係り受け関係を用いた新聞記事の見出し生成
Generating Headlines for News Articles using Dependency Structure

奥村　直也†　　白井　匡人††　　三浦　孝夫†

† 法政大学理工学研究科　〒 184–8584 東京都小金井市梶野町 3-7-2
†† 法政大学マイクロ・ナノテクノロジー研究センター　〒 184–0003 東京都小金井市緑町 3-11-15
E-mail: †naoya.okumura.7s@stu.hosei.ac.jp, ††{masato.shirai.37,miurat}@hosei.ac.jp

あらまし　本研究では，新聞記事の見出し生成手法を提案する．新聞記事の見出しは，特徴的な表現（用語や言い回し）で記述されていることが多い．従来手法では，記事本文からのキーワード抽出やパターン抽出手法が提案されている．しかし，特徴的な表現を考慮することができないという問題がある．本研究では，特徴的表現を考慮するために，類似した記事は類似した見出しが対応すると仮定し，類似した見出しのパターンを使用することで，その対応を考察する．潜在意味解析を利用することにより，類似した記事を選択し，対応する見出しに対して係り受け関係を使用し，キーワードを入れ替えることにより見出し生成を行う．

キーワード　潜在意味解析, 見出し生成, 係り受け関係, 特徴抽出

1. 前　書　き

近年，インターネットから電子書籍や新聞記事など様々な文書を容易に入手できる．しかし、短期間のうちにあまりに多数の文書が生じるため，何ら利用されることなく流れ去っている．これらの文書は，多様で大量のデータ量を含み，内容やその傾向を理解することは困難であり時間を要する．このため，文書内容を素早く理解するために文書要約などの特徴抽出手法が必要となる．

新聞記事には内容を表した要約や見出しが付与されている．要約や見出しは，記事本文の特徴を表現したものであり，文書の素早い理解に有効である．利用者は，正しい要約を読むことで，記事本文を読まず，1 記事の全貌を把握することができる．このため，要約により文書の内容を明確化することで，分類や検索が容易かつ効率的に行える．文書集合が大規模である場合，人手による文書要約はコストが大きいため，自動的に文書を要約できることが望ましい．しかし，1 記事から要約として文章を生成することは困難である．要約の文章数が多いほど，その文書がどういった内容を表すのか直感的に理解しずらい．1 文章で書かれた見出しは，一目で見て素早く記事の内容を把握することができる．

記事の要約には，様々な研究がなされている．記事の要約は大きく文章生成型要約と文書抽出型要約に分けられる．Wan ら [10] は，ビタビアルゴリズムを用い，文章生成型要約手法を提案している．文章生成型要約では，記事の内容に基づき要約文として文章を生成する．しかし，自然な文章を生成することは困難である．次に，Wan ら [2] は，ページランクを用いて，ランキングベースの文章抽出型要約を提案している．文章抽出型要約では，要約文を記事から直接抽出することで要約を行う．Ouyang [3] は，Support Vector Regression (SVR) を用いて，あらかじめ用意された学習データから対象の文書の要約に適切な文章を抽出する抽出型要約手法を提案している．学習データに基づく文章抽出型要約では，他に Deep learning [5] や潜在意味解析 (Latent Semantic Analysis) [4] を利用した文章抽出が行われている．学習データを使用した要約は，他の二つの手法と違い，文章の意味を考慮して記事集合から適切な文章を検索することができる．このため，本研究では，学習データに基づく文章抽出型要約を考慮する．

本研究では，内容が類似した記事には類似した見出しが付与されると仮定する．提案手法は，要約対象の記事に類似した記事を選択し，その見出しを使用することで，新たな見出しを生成する．ここでは，元の見出しの特徴語と要約対象の記事の特徴語を入れ替える．特徴語は文章中の最も重要な情報なため，見出しの特徴語が変われば見出しの意味も変化する．これより，1 文章による記事の要約を行うことができる．

2 章では記事内容の表現と見出しの表現の概要を述べ，本文検索についてを述べる．3 章では特徴語抽出について，そして提案手法についてを述べる．4 章では実験を行い，得られた結果に関する考察を述べる．5 章で結論とする．

