 国語教育上扱われる事項に関連する素性として, FC1~121 は「語と語や文と文との続き方に注意しながら, つながりのある文章を書くこと」(文部科学省 2008) という事項, FC122 は「自分の考えを明確に表現するため, 文章全体の構成の効果を考えること」(文部科学省 2008) という事項にそれぞれ関連する.

FC122: 並列的に展開して述べる接続関係が文章中に存在する場合に 1, 存在しない場合に 0 を素性値とする. 並列的な接続を明示する特定の表現 (人手で設定) の有無により, 同定する.

以下に, FC1~FC120 の素性値の算出方法を示す. 表 8 において entity grid で表される文章は, 文 1 と文 2 の間, 文 3 と文 4 の間がそれぞれ論理的結合関係, 文 2 と文 3 の間が多角的連続関係と捉えられている. 語句要素 1 は文 1 で主題, 文 2 で主語, 文 4 でその他の構文役割として, 語句要素 2 は文 1 と文 3 で目的語として捉えられている.

このとき, FC1~36 の素性値は以下のようにして求める. 2 文間での構文役割の遷移パターンは, 合計 36 通り (「HH」「HS」「HO」…「R-」「X-」「-」) ある. 論理的結合関係で接続する 2 文間では, 合計 4 カ所の遷移箇所のうち, 「HS」「-X」の遷移が 1 回ずつ, 「O-」の遷移が 2 回出現している. 従って, FC1~FC36 の素性のうち, 「HS」と「-X」の遷移確率を表す素性の値は 0.25, 「O-」の遷移確率を表す素性の値は 0.5 となる. その他の 33 個の素性の値は全て 0 となる. FC37~72 の素性値も同様にして求め, 「S-」と「-O」の遷移確率を表す素性の

表 8 結束性と局所的一貫性を同時に捉える entity grid の例

	前文との接続関係	語句要素 1	語句要素 2
文 1	—	H	O
文 2	論理的結合関係	S	—
文 3	多角的連続関係	—	O
文 4	論理的結合関係	X	—

値は 0.5, その他の 34 個の素性の値は全て 0 となる. 拡充的合成関係の接続関係は文章中に出現しないので, FC73~FC108 の素性値は全て 0 となる. FC109~FC114, FC115~FC120 はそれぞれ, 文章始端と文章終端にダミー要素 (始端を B, 終端を E) を持つとした時の遷移パターン「BH」~「B-」, 「HE」~「-E」の出現確率を FC1~FC108 と同様に算出した結果を値とする.

4.2.6 カテゴリ「モダリティ」の素性群

松吉ら (松吉, 江口, 佐尾, 村上, 乾, 松本 2010) は, 情報発信者の主観的な態度 (モダリティ) に真偽判断や価値判断などの情報を統合した「拡張モダリティ」を提案し, 体系化している. 拡張モダリティは主に「態度表明者」, 「相対時」, 「仮想」, 「態度」, 「真偽判断」, 「価値判断」の 6 項目から成る (文献 (松吉, 佐尾, 乾, 松本 2011) を参考にした). このうち態度⁸と真偽判断について文の特徴を捉え, 素性として扱う (表 9). 国語教育上扱われる事項に関連する素性として, FM1~8, FM12 は「事実と感想, 意見などとを区別すること」(文部科学省 2008) という事項にそれぞれ関連する.

松吉らは態度を 8 種類に, 真偽判断を 9 種類に分類している. この分類に基づいて, 機能表現辞書つづじ (松吉, 佐藤, 宇津呂 2007) に助動詞型機能表現として収録される機能表現と, 分類語彙表 (国立国語研究所 2004) に「精神および行為/心」として収録される用言を手がかりにしたルールベース手法で, 文の態度, 真偽判断について分類を行う. ただし, 真偽判断については, 肯否極性とアスペクトに関する区別をせずに判断の程度 (強さ) のみを扱うこととし, 断定, 推量, 判断程度不明の 3 種類を扱う. これら拡張モダリティ体系に準拠する素性 (FM1~FM11) は全て, 当該拡張モダリティカテゴリに分類される文が出現する回数を文の総数で割った値とする.

