文の選択と順序付けを同時に行う 評価文書要約モデル

Evaluative Text Summarization Model with Sentence Extraction and Ordering

西川 仁^{*1}
Hitoshi Nishikawa

日本電信電話(株)NTTサイバースペース研究所
NTT Cyber Space Laboratories, NTT Corporation
nishikawa.hitoshi@lab.ntt.co.jp

長谷川 隆明
Takaaki Hasegawa

(同上)
hasegawa.takaaki@lab.ntt.co.jp

松尾 義博
Yoshihiro Matsuo

(同上)
matsuo.yoshihiro@lab.ntt.co.jp

菊井 玄一郎^{*2} (同上)
Gen'ichiro Kikui

keywords: summarization, extractive summarization, sentiment summarization, sentence ordering, joint inference

Summary

In this paper we propose a novel algorithm for opinion summarization that takes account of informativeness and readability, simultaneously. We consider a summary as a sequence of sentences and directly acquire the optimum sequence from multiple review documents by extracting and ordering the sentences. We achieve this with a novel Integer Linear Programming (ILP) formulation. Our proposed formulation is a powerful mixture of the Maximum Coverage Problem and the Traveling Salesman Problem, and is widely applicable to text generation and summarization tasks. We score each candidate sequence according to its informativeness and readability. Since our research goal is to summarize reviews, the informativeness score is measured in terms of opinion information; the readability score is developed in training on the review document corpus. We evaluate our method using the reviews of restaurants and commodities. Our method outperforms existing opinion summarizers as indicated by its ROUGE score. We also show that our proposed method improves the readability of summaries by the results of human readability evaluation.

1. はじめに

インターネット利用者およびインターネットを通じた商取引の拡大に伴い、インターネット上には商品やサービスに対する消費者の意見や感想を記述したレビュー(以下、評価文書と呼ぶ)が数多く存在している。消費者は、新しく商品やサービスを購入しようとする際にこれらの評価文書を読むことによって、よりよい意思決定が可能になると考えられる。また、商品やサービスを開発する生産者側においても、これら消費者の意見や感想を利用することで、よりよい商品、サービスの開発が可能になると考えられる。

これらの評価文書は一つの商品やサービスに対して数多く存在している。人気のある商品やサービスに対しては数百の評価文書が付与されていることもあり、何らかの購買行動に関する意思決定を行う際に、これら全ての評価文書を読むことは現実的ではない。

数多く存在する評価文書を有効に利用する方法の一つは、これらの評価文書を文書要約技術を利用して要約することである。単一の商品やサービスに紐付いた複数の評価文書を要約することは複数文書要約タスクの一種とみなすことができ、実際に研究が行われている [Carenini 06, Lerman 09]。評価文書の要約の例を図 1 に示す。

既存の複数文書要約タスクの中心的なアプローチは、入力された文書集合を文に分割し、得られた文の集合の中から、与えられた要約長を満たし、要約に含めるべき重要な情報を網羅していると思われる部分集合を選び出すというものである。この方法を抽出型要約という [奥村 05]。しかし、単に重要と思われる文を抽出するだけのアプローチは、要約の内容性しか考慮しておらず、生成される要約の文章としての読みやすさ、可読性を考慮していないという問題点がある。すなわち、単に文を抽出するだけでは、読みづらい要約が生成される恐れがある。読みづらい要約は、文書の要点を短時間に把握することができるといった要約の利便性を損なうばかりでなく、読み手に誤った解釈を与えてしまう恐れもある。

^{†1} 現在、NTT メディアインテリジェンス研究所。

^{†2} 現在、岡山県立大学情報工学部情報システム工学科。

小町通りから一本入った路地裏にあるので、ちょっと道がわかりづらいですが、雰囲気は抜群です。料理も素晴らしく、特に鎌倉野菜をふんだんにサラダや相模湾で獲れた魚介類がおすすめです。お酒の種類も豊富で、特に日本酒が充実しています。

図1 飲食店に対する評価文書集合から生成された要約の一例。複数の評価文書からなる文書集合を短く簡潔に要約することができれば、評価文書にアクセスするユーザの利便性を大きく改善することができると期待できる。

要約の可読性を向上させる一つの方法は、選ばれた文を適切に順序付けることである^{*1}。選ばれた文を適切に順序付けることで、要約の可読性が向上することが知られている [Barzilay 02]。しかし、文を選んだのちに文を順序付けることには限界がある。すなわち、文を選んだのちに文を順序付けるとき、どのようにしてもうまく順序付けることのできない文が選択されている場合、この方法では文の可読性を改善することができない。

そこで、本論文では、文の選択と文の順序付けを同時に行う要約モデルを新しく提案する。提案する要約モデルは、文の選択と順序付けを別々に行うのではなく、同時に行う。本論文において要約は、文が含む評価情報を出来る限り要約に含め、かつ要約を構成する文が自然に並ぶように、入力された文書集合から適切な文の順列を探索する問題として定式化される。

また、提案手法の優位性を示すための実験を行い、その結果を示す。要約の自動評価尺度 ROUGE [Lin 04] を利用した実験によって、提案する要約モデルは、その内容性において既存の評価文書を対象にした要約手法を上回ることを示す。また、被験者実験によって、文の選択と順序付けを同時に行うことで、それらを別個に行った場合よりも可読性が向上することを示す。

本論文の構成は下記の通りである。第2章では関連研究について述べ、本研究の新規性を明らかにする。第3章では提案する要約モデルについて述べる。第4章では要約モデルのパラメータの推定について述べ、第5章では要約モデルのデコードについて述べる。第6章では評価実験について述べ、最後に第7章では本論文をまとめる。

2. 関連研究

2.1 複数文書要約

何らかの事物に紐付いた評価文書の集合を要約するため、複数文書要約の手法を利用することは既に行われている [Carenini 06, Lerman 09]。Carenini らは、Radev らが

提案した抽出型の複数文書要約手法 [Radev 04] を利用して評価文書の要約を行っている [Carenini 06]。Carenini らの手法は、まず、評価文書から評価情報を抽出する。評価情報は評価属性と評価極性からなる。例として「画質がよい」という表現を考える。この表現は、何らかの対象の属性である「画質」と、その評価表現「よい」、そして評価表現の極性からなる。次に、入力文書集合中での評価属性の出現頻度などから、評価属性と評価極性の組に重要度を与える。最後に、要約が含む、評価属性と評価極性の重要度を最大化するように文を選択し、要約を生成する。Lerman らも評価属性と評価極性を単位として抽出する点は変わらないが、要約が含む評価属性と評価極性の分布を、元の評価文書集合の分布と近づけるように要約を生成する方法を提案している [Lerman 09]。本論文においても、Carenini ら、Lerman らと同様に評価属性と評価極性を単位として要約を行う。しかしながら、これらのアプローチはあくまで要約にふさわしい文を選択することに焦点を当てているのに対し、我々の手法は文の順序まで同時に加味する点が大きく異なる。

入力を評価文書に限らない、汎用的な文書要約手法として、最大被覆問題に基づいた要約モデルの有効性が近年明らかになっている [Filatova 04, Yih 07, 高村 08, Gillick 09]。最大被覆問題に基づく要約モデルは、入力された文書集合が含む、ユニグラムやバイグラムといった何らかの概念を、できる限り要約が被覆するように文を選び出す要約モデルである。Filatova らは要約を最大被覆問題として表現し、性能保証付き貪欲法で解いた [Filatova 04]。Yih らはスタックデコーダーを利用したデコード法を提案している [Yih 07]。高村ら、Gillick らは最大被覆問題を整数計画問題として表現し、整数計画問題を解く汎用的なアルゴリズムである分枝限定法を用いることで最適な要約を得ることができることを示した [高村 08, Gillick 09]。本論文でも、最大被覆問題に基づくモデルを提案するが、高村ら、Gillick らの手法がユニグラムやバイグラムといった単位を被覆することによって要約を生成するのに対し、本論文では入力される評価文書集合の中に含まれる評価情報をより多く被覆するように要約を生成する。これは、ユニグラムやバイグラムの代わりに、情報抽出器によって抽出された評価情報を被覆の対象とすることで実現できる。