2. 記事内容の表現と見出しの表現

2.1 記事の表現

記事本文を検索や解析するためには，単語による表現方法である BagOfWords モデルと潜在意味解析が用いられている．BagOfWords モデルは，単語を特徴量の集合とみなし，出現頻度 TF (Term Frequency) や TF・IDF (Inverse Document Frequency) などの重みを特徴量として表現するモデルである．潜在意味解析は，文書とそこに含まれる語の共起性に注目し，特徴量としてまとめた値のベクトルで表現する手法である．BagOfWords モデルでは各文書を語ベクトルで表し潜在意味解析では語の意味ベクトルの集合で表現する．BagOfWords モデルでは，類義語や同義語などの類似した語を区別することがで

きない．潜在意味解析は，単語の概念として潜在意味を用いるため，潜在意味の類似により，類義語や同義語を考慮することができる．これらの BagOfWords モデルや潜在意味などの特徴量を表現するために，ベクトル空間モデルが用いられる．

2.2 見出しの表現

記事の見出しは，本文とは違い，明確な構文形式を持たず，特徴的な表現（単語や言い回し）でで記述されていることが多い．見出しを作成するために，キーワードや要約として文章を抽出する手法があるが，記事内の文章を引用することは少ない. [8]

2.3 潜在意味抽出

潜在意味解析の基本となる主成分分析 (PCA) は，文書内の語の共分散を最大化し，主成分と呼ばれる特徴量にまとめる手法である．主成分分析では，主成分を用いて個々のデータを記述し，互いの差異が極大化されるため，特徴部分の検出が容易となる．この利点を受け継いるものが潜在意味解析である．潜在意味解析では，主成分を潜在意味と呼び，2 項の軸（単語，記事）に対してそれぞれの潜在意味（主成分）で表現することができる．これらの潜在意味は，ベクトル表現により，1 つの特徴量として表現することができる．

記事内容の類似性を推定するために，潜在意味解析を利用する．潜在意味解析とは，文書集合とその文書に含まれる単語の共起性に基づいて潜在意味を抽出し，語および文書を潜在意味のベクトルで表現し，その関係を分析する技術である．

3. 提 案 手 法

3.1 本 文 検 索

記事は本文と見出しからなる．提案手法は，記事の見出しを推定するために類似した記事の見出しを利用する．類似した記事は，類似した見出しを持つと仮定し，類似記事の検索を行う．記事類似性を判定するために，潜在意味解析を用いる．

本文検索は以下のように行う．未知記事 q と見出しあり既知記事集合：D 件 ($|D| > 1$) を準備する．潜在意味解析を行うために，D に対して特異値分解（SVD）を行う．特異値分解は，潜在意味で表現された単語・単語行列 U，単語と文書の潜在意味行列 S，潜在意味で表現された文書・文書 V に分解できる．単語・文書行列 D の特異値分解は，以下の式で表される．

$D = U \times S \times V^t$

質問記事 q に対し，潜在意味空間へ射影する．潜在意味空間へ射影すると以下の式になる．

$q^t U(S^{-1})^t$

射影した q と既知記事集合のそれぞれの文書を潜在意味で表現された V を検索する．行列内のすべての記事と類似度を計算する．射影した q と V の中の文書 i の類似度 $cos(q, i)$ は以下のように定義される．

$\cos(q, i) = \frac{(q^t U(S^{-1})^t, V_i)}{|q^t U(S^{-1})^t||V_i|}$

これにより，提案手法は類似度の高い順に q の候補記事として抽出できる．候補記事の見出しが q の見出し候補となる．

3.2 係り受け関係

次に提案手法では，特徴語抽出と見出し生成の際に，係り受け関係を利用する．見出し生成では，見出しを 1 から生成するため，構文解析を必要とする．これは，生成した見出しを正しい文法に従わせ，見出しとして意味が成り立つようにするためである．構文解析は複雑であり，本研究では，構文解析を係り受け関係で代用する．

本研究で使用する係り受け関係は，述語に直接係る名詞（名詞句）を含んだ関係を使用する．図 1 から，述語（始まった，宣誓し，就任した）に直接かかっている名詞（名詞句）は,（就任式が，２０日，連邦議会議事堂前で，トランプ氏が，第４５代大統領）となる．係り受け関係は,（就任式が→始まった，２０日→始まった，連邦議会議事堂前で→始まった，トランプ氏が→宣誓し，第４５代大統領に→就任した）となる．

