情報処理学会論文誌 Vol.57 No.2 769–782 (Feb. 2016)

識別的半マルコフモデルによるテキスト結束性を考慮した単一文書要約

西川 仁[1,†1,a)] 有田 一穂[1,b)] 田中 克己[1,c)] 平尾 努[1,d)] 牧野 俊朗[1,e)] 松尾 義博[1,f)]

受付日 2015年4月16日, 採録日 2015年11月5日

概要：本論文では，隠れ半マルコフモデルによる単一文書要約手法を提案する．我々は，単一文書要約を，長さに関する制約の下で，所与の目的関数を最大化する文の系列を，入力文書から得た文集合から探索する問題と見なす．提案する手法は文を選択する際に文間の結束性を考慮することができ，さらに文短縮を組み込むこともできる．提案手法の有効性を評価するため，我々は 12,748 組もの文書とその参照要約からなる大規模な単一文書要約コーパスを用意した．提案手法は他のベースラインの多くを ROUGE および言語的品質による評価において上回った．提案手法によって生成された要約の約 20%が人間が作成した参照要約と完全に同一であった．また，大規模な単一文書要約コーパスが要約器の訓練に際して重要であることも明らかにした．

キーワード：自動要約，単一文書要約，隠れ半マルコフモデル

Learning to Generate Cohesive Summary with Discriminative Hidden Semi-Markov Model

Hitoshi Nishikawa[1,†1,a)] Kazuho Arita[1,b)] Katsumi Tanaka[1,c)] Tsutomu Hirao[1,d)] Toshiro Makino[1,e)] Yoshihiro Matsuo[1,f)]

Received: April 16, 2015, Accepted: November 5, 2015

Abstract: In this paper we introduce a novel single-document summarization method based on a hidden semi-Markov model. This method can naturally model single-document summarization as the optimization problem of selecting the best sequence from among the sentences in the input document under the given objective function and knapsack constraint. This advantage makes it possible for sentence selection to take the cohesion of the summary into account. In addition our method can also incorporate sentence compression into the summarization process. To demonstrate the effectiveness of our method, we conduct an experimental evaluation with a large-scale corpus consisting of 12,748 pairs of a document and its reference. The results show that our method significantly outperforms the competitive baselines in terms of ROUGE evaluation, and the linguistic quality of summaries is also improved. Our method successfully mimicked the reference summaries, about 20 percent of the summaries generated by our method were completely identical to their references. Moreover, we show that large-scale training samples are quite effective for training a summarizer.

Keywords: automatic summarization, single-document summarization, hidden semi-Markov model

1. はじめに

単一文書要約は，商業的な文脈において，より良い情報アクセスを提供するための鍵となる技術になりつつある．

[1] 日本電信電話株式会社
Nippon Telegraph and Telephone Corporation, Yokosuka, Kanagawa 213–0033, Japan
[†1] 現在，東京工業大学大学院情報理工学研究科計算工学専攻
Presently with Department of Computer Science, Graduate School of Information Science and Engineering, Tokyo Institute of Technology
[a)] hitoshi@cs.titech.ac.jp
[b)] arita.kazuho@lab.ntt.co.jp
[c)] tanaka.katsumi@lab.ntt.co.jp
[d)] hirao.tsutomu@lab.ntt.co.jp
[e)] makino.toshiro@lab.ntt.co.jp
[f)] matsuo.yoshihiro@lab.ntt.co.jp

The Financial Times [*1] と CNN [*2] は，ウェブサイトに掲載している記事にその要約をあわせて提示するようにしており，ウェブサイトにユーザを誘引する手段にしている．また，Yahoo! に買収された Summly は，ウェブから収集した記事に自動生成した要約を添えることで注目を浴びた．人手で要約を作成する費用は大きいため，自動的に人間に近い品質の要約を生成できれば，インターネットを利用する多くのユーザの情報アクセスを大きく改善することができると期待できる．

人間が作成した要約を模倣するための 1 つの重要な要素として，テキストの結束性がある [29]．テキスト結束性は複数文書要約の文脈でこれまで広く取り上げられてきたものの [7], [44]，単一文書要約においては必ずしも十分に取り上げられてこなかった．本論文では，単一文書要約におけるこのテキスト結束性の問題を取り上げ，これを利用し，情報量に富み，また言語的品質の高い要約を生成する手法を提案する．

本論文では，隠れ半マルコフモデルに基づく新しい単一文書要約手法を提案する．この要約手法は，単一文書要約における代表的な要約手法である，ナップサック問題に基づく要約手法と，隠れマルコフモデルに基づく要約手法の性質を兼ね備えたものである．前者は，要約としてふさわしい文の最良の組合せを求めることができ，後者は要約としてふさわしい文を選択する際に文の周辺の文脈を考慮することができる．これらの性質を兼ね備えた要約手法を用いることによって，我々の要約器はテキストの結束性を考慮しつつ，要約長以内の要約を生成することができる．

本論文の新規性および貢献は以下の点にある．

- 我々は単一文書要約を隠れ半マルコフモデルに基づく組合せ最適化問題として定式化し，解を得るためのアルゴリズムを提案する．
- 我々の提案する要約手法はテキスト結束性に関する様々な種類の特徴量を考慮することができ，識別学習を用いることで言語的品質においてより優れた要約が生成できることを示す．
- 我々は大規模な単一文書要約コーパスを用意し，これを利用して要約器を訓練することで高品質な要約が生成できることを示す．

本論文は以下のように構成される．2 章では，関連研究について述べる．3 章では，我々の提案する要約手法を詳説する．また，パラメータの学習についても説明する．4 章では，文短縮によって文の亜種がどのように生成されるかを説明する．5 章では，解の探索の方法について説明する．6 章では，提案手法を評価するための実験の設定について述べる．7 章では，実験の結果について述べる．8 章では本論文をまとめる．

[*1] http://www.ft.com/
[*2] http://www.cnn.com/

2. 関連研究

2.1 単一文書要約

単一文書要約は，多くの場合，要約としてふさわしい文（以降，重要文とする）を選択する問題として定式化される [29]．要約の対象となる文書は文分割器によって文の集合に分解され，要約器は，その集合の部分集合の中で，要約の長さに関する制約（以降，要約長とする）を満たし，かつ何らかの目的関数を最大にするものを要約として選び出す．

McDonald は，自動要約がナップサック問題として定式化できることを指摘した[*3] [27]．ナップサック問題は組合せ最適化問題の 1 種である [20]．価値と容量を持つ何らかの要素の集合が与えられたとき，ナップサック問題は，別途与えられた，容量に関する制約であるナップサック制約を満たし，かつ，要素の価値の和が最大になるように，要素の集合の部分集合を選び出す問題である．自動要約の文脈においては，要約長をナップサック制約として，与えられた文集合の要素となる文それぞれに何らかの重要度を与えれば，その集合から，重要度の和が最大となる部分集合を選び出す問題として，ナップサック問題に基づく要約手法（以降，ナップサック手法とする）を考えることができる [43], [46]．ナップサック問題は動的計画法を用いて擬似多項式時間での求解が可能である [20] ため，実用上は十分高速に最適解が得られる．平尾らはナップサック手法を，文短縮によって生成した元の文の亜種を組み込めるように拡張し，その有効性を示した [46]．西川らはナップサック手法を複数文書要約に用いるため，ナップサック手法において問題となる冗長性を削減するための仕組みをナップサック手法に組み込んだ手法を提案した [44]．これらの手法の問題点は，文間の結束性を考慮する仕組みを持たない点にあり，そのため生成された要約の言語的品質が劣化することがある．本論文で提案する隠れ半マルコフモデルに基づく要約モデルはこの問題を解決する．隠れ半マルコフモデル [39] はナップサック問題を拡張したものとみることができ，我々はこの性質を単一文書要約に利用する．ナップサック問題と隠れ半マルコフモデルの関係については 3 章で詳述する．

要約の言語的品質を向上させるための 1 つの方法は要約の談話構造を整理することである [29]．談話構造は大きく一貫性と結束性の 2 つの観点から議論することができる [42]．この一貫性と結束性にそれぞれ対応する形で，談話構造を考慮した要約を作成するためには大きく 2 つの方法が存在する．1 つは文書の木構造表現に基づく方法であり，もう 1 つは文書の結束性に基づく方法である．

[*3] ただし，McDonald は複数文書要約の文脈でこの点を指摘しており，彼自身は単一文書要約にナップサック問題に基づく要約手法を用いてはいない．

文書の木構造表現に基づく方法は，修辞構造理論 [25] に基づいて構築された文書の木構造表現を手がかりとして要約を生成する [11], [14], [18], [26]．この方法は，文書を構成する各文の文書全体における機能的な役割を考慮して要約を生成することができる．そのため，生成された要約は首尾一貫したものになりやすい．一方，この方法の弱点は，事前に談話構造解析器を用いて文書の木構造表現を明らかにしなければならない点にあり，そのためこの方法に基づく要約の精度は談話構造解析器の精度に強く依存する．

もう 1 つの方法は，大域的な文書の構造を利用するのではなく，局所的な文書の結束性を利用するものである [6], [35]．この方法の強みの 1 つは，木構造表現に基づく方法とは異なり，談話構造解析を必要としない点にあり，したがって前者の方法よりも頑健であることが期待できる．この理由から，本論文では文書の局所的結束性に基づく方法を採用する．