2.2 文の順序付け

要約として選択された文を適切に順序付けることによって、要約の可読性が向上することが知られている [Barzilay 02]。Barzilay らは、要約として選択された文が含まれている文書の出版日に従って、選択された文を順序付ける方法を提案している [Barzilay 02]。文を順序付ける方法として、Lapata は事前に大規模なコーパスを利用して、ある単語が出現したのちに次の文で出現しやすい単語を求め、この特徴量を用いて文を順序付けることを提

^{*1} もちろん、文の順序以外にも、要約の可読性に影響を与える要因は存在する。例えば、米国 National Institute of Standards and Technology が開催した評価型ワークショップ Document Understanding Conference 2007 では、可読性を構成する要素として、文法性、非冗長性、照応関係の明かさ、焦点、構造と一貫性の5つの尺度が示されている [National Institute of Standards and Technology 07]。文の順序付けは、特に5番目の構造と一貫性の尺度に関連すると考えられる。

案している [Lapata 03] . 本論文でも, 文を順序付ける際の特徴量として Lapata らと同様に文が含む内容語や固有表現を利用する .

文章の可読性を評価するための, 文の順序に基づく特徴量として, Barzilay らは Entity Grid を提案している [Barzilay 05, Barzilay 08] . これは, ある文において主語, 目的語といった何らかの役割において出現している単語は次の文においても何らかの役割を伴って出現しやすいという統計的な傾向を捉えるもので, Barzilay らはこの特徴量によって適切に文が順序付けられた文章とそうでない文章を高精度に弁別できることを示した . 横野らは Entity Grid を日本語の文章に適用し, その結果を報告している [横野 10] . Lapata の手法 [Lapata 03] が全ての文の間それぞれに文同士のつながりの良さを定義しその和を以って文の順序の良さを評価しているのに対し Entity Grid は文章全体から一つの特徴ベクトルを構成し, 文の順序の良さを評価している . そのため, Entity Grid を特徴量として文を順序付けることも可能であるとは考えられるものの, 本論文で提案する要約モデルのように個別の文の間に特徴ベクトルを構成する方法に直接適用することができない^{*2} .

文を順序付ける際には, 与えられた文の集合からそれらの適切な順列を求める必要がある . Barzilay らや Lapata が近似解法によってこれらの順列を求めたのに対し [Barzilay 02, Lapata 03] , Althaus らはこの問題が巡回セールスマン問題と等価であることを指摘し, 整数計画問題として問題を表現することでこれを解いた [Althaus 04] . これらはあくまで文を順序付けることを目的としているのに対し, 本論文では文の選択も同時に行う点が大きく異なる .

2.3 ヘッドライン生成

ある集合中から要素を選択し, それらを同時に順序付けるタスクとして, ヘッドライン生成がある [Banko 00, Deshpande 07] . Banko らはこれをビームサーチを利用して解いたが, Deshpande は整数計画問題としての定式と, 乱択法を利用した解法を報告している . これらは単語を要素として選択と順序付けを行うのに対し, 本論文は文を単位とする点が異なる .

3. 要約モデル

本章では, 本論文で用いる要約モデルについて述べる .

例として, 1 つ以上の評価文書を構成し, 既に文に分割された, 9 個の文の集合 $D = s_1, s_2, \dots, s_9$ が入力として与

図2 要約のグラフ表現 . 要約の先頭を表すノード s_0 と要約の末尾を表すノード s_{10} の間に, 要約として選択された文 s_1 , 文 s_6 , 文 s_9 が有向辺を通じて連結されたものとして抽出型の要約は表現できる .

えられたと考える . 次に, 要約が 3 つの文 s_1, s_6, s_9 から構成されており, またそれらが $s_1 \rightarrow s_6 \rightarrow s_9$ の順に順序付けられているとする . 順序の先頭に記号 s_0 を加え, また末尾に記号 s_{10} を加え, 要約を $S = \langle s_0, s_1, s_6, s_9, s_{10} \rangle$ と表現する . このとき, 要約 S は, 図2に示すように, s_0 を始点, s_{10} を終点とする有向グラフとして表現することができる .

入力文集合 D に含まれる文 s_i から文 s_j への有向辺を $a_{i,j}$ と表現する . このとき, 図2に示すノード $s_0, s_1, s_6, s_9, s_{10}$ の間の有向辺は, $a_{0,1}, a_{1,6}, a_{6,9}, a_{9,10}$ の4本である .

本論文では, 文を適切に順序付けるために, 隣接する文同士の接続されやすさを数値として各有向辺の上に定義する . 具体的には, 有向辺 $a_{i,j}$ に接続スコア $c_{i,j}$ を与える . 本論文では, 要約に含まれる文を接続する有向辺の集合を $A(S)$ として, それらの接続スコアの和 $\sum_{a_{i,j} \in A(S)} c_{i,j}$ が高くなるにつれ, 要約の可読性が向上すると仮定する .

次に, 入力文書集合 D が含む概念 e を考える . 概念 e が要約に被覆されるためには, 概念 e を含む文を少なくとも1つ要約は含む必要がある . 例を表1に示す . 表1によれば, 文書集合 D は8つの概念 $e_1, e_2, \dots, e_7, e_8$ を含んでいる . 文 s_1 は概念 e_1 と e_6 を含んでいる . また, 文 s_2 は概念 e_2 を含んでいる . このように, 文 s_i と概念 e_k の包含関係は行列として表現することができ, 文 s_i が概念 e_k を含むときには文-概念行列 M の i 行 k 列の要素が1に, 含まないときには0になるとする . 本論文では, それぞれの概念 e_k はその重要度に応じてスコア w_k を持つとする . 本論文ではこれを内容スコアと呼び, 概念 e_k が高い内容スコアを持つ場合, その概念は要約にとって重要なものであり, 要約に含めるべき情報と考える . 要約に含まれる概念の集合を $E(S)$ として, それらの内容スコアの和 $\sum_{e_k \in E(S)} w_k$ が高くなるにつれ, 要約が持つ内容性が向上すると仮定する .

以上の定義の下で, 本論文では以下のように目的関数 $\mathcal{L}(S)$ を定義する .

$$\mathcal{L}(S) = \lambda \sum_{e_k \in E(S)} w_k + (1 - \lambda) \sum_{a_{i,j} \in A(S)} c_{i,j} \quad (1)$$

^{*2} もちろん, 個別の文の間に Entity Grid に基づく特徴ベクトルを定義することも可能であるものが, それには Barzilay らによって提案された Entity Grid に若干の変更を加えることになるため, それは将来の課題としたい .

