

文章構造解析に基づく小論文の自動評価

Automatic Essay Scoring based on Text Structure Analysis

藤田 彬† 五條 善雅†† 田村 直良†††

† 横浜国立大学教育人間科学部 †† 東芝テック株式会社

††† 横浜国立大学大学院環境情報研究院

{†tactica, †††tam}@tamlab.ynu.ac.jp , †† Yoshimasa.Gojo@toshibatec.co.jp

1 はじめに

本論文では、論説文における筆者の主張や論理展開に着目した文章構造と、それに基づく小論文の自動評価手法について述べる。

近年、e-learning システムのように無人でシステムとのインタラクションを進めていく形態が求められているが、その中で自由記述式問題の自動評価システムに対するニーズが高まっている。本研究では、自由記述式問題の自動評価システムの中でも、小論文問題の自動評価システムに着目する。小論文問題とは、与えられた問題文に対する回答者の意見や考えを問い、文章で回答を求める問題形式である。

小論文自動評価の先行研究は [2] のサーベイが詳しい。英文の自動評価においては、PEG[3] に始まり、e-rater[1] などのシステムが代表的である。また、和文の自動評価においては石岡らが e-rater を参考に Jess というシステムを開発している [5]。これらのシステムは、表層的な評価が主で論理構造の評価に関しては多角的な評価が十分とは言えない。

本論文では、文章の論理構造を多角的に評価するため、文章の構造解析を行なう。そのためにまず、論説文において、論理展開の全体的な流れである主題の連鎖構造と主張の展開、論理展開の局所的な特徴である文の接続表現、文間の関係、文間の結束性に基づいて文章を構造化した文章構造モデルを提案する。また、小論文を人手で評価する際に用いられる評価基準を検討し、これに基づき文章の構造上の特徴量を評価パラメータとして定義し、その評価手法について述べる。

2 文章構造モデル

2.1 文章構造モデルの提案

筆者の主張の展開を把握するために文を 2 種類に分類し、以下のように定義する。

主張文 筆者の主張が述べられている文

叙述文 筆者の主張を含まない文

これら主張文と叙述文が混在する文章から筆者の主張や要望、論理展開などの意味的情報を抽出するためには、意味的情報の断片となる要素を抽出する必要がある。この際、文毎に筆者の主張や要望に着目した文構造とそれらを結合させた文章の構造化により要素の抽出が可能になる。また、その文構造は筆者の主張や要望に関する要素の抽出に特化することで、筆者の主張の展開をよりモデル化した意味的情報の抽出が可能となる。

2.1.1 文構造

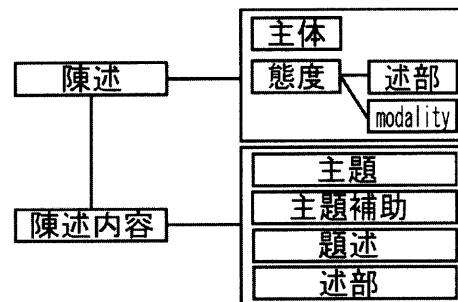


図 1: 文構造

図 1 に示したような文構造を用いる。以下でこの文構造について述べる。

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 主体の活動を表す述部

態度 / modality 発話者の判断や心的態度

陳述内容部 陳述内容についての記述

主題 文の中で筆者が話題の中心として取り上げている対象

主題補助 主題を直接、または間接的に修飾する語

題述 主題以外の語

述部 主題が係り、様相を含む述部

主張文では、陳述部に陳述内容に対する発話者について、陳述内容部に陳述内容について記述し、叙述文では、陳述内容部に陳述内容について記述する(陳述部は未定義)。

2.1.2 文間関係

文間の主題の連鎖関係、および接続表現等による隣接文間の関係を文間関係とする。

そのうち主題連鎖関係は、以下の2種とする。

主題連鎖 2文の主題が同一の語または類似する語どうしである関係

焦点 - 主題連鎖 ある文の題述またはその類似する語が後の文の主題になっている2文間の関係

接続関係は、接続詞に着目し、理由、例示、換言、強調、累加、説明、逆接、並列、選択、対比、提起、因果、結論、順接、相反、一般化、根拠、予測、条件、展開の20種類に分類する。図2に文間関係の例を示す。

