

## 話題の推移に基づく続報記事の自動抽出

福本文代<sup>†</sup> 鈴木良弥<sup>†</sup> 山田寛康<sup>††</sup>

ニュースなどの報道記事から特定の時間に特定の場所で生じた出来事に関する続報記事を自動的に抽出するためには、出来事に関する話題が時間の経過とともに変化する過程を的確にとらえる必要がある。本稿では、日付順に並べられた記事に対して、話題の推移を考慮した続報記事の抽出手法を提案する。我々は、特定の日に報道された記事と同様の話題を持つ記事（記事集合）は、記事の話題が出来事の生起を伝える第1記事のほか、その記事と時間的に近い時期に比較的多く存在すると仮定し、類似した話題を持つ記事集合をその時点から過去に遡り推定するという方法を用いた。続報記事の抽出は、得られた記事集合を基に学習を行った後、テスト記事が続報であるか否かを判定することにより行われる。この処理はテスト記事が続報であると判定されるたびに、その記事を訓練記事集合に追加し、再度最適な記事集合を推定するという過程を繰り返すことにより行われる。TDT1コーパスを用いて実験した結果、訓練記事中の正例記事数が16記事の場合に、F値0.78の精度で続報記事を判定でき、また、異なる話題を多く含む出来事に対しても高精度で判定できることが分かった。

### Detecting Shifts in News Stories for Event Tracking

FUMIYO FUKUMOTO,<sup>†</sup> YOSHIMI SUZUKI<sup>†</sup> and HIROYASU YAMADA<sup>††</sup>

For event tracking where data is collected over an extended period of time, the discussion of an event, i.e., the *subject* in a story changes over time. This paper focuses on subject shift and presents a method for event tracking on broadcast news stories to recognise and handle subject shift. The basic idea is to automatically adjust the optimal window size in the training data of the target event so as to include only the data which are sufficiently related to the current subject. The method was tested on the TDT1 which has been developed by the TDT (Topic Detection and Tracking) project, and the results show that when the optimal window size is selected, the system works well even for a large number of stories, each of these discusses a new subject related to the target event.

#### 1. はじめに

インターネットの普及により、日々報道されるニュースがオンライン上で利用可能となっている。ニュースなどで報道される出来事を容易に把握するための手段の1つとして、ユーザが指定した出来事に対しその経過を示す一連のニュース記事を自動的に提示する技術への関心が高まっている<sup>1)~3)</sup>。その先駆者的な存在としてAllanらを中心とした米国DARPA主催のTDT (Topic Detection and Tracking) プロジェクトがある<sup>4)</sup>。TDTプロジェクトではCNN、およびロイター通信 (Reuters) などの報道記事を対象とし、(1) 記事の

話題による分割 (Segmentation)、(2) 特定の時間に特定の場所で生じた出来事の抽出 (Detection)、(3) 出来事の経過を示す記事 (続報記事) の抽出 (Tracking) から成る3つのタスクを設定している。このうち、(3) の続報記事の抽出タスクでは、関西地区で起きた大震災やニューヨークの地下鉄爆破事件など、あらかじめ用意された出来事に対し、一定量の訓練記事を用いて学習を行った後、テスト記事が対象とする出来事の続報か否かを判定する。本研究はTDTプロジェクトで設定された3つのタスクのうち、(3) のタスク、すなわち続報記事を自動的に抽出することを目的とする。

続報記事の抽出手法は、従来より統計手法や機械学習を用いて記事中の語の特徴抽出を行った結果を利用して、テスト記事が続報か否かを判定する方法が多く提案されている<sup>5)~9)</sup>。Allanらは、日付順に並べられた記事において、特定の時期にそれまでは出現しなかった人物や場所などを表す語が出現した場合、これらをSurprising wordsと呼び、これらの統計量を用いてテ

<sup>†</sup> 山梨大学工学部コンピュータ・メディア工学科  
Department of Computer Science and Media Engineering,  
Faculty of Engineering, Yamanashi University

<sup>††</sup> 北陸先端科学技術大学院大学情報科学研究科  
Department of Information Science, Japan Advanced  
Institute of Science and Technology

スト記事が続報記事であるか否かを判定する手法を提案している<sup>10)</sup>。

一方、機械学習を用いて続報記事判定に必要なモデルを学習し、その結果を利用してテスト記事が続報記事であるか否かを判定する方法も多数報告されている<sup>4)</sup>。Carbonellらは、*k-Nearest Neighbor* (*kNN*)と決定木学習の2種類の方法を用いてテスト記事が続報であるか否かを判定している<sup>11)</sup>。しかし、TDTによる続報記事抽出タスクでは、出来事を示す訓練記事(正例記事)の個数は、その出来事でない記事(負例記事)数に比べ極端に少ない。したがって、訓練データに対してこれら機械学習を適用し学習を行う場合、大量の負例記事の中から続報記事判定に必要な負例記事のみを抽出し使用することが望ましい。Yangらはこの点に注目し、*kNN*法とRocchio法を組み合わせた新しい機械学習法を提案し、これを用いて続報記事の判定を行っている<sup>12)</sup>。2つの機械学習を組み合わせることで訓練データの中から最適な負例記事集合のみ抽出し、これを用いてテスト記事の判定を行っている。

一般に特定の時間に特定の場所で生じた出来事とその続報を伝える報道記事において、ある出来事が生じたことを最初に伝える記事では、出来事が起こったことがその記事の話題であるため、出来事と話題は一致する。その後出来事は推移し、その結果が続報記事の話題として報告されるため、続報記事では出来事は記事の背景となり、記事の話題とは異なる場合が多い。たとえば日付順に並べられたTDT1コーパスの関西地区での地震に関する記事とその続報記事では、第1記事において神戸で地震が生じたことを伝え、続く記事で地震の規模や被害の状況について触れ、さらに時間の経過とともに、地震調査者が被害状況を本格的に調査する内容へと話題が推移する。したがって、Yangらの手法のように、与えられた訓練記事中のすべての正例記事を用いる場合、長期間にわたり報道されているような出来事に対して、続報記事が正確に抽出できない場合が生じる。

本稿では、話題の推移を考慮した続報記事の抽出手法を提案する。本稿の主旨は、ある時点で報道された記事の話題を現在の話題とすると、それ以降に出現する記事が続報か否かを判定するために、現在の話題と関係する記事(記事集合)のみを訓練記事から抽出して用いるということである。ここで記事の話題は、記者がその記事を通して読者に伝えたいことと定義する。また、ある時点で報告された記事*d*と同じ話題を持つ記事、または記事集合は、出来事が生じたことを伝える第1記事のほか、それまでの出来事に関する記

