

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†
Akira Fujita

田村直良‡
Naoyoshi Tamura

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の中でも、小論文の自動評価を対象とする。小論文とは、与えられた問題文に対して回答者が意見や考えを述べる論説文である。

小論文自動評価の先行研究は[1]のサーベイが詳しい。英文の自動評価においては、PEG[2]に始まり、e-rater[3]などのシステムが代表的である。特に e-rater は専門家の採点結果との一致度が 97%という精度で自動評価を行う。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[4]。しかし和文の自動評価は、表層的な評価が主で論理構造の評価に関して多角的な評価が十分とは言えない。

本研究では、日本語文章を対象として、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、論理展開の局所的な特徴である文の接続、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、これを文章の構造上の特徴量により定義し、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、まず文のタイプを2種類に分類し、以下のように定義する。

- **意見文** 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

これら意見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主題や文末モダリティなどの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

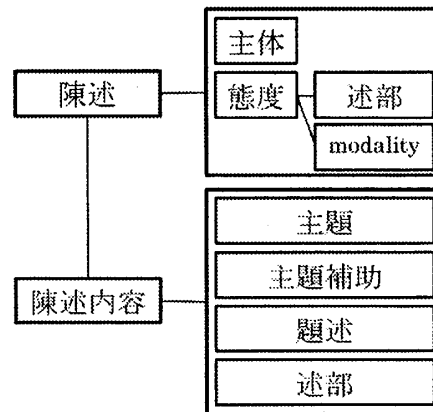


図1: 文構造

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 陳述行為と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 文の中で何について述べられているかを示す部分[5]

主題補助 主題を修飾する語

題述 主題以外の語

述部 様相を含む述部

意見文では、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するが、叙述文では、陳述内容部に陳述内容について記述する（陳述部は主体が筆者、態度/modality が叙述である場合が普通であるので省略とする）。

2.1.2 文間関係

文章中の語彙連鎖関係、主題の連鎖関係、および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3種とする。

- **主題維持** ある文の主題またはそれに類似する語

† 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

が直後の文の主題になっているか、または隣接する2文の主題のうち後文の主題が省略されている2文間の関係

- **主題変化** ある文の題述またはそれに類似する語が直後の文の主題になっている2文間の関係
- **主題回復** ある文の主題またはそれに類似する語が後の文(直後の文を除く)の主題になっている2文間の関係

接続関係は、修辞構造理論[6]を基に、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種類とする。

図2に主題連鎖関係の例を示す。

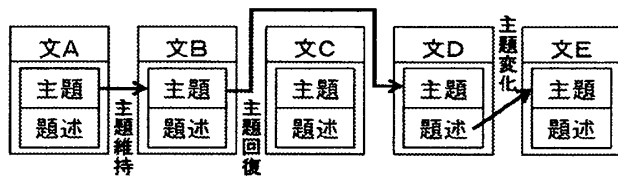


図2：主題連鎖関係の例

3. 文章構造モデルの各要素の抽出方法

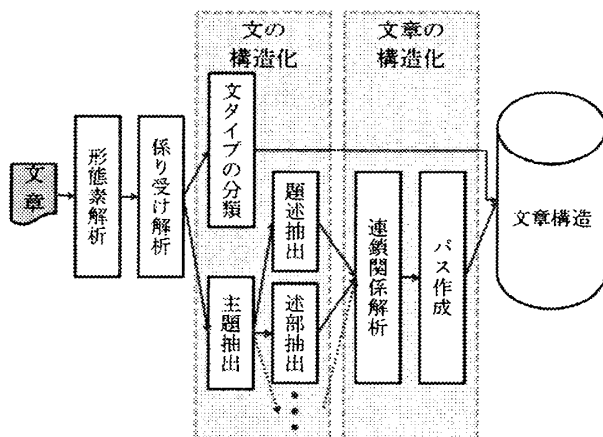


図3：文章構造解析の手順

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、係り受け解析処理には Cabocha を用いている。その他の処理には、独自のプログラムを用いている。

3.1 意見文と叙述文の自動分類

文末の表現に着目して、意見文と叙述文の分類を行なう。意見文の判定は、NTCIR をはじめ所々で取り上げられているが[7]、本研究では単純化のため、文の文末表現により意見文と叙述文の分類をする。この際、意見モダリティを意見、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。ただし、曖昧性を含むものに対しては全て叙述に分類している。

3.2 文構造の各要素の抽出方法

3.2.1 陳述部

以下で文構造の各要素のうち、陳述部に含まれる要素の抽出方法について述べる。

- **主体** 文が意見文と分類されている場合、主体を筆者とする。
- **態度/述部** 文が意見文と分類されている場合、態度/述部をモダリティを含んだ文末表現とする。
- **態度/modaliy** 文が意見文と分類されている場合、態度/modality を、意見文と叙述文の自動分類を行った際に分類したモダリティとする。

