

# トップダウン・ボトムアップな文章構造作成のための 推敲支援システム

Text Writing Support System for Creating Top-Down/Bottom-Up Structure

山手 砂都美  
Satomi Yamate

砂山 渡  
Wataru Sunayama

広島市立大学大学院情報科学研究科  
Graduate School of Information Sciences, Hiroshima City University

We have many opportunities to write a text. However, relationship among paragraphs are hard to be grasped. Therefore, we focused on top-down structures and bottom-up structures between paragraphs. Top-down means texts that a paragraph including a conclusion or what the writer wants to say comes at first and bottom-up means texts that such a paragraph comes lastly. In this paper, a system that supports polish of text structure is proposed. Relationships between paragraphs are expressed as a tree structure and writers can confirm whether a text is top-down or bottom-up. Users of the system can polish their texts to be top-down or bottom-up by seeing output of the system.

## 1. はじめに

文章を書く機会はさまざまところである。文章を書いた本人が分かりやすく書いたつもりでも、他人が読むと分かりにくい場合、間違った解釈をしてしまう場合もある。自分の伝えたい意図を正しく伝えるためには、現状の文章の構造を把握するのはもちろんのこと、文章の構造を推敲する必要がある。しかし、自分で書いた文章の構造を正確に理解し推敲するのは難しく、自分で推敲するには限界がある。自分で簡単に文章構造の推敲が可能な環境があることが望まれる。

そこで本研究では、文章を段落間のつながりに着目した段落間関係の評価に基づく文章構造推敲支援システムの構築を目的とする。

## 2. 関連研究

### 2.1 段落間の関係を評価する研究

本節では、段落間の関係を評価する研究について述べる。

単語の概念関係を用いて段落の一貫性を解析する研究 [1][2] がある。これらの研究は、あらかじめ一貫した形で構成されていると考えられる技術文章を対象とし、単語間の意味類似度を用いて提案する段落一貫度が有効であるかどうかを検証したものである。本研究では単語間の意味には着目せず、条件付き確率で段落間の関係を評価していく。

### 2.2 文章の構造化に関する研究

本節では、文章の構造化に関する研究について述べる。

情報理論による、文章の理論的構造の解析をする研究 [3] がある。この研究では、文章の理論構造を分析して可視化を行っている。本研究では、理論構造の分析を行わず、条件付き確率の大きい段落間をつなぐことで構造化を行っている。

また、理工系学生を対象とした技術文書作成支援システムを作成した研究 [4] がある。この研究では、あらかじめ、論文のルールに沿って書かれていない文、意図が分かりにくい文を検出する機能があり、分かりにくい文をクラス図で可視化してい

る。文章の構造を可視化している点は同じだが、本研究では、段落間を対象とした構造を可視化している。

更に、以上の関連研究では、話の分岐として、話を広げている段落（トップダウン構造）、話をまとめている段落（ボトムアップ構造）に着目していない点で異なる。

### 2.3 文章の推敲に関する研究

本節では、文章の推敲に関する研究について述べる。

文の理論構造デザインツールの試作と校正・推敲支援ツールを作成した研究 [5][6] がある。これらの研究では、推敲は文章の理解向上を指しており、句読点の統一チェック、長文チェック等、細かい点に着目している。本研究では、推敲に関して細かい点に着目せず、文章の構造に着目している点が異なる。

## 3. 本研究で扱う文章構造

本章では、本研究で扱う文章構造について述べる。

本研究では、トップダウン構造とボトムアップ構造に着目する。トップダウン構造とは、結論や話の全体の構造に関する説明を先の段落で述べた後、その詳細を後の段落で述べている構造のことを指し、ボトムアップ構造は、トップダウン構造とは逆に、先の段落で部分的な話の詳細を述べ、結論や全体のまとめを後の段落で述べている構造のことを指す。

図 1 にトップダウン構造とボトムアップ構造の具体例を示し、図 2 に図 1 の各段落に使われている単語を示す。

まず、トップダウン構造に着目する。第 1 段落とつながっている第 2 段落では単語 A のみ共通に使われており、それぞれの段落の単語の使用割合は、第 1 段落では、3 単語中に 1 単語、第 2 段落では 7 段落中に 1 単語である。第 1 段落よりも第 2 段落では使用単語が増えていることから、第 1 段落から第 2 段落に単語 A について話が展開していると捉えることができ、矢印の向きを付けることが出来る。第 3 段落、第 4 段落にも着目すると、第 3 段落では単語 B について、第 4 段落では単語 C について話が展開していると捉えられる。