事の中で、記事*d*と時間的に近い時期に比較的多く存在すると仮定する。したがって、次に報道される記事が続報記事であるか否かを判定するためには、それまでに報道された記事のうち、記事*d*と同じ話題を持つ記事(記事集合)を用いて判定すればよい。我々は、日付順に並べられた出来事に関する訓練記事に対し、テスト記事が続報であるか否かを判定するために必要な記事集合、および続報記事判定に必要なモデルを機械学習SVMs(Support Vector Machines)を用いて学習した。続報記事を判定するために必要な記事集合は、*Leave-One-Out* (*LOO*)と呼ばれる処理の結果得られるエラー率を利用した。*LOO*は、*l*個の訓練記事集合の中から1つをテスト記事として抽出し、残りの*l-1*個の訓練記事集合を用いて学習するという作業をすべての訓練記事について*l*回繰り返す方法である<sup>13)</sup>。このときに得られるエラー率は、複数の学習モデルの中から、未知のテストデータに対してより効果的な推定を行うモデルを選択する指標として使用される。我々は、*LOO*におけるエラー率を利用し、記事*d*と同じ話題を持つ記事(記事集合)を抽出し、これを用いて続報記事の判定に必要なモデルを学習した。テスト記事が続報記事であるか否かは、このモデルを用いて判定される。テスト記事が続報記事であると判定されるたびに、その記事を訓練記事集合に加え、再度最適な記事集合を推定するという処理を繰り返すことで続報記事の判定を行った。

2章では、Vapnikにより提案された機械学習SVMs、およびエラー率推定のために用いる*LOO*について説明する。3章ではSVMsを用いて続報記事判定のための最適な記事集合を推定する手法について述べる。4章では実験について報告し、終章でまとめと今後の課題について述べる。

## 2. SVMs

Support Vector Machines(SVMs)は、Vapnikにより提案された学習アルゴリズムであり、*n*次元空間上の正(+1)と負(-1)から成る訓練事例集合を2分する分離超平面 $w \cdot x + b = 0$ を見つけるためのアルゴリズムである( $w, x \in \mathbb{R}^n$ )<sup>14),15)</sup>。正負のクラスを分離する超平面は多数存在する。SVMsは図1の破線で示されるように、求める分離超平面と並行で等間隔の2つの平面間の距離(マージン)を最大にするような*w*, *b*を求める。

今*l*個の訓練事例集合 $L = \{(x_1, y_1), (x_2, y_2), \dots, (x_l, y_l)\}$ が与えられたとする( $x_i \in \mathbb{R}^n, y_i \in \{+1, -1\}$ )。マージンを最大にする平面は、式(1)で示さ

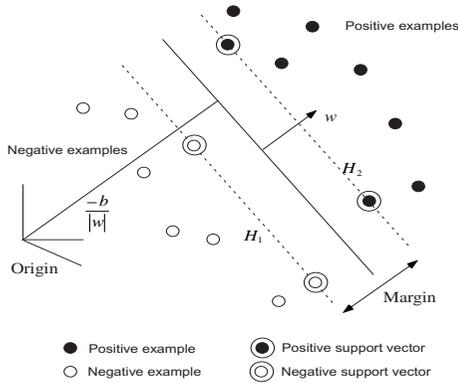


図 1 線形の場合における分離超平面

Fig. 1 The decision surface of a linear SVM.

れた値を式 (2) で示す条件で最大化する制約付き二次計画問題を解くことで得られる。式 (2) は与えられた訓練事例をできるだけ正確に分離することを意味する制約である。

$$\text{minimize} \quad \frac{1}{2} \|\mathbf{w}\|^2 \quad (1)$$

$$\text{s.t.} \quad y_i (\mathbf{w} \cdot \mathbf{x}_i + b) \geq 1, \forall i \quad (2)$$

この問題に対し Lagrange の未定乗数を使用し双対問題を解くことで、最終的に式 (3) で示す判別関数を得る。

$$\begin{aligned} f(\mathbf{x}) &= \mathbf{w} \cdot \mathbf{x} + b \\ &= \sum_i \alpha_i y_i \mathbf{x}_i \cdot \mathbf{x} + b \end{aligned} \quad (3)$$

ここで  $\alpha_i$  は式 (2) で示した制約に対する Lagrange 乗数である。SVMs は未知の事例  $\mathbf{x}$  のラベルを、式 (3) の符号により決定する。図 1 の 2 重で囲まれた事例を Support Vector と呼ぶ。Support Vector は式 (3) で  $\alpha_i > 0$  となる事例を指す。式 (3) から  $\alpha_i = 0$  となる訓練事例は分類にまったく影響を与えないことが分かる。

我々は、日付順に並べられた出来事に関する訓練記事に対し、テスト記事が続報であるか否かを判定するために必要な記事集合、および続報記事判定に必要なモデルを SVMs を用いて学習する。続報記事を判定するために必要な記事集合は、Leave-One-Out (LOO) と呼ばれる処理の結果得られるエラー率を利用した。LOO は、与えられた  $l$  個の訓練記事集合  $L$  の中から 1 つをテスト記事、残り  $l-1$  を訓練記事とし学習および分類を行い、この処理をすべての訓練記事に対し  $l$  回繰り返し行う手法である。このときのエラー率を  $E_{LOO}(L)$  とすると、 $E_{LOO}(L)$  は学習モデルが未

知のデータに対してどの程度精度良く判定できるかを測る指標として使用できる。しかし、LOO を用いてエラー率を求める場合、与えられた訓練事例の数だけ繰り返し学習を行う必要があり、非効率である。

SVMs では Support Vector 以外の訓練事例、すなわちマージンの外にある訓練事例は、LOO を行う際に、テスト事例として選択したとしても、分離平面はまったく変化せず、エラーとはならない。したがって、任意の Support Vector について、各々をテストデータと見なし判別関数を用いて判定した結果、それらすべてが誤って判定された場合に、LOO のエラー率は上限値となる。いま  $l$  個の訓練記事の集合  $L$  を用いて SVMs により学習した結果、 $m$  個の Support Vector を得たとする。 $m$  個の Support Vector に対して LOO を適用した結果得られるエラーの個数を  $m'$  個とすると、LOO によるエラー率  $E'_{LOO}(L)$  は、式 (4) で示される。

$$E'_{LOO}(L) = \frac{m'}{l} \quad (4)$$

我々は候補となる 2 種類の訓練記事集合の  $E'_{LOO}(L)$  を比較することにより、現在話題となっている記事と同じ話題を持つ記事 (集合) を抽出した。

### 3. 記事集合の推定と続報記事の判定

日付順に並べられた訓練記事中の正例記事集合  $\{d_1, \dots, d_{p-1}, d_p\}$  において、記事  $d_p$  の話題を現在の話題とすると、それ以降に出現する任意のテスト記事が与えられた出来事の続報記事であるか否かは、 $\{d_1, \dots, d_{p-1}, d_p\}$  のうち、記事  $d_p$  の話題と類似した話題を持つ記事のみを用いて判定する。我々は、任意のテスト記事が、与えられた出来事の続報記事であるか否かを判定するため、正例記事集合に対して、以下の 3 点を仮定した。

- 仮定 1: 特定の時間に特定の場所で生じた出来事に関する話題は、時間の経過とともに推移する。
- 仮定 2: 記事  $d_p$  と同じ話題について述べている記事、または記事集合は、 $d_p$  と時間的に近い時期に報道された記事 (集合) である可能性が高い。
- 仮定 3: 記事  $d_1$  は、出来事が生じたことを伝える記事であると同時にそれ以降に出現する続報記事の背景を示す記事となるため、続報記事集合  $\{d_2, \dots, d_{p-1}, d_p\}$  の各記事すべてに関係する。

今、 $d_1, \dots, d_{p-1}$ 、および  $d_p$  を日付順に並べられた訓練記事中の正例記事とし、 $\bar{d}_1, \dots, \bar{d}_{q-1}$ 、および  $\bar{d}_q$  を負例記事とする。訓練記事中に属する正例記事集合の中で、テスト記事が続報であるか否かを判断す

```

begin
do (1) k-means の再帰的な利用による訓練記事の分類
do (2) 最適な正例記事集合の推定
for i=1 to n
do (3) 続報記事の判定;
if i が続報(正例)記事である
then i を bottom cluster に追加
do (2) 最適な正例記事集合の推定
end_if
end_for
end
end
    
```

図 2 正例記事集合の推定と続報記事判定のアルゴリズム  
 Fig. 2 Estimating optimal positive training data and tracking.