次に後処理を行う．後処理では，係っている名詞の助詞を削除し，述語を原形に変更する．係っている名詞の助詞をとる理由は，見出しの特徴語を入れ替えるとき，名詞のみを入れ替えることで，必要な助詞は特徴語に付与されている助詞を使うことで，文の意味の繋がりを考えれる．さらに，述語を原形の理由は，係り受け関係の同じものを探索するとき，時制だけ違い，述語の語尾が違う形のものを拾うために，原形に直す必要がある．最後に，係り受け関係は,（就任式→始まる，２０日→始まる，連邦議会議事堂前→始まる，トランプ氏→宣誓，第４５代大統領→就任）となる．

本研究の特徴は，係り受け関係を利用することで，見出し生成に関わる特徴語の自動抽出とこれらを対応付けた見出しの生成にある．

表 1 見出し既知記事

見出し
1：トランプ氏が宣誓、第４５代大統領に就任
本文
米共和党のドナルド・トランプ新大統領の就任式が２０日、 首都ワシントンの連邦議会議事堂前で始まった。 トランプ氏が宣誓し、第４５代大統領に就任した。

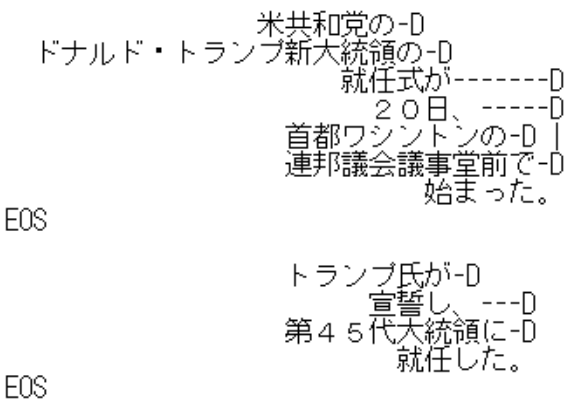

図 1 既知記事の係り受け関係

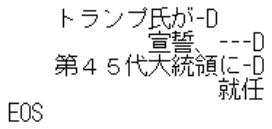

図 2　既知記事の見出しの係り受け関係

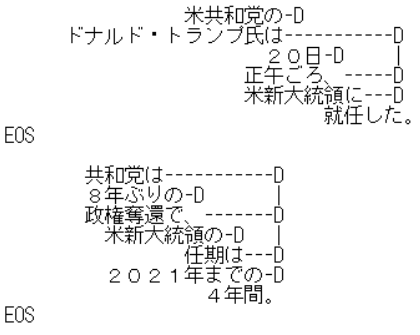

図 3　未知記事の係り受け関係

表 2　見出し未知記事

本文
ドナルド・トランプ大統領は２０日正午ごろ、米新大統領に就任した。 共和党は８年ぶりの政権奪還で、米新大統領の任期は２０２１年までの４年間。

3.3　特徴語抽出

見出しは記事の内容を特徴付ける表現であるから，その自立語はすべて（記事の）特徴語である．特徴語が異なると，見出しの意味も変化する．見出しは特徴語の配置を扱う枠組であるとみなせば，本文には特徴語を述べる表現を含むであろう．逆に，そのような表現は記事を表す新たな特徴語を定義している．

記事の要約や見出しは，記事の第１段落の文章または特徴語を使用するケースが多い．そのため,「類似記事の第１段落に出現する単語」と関連する「見出し未知記事の単語」を特徴語として選択する．その特徴語を見出しの特徴語を入れ替え，新たな見出しを生成する．その際，見出しのどの特徴語を入れ替えるかの判断が困難となる．そこで，単語同士の係り受け関係を用いて，特徴語の入れ替えを図る．

表３より，特徴語の抽出手順を例題で示す．

[1] 類似記事の見出し [0] と本文からそれぞれ係り受け関係を抽出．図２より，見出し係り受け関係（トランプ氏→宣誓,　第４５代大統領→就任）．図１より，本文係り受け関係（就任式→始まる，２０日→始まる，連邦議会議事堂前→始まる，トランプ氏→宣誓，第４５代大統領→就任）．