次に、ボトムアップ構造に着目する。第 5 段落とつながっている第 2 段落間をトップダウン構造と同様に考えると、単語 A を共通に使われ第 2 段落から第 5 段落へ単語 A について話がまとまっていると捉えることが出来る。同様に、矢印の向きを付けることが出来る。ただし、文章は小さい段落番

連絡先: 山手 砂都美, 砂山 渡, 広島市立大学大学院情報科学研究科システム工学専攻, 広島市安佐南区大塚東三丁目 4 番 1 号, {yamate, sunayama}@sys.info.hiroshima-cu.ac.jp

号から大きい段落番号へ話が進むため、この場合、第5段落から第2段落へ話が広がっているのではなく、第2段落から第5段落へ話がまとまっていると考える。

結論が先に書かれている場合、第5段落を取り除くとトップダウン構造の文章、逆に結論が最後に書かれている場合、第1段落を取り除くとボトムアップ構造の文章になる。

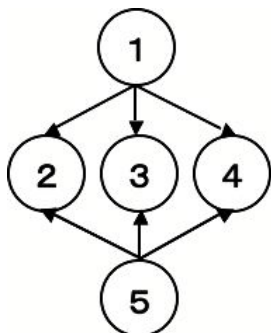


図 1: トップダウン構造とボトムアップ構造の例



図 2: 図 2 の各段落で使われている単語

#### 4. 段落間関係の評価に基づく文章構造推敲支援システム (SAT)

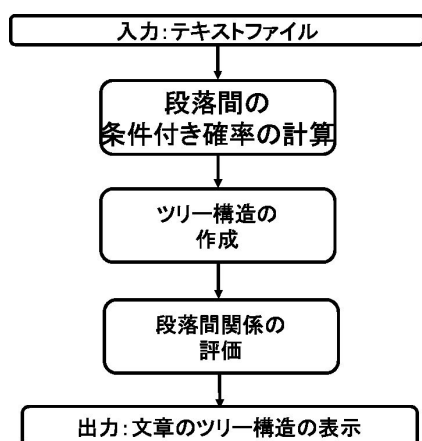


図 3: 段落間関係の評価に基づく文章構造推敲支援システム (SAT) の構成

##### 4.1 段落間関係の評価に基づく文章構造推敲支援システム (SAT) の構成

本節では段落間関係の評価に基づく文章構造推敲支援システム (SAT: Segment Association Tree) の構成について述べる。以下、SAT システムと呼ぶ。

図 3 に SAT システムの全体の構成を示す。SAT システムにテキストファイルを与え、それを元に段落間の条件付き確率を計算を行い、リンクをつなぐ段落間を決定する。次に、段落間の条件付き確率を元にリンクをつなぎ、文章のツリー構造を作成し段落間関係の評価を行う。

以下、各処理について述べる。

##### 4.2 段落間関係の評価

本節では、トップダウン構造とボトムアップ構造の2つで表される段落間関係の評価方法について述べる。

###### 4.2.1 段落間の条件付き確率の計算

本項では、段落間の条件付き確率の計算方法について述べる。以下の式 (1) で全ての段落間の条件付き確率  $Relation(A, B)$  を計算をする。段落  $A$  で使われている単語集合を  $W_A$ 、段落  $B$  で使われている単語集合を  $W_B$  とし、それらを数える関数  $n$  へ与える。段落  $A$  で使われている単語集合中、段落  $A$  と段落  $B$  で共通して使われている割合を求めることによって、段落間のリンクに向きをつけることが出来ると考えられる。

$$Relation(A, B) = \frac{n(W_A \cap W_B)}{n(A)} \quad (1)$$

###### 4.2.2 段落間関係の評価

本項では、段落関係の評価方法について述べる。

以下、後述する段落間の関係を表すツリー構造 (段落をノードとし、枝には条件付き確率  $Relation(A_i, A_j)$  が与えられる) が作成されたとき、 $n$  段落構成の文章のトップダウン構造の評価値を与える際のアルゴリズムを示す。

1. 各段落  $A_i$  について、枝に分かれている枝数  $Branch_i$  をカウントする。
2. 段落  $A_i$  の各枝  $j$  について、最も長い葉ノードまでの枝に与えられている、条件付き確率  $Relation(A_i, A_j)$  に +1 したものを加算し、 $BValue_{ij}$  とする。
3. 各段落  $A_i$  について、すべての  $BValue_{ij}$  を掛け合わせた  $NValue_i$  を求める。