るために必要となる正例記事の集合, すなわち  $d_p$  との話題と類似した話題を持つ正例記事の集合を最適な正例記事集合と呼ぶことにする. 最適な正例記事集合の推定と続報記事判定のアルゴリズムを図 2 に示す.

(1)  $k$ -means<sup>16)</sup> の再帰的な利用による訓練記事の分類

TDT の続報記事抽出タスクで設定された負例記事の個数は, 正例記事の個数に比べ, 極端に多い<sup>4)</sup>. したがって, 与えられた負例記事を効果的に用いるためには, 負例記事と正例記事との類似性, すなわち負例記事が正例記事にどのくらい似ているかという基準を用いてあらかじめ分類し, その結果を続報記事抽出に利用することが望ましい<sup>12)</sup>. 我々は与えられた正例と負例から成る訓練記事に対して  $k$ -means ( $k = 2$ ) を再帰的に適用することで, 記事集合をノードとする階層構造を作成し, テスト記事判定に利用した. 階層構造は以下のようにして作成される.

- (a)  $k = 2$  とし, 2 つの代表点を決める. 本手法では, 負例記事が正例記事にどのくらい似ているかという観点でクラスタリングを行う. したがって, 代表点のうちの 1 つは正例記事 (第 1 記事) とし, 残りの 1 点は, 正例記事集合の重心  $g$  と各負例記事との距離を計算し, 重心からの距離が最も離れている負例記事とした. 訓練記事中の単語の総異なり数を  $n$  とし, 各記事を  $n$  次元ベクトルで表現する.

$$\cos(\mathbf{X}, \mathbf{Y}) = \frac{\mathbf{X} \cdot \mathbf{Y}}{\|\mathbf{X}\| \|\mathbf{Y}\|}$$

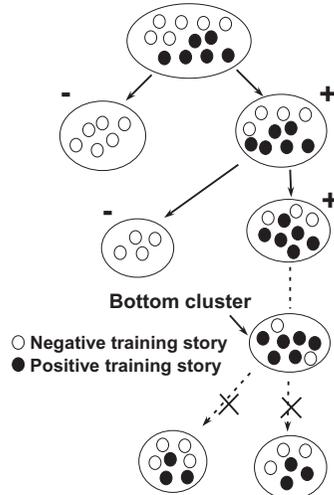


図 3  $k$ -means により得られた階層構造  
 Fig. 3 Hierarchical structure with  $k$ -means ( $k = 2$ ).

$$= \frac{\sum_{w=1}^n tf_X(w) \cdot tf_Y(w)}{\sqrt{\sum_{w=1}^n tf_X^2(w)} \cdot \sqrt{\sum_{w=1}^n tf_Y^2(w)}} \quad (5)$$

式 (5) において  $tf_X(w)$  と  $tf_Y(w)$  は記事ベクトル  $\mathbf{X}$  と  $\mathbf{Y}$  における単語  $w$  の出現頻度を表す. 我々は, 各負例記事と正例記事集合の重心ベクトルとの余弦尺度  $\cos(g, \bar{d}_j)$  ( $1 \leq j \leq q$ ) を用いて 2 点間の距離を計算した. ここで  $g$  は, 式 (6) で示される.

$$g = (g_1, \dots, g_n) = \left( \frac{1}{p} \sum_{i=1}^p d_{i1}, \dots, \frac{1}{p} \sum_{i=1}^p d_{in} \right) \quad (6)$$

$d_{ij}$  ( $1 \leq j \leq n$ ) は正例記事ベクトル  $d_i$  における単語  $j$  の出現頻度数を示す.

- (b) 訓練記事集合に対して  $k$ -means ( $k = 2$ ) を適用する.
- (c) 正例記事集合が 2 つのクラスタに分割されたら終了, そうでなければ 2 つのクラスタのうち, 正例記事を含むクラスタに対して (a) および (b) を繰り返す.

図 3 は  $k$ -means ( $k = 2$ ) により得られた階層構造を示す. 次処理である '(2) 最適な正例記事集合の推定' は, 図 3 の bottom cluster で示される記事集合に対して行われる.

(2) 最適な正例記事集合の推定

図 2 においてテスト記事の個数を  $n$  とすると, 各テスト記事が続報記事であるか否かは,

$\{d_1, \dots, d_{p-1}, d_p\}$  の各記事の中で記事  $d_p$  と同じ話題である記事の集合を用いて判定される。これは、仮にテスト記事が続報記事である場合、テスト記事は時間的に最も近い時期に報道された記事  $d_p$  と類似していると考えられるためである。最適な正例記事集合は、 $d_p$  が (i) それより以前の正例記事と同じ話題であるか、(ii) 新しい話題であるかを判定することにより求められる。我々は、式 (4) で示されるエラー率  $E'_{LOO}(L)$  を用いてこの判定を行った。