[2] 類似記事の見出しと本文の係り受け関係 [1] から一致する関係を抽出．一致する係り受け関係（トランプ氏→宣誓,　第４５代大統領→就任）．

[3] 一致する係り受け関係 [2] から述語を抽出．述語（□→宣誓,　□→就任）．

[4] 述語 [3] に関する係り受け関係を見出し未知記事の本文から抽出．図３より，述語 [3] に関する係り受け関係（ドナルド・トランプ氏→就任,　米新大統領→就任）．

そして，[4] で得られた単語を見出し未知記事の特徴語として選択．特徴語（ドナルド・トランプ氏,　米新大統領）．

表 3　特徴語抽出の例題

[0] 類似記事の見出し
１：トランプ氏が宣誓、第４５代大統領に就任
[1] 類似記事の見出しの係り受け関係
トランプ氏→宣誓,　第４５代大統領→就任
[1] 類似記事の本文の係り受け関係
就任式→始まる，２０日→始まる 連邦議会議事堂前→始まる トランプ氏→宣誓，第４５代大統領→就任
[2] 一致する係り受け関係
トランプ氏→宣誓,　第４５代大統領→就任
[3] 述語
□→宣誓,　□→就任
[4] 述語に関する係り受け関係
ドナルド・トランプ氏→就任,　米新大統領→就任
見出し未知記事の特徴語
ドナルド・トランプ氏,　米新大統領

3.4　見出し生成

見出し生成では，構文解析にあたる部分を，係り受け関係によって補う．そのため，元となる見出しが必要となる．そこで，類似記事から見出しを選択する．類似記事の見出しをそのまま使用しても，類似記事専用の見出しのため，そのまま使用することはできない．このため，見出しの特徴語を入れ替える必要があるが，適切な特徴語を入れ替えなければ見出しの意味が不明になる．見出し生成は，係り受け関係を用いることで，選択した特徴語を入れ替え，真の見出しの意味に近づけることを目標にする．

表４より，見出しの生成手順を例題で示す．

[5] 特徴語の頻度を計算．図３より，特徴語の頻度は（ドナルド・トランプ氏＝ 1, 米新大統領＝ 2）．

[6] 特徴語の頻度 [5] を利用し，頻度の高い順に，述語に合う特徴語を選択．述語に合う特徴語（米新大統領→就任）．

[7] 述語に合う特徴語 [6] を使用し，類似記事の見出し [0] の特徴語と入れ替え，見出し生成．生成見出し (１：トランプ氏が宣誓、米新大統領に就任).

そして，[7] で得られた生成見出しを，見出し未知記事の見出しとして使用．

表 4　見出し生成の例題

[5] 特徴語の頻度
ドナルド・トランプ氏＝ 1, 米新大統領＝ 2
[6] 述語に合う特徴語
米新大統領→就任
[7] 生成見出し
１：トランプ氏が宣誓、米新大統領に就任

4. 実　　験

本稿は実験環境と実験手順と評価方法を述べ，実験により得られた結果を示す．得られた結果に関する考察を述べる．本研究では，実験 1 : 特徴語抽出と実験 2 : 段落による特徴語抽出の二つの実験を行う． 実験 1 : 特徴語抽出では，提案手法とベースラインを比較評価するために，記事に合った特徴語が抽出できたかと特徴語として正しいかを評価する．新聞記事は，これまで「第 1 段落では全体を要約することが多い」と言われている．そこで実験 2 : 段落による特徴語抽出では，第 1 段落のみの記事が検索範囲として有効かを確かめるために，第 1 段落と第 2 段落を含めた記事を使用し，記事に合った特徴語を抽出できたか，特徴語として正しいかを評価する．

4.1 実 験 環 境

実験に使用するコーパスは，毎日新聞記事データ集 2012 より 1 月から 12 月分の 1 年分の記事 50557 件を用いる．質問記事として，全体の新聞記事 50557 件から記事 5055 件を使用する．形態素解析ソフト Mecab を用いて，形態素解析を行い，文書毎に名詞を抽出する．特徴語を抽出する際の係り受け関係は，係り受け解析ソフト Cabocha を用いて抽出する．ベースラインは，類似記事から単語列のパターンを用いて，特徴語候補を選択し，質問記事に存在する単語を特徴語として抽出する．