表 1 文-概念行列．文と、文が含む概念の包含関係は行列として表現できる．文 s_i が概念 e_k を含む場合には、行列の i 行 k 列の要素が 1 となる．

	e_1	e_2	e_3	...	e_6	e_7	e_8
s_1	1	0	0		1	0	0
s_2	0	1	0		0	0	0
s_3	0	0	0		0	0	1
$\vdots$				$\ddots$			
s_7	0	0	0		0	0	0
s_8	0	0	1		0	1	0
s_9	0	0	0		0	0	1

$$\text{s.t. length}(S) \leq L$$

$\text{length}(S)$ は要約 S の長さを返す関数とし、 L は要約長とする． λ は内容スコアと接続スコアを調整するパラメータである．すなわち、本論文では、与えられた要約長以内で、内容スコアを高めるように概念を要約に含め、また接続スコアを高めるように順序付けられた文の順列 S を探索する問題として、文書要約を定式化する．

この問題は NP 困難である．それぞれの文が 1 つの異なる概念のみを含むと考えた場合^{*3}，上の目的関数は景品収集巡回セールスマン問題 [Balas 89] となる．本論文ではこの問題を整数計画問題として表現し、汎用ソルバーを用いて解を求める．デコードについては 5 章で述べる．

4. パラメータ推定

本論文で提案するモデルは 3 つのパラメータを必要とする．入力文書集合に含まれる概念 e の内容スコア w と、文間に定義される有向辺 a に定義される接続スコア c ，そして λ である．ここでは内容スコア w と接続スコア c の推定について述べる．

4.1 内容スコア

本論文の目的は評価文書の要約である．そのため、前述したように、評価文書が含む評価情報を単位として要約を生成する．

まず、評価情報について述べる．本論文では、評価情報を評価属性 $aspect$ と評価極性 $polarity \in \{-1, 0, 1\}$ の組 $e = (aspect, polarity)$ として考える．評価属性は、何らかの対象が評価される際の属性を示し、評価極性は、対象が評価属性に関してポジティブな評価をされているときに 1、ネガティブな評価をされているとき -1、どちらとも言えないとき 0 の 3 値を取るものとする．一例として、「このデジタルカメラは画質がよい」という表現を考える．この表現の評価属性は「画質」である．また、評

価表現「よい」はポジティブな評価であることから、評価極性は 1 となり、評価情報 $e = (\text{画質}, 1)$ が得られる．

評価情報は以下のように抽出する．

- (1) 文に係り受け解析を行う．
- (2) 評価表現とその評価極性の組を格納した辞書^{*4}と係り受け解析の結果を照合し、評価表現とその評価極性を得る．
- (3) 係り受け木で、評価表現に係っている名詞を評価属性とする．
- (4) 得られた評価属性と評価極性を出力する．

評価属性と評価極性の組を 1 つの概念 e_k として、それらに対して内容スコア w_k を与える．本論文では、入力文書集合中での e_k の出現回数をそのまま w_k とする．

4.2 接続スコア

次に、接続スコア c について述べる．前述したように、本論文では 2 つの文 s_i, s_j の間の有向辺 $a_{i,j}$ に対して接続スコア $c_{i,j}$ を与える．2 つの文 $d = \{s_i, s_j\}$ とその順序 $o = \langle s_i, s_j \rangle$ が与えられたとき、それらの接続スコア $c_{i,j}$ を以下のように表現する．

$$c_{i,j} = \mathbf{w} \cdot \phi(d, o) \quad (2)$$

$\mathbf{w}$ は重みベクトルであり、 $\phi(d, o)$ は 2 つの文 $d = \{s_i, s_j\}$ とその順序 $o = \langle s_i, s_j \rangle$ の特徴ベクトルである． $\mathbf{w} \cdot \phi(d, o)$ は $\mathbf{w}$ と $\phi(d, o)$ の内積である．

特徴ベクトルの要素として、2 つの文に含まれる、名詞、動詞、形容詞の表層形、それらの品詞、および LOC や ORG といった固有表現タグの直積集合を特徴量として用いる．例えば、図 1 に含まれる文「小町通りから一本入った路地裏にあるので、ちょっと道がわかりづらいですが、雰囲気は抜群です」と文「料理も素晴らしく、特に鎌倉野菜をふんだんにサラダや相模湾で獲れた魚介類がおすすめです」という文からは、前の文に含まれる「道」という単語と後の文に含まれる「料理」という単語から順序対 $(\text{道}, \text{料理})$ を得ることができ、これが特徴ベクトルを構成する特徴量となる．文章の先頭と末尾を表す擬似的な文 s_0 および s_{n+1} は、それぞれ文章の先頭と末尾を表す $\langle d \rangle$ および $\langle /d \rangle$ というタグとして表す．例えば、要約の最初の文が「道」という単語を含む場合には、特徴量 $(\langle d \rangle, \text{道})$ が抽出される．同様に、要約の末尾の文が「料理」という単語を含む場合には、特徴量 $(\text{料理}, \langle /d \rangle)$ が抽出される．これらの特徴量は [Lapata 03] によって提案されたものである．

^{*4} 辞書は 57,238 項目からなり、それぞれの項目は評価表現と評価極性の組からなる．評価表現は「良い」「悪い」「綺麗」といった表現とその品詞情報であり、評価極性は前述のように 1, 0, -1 の 3 値からなる．類似した辞書として高村らによる辞書 [Takamura 05] がある．高村らの辞書においては評価極性を 1 から -1 までを値域とする実数としているのに対し、本論文で用いている辞書は評価極性を 3 値としている点が異なる．

^{*3} すなわち表 1 が対角行列である場合．

要約全体の接続スコアは、個々の文間の接続スコアの和とする。すなわち、文の集合 $d = \{s_0, s_1, \dots, s_n, s_{n+1}\}$ と、その順序 $o = \langle s_0, s_1, \dots, s_n, s_{n+1} \rangle$ の特徴ベクトル $\Phi(d, o)$ を以下のように表現する。

$$\Phi(d, o) = \sum_{d, o} \phi(d, o) \quad (3)$$

従って、要約全体の接続スコアは、重みベクトル w と特徴ベクトル $\Phi(d, o)$ の内積 $w \cdot \Phi(d, o)$ として表現される。

重みベクトル w は訓練事例から推定する。訓練事例として、文の集合 d_t とその正しい順序 o_t の T 個の組 $\{d_t, o_t\}_{t=1}^T$ が与えられるものとする。本論文では、先行研究 [Lapata 03] と同様に、人間が書いた文書を構成する文は正しく順序付けられているという仮定を置く。そのため、何らかの文書に対して文分割、形態素解析、固有表現抽出を実施することによって、人手によるアノテーションを必要とせずに訓練事例を得ることができる。訓練事例の詳細は 6 章で述べる。

訓練事例が与えられた際、重みベクトル w が、文の集合 d_t の正しい順序 o_t に対して、誤った順序 $\hat{o}$ *5 よりも高いスコアを与えることができたとする、その重みベクトル w は自然な文の順序に対して不自然な文の順序よりも高いスコアを与えることができると期待できる。この仮定に基づき、本論文では、訓練事例に含まれる、正しく並べられた文の順序に高いスコアを与えるように重みベクトル w を学習する。

重みベクトル w を推定するため、本論文では Passive-Aggressive algorithm [Crammer 06] を用いる。このアルゴリズムはオンライン型の学習アルゴリズムであり、一度に 1 つの訓練事例を取り上げ、現在の重みベクトルを利用して得られる解と、訓練事例に含まれる正しい解を比較する。それら 2 つの解が異なる場合は、損失関数に従って重みベクトルを更新する。本論文の場合、訓練事例に含まれる正しい順序 o_t と、現在の重みベクトルによって得られる順序 $\hat{o}$ を比較し、それらが異なる場合は正しい順序により高いスコアが与えられるように重みベクトルを更新する。重みベクトルの更新式を以下に示す。

$$w' = \arg \min_{w'} \frac{1}{2} \|w' - w\|^2 + C\xi^2 \quad (4)$$

$$\text{s.t. } w' \cdot \Phi(d_t, o_t) - w' \cdot \Phi(d_t, \hat{o}) + \xi \geq \ell(\hat{o}; o_t)$$

ここで、 w は更新前の重みベクトル、 w' は更新後の重みベクトル、 $\ell(\hat{o}; o_t)$ は損失関数である。 C は aggressiveness parameter と呼ばれ、重みベクトルの更新を制御するパラメータである。 ξ はスラック変数である。更新前の重みベクトル w は、 w の下で推定された順序 $\hat{o}$ が

