

RE-004

文章構造解析に基づく小論文の論理性についての自動採点

Automatic Scoring of Japanese Short Essays from a Viewpoint of Logicality based on Text Structure Analysis

藤田 彬[†] 田村 直良[‡]
Akira Fujita Naoyoshi Tamura

1. はじめに

本稿では、文間関係に基づいて文章をグラフ構造化し、文章構造から抽出した素性に基づいたモデルで文章を論理性に関して自動採点する手法について述べる。

近年、教育機関や企業における教育手段として、コンピュータ上で試験を行う e-learning システムが普及している。現在の日本の e-learning システムは、選択形式問題に対する回答を自動的に採点するシステムがほとんどである。しかし、選択形式問題のみの試験では教育手段として不十分であり、選択形式問題では測定できない論理的思考力や表現力を、自由記述形式問題で測定する試験も必要とされている。そのようなシステムの中で特に需要の高いシステムとして、小論文の自動採点システムが挙げられる。

小論文の自動採点に関する先行研究は[1]のサーベイが詳しい。代表的な小論文の自動採点システムとして、short essay (英語圏における小論文の呼び名) を自動的に採点する ETS の e-rater が挙げられる[2][3]。e-rater は文章上から多数の素性を抽出し、重回帰分析によって得られた回帰式にしたがって short essay を採点する。その精度は、97%である。一方、和文の小論文を対象にしたシステムとしては、石岡らが e-rater を参考に Jess というシステムを開発している[4]。

小論文を論理面から採点するという観点からいえば、Jess は接続詞などの表層表現を手掛かりにして隣接 2 文間の接続関係に着目することで、小論文を論理性に関して採点することができる。しかし、Jess における手法では、接続表現を含まない文や遠隔的な論理展開を含む文章への対応が問題となる。また、e-rater も論理性についての採点が可能である。しかし、e-rater において用いられている手法は英文を対象にした手法であるため、修辭的側面において英文とは異なる和文を対象にして十分な性能を発揮できるとは限らない。

そこで我々は、接続表現に限らず、語彙連鎖や照応関係、省略関係に着目して、より高い精度で文間関係を捉える手法を提案する。文間関係を捉える対象は、隣接する 2 文間の接続関係に限らず、離れた 2 文間の関係も含む。また、捉えた文間関係を基に文章を構造化して、構造に関する素性を含めて文章を論理性について採点する手法を提案する。

2. 文章の構造化のためのモデル

2.1 文のモデル

永野は文章構造を解明するための観点として、「文の接続および段落の接続」、「主語 (主題主語と主格主

語)の連鎖」、「陳述の連鎖」、「主要語句の連鎖」、「統括」を挙げた[5]。また、Halliday は節をメッセージとして組織化するものを「主題(theme)+題述(rheme)の結合体」であるとした[6]。本研究では、永野による主題主語と Halliday による主題¹を同質のものとし、永野が挙げた「文の接続」、「主語の連鎖」、「陳述の連鎖」を文章中から検出できるように、以下のようなモデル²で文を捉える (図1)。

- ◆ 陳述部 言表態度に関する記述。
 - 主体 陳述の主体
 - 態度/述部 陳述する態度を表明する表現
 - 態度/陳述形式 主体が陳述する態度の形式
- ◆ 内容部 言表事態に関する記述
 - 主題 何について述べられているかを示す語 (体言)
 - 主題修飾語 主題を修飾する語 (体言)
 - 題述 主題を展開する語 (体言または用言)

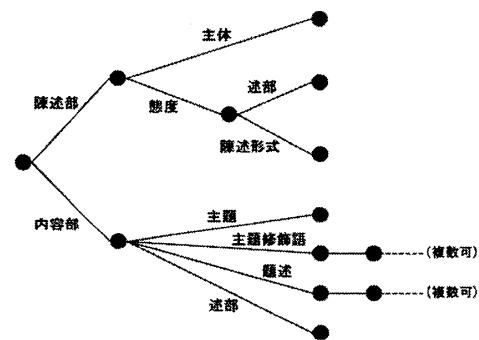


図1: 文のモデルのイメージ

たとえば、「私は民主主義の社会は世論によって形成されると考える。」という文を、提案するモデルで捉えんとする。このとき陳述部には「私は」と「と考える」が該当し、主体は「私」、態度/述部は「考える」、態度/陳述形式は「意見」である。また、内容部には「民主主義の社会は世論によって形成される」が該当し、主題は「社会」、主題修飾語は「民主主義」、題述は「世論」と「形成される」である。