今、 $q$  個から成る負例記事が図 3 で示される bottom cluster に含まれているとする。各負例記事と正例記事集合の重心との距離を求め、その値が昇順になるように並べた負例記事を  $\bar{d}_1, \dots, \bar{d}_{q-1}$ , および  $\bar{d}_q$  とする。さらに、 $Set_1$  を  $d_1, d_p, \bar{d}_1$ , および  $\bar{d}_2$  から成る集合とし、 $Set_2$  を  $d_{p-1}, d_p, \bar{d}_1$ , および  $\bar{d}_2$  から成る集合とする。ここで、エラー率を判定する際に用いる負例記事の個数はつねに正例記事の個数に合わせた。今、仮に  $E'_{LOO}(Set_2)$  の値が  $E'_{LOO}(Set_1)$  よりも小さい場合、 $d_p$  はその直前の正例記事  $d_{p-1}$  と同様の話題を持つと判定する。これは、 $Set_2$  を用いて学習したモデルは、 $Set_1$  で学習したモデルよりも未知のテストデータに対して低いエラー率であると推定されているためである。この場合、 $d_p$  は (i) であると判定されたため、我々は、仮定 2 より、正例記事集合の中から  $d_p$  と同じ話題を持つ正例記事を現在から過去に遡り 1 記事ずつ増やすことで、最適な正例記事集合を推定した。図 4 に推定のためのアルゴリズムを示し、図 5 に最適な正例記事集合の選択例を示す。図 4 において  $Set_a$  および  $Set_b$  は仮定 3 よりつねに記事  $d_1$  を含む。また、 $E'_{LOO}(Set_a)$ , および  $E'_{LOO}(Set_b)$  は式 (4) により求めた LOO におけるエラー率を示す。図 5 中、 $a$  と  $b$  は、それぞれ図 4 における  $Set_a$  と  $Set_b$  から  $d_1, \bar{d}_1, \dots, \bar{d}_{k+1}$ , および  $\bar{d}_{k+2}$  を略した集合を示す。図 5 においてたとえば  $k = 2$  における  $a$  と  $b$  の比較で  $a$  が選択された場合には最適な正例記事集合は  $\{d_1, d_{p-2}, d_{p-1}, d_p\}$  となる。一方、 $E'_{LOO}(Set_2)$  の値が  $E'_{LOO}(Set_1)$  よりも大きい場合には、 $d_p$  は (ii), すなわち新しい話題であると判定される。この場合、我々は、すべての正例記事を最適な正例記事集合として使用した。これは、記事  $d_p$  は新しい話題に関する

```

begin initialize Set ← φ
for k = 1 to p-3
  Seta および Setb をそれぞれ,
  Seta = {d1, dp-k, dp-k+1, ..., dp,  $\bar{d}_1, \dots, \bar{d}_{k+1}, \bar{d}_{k+2}$ }.
  Setb = {d1, dp-k-1, dp-k, ..., dp-1,  $\bar{d}_1, \dots, \bar{d}_{k+1}, \bar{d}_{k+2}$ }
  とする.
  if E'_{LOO}(Seta) < E'_{LOO}(Setb)
    then Set ← Seta exit loop
  end_if
end_for
if Set = φ
  then Set ← {d1, d2, ..., dp,  $\bar{d}_1, \bar{d}_2, \dots, \bar{d}_p$ }
end_if
end
    
```

図 4 LOO による正例記事集合の推定アルゴリズム  
Fig. 4 Flow of optimal positive training data with LOO.

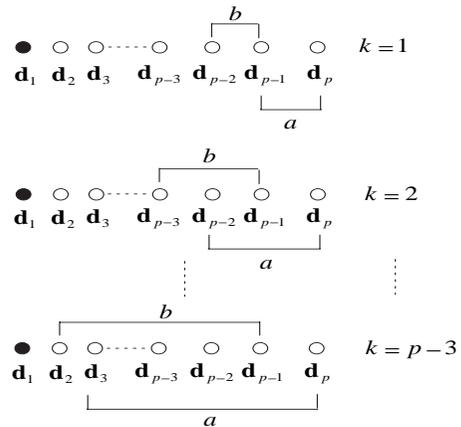


図 5 正例記事集合の選択  
Fig. 5 Extracting positive training data.

記事であるため、 $d_1, \dots, d_{p-2}$  および  $d_{p-1}$  に共通して出現する語は出来事背景を示す語のみとなる可能性が大きいためである。したがって、記事  $d_p$  のみを用いて後続する未知のテスト記事が続報か否かを判定するよりも、より多くの正例記事を最適な正例記事集合として用いた方が  $d_p$  に出現する語をより多く含む可能性があり、結果的にデータスパースネスによる判定の誤りを防ぐことができると考えたためである。

(3) 続報記事の判定

図 4 において、推定の結果得られた記事集合 ( $Set$ ) を bottom cluster における学習用の訓練データとする。図 3 で示される階層構造の各分岐点で得られる 2 つのクラスタに対し、SVMs を適用しモデルを作成する。各分岐点における SVMs の訓練データは、2 つのクラスタのうち、正例記事を含むクラスタに含まれる各記事を正例、他方のクラスタに含まれる記事を負例とす

表 1 出来事  
Table 1 Event name.

Event 番号	出来事	Event 番号	出来事
1	Aldrich Ames	14	Karrigan/Harding
2	Carlos the Jackal	15	Kobe Japan quake
3	Carter in Bosnia	16	Lost in Iraq
4	Cessna on White House	17	NYC Subway bombing
5	Clinic Murders	18	OK-City bombing
6	Comet into Jupiter	19	Pentium chip flaw
7	Cuban riot in Panama	20	Quayle lung clot
8	Death of Kim Jong Il	21	Serbians down F-16
9	DNA in OJ trial	22	Serbs violate Bihac
10	Haiti ousts observers	23	Shannon Faulkner
11	Hall's copter	24	USAir 427 crash
12	Humble, TX, flooding	25	WTC Bombing trial
13	Justice-to-be Breyer		

る．テスト記事は root ノードから下位の方向にモデルを適用することで判定する．テスト記事が任意のモデルにおいて負であると判定された場合には終了し，そうでない場合には下位に位置するモデルを用いて判定を行う．最下位である bottom cluster におけるモデルを用いて判定した結果，正例（続報）であると判定されたとき，その記事を bottom cluster に追加する．bottom cluster の記事集合に対して，‘(2) 最適な正例記事集合の推定’で示される処理を行い，終了する．

‘(3) 続報記事の判定’で示される処理を最終テスト記事  $n$  の判定が終わるまで繰り返すことで，続報記事の抽出を行う．

#### 4. 実験

提案手法の定量的評価を行うため，TDT1 コーパスを用いて実験を行った．評価は，以下の4点について行った．

- 訓練データ中の正例記事数の違いによる精度への影響
- 続報記事判定に用いる正例記事選択の有効性
- 他の Data management 手法との比較
- 関連研究との比較

##### 4.1 データ

実験では TDT プロジェクトで作成された TDT1 コーパスを用いた．TDT1 コーパスはロイター通信 (Reuters) 7,965 記事と CNN 7,898 記事から成り，これらの記事 (総計 15,863) は日付順に id 番号が付与されている．各記事は，あらかじめ設定された 25 の出来事の各々に対して，その続報記事である，すなわち

正例記事 (‘YES’) か，あるいは続報記事でない，すなわち負例記事 (‘NO’) であることを示すフラグが付与されている．日付順に並べられた記事に対して，先頭から一定の記事数を訓練データとする．訓練データは，正例記事と負例記事から成る．正例記事は，出来事の生起を伝える第 1 記事とその続報記事が含まれる．訓練で使用できる正例記事の個数は，1, 2, 4, 8, および 16 記事となっている．たとえば 16 個から成る正例記事 (出来事を伝える第 1 記事と 15 個から成る続報記事) の場合，訓練データは，これら 16 の正例記事のほか，負例記事として 16 番目の正例記事の日付以前に報道されたすべての負例記事から成る．テスト記事は，正例記事の個数によらず，つねに 16 番目の日付以降，最後の記事までとする．これは，訓練データ中の正例記事数の違いによる精度を比較するためである．表 1 に，25 種類の出来事を示す．

実験では，訓練，およびテストデータの各記事に対し，Tree tagger<sup>17)</sup>を用いて品詞付けおよび派生処理を行った結果から記号を除く全単語を抽出し，学習，および分類のための素性として用いた．実験結果の評価は TDT プロジェクトで作成された評価ツールを用いて行った．