誤っていた際、すなわち $o_t \neq \hat{o}$ であるときに、式 (4) を解くことによって得られる以下の更新式によって更新される。

$$w' = w + \eta(\Phi(d_t, o_t) - \Phi(d_t, \hat{o})) \quad (5)$$

$$\eta = \frac{\ell(\hat{o}; o_t) - \Phi(d_t, o_t) + \Phi(d_t, \hat{o})}{\|\Phi(d_t, o_t) - \Phi(d_t, \hat{o})\|^2 + \frac{1}{2C}} \quad (6)$$

重みベクトルを更新するには、現在の重みベクトル w の下で最もスコアの高い順序 $\hat{o}$ を求める必要がある。この操作は、文の集合からそれらの最良の順序を求める問題、すなわち巡回セールスマン問題 [Althaus 04] となる。この操作は NP 困難であることが知られており [Korte 08]、従って最適解の求解は容易ではない。もちろん、巡回セールスマン問題も整数計画問題として表現することができ、分枝限定法などのアルゴリズムによってその厳密解を求めることができるが、本論文では以下の 2 つの理由により、ビームサーチを利用して最適解ではなく近似解を求めることにする。

- (1) 学習の際には、訓練事例の数とイテレーションの数をかけた回数だけ巡回セールスマン問題を解かなければならず、厳密解を求めている学習に時間がかかり過ぎる恐れがある。
- (2) オンライン型のアルゴリズムは、近似解からでも良好な重みベクトルを推定できることが知られており [Zhang 08]、従って最適解を求めずとも適切な重みベクトルが得られると期待できる。

損失関数 $\ell(\hat{o}; o_t)$ に、本論文ではケンドールのタウを用いる。

$$\ell(\hat{o}; o_t) = 1 - \tau \quad (7)$$

$$\tau = 1 - 4 \frac{\text{Swap}(\hat{o}, o_t)}{N(N-1)} \quad (8)$$

τ はケンドールのタウである。 $\text{Swap}(\hat{o}, o_t)$ は、順序 $\hat{o}$ を順序 o_t に変換するために必要な、隣接する要素を交換する操作の回数である。 N は要素の数、すなわちこの場合では文の数である。Lapata らによれば、ケンドールのタウは文の順序の良さに関して、人間による評価と高い相関があり [Lapata 06]、従ってケンドールのタウによる損失を最小化することによって、人間によって自然な文の順序を再現する重みベクトルを得られると期待できる。

5. デ コ ー ド

本節では、3 章で述べた要約モデルをデコードする方法について述べる。すなわち、式 (1) を最大化する要約 S を探索する方法について述べる。本論文では、式 (1) を整数計画問題として表現し、これを整数計画問題に対

*5 誤った順序 $\hat{o}$ は、順序を構成する文の数を n としたとき、 $nP_n - 1$ すなわち $n! - 1$ 通り存在する。なお、順序の先頭と末尾を表す記号である s_0 および s_{n+1} はそれぞれ必ず順序の先頭と末尾に存在するものとし、順序の推定の対象とはしない。

応した汎用的なソルバーを用いて解くことで、推定されたパラメータの下での最良の要約を得る．

式 (1) を整数計画問題として表現するため、3 つの決定変数 x, y, z を用意する．決定変数 x_i は文書集合 D に含まれる文 s_i が要約に含まれるときに 1, 含まれないときに 0 となるものとする．決定変数 $y_{i,j}$ は、文 s_i および文 s_j が共に要約に含まれており、かつ文 s_i の直後に文 s_j が続いている際に 1 となり、そうでないときに 0 となるものとする．決定変数 z_k は概念 e_k が要約に含まれる際に 1, 含まれないときに 0 となるものとする．図 2 を例とすると、文に関する決定変数 $x_0, x_1, x_6, x_9, x_{10}$ および辺に関する決定変数 $y_{0,1}, y_{1,6}, y_{6,9}, y_{9,10}$ が 1 となる．決定変数 z_k は、表 1 に例を示した文-概念行列 M に従って、概念 e_k が要約に含まれるときに 1, 含まれないときに 0 となるものとする．

5.1 目的関数

整数計画問題の目的関数は以下ようになる．

$$\max \left\{ \lambda \sum_k w_k z_k + (1 - \lambda) \sum_{i,j} c_{i,j} y_{i,j} \right\} \quad (9)$$

式 (9) は、概念の重要度に関するパラメータ w_k と文の順序の自然さに関するパラメータ $c_{i,j}$ の下で、重要な概念を可能な限り要約に含めるように文を選択しつつ、それらを出来るだけ自然に並べることが目的としている． λ は w と c のバランスを調整するパラメータである．

5.2 制約

式 (9) を解くことによって出力される解が、要約として必要な性質を満たすことを保証するためには、いくつかの制約をを与える必要がある．

まず、3 つの決定変数 x, y, z の値域を以下のように制限する．

$$\begin{aligned} x_i &\in \{0, 1\} \quad \forall i \\ y_{i,j} &\in \{0, 1\} \quad \forall i, j \\ z_k &\in \{0, 1\} \quad \forall k \end{aligned}$$

すなわち、3 つの決定変数はいずれも 0 あるいは 1 を値として取る二値変数である．本論文の提案する要約モデルは、同じ文を 2 回含むことを許さない．辺についても同様である．概念については、同じ概念が 2 回以上要約に含まれることを許容するものの、目的関数上では、同じ概念の重要度は 1 度しか目的関数に寄与しない．例えば、表 1 を例とすると、文 s_3 と文 s_9 が要約に選ばれているとき、要約には概念 e_8 が 2 つ含まれていることになるが、決定変数 z_8 はバイナリ変数であるため、その値は 1 である．この、同じ概念は一度しか目的関数に寄与

しないという制約によって、要約の冗長性を削減することができる [高村 08] ．

また、出力される要約の長さは必ず要約長 L 以内とならなければならない．そのため、長さに関する不等式制約を与える．

$$\sum_i l_i x_i \leq L$$

l_i は文 s_i の長さを示し、 L は要約長を示す．

次に、文と概念の包含関係に関する制約を与える．

$$\sum_i m_{i,k} x_i \geq z_k \quad \forall k$$

この不等式制約は表 1 に例を示した文-概念行列 M を表現するものである． $m_{i,k}$ は文-概念行列 M の i 行 k 列の要素を示し、文 s_i に概念 e_k が含まれている場合に 1 となるものである．この不等式制約によって、文 s_i が要約に含まれていない場合には文 s_i が含む概念もまた要約には含まれない．

要約の先頭と末尾を表現するため、要約の先頭を表すノード s_0 に対応する決定変数 x_0 と、要約の末尾を表すノード s_{n+1} に対応する決定変数 x_{n+1} を導入する．

$$x_0 = 1$$

$$x_{n+1} = 1$$

n は入力文書中の文の数である．要約は、文の順列として、必ず先頭と末尾を必要とするため、これらの決定変数は必ず 1 である．

文や概念と同様に、辺についても制約を与える．要約の先頭を表すノードは、自身のノードに向かう辺を持たないが、必ずどこか 1 つ別のノードに向かう辺を持つ．同様に、要約の末尾を表すノードは他のいずれかのノードから自身に向かう辺を持つが、外に向かう辺をもたない．

$$\sum_i y_{0,i} = 1$$

$$\sum_i y_{i,0} = 0$$

$$\sum_i y_{n+1,i} = 0$$

$$\sum_i y_{i,n+1} = 1$$

要約の先頭を表すノードと末尾を表すノード以外のノード、すなわち要約に含まれる文は、必ず別のノードから自身に向かう 1 つの辺と、自身から別のノードに向かう 1 つの辺を持つ．