FM12: 最終文節で思考動詞, 知覚・感覚動詞が用いられる文の出現回数を文の総数で割った値を素性値とする. 思考動詞, 知覚・感覚動詞であるか否かの判断は, 分類語彙表 (国立国

表 9 カテゴリ「モダリティ」の素性群

ID	素性名	ID	素性名
FM1	態度「叙述」出現率*	FM7	態度「許可」出現率*
FM2	態度「意志」出現率*	FM8	態度「問いかけ」出現率*
FM3	態度「欲求」出現率*	FM9	真偽判断の程度が「断定」の文の出現率*
FM4	態度「働きかけ—直接」出現率*	FM10	真偽判断の程度が「推量」の文の出現率*
FM5	態度「働きかけ—間接」出現率*	FM11	真偽判断の程度が不明な文の出現率*
FM6	態度「働きかけ—勧誘」出現率*	FM12	文末思考知覚感覚動詞使用率*

⁸ 言語学における「表現類型のモダリティ」(益岡 1991) に相当する.

語研究所 2004) を典拠とする.

4.2.7 カテゴリ「内容」の素性群

文章の内容(筆者により書かれた行動, 出来事, 状態)について, その正しさを意味面で判断することは, 本研究の目的としない. その代わりに, 与えられた論題に対して適合した語彙が使用されていることを捉えるための素性を用意する. 論題に含まれる名詞と文章中の名詞が, EDR 電子化辞書(日本電子化辞書研究所 2010)において同一の概念識別子を持つ, もしくは所属概念が直接の上位または下位関係にある場合, 論題に適合する語彙として判断する. このように判断される語彙が文章中の全名詞中で占める割合を素性(FS1)とする.

4.3 評価モデルの構築とそのための素性値の正規化

SVR は, 線形カーネルを使用して学習する場合, 回帰係数 w の成分値(以下, 成分値)を参照することで, 各素性がスコア推定モデルに寄与する度合を知ることができる. これにより, 教師データにラベル(評点)をつけた評価者が, 各素性に対して「どの程度の配分で評価していたか」, また「加点要素としたか減点要素としたか」を定量化することができる. 前者は成分値の絶対値, 後者は成分値の正負に着目することでそれぞれ明らかになる. この方法により個々の評価者の評価モデルにおいて重要な変量を明らかにする.

これらの特徴を素性間で比較する目的で, 下記の2通りの方法で素性値を正規化する.

$$\textcircled{1} \quad x_{\text{regularized}} = (x_{\text{original}} - \min X) / (\max X - \min X)$$

$$\textcircled{2} \quad x_{\text{regularized}} = (x_{\text{original}} - Q_1) / (Q_3 - Q_1)$$

x_{original} は任意の文章データにおける任意の素性の素性値, X は全教師データにおける任意の素性の素性値の集合, Q_1 は集合 X 中の第1四分位値, Q_3 は集合 X 中の第3四分位値である.

①は, 全教師データ中の素性値が0から1の間に分布するように正規化する方法である. 一方②は, 全教師データの四分位数範囲に位置する素性値が0から1の間に分布するように正規化する方法である.

5 実験・考察

5.1 設定

提案手法の評価のために実験を行う. SVR の学習には SVMlight⁹を用いる. また, 素性の抽

⁹ <http://svmlight.joachim.org/>

出には形態素解析器 MeCab¹⁰, 係り受け解析器 CaboCha¹¹を用いる¹².

教師データには第3章で用いた高校生による小論文を電子化したものを用いる. これらの小論文は, 論題 α 「小学校の授業における, 英語の早期教育は必要であるか否かに対して意見を述べよ」, 論題 β 「グラフと説明文を読み, 日本人の子育ての態度に関してどのような特色が読み取れるかに関して述べよ」という2種類の論題に沿って書かれている. また, 400字以内と800字以内の2種類の字数制限が存在する. 事例は合計で584事例あり, 論題と字数制限毎の内訳は表10に示す通りである.

これらの584事例に対して4人の評価者が総合的につけた10段階の評点を, 各教師データのラベルとする. ラベルとなる評点は, 下記に示す線形変換を行うことで, 評価者系列毎に評点分布の平均が0, 分散が1になるよう正規化する.

$$score'_i = \frac{score_i - \overline{score}}{\sqrt{\frac{1}{n} \sum_{i=1}^n (score_i - \overline{score})^2}}$$