3.2.2 陳述内容部

以下で文構造の各要素のうち、陳述内容部に含まれる要素の抽出方法について述べる。

- **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[8]による。
- **主題補助** 主題を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、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

抽出された各文の主題と題述の情報と、新たに抽出する接続詞の情報を基に、語彙連鎖関係と主題連鎖関係と接続関係の解析を行う。

3.3.1 語彙連鎖関係

ある2文間の題述に5組以上の語彙連鎖関係がある場合は、その2文間を文間関係リンクで結ぶ。「5組の語彙連鎖関係」という文間関係の閾値は経験的に得たものである。

3.3.2 主題連鎖関係

まず、ある文(以下、該当文)の主題に着目し、一文ずつ開始方向へ調べていく。該当文の主題と調べている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文間を**主題回復リンク**で結ぶ。ただし、その2文が隣接している場合は、主題回復リンクではなく**主題維持リンク**で結ぶ。また、同様に調べている文の題述と同一か類似関係にあった場合は、2文間を**主題変化リンク**で結ぶ。また、該当文の主題が省略されている場合は、直前の文との間を**主題維持リンク**で結ぶ。

3.3.3 接続関係

文間の関係は、本来接続詞の有無に関係なく存在するのであるが、文章中の接続詞を手がかりに、文間の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、2文間を**接続関係リンク**で結ぶ。

3.3.4 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. 評価観点

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、「話題の発展性」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。

意見文の割合 $A =$ 意見文の個数、 $S =$ 文の数

$$E_1 = \frac{A}{S}$$

主張の正当性 $D_i =$ 意見文に至るパスの距離

$$E_2 = \frac{1}{A} \times \sum_{i=1}^A D_i$$

話題の発展性 $L_{max} =$ 意見文間のパスの最長値

$$E_3 = \frac{L_{max}}{S}$$

結論との結束性 $L_c =$ 結論段落とリンクがある文の個数

$$E_4 = \frac{L_c}{S}$$

文章全体の結束性 $L =$ リンクの総数

$$E_5 = \frac{L}{S(S-1)}$$

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

$$E_6 = \frac{L_p}{P}$$

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

$$E_7 = \frac{1}{P} \times \sum_{n=1}^P \frac{L_n}{S_n}$$

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

$$E_8 = \frac{N_s}{N}$$

文章構成の多様性 $v_s =$ 文章中に出現する態度/modalityの種類の数、 $V_s =$ 本研究で定義した態度/modalityの種類数の総数(=17)

$$E_9 = \frac{v_s}{V_s}$$

文章展開の多様性 $v_l =$ 文章中に出現する文間関係の種類の数、 $V_l =$ 本研究で定義した文間関係の種類数の総数(=24)

$$E_{10} = \frac{v_l}{V_l}$$

5. 実験と考察

5.1 意見文判定

システムの評価に先立って、解析過程で用いた意見文判定について、その判定結果を評価しておく。

評価実験で用いる6編の文章(合計95文)に対して人手で意見文判定を行い、その結果を正解データとする。

システムの意見文判定精度を、表1に示す。

表1: 意見文判定精度

適合率	再現率	F値
0.833	0.816	0.824

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。また、 α と γ には模範文で用いられている重要語句を出現順に示した語群(順序付語群)を順序通りに用いるという条件が与えられ、 β と δ には順序付語群を無作為に並び変えた語群を任意に用いるという条件が与えられている。なお、被験者全体に対して、600字前後で記述すること、だ調で書くこと、突飛な表現を避けることが条件として与えられている。

5.2.1 手法と結果

自動評価システムと評価観点の具体的な算出方法を知らない人間の評価者2名と知っている人間の評価者1名にそれぞれ6編の文章を評価観点ごとに0から1の実数値で評価させ、3名の評価結果の平均値とシステムの評価結果とを比較した。システムと人間の評価結果について、観点ごとにシステムと人間との間での相関係数を算出した。二値的な判定の評価ではなく、実数値による判定の評価であるため、相関係数を用いて評価を行うこととした。結果を表2、3、4に示す。

5.2.2 考察

E_1 、 E_3 、 E_5 、 E_7 、 E_8 、 E_9 、 E_{10} は正の相関が生じており、人間の評価手法と近似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いる。特に E_3 (話題の発展性)は筆者の論理展開を文章構造から評価するものであり、所期の目標に達している。

しかし、 E_2 、 E_4 、 E_6 は人間の評価結果と近似しているとは言いがたい。

- E_2 (主張の正当性)