##### 4.2 実験結果

###### 4.2.1 訓練データ中の正例記事数の違いによる精度への影響

実験結果を表 2 に示す．‘ $N_t$ ’は訓練データ中の正例記事数を示す．‘個数’は 25 種類の出来事のうち，‘ $N_t$ ’で示される正例記事数が訓練データ中に存在する出来事の個数を示す．‘Miss’は続報記事を続報記事でないとして判定した記事の割合を示し，‘F/A’は続報記事でない記事を続報記事であると誤って判定した割合を示す．

‘Prec’( $p$ ) は、本手法により続報記事であると判定された記事のうち実際に続報記事であった割合を示し、‘F’ は

$$F(r, p) = \frac{2rp}{r + p}$$

で示される。式中、 $r$  は Recall を示し、続報記事の中で実際に正しく続報記事であると判定された記事数の割合を示す。‘Miss’、‘F/A’、‘Prec’、および F の各値は、‘個数’で示されたすべての出来事に対する micro-average の値を示す。

表 2 より、F 値は正例記事数が 16 記事のときに最

表 2 続報記事の抽出結果  
Table 2 Tracking results.

$N_t$	個数	Miss	F/A	Prec	F
1	25	40%	0.19%	69%	0.64
2	25	31%	0.14%	76%	0.73
4	25	28%	0.14%	77%	0.74
8	23	25%	0.14%	77%	0.76
16	15	22%	0.13%	79%	0.78

高となり 0.78 であった。正例記事数が 2 記事以上の場合において、記事数の増加にともない Miss 率、および F/A が減少している。しかし F/A の減少率は 0.13~0.14 でありわずかであることから、本手法は、正例記事数が少ない場合にも続報でない記事を正しく判定することができているといえる。

#### 4.2.2 続報記事判定に用いる正例記事選択の有効性

正例記事選択の有効性を検証する。表 3 は  $N_t = 16$  において抽出された正例記事集合を用いて判定を行った結果を示す。出来事は ‘Kobe Japan Quake’ であり、日付順に並べられたテスト記事において正例と判定されるべき記事のうち、最初の 30 記事の結果を示す。1 から 3 列目はそれぞれ、id 番号、日付、および記事のタイトルを示す。たとえば ‘id = 1’ は、‘Kobe Japan Quake’ の第 1 記事を示す。4 列目と 5 列目はそれぞれ本手法の結果 (system) と人手による評価結果 (actual) を示す。人手による評価では、新しい話題か否かについて 3 名の多数決により判定を行った。本手法と評価者における ‘YES’ はテスト記事がそれ

表 3 ‘Kobe Japan Quake’ における続報記事の判定結果  
Table 3 The tracking result in Event 15, ‘Kobe Japan Quake’.

id	日付	タイトル	話題の推移		正例記事集合の精度		
			system	actual	recall	precision	F
1	01/17/95	Kobe Residents Unable to Commence Rescue Operations	NEW	NEW	-	-	-
2	01/17/95	Emergency Efforts Continue After Quake in Japan	NEW	YES	-	-	-
3	01/17/95	Japan Helpline Worker Discusses Emergency Efforts	NEW	NEW	-	-	-
4	01/17/95	U.S. Businessman Describes Japan Earthquake	NEW	YES	-	-	-
5	01/17/95	Osaka, Japan, Withstands Earthquake Better Than Others	NEW	NEW	-	-	-
6	01/17/95	President Clinton Drums Up Support in Humanitarian Trip	NEW	NEW	-	-	-
7	01/17/95	Engineer Examines Causes of Damage in Japan Quake	NEW	NEW	-	-	-
8	01/18/95	Mike Chinoy Updates Japan’s Earthquake Recovery Efforts	NEW	YES	-	-	-
9	01/18/95	Smoke Hangs in a Pall Over Quake-, Fire-Ravaged Kobe	NEW	NEW	-	-	-
10	01/18/95	Japanese Wonder If Their Cities Are Really ‘Quakeproof’	NEW	NEW	-	-	-
11	01/18/95	Cooperative Plan Allows U.S. to Help Quake Victims	NEW	NEW	-	-	-
12	01/18/95	Vivid Descriptions Tell Story of Kobe Earthquake	YES	YES	0.50	0.29	0.36
13	01/18/95	Kobe Neighbors Get a Lot Closer in Quake’s Aftermath	YES	YES	0.75	0.43	0.55
14	01/18/95	Worldwide Help Offered to Japan in Quake’s Aftermath	YES	YES	1.00	0.25	0.40
15	01/18/95	Kobe Begins Long Process of Digging Out, Rebuilding	YES	YES	0.80	0.44	0.57
16	01/18/95	Kobe Rescue Operations Produce Some Miracles	YES	YES	0.80	0.40	0.53
17	01/18/95	Kobe, Japan Quake Survivors Recovering With Stride	YES	YES	0.83	0.45	0.58
18	01/18/95	Traffic Jams Hamper Japan Quake Recovery Efforts	YES	YES	0.86	0.50	0.63
19	01/18/95	Worldwide Help Offered to Japan in Quake’s Aftermath	YES	YES	0.66	0.15	0.25
20	01/18/95	Seismologist Concerned at Lack of Quake Predictability	YES	NEW	-	-	-
21	01/18/95	Japanese Rescue Teams Intensify Search for Survivors	YES	YES	0.66	0.22	0.33
22	01/18/95	Japanese Firefighters Attempt to Control Quake Blazes	YES	YES	0.86	0.50	0.63
23	01/18/95	Japan quake death toll reaches 3,081	YES	NEW	-	-	-
24	01/18/95	Death Toll in Kobe Quake Exceeds 3,000	YES	YES	0.33	0.33	0.33
25	01/18/95	Kobe Earthquake Shares Pattern With California Quakes	YES	YES	0	0	0
26	01/18/95	Reporter Says Damage in Kobe, Japan, Not Uniform	YES	YES	0.88	0.40	0.67
27	01/19/95	Sights and Sounds of Kobe After the Earthquake	YES	YES	0.89	0.57	0.70
28	01/19/95	US forces to fly blankets to Japan quake survivors	NO	YES	-	-	-
29	01/19/95	Many People in Kobe, Japan, Had No Insurance	NEW	NEW	-	-	-
30	01/19/95	Japanese Prime Minister Visits Quake-Stricken Kobe	NEW	NEW	-	-	-
Average					0.79	0.58	0.54

以前の記事と同様の話題であると判定されたことを示し, ‘NEW’ はテスト記事が新しい話題であると判定されたことを示す. 本手法における ‘NO’ はテスト記事が続報記事でないと判定されたことを示す.

