2G3-04

複数の新聞記事から抽出した文の並び順の検討

An Investigation of Sentence Ordering Extracted from Multiple Newspaper Articles

岡崎 直観　　石塚 満
Naoaki Okazaki　Mitsuru Ishizuka

東京大学情報理工学系研究科
Graduate School of Information Science and Technology, The University of Tokyo

It is necessary to work out a nice arrangement of sentences extracted from multiple documents when we generate a well-organized summary. In this paper we describe our approach to coherent sentence ordering for summarizing newspaper articles. Since there is no guarantee that chronological ordering of extracted sentences, which is widely used by conventional summarization system, arranges each sentence behind presupposed information of the sentence, we improve chronological ordering by resolving antecedent sentences of arranged sentences. Combining the refinement algorithm with topical segmentation and chronological ordering, we address our experiment to test the effectiveness of the proposed method. The results reveal that the proposed method improves chronological sentence ordering.

1. はじめに

電子化文書やその検索技術の進歩により，欲しい情報を簡単に手に入れられるようになった．このような状況は非常に便利である反面，情報が多すぎて人間が手に入れた情報を活用できない情報過多として問題視されている．この問題に対し，文書自動要約 [Mani 01] は元の文書の代わりに縮約した文章を提供することで，氾濫する情報の活用を目指している．文書自動要約システムの多くは，元の文書の中から重要な箇所を統計的手法などを用いて推定し，重要と思われる箇所を文や段落を単位として抽出する．抜き出した文や段落は，文の並び順を決定したり，文中で不要な箇所を削除したり，逆に必要な表現を追加するなどの処理を行い，要約として纏める．文書の中から重要な箇所を抜き出す重要箇所抽出に関しては，自然言語処理研究の初期の段階から盛んに研究されてきたが [Luhn 58]，抽出した文を纏めて要約を作成する研究は，比較的数が少ない．要約文の並び順の決定方法としては，元文書中における並び順をそのまま採用する方法が用いられてきたが，要約対象が複数の文書 [McKeown 99] になった場合は，文書間の関係も同時に考慮しなければならないため，元の文書中における並び順だけで要約文の順序を決定することはできない．

そこで，本発表では複数の新聞記事から抜き出した文を並べ，要約文章を作成するための手法を提案する．提案手法は，時間の順序に基づいて並べられた文に対し，それぞれの文が前提としている情報が以前の文で述べられているかを調べ，時間順による文の並びを改善するものである．我々が行った実験では，提案手法によって時間序で並べられた文の並び順を改善する効果が認められた．

2. 文の並び順の決定方法

2.1 関連研究

我々の目標は幾つかの文が与えられたとき，その尤もらしい並び順を決定することである．人間は自分の頭の中にある考えを文章に書き下すように，このタスクを簡単に遂行することができる．しかし，計算機は物事の順序を理解しているわけではないので，このタスクについて検討する必要がある．修辞構造 [Mann 88] や整合関係 [Hobbs 90] に代表されるような談話理論は，この問いに役立つものである．Hume [Hume 48] は，物事の一貫性が生まれる根源として，類似性，時空間的連続性，因果関係の３つを挙げている．このことは，幾つかの文から文章を構成するには，話題の類似性，時間的なつながり，因果関係に着目すべきであるというヒントを与えている．

Barzilay ら [Barzilay 02] は，複数文書要約というタスクの中で，文の並び順の決定方法に関する問題を提起し，文の並び順が文章の読みやすさに与える影響の大きさを実験で示した．次に，大多数順 (majority ordering) *1 と時間順 (chronological ordering) *2 という２つの単純な手法を提案・評価したが，この２つの手法では満足のいく結果が得られなかったと報告している．これに対し，Barzilay らは元文書の中に含まれている話題を幾つか認識し，認識された話題の中で文を時間順に並べる方法を最終的に提案している．Lapata [Lapata 03] は，ある文がすでに並べられた文の次に来る確率を定式化し，その確率モデルに基づく文の並び順の決定方法を提案している．ある文が配置される確率を直前の文との二項関係のみで近似し，その確率を計算するために，動詞の順序性や名詞の同一性，文の構造などを用いている．

