

文章構造解析に基づく小論文の論理性の自動採点

Automatic Scoring of Japanese Short Essays from a viewpoint of logicity based on Text Structure Analysis

藤田 彬† 田村直良‡

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‡

E-mail: fujita@tamlab.ynu.ac.jp† tam@ynu.ac.jp‡

1 はじめに

本稿では、文間の連鎖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、教育手段としては選択式問題のみの試験では不十分であり、選択式問題では測定できない論理的思考力や表現力を、自由記述式問題で測定する試験も必要とされている。このような自由記述式問題を自動的に採点するシステムの中で、特にニーズが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。ただし小論文とは、一般的に、与えられた問題に対して回答者が意見を述べる論説文のことを呼ぶ。

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、英文を対象にした ETS の e-rater が挙げられる[2]。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たモデルにしたがって小論文(essay)を採点する。その精度は 97%である。一方、和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[3]。jess は、接続表現を手掛かりにして隣接 2 文間の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。

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、構造から抽出した素性を用い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

本研究では、図 1 に示すモデルで文の構造を捉える。以下でこのモデルについて述べる。

- ◆ **陳述部** 陳述内容に対する発話者についての記述
 - **主体** 発話者
 - **態度/述部** 主体の陳述態度に関する表現
 - **態度/modality** 主体の陳述態度
- ◆ **陳述内容部** 陳述内容についての記述
 - **主題** 文中で主体が話題の中心としている対象
 - **主題補助** 主題を修飾する語
 - **題述** 主題と主題補助以外の語
 - **述部** 用言

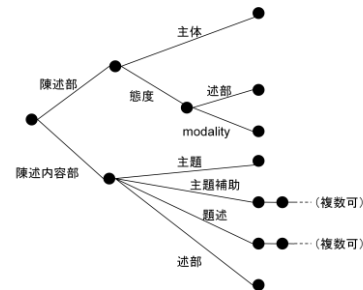


図 1: 文の構造を捉えるモデル

2.2 文間関係

文章を構造化するためには、文章の cohesion を捉える必要がある。cohesion を作り出す方法として Halliday は「照応(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[4]。本研究では、文章中の 2 文間の「省略」「語彙的結束性」「照応」「接続」をそれぞれ、「省略」「語彙的結束性」を「主題連鎖関係から、「語彙的結束性」を語彙連鎖関係から、「照応」を照応関係から、「接続」を接続関係から捉える。「省略」「語彙的結束性」「照応」「接続」のいずれかの存在が認められる 2 文間には、リンクがあると呼ぶこととし、リンクによって連鎖的に接続される文の集合をパスと呼ぶことにする。また、リンクがある 2 文間の文脈上における意味的な関係を文間関係と呼ぶことにする。

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 種類に分類している[5]。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 種類に大分類している[6]。本研究では、「論理的結合関係」と「多角的連続関係」を一元化した上で、「論理的関係」と「拡充的關係」の 2 種類に文間関係を大分類する。「論理的関係」は、二つの事柄を論理的に結び付ける関係とする。また、「拡充的關係」は、一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

本研究では、文章構造をグラフ構造で表す。この構造では、文章中の文をノード、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、L エッジ(Logical Edge)と A エッジ(Additional Edge)の 2 種類があるとする。L エッジは論理的関係を表す。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持つ文と、演繹の結果となる内容を持つ文であるとする。また、A エッジは拡充的關係を表す。A エッジが張られる 2 ノードはそれぞれ、拡充的な内容を持つ文と、拡充される文であるとする。

3 小論文自動採点システム

文章構造に基づいたモデルで論理性について小論文を自動採点する手法について述べる。最初に、文章(小論文)に対して形態素解析、係り受け解析を行う。次に文章構造解析を行う。最後に文章構造上の素性を用いたモデルで、文章を論理性について自動的に採点する。ただし、形態素解析器には茶莖、係り受け解析器には南瓜を用いる。