5.2 実験

実験1. 評点推定の性能

教師データを用いてSVRを構築し, 各評価者がつけた評点とSVRによる評点推定結果の間の差について検討する. また, 本手法とベースラインで評点推定性能を比較することで, 素性設計の妥当性についても検証する. SVRによる評点推定の評価指標には平均二乗誤差 (Mean Square Error)¹³を用いる. また素性値の正規化に, 4章で述べた2通り (①, ②) の方法を適用し, それぞれを教師データに用いて構築したSVRのMSEを比較する. 評点推定性能のベースラインには, 本研究において独自に設定した素性を除いた素性のみでSVRを構築した場合のMSEを設定する. なお, ベースラインにおける正規化手法には②の手法を用いた. SVR構築に以降全て, 線形カーネル, コストパラメータ $C = 10$ を用いる. それぞれ表11に全教師データを用いた場合 (ALL), 論題 α のみ (only α), 論題 β (only β) のみの教師データを用い

表10 教師データ事例数の内訳

		字数制限	
		400字以内	800字以内
論題	α	153	147
	β	140	144

¹⁰ <http://mecab.sourceforge.net/>

¹¹ <http://sourceforge.net/projects/cabocha/>

¹² 標準のIPA辞書のほか, 我々が独自に拡張したユーザ辞書 (藤田, 藤田, 田村 2011) を用いている.

¹³ $\frac{1}{n} \sum_{i=1}^n (y_i - f_i)^2$ y_i を評価者による評点, f_i をSVRにより推定される評点, n を事例数とする.

表 11 提案手法の評点推定性能 (MSE)

評価者	素性セット	正規化手法	教師データセット		
			ALL	only α	only β
A	提案手法	①	0.635	0.477	0.713
		②	0.443	0.451	0.481
	baseline		0.522	0.550	0.528
B	提案手法	①	0.618	0.690	0.621
		②	0.517	0.591	0.507
	baseline		0.629	0.646	0.612
C	提案手法	①	0.642	0.686	0.650
		②	0.528	0.586	0.521
	baseline		0.578	0.655	0.664
D	提案手法	①	0.683	0.681	0.685
		②	0.514	0.584	0.546
	baseline		0.677	0.657	0.616

baseline は独自に設定した素性を除いた素性のみで学習した場合の MSE

た場合の 5 分割交差検定の結果を示す。なお以降の実験では素性値の正規化に②の手法を用いる。

実験 2. 評価モデルの構築

教師データを全て訓練に用いて SVR を構築し、回帰係数の各成分値について検討する。表 12 に、成分値の絶対値が大きいもの上位 5 素性を正負別に示す。この実験により、提案手法が国語教育で参照される言語的要素を用いて評価モデルを表現でき、なおかつ個人間での重みの配分の違いを明示できることを確認する。ただし、FC1 から FC120 までの素性については、素性が示す意味自体を ID として表記する。これらの素性 ID はそれぞれ、1 文字目が前文、2 文字目が後文の構文役割を、3 文字目が 2 文間の接続関係を表す。構文役割「B」は文章始端、「E」は文章終端のダミー要素に、「N」は前述の「—: 出現せず」に対応する。また接続関係は、1 が論理的結合関係、2 が多角的連続関係、3 が拡充的合成関係に対応する。例えば、FC1~FC36 の素性群（論理的結合関係で接続する文間の構文役割遷移確率）のうち、前文での構文役割が「H」、後文での構文役割が「N」である語句要素の 2 文間での遷移確率についての素性は「HN1」と表記する。

