

文章構造を取り入れた文間類似度グラフに基づいた 一貫性に関する評価指標の提案

Proposal of Metrics Measuring Coherence
Based on Sentence Similarity Graph with Document Structure

庵 愛^{*1} 富永 敦子^{*2} 竹川 佳成^{*2} 平田 圭二^{*2}
Mana Ihori Atsuko Tominaga Yoshinari Takegawa Keiji Hirata

^{*1}公立はこだて未来大学大学院 ^{*2}公立はこだて未来大学
Future University Hakodate Graduate School Future University Hakodate

Coherence of the document is having unity of semantic, further It is necessary to write a document that is easy to understand. It is difficult to evaluate coherence of long document, since demand of automatic evaluation of coherence is high. Also, previous method focused on adjacent sentences to evaluate coherence. However, it is necessary to consider document structure, since document consists of multiple segments. Therefore, in this research, we propose metrics measuring coherence based on sentence similarity graph with document structure. We performed comparative experiments with previous method. As a result, some output accuracies comparable to previous method, it is assumed that proposed method is valid to evaluate coherence.

1. はじめに

一貫性とは、文章の意味的なまとまりの良さであり [田窪 04], わかりやすい文章の執筆に必要なものである。しかし、一般に文章が長くなるほど人手での一貫性の評価は難しいため、一貫性の自動評価における必要性は高い。

ここで、文章には章や段落といった粒度の異なる構造が多数存在する。例えば、いくつかの文が部分的にまとまって段落を成し、小さい段落が互いに結合し次第に章などの大きな構造を成す。これによって、最終的に 1 つの文章を成している。砂山らは、理解しやすい文章とは前後の構造間の関係が明確であり、全体を通して一貫した話の流れが存在すると述べている [砂山 06]。そのため、文章全体における一貫性を評価するためには、前後する 2 文間における一貫性から段落、節、章など構造間の一貫性の情報を含む必要がある。

本研究の目的は、論文執筆者によって書かれた文章の一貫性に関する自動評価である。本研究で対象とする論文における文章は、文章構造が顕著に表れているという特徴がある。本稿では、文章の構造を取り入れた一貫性に関する評価指標を提案し、その評価指標について従来手法との比較実験および考察を行う。

2. グラフに基づく一貫性モデル

Barzilay らは、前後する 2 文間における局所的な一貫性に関して、Entity grid モデルを提案した [Barzilay 05]。Entity grid とは、文章中における名詞句などの談話要素の分布パターンを捉える表現方法である。Entity grid の構成要素は各談話要素の構文役割であり、例えば以下の文章では表 1 のように表現される。

- S1 : エクセルには文字の置換機能があります。
S2 : 置換では文字を置き換えることができます。
S3 : しかし、「すべて置換」には注意が必要です。

Entity grid モデルは、Entity grid を用いて文章中出现する談話要素がどのように遷移するかを捉えることで一貫性を連絡先: 庵 愛, 公立はこだて未来大学, 〒041-8655 北海道函館市亀田中野町 116-2, 0138-34-6462, g2117008@fun.ac.jp

表 1: Entity grid

	エクセル	文字	置換	機能	注意
S ₁	S	X	O	O	-
S ₂	-	X	S	-	-
S ₃	-	-	S	-	O

S = 主語, O = 目的語, X = その他, - = 出現なし

評価している。このモデルは高い評価性能を算出したが、複雑な計算が必要であることや前後する 2 文間の一貫性しか評価できないといった問題があった。

そこで、Guinaudeau らは Entity grid のグラフ表現に基づくモデルを提案した [Guinaudeau 13]。この手法では、構文役割が数値（主語：3，目的語：2，その他：1，出現なし：0）で表された Entity Grid を用いてグラフを作成し、そのグラフの特徴から一貫性を評価している。

以下に Guinaudeau らのモデルの実装方法を示す。表 1 から、要素を共有する文同士のみを用いて図 1 の従来手法のようなグラフを作成する。そのグラフから、枝の重みを利用することで一貫性の評価値を算出する。ここで、枝の重みづけは以下の 3 手法が提案されている；(1) Unweighted: 共有の要素が存在する場合に重みを 1 とする, (2) Weighted: 共有の要素の数を重みとする, (3) Acc: 共有の要素の構文役割における数値を掛け合わせたものを重みとする。最終的な重みは、隣り合わない文同士の重要度を下げるため、各値と文同士の距離との除算によって得られる。文章 D における最終的な一貫性の値は、以下の通りに計算される (N=文章を構成する文の総和, S_i=i 番目の文における枝の重み)。この値が高いほど、一貫性も高くなる。

$$LocalCoherence(D) = \frac{1}{N} \sum_{i=1}^N OutDegree(S_i)$$

Guinaudeau らのモデルでは、より簡単な計算で従来の Entity grid モデルに匹敵する性能を算出し、非隣接の 2 文間における一貫性を考慮することが可能になった。しかし、Guinaudeau らのモデルでは単純な文同士の繋がりにしか着目して