表 3 によると, 30 記事中, 24 のテスト記事を正しく判定している. テスト記事が続報記事にもかかわらず, これを正しく判定できなかった記事は 1 記事 (id = 28) のみであった. 一方, テスト記事が続報であることは判定できたものの, 新しい話題であるか否かについては, 5 記事のテスト記事が正確に判定されなかった. 本手法では, 現在の話題である記事  $d_p$  が新しい話題であるか否かを判定するために  $d_p$  と時間的に最も近い時期に報道された記事  $d_{p-1}$  を用いている. したがって, 不連続な場合, すなわち  $d_p$  がその直前である  $d_{p-1}$  とは異なる話題であり,  $d_{p-1}$  よりも前に報道された記事と同じ話題ある場合には, 結果的に  $d_p$  は新しい話題であると判定されてしまう. ‘id = 2, 4, および 8 は, いずれもこの原因により正確に判定できなかった. 今後は, 過去に報道された記事について不連続な場合も扱えるよう枠組みを拡張する必要がある.

次に本手法により得られた正例記事の精度, すなわち, 得られた正例記事が互いに同様の話題から成る記事であるか否かを検証する. 6 列目から 8 列目は, テスト記事がそれ以前に出現する記事と同様の話題であると正しく判定されたとき, すなわち, ‘system = YES’, かつ ‘actual = YES’ であるときに用いられた正例記事集合の精度を示し, それぞれ *recall*, *precision*, および F 値を示す. *Recall* は, 実際に同じ話題である記事の個数のうち, 本手法により同じ話題であると正しく判定された記事数の割合を示し, *precision* は, 本手法により同じ話題であると判定された記事数の中で, 実際に同じ話題である記事の個数の割合を示す.

表 3 よりテスト記事が訓練記事中の正例記事と同じ話題について述べていると正しく判定された場合 (‘system = actual = YES’) の micro-average F 値は 0.54 であった. 表 3 の各テスト記事における *precision* は, *recall* よりも低いことから, 同じ話題の記事の取りこぼしは少ないものの, 同じ話題でない記事も抽出する傾向がある. そこで, *precision* 値が最も高い ‘id = 27’ とゼロであった ‘id = 25’ に対して, 本手法により得られた記事集合を用いて学習した結果得られる分離超平面  $w \cdot x + b$  における  $w$  の各要素の値を求めることで, 抽出された正例記事集合を特徴づけている単語の検証を行う.  $w$  は式 (3) により計算される.  $w$  の各要素は各単語に対応しており, その値が

表 4 分離超平面  $w \cdot x + b$  におけるベクトル  $w$  を構成する単語 (id = 25, 27)

Table 4 Words in  $w$  of  $w \cdot x + b$  (id = 25, 27).

Event 15 (‘Kobe Japan Quake’)	
id	単語
25	Kobe, central, injure, Tuesday, early, city, Japan, national, earthquake, tear
id	単語
27	building, Kobe, rescue, house, collapse, search, work, lot, hard, crew

正の場合は正例に, 負の場合は負例を分類することに貢献しており, 絶対値が高いほど貢献度が高いといえる. 表 4 において ‘id = 25’, あるいは ‘id = 27’ と同じ話題であると判定された正例記事集合, および判定の際に使用した正例記事集合と同数からなる負例記事集合を用いて学習した結果得られる分離超平面における  $w$  の各要素となる単語の値を示す.

‘id’ は記事番号を示し, ‘語’ は, SVMs による学習の結果得られる分離超平面  $w$  の各要素において高い重みを持つ単語のうち上位 10 語を示す.

‘id = 25’ は地震調査者の調査結果に関する記事であり, 神戸で発生した地震はカリフォルニアで起きた地震と性質が似ていることを述べている. この記事は, 地震調査者による調査が始まったことを報告する ‘id = 20’ の記事と同じ話題である. 一方, ‘id = 27’ は, 地震が勃発してからの最近の被害状況, および救助活動について述べている記事であり, ‘id = 21 ~ 24’, および ‘26’ の記事と同様の話題である. 表 4 において, ‘id = 27’ では, 地震の被害や救助活動を示す ‘rescue’, ‘collapse’, ‘search’ などが出現している. 一方, ‘id = 25’ の場合には ‘California’ のほか, ‘San Francisco’ や ‘Los Angeles’ などの単語が文書中に出現していた. しかし, 単語の表層情報のみを使用して判定を行っている本手法ではこれらの単語を同一と見なすことができず, 結果的に ‘id = 20’ と同じ話題として正しく認識することができなかった. 今後は, コーパスに対しクラスタリング手法を用いて場所や地名を示す語の類似性を求めることで, 類似した話題に対する判定精度を向上させる必要がある.

#### 4.2.3 他の Data management 手法との比較

本手法は, テスト記事が続報記事であると判定されるごとに, その記事を訓練記事集合に加え, 再度, 最適な正例記事集合の推定を行っている. そこで, 本手法の有効性を検証するため,  $k$ -means の再帰的な利用による訓練記事集合の分類手法を用いたものと用いないものに対して, それぞれ以下で示す Data management 手法を適用した結果との比較を行った.

表 5 他の data management と本手法との精度比較  
Table 5 F-measure with several data management approaches.

high						
期間	event 番号	<i>k</i> -means による分類なし		<i>k</i> -means による分類あり		本手法
		fix	full	fix	full	
large	6	0.93	0.98	0.96	0.96	0.98
	8	0.32	0.37	0.30	0.37	0.68
	9	0.55	0.50	0.57	0.61	0.78
	24	0.41	0.54	0.70	0.76	0.82
	25	0.73	0.86	0.72	0.82	0.92
small	3	0.55	0.54	0.75	0.80	0.86
	11	0.50	0.75	0.88	0.89	0.89
	15	0.54	0.64	0.77	0.81	0.87
	16	0.59	0.81	0.55	0.58	0.73
	22	0.73	0.71	0.67	0.72	0.82
Average		0.58	0.67	0.69	0.73	0.84

low						
期間	event 番号	<i>k</i> -means による階層なし		<i>k</i> -means による階層あり		本手法
		fix	full	fix	full	
large	18	0.48	0.61	0.58	0.67	0.72
small	5	0.81	0.81	0.75	0.71	0.67
	12	0.29	0.29	0.29	0.29	0.29
	17	0.50	0.67	0.67	0.78	0.82
	21	0.39	0.55	0.60	0.77	0.78
Average		0.49	0.59	0.58	0.64	0.66
Total Average		0.55	0.64	0.65	0.70	0.78

- 1 ‘Fixed Memory’: 初期に与えられた訓練記事集合のみを用いて、すべてのテスト記事の判定を行う。
- 2 ‘Full Memory’: テスト記事が正例記事であると判定された時点で、それまでの負例記事と正例記事、および正例と判定されたテスト記事を用いて、新たなテスト記事の判定を行う。

実験結果を表 5 に示す。表 5 は  $N_t = 16$  の結果を示す。‘High’における event 番号は、異なる話題が多く含まれている出来事を示し、‘low’は同じ話題が多く含まれている出来事を示す。‘期間’における ‘large’は正例のテスト記事が 2 カ月以上の期間にわたって出現することを示し、‘small’は 2 カ月以内に出現することを示す。‘*k*-means による分類なし’、および ‘*k*-means による分類あり’はそれぞれ、*k*-means の再帰的な利用により訓練記事集合の分類を行った結果と行わなかった結果を示す。‘fix’、および ‘full’はそれぞれ fixed memory、および full memory の結果を示す。‘high’および ‘low’における ‘Average’は、それぞれに属する出来事の micro-average F 値を示し ‘Total average’はすべての出来事に対する micro-average F 値を示す。