実験 3. 教師データ数と評点推定性能の関係

教師データの増加に伴う評点推定性能の推移について検討する。図 2 に 584 事例の教師データの部分集合（無作為抽出）を用いた 5 分割交差検定の結果を示す。

表 12 評価者別の回帰係数 w の成分値

		順位	評価者 A		評価者 B		評価者 C		評価者 D	
ALL	+	1	FV2*	0.091	FV2*	0.117	FV2*	0.247	FV2*	0.225
		2	NE3	0.026	ON3	0.071	FW6*	0.090	FW6*	0.137
		3	FV5	0.023	FW1	0.068	FV1	0.084	NO1	0.112
		4	FV10	0.022	RN2	0.067	OX3	0.079	FM2*	0.107
		5	FD6*	0.022	FD8*	0.050	BS3	0.077	FV7	0.082
	-	5	FW9*	-0.016	RN3	-0.050	BR3	-0.073	FF1*	-0.054
		4	FW7*	-0.016	FF3*	-0.050	FD8*	-0.075	FM12*	-0.055
		3	OE3	-0.020	FM6*	-0.052	FW9*	-0.082	FF8*	-0.071
		2	FV8	-0.022	FF1*	-0.063	FD1*	-0.102	FW4*	-0.076
		1	FM12*	-0.024	OE3	-0.066	FV15*	-0.106	FW9*	-0.099
only α	+	1	FV2*	0.057	FM5*	0.058	FV2*	0.176	FV2*	0.151
		2	FV13*	0.027	FV2*	0.052	XN3	0.068	NO1	0.123
		3	FV8	0.017	FM9*	0.033	OX3	0.056	FM2*	0.099
		4	NN3	0.017	FF1*	0.032	FW6*	0.052	FW6*	0.096
		5	FW9*	0.015	FV13*	0.022	HE3	0.051	BX3	0.075
	-	5	FV10	-0.014	NS2	-0.023	RN1	-0.062	HN3	-0.066
		4	FM10*	-0.015	FM10*	-0.024	FW9*	-0.071	FF1*	-0.067
		3	FD10	-0.016	FF8*	-0.024	HN3	-0.074	BR3	-0.068
		2	FW2	-0.018	FM3*	-0.025	FV15*	-0.081	FF8*	-0.074
		1	FV6	-0.019	FM12*	-0.032	HN1	-0.097	NH1	-0.109
only β	+	1	FV2*	0.105	FV2*	0.057	FV2*	0.217	FV2*	0.247
		2	FW1	0.076	FW1	0.053	FM9*	0.103	FM9*	0.059
		3	FV10	0.051	ON3	0.032	NS2	0.079	FD8*	0.057
		4	FD13*	0.047	RN2	0.028	FV1	0.077	FD5*	0.055
		5	FD3*	0.031	FV10	0.025	FF10*	0.067	ON3	0.054
	-	5	BR3	-0.028	OE3	-0.025	FD14*	-0.071	XX3	-0.067
		4	FW9*	-0.032	HE3	-0.031	FW2	-0.076	NH2	-0.069
		3	FM12*	-0.036	BX3	-0.031	SN1	-0.085	FW4*	-0.086
		2	FW4*	-0.043	FF3*	-0.033	FV15*	-0.089	FV5	-0.096
		1	FV8	-0.049	FF1*	-0.058	FD15*	-0.119	FD12	-0.105

素性 ID と数値は対応する素性と回帰係数における成分値, { +, - } は成分値の正負

(*: 本研究において独自に設定した素性)

5.3 考察

実験 1 の考察

表 11 より, 素性値の正規化手法には, ①の手法に比べて②の手法を適用した場合の方が良い結果が得られる傾向がある. 全ての結果において, 設定したベースラインよりも低い MSE が得られている. 従来の文章自動評価で用いられて来た素性のみを用いる場合に比べ, 本研

究において独自に設定した素性を追加する場合の方が、より高い精度で個人の評価モデルを推定できたことがわかる。このことから、本研究で新たに提案する素性は、評価基準のモデル化に有用な素性であるといえる。

教師データのラベル系列（評価者）別に MSE を比較すると、差がみられる。これは、評価者が評価の依拠とした言語的要素の中に、本実験で設定した素性群に含まれないものが存在したことに因るものと考えられる。

論題別の評点推定性能に関して、論題 α に沿って書かれた小論文の評点性能は、相対的に低い傾向にある。論題 α は、ある事柄に関して是非を問う類の論題であり、回答は賛成もしくは反対いずれかの立場をとる 2 種類に分類される。そのため、回答の方向性が定められていない論題 β に比べて表現手法や構成に多様性がなく、教師データが偏りやすいと考えられる。

実験 2 の考察

表 12 より、「文字数制限の達成度 (FV2)」「複合名詞の使用率 (FW6)」「真偽判断の程度が『断定』の文の出現率 (FM9)」は加点要素、「文末思考知覚感覚動詞使用率 (FM12)」「オノマトペ使用率 (FW4)」「文末の単調さ (FF8)」は減点要素として、それぞれ一部の評価者の間で共通した傾向があることがわかる。「HN-」「NH-」の素性は、「ある文で主題として出現する語が隣接する前後の文で出現しない」という文章のつながりの悪さを示す素性である。これらの素性についても、減点要素とされる傾向にあることがわかる。