2.2 時間順の問題点

このような先行研究に対し，我々はそれぞれの文が元の文書の中でおかれていた環境に着目して文を並べる手法を提案する．例として，図 1 に示すようなクローン羊ドリーに関する３文の妥当な並び順を決定する方法について考える．ここで，図 1 にある３文 a, b, c を時間順に並べたとき，[a-b-c] という並び順が得られるとしよう．時間順の通りに文を並べるという従来の手法に従い，この３文を [a-b-c] という順序で読むと，文 b の位置が不適当であることに気付くであろう．これは文 b を読むために必要な前提知識である，「ドリーは子供を生んだ」という情報を読者に伝えていないのにもかかわらず，ドリーの子供の父親やドリーの妊娠に関する文 b を配置してしまったためである．ある文が前提としている情報を伝えないまま文を配置すると，その文が何について言及しているのか曖昧になり，読者を混乱に陥れてしまう．ここで文 c の内容に着目すると，文 c はまさに文 b の前提となる情報について言及しており，文 b は文 c の詳細化であると捉える方が自然である．

文 b の中途半端な情報から想像がつくかもしれないが，元文書で文 b が配置されていた箇所を調べると，文 b の前に幾

*1　元の文書の中でよく出現する順序に従って文を並べる手法
*2　記事が書かれた日時を使い，その順序の通りに文を並べる手法

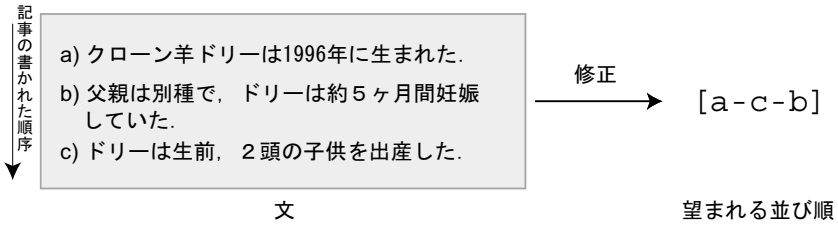

図 1: 時間順に基づく文の並び順では不十分な例.

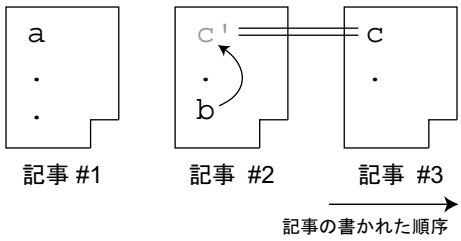

図 2: 並べる文とその先行文の関係.

つかの先行文が存在し，それらの先行文の内容を受けて文 b が配置されていた．図 2 では，文 b の前に文 c' なる文が存在していることを示している．もし，文 a が文 c' とほぼ同様の情報を伝達しているのであれば，文 b の前提情報は文 a で述べられているので，文 b を文 a の直後に配置しても問題ない．しかし，図 2 のように，文 c' の内容が文 a ではなく文 c とほぼ等価である場合，文 b の前提とする情報は文 a では伝達されないので，文 c を文 b の前に配置する方が望ましい．このように文 b の前に文 c を挿入し，最終的な並び順 [a-c-b] を得る．

2.3 提案手法による時間順の改善

図 3 は，a,b,...,f で示された 6 つの文の時間順序 [a-b-c-d-e-f] を提案手法によって改善する様子を示したものである．Start と End の間に存在するノードは文を示しており，最終的に実線で示されたエッジを辿り，[a-b-e-c-d-f] という並び順を得る．ここで，すでに配置された複数の文と，これから配置しようとしている文 x との「距離」を，すでに配置された複数の文と，文 x の先行文の内容との非類似性[*3]で定義する．例えば，ある文に先行文があり，その内容がすでに並べられた文の中で述べられていない場合は，この「距離」は大きくなる．逆に，ある記事の先頭に出現する文に対しては，この「距離」を 0 と定義する．

