

単独記事フィルタリングを用いた 時系列ニュース記事分類法の提案 Clustering Temporal Internet News Articles with Filtering Single Articles

中村 智浩* 平野 孝佳*
平手 勇宇† 山名 早人‡

Tomohiro NAKAMURA* Takayoshi HIRANO*
Yu HIRATE† Hayato YAMANA‡

インターネット上のニュース記事を対象としたクラスタリングは、関連記事や最新話題語などの検出を目的として、アメリカの TDT プロジェクトをはじめ、広く研究されている。ニュース記事において、他記事との関連性を持たない単独記事は頻出するが、従来のクラスタリング手法では単独のクラスタとして検出することが困難である。そこで本稿では、事前に固有名詞情報や地理的情報などを用いて単独記事を検出し、検出した単独記事をニュース記事集合から除いたうえでクラスタリングする手法を提案する。評価では、半年分の日本のニュース 2,136 記事を用いた。最短距離法でクラスタリングする場合、本手法を適用することにより、精度を 10.2%改善できると共に計算時間を約 1/3 に抑えることができた。

Clustering of internet news articles makes it possible to detect various useful information, for example, related articles, and latest topic words. From the TDT project down, this area is widely researched. Conventional clustering methods have difficulties to detect single article as a single cluster even though many single articles exist. In this paper, we propose a method to cluster news articles from which single articles are excluded in advance by using proper noun information, topographic information and other characteristics between single and non-single articles. In evaluation, we use half a year Japanese news articles. Compared to the Single-Link Method, which alone is difficult to judge articles single, our proposing method improves precision 10.2% and reduces the computation time to approximately a third.

1. はじめに

近年、社会の情報化に伴い電子化された文書が増大している。この膨大な数の文章を手により処理することは大変時間のかかる作業である。このような背景から、自動的に内容

* 学生会員 早稲田大学大学院基幹理工学研究科

{naka.hirano}@yama.info.waseda.ac.jp

† 正会員 早稲田大学メディアネットワークセンター

hirate@yama.info.waseda.ac.jp

‡ 正会員 早稲田大学理工学術院, 国立情報学研究所

yamana@waseda.jp

ごとに文書を分類する文書クラスタリングの手法が多く研究されている。この中でも、時系列のニュース記事群に対する研究としては、アメリカでの TDT (Topic Detection and Tracking) をはじめ数多く研究されている [1][2]。TDT は、あるイベント (できごと, 事件など) についての情報を、ニュースの文書などから成るコーパスから自動生成する手法である。TDT は、主にトピック検出 (Topic Detection) とトピック追跡 (Topic Tracking) から成り、Topic Detection の部分でクラスタリングを行っている。

Hatzivassiloglou ら [3] は TDT プロジェクトが提供するコーパスを用いて、複数のクラスタリング手法を用いた評価実験を行っている。用いたクラスタリング手法の中で評価値が最も高かった手法は群平均法 (GroupWise-Average Method) である。しかし、コーパス内には単独の記事も多いため、こうした単独記事を事前にクラスタ対象外にすることでさらに精度を上げることができると考えられる。

さらに、文書クラスタリングの精度以外の問題として、計算量の問題がある。クラスタリング手法は大きく階層型のものとは非階層型のものに分けられるが、階層型の場合計算量は少なくとも文書数 N に対し $O(N^2)$ となる。k-means 法に代表される非階層型のクラスタリングの場合、事前にクラスタ数を決定しなければならないという欠点がある。非階層型の手法を、ニュース記事に適用することを考えると、事前にクラスタ数を決定することが難しいという問題がある。

また、次々に追加されるニュース記事群に対しクラスタリングを行うため、インクリメンタルなクラスタリングを行う研究もある [1][2][4]。しかし、その場合首相、大統領の人名や、会議などのイベント名称など主題となりえる重要な単語が記事の先頭付近に 1 度だけ出現するような場合も多く、単語の出現頻度 TF と、単語の重要性との相関が必ずしも等しくないニュース記事の文書ベクトルに、IDF 値が使いつらい [5]。これを解決するには、インクリメンタルなクラスタリングを行う前に、初期集合に対して階層型クラスタリングを行った上で、インクリメンタルなクラスタリングを行うなどの方法があり、ニュース記事群に対して階層型クラスタリング手法が部分的に有効と考えられる。実際に TDT に参加する一部の研究者は、既に存するニュース記事集合に対する Topic Detection (Retrospective Detection) と、リアルタイムに入ってくるニュース記事を分類する Topic Detection (On-line Detection) において用いるアルゴリズムを変えている [1]。

以上述べたように、ニュース記事のクラスタリング手法では精度の問題がある。また、インクリメンタルなクラスタリングを含めた非階層型クラスタリングに対して階層型クラスタリングは精度の向上を臨みやすいが、計算量は少なくとも文書数 N に対し $O(N^2)$ となる。本稿では、精度と計算量が両立しない問題に対応するため、事前に単独の記事をフィルタリングし、精度の向上と計算時間の減少を試みる。