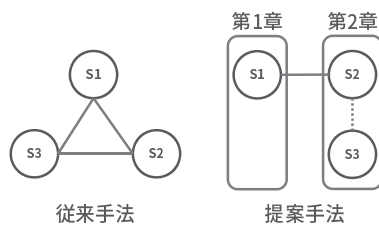


図 1: 従来手法と提案手法のグラフ表現

ならず、上記で述べた構造間の一貫性は考慮されていない。

3. 提案手法

3.1 文間類似度グラフ

本研究で用いる文間類似度グラフとは、文間のコサイン類似度に基づいた連結性行列を隣接行列として作成するものである。このグラフは複数文書要約手法などに用いられており、文書の全体像をまとめることに適している [Erkan 04]。本研究では、tf 値によって構成された文ベクトルを用いて、以下の式でコサイン類似度を計算する。

$$\text{sim}(v_x, v_y) = \frac{\sum_{i=1}^n v_{xi} \cdot v_{yi}}{\sqrt{\sum v_{xi}^2} \cdot \sqrt{\sum v_{yi}^2}} \quad (1)$$

ここで、 v_x , v_y は文ベクトルを、 n は文章中出现するすべての単語数を、 v_{xi} , v_{yi} は i 番目の単語における tf 値を表している。文ベクトルは 1 文という情報量の少ない単位から構成されるため、正確に算出できる値として tf 値を用いている。本研究では、文章における文に対して式 (1) を用いて類似度を計算し、類似度が 0.1 以上になった文のみを用いてグラフを作成する。Guinaudeau らのモデルでは、談話要素が 1 つでも共有されているとグラフを作成していたため、本研究でもグラフが作成される条件を低く設定するために基準値を 0.1 とした。

ここで、Guinaudeau らのモデルでは取り入れられていなかった、文章構造の概念をグラフ作成時に用いる。S1 が第 1 章、S2, S3 が第 2 章に属する文であると仮定すると、図 1 の提案手法のようなグラフが作成される。従来手法では、文がどの章、段落といった構造に含まれているかが考慮されていなかったが、提案手法では、章や段落などの構造で区切られたグラフを作成することで、文章構造をグラフに取り入れる。

また、S2 と S3 は同じ第 2 章に属する文だが、同じ章でも違う段落に属している可能性がある。そのため、上位構造で同じ構造に含まれる文では、直下の構造に粒度を落として着目し、異なる構造に属するようになるまで構造の粒度を落としてグラフを作成する (図 2)。つまり、類似度が 0.1 以上の文同士が存在する限り、文章が持つ構造の数だけグラフが作成されることになる。

3.2 グラフごとの一貫性指標

前節で作成したグラフを用いて、グラフごとの一貫性を算出する。グラフ中の一貫性を表す特徴量として、我々は、構造網羅割合、枝の重み、構造同士の距離、枝数に着目した。構造網羅割合とは、グラフが直下の構造をどの程度網羅しているかを表した値である。例えば、図 1 における提案手法のグラフが 5 章構成の文章から作成されたとすると、出現する章の数

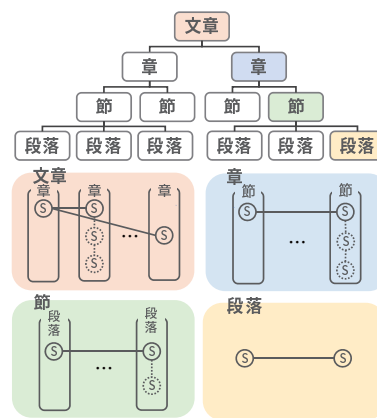


図 2: グラフの作成方法

は 2 である。従って、このグラフでの構造網羅割合は、 $2/5$ で 0.4 となる。多くの構造を網羅するほど一貫性が高くなると考えられるため、この値が 1 に近づくほど評価値は高くなる。