このような距離を導入すると，抽出された文の前提情報を考慮して文を並べる問題は，Start ノードから End ノードまでの距離が最短である経路を求める問題として定式化される．ただ，この経路探索問題は NP 困難であることと，時間的な順序性を必要以上に破壊することを避けるため，次のような方法で近似解を求める．まず，文の時間順序に従い，文 a を取り出す．文 a は元記事の中で最初に出現する文（リード文）であるとすると，上で定義した距離は 0 になるので，文 a をそのまま出力して並び順 [a] を作成する．次に，まだ並べていない残りの文の中から最も時間的順序が早い文 b を取り出し，やはりこの文もリード文であるとすると，そのまま文 b を出力して並び順を [a-b] に延長する．しかし，次に取り出した文 c には元記事の中で先行する文があるとすると，まず最初にその先行文の内容と文 a,b の内容の距離を計算する．図 3 に示した例では，文 a,b 直後に文 c を配置するための距離は 0.7 と計算されており，文 c の前提とする情報が文 a，b では十分に述べられていないことが分かる．そこで，文 a，b から文 c に至る短い経路を文 c から逆向きに最良優先探索で求める．図 3 では，文 c の前に文 e を配置すると，文 a,b から文 c まで距離 0.2 で到達できることが分かるので，文 c の前に e を挿入し，並び順を [a-b-e-c] に延長する．そして，まだ並べていない残りの文の中から最も時間順序が早い文 d を取り出すと，この文も先行文があるので文 a,b,e,c から文 d に至る距離を計算する．今回は文 a,b,e,c から文 d に直接至るパスの方が文 f を経由するよりも短いので，文 d をそのまま出力し，並び順を [a-b-e-c-d] に更新する．最後に，残りの文 f を出力し，並び順を [a-b-e-c-d-f] と確定する．

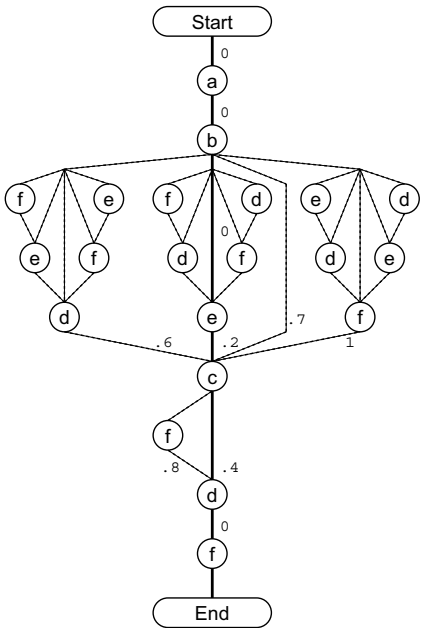

図 3: 時間順に並べられた文の並び順を改善する様子

2.4 文の並び順を決定するシステム

これまで述べてきた提案手法を用いて，文の並び順を決定するシステムの構成を図 4 に示した．ここでは，a,b,...,i で示される 9 つの文の並び順の決定手順を例に説明する．ま

*3 詳しい説明は紙面の都合で割愛するが，すでに配置された複数の文に含まれる単語から作成した単語ベクトルと，文 x の先行文に含まれる単語から作成した単語ベクトルの類似度を内積を用いて計算し，その値を 1 から減算することで距離とした．

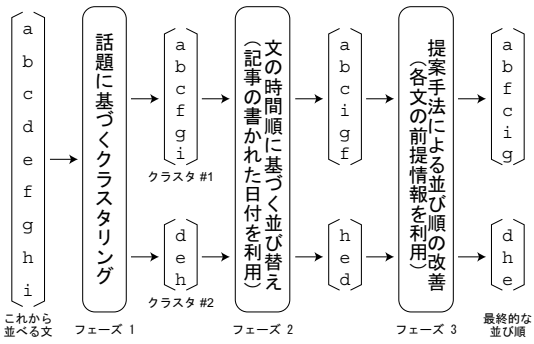

図 4: 文の並び順を決定するシステムのダイアグラム.