4.2 評 価 方 法

特徴語抽出と段落よる特徴語抽出の実験についてそれぞれ行う．それぞれの実験では，記事に合った特徴語を抽出できたかと特徴語として正しいかを評価する．質問記事に対して，それぞれ特徴語を抽出する．このとき，正解見出しの名詞と同じ単語が生じていれば正しい特徴語とする．(1) 抽出した特徴語がどれだけ正解見出しの名詞に生じているかの割合を適合率という．(2) すべての抽出した特徴語が，特徴語が正しいと判断された割合を再現率という．本論文では，特徴語の適合率・再現率を評価対象とする．

4.3 実 験 結 果

実験 1 で得られた結果を表 5 に示す．

表 5　実験 1 の適合率と再現率（全体）

	提案手法	ベースライン
適合率	45.56	39.29
再現率	46.77	41.39

実験 2 で得られた結果を表 6 に示す．

表 6　実験 2 の適合率と再現率（全体）

	提案手法（第 1 段落）	提案手法（第 1 段落＋第 2 段落）
適合率	45.56	36.42
再現率	46.77	40.01
	ベースライン（第 1 段落）	ベースライン（第 1 段落＋第 2 段落）
適合率	39.29	.30.19
再現率	41.39	35.49

実験 1 では，適合率は表 5 のようになり，提案手法とベースラインを比べると適合率が 6.27%，再現率は 5.38%の向上が見られる．実験 2 では，適合率は表 6 のようになり，提案手法（第 1 段落）と提案手法（第 1 段落＋第 2 段落）を比べると 9.12%の向上が見られ，再現率は 6.71%の向上が見られる．

4.4 考　　察

見出し未知記事（5055 件）に対して，正解見出しの特徴語を抽出記事数を確認する．表 10 より，実験 1 では，提案手法とベースラインを比べると，正しい特徴語を抽出できた記事は，4.84%向上しており，提案手法が有効であることが言える．表 10 より，実験 2 では，提案手法 (第 1 段落) と提案手法（第 1 段落＋第 2 段落）を比べると，正しい特徴語を抽出できた記事は，2.58%向上しており，提案手法（第 1 段落）が有効と言える．

実験 1 の係り受け関係を利用した特徴語抽出について，考察を行う．表 11 より，優れている結果を見てみると，正解見出しに生じる 6 特徴語（映像, 番組, バラエティー, ＴＢＳ, 音楽, 系）中の正解見出しに含まれる特徴語 6 単語（番組/休養, ＴＢＳ/休養, バラエティー/休養, 音楽/休養, 系/休養, 映像/継続）が選ばれ，最大適合率 100%となる．生成された見出しでは，係り受け関係を使って入れ替えた単語（番組/休養，曲/継続）であり，正解の特徴語が含まれているので，正しく見出しを作成されたと言える．

表 12 より，正解見出しに生じる 3 特徴語（シリア, 人, ファイル）中の正解見出しに含まれる特徴語 1 単語（シリア/視察）が選ばれ，適合率 33.33%となる．ベースラインでは，正解見出しに含まれる特徴語 2 単語（シリア, 人）が選ばれ，適合率 66.67%となる．生成された見出しでは，係り受け関係を使って入れ替えた単語（シリア/視察）であり，正解見出しの特徴語を選んでいるのにも関わらず，ベースラインより劣っている．これは，係り受け関係で抽出できない単語がベースラインよりも多くなったため劣る結果になったと考えられる．

実験 2 の提案手法（第 1 段落）と提案手法（第 1 段落＋第 2 段落）について，考察を行う．表 13 より，優れている結果を見てみると，正解見出しに生じる 8 特徴語（エレベーター, 咲, 洲, 緊急, 庁舎, 大阪, 府, 強風）中の正解見出しに含まれる特徴語 8 単語（エレベーター/停止, 庁舎/停止, 大阪/停止, 強風/停止, 咲/停止, 洲/停止, 緊急/停止, 府/停止,）が選ばれ，最大適合率 100%となる．第 1 段落から生成された見出しでは，係り受け関係を使って入れ替えた単語（エレベーター/停止, 庁舎/停止）であり，正解の特徴語が含まれているので，正しく見出しが生成され，第 1 段落の方が検索範囲として有効であると言える．