またボトムアップ構造の評価値は、トップダウン構造の評価値の計算アルゴリズムと同様で、作成されるツリー構造の天地逆転してできるツリーのトップダウン構造の評価値と同じとする。段落間関係の評価値は、トップダウン構造の評価値とボトムアップ構造の評価値の対比を取って求めている。トップダウン構造の評価値を  $N_PValue$ 、ボトムアップ構造の評価値を  $N_BValue$  とすると、求める評価値  $N_{PB}Value$  は、式 (2) となる。トップダウン構造の評価値が高い文章構造は、最初の段落で簡潔に述べ、以降の段落で詳細について述べる形である。

$$N_{PB}Value = N_PValue / (N_PValue + N_BValue) \quad (2)$$

##### 4.3 文章構造の作成

本項では、文章構造を作成する方法について述べる。本研究で作成する文章構造は、 $n$  段落構成の文章について、各段落をノード、ノード間のリンクを条件付き確率  $Relation(A_i, A_j)$  として、最少の強いリンクのみで作成したものを指す。

###### 4.3.1 トップダウン構造の作成

本項は、トップダウン構造の作成について述べる。以下、トップダウン構造を作成するアルゴリズムを示す。

1. 全ての段落  $A_i, A_j (0 \leq i, j \leq n, i \neq j)$  間の条件付き確率  $Relation(A_i, A_j)$  を計算する。

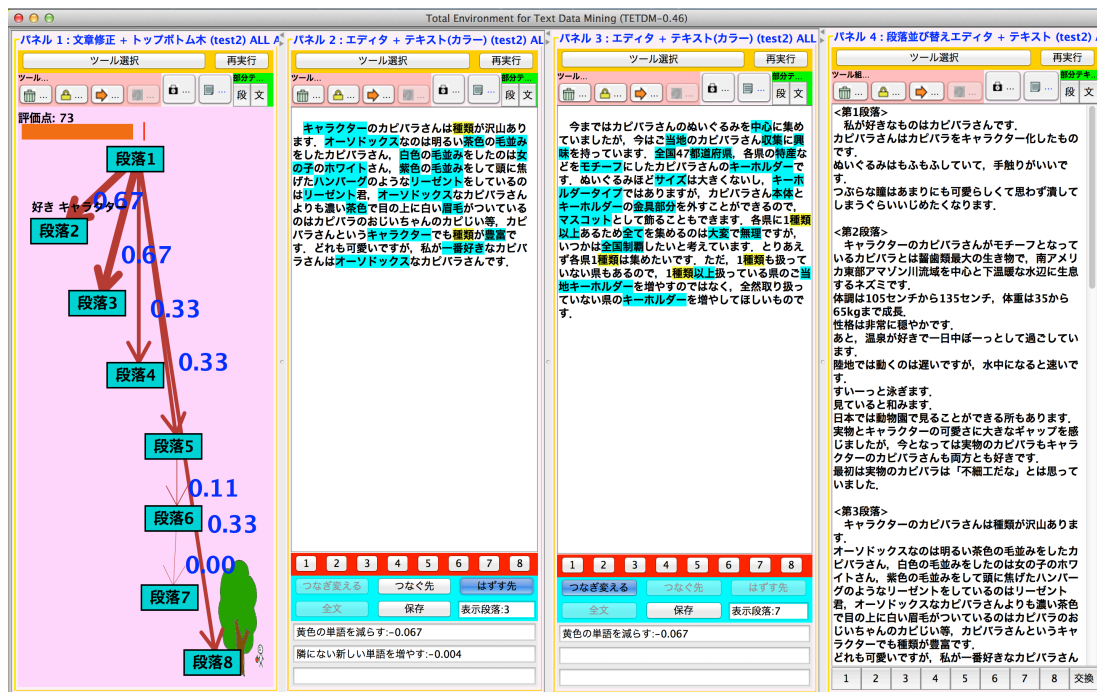


図 4: 文章構造推敲支援システムの使用例

2. 各段落  $A_j$  に着目し、下向きの矢印が入ってくる段落  $A_i (i < j)$  の中で、条件付き確率  $Relation(A_i, A_j)$  が最も高い段落にリンクを生成する。
3.  $n$  個の段落のすべてがリンクでつながっていれば（リンクによる段落間のパスが存在すれば）5.へ。そうでなければ、全段落間の条件付き確率  $Relation(A_i, A_j)$  を大きい順にソートする。
4. ソートされた条件付き確率  $Relation(A_i, A_j)$  の大きい順に、段落  $A_i$  と段落  $A_j$  との間にリンクを生成し、すべての段落がリンクでつながるまでリンクの生成を繰り返す。
5. 第1段落を根ノード（親ノード）とし、つながっているノードを順に子ノードとしたツリー構造を生成する。
6. リンク生成に用いた条件付き確率  $Relation(A_i, A_j)$  をもとの、リンクに  $A_i \rightarrow A_j$  の向きを与える。