提案手法で独自に設定した素性のうち、カテゴリ「表層」「語」「文体」「モダリティ」に該当する素性に有効な（一部の評価者の間で共通して有効な傾向があるとは限らない）素性が多い傾向がみられた。特に、「文字数制限の達成度 (FV2)」「オノマトペ使用率 (FW4)」「文末の単調さ (FF8)」などの、国語教育に関連する素性に有効なものが多い。本手法で用いる素性には、既存手法 (Attali and Burstein 2006) (Ishioka and Kameda 2006) でも共通して用いられるものも含まれているが、そのうち回帰係数の大きいものは「文字数 (FV1)」「自立語の最大長 (FW1)」などの一部分に限られた。本手法のように総合面での文章評価を国語教育上扱われる言語的要素を用いてモデル化する手法と、そうでない手法では、有効な素性が異なると考えられる。

一方、ほとんど影響を持たない素性も存在する。カテゴリ「文章のまとまり」に関する素性群には「OE3」「HN3」など、成分値の絶対値が大きくかつ評価者間で極性が一致する傾向にある素性もあるが、大半の素性は成分値が 0 に近い。カテゴリ「内容」に関する素性も、全評価者において成分値の絶対値が小さく、あまり評点に影響を与えていないことがわかる。なお、ここでは絶対値 0.01 以上の重みをもつ素性を絶対値の大きい素性として扱う。

論題別の評価モデルに関して、論題 α と論題 β では重要な素性が異なる傾向にあることがわかる。論題 α では、評価者 A, B の評価モデルの推定結果に「真偽判断の程度が『推量』の文の出現率 (FM10)」が大きな減点要素として含まれている。一方、論題 β では評価者 A, B

ともに加点要素にも減点要素にも含まれていない。論題 β はデータを参照した上で推量される事柄を述べる性質の論題であるため、推量表現は一般的に用いられる。他方、論題 α は賛成か反対かを問う論題であるため、断定的な態度をとった文章に比べて、婉曲的な態度の文章は論旨が不明確になりやすいと考えられる。

小論文試験で出題される論題間の異なりは、「内容の異なり」と「形式の異なり」の二種があると考えられる。本研究では内容面について「論題に示された語と関連する語を使っているかどうか」のみを素性として導入している。したがって、出題内容の異なりが評価モデルに影響を与えることはないと考えられる。しかしながら、実験結果に示されたように出題形式の異なりが評価モデルに影響を与えることがある。

評価モデルをより汎用的なものとするために、論題の出題形式を一定の粒度で分類した上で、同類の出題形式の論題に対して同じ評価モデルを適用することが妥当であるかを今後確認する必要がある。

実験 3 の考察

図 2 より、全体的にデータの増加に対する MSE の変化は収束傾向にあることがわかる。

6 おわりに

本稿では、国語教育的評価項目を「表層」「語」「文体」「係り受け」「文章のまとまり」「モダリティ」「内容」というカテゴリに分けられる素性群で表し、機械学習を用いて個人の評価者の評価モデルを学習する手法について述べた。また評価モデルにおいて重要な変量を明らかにす

図 2 教師データの増加に伴う MSE の推移¹⁴

¹⁴ 教師データを最小数から始めて 25 編ずつ追加し、それぞれについての 5 分割交差検定の結果を表示した。

る手法について述べた。この手法により、個々の評価者について「総合的評価（採点）処理のシミュレーション」、「国語教育上扱われる言語的要素を用いての評価モデルの説明」、「他者との評価モデルの違いの定量的明示」が可能になった。予め手本となる評価者を設定し、その評価者による採点結果を本手法で分析すれば、他の評価者による評価がどの要素についてどの程度手本から離れているかを照らし合わせることができる。このように、本手法により得られる情報は、多人数での評価作業において公平性を損なう要因となる「評価基準の個人差」についての問題解消に役立つと考えられる。

しかし、本稿で言及した素性のうち「文章のまとまり」「内容」といったカテゴリの素性群は、ほかの素性群に比べ影響が小さいことが分かった。今後、特にカテゴリ「文章のまとまり」を評価の観点に加えるには、修辞構造等の文章構造を大域的にとらえたものを素性として加える必要があると考えられる。

