

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

勝又大介[†] 藤田彬[§] 田村直良[‡]

[†] 横浜国立大学大学院 環境情報学院

[§] 成長戦略研究センター 産学連携研究院

[‡] 横浜国立大学大学院 環境情報研究院

{ katsumata-daisuke-rx[†], fujita-akira-rm[§], tam[‡] }@ynu.ac.jp

1 はじめに

本稿では、文間関係に基づいて文章を有向グラフにより構造化し、小論文の論理構成の整然さに関して自動評価する手法を提案する。

近年、教育機関における大規模な記述式試験の実施にかかる人的コストを減らすことや、e-learningにおける自由記述式問題の学習支援を目的とした自由記述式問題の自動採点技術に対する需要が高まっている。その中でも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。和文を対象にしたシステムとしては、石岡らが e-rater [2] を参考に開発した Jess [3] が挙げられる。

しかし Jess では、文脈展開に関して評価する「論理構成」という観点において、隣接 2 文間の接続表現に着目した評価を行なうのみであり、文脈展開を捉えて評価を行なうという点で不十分である。また、Jess の問題点に対して、藤田らは離れた 2 文間の照応や語彙連鎖による関係も考慮した文章構造解析に基づく評価手法を提案している [4]。しかし、藤田らの手法でも、文間関係における順序関係を考慮しないため、文章構造は無向グラフで表現され、前提から演繹的に主張が導かれる展開を正確に捉えることができない。

そこで本研究では、順序関係を考慮した文間関係の解析により文章を有向グラフにより構造化し、小論文の内容的構成が論説として十分であり、それが自然な順序で記述されているか (以下、**論理構成の整然さ**) について自動評価する手法を提案する。

2 本研究で解析する文章構造

2.1 文間関係

本研究では、2 文間にある修辭的な関係を文間関係と呼ぶ。文章の修辭的な関係を捉える枠組みとしては、修辭構造理論 (Rhetorical Structure Theory: RST) などが提案されている [10]。RST では、文章中の隣接する文またはセグメント間に 25 種類の修辭関係を規定している。修辭関係は、二項間を核 (nucleus) と核に従属的な衛星 (satellite) として関係付ける単核 (mononuclear) 関係

と、二項間を等位的に関係付ける多核 (multinuclear) 関係に大別される。

本研究では、小論文の論理構成を正確に捉えるために RST で定義される核関係 (nuclearity) と、論述において不可欠な根拠などの前提を捉える修辭関係に基づいて、「単核因果関係」「単核拡充関係」「多核関係」という 3 種類の文間関係を定義する。「単核因果関係」は「一方の事柄が他方の事柄の前提となる関係」とし、「単核拡充関係」は「一方の事柄が他方の事柄の補足的な内容となる関係」、「多核関係」は「2 つの事柄を別々に結び付ける関係」とする。RST で規定される関係と本研究で定義する関係の対応を表 1 に示す。

2.2 文章構造

本研究では、文章構造を有向グラフで表す。グラフにおいて文をノード、文間関係を有向エッジで表す。各文間関係は次の 3 種類のエッジで表す。

- **C エッジ (Causality Edge)** 単核因果関係を表す。エッジの始点には RST における衛星、終点には核が対応する。
- **A エッジ (Additional Edge)** 単核拡充関係を表す。エッジの始点には RST における衛星、終点には核が対応する。
- **M エッジ (Multinuclear Edge)** 多核関係を表す。2 文間を双方向エッジで結ぶ。

RST では、文章構造は再帰的な階層構造として定義され、二項関係同士は交差しないという構造的な制約がある。しかし、RST の制約は一貫性のある文章に対して成立するものであり、一方で本研究では一貫性の無い文章も解析対象に含まれる。したがって、本研究では文章構造を有向グラフと定義し、RST の制約が必ずしも満たされないことを前提に文章構造解析を行なう。

3 論理構成の自動評価手法

文章構造に基づいて小論文の論理構成における整然さを自動評価する手法を提案する。提案手法では入力

表 1: 本研究で使用する文間関係

文間関係	対応する RST の関係
単核因果関係	Condition, Evidence, Justify, Motivation, Purpose, Non-volitional Cause, Non-volitional Result, Volitional Cause, Volitional Result
単核拡充関係	Antithesis, Background, Circumstance, Concession, Elaboration, Enablement, Evaluation, Interpretation, Otherwise, Restatement, Solutionhood, Summary
多核関係	Contrast, Joint, List, Sequence