3.1 文章構造解析

文章構造解析は以下のような段階に分かれている。

1. 文構造上の要素を手掛かりに、リンクを解析する。(リンク解析)
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 リンク解析

リンク解析では、文章中のある 2 文間において、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。

主題連鎖関係に関する素性には、主題連鎖の種類を用いる。本研究では、主題連鎖の種類を以下の 4 種類とする。

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

これらの主題連鎖の有無を素性とする。

語彙連鎖関係に関する素性には、2 文から抽出した名詞間の総当たりでの類似度を用いる。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

文間関係を判定する処理は、ある 2 文間の文間関係を離散的に判定(「論理的関係」「拡充的關係」「無関係」の 3 値判定)するのではなく、文間関係が論理的関係であるか拡充的關係であるかを実数値で示すことを目的とする。この実数値を **LA 値** と呼ぶことにする。LA 値を以下のように定義する。

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

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リンクの結び付きの強さを表す。絶対値が 1 に近いほどリンク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 とし訓練事例を作成する。事例ベクトルは、リンク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。

$$\begin{aligned} \min \quad & \sum_{i \in N} \sum_{j \in N} (Ecost_{ij} + Rcost_{ij}) \\ \text{s.t.} \quad & e_{ij} + e_{ji} \leq 1, \forall i, \forall j \\ & \sum_{j \in N} (e_{ij} + e_{ji}) \geq 1, \forall i \\ & \sum_{i \in N} e_{ii} < 1 \\ & Ecost_{ij} = -e_{ij} \cdot |LA_{ij}| \\ & Rcost_{ij} = \omega e_{ij} \cdot |i - j| \end{aligned}$$

N はノードの集合を、 i, j はそれぞれノードを示す。 i, j 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[7]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、文章を論理性について採点する。論理性についてのスコアは「構成」「簡潔」「一貫」「説得」「明確」の5種類があり、それぞれ1から5の実数で表される。スコアは大きいほど、良いことを示す。

採点するモデル(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。本研究では、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。異なる採点者が採点したスコアの間から従属的な関係があるとは限らないことから、他者間でスコアの合意をとることに妥当性があるとは必ずしも言えない。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。

図2に小論文自動採点システムのフローチャートを示す。

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。

まず、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

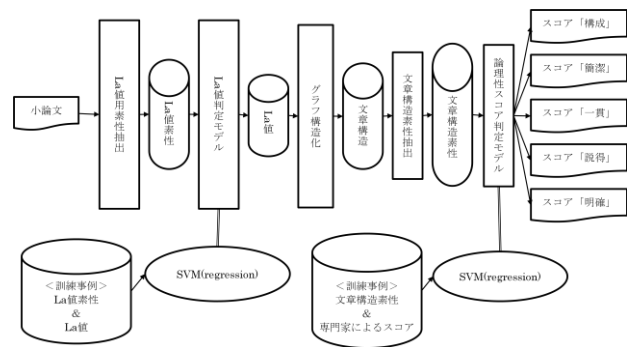


図2：小論文自動採点システムのフローチャート

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。これらの73編の小論文に対して、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。また、全ての小論文について、大学生2名が人手でグラフ構造を作成した(以下、人手によるグラフ構造)。人手によるグラフ構造中には、文間関係の事例(LA値の事例)が6954事例含まれている。ただし、グラフ構造の作成者には「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」のみ伝えた。

次に、73編の小論文について、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。

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

提案システムの訓練は、交差検定において訓練集合となる小論文事例の正解LA値と正解論理性スコアを用いて文間関係判定器とスコア判定器の両者を訓練することで、行なわれる。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(事前調査を行なった結果、0.1が最適値と判断された)。

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、表1は専門家Aが採点したスコアを、表2は専門家Bが採点したスコアを、表3は専門家Cが採点したスコアを、表4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の形状の対応を示す。