本稿では総合面での評点を順序関係付きラベルとして個人の評価者の評価モデルの学習を試みているが、今後個々の観点に関して素性の適性の検討を行う上では観点別の評点をラベルとして学習を行う必要がある。我々が研究対象とした高校生の小論文データには、「総合」評価のほかに「語句」「表現」「語彙」「課題」「簡潔」「明確」「構成」「一貫」「説得」「独創」という観点からの評価も存在する。今後、「文章のまとまり」「内容」といったカテゴリの素性群の影響が大きく反映される観点（「課題」、「一貫」等）からの評価モデルの学習を試みた上で、各素性についてさらに検討を行う必要があると考えられる。

今後、個人の評価モデルについての情報が評価基準の個人差の解消にもたらす効果について、臨床的な実験（ある評価者が、他者の評価モデルとの差異についての情報に基づいて自身の基準を再考した後に、再評価した結果を考察する実験）による検証を行いたい。

また本稿では、複数の評価者の評価モデルによる評価から最終的な評価判断を導き出すことについて扱っていないが、受験者の最終的な評価を決定することは、自動評価手法一般に期待される機能であるといえる。今後これらの機能について検討を行う必要がある。

謝 辞

本研究については、公益財団法人博報児童教育振興会の児童教育実践事業についての研究助成事業、「学習指導要領に立脚した児童作文自動点検システムの実現」（助成番号：11-B-081、研究代表：藤田彬）の援助を受けた。

また、高校生の小論文答案をお貸しいただき、研究利用を認めて下さった揚華氏、宇佐美慧氏、東京工業大学大学院社会理工学研究科の前川眞一教授に感謝の意を表す。

参考文献

- Attali, Y. and Burstein, J. (2006). “Automated Essay Scoring With E-rater v.2.0” *Article in Journal of Technology, Learning, and Assessment*, **4** (3).
- Attali, Y. and Powers, D. (2008). “A Developmental Writing Scale” Tech. rep. ETS Research Report No. RR-08-19, Educational Testing Service.
- Barzilay, R. and Lapata, M. (2008). “Modeling Local Coherence: An Entity-based Approach.” *Computational Linguistics*, **34** (1), pp. 1–34.
- Burstein, J., Kukich, K., Wolff, S., Lu, C., Chadorow, M., Braden-Harder, L. C., and Harris, M. D. (1998). “Automated Scoring Using A Hybrid Feature Identification Technique.” In *ACL '98 Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics*, Vol. 1.
- Burstein, J. and Wolska, M. (2003). “Toward Evaluation of Writing Style: Finding Overly Repetitive Word Use in Student Essays.” In *EACL '03 Proceedings of the 10th conference on European chapter of the Association for Computational Linguistics*, Vol. 1.
- Elliot, S. (2003). “How Does IntelliMetric Score Essay Responses?” Tech. rep. RB-929, Vantage Learning, Newtown, PA.
- Ellis, B. P. (1966). “The Imminence of Grading Essays by Computer.” *The Phi Delta Kappan*, **47** (5).
- Ellis, B. P. (1994). “New Computer Grading of Student Prose, Using Modern Concepts and Software.” *Journal of Experimental Education*.
- Foltz, P. W., Laham, D., and Landauer, T. K. (1999). “Automated Essay Scoring: Applications to Educational Technology.” *Proc. EdMedia '99*.
- 藤田央, 藤田彬, 田村直良 (2011). Wikipedia から抽出した語彙関係リソースの小論文自動評価タスクへの適用. 第 10 回情報科学技術フォーラム (FIT2011), E-053, 2 巻, pp. 341–344.
- Halliday, M. A. K. and Hasan, R. (1976). *Cohesion in English*. Longman, London.
- Ishioka, T. and Kameda, M. (2006). “Automated Japanese Essay Scoring System based on Articles Written by Experts.” In *Proc. 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics*, pp. 233–240.
- 石岡恒憲 (2008a). 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, **23** (1), pp. 17–24.
- 石岡恒憲 (2008b). 日本語小論文の論理構成の把握とその図式表現. 人工知能学会論文誌, 特集論文「情報編纂：要素技術と可能性」, **23** (5), pp. 303–309.
- 石岡恒憲, 亀田雅之 (2003). 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, **16** (1), pp. 3–18.
- 岩淵悦太郎 (1979). 第三版 悪文. 日本評論社.
- 国立国語研究所 (2004). 分類語彙表増補改訂版データベース.
- Kresel, U. (1999). *Pairwise Classification and Support Vector Machines Methods*. MIT Press.
- Krippendorff, K. (1980). *Content Analysis: An introduction to its methodology*. Sage Publications.
- Landauer, T. K., Laham, D., and Foltz, P. W. (2003a). “Automated Scoring and Annotation of Essays with the Intelligent Essay Assessor.” In Sherims, M. and Burstein, J. (Eds.), *Automated Essay Scoring: A Crossdisciplinary Perspective*, pp. 87–112 Hillsdale, NJ. Lawrence Erlbaum Associates.
- Landauer, T. K., Laham, D., and Foltz, P. W. (2003b). “The Intelligent Essay Assessor, the Debate on Automated Essay Grading.” *IEEE Intelligent Systems*, **15** (5), pp. 27–31.
- 益岡隆志 (1991). モダリティの文法. くろしお出版.
- 松吉俊, 江口萌, 佐尾ちとせ, 村上浩司, 乾健太郎, 松本裕治 (2010). テキスト情報分析のための判断情報アノテーション. 電子情報通信学会論文誌 D, **J93-D** (6).
- 松吉俊, 佐尾ちとせ, 乾健太郎, 松本裕治 (2011). 拡張モダリティ付与コーパス作成の基準 version 0.8 β , <http://cl.naist.jp/nltools/modality/manual.pdf>.
- 松吉俊, 佐藤理史, 宇津呂武仁 (2007). 日本語機能表現辞書の編纂. 自然言語処理, **14** (5), pp. 123–146.
- 文部科学省 (著作権所有) (2008). 小学校学習指導要領解説 国語編. 東洋館出版社.
- 森岡健二 (1963). 文章構成法 文章の診断と治療. 至文堂.
- 日本電子化辞書研究所 (2010). EDR 電子化辞書 V4.0. 独立行政法人情報通信研究機構.
- Nisbett, R. and Wilson, T. (1977). “The halo effect: Evidence for Unconscious Alternation of Judgements.” *Journal of Personality and Social Psychology*, **35** (4), pp. 250–256.
- 岡野原大輔, 辻井潤一 (2007). レビューに対する評価指標の自動付与. 自然言語処理, **14** (3), pp. 273–295.
- Smola, A. and Sch, B. (1998). “A tutorial on Support Vector Regression.” Tech. rep. NeuroCOLT2 Technical Report NC2-TR-1998-030.
- 田窪行則, 西山佑司, 三藤博, 亀山恵, 片桐恭弘 (2004). 談話と文脈, 言語の科学 7. 岩波書店.
- 宇佐美慧 (2011). 小論文評価データの統計解析—制限字数を考慮した測定論的課題の検討—. 行動計量学, **38** (1), pp. 33–50.
- Walker, M., Iida, M., and Cote, S. (1994). “Japanese Discourse and the Process of Centering.” *Computational Linguistics*, **17** (1), pp. 21–48.
- 横野光, 奥村学 (2010). テキスト結束性を考慮した entity grid に基づく局所的一貫性モデル. 自然言語処理, **17** (1), pp. 161–182.