次に、類似度グラフを用いているため、枝の重みは文同士の類似度とする。この重みは、先行研究が文同士の距離との除法によって算出されるのに対して、本研究では構造同士の距離との除法によって算出される。例えば、図 2 における節で作成されたグラフに着目すると、2 つの段落をまたぐグラフが形成されている。ここで、段落が第 3 段落と第 7 段落だとすると、構造同士の距離は $7-3$ で 4 となる。以上の 2 つの特徴量は、文同士の類似度が高く、近くの構造に属しているほど評価値は高くなる。ここで、類似している文の総数はグラフごとに異なるため、グラフごとの一貫性の評価値を等しく見るためにグラフの総枝数で正規化する。以上の特徴量を踏まえ、 n 個の枝を持つグラフ S の一貫性の値は以下の式で計算する。

$$\text{LocalCoherence}(S) =$$

$$\text{構造網羅割合} \times \frac{1}{n} \sum_{i=1}^n \frac{\text{類似度}}{\text{構造同士の距離}} \quad (2)$$

3.3 文章全体における一貫性指標

式 (2) では、構造ごとに作成されるグラフにおける一貫性の値が計算される。文章が複数の粒度の異なる構造から成り立つように、各構造ごとの一貫性の集合によって文章全体の一貫性が算出されると考える。我々は、グラフごとの一貫性を文章構造における特徴を踏まえて値を集約する方法を提案する。

図 2 に示すように、文章構造は章の下に節があり、節の下に項がある構造になっていることが多い。そのため、文章全体の一貫性を算出する際には下位構造の値の平均を掛け合わせていく。例えば、図 2 中の節は、3 つの段落を持ち合わせている。そのため、節の一貫性の値は、もともとの節の一貫性の値に、3 つの段落の一貫性の平均値を掛け合わせたものとなる。我々は、下位構造の一貫性の値が 0 であった場合を考慮するため、提案式に乗法を用いた。ここで、本研究における文章の構造は、章の下に節、節の下に項、項の下に段落、段落の下に文があることを基本と考える。

以下に、下位構造を考慮した一貫性を算出する式を示す。ここで、 S_u は現在着目している構造を、 S_d はその構造が持つ m 個の下位構造を表している。この計算を文章レベルまで繰り返して計算することで、文書全体にける一貫性の値を算出で

きる。

$$Coherence(D) =$$

$$LocalCoherence(S_u) \times \frac{\sum_{i=1}^m LocalCoherence(S_d)}{m} \quad (3)$$

3.4 上位/下位構造を重視するパラメータの付与

本研究で提案する評価指標では、上位/下位構造における評価値の平均を掛け合わせることで文章全体の評価値を計算する。そのため、上位構造と下位構造それぞれに重視する構造を選択可能なパラメータを設定することで、どちらかの値を重視した一貫性の値を算出できる。例えば、上位構造では文章全体における一貫性を、下位構造では文同士における一貫性を重視した一貫性の値を算出する。我々は、このパラメータを調節することで、重視する構造を選択できると考えらる。そこで、式(3)に以下の様なパラメータ α (0~1) を付与する。今回は、任意の値として重視する構造が上位の場合には 0.9 を、下位の場合には 0.1 を設定した。また、それぞれの構造を等しく重視する場合には、0.5 を設定した。

$$Coherence(D) =$$

$$LocalCoherence(S_u)^\alpha \times \left(\frac{\sum_{i=1}^m LocalCoherence(S_d)}{m} \right)^{1-\alpha} \quad (4)$$

4. 評価実験

以上の評価指標の性能を測るために従来手法との比較実験を行う。比較実験では、先行研究に従って、Discrimination [Barzilay 08] と Insertion [Elsner 11] の 2 種類の手法で人工的に一貫性の低い文章データを作成し、評価値を計算する。これらの手法では、元の文章データの文の順番を入れ替えることで一貫性の低い文章データを作成するため、文数や文章の構造が異なるデータの比較を防止できる。

まず、Discrimination とは元の文章における文の順番をランダムに入れ替えたデータを作成する手法である。本手法では、Discrimination によって作成された文章と元の文章において一貫性の評価値を算出し、元の文書における評価値が高い場合に評価指標の出力は正しいということになる。本研究では、先行研究に従って Discrimination を用いて 20 件の一貫性の低いデータを作成し、元の文書との一貫性の評価値を比較する。20 件のデータ中、元の文書より評価値が低くなったデータ数を正解データ数とし、F 値を算出することで従来手法との比較を行う。

次に、Insertion とはある 1 文を文章から抜き出し、抜き出した文の元の位置を検索する能力を評価する手法のことである。具体的には、文書中の各文を順番に抜き出し、可能性のあるすべての位置に挿入する。すべての位置に挿入するごとに、一貫性の評価値を算出する。本手法では、最も高い評価値が出力されたときの文の位置が元の位置と同じである場合、システム出力は正しいということになる。本研究では、実験時間の関係から各文ではなく、ランダムに選択した 20 文について、可能性のあるすべての位置に挿入する。ランダムに選択された 20 文のうち、最も高い評価値を出力したのが、元の位置と同じであった場合の割合を Insertion score (Ins) として算出することで従来手法との比較を行う。