5 考察

採点モデルの構築対象となる専門家のスコアをテスト事例とした場合、平均絶対誤差が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、スコア「説得」の精度が訓練パターンに限らず他のスコアに比べて低いことから、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、文章構造をパターン化し、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、小論文を論理性について採点する目的に対して有用であることが示された。

今後の課題として、文間関係の解析にオントロジーを適用することで、より意味面を考慮した文間関係の抽出を可能とすることが挙げられる。

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, pp.17-24(2008).
- [2] AttaliYigal, Burstein Jill. Automated essay scoring with e-rater v.2. Journal of Technology, Learning and assesment, (2007).
- [3] 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, pp.3-18(2003).
- [4] M. A. K. Halliday. An Introduction to Functional Grammer. (1994).
- [5] 市川孝. 国語教育のための文章論概説. 教育出版, (1978).
- [6] 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).
- [7] A.Makhorin. Reference manual of gnu linear programming kit, version 4.9. Technical report(2006).

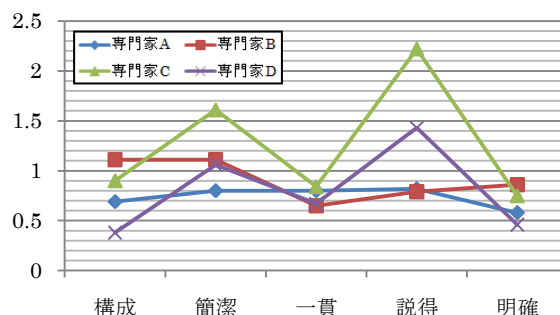


図3：専門家Aが採点したスコアで判定器を訓練した場合の提案システムの評価実験結果

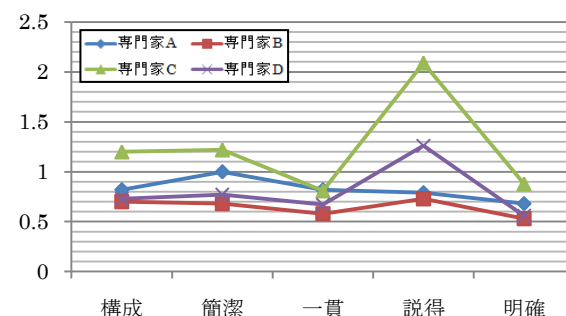


図4：専門家Bが採点したスコアで判定器を訓練した場合の提案システムの評価実験結果

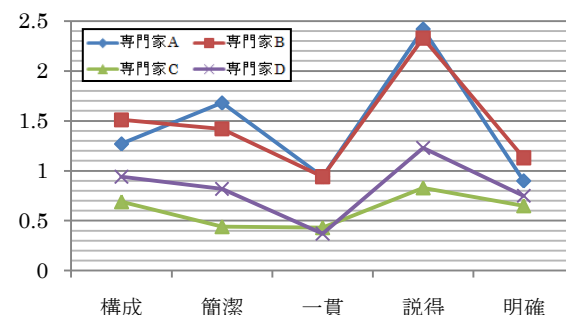


図5：専門家Cが採点したスコアで判定器を訓練した場合の提案システムの評価実験結果

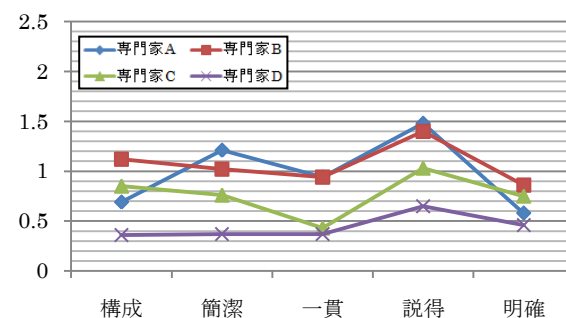


図6：専門家Dが採点したスコアで判定器を訓練した場合の提案システムの評価実験結果