ず，文 a,b,...,i の話題に基づいてクラスタリングを行い，文 a,b,c,f,g,i と文 d,e,h の 2 つのグループ（クラスタ）に分ける（以下，フェーズ 1 と呼ぶ）．ある記事は一つの話題に関して述べられていると仮定し，すべての記事に対して，単語頻度を要素とする記事ベクトルを作成し，それらのベクトルを最短距離法 [Cover 67] を用いてクラスタリングする．そのクラスタリング結果を元に，これから並べようとしている文を話題に関して分類し，別の話題について述べている文が一緒に配置されないようにしておく．

次に，話題によって区切られた各グループの文を，時間順序に従って並び替える．それぞれの文には，その元記事が書かれた日時に基づいてタイムスタンプを割り当て，そのタイムスタンプの早い順に文を並べる（フェーズ 2）．図 4 の例では，それぞれのクラスタに対して [a-b-c-i-g-f] と [h-e-d] という並び順が得られている．これまで説明してきた提案手法は，この並び順に対して改善を施すものである（フェーズ 3）．

3. 評価

提案手法の有効性を評価するために，複数文書を対象とした要約システムを構築し，重要文抽出で抜き出された文をどのくらい読みやすく並べるのかを測定する実験を行った．実験には TSC-3 コーパス [Hirao 04] を用い，その中の 28 セットの複数文書要約タスク[*4]を用いた．複数文書向けの重要文抽出法 [Okazaki 04] を適用し，元の文書の約 10%の量になるように文抽出を行い，その抽出された文の範囲内で並び順を決定して，読みやすい要約を作成することを課題とした．抽出された文は次に示すような 6 つの手法を用いて並び替え，その良し悪しを比較することとした: **HO** (人間に並び替えてもらう方法で，評価値の上限として用いる); **RO** (文の並び順をただランダムに決定する方法で，評価値の下限として用いる); **CO** (文を記事の書かれた順番に並べる方法で，図 4 のフェーズ 2 のみを行う方法); **COT** (図 4 のフェーズ 1, 2 を適用する方法で，元文書の中に含まれる話題を認識し，その話題の中で記事の書かれた順番に文を並べる方法で，Barzilay らの手法 [Barzilay 02] と同じ); **PO** (提案手法であり，図 4 のフェーズ 2, 3 を行う方法); **POT** (図 4 のフェーズ 1, 2, 3 をすべて行う方法)．以上の 6 つの手法を用いて並べられた文を人間の被験者 3 人に評価してもらい，文の並び順の採点と，修正すべき並び順に対しては添削結果を得た．

*4 TSC-3 コーパスは 30 セットの複数文書要約タスクで構成されているが，元文書の量が多く 10%の要約率でも要約文の数が 30 前後になってしまう 2 セットは，並び順の評価が難しくなるので除外した．

表 1: 採点による並び順の評価.

	優	可	悪	不可
RO	0.0	0.0	6.0	94.0
CO	13.1	22.6	63.1	1.2
COT	10.7	22.6	61.9	4.8
PO	16.7	38.1	45.2	0.0
POT	15.5	36.9	44.0	3.6
HO	52.4	21.4	26.2	0.0

3.1 被験者の採点による評価

表 1 は，以上で述べた 6 種類の戦略で並べた文章に対して，被験者が 4 段階評価を行った際の平均点[*5]である．採点基準は，4 (優: 並び順を変えてもこれ以上読みやすさの改善が行えないもの); 3 (可: 並び順を修正することで可読性が向上するが，文章としての筋は通っており修正する必要のないもの); 2 (悪: 1 箇所以上の場所で文の順序関係に問題があり，文章としての筋が通っておらず，多少の修正を加えることで可レベルに到達できるもの); 1 (不可: 文の並び順にかなり問題があり，部分的な修正では並び順を改善できず，一から並び順を検討する必要のあるもの) とあらかじめ決めておいた．