表 5 によると、Total average は fix および full memory のいずれの場合にも ‘*k*-means による分類あ

り’の方が ‘*k*-means による分類なし’よりも高い精度が得られていることから、*k*-means を用いた訓練記事集合の分類は精度に貢献しているといえる。しかし、event 番号が 5、および 16 の場合には、fix および full memory のいずれの場合にも *k*-means を用いることで精度が低下している。原因の 1 つに、*k*-means の終了条件が考えられる。本手法では、正例記事集合が 2 つのクラスに分割されるまで、訓練記事集合に対して再帰的に *k*-means を適用している。したがって負例記事が意味的に正例記事と類似した内容である場合にも、正例記事集合が 2 つのクラスに分割されない限り、分割を繰り返すため、結果的に正例であるテスト記事を誤って負例であると判定する可能性がある。実際、event 番号が 5 および 16 の場合、*k*-means を用いた場合と用いなかった場合の精度差は、続報記事であるテスト記事を続報記事でないと誤って判定したことによるものであり、そのすべてが、bottom cluster の 1 つ上のクラスで負例から成るクラスに分類されていた。表 6 は event 番号 5 (Clinic Murders) において、bottom cluster の 1 つ上のクラスを構成する記事集合を用いて学習した結果得られる分離超平面  $w \cdot x + b$  における  $w$  の各要素の値を示す。

‘単語’で示される語は、SVMs による学習の結果得られる分離超平面  $w$  の各要素において正、および負

表 6  $w \cdot x + b$  において  $w$  を構成する単語 ( $k$ -means による分類あり)

Table 6 Words in  $w$  of  $w \cdot x + b$  (with  $k$ -means).

Event 5 ('Clinic Murders (Salvi)')	
正例負例	単語
正	police, Friday, State, United, fire, person, race, clinic, street, abortion
負	health, president, house, move, committee, senate, care, Clinton, characterize, constructive

でかつ絶対値が高い上位 10 語を示す。表 6 によると、'president'、'Clinton' など、'Clinic Murders' に関係する語 が負例の中に含まれている。今後は、 $k$ -means に教師つき学習を用いることで、クラスタリングの終了条件を改良する必要がある。

次に ' $k$ -means による分類あり' における fix memory、および full memory による精度を比較する。'low' の結果、すなわち、同じ話題について述べている記事が多い場合には、 $k$ -means を用いた場合、fixed memory よりも full memory による手法が優れている。これは、同じ話題を多く含む場合には、より多くの正例記事を訓練記事として用いた方がよいということを裏付ける結果となっている。'high' の場合、すなわち、異なる話題を多く含む出来事である場合、fix と full memory との精度差は 0.04 (0.73 - 0.69) であり 'low' 0.06 (0.64 - 0.58) よりも小さいものの、'low' と同様、full memory による結果の方が優れている。さらに、'low' と同様、'high' において最も高い精度が得られたのは '本手法' であり、両者における本手法と 'full' との精度差を比較すると、'low' では 0.02 (0.66 - 0.64) であるのに対し、'high' における精度差は、0.11 (0.84 - 0.73) であることから、本手法は、話題が推移している出来事に対してより有効であるといえる。

#### 4.2.4 関連研究との比較

本項では、関連研究との比較について述べる。関連研究として、カーネギーメロン大学、Dragon Systems、およびマサチューセッツ大学との比較結果を示す。カーネギーメロン大学の Carbonell らは、 $kNN$  手法および決定木学習の 2 種類の方法を用いて続報記事を判定する方法を提案している<sup>11)</sup>。訓練データ中の各正例記事と負例記事、およびテスト記事を単語を次元とするベクトルで表現し、テスト記事と訓練データとの各記事間の類似度を余弦尺度を用いて計算する。類似度の

表 7 関連研究との比較結果

Table 7 The result of comparative experiment.

Run	Miss	F/A	Prec	F
CMU $kNN$	29%	0.40%	61%	0.66
Dragon	71%	0.21%	60%	0.39
UMass nonRF-comb	55%	0.10%	88%	0.60
UMass nonRF-20T	13%	0.35%	27%	0.41
UMass RF100	39%	0.27%	62%	0.62
本手法	28%	0.14%	77%	0.74

値が最も高い  $k$  番目までの訓練記事を抽出し、正例記事の個数が負例記事数よりも多い場合にテスト記事は続報記事であると判定している。Dragon Systems の Yamron らは、Kullback-Leibler (KL) による類似度尺度に基づき訓練データをクラスタに分類し、テスト記事が訓練データ中のどのクラスタに属するかにより、続報記事であるか否かの判定を行っている<sup>8)</sup>。マサチューセッツ大学の Allan らは、適合性フィードバックを用いて続報記事を判定する手法を提案している<sup>4)</sup>。訓練データ中の正例記事と負例記事は検索要求文を作成するために用いられ、テスト記事が要求文を満たすときにその記事は続報記事であると判定している。実験では、以下で示す 3 種の手法を検証している。

- (1) 検索要求文を作成する際に適合性フィードバックを用いず、訓練データから名詞と名詞句を抽出して使用した場合 (nonRF-comb 法)
- (2) 適合性フィードバックを用いず、訓練データ中の単語の中で頻度数が上位 20 の単語を使用した場合 (nonRF-20T 法)
- (3) 適合性フィードバックを用い、訓練データ中の負例記事数をたかだか 100 記事使用した場合 (RF100 法)

各手法と本手法との比較結果を表 7 に示す。表 7 における関連研究の実験結果は Allan らにより報告されたものであり、CMU、Dragon Systems、および UMass はそれぞれカーネギーメロン大学 ( $kNN$  法)、Dragon Systems、およびマサチューセッツ大学の結果を示す<sup>4)</sup>。表の各値は、 $N_t = 4$  の結果を示し、総計 23 種類の出来事に対する micro-average の値を示す<sup>4)</sup>。'UMass nonRF-comb'、'UMass nonRF-20T' および 'UMass RF100' はそれぞれ nonRF-comb 法、nonTF-20T 法、および RF100 法の結果を示す。

Miss 率、および F/A を比較すると、CMU  $kNN$ 、Dragon、および UMass RF100 の 3 手法はいずれも本手法よりも高い Miss 率と F/A 値となっている。一方、UMass nonRF-20T は本手法よりも Miss 率は低いものの F/A すなわち、続報記事でない記事に対して、続報記事であると誤って判定した割合が高い。UMass

大統領が、Clinic Murders に対して対応策を講じるよう命じた。

表 8 正例記事数と F 値の変化  
Table 8  $N_t$  values and F-measure.