2.2 結束性(cohesion)を捉えるモデル

Halliday は文法的もしくは語彙的要素による言語表現レベルの関連性を結束性(cohesion)と呼んでいる[7]。文章を構造化するためには cohesion を捉える必要がある。Halliday は cohesion を作り出す方法として「照応

¹ 永野は、文章の中で述べようとする (または、述べられている) 中心的な事柄の謂を、「(文章の) 主題」と呼んでいる。

² 現時点で本研究は、文のモデルの内容部のみに着目して主に「文の接続」と「主語の連鎖」の観点から文章構造を解明するにとどまっている。今後、「陳述の連鎖」を文章構造の解明を行う際の観点として追加するときには、陳述部にも着目するような拡張を行う。

[†] 横浜国立大学大学院環境情報学府
[‡] 横浜国立大学大学院環境情報研究院

(reference)」、「省略(ellipsis)」、「接続(conjunction)」、「語彙的結束性(lexical cohesion)」の4種類を挙げている[6]。本研究では、文脈上の「照応」、「省略」、「接続」、「語彙的結束性」を捉えるために、2文間の主題連鎖関係、語彙連鎖関係、照応関係、接続関係に着目する。

主題連鎖関係からは「語彙的結束性」と「省略」を、語彙連鎖関係からは「語彙的結束性」を、照応関係からは「照応」、接続関係からは「接続」をそれぞれ捉えることができる。ただし、主題連鎖関係には、本研究で提案する文のモデルに基づいた以下の4種類があるとする。

- ・**主題維持** ある文の主題が直後の文の主題と一致または類似するか、隣接する2文のうち後の文の主題が省略されている2文間の関係

- ・**主題変化** ある文の題述が直後の文の主題と一致または類似する2文間の関係

- ・**主題回復** ある文の主題が(直後ではない)後の文の主題と一致または類似する2文間の関係

- ・**遠隔主題変化** ある文の題述が(直後ではない)後の文の主題と一致または類似する2文間の関係

上記の4種類の主題連鎖関係からは「語彙的結束性」を捉えることができるが、主題維持からは「語彙的結束性」と共に「省略」も捉えることができる。

「照応」、「省略」、「接続」、「語彙的結束性」のいずれかの存在が認められる2文間には、**文間結合**があると呼ぶこととする。また、文間結合がある2文間の文脈上における修辭的な関係を**文間関係**と呼ぶこととする。

市川は文間関係を「順接」、「逆接」、「添加」、「対比」、「転換」、「同列」、「補足」、「連鎖」の8種類に分類した[8]。横野らは文脈形成の観点に基づいて、市川の分類を「論理的結合関係」「多角的連続関係」「拡充的合成関係」の3種類に大分類した[9]。本研究では、「論理的結合関係」と「多角的連続関係」を一元化して**論理的関係**とし、「拡充的合成関係」を**拡充的關係**と言い換えることとする。論理的関係は「2つの事柄を論理的に結び付ける関係」とし、拡充的關係は「1つの事柄に対して拡充して述べる関係」とする。

2.3 文章構造のモデル

本研究では、文章構造をグラフ構造で表す。この構造では、文をノード、文間関係をエッジとして文章全体をグラフ構造で表現する。エッジには次のような2種類がある。

- **Lエッジ(Logical Edge)**

論理的関係を表す。Lエッジが張られる2ノードはそれぞれ、演繹の条件となる内容をもつ文と、演繹の結果となる内容をもつ文であるとする。

- **Aエッジ(Augmentative Edge)**

拡充的關係を表す。Aエッジが張られる2ノードはそれぞれ、拡充的内容をもつ文と、拡充される文であるとする。

3. 小論文自動採点システム

文章構造に基づいて小論文を論理性的の面で自動的に採点する手法について提案する。本手法では、最初に文章(小論文)に対して形態素解析、係り受け解析を行なう。次に、文章構造解析をする。最後に文章構造上の素性を