表 14 より，正解見出しに生じる 8 特徴語（新宿, 年頭, 東京, 警視庁）中の正解見出しに含まれる特徴語 2 単語（副作用, 年, 日本, 脳炎, 件, ワクチン, 改良, 型）が選ばれ，適合率 50%となる．ベースラインでは，正解見出しに含まれる特徴語 7 単語（ワクチン/死亡, 脳炎/死亡, 副作用/死亡, 日本/死亡, 型/中止, 年/中止, 件/中止,）が選ばれ，適合率 87.5%となる．第 1 段落から生成された見出しでは，係り受け関係を使って入れ替えた単語（脳炎/死亡）であり，第 1 段落＋第 2 段落から生成された見出しでは，入れ替えた単語（ワクチン/死亡, 労働省/中止）である．正解見出しに含まれる特徴語を正しく選んでいるのにも関わらず，第 1 段落＋第 2 段落から生成された見出しの方が優れている結果となっている．見出しの特徴語を入れ替える箇所が，1 箇所と少なかったため，選ばれた見出しが悪かったと

表 7　生成見出しの正解例

見出し番号	生成見出し
1	ギリシャ：債務削減、８５％が参加　実行ライン超す　ギリシャ強制措置へ
2	大阪市：職員３条例提案　政治活動、懲戒免職も　【大阪】
3	ＭＢＳ漫才アワード：かまいたちが優勝
4	葬儀：栄一さん＝２日死去
5	堺・資産家２人殺害：女性遺棄　「遺体一晩中焼いた」　容疑者、入念に隠滅か　【大阪】
6	サッカー：ロンドン五輪アジア最終予選　大津を日本招集
7	広島国際アニメーションフェスティバル：映画６６作選ぶ――２３日から
8	交通事故：列車追突、２７人重軽傷　車線変更、トラックに――岡山・倉敷の山陽道　【大阪】
9	サッカー：ロンドン五輪アジア最終予選　Ｕ２３日本代表、２月試合でイラク破る
10	東京・池袋の女性殺害事件：９３年の殺人、石垣タイ潜伏　代理処罰を要請
見出し番号	正解見出し
1	ギリシャ：債務削減、保険金支払いへ　国際協会、デフォルトと認定
2	野田首相：大阪市の職員アンケート、評価避ける――参院予算委　【大阪】
3	ＮＨＫ上方漫才コンテスト：かまいたち優勝
4	お別れの会：松本栄一さん＝２月１日死去
5	堺・資産家２人殺害：西口被告、強盗殺人容疑で再逮捕
6	サッカー：ロンドン五輪アジア最終予選　日本４－０マレーシア　４発快勝
7	ベルリン国際映画祭：東日本大震災描いたアニメ、１４～１７歳向け表彰
8	鉄道事故：２トントラックとＪＲ特急が衝突　９人けが――兵庫・明石
9	サッカー：ロンドン五輪アジア最終予選　Ｕ２３代表、ドーハへ出発　【大阪】
10	沖縄・尖閣諸島：石垣市議ら４人、魚釣島に上陸

表 8　生成見出しの失敗例

見出し番号	生成見出し
1	きものの話：名匠の技、時を超え　染め・織り、貴重な映像ＤＶＤ化
2	取り調べ可視化：試行拡大を提言　「全面」は賛否――警察庁研究会
3	東日本大震災：行方不明の息子手掛けた「農業体験」引き継ぐ　「児童との約束守りたい」――宮城・南三陸町
4	シリア：米仏２記者死亡
5	東京・多摩地域の下水道談合：ゼネコン側逆転敗訴　「合意は不当取引制限」――最高裁
6	ドイツ：大統領にガウク氏選出へ　与野党合意、統一後初の「東」出身
7	貿易統計：赤字１．４兆円　１月過去最大、リーマン後上回る
8	電視：“食”で表した人となり　映画「キツツキと雨」の沖田監督　【大阪】
9	ゴルフ：ノーザントラスト・オープン　遼、予選通過確実　第２Ｒ通算１オーバー　【大阪】
10	原発検査：「丸写し」０３年設立以来　第三者委、あす改善要請　報告書「理解と意識、希薄」
見出し番号	正解見出し
1	小学館：ドラえもん映像修正
2	国家戦略会議：委員定年制を提言　２０５０年の将来像――フロンティア分科会
3	海難事故：作業船が沈没、２８歳子供不明――高松港沖　【大阪】
4	シリア：治安被弾、１０人が死亡――首都近郊
5	生存権訴訟：生活保護の老齢加算廃止は「合憲」　最高裁初判断、東京敗訴が確定
6	ロシア：大統領大統領就任　抗議デモ、１００人以上拘束
7	貿易概況：近畿、貿易赤字１２６１億円　過去２番目、赤字連続　【大阪】
8	映画：仏映画「サラの鍵」のブレネール監督　迫害ユダヤ人の一人一人にドラマ　沖田価値届けたかった
9	ゴルフ：ノーザントラスト・オープン　石川、乱れて７２位　【大阪】
10	福島第１原発：委員規定違反、当初の再発防止策「不十分」――保安院　【大阪】