本論文で提案する手法に最も近い手法は，Barzilay らが提案した隠れマルコフモデルに基づく手法 [6] と Shen らが提案した条件付き確率場に基づく手法である [35]．これらの手法は要約を線形タグ付け問題と見なすことにより，文書の局所的な結束性あるいは構造を加味しつつ重要文を同定する．Barzilay らは隠れマルコフモデルを利用し，特定の分野の文書の局所的な構造を学習し，学習したモデルを重要文の選択に利用した [6]．Shen らは隠れマルコフモデルに基づく手法を拡張し，条件付き確率場 [21] に基づく手法を提案した [35]．条件付き確率場は重要文を同定するために様々な特徴量を導入することができ，彼らはその有効性を示している．また，Wu らは，条件付き確率場を半マルコフ過程へと拡張した半マルコフ条件付き確率場を用いた要約の手法を提案している [38]．

これらの手法の弱点は，これらの手法が本質的には文を単に分類しているにすぎず，要約長に関する制限を直接考慮できない点にある．これらの手法は文の系列に対してある種の重要度を与えるにすぎず，重要文として選択された文を，要約長を度外視して出力することになる．したがって，これらの手法においては，隠れマルコフモデルや条件付き確率場は要約長を加味せずに最適化されており，実際に要約長が与えられて要約を生成するときに向けて最適に学習されているとは限らない．これらの手法とは異なり，我々の提案する要約手法は要約長の制限を自然に考慮しつつ，文間の結束性を加味することができる．

これらと別に，Clarke らは，センタリング理論を援用して文書の品質をできる限り落とさずに文書を圧縮する手法を提案している [8]．この手法は，彼ら自身も指摘するように，要約長を陽に考慮する仕組みを持たず，したがって何文字以内の要約を生成するといったことができない．そのため，彼ら自身，その手法を要約ではなく圧縮と呼んでいる．彼らの手法はいわゆる要約手法ではないことから，本論文では比較対象とはしない．

表 1 結束性に問題のある要約の例
Table 1 A summary with a cohesive problem.

入力文書	A 首相は 20 日未明，B 国に到着した．明日 21 日には同国 C 大統領との会談が予定されている．D 官房長官は会見にて，「これは同国の友好関係を更なる段階へ進歩させる最初の一歩となる」と述べた．
要約	明日 21 日には同国 B 大統領との会談が予定されている．D 官房長官は会見にて，「これは同国の友好関係を更なる段階へ進歩させる最初の一歩となる」と述べた．

まとめると，要約長を考慮して要約を生成できるものの文間の結束性を加味できないナップサック問題に基づく要約手法と，文間の結束性を加味できるが要約長は考慮できない線形タグ付け問題に基づく要約手法とがあり，我々の提案する隠れ半マルコフモデルに基づく要約手法はこれらのトレード・オフを両立させるものになっている．

2.2 テキスト結束性

テキストの結束性 [42] は複数文書要約の文脈において頻繁に取り上げられてきた．これは，複数文書要約においては，複数の異なる文書から重要文が選び出されるため，これらを 1 つの文章としてまとめあげるためにテキストの結束性が重要になるためである [29]．自動要約において結束性が問題となる典型的な例は指示表現に関するものである．**表 1** に例を示す．例では，入力文書は 3 つの文から構成されており，要約は 3 つの文のうち 2 文めと 3 文めから構成されている．要約の 1 文めは，入力文書の 2 文めであり，そのため「会談」の主語が省略されており，B 大統領と会談を行う人物が不明である．このような要約からは適切に情報を読み取ることができないため，こういった要約が生成されることを避ける何らかの工夫が必要である．

要約の結束性の向上させる典型的な方法は，重要文として選択された文に何らかの方法で順序を与えるものであり，これは文の順序付けと呼ばれる課題である．文の順序付けを行うため，これまで様々な特徴量が提案されている [1], [4], [17], [32]．

いくつかの研究は複数文書要約において重要文の選択と順序付けを同時に行うことを提案している [7], [44]．重要文の選択と順序付けを同時に行う問題は巡回セールスマン問題と見なすことができる [2]．そのため，最適解の探索が難しく，解の探索は重要な課題である．西川らは問題を整数計画問題（以降，ILP とする）として表現し ILP ソルバを利用して問題を解き [44]，Christensen らは局所探索を用いて近似解を求めた [7]．前者の手法は最適解を発見できるものの，一般に多大な時間を要し[*4]，後者の手法は素早く解を求めることができるものの発見された解が最適解である保証がない．これらの手法とは異なり，我々の提案す

[*4] もちろん，探索の途中の解を出力することはできる．

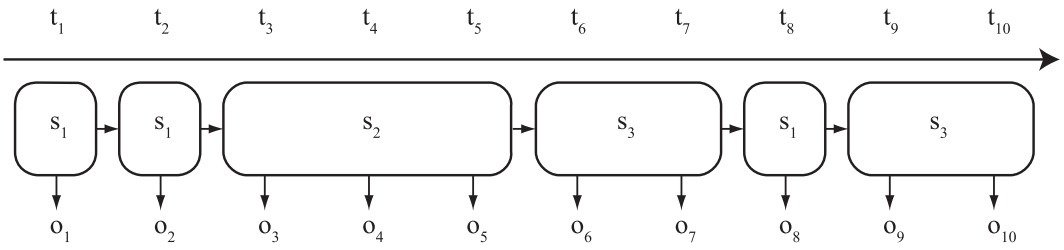


図 1 隠れ半マルコフモデルの例．システムが時間 $t_1 ... t_{10}$ にわたって 10 個の記号 $o_1 ... o_{10}$ を観測し，3 種類の状態 $s_1 ... s_3$ の間を遷移するものとする．通常の隠れマルコフモデルとは異なり，隠れ半マルコフモデルにおいては，それぞれの状態は単位時間を超えて継続することができる．この図では，状態 s_2 と状態 s_3 が単位時間を超えて継続しているため，システムが 10 個の記号を観測したにもかかわらず，システムが遷移した状態の延べ数は 6 つにとどまっている

Fig. 1 An example of a hidden semi-Markov model.

る，動的計画法に基づくアルゴリズムは，実用上十分高速に解を求めることができる[*5]．

3. 隠れ半マルコフモデルに基づく要約モデル

本章ではまず単一文書要約を自然に模擬できるナップサック問題に基づく要約手法を説明する．次に，隠れ半マルコフモデルに基づく要約手法を説明し，これがナップサック問題の拡張となっていることを示す．そののち，我々の提案する要約手法について詳説する．

3.1 ナップサック問題に基づく要約モデル

単一文書要約はナップサック問題として定式化できる．入力された文の最適な組合せは，文の重要度と長さを加味し，与えられた要約長以内で重要度が最大となる文の集合を，動的計画ナップサックアルゴリズムで選び出すことによって得られる．

3.2 隠れ半マルコフモデル

隠れ半マルコフモデル[*6]は隠れマルコフモデルを拡張したものである [39]．一般的な隠れマルコフモデルでは，1 つの状態と単位時間が直接結び付いており，1 つの状態は 1 単位時間のみ継続する．たとえば，システムが 10 の離散的な記号からなる系列を観測したとすると，システムはそれに対応する 10 の隠れ状態を遷移する．一方，隠れ半マルコフモデルにおいては，1 つの状態が 1 単位時間以上継続することが許される．たとえば，あるシステムが 10 の離散的な記号からなる系列を観測したとき，1 つの隠れ状態がたとえば 2 単位時間継続するということが可能であるため，この場合であればシステムは 5 つの隠れ状態を遷移する．例を**図 1** に示す．このシステムは，10 の記号からなる系列を観測し，3 種類の隠れ状態の間を遷移している．図 1 では，状態 s_2 と s_3 とが 1 単位時間以上継続している．そのため，このシステムは 10 の記号からなる系列を観測したにもかかわらず，延べ 6 つの状態のみを遷移している．この性質は線形タグ付け問題に広く利用されており，固有表現認識 [34] や文字認識 [36]，音素認識 [19] などに応用されている．

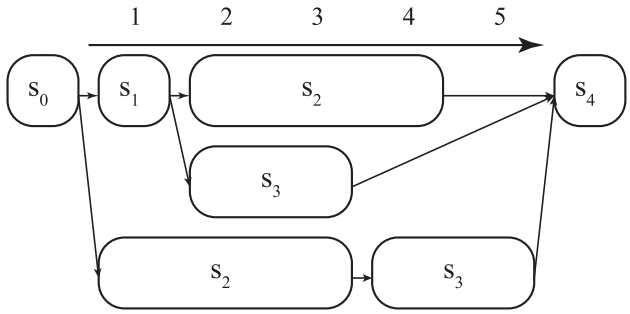


図 2 隠れ半マルコフモデルによる文選択の例

Fig. 2 An example of sentence selection with a hidden semi-Markov model.