された小論文に対して最初に文章構造解析を行ない、文章構造から抽出した素性を用いた評価モデルにより小論文を自動評価する。

3.1 文章構造解析

3.1.1 使用する素性

本研究では機械学習手法を用いて文章構造解析を行なう。使用する主な素性を以下に示す。

結束性 (cohesion) を捉えるための素性

文間の照応関係の有無、後文の接続表現の有無、主題連鎖関係、語彙連鎖関係など

文間関係を捉えるための素性

各文に出現する文末機能表現、接続表現、手がかり語などの表層表現

文の統括力を捉えるための素性

各文のモダリティ、接続表現などの分類、提題表現の有無、主題が共通する文の数

文章中での位置関係に関する素性

各文の所属する段落の位置、段落内での位置、2文間の段落間距離、文間距離など

結束性 (cohesion) を捉える素性に関しては Halliday [7] の挙げる「指示 (reference)」「省略 (ellipsis)」「接続 (conjunction)」「語彙的結束性 (lexical cohesion)」を捉える。照応表現には指示表現を省略主題を抽出し、先行詞の推定を村田らのルールベース手法 [11] により行なう。主題連鎖関係と語彙連鎖関係については、藤田らと同一の素性を使用する [4]。文間関係を捉えるために使用する手がかり語には林 [5] の「思考運用のための用語」と小林らの評価極性辞書 [6] を参考にした。手がかり語には、「根拠」や「結論」など筆者が論旨を展開するために使用する語が含まれる。

また、核関係を捉えるための素性には永野 [8] の規定する「統括」という観点を参考にした。「統括」とは、文章において話題を締めくくる文の特徴 (統括機能) に着目して、文章の完結性と統一性について分析するための観点である。核関係の解析において、文の統括機能を捉えることは重要であると考えられる。佐久間 [9] は分析を基に文の統括機能を示す言語形式として「接続表現」「指示表現」「反復表現」「省略表現」「提題表現」「叙述表現」を挙げている。本研究では、「接続表現」「反復表現」「提題表現」「叙述表現」に関する統括に関する言語形態を参考にし、素性とする。

3.1.2 文章のグラフ構造化

本研究で扱う文章構造は、RST で規定される構造的制約に反する構造も許容するが、RST に基づいた文間関係を定義している以上、グラフ構造上で部分的に RST の制約を満たす箇所があると考えられる。そのため、本研究では Markov Logic Networks (MLNs) [12] を使用することにより RST の制約を解析の手がかりと捉え、学習することで重み付き制約として使用し、解析モデルを構築する。

MLNs は、Markov Networks に一階述語論理を組み合わせた統計的關係学習 (Statistical Relational Learning) という枠組みの 1 つである [12]。MLNs のモデル構築には、解析対象である潜在述語 (hidden predicate)、既知情報である観測述語 (observed predicate) を定義し、それらを基に推論に有効な論理式を定義し、論理式の重みを学習する。また、論理式には潜在述語が 1 つしか含まれない局所論理式 (local formula) と潜在述語が複数含まれる大域論理式 (global formula) があり、重みを学習によって与えられる式 (soft constraint) の他に、厳密制約 (hard constraint) として使用する式を定義することもできる。

本研究で解析対象とする潜在述語を次のように定義する。

nuclearity(i, j, n)

文 i は文 j と関係を持ち、その核関係は n である

relation(i, j, r)

文 i は文 j と関係を持ち、その関係は r である

潜在述語 nuclearity は核関係を示す述語であり項 n は “N-S” と “S-N”, “N-N” のいずれかをとる。潜在述語 relation は文間関係を示す述語であり項 r は “C” と “A”, “M” のいずれかをとる。ただし、relation に関しては C エッジについてののみ学習し、解析の際には C エッジでない場合は nuclearity に基づき M エッジか A エッジに分類する。

この規則は大域論理式により hard constraint 定義する。その他に、大域論理式では RST の制約について学習するために、ノードからの出次数がたかだか 1 であることと、根ノード (出次数 0 のノード) が複数存在しないこと、そしてエッジが交差しないことを soft constraint として定義する。例えばノードからの出次数がたかだか 1 であることは以下のように定義した。

$$\begin{aligned} & \text{nuclearity}(i, j, \text{"S-N"}) \wedge j \neq k \\ & \Rightarrow \neg \text{nuclearity}(i, k, \text{"S-N"}) \end{aligned} \quad (1)$$

これは、文 i が複数の文 (文 j と文 k) に対する satellite とならないことを表す。このような論理式により、RST の構造的制約を選好性として学習することで、文章構造解析のための MLNs を構築した。