表 9　考察-実験 1（全体）

提案手法	正解数	正解率
特徴語を抽出できた記事	4712/5055	93.21
ベースライン	正解数	正解率
特徴語を抽出できた記事	4467/5055	88.37

表 10　考察-実験 2（全体）

提案手法（第１段落）	正解数	正解率
特徴語を抽出できた記事	4712/5055	93.21
提案手法（第２段落）	正解数	正解率
特徴語を抽出できた記事	4581/5055	90.62

考えられる．

5. 結　　論

本稿では，潜在意味解析を用い，類似した本文を選択し，その記事の見出しを利用することで新たな見出しを生成する手法を提案した．実験では，提案手法として係り受け関係を利用した特徴語抽出を行ったが，ベースラインに優る結果となり，見出しと特徴語を組み合わせることにより，適合する見出しを提示できた．実験 1 では，係り受け関係を用いる場合，最大適合率が 50%と向上することができた．また，提案手法（第 1 段落）と提案手法（第 1 段落＋第 2 段落）を比べることで，検索空間として有効なものは，短い内容なのか，長い内容なのかを確認した．提案手法（第 1 段落）が提案手法（第 1 段落＋第 2 段落）よりも最大適合率 63%向上し，短い内容が優れていることを確認できた．

文　　献

[1] Stephen Wan, Robert Dale Mark, and Ccile Paris. ”Statistically generated summary sentences: A preliminary evaluation using a dependency relation precision metric.”

[2] Wan, Xiaojun, and Jianmin Zhang. ”Ctsum: extracting more certain summaries for news articles.”

[3] Ouyang You, Sujian Li, and Wenjie Li. ”Developing learning strategies for topic-based summarization.”

[4] Dingding Wang, Tao Li, Shenghuo Zhu, and Chris Ding. ”Multi-Document Summarization via Sentence-Level Semantic Analysis and Symmetric Matrix Factorization.”

[5] Yan Liu, Sheng-hua Zhong, and Wenjie Li. ”Query-Oriented Multi-Document Summarization via Unsupervised Deep Learning.”

[6] Das, D. and Martins, A.F.T: A Survey on Automatic Text Summarization Tech. Report, Univ. of Duisburg-Essen, 2007

[7] Kleinbaum, D.G. and Klein, M. : Logistic Regression: A Self-Learning Text (Statistics for Biology and Health), Springer-Verlag, 2010

[8] 奥村 学, 難波 英嗣: テキスト自動要約 (知の科学)

[9] Pedersen, T., Patwardhan, S., and Michelizzi, J. (2004,

表 11　考察-実験 1-優れている例（提案手法“係り受け関係”VS ベースライン）

正解見出し	ＴＢＳ系番組：音楽バラエティー、映像途切れる
正解特徴語	映像, 番組, バラエティー, ＴＢＳ, 音楽, 系
選択見出し	やしきたかじんさん：レギュラー番組を休養　「委員会」不在で放送継続　【大阪】
作成見出し	やしきたかじんさん：番組 TBS を休養　「委員会」不在で曲継続　【大阪】
提案手法の特徴語	番組/休養, ＴＢＳ/休養, バラエティー/休養, 音楽/休養, 系/休養, 曲/継続, 映像/継続, 日/継続, 一時/継続, 夜/継続, 中/継続, 火曜/継続,
ベースラインの特徴語	番組, ＴＢＳ, 広報, 分, 画面, 一時, 音声, 黒, 事故, バラエティー, トラブル
	適合率
提案手法	6 語/6 語
ベースライン	3 語/6 語