隠れ半マルコフモデルはナップサック問題と密接に関係している．システムが観測した記号系列の長さ K はナップサック制約と見なすことができる．隠れ半マルコフモデルにおいては，システムは，観測した記号列の長さが許す限り，尤度を最大化するように隠れ状態を遷移しながら，隠れ状態を記号系列に詰め込んでいるものと見なすことができる．たとえば，10 文からなる文書を要約することを考え，文の長さは文の文字数で測られるものとする．この場合，システムは 10 の文に対応する 10 種類の状態を，要約長の制限が許す限り遷移し，遷移先となった文はすなわち重要文となる．それぞれの状態の継続時間は対応する文の長さとなり，システムが新しい状態に遷移するたびにその状態に対応した長さだけシステムは記号系列上を前進する．

図 2 に隠れ半マルコフモデルによる文選択の例を示す．s_1，s_2，s_3 の 3 つの文が選択の対象としてあり，s_0 と s_4

[*5] これは，複数文書要約とは異なり，単一文書要約では通常，文の順序を入れ替えることを考える必要がないため，探索空間が劇的に小さくなるためである．

[*6] 隠れ半マルコフモデルは Hidden Semi-Markov Model の訳であるが，「半分」を意味する接頭辞 Semi の訳出については一貫しておらず，半，準，他にそのまま片仮名でセミとするものがある．本論文では本来の語義に鑑み，また Semi-Supervised Learning は半教師あり学習と訳出されることが多いことから，半とした．

はそれぞれ文書の先頭と末尾を示す仮想的な文とする．さらに，文 s_1，s_2，s_3 はそれぞれ長さ 1，3，2 を持つとする．このとき，長さ 5 の要約を文選択に基づいて作成することを考えると，選択可能な文の組合せは s_1 と s_2，s_1 と s_3，s_2 と s_3 の 3 種類である．文を長さの制限の中で選択する過程は図 2 に示すように，それぞれの文を 1 つの状態と見なし，状態を遷移していく過程と見なせる．上で述べたように，隠れ半マルコフモデルは各状態が 1 単位時間以上継続することが許されており，このことはすなわち各文，すなわち各状態が対応する長さの分だけ遷移の過程で継続することに対応している．各要約のある種のスコアは，各状態の出力確率および遷移確率から計算することができる．各状態の出力確率はそれぞれの状態，すなわち文の重要度と見なすことができる．また，各状態間の遷移確率，すなわちある文から他の文への遷移確率は，文間の結束性を表すものと見なすことができる．このように，隠れ半マルコフモデルは文選択を状態間の遷移として自然に表現することができる．

なお，本来，隠れマルコフモデルは生成モデルであり，出力確率および遷移確率は同時確率に基づいて求められるが，Collins はこれを識別的に拡張し出力確率および遷移確率を求める方法を提案している [9]．隠れ半マルコフモデルも同様に識別的に拡張することができ [19]，我々も最大マージン法によって識別的に訓練した隠れ半マルコフモデルを用いる．

3.3 定式化

要約の対象となる文書 x を構成する n 個の文 $s_1, s_2, ..., s_n$ があるものとする．これらの文の長さはそれぞれ $\ell_1, \ell_2, ..., \ell_n$ で表されるものとする．文の長さは文字数，単語数，バイト数などで測ることができるが，本論文では文字数で測るものとする．これは今回用意した単一文書要約コーパスを構成する参照要約の長さが文字数によって規定されているためである．それぞれの文は重要度 $w_1, w_2, ..., w_n$ を持ち，この重要度が高い文ほど要約を構成するにふさわしいものであるとする．それぞれの文には文短縮器や言い換え器などによって生成される亜種があるものとし，文 s_i の m_i 種類の亜種をそれぞれ $s_{i,1}, s_{i,2}, ..., s_{i,m_i}$ と書く．これらの亜種も同様に長さと重要度を持っているものとし，それらはそれぞれ $\ell_{i,1}, \ell_{i,2}, ..., \ell_{i,m_i}$ と $w_{i,1}, w_{i,2}, ..., w_{i,m_i}$ で表される．以降，簡単のために元の文 $s_1, s_2, ..., s_n$ をそれぞれ $s_{1,0}, s_{2,0}, ..., s_{n,0}$ と表記する．したがって，元の文 $s_{i,0}$ とその亜種 $s_{i,1}, s_{i,2}, ..., s_{i,m_i}$ が存在することになる．$s_{0,0}$ と $s_{n+1,0}$ をそれぞれ文書の先頭と末尾を表す仮想的な文とする．それぞれの文の間には文どうしの結束性の強さを表す結束度が定義されるものとし，文 $s_{g,h}$ と文 $s_{i,j}$ の結束度を $c_{g,h,i,j}$ と表す．出力となる要約は文の系列 y として表現され，入力された文の集合とそれらから生成された文の亜種の集合から生成されうる文の系列の全体を Y とする．すなわち $y \in Y$ である．最後に，K を要約長とする．これらの記法に基づくと，我々の提案する要約手法は以下のように表現される：

$$y^* = \underset{y \in Y}{\operatorname{argmax}} \sum_{s_{i,j} \in sent(y)} w_{i,j} + \sum_{(s_{g,h}, s_{i,j}) \in adj(y)} c_{g,h,i,j} \quad (1)$$

$$s.t. \sum_{s_{i,j} \in sent(y)} \ell_{i,j} \leq K. \quad (2)$$

ここで，$sent(y)$ と $adj(y)$ はそれぞれ文の系列 y を構成する文の集合と，文の系列 y の中で隣接している 2 つの文の組の集合を示すものとする．すなわち，我々の提案する要約手法では，入力された文の集合とそれらから生成される亜種の集合から構成されるうる系列のうち，要約長に関する制約を満たし，かつ重要度と結束度の和が最大のもの y^* を最良の要約と見なす．

3.4 パラメータの最適化

我々の提案する要約手法は 2 種類のパラメータ $w_{i,j}$ と $c_{g,h,i,j}$ とを持っており，これらの値を何らかの方法で定める必要がある．本節ではこれを定める方法について述べる．まず，$w_{i,j}$ と $c_{g,h,i,j}$ とを以下のように定める：

$$w_{i,j} = \mathbf{w}_w \cdot \mathbf{f}_w(x, s_{i,j}) \quad (3)$$

$$c_{g,h,i,j} = \mathbf{w}_c \cdot \mathbf{f}_c(x, s_{g,h}, s_{i,j}). \quad (4)$$

ここで，$\mathbf{f}_w$ と $\mathbf{f}_c$ はそれぞれ，d_w 次元と d_c 次元の，文書 x における文，文の間の結束性を表す特徴ベクトルである．$\mathbf{w}_w$ と $\mathbf{w}_c$ はそれぞれ d_w 次元と d_c 次元の重みベクトルである．パラメータの最適化の目的はこれら $\mathbf{w}_w$ と $\mathbf{w}_c$ の値を，特徴ベクトル $\mathbf{f}_w$ と $\mathbf{f}_c$ に基づいて定めることである．簡単のため，y を要約を表すものとし，$\mathbf{f} = \langle \mathbf{f}_w, \mathbf{f}_c \rangle$ を $(d_w + d_c)$ 次元の要約 y 全体の特徴ベクトルとする．$\mathbf{w} = \langle \mathbf{w}_w, \mathbf{w}_c \rangle$ は $(d_w + d_c)$ 次元の重みベクトルとする．このとき，要約 y の目的関数の値は $\mathbf{w} \cdot \mathbf{f}(x, y)$ となる．

パラメータを最適化するために，本論文では構造学習手法として広く利用されている Passive-Aggressive アルゴリズム [10] を用いる．損失関数に ROUGE [23] を用いると，パラメータ最適化の学習は以下の等式を繰り返し解くことで行われる：

$$\mathbf{w}^{new} = \underset{\mathbf{w}}{\operatorname{argmin}} \frac{1}{2} ||\mathbf{w} - \mathbf{w}^{old}||^2 \quad (5)$$

$$s.t. \ \mathbf{w} \cdot \mathbf{f}(x, r) - \mathbf{w} \cdot \mathbf{f}(x, y) \geq \text{loss}(r, y).$$

ここで，$\mathbf{w}^{new}$ は更新後の重みベクトルであり，$\mathbf{w}^{old}$ は更新前の重みベクトルである．r は参照要約であり，loss は損失関数である．前述したように，損失関数は ROUGE に基づいて，$\text{loss}(r, y) = 1 - \text{ROUGE}(r, y)$ と定義する．$\text{ROUGE}(r, y)$ は参照要約 r が与えられた際の，y の ROUGE スコアである．ROUGE にはいくつかの亜種があるが，そのなかの ROUGE-1 を用いる

3.4.1 文の重要度の特徴量

ここでは文の重要度 $w_{i,j}$ を定めるために用いる特徴量について述べる．

単語の頻度：単語の頻度は自動要約の古典的な特徴量である [24]．本論文では文を構成する内容語それぞれの入力文書中での頻度を数え，それらの和を特徴量として用いる．すなわち，入力文書において頻繁に出現する単語を多く持つ文は，この特徴量において高い値を持つ．

単語の表記：文を構成する内容語の表記およびそれらの品詞も特徴量として用いる．

固有表現：人名や組織名といった固有表現とそのクラス名を特徴量として用いる．

文の長さ：文の長さも特徴量として用いる．文の長さは文字数によって数えられる．

文の位置：単語の頻度と同様に，文の位置も自動要約の古典的な特徴量である [24]．それぞれの文の，文書中での位置，段落の先頭にあるか否か，段落中での位置の 3 つを文の位置に関する特徴量として用いる．文書中での位置および段落中での位置は 0 から 1 の値に正規化した．

3.4.2 文間の結束度の特徴量

ここでは文間の結束度 $c_{g,h,i,j}$ を定めるために用いる特徴量について述べる．

語彙の遷移：Lapata は，隣接する 2 つの文それぞれを構成する単語の集合の直積集合を考え，これを用いて文を適切な順序に並べることができることを示した [22]．本論文でもこの特徴量を文間の結束性をとらえるために用いる[*7]．具体的には，文 $s_{g,h}$ と文 $s_{i,j}$ からそれぞれ内容語を取り出し，これらの直積集合の要素を特徴量として利用する．