3.2 論理構成の整然さの自動評価

3.2.1 使用する素性

グラフ上のパス・エッジに関する素性

平均・最長パス長、C エッジを含まないパス数、終点が意見文のパス数、始点が叙述文であるパス数、交差するエッジの数など

ノードの次数に関する素性

意見文・叙述文の入次数、意見文・叙述文の出次数、最大入次数、入次数が最大である文に係るエッジの種類、出次数 0 の文の数、入次数 0 の文の数、出次数が 2 以上の文の数など

グラフ構造から抽出する素性は、個々の話題を掘り下げていることや、モダリティや文間関係から見た展開の適切さや、RST の制約に反している箇所などに着目した観点に基づいている。

また上記の素性の中には、自動評価へ寄与の仕方が文章の頭括型、尾括型などの構成の違いにより変動する素性があると考えられる。そのため本研究では、文章構成によらず共通して上記の素性を学習する他に、「統括型」「中括型」「尾括型」という 3 種類の文章構成ごとに区別した上記の素性を用意することで、文章構成の違いを考慮したモデルを構築した。文章構成の判断は、グラフ構造上で最も多くの文から導出可能な文を「主題文」とし、その主題文の位置が文章の先頭から $\frac{1}{4}$ 以内の位置に出現する場合を「頭括型」、最後から $\frac{1}{4}$ 以内の位置に出現する場合を「尾括型」、それ以外を「中括型」に分類した。

3.2.2 スコア推定モデル

解析した文章構造上の素性を用いて、機械学習手法により文章の論理構成における整然さについて評価する。学習器には、SVM-Light(Regression-mode) を使用し、一人の評価者が評価したスコアと素性を事例として、学習する。学習するスコアは 1 から 5 の整数値とし、システムからは 1 から 5 の実数値が出力され、値は大きいほど文章が優れていることを示す。採点者ごとに採点傾向が異なるため、他者間でスコアの合意をとることに妥当性があるとは言えない。したがって、各採点者ごとに採点モデルを構築する方針をとる。採点に使用する主な素性を以下に示す。

4 評価実験・結果

提案手法を用いて文章構造解析と、小論文を論理構成について自動評価する実験を行なった。

実験に使用したデータは、高校生により書かれた 100 編の小論文と、それらに対するグラフ構造データ、

4 名の専門家 (専門家 A, B, C, D) それぞれにより付与されたスコアである。小論文は 2 種類の論題 (「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ついて、あなたの意見と根拠が明確になるように論じなさい」「日本の親の子育ての態度に関してどのような客観的特徴がデータから読み取れるか、また望ましい子育てのあり方についてあなたの考えをまとめて論述しなさい」) の回答により構成されており、いずれの論題も字数制限は 800 字である。グラフ構造データは 1 名のアノテータにより作成されたものである。

4.1 文章構造解析実験

文章構造解析の実験では、潜在述語の解析に関する精度、再現率と F 値について 2 分割交差検定による評価を行なう。ここで、エッジの有無を示す述語 relate を、解析には寄与しないがエッジの存在を正しく推定できるかを評価するために定義した。なお、MLNs の学習器には Markov thebeast¹ を使用する。実験の結果を表 2 に示す。

4.2 自動評価実験

自動評価の実験では、各評価者の評点を用いて学習したモデルについて、点数推定の精度と平均絶対誤差 (MAE) を評価する 5 分割交差検定を行なう。ベースラインには、文章構造の核関係を考慮しないで素性抽出して学習したモデルを、藤田らのモデルを模倣した手法として用意し、提案手法と比較する。また、現時点で文章構造解析の精度が、自動評価へ応用する上で十分とは言えない。そのため、評価モデルの構築にはアノテータにより作成されたグラフ構造データを使用して実験を行なった。精度の評価では専門家により与えられたスコア (実測値) とシステムの推定したスコア (予測値) の絶対誤差が 0.5 以内である事例を一致とみなし、MAE は以下の式により算出する。ただし、 y_i は実測値、 f_i は予測値、 n は小論文の数をそれぞれ示す。

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - f_i| \quad (2)$$

各評価者における実験の結果を表 3 に示す。

5 考察

文章構造解析の実験では文間関係を持つ 2 文間の対応を約 7 割の再現率で解析することができた。しかし、核関係の解析と文間関係の解析それぞれについては約 5 割の再現率となり、最終的に共に正しく解析できたエッジについては 41% となった。

正しく解析できなかった事例については、表層的な手がかりが少ない事例が多かった。このような事例に

¹<http://code.google.com/p/thebeast/>

表 2: 文章構造解析結果

	relate	nuclearity	relation	nuclearity ∧ relation
精度	0.71	0.63	0.55	0.34
再現率	0.70	0.56	0.53	0.41
F 値	0.70	0.60	0.54	0.37