また、本研究では文章の構造を考慮した一貫性指標を提案した。そのため、従来研究では Discrimination と Insertion を

表 2: 各論文における構造ごとの枝数の平均

	文章	章	節	項	段落
dataA	3522.00	737.57	313.43	43.50	206.57
dataB	3116.43	841.86	627.29	149.14	182.86

文のみを対象として行っていたが、ここでは章を対象としても同様の実験を行う。各評価も、同様の方法で実施する。

本研究では、ジャーナル論文を対象としてこれらの実験を行う。ジャーナル論文は、厳格な査読を経ているため一般的な論文よりも一貫性が高いことが期待される。本実験では、同じ著者によって書かれた教育学系のジャーナル論文 7 件、情報工学系のジャーナル論文 7 件を用いた。表 2 に、各ジャーナル論文から作成されるグラフの枝数を示す。なお、教育学系の論文データセットを dataA、情報工学系の論文データセットを dataB とする。ここで、本研究で作成されるグラフは、図 2 に示すように構造をまたいで形成されるものと、同じ構造内で形成されるものがある。各データセットに対してグラフの形成に異なる性質があるかを調査するために、各構造間、構造内における枝数の差について対応なしの t 検定を行った結果、文章と章において有意差が認められた ($t(12)=2.21$, $p<.05$, $t(12)=2.26$, $p<.05$)。そのため、dataA で形成されるグラフでは、主に上位構造において dataB よりも構造間の枝数が多いといえる。

5. 結果

まず、文を対象とした Discrimination と Insertion を用いた従来手法との比較実験の結果を表 3 に示す。Discrimination の結果について、従来手法では、各データのすべての場合において最も高い F 値を算出した。一方、提案手法では、各データにおいて $\alpha = 0.1$ と $\alpha = 0.5$ で比較手法に匹敵する F 値であったが、 $\alpha = 0.9$ で著しく値が低かった。Insertion の結果について、従来手法では、各データにおいて Weighted で最も高い Ins を算出した。一方、提案手法では、各データにおいて $\alpha = 0.1$ で高い Ins を算出したが、Weighted より低い値であった。そこで、Weighted と $\alpha = 0.1$ の Ins について、対応なしの t 検定を行った結果、有意な差は認められなかった ($t(12)=1.80$, $p=0.1$, $t(12)=1.80$, $p=0.1$)。つまり、文を対象とした入れ替えでは、提案手法のうち $\alpha = 0.1$ では従来手法に匹敵する精度で評価できるといえる。

次に、章を対象とした Discrimination と Insertion を用いた従来手法との比較実験の結果を表 4 に示す。Discrimination の結果について、従来手法では、各データにおいて Unweighted 以外で高い F 値を算出した。一方、提案手法では、各データにおいて $\alpha = 0.1$ で最も高い F 値を算出し、従来手法を上回った。Insertion の結果について、従来手法では、dataA については Acc で、dataB については Unweighted で高い Ins を出力した。一方、提案手法では、各データについて $\alpha = 0.9$ の場合に高い Ins を出力し、dataB においては従来手法を上回った。つまり、章を対象とした入れ替えでは、提案手法で従来手法を上回る精度で評価できるといえる。

6. 考察

まず、文を対象とした実験に関する考察を行う。実験の結果、Discrimination では、提案手法の $\alpha = 0.9$ は他と比べて著し

表 3: 文を対象とした実験の F 値と Ins の平均値

	F 値		Ins	
	dataA	dataB	dataA	dataB
Unweighted	1.0	1.0	0.086	0.036
Weighted	1.0	1.0	0.107	0.100
Acc	1.0	1.0	0.064	0.050
$\alpha = 0.1$	1.0	1.0	0.036	0.043
$\alpha = 0.5$	1.0	1.0	0.021	0.007
$\alpha = 0.9$	0.19	0.14	0.007	0

表 4: 章を対象とした実験の F 値と Ins の平均値

	F 値		Ins	
	dataA	dataB	dataA	dataB
Unweighted	0.71	0.71	0.51	0.69
Weighted	0.86	0.86	0.57	0.65
Acc	0.86	0.86	0.66	0.60
$\alpha = 0.1$	1.0	1.0	0.32	0.17
$\alpha = 0.5$	0.36	0.24	0.05	0.02
$\alpha = 0.9$	0.65	0.71	0.43	0.71