語彙的結束性：Pitler らは 2 つの文の類似度がテキストの言語的品質を評価するための強い指標になっていることを示した [33]．この知見に基づき，我々も 2 つの文のコサイン類似度を計算し，その値を特徴量として用いる．

エンティティ・グリッド：Barzilay らは文中に現れる名詞句の統語的な役割の変化がテキストの結束性[*8]を評価するための重要な指標になることを指摘した [5]．Pitler らはこれをテキストの言語的品質を予測するモデルの特徴量として用い，この指標の有効性を実証している [33]．本論文でもこれを特徴量として用いる．エンティティ・グリッドは元来，英語で書かれた文のために考案されたものだが，日本語で書かれた文での利用については横野らの提案がある [41]．本論文ではこれに基づいて特徴量を実装した．なお，本来のエンティティ・グリッドは一度文書全体の名詞句の役割変化の分布を求め，この分布をベクトルとして表現するが，本論文では局所的な 2 つの文に対してこのベクトルを作成し，これを特徴量とした．すなわち，隣接する 2 つの文それぞれに現れる名詞句の役割がどのように変化したかを特徴量として用いる．これは，隠れ半マルコフモデルにおいて長距離の依存関係をとらえようとすると著しくデコードが困難になるためである[*9]．

[*7] Barzilay らは密接な関係を持つ 2 つの単語の組は重要文選択に寄与することを示しており [3]，そのためこの特徴量は文選択にも寄与するものと考えられる．

[*8] Barzilay らは文献 [5] においてエンティティ・グリッドがとらえる性質を局所的一貫性 Local Coherence の語で表現しているが，本論文では田窪らの分類 [42] および横野らの記述 [41] に従い，この性質を結束性に関するものと見なす．

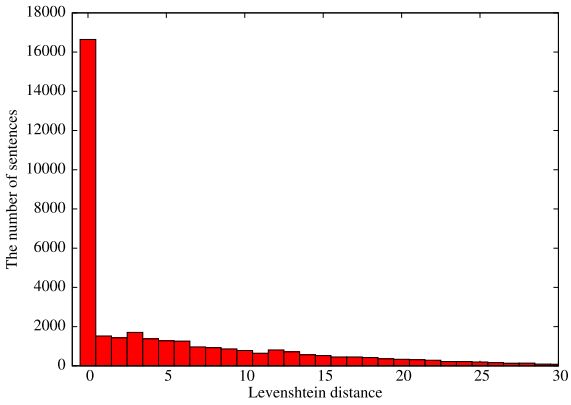


図 3 対応づけられた文どうしにおける編集距離の分布．参照要約を構成する 36,413 文のうち 16,643 文は入力文書のそれとまったく同一，すなわち編集距離が 0 であった

Fig. 3 Distribution of edit distance between aligned sentences.

4. 文の亜種の生成

本論文で提案する要約モデルは元の文とは異なる文を選択の対象とすることもできる．そこで，平尾ら [46] と同様に，文の亜種を生成しておき，これを選択の対象に含める．

本論文では 2 種類の方法で文の亜種を生成する．文が含む括弧およびその内部のテキストを除去する方法と，元の文の係り受け木を枝刈りする方法である．これら方法は単純なものであるが，参照要約を調査したところ，この比較的単純な操作でも参照要約に近いものが生成できると思われた．調査は入力文書に含まれる文と，参照要約に含まれる文を対応づけることで行った[*10]．調査によれば，参照要約の作成者は文に対してあまり大きな修正を加えておらず，参照要約を構成する 36,413 文のうち，半数弱の 16,643 文は入力文書に含まれる文とまったく同一であった．また，**図 3** に示すように，書き換えが加えられている文においても，多くの文は，要約長に収まるように数語程度，重要ではないと思われる情報が除去されている程度であり，大き

[*9] なお，本来のエンティティ・グリッドや，2 章で述べた文書の木構造表現など，大域的な特徴量を組み入れる方法としては，局所的な特徴量のみでデコードを行い，デコードの結果から n-best 解を得て，それらを大域的な特徴量でリランキングするといった方法が考えられる．一方で，n-best の n の値の設定など，別途設定が必要なハイパーパラメタが生じるため，本論文ではエンティティ・グリッドを局所的な特徴量として組み入れる形とした．

[*10] 文の対応づけは，まず入力文書のすべての文と参照要約すべての文の編集距離を計算し，そののち編集距離の和を最小にするように文どうしを対応づけた．対応づけは動的計画法で行うことができる．

く修正されているものは少ない．

具体的には以下のように文の亜種を生成する．

(1) 文内の，括弧およびその内部のテキストを除去する．いくつかの文は，付加的な情報を読者に提示するために括弧で囲われたテキストを含んでいる．

(2) 文の係り受け木を枝刈りする．この方法は Nomoto の提案した方法と基本的には同一であるが [31]，必須格を除去した場合に非文が生成される恐れがあるため，動詞とその必須格に関する情報を保持した辞書を用意し，必須格が除去されないようにした．ある係り受け木の部分木は無数に存在するため，N グラムに基づく言語尤度と，係り受けに基づく言語尤度を用いて，これらが高いものから順に 1 位から 10 位までを出力させるようにした．

5. 動的計画法による解の探索

式 (1) を，式 (2) の制約の下で解くことが本章の主題である．動的計画法に基づくアルゴリズムの疑似コードを Algorithm 1 に示す．1 行めから 7 行めでは利用する変数の初期化を行う．ベクトル $\mathbf{z} = \langle z_0, ..., z_{n+1} \rangle$ は，どの文が要約として選択されているかを格納する．もし $z_3 = 2$ であれば，文 $s_{3,2}$ が要約に含まれる．$z_3 = -1$ であれば，文 $s_{3,0}$ およびその文の亜種は要約に含まれない．V と P，S は 2 次元の配列であり，それぞれアルゴリズムの途中の結果を保存する．$V[i][k]$ は 2 次元の配列 V の i 行 k 列の時点で文のどの亜種が選択されるか，あるいは文が選択されないかを保存する．もし $V[i][k] = 0$ であれば，元の文 $s_{i,0}$ が i 行 k 列の時点で採用される．もし $V[i][k] = -1$ であれば，文は選択されない．$P[i][k]$ は現在の文の直前に選択されている文へのポインタを格納する．$S[i][k]$ は i 行 k 列時点での目的関数の値である．

8 行めから 28 行めまでの疑似コードは，以下の漸化式を解いていることに相当する：

$$S[i][k] = \begin{cases} \max_{h=0...i-1, v=0...m_i} S[h][k-\ell_{i,v}] + \\ \qquad w_{i,v} + c_{h,V[h][k-\ell_{i,v}],i,v} & \text{(A)}, \\ S[i][k-1] & \text{(B)}. \end{cases}$$

ここで，A は $\ell_{i,v} \leq k \wedge S[i][k-1] \leq S[h][k-\ell_{i,v}] + w_{i,v} + c_{h,V[h][k-\ell_{i,v}],i,v}$ の場合であり，B はそれ以外の場合である．

A の場合は，i 行 k 列の時点において，要約文 $s_{i,v}$ を要約に追加する場合である．動的計画法の過程において，i 行 k 列に位置する際に行うことは，文 1 から文 i までのみを用いて，要約長 k[*11]の要約を作成する際に，最後の文として文 $s_{i,v}$ を採用するかを決定することに相当する．まず，長さ $\ell_{i,v}$ の文 $s_{i,v}$ を要約に追加するためには，要約長 k が文 $s_{i,v}$ の長さ $\ell_{i,v}$ 以上である必要がある．$\ell_{i,v} \leq k$ は

[*11] 本来の要約長 K ではなく，あくまで動的計画法の途中であるため，k であることに注意されたい．

Algorithm 1 動的計画法によるデコード

```
 1: Z = ⟨z_0, ..., z_{n+1}⟩
 2: for i = 0 to n + 1 do
 3:     z_i = −1
 4:     V[i][0] ← −1
 5:     P[i][0] ← −1
 6:     S[i][0] ← 0
 7: V[0][0] = 0
 8: for k = 1 to K do
 9:     for i = 1 to n do
10:         V[i][k] ← V[i][k − 1]
11:         P[i][k] ← P[i][k − 1]
12:         S[i][k] ← S[i][k − 1]
13:         for v = 0 to m_i do
14:             if ℓ_{i,v} ≤ k then
15:                 for h = 0 to i − 1 do
16:                     u = V[h][k − ℓ_{i,v}]
17:                     if u ≠ −1∧S[h][k−ℓ_{i,v}]+w_{i,v}+c_{h,u,i,v} ≥ S[i][k]
                        then
18:                         V[i][k] ← v
19:                         P[i][k] ← h
20:                         S[i][k] ← S[h][k − ℓ_{i,v}] + w_{i,v} + c_{h,u,i,v}
21: V[n + 1][K + 1] ← 0
22: P[n + 1][K + 1] ← 0
23: S[n + 1][K + 1] ← 0
24: for h = 1 to n do
25:     u = V[h][K]
26:     if S[h][K] + c_{h,u,n+1,0} ≥ S[n + 1][K + 1] then
27:         P[n + 1][K + 1] ← h
28:         S[n + 1][K + 1] ← S[h][K] + c_{h,u,n+1,0}
29: i ← n + 1
30: k ← K + 1
31: while i ≠ 0 do
32:     v ← V[i][k]
33:     z_i ← v
34:     j ← k
35:     k ← k − ℓ_{i,v}
36:     i ← P[i][j]
37: z_0 ← 0
38: return z
```