ついて正しく解析するには、事象や語彙に関する知識をさらに活用してより意味を考慮した解析を行なうことが挙げられる。また、離れた 2 文間についても関係解析の精度が低かった。自動評価という応用を考慮すると、離れた文間の関係を正しく解析することは非常に重要である。本研究では RST の構造的な制約を手がかりとして使用したが、周辺文脈をより多く考慮して集合的に解析を行なうことが離れた文間の解析精度に必要であると考えられる。

小論文の自動評価の実験では、ベースラインと比較して精度と MAE において共に提案手法が優れた結果が得られた。専門家 D のスコアを用いた実験結果が他の結果と比較して精度が高いのは、専門家 D のスコアの分布が他の専門家と比較して狭かったためである。文間の核関係を考慮することで、グラフ構造上では葉となるノードには主張の前提となる文が現れ、根となるノードには筆者の強調する主張が現れる。したがって、グラフ上のパスや次数にを基に、筆者の主張と前提のような文の役割が区別できるようになり、それにより各種素性の学習性能が良くなったために精度が向上したと考えられる。

しかし、実用化を考慮するとさらなる精度向上が必要である。精度を向上させるためには、より多くのサンプルを学習することが必要であると考えられる。今回の実験では 100 編の小論文を使用したため、1 点から 5 点についての精度が低かった。これは、1 点と 5 点の事例が少なく、点数のスケールリングが十分に行えなかったためであると考えられる。また、さらに文間関係を詳細化することも、論理構成をより正確に捉えるために必要であると考えられる。

6 おわりに

本稿では、文間関係に基づいて文章を有向グラフにより構造化する手法と、解析したグラフ構造に基づいて小論文の論理構成における整然さを自動評価する手法を提案した。文間関係の順序関係も解析するため、前提から主張が導かれる展開を正確に捉えることができ、主張となる文を中心とした文章の論理構成を把握した評価を行なうことができる。提案手法により小論文を自動評価した結果、文間関係における順序関係が評価精度の向上に寄与することが確認できた。

今後の課題としては、文章構造解析と小論文自動評価の一貫処理について評価実験を行なうことが挙げられる。また、その際に十分な性能を実現するためにも文章構造解析のさらなる精度向上が必要である。因果関係などを高い精度で解析するには、語彙に関する知識を活用する必要がある。そして文章構造解析の精度

表 3: 自動評価結果

	評価者	A	B	C	D
ベースライン	精度	0.40	0.39	0.44	0.45
	MAE	1.21	1.39	1.23	0.81
提案手法	精度	0.49	0.41	0.49	0.64
	MAE	0.96	1.34	1.08	0.72

が向上することで、より細かい文間関係を捉えた評価も可能となり、自動評価の精度向上が期待できる。

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(2008).
- [2] Attali Y., Burstein J., “Automated essay scoring with e-rater v.2”, Journal of Technology, Learning and assessment(2006).
- [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(2003).
- [4] 藤田彬, 田村直良, “文章構造解析に基づく小論文の論理性の自動採点”, 言語処理学会 第 16 回年次大会 発表論文集, pp.796-799(2010).
- [5] 林四朗, “語彙調査と基本語彙”, 電子計算機による国語研究 III, 国立国語研究所報告 39, pp.1-35(1971).
- [6] 小林のぞみ, 乾健太郎, 松本裕治, 立石健二, 福島俊一, “意見抽出のための評価表現の収集”, 自然言語処理, Vol.12, No.2, pp.203-222,(2005).
- [7] M. A. K. Halliday, “An Introduction to Functional Grammar”(1994).
- [8] 永野賢, “文章論総説”(1986).
- [9] 佐久間まゆみ, “現代日本語の文章構造類型”, 日本女子大学紀要 文学部, Vol.48, pp.1-28(1999).
- [10] Mann W. C., Thompson S. A. “Rhetorical structure theory: Description and construction of text structures”, In G. Kempen Ed., Natural Language Generation, Vol.135, pp.85-96, Martinus Nijhoff Publishers(1987).
- [11] 村田真樹, 長尾眞, “用例や表層表現を用いた日本語文章中の指示詞・代名詞・ゼロ代名詞の指示対象の推定”, 自然言語処理, Vol.4, No.1, pp.87-109(1997).
- [12] Richardson M., Domingos P. “Markov Logic Networks”, Machine Learning, Vol.62, No.1-2, pp.107-136(2006).