$$\sum_i y_{i,j} + \sum_i y_{j,i} = 2x_j \quad \forall j$$

$$\sum_i y_{i,j} = \sum_i y_{j,i} \quad \forall j$$

これらの制約だけでは、グラフに閉路が発生する恐れがある。閉路を防ぐために、ネットワーク・フローに基づく以下の制約を加える。

$$\sum_i f_{0,i} = n$$

$$\sum_i f_{i,0} \geq 1$$

$$\sum_i f_{i,j} - \sum_i f_{j,i} = x_j \quad \forall j$$

$$f_{i,j} \leq ny_{i,j} \quad \forall i, j$$

フロー $f_{i,j}$ は辺 $a_{i,j}$ 上に定義される。上の制約では、 n 単位のフローが、 s_0 を始点、 s_{n+1} を終点として、始点から終点に向かって流れていくことを表現している。フローは各ノードで 1 単位ずつ消費される。終点には必ず、1 単位以上のフローが届かなければならない。もし閉路が発生していると、あるノードに n 単位のフローが流入しているときには、そのノードからは $n-1$ 単位のフローが流出するという制約が満たされなくなるため、この制約によって閉路の発生を防ぐことが出来る。この制約はヘッドライン生成においても閉路を除去するために利用されている [Deshpande 07]。

6. 実験

本節では、提案手法の有効性を検証するために行った 2 つの実験の結果について述べる。生成された要約が、入力された文書集合の内容を適切に反映しているか（要約の内容性）、読みやすい要約になっているか（要約の可読性）、これらの 2 点の評価を行う。要約の内容性に関する評価実験については 6.1 節、要約の可読性に関する評価実験については 6.2 節で述べる。以下のいずれの実験においても、要約を生成する際には高速な商用ソルバーである IBM ILOG CPLEX *6 を利用した。要約長 L は 300 バイトとした。パラメータ λ は 0.5、すなわち内容スコアと接続スコアを等しく重み付けるように設定した*7。

*6 <http://www-06.ibm.com/software/jp/websphere/ilog/optimization/core-products-technologies/cplex/>

*7 実験に用いたコーパスとは別に開発セットを用意し、開発セットに含まれる文書から生成された要約を著者が確認することで λ を調整した。理想的には、要約の内容性および可読性に関する評価を包含する、総合的な要約の評価尺度を用意し、その尺度による評価を最大化するように λ を調整するべきである。しかし、現時点では、そのような評価には人手による大規模な評価実験が必要となり、多大なコストを要する。そのため、本論文では著者らの目視によって λ を調整した。 λ の最適化は、要約の総合的な自動評価手法の確立も含め、興味深い課題である。

表 2 コーパスの統計量。平均文書数は一つの飲食店あるいは商品に紐付いている評価文書数の平均を取ったもの。平均サイズはそのサイズである。

	平均文書数	平均サイズ (byte)
飲食店	29.40	5,343
商品	44.75	10,173

形態素解析には Fuchi らによる形態素解析器 [Fuchi 98] を、係り受け解析には Imamura らによる係り受け解析器を利用した [Imamura 07]。評価表現辞書は Asano らによる辞書を利用した [Asano 08]。

パラメータ $a_{i,j} \in A$ の推定のため、評価に利用するコーパスとは別に、飲食店に関する評価文書を 4,390 文書、商品に関する評価文書を 47,570 文書を収集した*8。式 (6) のパラメータ C は 100 とした。

6.1 要約の内容性

要約の内容性の評価のため、何らかの飲食店に関する評価文書の集合、および何らかの商品に関する評価文書の集合をそれぞれ 100 セットずつ用意した。すなわち、ある飲食店に関する評価が書かれた文書の集合 100 店舗分、ある商品に関する評価が書かれた文書の集合 100 商品分である。コーパスの統計量を表 2 に示す。また、評価のため、それぞれの評価文書集合に対して 4 つの参照要約を付与した。参照要約は、対応する評価文書集合を読んだ人間によって書かれたもので、同一の飲食店あるいは商品に紐付いている参照要約は全て異なる人間によって記述されている。参照要約の書き手に著者らは含まれない。

要約の内容性の評価には、要約の自動評価尺度である ROUGE [Lin 04] を用いた。ROUGE は、機械によって生成された要約と人間に記述された参照要約のある種の類似度を計算するものである。評価のため、ROUGE の中でも ROUGE-2、ROUGE-SU4、ROUGE-SU9 を利用した。

それぞれの要約の ROUGE 値を求める際には、4 つの参照要約全てに対して ROUGE 値を求め、最も高い値を ROUGE 値とした。なお、平尾らの報告によれば、日本語の要約と参照要約の ROUGE 値を計算する際には、文を構成する全ての単語ではなくて、名詞、動詞、形容詞、未知語のみを用いて計算を行った方が人間による評価と相関が高い結果が得られる [平尾 06]。そのため、本論文でも ROUGE 値を計算する際には名詞、動詞、形容詞、未知語のみを対象とした。

*8 飲食店に比べ、商品はその内容において多様と考えられることから、文の接続に対してもより多数のパラメータを推定しなければならないと考えられる。そのためより多くの文書を用意した。

§1 比較手法

提案手法の内容性の評価のため、以下の6つの手法を比較する。

- **Content:** 被覆を行う対象を、評価情報ではなく内容語にしたもの。具体的には、形態素解析器によって名詞、動詞、形容詞、未知語とされた単語を被覆の対象として要約を行う。内容語の重みは、単語の入力文書集合中での出現頻度とした。これを手法 Content とする。
- **Carenini:** Carenini らによる手法 [Carenini 06] を実装したもの。これを手法 Carenini とする。手法の詳細は後述する。
- **Lerman:** Lerman らによる手法 [Lerman 09] を実装したもの。これを手法 Lerman とする。手法の詳細は後述する。
- **手法 (1):** 文の選択と文の順序付けを別々に行うもの。これを手法 (1) とする。具体的には以下の手順によって要約を行う。
 - (1) まず、式 (1) の第一項のみを目的関数とし、与えられた要約長を満たすよう文を選択する。すなわち、内容スコアのみを用いて文を選択する。文の順序付けは行わない。
 - (2) 要約に用いる文を選択したら、それらを式 (1) の第二項のみを用いて順序付ける。すなわち、接続スコアのみを用いて文を順序付ける。
- **手法 (2):** 提案手法。文の選択と文の順序付けを同時に行うもの。これを手法 (2) とする。
- **Human:** 人手による4つの参照要約のうち、最も ROUGE が高い2つの組の ROUGE を示したもの。これが評価の上限となる。

なお手法 Content、手法 Carenini、手法 Lerman で選ばれた文は、投稿された日時が早い文書から選ばれた文が前に、また同じ文書から選ばれた文は文章の先頭に近い文が前になるように並べられて出力される。

これらの比較によって明らかにしたいことは、以下の3点である。

- 評価情報を被覆の対象として用いることの影響。評価情報を概念として扱い要約を生成することで、評価文書を要約の対象とした際には内容語等を概念として用いるよりもよい要約が生成できると期待する。これを手法 Content と手法 (1)、手法 (2) を比較することで行う。
- 内容性に関する、既存手法への優位性。前に述べたように、本論文の主たる貢献は可読性に優れた要約を生成するところにある。しかし、提案手法によって可読性が向上したとしても、既存の評価文書の要約手法に比べて提案手法の内容性が低い場合、提案手法は要約技術として必ずしも有意義とはいえない。そのため、既存手法と提案手法の内容性を比較した

い。これは、Carenini らによる手法、Lerman らによる手法と、手法 (2) を比較することで行う。

- 文の選択と順序付けを同時に実施する手法の内容性に対する影響。文の選択のみをまず行ったのち文の順序付けを行う手法と、文の選択と順序付けを同時に行う手法を比較したとき、後者は文を選択する際に可読性も考慮して選択を行うため、前者に比べて後者の要約は内容性が低下している恐れがある。そのため、前者と後者の内容性を比較したい。これは、手法 (1) と手法 (2) を比較することで行う。