これを意味する．次に，$S[i][k-1]$ は文 1 から文 i までのみを用いて，要約長 $k-1$ の要約を作成するときに得られる要約の目的関数値である．すなわち，要約長が 1 だけ短いときの要約の目的関数値である．文 $s_{i,v}$ を要約として追加するときには，文 $s_{i,v}$ を挿入することによって，この要約長が 1 だけ短いときの要約の目的関数値 $S[i][k-1]$ を上回る目的関数値 $S[h][k-\ell_{i,v}] + w_{i,v} + c_{h,V[h][k-\ell_{i,v}],i,v}$ が得られる必要がある．さもなくば，過去に作成した要約より好ましくない要約を得ることになる．$S[i][k-1] \leq S[h][k-\ell_{i,v}] + w_{i,v} + c_{h,V[h][k-\ell_{i,v}],i,v}$ はこれを意味する．文 $s_{i,v}$ を要約に追加したときの目的関数値は，過去に求めた，要約長 $k-\ell_{i,v}$ の要約[*12]の目的関数値 $S[h][k-\ell_{i,v}]$ に，新しく文 $s_{i,v}$ を追加したときの目的関数値を計算すること

[*12] 正確には要約長が $k-\ell_{i,v}$ と指定された際に作成しうる要約であり，長さが $k-\ell_{i,v}$ より短いこともある．たとえば，要約長 100 のときに，得られる最良の要約の長さが 97 である場合などである．ここでは，簡単のため，要約長 $k-\ell_{i,v}$ の要約と説明する．

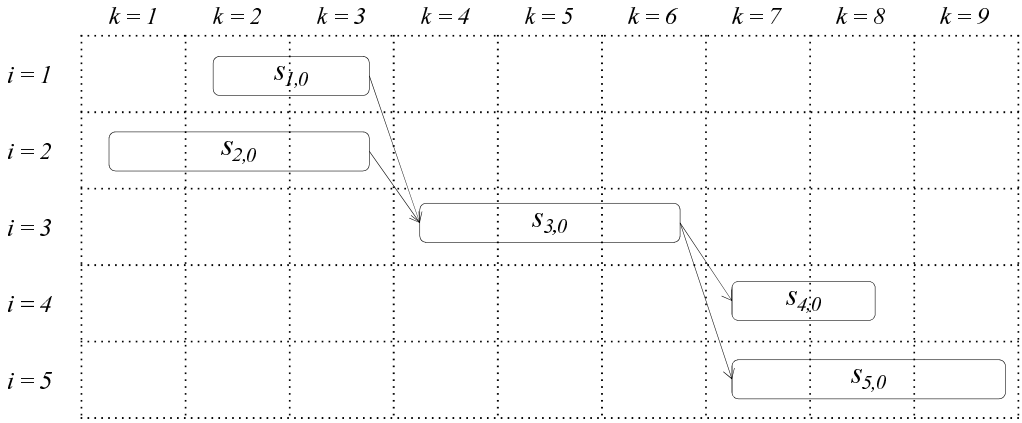


図 4 要約文を要約に追加する場合の処理の一例．図は $i = 3$ 行 $k = 6$ 列の時点での処理を示している．$i = 3$ 行 $k = 6$ 列の時点で，文 $s_{3,0}$ が動的計画法の過程を保存する 2 次元配列に挿入可能であるかを検査するとき，この例では s_3 の長さは 3 であるため，配列に挿入することが可能である．文 $s_{3,0}$ はそれ以前に挿入された $s_{1,0}$ あるいは $s_{2,0}$ と接続することが可能であり，動的計画法の過程で，より目的関数の値が高まる方と接続されることになる．同様に，文 $s_{3,0}$ は，将来 $s_{4,0}$ や $s_{5,0}$ が配列に挿入される際に接続可能な文として，接続の候補になる

Fig. 4 An example of a process adding a summary sentence into a summary.

で得られる．新しく文 $s_{i,v}$ を追加したときの目的関数値の増加分は，文 $s_{i,v}$ の重要度および，文 $s_{i,v}$ と文 $s_{h,V[h][k-\ell_{i,v}]}$ の結束度の和である．文 $s_{i,v}$ と接続可能な文を持つ要約長 $k - \ell_{i,v}$ の要約は，文 1 のみを利用して要約を作成する場合から，文 1 から文 $i-1$ までを利用する場合までの $i-1$ 種類存在する．これらの要約の中で，接続した際に最も目的関数値が高くなる文は，文 1 から文 h までを利用した際に得られる要約長 $k - \ell_{i,v}$ の要約に最後に追加された文 $s_{h,V[h][k-\ell_{i,v}]}$ である．$V[h][k - \ell_{i,v}]$ には h 行 $k - \ell_{i,v}$ 列時点で用いられた文の亜種の種類が格納されている．

要約文を要約に追加する場合の処理の一例を**図 4** に示す．2 次元配列におけるすべての要素，すなわち $i \in 1...n$ 行 $k \in 1...K$ 列において，アルゴリズムは文 $s_{i,0}$ あるいはその亜種 $s_{i,1}...s_{i,m_i}$ が 2 次元配列中に挿入可能であるかを検査する．$i = 3$ 行 $k = 6$ 列の時点では，文 $s_{3,0}$ の長さは 3 であるため，配列に挿入することが可能である．このとき，文 $s_{3,0}$ はそれ以前に挿入された $s_{1,0}$ あるいは $s_{2,0}$ と接続することが可能であり，動的計画法の過程で，より目的関数の値が高まる方と接続されることになる．同様に，文 $s_{3,0}$ は，将来 $s_{4,0}$ や $s_{5,0}$ が配列に挿入される際に接続可能な文として，接続の候補になる．

B の場合は，A の場合を満たさない場合，すなわち，要約の長さが足りないために要約に文 $s_{i,v}$ を追加できない場合，あるいは，要約に文 $s_{i,v}$ を追加しても目的関数値が改善されない場合である．

擬似コードに戻ると，まず 10 行めから 12 行めにかけては目的関数値が改善されない場合を想定して探索を先に進める．そののち，13 行めから 20 行めにかけて，目的関数値が改善可能かを調べ，改善可能であれば目的関数値を改善する．13 行めでは，i 行 k 列時点で，文 s_i の m_i 種類の亜種それぞれを追加の候補として考慮するため，繰返しを行う．14 行めでは，現在追加の候補となっている文 $s_{i,v}$ が要約に追加可能であるかを調べる．追加可能である場合，文 $s_{i,v}$ には $i-1$ 種類の接続候補となる文が存在するため，15 行めで繰返しを行い，それぞれの接続候補と接続した際の目的関数値を調査する．接続候補となる文の亜種の種類は $V[h][k - \ell_{i,v}]$ に格納されているため，16 行めで変数 u にこれを格納しておく．17 行めでは，目的関数値が改善されるか検査する．目的関数値が改善される場合は，用いた亜種を 18 行めで，接続した文を 19 行めで格納する．改善された目的関数値は 20 行めで格納する．21 行めから 23 行めにかけては，文書の末尾を表す仮想的な文を要約の最後に追加する．24 行めから 28 行めにかけては，文書の末尾を表す仮想的な文との結束度を，それまでに作成された要約候補の末尾の文と接続し，目的関数値を計算する．

先に述べた漸化式は，動的計画法の途中の結果を保存する 2 次元配列 V の i 行 k 列の値は h 行 $k - \ell_{i,v}$ 列の結果から計算できるということを示している．そのため，最初の状態である 0 行 0 列の結果から，最終的な結果である $n+1$ 行 $K+1$ 列の結果を計算することができる．最終的な結果が格納された $n+1$ 行 $K+1$ 列から，2 次元配列 P に格納されたポインタを逆にたどることで，その結果を得るための過程，すなわち最良の解を構成する文の系列が得られる．

29 行めから 38 行めは文の系列を復元するためのもので，ポインタを逆にたどり，格納されている文あるいは文の亜種を 1 つずつ z_i に格納する．アルゴリズムは最終的に $\mathbf{z}$ を出力する．

6. 実験

6.1 コーパス

実験のために，12,748 組の入力文書と参照要約を用意した．入力文書はすべて新聞記事であり，参照要約はそれぞれの記事についての人手で作成された要約である．すべての参照要約は 150 文字以内で書かれている．コーパスの統計量を**表 2** に示す．表に示すように，本論文で扱う要約課題に求められる要約率[*13]は約 30%である．

6.2 評価尺度

以下の 2 種類の評価尺度を用いて評価を行う．

内容性：生成された要約の内容性，すなわち生成された要約に要約として含めるべき重要な情報が保持されているかという観点については ROUGE [23] を用いて自動的に評価した．本論文では，ROUGE の亜種のうち，ROUGE-1 と ROUGE-2 を評価に用いる．ROUGE を計算する際には，平尾ら [45] の知見に基づき，内容語[*14]のみを用いた．

言語的品質：要約の言語的品質を評価するため，人手による評価を実施した．人手による評価の際には米国国立標準技術研究所による評価尺度 [28] を用いた．この評価尺度は複数文書要約の評価のために設計されたものであり，本論文では単一文書の要約を対象とするため，文法性，照応関係の明瞭さ，構造と一貫性，全体の 4 つの観点から評価を行った[*15]．評価では，それぞれの観点において，1 点から 5 点までの得点が付与された[*16]．評価のため，各手法によって生成された要約からランダムに 100 文章を抽出し，7 人の評価者でこれを評価した．評価者に著者らは含まれていない．