表 12　考察-実験 1-劣っている例（提案手法“係り受け関係”VS ベースライン）

正解見出し	ファイル：「シリア弾圧で４００人死亡」
正解特徴語	シリア, 人, ファイル
選択見出し	シリア：国連事務次長、ホムスを視察　【大阪】
作成見出し	シリア：国連事務次長、シリアを視察　【大阪】
提案手法の特徴語	シリア/視察, 先月/視察, アラブ/視察, 団/視察, 下旬/視察, 連盟/視察, 氏/視察
ベースラインの特徴語	国連, 先月, シリア, 人, 下旬, 理事, 連盟, 氏, 政治
	適合率
提案手法	1 語/3 語
ベースライン	2 語/3 語

表 13　考察-実験 2-優れている例（第 1 段落 VS 第 1 段落+第 2 段落）

正解見出し	大阪府咲洲庁舎：強風で揺れる　エレベーターが緊急停止
正解特徴語	エレベーター, 咲, 洲, 緊急, 庁舎, 大阪, 府, 強風
第 1 段落選択見出し	フジテレビ：映像一時停止
第 1 段落作成見出し	フジテレビ：エレベーター庁舎停止
提案手法第 1 段落の特徴語	エレベーター/停止, 庁舎/停止, 大阪/停止, 強風/停止, 昨年/停止, しま/停止, 市/停止, 日/停止, 分間/停止, 東日本/停止, 原因/停止, 住之江/停止, 咲/停止, 洲/停止, こと/停止, 緊急/停止, 基/停止, 区/停止, 大震災/停止, 府/停止, さき/停止,
第 2 段落選択見出し	トラブルで停止
第 2 段落作成見出し	大阪で停止
提案手法第 2 段落の特徴語	大阪/停止, 強風/停止, 昨年/停止, 市/停止, こと/停止, 日/停止, 基/停止, 区/停止, 原因/停止, 府/停止,
	適合率
提案手法	8 語/8 語
ベースライン	3 語/8 語

表 14　考察-実験 2-劣っている例（第 1 段落 VS 第 1 段落+第 2 段落）

正解見出し	日本脳炎：ワクチン副作用１１件　０９年改良型に切り替え後
正解特徴語	副作用, 年, 日本, 脳炎, 件, ワクチン, 改良, 型
第 1 段落選択見出し	日本脳炎：予防接種で死亡・重症　厚労相「調査迅速に」
第 1 段落作成見出し	日本脳炎：脳炎で死亡・重症　厚労相「調査迅速に」
提案手法第 1 段落の特徴語	脳炎/死亡, 年/死亡, 日/死亡, 件/死亡, 性/死亡, 会/死亡, 型/死亡
第 2 段落選択見出し	日本脳炎：予防接種死亡　原因究明へ接種中止を
第 2 段落作成見出し	日本脳炎：ワクチン死亡　原因究明へ労働省中止を
提案手法第 2 段落の特徴語	ワクチン/死亡, 脊髄/死亡, 脳炎/死亡,, 副作用/死亡, 完治/死亡, 日/死亡, 日本/死亡, 性/死亡, 労働省/中止, 型/中止, 急性/中止, 厚生/中止, こと /中止, せき/中止, 年/中止, 安全/中止, 件/中止
	適合率
提案手法	4 語/8 語
ベースライン	7 語/8 語

May). WordNet:: Similarity: measuring the relatedness of concepts. In Demonstration Papers at HLT-NAACL 2004 (pp. 38-41). Association for Computational Linguistics

[10] Stephen Wan Robert Dale Mark, and Ccile Paris. ”Statistically generated summary sentences: A preliminary evaluation using a dependency relation precision metric.”

[11] Saranyamol, C.S. and Sindhu, L.: A Survey on Automatic Text Summarization International Journal of Computer Science and Information Technologies (IJCSIT), Vol. 5 (6), 2014, pp. 7889-7893

[12] Yuen-Hsien Tseng, Chi-Jen Lin, Hsiu-Han Chen, Yu-I Lin: Toward Generic Title Generation for Clustered Documents, Third Asia Information Retrieval Symposium, AIRS 2006, Singapore, October 16-18, 2006, pp.145-157