く F 値が低くなっていた。このことから、上位構造を重視することに問題があると考えられる。その原因として、文を対象とした Discrimination ではグラフの形成が大きく変化することが挙げられる。例えば、S1~S5 の 5 文を持つ文章のうち S1 と S3 を入れ替える実験を行う。ここで、元の文章におけるグラフと文を入れ替えた後におけるグラフを図 3 に示す。従来手法では文同士の距離しか変わらないのに対して、提案手法ではグラフの構造自体が変わってしまう。具体的には、元の文章では S2 と S3、S4 と S5 が同じ構造内でグラフを形成しているのに対して、入れ替えた文章では同じ構造内でグラフは形成されない。このように、提案手法では文を入れ替えることによって下位構造のグラフが消失する可能性がある。そのため、下位構造に関する特徴量が減り上位構造に吸収されるために、上位構造の評価値が高くなり、下位構造の評価値が低くなると考えられる。これより、下位構造の変化を重視した方が一貫性の変化も捉えやすいと考えられる。

次に、章を対象とした実験に関する考察を行う。実験の結果、Discrimination では、提案手法の $\alpha = 0.1$ の場合に従来手法を上回る F 値を出力した。このことから、文章構造を取り入れた本評価指標では、章などの構造の入れ替えに対する一貫性の評価に従来手法よりも適していると考えられる。また、Insertion を行った結果では、dataA と dataB で異なる傾向が見られた。dataA では、従来手法の Acc で最も高い Ins を出力したのに対して、dataB では提案手法の $\alpha = 0.9$ で最も高い Ins を出力した。この結果と、dataA と dataB でグラフの特徴が異なることから、本評価指標で評価するのに適したグラフがあると考えられる。さらに、 $\alpha = 0.5$ ではどちらの実験においても評価値が高くなかったことから、本評価指標では上位/下位構造のどちらかに重みを置いた方が高い性能で評価できると考えられる。今後、これらの要因を踏まえて評価指標について再検討を行う必要がある。

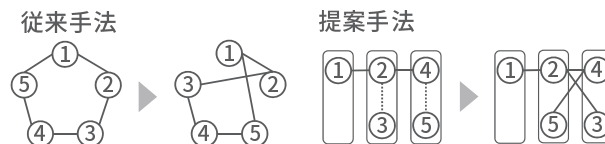


図 3: グラフの形成

7. おわりに

本研究は文章構造を考慮した一貫性の自動評価を目的に、文間類似度グラフに基づく評価指標を提案した。具体的には、文間の類似度を用いて章や段落といった文章の構造ごとにグラフから構造網羅割合、文同士の類似度、構造同士の距離、枝数を用いて一貫性を算出する式を作成し、章、節、項、段落といった粒度の異なる構造を考慮してすべての一貫性を掛け合わせる手法を提案した。その際に、上位構造、下位構造のどちらかを重視できるパラメータの設定も提案した。提案手法における評価性能の調査のため、従来手法との比較実験を行った。その結果、ある条件においては従来手法を上回る結果を出力した。

今後の課題として、入れ替える構造によっては、本評価指標に適していないデータも存在すると考えられるため、グラフの特性と評価指標の関係を分析する必要がある。また、今回は α の値を任意で決めていたが、最適な値を検討する実験を行うべきだと考える。

参考文献

- [Barzilay 05] R. Barzilay and M. Lapata. "Modeling local coherence: an entity-based approach," Computational Linguistics, pp. 141-148, 2005.
- [Barzilay 08] R. Barzilay and M. Lapata. "Modeling local coherence: An entity-based approach," Computational Linguistics, Vol. 34, No. 1, pp. 1-34, 2008.
- [Elsner 11] M. Elsner and E. Charniak. "Extending the entity grid with entity-specific features," In Proceedings of the ACL 2011 Conference Short Papers, pp. 125-129, 2011.
- [Erkan 04] G. Erkan and D. R. Radev, "LexRank: Graph-based Lexical Centrality as Salience in Text Summarization," Journal of Artificial Intelligence Research, Vol. 22, pp. 457-579, 2004.
- [Guinaudeau 13] C. Guinaudeau and M. Strube. "Graph-based Local Coherence Modeling," Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics, pp. 93-103, 2013.
- [砂山 06] 砂山渡, 橘啓八郎, "サブトピックモデルに基づく文章の流れの評価指標の提案," 日本知能情報ファジィ学会誌, Vol. 18, No. 2, pp. 280-289, 2006.
- [田窪 04] 田窪行則, 西山佑司, 三藤博, 亀山恵, 片桐恭弘. 談話と文脈 言語の科学 7 岩波書店, pp. 193-232, 2004.