6.3 比較手法

本論文では以下の 8 種類の手法を比較した．最初の 6 種類がベースラインであり，最後の 2 種類が本文の提案する手法である．

ランダム（RANDOM）：この手法は入力文書を構成する文をランダムに選択する．選択は，ランダムに選択された文の長さとそれまでに選ばれた文の長さの和が要約長を満たす限り行われる．

リード（LEAD）：リード法は自動要約における古典的なベースラインである．この手法は入力文書の先頭から要約長の分だけ文字列を切り出すことで要約を生成する．本論文では単純に入力文書の先頭から 150 文字を切り出すことでこの手法を実装した[*17]．

ナップサック（KP）：前述したようにナップサック問題は単一文書要約を表現する典型的なモデルである．このベースラインは，文に対して重要度を与え，あとは要約長をナップサック制約としてナップサック問題を解くものである．文の重要度は，Nenkova らの提案に基づき，文を構成する内容語の生起確率の平均とした [30]．内容語の生起確率は，入力文書を構成する単語の延べ数に対するある単語の出現回数の割合として計算した．すなわち，入力文書において頻繁に出現する単語が多く文に含まれるほど，その文の重要度は高まる．

ナップサック（教師あり）（KP(S)）：このベースラインでは，文の重要度を単語の生起確率ではなく，機械学習で定める．3.4.1 項で述べた文そのものに関する特徴量 $\mathbf{f}_w$ のみを用いて学習を行い，文の重要度を求めた．すなわち，文間の結束性を加味せずに要約が行われた場合に相当する．

条件付き確率場（CRF）：このベースラインでは，条件付き確率場を用いて文の重要度を求め，その重要度を用いてナップサック問題を解き要約を生成する[*18]．文の重要度は条件付き確率場によって求められた各文の周辺確率とした．条件付き確率場の学習のために，入力文書の各文に対して，重要文か否かの情報を付与する必要がある．そのため，4 章で説明した方法で入力文書の文と参照要約の文を対応づけ，対応づけられた文を重要文と見なした．

隠れ半マルコフモデル（HSMM）：提案手法．この手法では，文の亜種を用いず，入力文書を構成する元の文のみを選択する．

隠れ半マルコフモデル（文亜種あり）（HSMM(C)）：提案手法．上の隠れ半マルコフモデルに基づく要約手法を，文の亜種も加味するようにしたもの．

参照要約（HUMAN）：言語的品質の上限を調査するため，言語的品質の評価の際には参照要約を加えた．

学習の際には 10 分割交差検定を行った．手法間の有意差の検定には，ウィルコクソンの符号付き順位検定を用いた [37]．なお，多重比較となるため，ホルム法 [15] を用い

表 2 コーパスの統計量
Table 2 Corpus statistics.

	入力文書	参照要約
平均文字数	476.2	142.0
平均単語数	298.6	88.3
平均文数	9.7	2.9

[*13] 要約率は要約の長さを入力文書の長さで割った値である [40].
[*14] 名詞，動詞および形容詞．
[*15] それぞれの観点の評価方法については付録 A.1 に示す．
[*16] 1 点：非常に悪い，2 点：悪い，3 点：許容できる，4 点：良い，5 点：非常に良い．
[*17] リード法の実装においてはいくつかの異なる形態を考えることができる．たとえば文字単位ではなく文単位で先頭から要約長を満たす範囲で文を選びだすということもできる．我々は文単位の実装も試したものの，文単位で選択を行うと往々にして要約長に大きな余裕を残して選択が終了し，文字単位と比べ ROUGE において著しく劣ることが多く，したがって本論文では文字単位の実装を採用した．
[*18] なお，Shen らによる方法 [35] は文を逐次的に選択しているにすぎないため，本論文のベースライン手法 CRF とは異なるものとなっている．

て有意水準を調整した．

6.4 ツール

文分割器はコーパスに合わせて独自に実装した．特徴量として用いる段落境界に関する情報も文分割の際に規則を用いて抽出した．形態素解析器は Fuchi らによるもの [13] を，係り受け解析器は Imamura らによるもの [16] をそれぞれ用いた．ROUGE の計算は，Lin の文献 [23] と平尾らの文献 [45] に基づき独自に実装したスクリプトで行った．

7. 結果と考察

実験の，ROUGE による評価の結果を**表 3** に，言語的品質の評価の結果を**表 4** にそれぞれ示す．ここでは，まず ROUGE による内容性の評価の結果について議論し，そののちに言語的品質の評価の結果について議論する．

ROUGE による評価においては，RANDOM を除いたすべての手法が良好な結果を示している．これは，4 章で示したように，多くの参照要約は入力文書を構成する文とまったく同一の文によって構成されているため，重要文を適切に同定することさえできれば良好な要約を生成することができるためである．また，本論文で用いた単一文書要約コーパスの要約率はおおむね 30%と比較的高く，したがって重要文を同定しやすい．確率的には，10 文のうち 3 文は重要文であることになり，したがってランダムに文を選択した場合においても一定の結果が得られている．比較手法した手法の中では，LEAD と HSMM(C) が最も良い結果を示した．これらの間には有意差はなかった．この LEAD の性能は必ずしも驚くべきものではなく，新聞記事は基本的には冒頭部分に重要な情報が記述されており，したがって先頭から 150 文字分の文字列を切り出す LEAD によって生成された要約が ROUGE による評価において高い結果を示すのは自然なことである．一方で，詳しくは後述するが，LEAD は高い ROUGE の代償として言語的品質において劣っており，また参照要約とまったく同一のものを作成する能力も 4.4%と低い[*19]．LEAD とは対照的に，HSMM(C) によって生成された要約の約 20%は参照要約とまったく同一である．文の亜種の生成を行わなかった HSMM が HSMM(C) および LEAD に続く結果を示した．HSMM(C) と HSMM の間には有意差があることから，文の亜種を文選択の候補に含める重要性が示されている．特に，冒頭の 1 文には非常に長いものが稀に存在しており，人間の要約作成作業者はそのような文に含まれる不要な節を除去し文を短く書き換えることがある．文短縮はこの作業を模倣することができたと考えられる．ROUGE だけでなく，人間の参照要約と同一の要約を生成する点においても文の亜種の生成は非常に有効に働いている．

HSMM は HSMM(C) および LEAD 以外の手法に対して ROUGE において統計的に有意に優越している．CRF も HSMM が用いるすべての特徴量を用いることができるが，HSMM が結束性を考慮したデコードを行うことができるのに対し，CRF は文の重要度で結束性を考慮してはいるものの，デコードはあくまでナップサック問題を解くことによって行われている．そのため，CRF は要約長と文間の結束性を学習時とデコード時で一貫して扱うことができないが，HSMM は要約長と文間の結束性を学習時とデコード時で一貫して扱うことができる．この差異が CRF と HSMM の差となったものと考えられる．HSMM と KP(S) の間にも有意差がある．この結果は，文間のテキスト結束性を考慮することで文選択においても良い影響があることを示している．ベースラインの中では CRF と KP(S) の間に ROUGE-2 において有意差があった．このことも同様に，文を選択する際にテキスト結束性を考慮する重要性を示すものといえる．KP(S) と KP の間にも有意差があった．このことは単一文書要約における学習の効果を示すものといえる．

隠れ半マルコフモデルの学習曲線を**図 5** に示す．これは，2,748 組の入力文書と参照要約を固定し，学習に用いる入力文書と参照要約の組を 100，250，500，1,000，2,500，

表 3 ROUGE による内容性の評価の結果．Idt. とある列は，各手法によって生成された要約が参照要約とまったく同一であった割合である．表中の ^{C,S,U,L,R} の記号は，それぞれ条件付き確率場，ナップサック（教師あり），ナップサック，リード，ランダムに対する有意差の有無を示す

Table 3 A result of content quality evaluation with ROUGE.

Method	ROUGE-1	ROUGE-2	Idt.
RANDOM	0.417	0.291	1.2%
LEAD	$0.779^{C,S,U,R}$	$0.727^{C,S,U,R}$	4.4%
KP	0.704^{R}	0.611^{R}	9.3%
KP(S)	$0.729^{U,R}$	$0.647^{U,R}$	10.4%
CRF	$0.741^{U,R}$	$0.675^{S,U,R}$	11.3%
HSMM	$0.769^{C,S,U,R}$	$0.703^{C,S,U,R}$	15.2%
HSMM(C)	$0.785^{C,S,U,R}$	$0.722^{C,S,U,R}$	20.4%

表 4 言語的品質の評価の結果．値の値域は 1（非常に悪い）から 5（非常に良い）まで [28] である．有意差に関する記法は表 3 と同様である

Table 4 A result of linguistic quality evaluation.