略歴

藤田 彬：2008 年横浜国大教育人間科学部卒業。2009 年同大学院環境情報学府博士課程前期短縮修了。同年より同大学院環境情報学府博士課程後期。現在に至る。自然言語処理、特に文章の自動評価、教育分野への応用の研究に興味を持つ。公益財団法人博報児童教育振興会第六回児童教育実践についての研究助成事業優秀賞。言語処理学会、情報処理学会、全国大学国語教育学会各会員。

藤田 央：2011 年電通大電気通信学部情報通信工学科卒業。同年より横浜国大学院環境情報学府博士課程前期。現在に至る。自然言語処理の研究に従事。情報処理学会員。

田村 直良：(正会員) 1985 年東工大大学院理工学研究科情報工学専攻博士課程了。工博。同年東工大助手。1987 年横浜国大工学部講師。助教授。1996 年同大教育人間科学部教授を経て、2001 年同大学院環境情報研究院教授。自然言語処理、特に文章構造の解析、文章理解の研究、音声合成、認識の応用、福祉情報工学の研究に興味を持つ。FIT2006 論文賞。2011 年電子情報通信学会論文賞。言語処理学会、情報処理学会、人工知能学会、電子情報通信学会各会員。

(2011 年 10 月 27 日 受付)

(2011 年 12 月 30 日 再受付)

(2012 年 1 月 6 日 再々受付)

(2012 年 7 月 19 日 再々々受付)

(2012 年 8 月 24 日 採録)