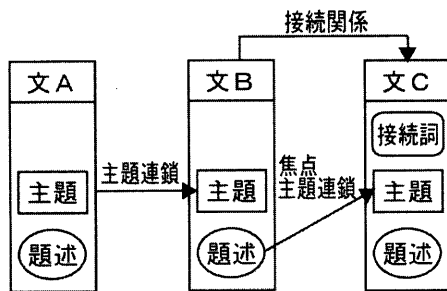


図 2: 文間関係の例

3 文章構造モデル各要素の抽出手法

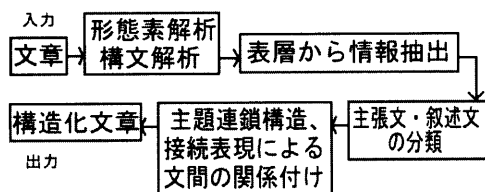


図 3: 文章構造解析の手順

本研究では、図3で示したように文章構造解析を行う。形態素解析処理には ChaSen[7]、構文解析処理には Cabocha[8] を用いている。

3.1 主張文と叙述文の自動分類

文末の表現に着目して、主張文と叙述文の分類を行なう。「意見文」の判定は、NTCIRでタスクとして取り上げられるなど[4]独立したテーマで

あるが、本研究では単純化のため、文の文末部のマッチングから、主張文と叙述文の分類をする。この際、主張モダリティを意見、問掛、推量、断定、理由、判断、要望、義務の8種類に、叙述モダリティを、叙述、可能、伝聞、様態、存在、継続、状態、使役、例示の9種類に分ける。

3.2 文構造の各要素の抽出手法

3.2.1 陳述部

主体 文が主張文と分類されている場合、主体を筆者とする。ただし、文が叙述文と分類されている場合は主体は未定義とする。

態度 / 述部 文が主張文と分類されている場合、文の文末部を態度 / 述部とする。ただし、文が叙述文と分類されている場合、態度 / 述部を未定義とする。

態度 / modality 主張文と叙述文の自動分類を行った際に、決定したものを態度 / modalityとする。ただし、文が叙述文と分類されている場合、態度 / modality を未定義とする。

3.2.2 陳述内容部

主題 主題を提示する助詞を手がかりに抽出を行う。主題を提示する助詞は、永野[6]による。

主題補助 主題を直接、または間接的に係り先とする句に含まれる名詞を主題補助とする。

題述 主題以外かつ主題補助以外の名詞を題述とする。

述部 主題を含む句の係り先である用言を述部とする。

3.3 文間関係モデルの自動構築

3.3.1 主題連鎖関係

まず、ある文の主題に着目する。その主題と前方の文の主題とを照合し、もし同一か類似関係にあった場合は、2文間に**主題連鎖リンク**を与える。また、同様に前方の文の題述と同一か類似関係にあった場合は、2文間に**焦点 - 主題連鎖リンク**を与える。

3.3.2 接続関係

文章中から接続詞を抽出し、接続関係を決定する。もし、接続関係があると認められた場合は、接続詞を含む文とその直前の文の2文間に**接続関係リンク**を与える。

3.3.3 リンクの統合

今回のシステムは文間関係を統合して扱う。2文間の関係として与えられたリンクを連鎖して、リンクによって直接的、間接的に接続されたノードの集合を、パスと呼ぶことにする。

4 評価パラメータ

評価に際し、どのような観点で文章を評価するのかを検討した結果、「主張の正当性」など以下の10種類を用いる。さらにこれらを、前述の文章構造上で以下のように定式化する。なお、段落とは形式段落を指す。

主張文の割合 A = 主張文の個数、 S = 文の数

$$E_1 = \frac{A}{S} \quad (1)$$

主張の正当性 D_a = 主張文とリンクがある叙述文の個数

$$E_2 = \frac{D_a}{A} \quad (2)$$

話題の発展性 L_{max} = パスの最長値

$$E_3 = \frac{L_{max}}{S} \quad (3)$$

結論との結束性 L_c = 結論部となる段落とリンクがある文の個数

$$E_4 = \frac{L_c}{S} \quad (4)$$

文章全体の結束性 L = リンクの総数

$$E_5 = \frac{L}{S} \quad (5)$$

段落間の関連性 L_p = 異なる段落間のリンクの個数、 P = 段落の数

$$E_6 = \frac{L_p}{P} \quad (6)$$

段落内の結束性 $L_n=n$ 段落内のリンクの個数、 $S_n=n$ 段落の文の数

$$E_7 = \frac{1}{P} \times \sum_{n=1}^p \frac{L_n}{S_n} \quad (7)$$

テーマとの関連性 N_s = テーマと類似する名詞の個数、 N = 文章中で出現する名詞の総数

$$E_8 = \frac{N_s}{N} \quad (8)$$

文章構成の多様性 v_s = 出現する態度 / modality の種類の数、 V_s = 態度 / modality の数

$$E_9 = \frac{v_s}{V_s} \quad (9)$$

文章展開の多様性 v_l = 出現する文間関係の種類の数、 V_l = 文間関係の種類の数

$$E_{10} = \frac{v_l}{V_l} \quad (10)$$

5 実験と考察

5.1 主張文判定

システムの主張文判定性能について検証する。6編の文章（合計95文）に対して人手で主張文判定を行い、その結果を正解データとする。システムの主張文判定の結果を、表1に示す。

5.2 評価実験

小論文の自動評価手法についての評価実験を行なう。実験には、小論文の模範文集 [9] に掲載されている2テーマ2編の文章（模範文A、B）と、それと同一のテーマを与えて被験者4名による2テーマ4編（1テーマ当たり2編）の文章（文章 α 、 β 、 γ 、 δ ）、合計6編の文章を使用した。ただし、文章 α 、 β が模範文Aと、文章 γ 、 δ が模範文Bと同じテーマを与えられている。