表 1 の結果によると，被験者は 75% の人手による並び順 (HO) に対して「優」または「可」をつけているが，ランダムな並び順 (RO) の 95% を「不可」と採点している．時間順 (CO) は RO よりはかなり良いものの満足のいく結果が得られておらず，64% の要約に対して並び順の改善が必要とされている．Barzilay らの手法 [Barzilay 02] で良いとされた COT も満足のいく結果が出せず，むしろ CO よりも悪くなるという結果が出ている[*6]．これに対し，提案手法 (PO) は CO や COT よりもかなり良い結果を示している．特に「可」以上の採点を得る要約の数が 36% (CO) から 55% (PO) にまで改善される点に着目していただきたい．このことから，提案手法は時間的な順序による並び順 CO を改善する効果があることが分かる．

3.2 正解の並び順との比較による評価

以上で説明した採点による評価に加え，各要約の文の並び順が正解の並び順とどの程度近いのかを計測することによって，並び順の良し悪しを評価した．ある文集合に対してその正しい並び順は何通りか存在するので，採点による評価で「悪」と評価した並び順に対して，被験者には図 5 (1) に示すような添削を行っていただくこととした．このとき，添削に用いることの出来る操作は，「ある文の位置を別の位置に移動すること」のみに制限することで，元の並び順を出来るだけ残すように工夫した．ある文の並び順と，その添削結果が得られると，その文の並び順は図 5 (2) に示したように，ある並び順を表す置換 π とその正解の並び順を表す置換 σ で表現することができる．例えば，$\pi(1) = 5$ は，要約文の中で文 s_1 が 5 番目に並ぶことを表し，$\sigma(1) = 7$ は正解の並び順では文 s_1 が 7 番目に並ぶことを示している．スピアマンの順位相関係数 $\tau_s(\pi, \sigma)$ と，ケンドールの順位相関係数 $\tau_k(\pi, \sigma)$ は，2 つの順序関係がどのくらい近いのかを表す尺度として有名である．これらは -1 (逆順) から 0 (無相関の並び順) を通じて，1 (同一の並び順) までの値をとり，図 5 (2) の例では $\tau_s(\pi, \sigma) = 0.85$，$\tau_k(\pi, \sigma) = 0.72$

*5 28[トピック] × 3[人] = 84 個の評価値の平均である．

*6 原因としては，TSC-3 の要約タスクで要約対象とされた文章集合の話題による纏まりが良く，クラスタリングが効果を発揮できなかったことが考えられる．

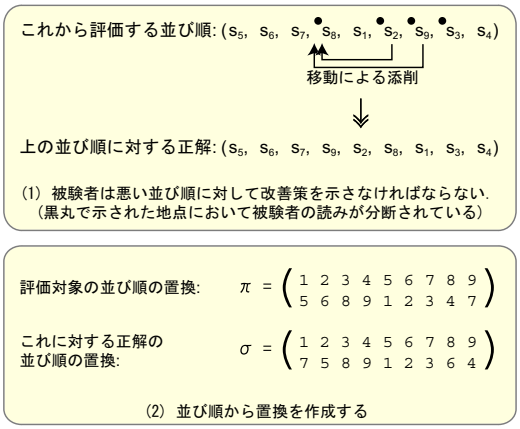

図 5: 並び順の添削とその評価方法.

表 2: 添削による並び順の評価.

	Spearman		Kendall		Continuity	
Method	AVG	SD	AVG	SD	AVG	SD
RO	0.041	0.170	0.035	0.152	0.018	0.091
CO	0.838	0.185	0.870	0.270	0.775	0.210
COT	0.847	0.164	0.791	0.440	0.741	0.252
PO	0.843	0.180	0.921	0.144	0.856	0.180
POT	0.851	0.158	0.842	0.387	0.820	0.240
HO	0.949	0.157	0.947	0.138	0.922	0.138

と計算される.

このほかに，文章を読むときの連続性に着目した連続性 $\tau_c(\pi,\sigma)$ という尺度を以下のように定義した:

$$\tau_c(\pi,\sigma) = \frac{1}{n}\sum_{i=1}^{n} \mathrm{eq}\left(\pi\sigma^{-1}(i), \pi\sigma^{-1}(i-1)+1\right). \qquad (1)$$

この尺度は 0 (非連続的) から 1 (同一の並び順) までの値をとる．図 5 の例では，文 s_7, s_1, s_2, s_9 の文を読んだ後に被験者は次に読むべき文を探さなければならず，その地点で読者の読みが分断されると考え，$\tau_c(\pi,\sigma) = (9-4)/9 = 0.56$ と計算する.