§2 ベースライン

実験に用いる2つのベースライン、手法 Carenini と手法 Lerman について述べる。

- **Carenini:** Carenini らは、評価文書要約のため、MEAD [Radev 04] を評価文書要約向けに改良した MEAD* と、自然言語生成器に基づく方法の2つを紹介している [Carenini 06]。前者は評価情報を抽出した後ならば言語に依存しない手法であり^{*9}模倣が容易である。一方、後者は自然言語生成器が必要となるため模倣が難しい。そのため本論文では前者の MEAD* を実装し利用する。なお Carenini らの報告によれば、これら2つの手法は内容性においては同等水準の性能を持っている。また、MEAD* は評価文書を要約の対象とした際には MEAD を上回る性能を持っている。

手法 Carenini は以下の手続きによって要約を生成する。

- (1) 入力文書集合に含まれる各文から評価情報を抽出する。
- (2) 各文を、それらの文が持つ評価情報の評価属性に応じて「バケット」に分類する。例えば評価属性「料理」を持つ文は「料理」バケットに入れられる。複数の異なる評価属性を持つ文は複数の異なるバケットにそれぞれ入れられる。
- (3) 各バケットから一つずつ文を選び出す。同一のバケットに属している文のうち、そのバケットの評価属性に関して、より顕著な評価極性を持っている文を選ぶ。具体的には、評価極性の絶対値（評価極性が +1 の場合も -1 の場合も 1 となる）の和が大きいものを選ぶ。例えば、ある文が評価属性「料理」に関してネガティブな評価極性を持つ評価情報を2つ持っている場合、この文の評価属性「料理」に関する評価極性の絶対値の和は2となる。もし評価極性の絶対値の和が同数の文が複数ある場合は、より短い文を選ぶ。バケットを選択する順序は、バケットに属している文の数が多いものから順に選ぶものとした。文の数が同数である場合はランダム

^{*9} 評価表現辞書は言語依存であるため、評価情報の抽出は言語に依存する。

に選択する．全てのバケットから文を選んだ場合は、最初に文を取りだしたバケットから再び文を取りだす．もちろん、同一の文が 2 回以上要約に選ばれることはない．

- (4) 手続き (3) を、選択された文の集合の長さの和が要約長に達するまで繰り返す．

● **Lerman:** Lerman らは、評価文書要約のため、要約に含まれる評価情報の分布と、入力文書集合に含まれる評価情報の分布の差異をカルバック・ライブラー情報量で計り、差異が少ない要約をよい要約とするモデルを提案している．Lerman らの手法も MEAD*と同様に、評価情報を抽出した後は言語に依存しない．彼らの提案するモデルの目的関数は非線形となっており、最適化が難しい．Lerman らは山登り法を用いてデコードを行っている．山登り法は容易に局所解に収束する恐れがあるが、Lerman らは大量の初期解を用意することでこれに対処している．モデルとデコードを切り離して議論するため、出来る限り最適解に近い解を生成したいが、Lerman らの提案する目的関数は最適化が容易ではないため、我々も同様に大量の初期解を生成し出来る限り良好な解を得ることを試みる．

手法 Lerman は以下の手続きによって要約を生成する．

- (1) 入力文書集合に含まれる各文から評価情報を抽出する．
- (2) 要約長を超えない範囲でランダムに文を選択し、初期解を生成する．
- (3) 生成した初期解が含む文のいずれかを別の文と交換し、目的関数を改善する．これを目的関数が改善する限り貪欲に繰り返し、目的関数の改善が不能になった際に停止する．
- (4) 手続 (2) と (3) を 10,000 回繰り返し、最良の解を出力する．

§ 3 結果と考察

結果を表 3 に示す．表 3 の示すように、内容性において、手法 (1)、手法 (2) は、内容語のみを被覆の対象とした手法 Content、手法 Carenini、および手法 Lerman を上回っている．ウィルコクソンの符号付順位検定を実施したところ、手法 (1)、(2) は手法 Content、手法 Carenini、および手法 Lerman らの手法を ROUGE-2 において有意 ($p < 0.05$) に上回った．ROUGE-SU4 および ROUGE-SU9 では、手法 (1) および手法 (2) は手法 Content、手法 Carenini に対しては有意に上回ったが、手法 Lerman に対しては有意性を確認できなかった．

内容語のみを被覆の対象とした手法 Content は、評価文書の要約に含まれるべき情報である評価情報を正しく捉えることができない．

Carenini らの手法は、重要な評価属性をできる限り満

遍なく要約に含めることを意図しており、その点で本論文にて我々が提案する目的関数と近い意図を持っている．しかし、Carenini らは陽に目的関数を記述しておらず、要約を生成する手続きもアドホックなものである．それに対し、我々が提案する手法の利点は、評価文書の要約問題を、整数計画問題の一種である最大被覆問題として陽に表現することによって、最適化の分野で開発されてきた手法を用いて最適化を行う点にある．結果として、表 3 に示すように、既存の評価文書要約手法に比べて高い内容性を得ることに成功している^{*10}．

Lerman らの目的関数の一つの問題は冗長性である．彼らの目的関数は、評価情報の極性の分布だけでなく、評価情報の数の分布に関しても、入力文書集合と要約を近づけようとしている．例えば、入力文書集合に含まれる全ての評価情報の 4 割を、ある単一の評価属性に関する評価情報が占めていたとする．そのような場合、その評価属性に関する評価情報が 4 割を占めるような要約を Lerman らが提案した目的関数は高く評価する．同一の評価情報ばかりで構成された要約は冗長である．対して提案手法は重要な評価情報を過不足なく網羅することを意図しており、冗長な要約の生成を避けることができる．

提案手法 (1)、(2) の間に有意差はなかった．このことから、提案した文の選択と順序付けを同時に行う要約手法は、文の選択のみをまず行う場合と比べても、内容性において同様の品質の要約を生成することができることがわかる．

参照要約に対しては、表 3 が示すように、いずれの手法も及ばなかった．これは、人手によって作成された要約は、単に評価情報を網羅するものではなく、その評価の根拠など、評価情報に付随する他の情報も同時に要約に含めるためと思われる．例えば、図 1 に示した要約では「道がわかりづらい」ことの根拠として「小町通りから一本入った路地裏にある」といった情報が含まれているが、提案した要約モデルはこのような情報を陽に取り込む仕組みを持たない．人間に近い要約を生成するためには、要約に含めるべき情報の更なる吟味とそれを文書中から抽出するための仕組みが必要になるだろう．

6.2 要約の可読性

要約の可読性の評価のため、被験者実験を実施した^{*11}．6.1 節で説明した飲食店と商品各 100 セットの評価文書集合に対して、後述する 2 つの手法で要約を生成した．

評価は、生成した 2 つの要約を評価者に同時に提示し、より可読性に優れている要約を選択させる一対比較法に

*10 もちろん、モデルとデコードは切り離されて評価されるべきである．しかし、陽に目的関数が記述されていない場合や、目的関数の形式次第では最適化が困難となる．陽に、既存の最適化問題の一種として表現することも本手法の優位性である．

*11 要約の可読性の自動評価は端緒にすぎない [Pitler 10]、まだ確立したものとはいえない．

表 3 内容性の評価

	飲食店			商品		
	ROUGE-2	ROUGE-SU4	ROUGE-SU9	ROUGE-2	ROUGE-SU4	ROUGE-SU9
Content	0.240	0.278	0.255	0.161	0.199	0.185
Carenini	0.249	0.281	0.258	0.158	0.202	0.186
Lerman	0.260	0.296	0.268	0.205	0.247	0.227
手法 (1)	0.285	0.312	0.283	0.231	0.261	0.240
手法 (2)	0.283	0.313	0.281	0.233	0.258	0.245
Human	0.357	0.367	0.335	0.384	0.405	0.361