用いた採点モデルで、文章を論理性について自動的に採点する。形態素解析には茶茎、係り受け解析には南瓜を用いている。

3.1 文章構造解析

文章構造解析は以下のように行われる。

1. 文のモデルに基づいて抽出した各文の主題と題述を手掛かりに文間結合の解析をする。(文間結合解析)
2. 教師あり機械学習モデルで、2文間(文章内総当たり)の文間関係を判定する。(文間関係判定)
3. 最適化問題として定式化したモデルで文章をグラフ構造化する。(グラフ構造化)

3.1.1 文間結合解析

文間結合解析では、文章中の全ての2文間に対して、前述の4種類の関係(主題連鎖関係、語彙連鎖関係、照応関係、接続関係)に関する素性の抽出を行う。

主題連鎖関係は、各種類の主題連鎖の有無を素性とする。語彙連鎖関係は、2文から抽出した名詞間の全ての組み合わせに対してそれらの間の類似度を計算し、全組み合わせの類似度の平均と分散、表層が一致する名詞の有無を素性とする。

照応関係に関しては、その有無を素性とする。“連体詞(「そんな」、「こんな」、「その」、「この」)+名詞”という照応表現のパターンが出現する文以前の文において、照応表現に含まれる名詞と同じ名詞が含まれている場合、2文間に照応関係があるとする。

接続関係に関しては、その有無を素性とする。接続表現を手掛かりにして、接続表現をもつ文は直前の文との間に接続関係があるとする。本研究で接続表現と設定した表現は61種類ある。

3.1.2 文間関係判定

一般には、文間関係は論理的性質と拡充的性質の双方を持つと思われるが、本研究では論理的性質と拡充的性質が相反するものと仮定して一次元的に扱う。すなわち、文間関係が論理的関係であるか拡充的關係であるかを一つの実数値で表す。この実数値を**LA値**と呼び、以下のように設定する。

$$LA_{ij} \in [-1, 1]$$

$$i, j \in N$$

N はノード(文)の集合を表し、 LA_{ij} はノード*i*とノード*j*の間に張られるエッジ(文*i*と文*j*の文間関係)のLA値を表す。LA値は-1に近いほど文間関係が拡充的關係としての性質を持つことを示し、1に近いほど論理的関係としての性質を持つことを示す。また、LA値の絶対値はノード間の文間結合の結び付きの強さを表す。絶対値が1に近いほど文間結合の結び付きが強いことを示し、0に近いほど結び付きが弱いことを示す。

文間関係判定のモデル(以下、文間関係判定器)には、SVM-Light(Regression-mode)を使用する。判定器の訓練事例となる訓練用LA値と事例ベクトルは、訓練データ中の文セット(2文)から抽出される。訓練データには、人手によりグラフ構造化された文章を用いる。人手によるグラフ構造化の結果、ノード間に結ばれたエッジがLエッジであった場合LA値を1、Aエッジであった場合LA値

を-1, ノード間にエッジが結ばれなかった場合 LA 値を 0 として訓練事例を作成する. 事例ベクトルは, 文間結合解析で得られる素性を要素とする.

3.1.3 グラフ構造化

2 文間の関係である文間関係を基にして, 文章中に他の文と文間関係を持たない文を含まないという条件で, 文章のグラフ構造化を行う. 本研究では, このグラフ構造化を以下のような整数計画問題として定式化する.

$$\min \sum_{i \in N} \sum_{j \in N} (eCOST_{ij} + rCOST_{ij})$$

s.t.