システムによる評価手法では筆者が用いる論理展開の手法が適切か否かを評価しておらず、主張が述べ

られるまでの論理展開の長さのみを評価する。一方、人手による評価では論理展開の手法が適切であるか否かも含めて、主張の正当性を評価する。このように両者の評価手法に相違があり、評価結果が近似しなかったと考えられる。

● E_4 (結論との結束性)

尾括式の文章を想定してパラメータを形式化したため、システムによる評価手法では頭括式の文章のような始めに結論が述べられる文章に関して評価できていない。一方、人手による評価では文章において結論が述べられる位置に関わらず評価することができる。実験に用いた文章は6編のうち、3編の文章が頭括式の文章である。したがって、評価結果が近似しなかったと考えられる。

● E_6 (段落間の関連性)

システムによる評価手法では段落の総数で正規化するため、段落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。一方、人手による評価では段落の総数を評価に影響させず、むしろ段落が持つ構造に重点を置いて評価する傾向がある。したがって、評価結果が近似しなかったと考えられる。

E_2 、 E_4 、 E_6 は、形式化の方法を再考する必要があると考えられる。特に、 E_2 (主張の正当性) の考察で挙げた論理展開手法の適切さを評価することは、他の評価観点の強化にもつながると考えられる。

また、 E_1 、 E_3 、 E_5 、 E_7 、 E_9 、 E_{10} は、文章構造解析の段階で主題の抽出精度や意見文の判定精度の向上を図ることで、さらに人間の評価と近似する必要があると考えられる。

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、様々な観点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、人間の評価と近似する評価項目もあり、今後性能の向上が期待できる。意見文判定精度の向上を図ることや、省略された主題の抽出、照応詞の照合、複文への対応を実現することにより主題抽出精度の向上を図ることが必要である。

また、文間関係を使い分けた評価観点や論理展開手法の適切さを評価する評価観点の実現や、評価実験で使用する文章の数を増やすことなど、様々な改善点も今後の課題として挙げられる。

表2: 人間とシステムによる評価結果間の相関係数

E_1	E_2	E_3	E_4	E_5
0.73	-0.31	0.83	-0.89	0.34
E_6	E_7	E_8	E_9	E_{10}
-0.87	0.86	0.79	0.59	0.21

表3: 観点ごとの人間による評価結果(0~1.00)

	文章 A	文章 α	文章 β	文章 B	文章 γ	文章 δ
E_1	0.55	0.65	0.70	0.60	0.10	0.35
E_2	0.70	0.60	0.83	0.87	0.07	0.80
E_3	0.50	0.43	0.43	0.47	0.50	0.47
E_4	0.63	0.57	0.83	0.70	0.07	0.70
E_5	0.73	0.60	0.63	0.73	0.60	0.70
E_6	0.63	0.67	0.77	0.87	0.47	0.73
E_7	0.77	0.70	0.70	0.77	0.73	0.77
E_8	0.82	0.58	0.63	0.85	0.87	0.83
E_9	0.47	0.47	0.67	0.47	0.47	0.50
E_{10}	0.40	0.43	0.73	0.60	0.17	0.73

表4: 観点ごとのシステムによる評価結果(算出式の値)

	文章 A	文章 α	文章 β	文章 B	文章 γ	文章 δ
E_1	0.53	0.42	0.76	0.58	0.35	0.38
E_2	11.67	1.38	3.46	1.86	6.83	2.60
E_3	0.53	0.11	0.29	0.17	0.53	0.31
E_4	0.29	0.16	0.06	0.17	0.41	0.15
E_5	0.12	0.03	0.06	0.11	0.14	0.08
E_6	4.00	1.00	1.50	1.00	5.60	1.50
E_7	0.31	0.08	0.00	0.16	0.14	0.22
E_8	0.22	0.14	0.13	0.24	0.17	0.18
E_9	0.33	0.33	0.40	0.33	0.40	0.33
E_{10}	0.25	0.20	0.35	0.15	0.10	0.05

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]E.B Page. The immiense of grading essays by computer. *Phi Delta Kappan*, pp.238-243, 1966.
- [3]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*
- [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, pp.3-18, 2003.
- [5]長尾真. 自然言語処理. 岩波書店. 1996.
- [6]川本大輔, 田村直良. 論説文における論証構造の解析に関する研究. 横浜国立大学大学院. 2006.
- [7]関洋平, David Kirk Evans, Hsin-His 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 自然言語処理研究会, Vol.183, pp.51-58, 2008.
- [8]永野賢. 文章論総説-文法論的考察-. 浅倉書店, 1986.
- [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.