よって行った^{*12}。質問は、単にどちらかが可読性に優れているか選択させるのではなく、要約 A と B があったとしたとき、要約 A が優れている、どちらかというと言約 A が優れている、同程度、どちらかというと言約 B が優れている、要約 B が優れているの 5 段階から選択できるようにした。

評価者は 10 名の日本語母語話者であり、著者らは含まれていない。要約を評価する際には、それぞれの要約を生成した方法は評価者に通知されない。1 つの評価文書集合から生成された要約に対して 2 名の評価者が評価を行ったため、評価者は飲食店、商品それぞれのドメインにおいて 1 人あたり 40 個の要約対を評価した。

§ 1 比較手法

提案手法の可読性の評価のため、以下の 2 つの手法を比較する。

- 手法 (1): 文の選択と文の順序付けを別々に行う。これを手法 (1) とする。
- 手法 (2): 提案手法。文の選択と文の順序付けを同時に行うもの。これを手法 (2) とする。

これらの比較によって明らかにしたいのは以下の点である。

- 文の選択と文の順序付けを同時に行うことによる、要約の可読性への影響。文の選択と文の順序付けを同時に行うことで、それらを別々に行うことによって発生しうる読みづらい要約を減少させられる可能性がある。手法 (1) と手法 (2) を比較することで、それを検証する。

§ 2 結果と考察

結果を表 4 に示す。表 4 に示すように、飲食店、商品いずれの評価文書でも提案手法が可読性において大きく上回っていることが確認できる。

可読性が改善された要約の例を示す。図 3 に文の選択と順序付けを別個に行った場合の要約例、図 4 にそれらを同時に行った場合の要約例を示す。いずれも同じ入力から作られたものである。

*12 可読性の評価においては絶対評価が容易でないため、手法の差をより明らかにしやすい対比較法を選んだ。

表 4 可読性の評価

	飲食店	商品
手法 (1) がよい	18	21
どちらかというと言約 (1) がよい	20	17
同程度	32	41
どちらかというと言約 (2) がよい	43	23
手法 (2) がよい	87	98

おひつに入ったごはんなど、特別な気分も味わえ、お料理も和風で美味しかったです。小鉢も沢山でご飯の量も多くて大満足です。創作和食で食卓も美しく景色も美しい。お昼のおひつご飯もオススメ。いいお店。お店の雰囲気もよく、観覧車が目の前でとても綺麗でした。夜景がとっても綺麗です。

図 3 文の選択と順序付けを別個に行った場合の要約例。

文の選択と順序付けを同時に行うことで、最初に文の選択のみを行う場合とは異なる文が要約に選ばれていることがわかる。図 4 に示す、提案手法による要約例では最後に文「デートにお奨めの店です」が存在するが、図 3 にはその文は含まれていない。評価文書においては、末尾にその商品やサービスを勧められる状況等に関する言及がなされることがあり、その傾向を「お奨め」といった単語を通じて学習することで、より入手に近い要約を生成することができたものと考えられる。すなわち、接続スコアの学習によって、特徴量 (お奨め, < /d >) が高い重みを持つことによって、入手に類似した文の順序を再現することができる。文書の先頭と末尾以外の順序の接続スコアについては、特徴量 (良い, ランチ), (雰囲気, 個室), (雰囲気, デート) などが大きな重みを持っており、こういった特徴量を通じて文の順序が制御されている。つまり「雰囲気」という単語が使われている文の後に「個室」といった単語が使われる文が出現しやすいといった傾向を捉えることができています。

可読性を阻害する要因として、重複した話題の出現がある。図 5 に示す要約では「個室」に関する言及が 2 度要約に出現している。このような関連する話題が離れて出現することは読み手に対して奇異な印象を与えるが、

お店から見る夜景がとても綺麗でした。創作和食で食卓も美しく景色も美しい。小鉢も沢山でご飯の量も多くて大満足です。おひつに入ったごはんなど、特別な気分も味わえ、お料理も和風で美味しかったです。お店の雰囲気もよく、観覧車が目の前でとても綺麗でした。デートにお奨めのお店です。

図 4 文の選択と順序付けを同時に行った場合の要約例 1。

個室が多く、周りが騒がしくなくお酒が飲めるのが良かった。店の雰囲気もよかった。値段も手ごろ。ヘルシーな感じで体に良さそうな料理です。照明が暗く、デートにお勧めです。和食が大好きです。大事な接待で利用しました。噂どおりのすごい店。個室が素晴らしい。

図 5 文の選択と順序付けを同時に行った場合の要約例 2。要約の冒頭で述べられた「個室」に関する話題が要約の最後に再び出現してしまっている。

現在のモデルは文の順序に一次のマルコフ過程を仮定しているため、類似する話題が間に他の話題の文を挟んで再び出現する可能性がある。一つの措置は、「個室」に関する評価情報など、同一の評価情報は要約中に一度しか出現できないという制約を加えることである^{*13}。その場合、図 5 の最初の文あるいは最後の文いずれかのみが要約に含まれることになり、重複した話題の出現を避けることができる。しかし、この措置を加えることによって選択できる文の自由度が低下するため、内容性に対して悪影響を与える可能性がある。この悪影響は、図 5 の最初の文のような複文を単文に書き換えることで低減できる可能性があるが、これは将来の課題としたい。

より高次のマルコフ過程を仮定することでこれらの問題に対処することもできるが、これによって探索空間が指数的に広がるため、現実的ではない。他の解決策としてはドメイン特有の話題の順序を学習すること [Chen 09] が考えられるが、ドメインの粒度は自明ではない。

7. ま と め

本論文では、評価文書の要約をするための新しい要約モデルを提案した。本稿の新規性、および貢献を以下にまとめる。

- 文の選択と順序付けを同時に行う要約モデルを提案した。提案した要約モデルは要約が網羅する情報と、要約を構成する文同士の繋がりを最適化するものである。本論文では、提案した要約モデルを整数計画問題として表現し、これを解くことで要約を生成した。
- 提案した要約モデルを用いると、文の選択と順序付

けを別個に行う要約モデルと内容性において同等であり、可読性において優れた要約ができることを示した。

今後の課題としては、上で述べたように、まず評価情報以外の要約に含まれるべき情報の特定と、その抽出を組み込んだ要約モデルが挙げられよう。加えて、同一の評価情報が一度しか出現しないよう新しい制約を加えたモデルを評価する必要もある。

謝 辞

本論文に関して、NTT コミュニケーション科学基礎研究所の平尾努氏より様々なご助言を頂戴した。また、査読者の方々から有益かつ丁寧なご助言を頂戴した。記して感謝する。