	$N_t$ value				
	1	2	4	8	16
Dragon	-55%	-26%	-	+12%	+40%
CMU, $kNN$	-25%	-9%	-	+11%	+32%
UMass	-39%	-5%	-	+5%	+5%
本手法	-10%	-1%	-	+2%	+4%

nonRF-comb は F/A は本手法よりも低いものの Miss 率、すなわち続報記事を正しく続報記事であると判定できない割合が高い。以上のことから、本手法は、他手法と比較して安定した結果が得られているといえる。

表 8 は  $N_t$  で示される正例記事数と精度との関係を示す。表 8 において各数値は  $N_t = 4$  のときの精度をベースラインとしたときの増加（減少）率を示す。‘Dragon’、‘CMU’、および ‘UMass’ の結果は Allanらの報告によるものであり<sup>4)</sup>、それぞれ Dragon Systems、カーネギーメロン大学、およびマサチューセッツ大学の nonRF-comb 手法を示す。

表 8 より、本手法とマサチューセッツ大学の nonRF-comb 手法の結果はともに  $N_t = 2$  以上のときは精度に大きな変化が見られないことから、正例記事数が変化した場合にも安定した結果が得られているという点で類似しているといえる。

## 5. おわりに

本稿では、日付順に並べられた記事に対し話題の推移を考慮した続報記事の抽出手法を提案した。TDT1 コーパスを用いて実験した結果、訓練記事中の正例記事数が 16 記事の場合に、F 値 0.78 の精度で続報記事を判定でき、また、既存の手法と比較した結果、正例記事数が少ない場合にも安定した結果が得られた。さらに、実験の結果から以下の点が明らかになった。

- (1) fix および full いずれの Data management 手法においても  $k$ -means を用いて訓練記事を分類した結果を用いた方が用いない場合よりも高い精度が得られたことから、続報記事抽出タスクで与えられた負例記事の利用は続報記事抽出に貢献しているといえる。
- (2) 異なる話題が多く含まれている出来事 (high)、および同じ話題が多く含まれている出来事 (low) のいずれの場合にも、full の結果が fix の結果を上回っていることから、続報記事の抽出では話題の推移にかかわらず、正例と判断された記事を随時追加し、新たなテスト記事判定に用いた方が良い精度が得られる。

- (3) ‘low’ および ‘high’ における本手法と ‘full’ との精度差を比較すると、‘low’ では 0.02 であるのに対し、‘high’ における精度差は、0.11 であることから、本手法は、話題が推移している出来事に対してより有効であるといえる。

今後は、 $k$ -means における代表点の決定方法、および終了条件について改良を行うと同時に、過去に報道された記事について、話題が不連続に出現する場合にも扱えるよう枠組みを拡張する予定である。

謝辞 本研究は、文部省科学研究費補助金、人工知能研究振興財団、および柏森情報科学振興財団の援助を受けています。ここにそれらを記し謝意に代えさせていただきます。

## 参 考 文 献

- 1) Yang, Y., Carbonell, J.G., Allan, J. and Yamron, J.: Topic Detection and Tracking: Detection-Task, *Proc. Workshop of Topic Detection and Tracking* (1997).
- 2) Strzalkowski, T., Stein, G.C. and Wise, G.B.: GE. Tracker: A Robust, Lightweight Topic Tracking System, *Proc. DARPA Broadcast News Transcription and Understanding Workshop* (1999).
- 3) Swan, R. and Allan, J.: Automatic Generation of Overview Timelines, *Proc. 23rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp.49–56 (2000).
- 4) Allan, J., Carbonell, J., Doddington, G., Yamron, J. and Yang, Y.: Topic Detection and Tracking Pilot Study: Final Report, *Proc. DARPA Broadcast News Transcription and Understanding Workshop* (1998).
- 5) Allan, J.: Incremental Relevance Feedback for Information Filtering, *Proc. 19th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp.270–278 (1996).
- 6) Papka, R., Allan, J. and Lavrenko, V.: UMASS Approaches to Detection and Tracking at TDT2, *Proc. DARPA Broadcast News Transcription and Understanding Workshop* (1999).
- 7) Schultz, J.M. and Liberman, M.: Topic Detection and Ttracking using idf-Weighted Cosine Coefficient, *Proc. DARPA Broadcast News Transcription and Understanding Workshop* (1999).
- 8) Yamron, J.P., Carp, I., Gillick, L., Lowe, S. and Mulbregt, P.: Topic Tracking in a News Stream, *Proc. DARPA Broadcast News Tran-*

- scription and Understanding Workshop* (1999).
- 9) Lowe, S.A.: The Beta-Binomial Mixture Model and its Application to TDT Tracking and Detection, *Proc. DARPA Broadcast News Transcription and Understanding Workshop* (1999).
  - 10) Allan, J., Papka, R. and Lavrenko, V.: On-Line New Event Detection and Tracking, *Proc. 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp.37-45 (1998).
  - 11) Carbonell, J., Yang, Y., Lafferty, J., Brown, R.D., Pierce, T. and Liu, X.: CMU Report on TDT-2: Segmentation, Detection and Tracking, *Proc. DARPA Broadcast News Transcription and Understanding Workshop* (1999).
  - 12) Yang, Y., Ault, T., Pierce, T. and Lattimer, C.W.: Improving Text Categorization Methods for Event Tracking, *Proc. 23rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp.65-72 (2000).
  - 13) Lunts, A. and Brailovskiy, V.: Evaluation of Attributes Obtained in Statistical Decision Rules, *Engineering Cybernetics*, Vol.3, pp.98-109 (1967).
  - 14) Vapnik, V.: *The Nature of Statistical Learning Theory*, Springer (1995).
  - 15) Burges, C.J.C.: *A Tutorial on Support Vector Machines for Pattern Recognition*, Kluwer Academic Publisher (1998).
  - 16) Duda, R.O. and Hart, P.E.: *Pattern Classification and Scene Analysis*, John Wiley & Sons (1973).
  - 17) Schmid, H.: Improvements in Part-of-Speech Tagging with an Application to German, *Proc. EAACL SIGDAT Workshop* (1995).

(平成 14 年 5 月 8 日受付)

(平成 15 年 5 月 6 日採録)



福本 文代 (正会員)

1986 年学習院大学理学部数学科卒業。同年沖電気工業(株)入社。総合システム研究所勤務。1988 年より 1992 年まで(財)新世代コンピュータ技術開発機構へ出向。1993 年マンチェスター工科大学計算言語学部修士課程修了。同大学客員研究員を経て 1994 年より山梨大学工学部助手, 1999 年同学部助教授, 現在に至る。自然言語処理の研究に従事。理学博士。ACM, ACL, 言語処理学会各会員。



鈴木 良弥 (正会員)

1986 年山梨大学工学部計算機科学科卒業。1988 年山梨大学大学院工学研究科計算機科学専攻修了。同年木更津工業高等専門学校助手。1993 年東京工業大学大学院総合理工学研究科博士後期課程修了。1994 年より山梨大学工学部助手, 現在に至る。音声言語処理の研究に従事。工学博士。ACL, 電子情報通信学会, 日本音響学会, 言語処理学会各会員。



山田 寛康

1997 年山梨大学工学部電子情報工学科卒業。1999 年山梨大学大学院工学研究科博士前期課程修了。2003 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。同年北陸先端科学技術大学院大学助手。自然言語処理の研究に従事。工学博士。