下向きの矢印の条件付き確率  $Relation(A_i, A_j)$  に着目することによって、トップダウンの段落からつながっている段落元を探すことができる。

#### 4.3.2 ボトムアップ構造の作成

本項はボトムアップ構造の作成について述べる。

ボトムアップ構造のアルゴリズムは、4.3.1のトップダウン構造のアルゴリズムのステップ2が異なる。段落段落  $A_i$  に着目し、上向きの矢印が入ってくる段落  $A_j (i < j)$  の中で、条件付き確率  $Relation(A_i, A_j)$  が一番高い段落間にリンクを生成、他のステップは同様である。上向きの矢印の条件付き確率  $Relation(A_i, A_j)$  に着目することによって、ボトムアップ構造からつながっている段落元を探すことができる。

#### 4.4 出力：文章のツリー構造の表示

本節では、システムの出力：文章のツリー構造の表示について述べる。4.2段落間関係の評価で述べたものを TETDM[7]へ実装している。

SAT システムの出力例を図4の一番左パネルへ示す。文章は、著者の一人が書いた「私の好きなカピバラさんについて」を用いている。つながっている各段落間には、段落間のつながりと評価点、共通単語を表示する。段落間のつながりは矢印で表示され、つながりの強さが0から1の値で表示される。共通単語は、単語が表示されている段落より下にある全ての段落およびひとつ上の段落に共通して出現する単語を表す。評価点は100点満点で、文章がトップダウン構造の評価値が高いほど、値が高くなる。なお、このパネルで、段落をドラッグすると、段落をつなぎ替えることができる。

##### 4.4.1 文章構造推敲支援システム (SAT) の使用の流れ

本項は、文章構造をトップダウン構造に修正したい場合のSATシステムの使用の流れについて述べる。

1. 段落の並び替えを行う
2. 段落同士のつながりを変更する

以下の項で各ステップに使用するインタフェースについて説明する。

##### 4.4.2 段落の並び替え可能のインタフェース

本項は、段落の並び替え可能のインタフェースについて述べる。

図4の一番右パネルに段落並び替え可能のインタフェースを表示する。右側のパネルは文章を示しており、背景が赤いボタンに段落番号を表示している。並び替えを行いたい2つの段落を選んで「交換」ボタンを押すことによって段落を入れ替えることが出来る。同時に、一番左のパネルの文章構造と評価が変化し、適宜、段落の順番を並び替えてツリー構造と評価値を確認することができる。

##### 4.4.3 段落の文章表示するインタフェース

本項は、段落の文章表示するインタフェースについて述べる。図4の真ん中パネルに段落の文章表示するインタフェースを表示する。一番左のパネルで段落をつなぎ替えた結果、背景

が水色のボタンの「つなぐ先／はずす先 (の段落)」「つなぎ変える (段落)」を押すと、それぞれの段落の文章が表示される。また、背景が赤いボタンの番号を押すと、すでにつながっている、または任意の段落の文章を表示することも出来る。各段落内の文章の表示を行うと、表示している2つの段落に共に出現する単語 (名詞) が黄色、片方の段落にしか出現しない単語が水色、で表示される。

真ん中パネルの下部に、それぞれ、つなぎ変えるための修正アドバイスと修正によるつながりの強さの変化量が表示される。これらを参考に共通単語を増やす等をして文章の修正を行う。修正後に「保存」ボタンを押すと、ボタンを押したパネルの修正内容が他のパネルに反映される。

## 5. 評価実験

### 5.1 実験内容

本システムを使用することによって、文章構造推敲の支援を行えるか実験を行った。情報科学を専攻する大学生・大学院生8名に、「私が好きな○○について」というタイトルで、自分の好きなものの説明、ならびに好きな理由を含めて、約2000字で作文を書いてもらった。ただし、8段落前後で各段落は150字以上または5文以上を含むように書いてもらった。

次に、本システムを使用して、書いてもらった文章を要点や結論を後回しにするのではなく、先に述べる形式の文章に修正してもらった。システム上に表示される評価点が80点を越えると終了とした。修正は、以下の手順を繰り返すことで行った。