◇ 参 考 文 献 ◇

- [Althaus 04] Althaus, E., Karamanis, N., and Koller, A.: Computing Locally Coherent Discourses, in *Proceedings of the 42nd Meeting of the Association for Computational Linguistics (ACL'04), Main Volume*, pp. 399–406 (2004)
- [Asano 08] Asano, H., Hirano, T., Kobayashi, N., and Matsuo, Y.: Subjective Information Indexing Technology Analyzing Word-of-mouth Content on the Web, *NTT Technical Review*, Vol. 6, No. 9 (2008)
- [Balas 89] Balas, E.: The prize collecting traveling salesman problem, *Networks*, Vol. 19, No. 6, pp. 621–636 (1989)
- [Banko 00] Banko, M., Mittal, V. O., and Witbrock, M. J.: Headline generation based on statistical translation, in *ACL '00: Proceedings of the 38th Annual Meeting on Association for Computational Linguistics*, pp. 318–325, Morristown, NJ, USA (2000), Association for Computational Linguistics
- [Barzilay 02] Barzilay, R., Elhadad, N., and McKeown, K. R.: Inferring strategies for sentence ordering in multidocument news summarization, *Journal of Artificial Intelligence Research*, Vol. 17, No. 1, pp. 35–55 (2002)
- [Barzilay 05] Barzilay, R. and Lapata, M.: Modeling local coherence: an entity-based approach, in *ACL '05: Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics*, pp. 141–148, Morristown, NJ, USA (2005), Association for Computational Linguistics
- [Barzilay 08] Barzilay, R. and Lapata, M.: Modeling local coherence: An entity-based approach, *Computational Linguistics*, Vol. 34, No. 1, pp. 1–34 (2008)
- [Carenini 06] Carenini, G., Ng, R., and Pauls, A.: MultiDocument Summarization of Evaluative Text, in *Proceedings of the 11th European Chapter of the Association for Computational Linguistics (EACL)*, pp. 305–312 (2006)
- [Chen 09] Chen, H., Branavan, S., Barzilay, R., and Karger, D. R.: Global Models of Document Structure using Latent Permutations, in *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pp. 371–379 (2009)
- [Crammer 06] Crammer, K.: Online Passive-Aggressive Algorithms, *Journal of Machine Learning Research*, Vol. 7, No. Mar, pp. 551–585 (2006)
- [Deshpande 07] Deshpande, P., Barzilay, R., and Karger, D.: Randomized Decoding for Selection-and-Ordering Problems, in *Human Language Technologies 2007: The Conference of the North American Chapter of the Association for Computational Linguistics; Proceedings of the Main Conference*, pp. 444–451, Rochester, New York (2007), Association for Computational Linguistics
- [Filatova 04] Filatova, E. and Hatzivassiloglou, V.: A Formal Model for Information Selection in Multi-Sentence Text Extraction, in *Pro-*

*13 ただし、文書の冒頭で概要を述べ、その後詳細を述べるといった構造を持つ文書においては同一の情報が複数出現することが許容されており、同一の情報が複数回出現することを必ずしも避けるべきとは言えない。

ceedings of Coling 2004, pp. 397–403 (2004)

- [Fuchi 98] Fuchi, T. and Takagi, S.: Japanese morphological analyzer using word co-occurrence: JTAG, in *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and the 17th International Conference on Computational Linguistics (ACL-COLING)* (1998)
- [Gillick 09] Gillick, D. and Favre, B.: A Scalable Global Model for Summarization, in *Proceedings of the Workshop on Integer Linear Programming for Natural Language Processing*, pp. 10–18 (2009)
- [Imamura 07] Imamura, K., Kikui, G., and Yasuda, N.: Japanese Dependency Parsing Using Sequential Labeling for Semi-spoken Language, in *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics Companion Volume Proceedings of the Demo and Poster Sessions*, pp. 225–228 (2007)
- [Korte 08] Korte, B. and Vygen, J.: *Combinatorial Optimization*, Springer-Verlag, third edition edition (2008)
- [Lapata 03] Lapata, M.: Probabilistic Text Structuring: Experiments with Sentence Ordering, in *Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics*, pp. 545–552, Sapporo, Japan (2003), Association for Computational Linguistics
- [Lapata 06] Lapata, M.: Automatic Evaluation of Information Ordering: Kendall's Tau, *Computational Linguistics*, Vol. 32, No. 4, pp. 471–484 (2006)
- [Lerman 09] Lerman, K., Blair-Goldensohn, S., and McDonald, R.: Sentiment Summarization: Evaluating and Learning User Preferences, in *Proceedings of the 12th Conference of the European Chapter of the ACL (EACL)*, pp. 514–522 (2009)
- [Lin 04] Lin, C.-Y.: ROUGE: A Package for Automatic Evaluation of Summaries, in *Proceedings of ACL Workshop Text Summarization Branches Out*, pp. 74–81 (2004)
- [National Institute of Standards and Technology 07] National Institute of Standards and Technology, : The linguistic quality questions (2007), <http://www-nlpir.nist.gov/projects/duc/duc2007/quality-questions.txt>
- [Pitler 10] Pitler, E., Louis, A., and Nenkova, A.: Automatic Evaluation of Linguistic Quality in Multi-Document Summarization, in *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pp. 544–554 (2010)
- [Radev 04] Radev, D. R., Jing, H., Styś, M., and Tam, D.: Centroid-based summarization of multiple documents, *Information Processing and Management*, Vol. 40, No. 6, pp. 919–938 (2004)
- [Takamura 05] Takamura, H., Inui, T., and Okumura, M.: Extracting Semantic Orientations of Words using Spin Model, in *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL)*, pp. 133–140 (2005)
- [Yih 07] Yih, W.-t., Goodman, J., Vanderwende, L., and Suzuki, H.: Multi-document summarization by maximizing informative content-words, in *IJCAI'07: Proceedings of the 20th international joint conference on Artificial intelligence*, pp. 1776–1782 (2007)
- [Zhang 08] Zhang, Y. and Clark, S.: Joint Word Segmentation and POS Tagging Using a Single Perceptron, in *Proceedings of ACL-08: HLT*, pp. 888–896 (2008)
- [奥村 05] 奥村 学, 難波 英嗣: テキスト自動要約, オーム社 (2005)
- [横野 10] 横野 光, 奥村 学: テキスト結束性を考慮した entity grid に基づく局所的一貫性モデル, 自然言語処理, Vol. 17, No. 1, pp. 161–182 (2010)
- [高村 08] 高村 大也, 奥村 学: 最大被覆問題とその変種による文書要約モデル, 人工知能学会論文誌, Vol. 23, No. 6, pp. 505–513 (2008)
- [平尾 06] 平尾 努, 奥村 学, 磯崎 秀樹: 拡張ストリングカーネルを用いた要約システム自動評価法, 情報処理学会論文誌, Vol. 47, No. 6, pp. 1753–1766 (2006)

[担当委員: 奥村 学]

2012 年 6 月 4 日 受理

著 者 紹 介

西川 仁(学生会員)

2006 年慶應義塾大学総合政策学部卒業。2008 年同大学大学院政策・メディア研究科修士課程修了。同年日本電信電話株式会社入社。2012 年より奈良先端科学技術大学院大学博士後期課程在学中。自然言語処理の研究開発に従事。言語処理学会, 情報処理学会各会員。

長谷川 隆明(正会員)

1992 年慶應義塾大学理工学部電気工学科卒業。1994 年同大学大学院理工学研究科計算機科学専攻修士課程修了。同年日本電信電話株式会社入社。2003 年～2004 年ニューヨーク大学客員研究員。現在, NTT メディアインテリジェンス研究所主任研究員。博士(工学)。自然言語処理, 情報アクセス技術の研究開発に従事。言語処理学会, 情報処理学会各会員。

松尾 義博

1988 年大阪大学理学部物理学科卒。1990 年同大学大学院研究科博士前期課程修了。同年日本電信電話株式会社入社。機械翻訳, 自然言語処理の研究に従事。言語処理学会, 情報処理学会各会員。

菊井 玄一朗(正会員)

1984 年京都大学工学部電気工学科卒。1986 年同大学大学院工学研究科修士課程電気工学第二専攻修了。同年日本電信電話株式会社 (NTT) 入社。1990 年～94 年株式会社国際電気通信基礎技術研究所 (ATR)。2001 年 ATR 音声言語コミュニケーション研究所研究室長。2006 年 NTT サイバースペース研究所音声・言語基盤技術グループリーダー。2011 年岡山県立大学情報工学部情報システム工学科教授。現在に至る。博士(情報学)。言語処理学会, 情報処理学会各

会員。