$$e_{ij} + e_{ji} \leq 1, \forall i, \forall j$$

$$\sum_{j \in N} (e_{ij} + e_{ji}) \geq 1, \forall i$$

$$\sum_{i \in N} e_{ii} < 1$$

$$eCOST_{ij} = -e_{ij} \cdot |LA_{ij}|$$

$$rCOST_{ij} = \omega \cdot e_{ij} \cdot |i - j|$$

N はノードの集合を, ij はそれぞれノードを示す. ij の値は, 対応する文の文章中における順序を示す. e_{ij} は i と j の間のエッジの実体を表す変数である ($e_{ij} \in \{0, 1\}$). e_{ij} の値は, 1 のとき i と j の間にエッジがあることを, 0 のときにエッジがないことを示す. LA_{ij} は i と j の間の LA 値である. $eCOST_{ij}$ と $rCOST_{ij}$ はそれぞれ i と j の間のエッジの存在コストを示す. ω は $rCOST_{ij}$ に重みづけをする定数である. $eCOST_{ij}$ と $rCOST_{ij}$ の和から計算されるスコア関数が最小となるときに導き出されるグラフ構造を, 最適化された文章構造とする. ただし上記の整数計画問題は, GLPK パッケージ[10]を用いて解く.

3.2 自動採点

グラフ構造化の結果として導き出される文章構造上の素性を用いて, 機械学習により文章を論理性について採点する. 論理性についての観点は「構成」, 「簡潔」, 「一貫」, 「説得」, 「明確」の 5 種類があり, 観点毎にスコアと文章構造上の素性のセットを学習させて 1 から 5 の実数で点数を出力する. スコアは大きいほど文章が優れていることを示す.

機械学習器 (以下, スコア判定器) には, SVM-Light(Regression-mode)を使用する. スコア判定器は, 一人の人間の採点者が採点したスコアと文章構造上の素性を事例として, 訓練する. 予備実験¹により, 異なる採点者が採点したスコアの間共通した傾向があるとは限らないことが判明した. このことから, 他者間でスコアの合意をとることに必ずしも妥当性があるとは言えない. したがって, 個人の採点者のスコアを対象にして採点モデルを構築する方針をとる. すなわち, 複数の採点者間の合意をとったスコアを基にスコア判定器を訓練するのではなく, 一人の採点者のスコアを基にスコア判定器を訓練する. ただし, 分散の少ない採点データを得ることが

出来れば, 複数の採点者間の合意を取ってスコア判定器を学習させることは可能である.

4. 評価実験・結果

提案手法を用いて小論文を論理性について自動採点する実験を, 以下の手順で行なった.

1. 高校生により書かれた 73 編の小論文を用意した. これらの小論文は全て同じ課題 (「英語の早期教育に対して意見を述べよ」) に沿って書かれている. また, 2 種類の字数制限に沿って出題されており, 400 字以内が 37 編, 800 字以内が 36 編ある.
2.
 - (ア) 全ての小論文に対して, 4 人の専門家 (以下, 専門家 A, B, C, D) がそれぞれ, 5 種類の論理性のスコア (「構成」, 「簡潔」, 「一貫」, 「説得」, 「明確」) について 1 点から 5 点の整数で採点した. これを正解論理性スコアとする.
 - (イ) 全ての小論文に含まれる 6954 個の文間関係の事例に対して, 大学生 2 名が -1, 0, 1 の三値で LA 値を付与した (以下, 人手によるグラフ構造). この人手によるグラフ構造に含まれる LA 値の事例を正解 LA 値とする. ただし, グラフ構造の作成者 (大学生 2 名) には「文をノード, 文間関係をエッジとすること」と「L エッジと A エッジの定義」のみ伝えてある. また, 作成者間で合意をとることはしていない.
3. 正解 LA 値を用いて文間関係判定器を訓練し, 正解論理性スコアを用いて文間関係判定器を訓練する. また, グラフ構造化処理において $rCOST_{ij}$ の制約式に含まれる重み ω を, 0.1 に設定する (予備実験の結果, 0.1 が最適値と判断された).
4. 全ての小論文について, 提案システムにより出力されるスコア (以下, 予測スコア) と専門家によって採点されたスコア (以下, 実測スコア) との間の MAE (平均絶対誤差) を評価する 5 分割交差検定を行なう. MAE の定義は以下の通りである. ただし, y_i は実測スコア, f_i は予測スコア, n は小論文の数をそれぞれ示す.