Method	Gram.	Ref.	S./C.	Overall
LEAD	1.9	3.9	2.5	2.1
KP	4.1^{L}	3.7	3.4	3.5
KP(S)	4.2^{L}	3.6	3.5	3.6^{L}
CRF	4.1^{L}	3.9	3.7^{L}	3.6^{L}
HSMM	4.3^{L}	4.0	4.1^{L}	4.0^{L}
HSMM(C)	4.0^{L}	3.9	4.0^{L}	3.9^{L}
HUMAN	4.7^{L}	4.5	4.7^{L}	4.8^{L}

[*19] 逆に考えると，参照要約のうち 4.4%は単に冒頭の 150 文字を切り出しただけ，あるいは元々 150 文字以内の入力文書をそのまま転写しただけのものとなっている．

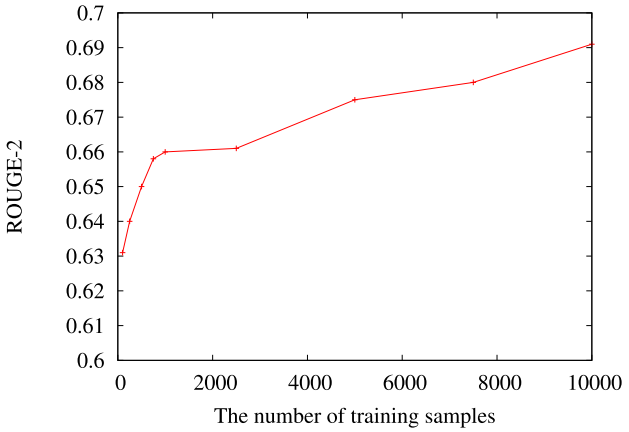


図 5 提案手法の学習曲線

Fig. 5 A learning curve of the proposed method.

5,000，7,500，10,000 と段階的に増やしたものである．グラフが示すように，大規模な訓練事例を用いることによって，要約の内容性が大きく改善されている．グラフを見る限りは，入力文書と参照要約の組の追加による ROUGE の向上はまだ収束していないように観察される．したがって訓練事例をさらに追加することによってさらなる ROUGE の改善が可能であろう．Filippova は文短縮課題において大規模な訓練事例の有効性を示しており [12]，本論文で示す結果も同様に自動要約における大規模な訓練事例の有効性を示すものといえる．過去の自動要約研究は数十から数百の比較的小規模なコーパスに頼ってきたが，本論文では大規模なコーパスを用いることで要約の品質が改良されることを示した．

次に，言語的品質の評価の結果について議論する．ROUGE による評価とは異なり，HSMM が最良の結果を得た．Nenkova らが指摘しているように [29]，文短縮は文法的でない文を生成することがある．そのため，文短縮は内容性を向上させるものの，往々にして言語的品質の劣化の原因となる．表 4 が示すように，HSMM(C) の文法性は HSMM より低い．HSMM と，LEAD を除く他のベースラインの間に言語的品質の評価において有意差は見いだせなかったものの，HSMM と HSMM(C) は構造と一貫性の観点において最良の結果を得ている．構造と一貫性は談話構造に関する評価の観点であり，結束性に関する特徴量を要約に導入することで，この観点において要約の言語的品質が改善されたことが分かる．ROUGE において高い性能を示した LEAD は，言語的品質では低い評価を得ている．これは，LEAD は要約長に達した際にそれ以降の文は文の途中であろうともすべて切除するため，往々にして要約の末尾に非文法的な文の断片を残すためである[*20]．

加えて，内容性と言語的品質の関係についても述べておく．表 3 と表 4 を比較すると，これらの間に一定の相関があることが観察される．この結果は必ずしも驚くべきものではなく，人間は文法的でまとまりの良い要約からより多くの情報を引き出すことができると考えるのは自然である．本論文で提案した手法は，学習の際に ROUGE に最適化されているが，この最適化は言語的品質についても貢献したものと思われる．人間の作業者の作成した参照要約は文法的でよくまとまっているものであり，これを模倣しようとすれば言語的品質も向上することが予想できる．

[*20] 6.3 節で述べたように，LEAD を実装する際には文字単位ではなく文単位にしてもよい．その場合，言語的品質は大きく向上すると思われるものの，一方で ROUGE に関しては，実際に評価したところ，文字単位の LEAD に比べ評価が著しく悪化した．

表 5 提案手法が学習したパラメータの一部

Table 5 Parameters learned in the proposed method.

特徴量	重み
文の位置：先頭の文	2.032796
文の位置：最初の段落に含まれる	0.607047
文の位置：段落の先頭	0.154139
エンティティ・グリッド：X1	0.019769
単語の表記：問題	0.016576
単語の表記：調査	−0.011098
単語の表記：主張	−0.013093
語彙の遷移：<s>-検討	−0.015425
エンティティ・グリッド：S1	−0.015466
文の位置：文書中での位置	−0.863246

提案手法が学習したパラメータについても触れておく．**表 5** に学習されたパラメータのうち重みが大きいものと小さいものをそれぞれ 5 つ示す．

重みが大きいものの 3 つと小さいものの 1 つは文の位置に関する特徴量である．文書中での各文の位置は文書の先頭の文が 0，末尾の文が 1 となるように正規化されており，文書の後半の文ほど値が大きくなっている．そのため，「文書中での位置の重み」が負であるのは文書の後半に位置する文ほど要約に選ばれにくいことを示している．また，先頭の文や最初の段落の文，段落の先頭に位置する文などは要約に非常に選ばれやすいことが分かる．特に先頭の文は著しく要約に選ばれやすくなっており，このことは単一文書要約におけるリード法の有効性を学習結果から示すものとなっている．

エンティティ・グリッドに基づく特徴量「X1」[*21]は，要約の末尾の文に主語や目的語以外の何らかの名詞句が含まれることを示す．この特徴量が大きな重みを持っていることは，要約の末尾の文には，「は」や「が」といった格助詞や「に」や「を」といった格助詞をともなって出現する名詞句よりも，それらをともなわずに出現する名詞句が多いことを示している．このことを裏付けるように，エンティティ・グリッドに基づく特徴量「S1」は非常に小さい重みを得ている．これは，要約の末尾の文においては「は」や「が」などの格助詞をともなう名詞句が出現しにくいこと

[*21] エンティティ・グリッドの詳細については横野らの文献 [41] を参照されたい．

を示しており，要約の末尾の文においては，主語となる名詞句が省略されやすいことを示唆している．

個別の単語の表記については，「問題」が頻繁に要約に出現すること，一方で「調査」「主張」は出現しにくいことを示している．「調査」「主張」といった単語が要約に出現しにくい理由は必ずしも明らかではないが，要約はあくまで記事の端的な結論のみを含むものであるため，これらの単語はそのような結論を示す文以外の文に頻繁に出現することが示唆される．

8. まとめ

本論文では，隠れ半マルコフモデルに基づく新しい単一文書要約モデルを提案した．提案した要約モデルは，ナップサック問題に基づく要約モデルと隠れマルコフモデルに基づく要約モデルを融合させたものとなっている．我々の提案するモデルは文間の結束性をとらえることができ，これを加味して重要文の選択を行うことができる．この性質は，生成される要約のテキスト結束性を担保するために重要であり，要約の内容性と言語的品質を両立させるために重要である．本論文では提案する要約手法の解を探索するためのアルゴリズムもあわせて提案した．また，大規模単一文書要約コーパスに基づく実験により，本論文で提案する要約手法は，他のベースラインに優越する性能を持つことが分かった．

謝辞 本論文で用いられた単一文書要約コーパスは株式会社毎日新聞社の所有物であり，日本電信電話株式会社に貸与されたものである．株式会社の毎日新聞社のご厚意に記して感謝する．

論文の採録に際しては，担当編集委員および 2 名の査読者の方々より有益なご助言を頂戴した．記して感謝する．

参考文献

[1] Althaus, E., Karamanis, N. and Koller, A.: Computing Locally Coherent Discourses, *Proc. 42nd Meeting of the Association for Computational Linguistics* (*ACL*), pp.399–406 (2004).

[2] Balas, E.: The prize collecting traveling salesman problem, *Networks*, Vol.19, No.6, pp.621–636 (1989).

[3] Barzilay, R. and Elhadad, M.: Using Lexical Chains for Text Summarization, *Proc.Intelligent Scalable Text Summarization Workshop* (*ISTS*), pp.10–17 (1997).

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

[5] Barzilay, R. and Lapata, M.: Modeling local coherence: an entity-based approach, *Proc. 43rd Annual Meeting on Association for Computational Linguistics* (*ACL*), pp.141–148 (2005).

[6] Barzilay, R. and Lee, L.: Catching the Drift: Probabilistic Content Models, with Applications to Generation and Summarization, *Proc. HLT-NAACL 2004*, pp.113–120 (2004).

[7] Christensen, J., Mausam, Soderland, S. and Etzioni, O.: Towards Coherent Multi-Document Summarization, *Proc. 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp.1163–1173 (2013).

[8] Clarke, J. and Lapata, M.: Modelling Compression with Discourse Constraints, *Proc. 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning* (*EMNLP-CoNLL*), pp.1–11 (2007).

[9] Collins, M.: Discriminative Training Methods for Hidden Markov Models: Theory and Experiments with Perceptron Algorithms, *Proc. 2002 Conference on Empirical Methods in Natural Language Processing* (*EMNLP*), pp.1–8 (2002).

[10] Crammer, K.: Online Passive-Aggressive Algorithms, *Journal of Machine Learning Research*, Vol.7, No.Mar, pp.551–585 (2006).

[11] Daume, III, H. and Marcu, D.: A Noisy-Channel Model for Document Compression, *Proc. 40th Annual Meeting of the Association for Computational Linguistics* (*ACL*), pp.449–456 (2002).

[12] Filippova, K.: Overcoming the Lack of Parallel Data in Sentence Compression, *Proc. 2013 Conference on Empirical Methods in Natural Language Processing* (*EMNLP*), pp.1481–1491 (2013).

[13] Fuchi, T. and Takagi, S.: Japanese morphological analyzer using word co-occurrence: JTAG, *Proc. 36th Annual Meeting of the Association for Computational Linguistics and the 17th International Conference on Computational Linguistics* (*ACL-COLING*), pp.409–413 (1998).