これらの 3 つの尺度 τ_s, τ_k, τ_c を用いて，作成した並び順と，添削後の並び順の相関の平均値 (AVG) と標準偏差 (SD) を示したのが表 2 である．この表からも，HO が一番良く，提案手法 (PO) は時間順 (CO) よりも良く，ランダム (RO) はかなり悪いという結果が読み取れる．RO, CO, PO, HO に対して一次分散分析を行うと，τ_s, τ_k, τ_c のすべての尺度において 4 つの異なる手法を用いることの差が認められた ($p < 0.01$). チューキーの多重比較を用いて，これらの評価値の平均値の比較すると，RO は τ_s, τ_k, τ_c のすべての尺度において，他の手法を用いる場合と比較して悪いことが統計的な有意性をもって示された．スピアマン τ_s，ケンドール τ_k の順位相関係数は検定の有意水準 ($\alpha = 0.05$) における CO と PO，HO の比較に失敗したが，連続性 τ_c のみが CO よりも PO の方が良い，CO よりも HO の方が良いことを示した.

4. 結論

本発表では，複数の新聞記事を要約するときに必要な，文の並び順の決定方法に関する我々の提案手法を紹介した．我々が行った実験によると，提案手法は時間的な順序に基づく並び順で問題のあるケースを 20% 削減することが判明した．今後は，さらなる精度向上を目指すとともに，新聞記事以外の要約ソースに対して文を並べる方法について検討したり，重要文抽出タスクと統合することで重要文抽出と文の並び順の両方の質を向上させる方法等について研究を行う予定である.

謝辞

本研究においては，国立情報学研究所の支援により開催されている評価型ワークショップ NTCIR-4 の自動要約タスク TSC-3 のコーパスを用いました.

参考文献

[Barzilay 02] Barzilay, R., Elhadad, E., and McKeown, K.: Inferring strategies for sentence ordering in multidocument summarization, *Journal of Artifical Intelligence Research (JAIR)*, Vol. 17, pp. 35–55 (2002)

[Cover 67] Cover, T. M. and Hart, P. E.: Nearest neighbor pattern classification, *IEEE Transactions on Information Theory*, Vol. IT-13, pp. 21–27 (1967)

[Hirao 04] Hirao, T., Fukusima, T., Okumura, M., and Nanba, H.: Text Summarization Challenge 3: text summarization evaluation at NTCIR Workshop4, in *Working note of the 4th NTCIR Workshop Meeting* (to appear in 2004)

[Hobbs 90] Hobbs, J.: *Literature and Cognition, CSLI Lecture Notes 21*, CSLI (1990)

[Hume 48] Hume, D.: *Philosophical Essays concerning Human Understanding* (1748)

[Lapata 03] Lapata, M.: Probabilistic text structuring: experiments with sentence ordering, in *Proceedings of the 41st Meeting of the Association of Computational Linguistics*, pp. 545–552 (2003)

[Luhn 58] Luhn, H. P.: The Automatic Creation of Literature Abstracts, *IBM Journal of Research and Development*, Vol. 2, No. 2, pp. 159–165 (1958)

[Mani 01] Mani, I.: *Audomatic Summarization*, John Benjamins (2001)

[Mann 88] Mann, W. and Thompson, S.: Rhetorical structure theory: Toward a functional theory of text organization, *Text*, Vol. 8, pp. 243–281 (1988)

[McKeown 99] McKeown, K., Klavans, J., Hatzivassiloglou, V., Barzilay, R., and Eskin, E.: Towards Multidocument Summarization by Reformulation: Progress and Prospects, in *Proceedings of 16th National Conference on Artificial Intelligence*, pp. 453–460, Orlando, Florida (1999)

[Okazaki 04] Okazaki, N., Matsuo, Y., and Ishizuka, M.: TISS: An Integrated Summarization System for TSC-3, in *Working note of the 4th NTCIR Workshop Meeting* (to appear in 2004)