5.2.1 手法と結果

自動評価システムと評価パラメータの具体的な算出方法を知らない人間の評価者1名にそれぞれ6編の文章を評価させ、システムと比較した。システムの評価者の評価結果についてまず、パラメータ毎に評価結果値の偏差値を算出した。次に、その偏差値に関するシステムと人間との間でのピアソン積率相関係数を算出した。結果を表2、3、4に示す。

5.2.2 考察

E_1, E_3, E_4, E_5, E_9 は、相関係数が0.3を超えており相関していると考えてよい。これらの観点の内には文章の展開に関するものも含まれており、一応所期の目標には達している。しかし、その他の評価パラメータは負の相関が生じるなど、相関していると言えない。

E_6 、 E_7 システムによる評価手法では段落の総数の多い文章を評価した際、低く評価され過ぎる傾向がある。

E_8 人手による評価で総合的に低い評価を得た文章でも関連がある語というだけで高い評価を得ていることから、定義を再考すべきである。

表 1: 主張文判定

適合率	再現率	F値
0.833	0.816	0.824

表 2: 自動評価システムの評価結果

	A	α	β	B	γ	δ
E1	0.76	0.53	0.42	0.35	0.58	0.38
E2	0.15	0.56	0.75	0.83	0.29	0.80
E3	0.71	0.41	0.47	0.29	0.25	0.23
E4	0.59	0.59	0.37	0.24	0.00	0.00
E5	0.65	0.65	0.42	0.35	0.17	0.23
E6	0.06	0.33	0.18	0.57	0.67	0.67
E7	0.93	0.86	0.53	0.41	0.21	0.32
E8	0.18	0.18	0.12	0.28	0.21	0.27
E9	0.40	0.33	0.33	0.40	0.33	0.33
E10	0.35	0.25	0.20	0.10	0.15	0.05

表 3: 人手による評価結果

	A	α	β	B	γ	δ
E1	0.60	0.80	0.70	0.10	0.40	0.20
E2	0.80	0.60	0.80	0.10	0.80	0.60
E3	0.90	0.70	0.50	0.40	0.70	0.70
E4	0.90	0.90	0.90	0.10	0.90	0.60
E5	0.80	0.80	0.80	0.70	0.80	0.70
E6	0.80	0.80	0.80	0.70	0.80	0.80
E7	0.90	0.90	0.90	0.80	0.90	0.90
E8	0.90	0.95	0.95	0.90	0.95	0.90
E9	0.70	0.60	0.30	0.60	0.70	0.60
E10	0.80	0.40	0.30	0.30	0.80	0.70

E10 接続表現が十分に網羅されていなかったため、システムによる評価が低い評価になったと考えられる。

全体として、主題抽出が適当に行なわれないと、題述の抽出や主題連鎖関係の抽出に影響を及ぼし、結果的に評価結果に影響するため、文章評価において主題抽出精度が重要であることがわかった。

6 おわりに

本論文では文構造のモデル化、また、文間関係に基づいてそれらをリンクし文章を構造化する手法、さらに、様々な要素から文章を自動評価する手法を提案した。

自動評価の結果としては、人間の評価と一致する評価項目もあり、今後性能の向上が期待できる。主張文判定や省略された主題の抽出、照応詞の照合を実現して主題抽出精度の向上を図る必要がある。また、語彙連鎖や主題連鎖、接続関係などの文間関係を使い分けた評価パラメータや、使用する連鎖関係の種類を拡張など、様々な改善点も今後の課題として挙げられる。

表 4: 相関係数

E1	E2	E3	E4	E5
0.50	-0.60	0.50	0.31	0.46
E6	E7	E8	E9	E10
-0.29	0.22	-0.66	0.35	0.15

参考文献

- [1] Kukich K. Wolff S. Lu C. Chodorow M. Braden-Harder L. Harris M.D Burstein, J. Automated scoring using a hybrid feature identification technique. *In the Proceedings of the Annual Meeting of the Association of Computational Linguistics, Montreal, Canada*, 1998.
- [2] 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.
- [3] E.B Page. The immience of grading essays by computer. *Phi DeltaKappen*, pp. 238-243, 1966.
- [4] 関洋平, David Kirk Evans, Hsin-Hsi Chen, Lun-Wei Ku, 神門典子. 意見分析タスク - 多言語テキストを対象とした意見抽出技術の評価 -. 情報処理学会 自然言語処理学会, Vol. 183, pp. 51-58, 2008.
- [5] 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.
- [6] 永野賢. 文章論総説 - 文法論的考察 -. 浅倉書店, 1986.
- [7] 奈良先端科学技術大学院大学 松本研究室. 形態素解析システム茶釜 version 2.3.3, 2001.
- [8] 奈良先端科学技術大学院大学 松本研究室. 日本語係り受け解析システム南瓜 version 0.52, 2001.
- [9] 藤森芳郎, 吉川裕昭. 小論文・作文 書き方と模範文例. 新星出版社, 1980.