[14] Hirao, T., Yoshida, Y., Nishino, M., Yasuda, N. and Nagata, M.: Single-Document Summarization as a Tree Knapsack Problem, *Proc. 2013 Conference on Empirical Methods in Natural Language Processing* (*EMNLP*), pp.1515–1520 (2013).

[15] Holm, S.: A Simple Sequentially Rejective Multiple Test Procedure, *Scandinavian Journal of Statistics*, Vol.6, No.2, pp.65–70 (1979).

[16] Imamura, K., Kikui, G. and Yasuda, N.: Japanese Dependency Parsing Using Sequential Labeling for Semi-spoken Language, *Proc. 45th Annual Meeting of the Association for Computational Linguistics Companion Volume Proceedings of the Demo and Poster Sessions*, pp.225–228 (2007).

[17] Karamanis, N., Poesio, M., Mellish, C. and Oberlander, J.: Evaluating Centering-Based Metrics of Coherence, *Proc. 42nd Meeting of the Association for Computational Linguistics* (*ACL*), pp.391–398 (2004).

[18] Kikuchi, Y., Hirao, T., Takamura, H., Okumura, M. and Nagata, M.: Single Document Summarization based on Nested Tree Structure, *Proc. 52nd Meeting of the Association for Computational Linguistics* (*ACL*), pp.315–320 (2014).

[19] Kim, S., Yun, S. and Yoo, C.D.: Large Margin Discriminative Semi-Markov Model for Phonetic Recognition, *IEEE Trans. Audio, Speech, and Language Processing*, Vol.7, No.19, pp.1999–2012 (2011).

[20] Korte, B. and Vygen, J.: *Combinatorial Optimization, 3rd edition*, Springer-Verlag (2008).

[21] Lafferty, J., McCallum, A. and Pereira, F.C.N.: Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data, *Proc. 18th International Conference on Machine Learning* (*ICML*),

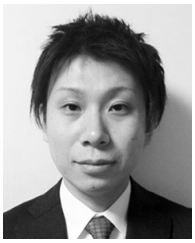
pp.282–289 (2001).
[22] Lapata, M.: Probabilistic Text Structuring: Experiments with Sentence Ordering, *Proc. 41st Annual Meeting of the Association for Computational Linguistics* (*ACL*), pp.545–552 (2003).
[23] Lin, C.-Y.: ROUGE: A Package for Automatic Evaluation of Summaries, *Proc. ACL Workshop Text Summarization Branches Out*, pp.74–81 (2004).
[24] Luhn, H.P.: The automatic creation of literature abstracts, *IBM Journal of Research and Development*, Vol.22, No.2, pp.159–165 (1958).
[25] Mann, W.C. and Thompson, S.A.: Rhetorical structure theory: Toward a functional theory of text organization, *Text*, Vol.8, No.3, pp.243–281 (1988).
[26] Marcu, D.: From discourse structure to text summaries, *Proc. ACL/EACL 1997 Summarization Workshop*, pp.82–88 (1997).
[27] McDonald, R.: A study of global inference algorithms in multi-document summarization, *Proc. 29th European Conference on Information Retrieval* (*ECIR*), pp.557–564 (2007).
[28] National Institute of Standards and Technology: The linguistic quality questions (2007), available from ⟨http://www-nlpir.nist.gov/projects/duc/duc2007/quality-questions.txt⟩.
[29] Nenkova, A. and McKeown, K.: *Automatic Summarization*, Now Publishers (2011).
[30] Nenkova, A. and Vanderwende, L.: The Impact of Frequency on Summarization, Technical Report, Microsoft Research (2005).
[31] Nomoto, T.: A Generic Sentence Trimmer with CRFs, *Proc. 46th Annual Conference of the Association for Computational Linguistics: Human Language Technologies* (*ACL-HLT*), pp.299–307 (2008).
[32] Okazaki, N., Matsuo, Y. and Ishizuka, M.: Improving Chronological Sentence Ordering by Precedence Relation, *Proc. 20th International Conference on Computational Linguistics* (*Coling*), pp.750–756 (2004).
[33] Pitler, E., Louis, A. and Nenkova, A.: Automatic Evaluation of Linguistic Quality in Multi-Document Summarization, *Proc. 48th Annual Meeting of the Association for Computational Linguistics* (*ACL*), pp.544–554 (2010).
[34] Sarawagi, S. and Cohen, W.W.: Semi-Markov Conditional Random Fields for Information Extraction, *Advances in Neural Information Processing Systems 17*, pp.1185–1192 (2004).
[35] Shen, D., Sun, J.-T., Li, H., Yanfg, Q. and Chen, Z.: Document summarization using conditional random fields, *Proc. 20th International Joint Conference on Artifical Intelligence* (*IJCAI*), pp.2862–2867 (2007).
[36] Weinman, J.J., Learned-Miller, E. and Hanson, A.: A discriminative semi-Markov model for robust scene text recognition, *Proc. 19th International Conference on Pattern Recognition* (*ICPR*), pp.1–5 (2008).
[37] Wilcoxon, F.: Individual comparisons by ranking methods, *Biometrics Bulletin*, Vol.1, No.6, pp.80–83 (1945).
[38] Wu, X. and Zong, C.: A New Approach to Automatic Document Summarization, *Proc. 3rd International Joint Conference on Natural Language Processing* (*IJCNLP*), pp.126–132 (2008).
[39] Yu, S.-Z.: Hidden semi-Markov models, *Artificial Intelligence*, Vol.174, No.2, pp.215–243 (2010).
[40] 奥村 学, 難波英嗣：テキスト自動要約, オーム社 (2005).
[41] 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).
[42] 田窪行則，西山佑司，三藤 博，亀山 恵，片桐恭弘：談話と文脈，岩波書店 (1999).
[43] 西川 仁，平尾 努，牧野俊朗，松尾義博，松本裕治：冗長性制約付きナップサック問題に基づく複数文書要約モデル，自然言語処理，Vol.20, No.4, pp.585–612 (2013).
[44] 西川 仁，長谷川隆明，松尾義博，菊井玄一郎：文の選択と順序付けを同時に行う評価文書要約モデル，人工知能学会論文誌，Vol.28, No.1, pp.88–99 (2013).
[45] 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.753–1766 (2006).
[46] 平尾 努，鈴木 潤，磯崎秀樹：最適化問題としての文書要約，人工知能学会論文誌，Vol.24, No.2, pp.223–231 (2009).

付 録

A.1 言語的品質の評価

6.2 節で述べた言語的品質の評価の際には，各評価項目において，以下の質問を要約とともに評価者に提示した．評価者は各項目について 1 点（非常に悪い），2 点（悪い），3 点（許容できる），4 点（良い），5 点（非常に良い）のいずれかの得点を付与した．以下に示す質問は基本的には米国国立標準技術研究所による言語的品質の評価尺度 [28] に基づくものである．

(1) **文法性**：要約に，明らかに非文法的な文（細切れになっている文や，主語や述語など文に不可欠な要素が欠けており文意が読み取れない文）が含まれていないか？ これらが含まれている要約は好ましくない要約である．
(2) **照応関係の明瞭さ**：要約に含まれる代名詞や名詞句が指示しているものは明瞭か？ 指示しているものが不明瞭な代名詞や名詞句を含む要約は好ましくない要約である．
(3) **構造と一貫性**：要約を構成する文どうしの論理関係を読み取ることができるか？ ある文とその次の文で話題が唐突に変化していないか？ 論理関係を読み取ることのできない文の組や，話題の唐突な変化を含む要約は好ましくない要約である．
(4) **全体**：文法性，照応関係の明瞭さ，構造と一貫性のそれぞれを加味したうえで，全体として要約から明瞭に意味を読み取ることができるか？



西川 仁 （正会員）

2006 年慶應義塾大学総合政策学部総合政策学科卒業．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．2013 年奈良先端科学技術大学院大学情報科学研究科博士後期課程修了．博士（工学）．NTT メディアインテリジェンス研究所研究員を経て，2015 年より東京工業大学大学院情報理工学研究科計算工学専攻助教．言語処理学会，人工知能学会，The Association for Computational Linguistics 各会員．

有田 一穂

1986 年鹿児島大学工学部卒業．1988 年同大学大学院工学研究科電子工学専攻修士課程修了．同年日本電信電話株式会社入社．現在，NTT サービスエボリューション研究所主任研究員．

田中 克己

1989 年早稲田大学電気工学部卒業．同年日本電信電話株式会社入社．現在，NTT サービスエボリューション研究所主任研究員．

平尾 努 （正会員）

1995 年関西大学工学部電気工学科卒業．1997 年奈良先端科学技術大学院大学情報科学研究科博士前期課程修了．同年株式会社 NTT データ入社．2000 年より NTT コミュニケーション科学基礎研究所に所属．博士（工学）．自然言語処理の研究に従事．言語処理学会，The Association for Computational Linguistics 各会員．

牧野 俊朗

1987 年東京大学工学部電子工学科卒業．1992 年同大学大学院博士課程修了．博士（工学）．同年日本電信電話株式会社入社．現在，NTT メディアインテリジェンス研究所主幹研究員．知識獲得，推論手法，言語処理等に関する研究開発に従事．

松尾 義博 （正会員）

1988 年大阪大学理学部物理学科卒業．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．現在 NTT メディアインテリジェンス研究所音声・言語基盤技術グループリーダ．機械翻訳，自然言語処理の研究に従事．言語処理学会会員．