$$\frac{1}{n} \sum_{i=1}^n |y_i - f_i|$$

図 3, 図 4, 図 5, 図 6 に, それぞれ縦軸を MAE, 横軸をスコアの種類として実験結果を示す. ただし, 図 3 は専門家 A が採点したスコアを, 図 4 は専門家 B が採点したスコアを, 図 5 は専門家 C が採点したスコアを, 図 6 は専門家 D が採点したスコアをそれぞれスコア判定器の訓練に用いた場合の実験結果である. また表中の凡例は, 実測スコアとするスコアの採点者名とグラフのマーカ形状の対応を示す.

5. 考察

訓練事例と評価事例の採点者が同一の場合, MAE が 1 以内であることが多く, 十分な精度で論理性スコアを判定できているといえる.

本実験において人手によるグラフ構造が付与された小論文の事例数は, 73 と十分とはいえない量である. 事例を十分に用意して判定器を訓練することで, MAE はより小さくなると考えられる. また, 「説得」に関するスコア判定器の精度が訓練パターンに限らず他の観点到に比べ

¹図 2 に, 高校生が書いた 602 編の小論文を 4 人の専門家が 5 つの観点に沿って 1~5 点で採点した結果を示す. 縦軸は採点された小論文の数を, 横軸は点数を表している.

て低いことから、本実験で用いた素性中に論理展開の説得力を捉える素性が不足していることが考えられる。論理展開の説得力を捉える方法として、文章の起承転結や頭括型であるか尾括型であるかなど、陳述のパターンを捉え、パターン毎に説得力を計るという方法が考えられる。しかし、文章構造を表現する文間関係のバリエーションが論理的関係(L)であるか拡充的關係(A)であるかの2種類と少なく、この方法を実現するには不足であると思われる。このことから、文間関係の種類を詳細化した上で多種の文間関係を捉える方法を考える必要がある。

6. おわりに

本稿では、文間関係に基づいて文章をグラフ構造化して文章構造を導き出す手法と、導き出した文章構造に基づいて小論文を論理性について採点する手法を提案した。遠隔的な論理展開も含めた上で文章構造を導き出して採点を行うため、小論文の筆者の意図を詳細に把握した採点が可能となる。提案手法を評価した結果、採点者を特定すればその採点者の採点をシミュレートした採点システムが構築できることが示された。

今後の課題として、「陳述の連鎖」を文章構造解析の観点として追加することが挙げられる。特にスコア「構成」とスコア「説得」の採点は陳述面の解析結果と密接に関連すると考えられるため重要な課題である。また、文章構造を表すグラフに関して文間の演繹の方向性を考慮せず無向グラフと設定しているが、有向グラフとして捉えなおすことで文章構造の表現が詳細になり、採点精度が向上することが考えられる。また、論理的関係と拡充的關係に従属的な関係があると仮定してLA値をスカラー値として設定したが、各文間関係が独立した概念であることを前提にしたモデル化を行うことで文章構造解析の精度が向上する可能性が考えられる。

謝辞

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楊華氏、東京工業大学大学院社会理工学研究科人間行動システム専攻の前川眞一教授、東京大学大学院教育学研究科の宇佐美慧氏に感謝の意を表します。

参考文献

- [1] 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, No.1(2008).
- [2] Attali Yigal, Burstein Jill, "Automated essay scoring with e-rater v.2", Journal of Technology, Learning and assessment(2006).
- [3] Attali Yigal, Powers Don, "A Developmental Writing Scale", ETS Research Report(2008).
- [4] 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, No.1(2003).
- [5] 永野賢, "文章論総説". (1986).
- [6] M. A. K. Halliday, "An Introduction to Functional Grammar". (1994). (邦訳: 機能文法概説, 山口登, 笈壽雄 訳(2001)).
- [7] 田窪行則, 西山佑司, 三藤博, 亀山恵, 片桐恭弘, "談話と文脈". (2004).