1. 段落の並び替えを行う
2. 段落同士のつながりを変更する

実験終了後には、以下のアンケートに答えてもらった。アンケートの各項目は5段階評価とし、各項目について理由も答えてもらった。

1. 最終的に完成した文章は、最初の文章に比べて読みやすくなったか
2. このシステムは、読みやすい文章の作成に役立てられるか
3. システムの各機能や表示は、文章の修正に役に立つか
  - (a) 段落間のつながり (矢印や評価値、共通単語) を表示して、段落間のつなぎ替えをアシストする機能
  - (b) 段落の並び替えを行う機能
  - (c) 段落間のつながりの強さを変える修正をアシストする (修正のアドバイスや、段落間の共通の単語や異なる単語を明示する) 機能
4. システムは全体的に使いやすかったか

### 5.2 実験結果と考察

被験者に答えてもらったアンケートの結果を図5に示す。縦軸は評価値、横軸は5.1に掲載したアンケートの項目番号を示す。

被験者に答えてもらったアンケートの結果の中で、3の(c)の「段落間のつながりの強さを変える修正をアシストする機能」は役立ったと感じた被験者が多かったことが分かった。つながりの強さを変える際に、どの単語を増やせばよいか、減らせばよいか、これらのアシストがあることによって、修正が容易だったと考えられる。この機能があることによって、文章構造の推敲を促すことができたと言える。

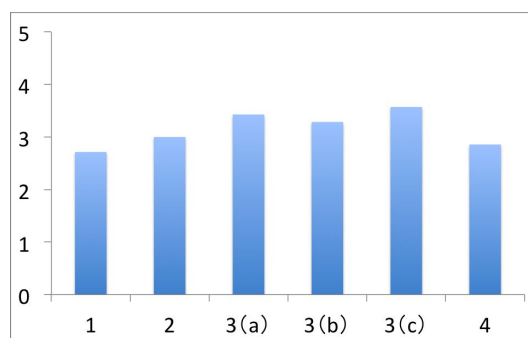


図5: 被験者のアンケートの回答

また、アンケート項目1に関して、完成した文章は、最初に比べて読みやすくなっていない、と答えた被験者の方が多かったことが分かった。これは、被験者が実際の文章の内容とつながりを意識せず、とにかく評価点が高くなる修正を行ったこと、また、画面上に現れないボトムアップ構造の点数が評価点に加えられていたため、被験者が思うように点数を上げられなかったことが原因と考えられる。今後、これらの点を踏まえた改善を行い、再度実験を行うことを検討している。

## 6. 結論

トップダウン・ボトムアップな文章構造作成のための推敲支援システムを提案した。今回の実験で、評価値の求め方に問題があることが分かったため、今後は、評価値の求め方を変更し、もう一度実験を行うことを検討している。

## 参考文献

- [1] 板倉由知, 白井治彦, 黒岩丈介, 小高知宏, 小倉久和: 単語の概念関係を用いた段落一貫性評価指標の有効性, 情報処理学会研究報告, NL-183, pp.107-113, (2008)
- [2] 板倉由知, 白井治彦, 黒岩丈介, 小高知宏, 小倉久和: 様々な文書を対象とした段落一貫性の解析: 情報処理学会研究報告, NL-192, pp.1-6, (2009)
- [3] 奥出 信一郎: 情報理論による、文章の理論的構造の解析, 電子情報通信学会技術研究報告, pp.1-5, (2008)
- [4] 松本章代, 山田未央佳, 山田翔, 鈴木雅人: 理工系学生を対象とした技術文書作成支援システム, 情報処理学会研究報告, CE-98, pp.91-96, (2009)
- [5] 大野博之, 稲積宏誠: 文の構造デザインツールの試作と校正・推敲支援ツールとの連携, 電子情報通信学会技術研究報告, ET, 教育工学 108(146), p.73-78, (2008)
- [6] 奥村有希, 大野博之, 稲積宏誠: 技術文章作成支援ツールの推敲支援機能の拡張—長い修飾節に起因する悪文の検出手法の提案 (次世代情報教育の構築に向けて/一般), 電子情報通信学会技術研究報告, ET, 教育工学 107(536), (2008)
- [7] 砂山渡, 高間康史, ダヌシカ ボレガラ, 西原陽子, 徳永秀和, 串間宗夫, 松下光範: テキストデータマイニングのための統合環境, 人工知能学会論文誌, Vol.26, No4, p483-493, (2011)