- [8] 市川孝, "国語教育のための文章論概説". (1978).
- [9] 横野光, 奥村学, "テキストの結束性判定のための entity grid モデルの素性の検討", 情報処理学会研究報告, 2009-NL-189(2009).
- [10] A.Makhorin, "Reference manual of gnu linear programming kit, ver 4.9". (2006).

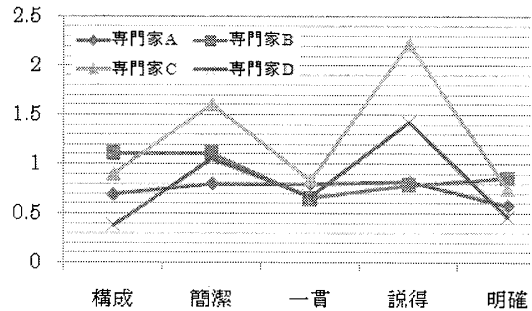


図3: 訓練事例を専門家Aの採点結果とした場合の評価実験結果

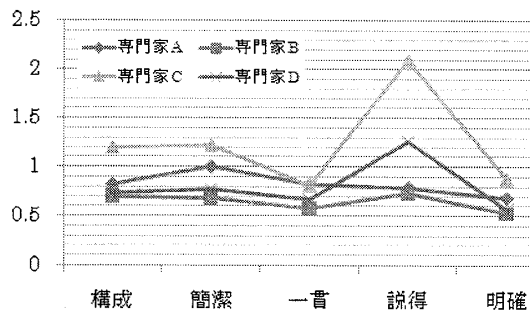


図4: 訓練事例を専門家Bの採点結果とした場合の評価実験結果

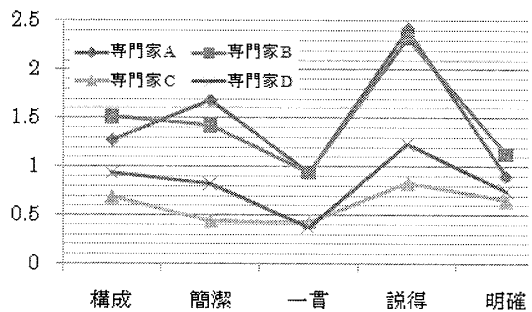


図5: 訓練事例を専門家Cの採点結果とした場合の評価実験結果

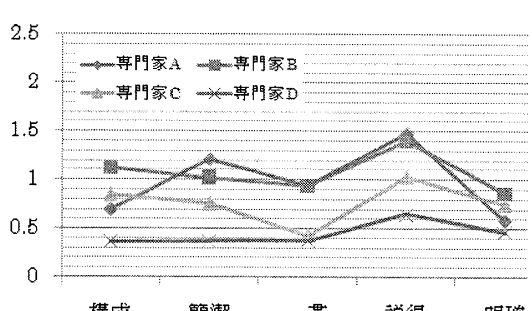


図6: 訓練事例を専門家Dの採点結果とした場合の評価実験結果

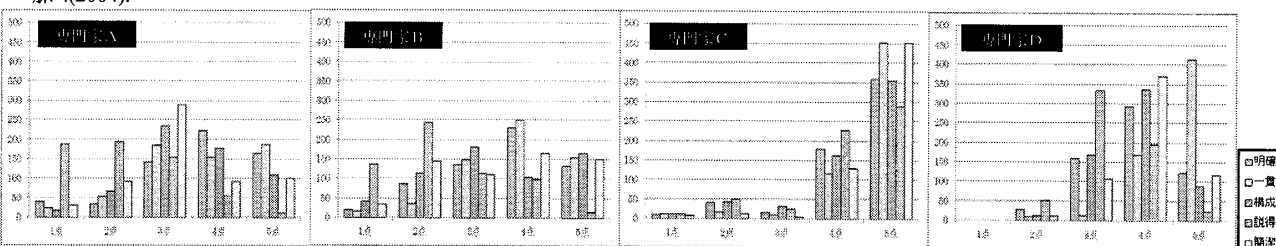


図2: 採点者間の採点傾向の違い (縦軸:小論文の数 横軸:点数)