本稿では、まず事前実験において集めたニュース記事群中において続報・派生記事の出ない記事 (以下単独記事) が多いことを確認する。例えばある窃盗事件に関する記事が出て、その後一度もその事件に関する記事が出なければその報道は単独記事となる。

次に、単独記事のフィルタリングを用いるクラスタリング手法を提案する。提案手法では、単独記事をフィルタリングするため、当該記事中に出現する固有名詞が予め設定した閾値以上であるものを単独記事の候補群とする。次に、記事中の地理的情報や、非単独記事固有の「～～問題で」「～～事

件で」という特徴的な表現を用いて、単独記事候補群から非単独記事の可能性のあるものを除外する。以上の手順で作成された記事群に対して既存の階層型クラスタリングを適用する。これによって、クラスタリングの対象となる記事数を減らすことができるため、計算時間の削減も可能となる。

以下、2節では、本手法に関する関連研究について述べる。3節では、提案手法について述べる。4節では、評価実験を行い考察する。最後に5節でまとめを述べる。

2. 関連研究

ニュース記事群を対象としたクラスタリング手法に関する研究は、これまでに数多く行われている。

アメリカの NIST が主催する TDT (Topic Detection and Tracking) [1] [2] は、何らかのイベントについての情報を、ニュースの文書などから構成されるコーパスから自動識別する試みである。

Hatzivassiloglou らは、TDT プロジェクトの提供するニュース記事コーパスを用いて、複数のクラスタリング手法を評価している。評価の結果、群平均法が精度の面から最も優れていることを示した [3]。しかし、Hatzivassiloglou らは群平均法のようなマルチパス・アルゴリズムではなく、単一パス・アルゴリズムを計算量の面で考慮すべきだとしている。この計算量の利点と、階層型クラスタリング手法の精度の利点を両立するため、TDT に参加する一部の研究者は、既に存在するニュース記事集合に対する Topic Detection

(Retrospective Detection) と、リアルタイムに入ってくるニュース記事を分類する Topic Detection (On-line Detection) において用いるアルゴリズムを変えている [1]。前者では、精度を重視するため、階層型クラスタリング手法を用いる場合が多いのに対し、後者については、計算量を減らすため、単一パス法のようなインクリメンタルなクラスタリングを用いている場合が多い。

コロンビア大学の McKeown らは、[3] の結果を受けてニュース記事群を群平均法により階層的にクラスタリングを行い、各クラスタに含まれている記事群の要約や、関連クラスタなどを表示するサイトを開発した [6]。なお、文書ベクトルは TFIDF をベースとしている。このサイト Columbia Newsblaster [7] は、要約記事の他、イベントを簡潔に要約する単語集合も表示する。すなわち、一つのイベントに対する2種類の要約を提供している。特徴的な点として、クラスタがどのような記事で構成されているのか (ある事件に関する記事とその続報記事集合か、ある人物に関する記事かなど) によってそれぞれ異なる要約器を用いて要約を作成することにより精度を高めている。

佐藤らは 2005 年に、新鮮で可能な限り多くの話題情報を、最新話題語と呼ぶキーワードの形態で抽出する手法を提案した [5]。

また、上嶋らは 2004 年に、タイムスタンプ (発行時間) を与えられていない文書集合にタイムスタンプを割り当てる (順序づけ) 手法を提案した [4]。

これらの研究では [1], [2] の一部と [4] を除き全て、階層型クラスタリング手法を用い精度向上に力点が置かれている。しかし、計算量の問題や、ニュース記事はリアルタイムに追加されていく問題を指摘し、インクリメンタルなクラスタリングを含めた、非階層型のクラスタリング手法を適用していることもある [1][2][4]。

本提案は、階層型クラスタリングの前処理として単独記事

のフィルタリングを行うことにより、精度の向上と計算時間の削減を目的としている。

3. 単独記事フィルタリングの提案

本節では、ニュース記事群に対しクラスタリングアルゴリズムを適用する前に、続報・派生記事の出ていない記事、つまりクラスタがその記事だけで構成される記事をフィルタリングする手法を提案する。以下、このような記事を単独記事と呼ぶ。なお、以下の全ての実験において、形態素解析器には Yahoo! 日本語形態素解析 Web サービス [8] を、群平均法のクラスタリングにはライブラリ SlothLib [9] を用いた。

3.1 事前実験

事前に集めたニュース記事群に対して、クラスタリングの正解セット作成を手で行ったところ、ニュース記事群には続報・派生記事が出ることのない記事が多いことが分かった。そのため、これらの記事を事前にフィルタリングすることにより、クラスタリングの改善が期待できるのではないかと考え、事前実験を行った。

事前実験では、コーパスとして YOMIURI ONLINE [10] の「社会」ジャンルの記事 (<http://www.yomiuri.co.jp/national/> 以下の記事) を用いた。期間は 2007/7/2~2007/9/9、記事数は 804 記事である。

実験では、群平均法、最短距離法、及び人による分類 (正解セット) の比較を行った。なお、閾値としては、群平均法、最短距離法、共にベクトルのコサイン類似度として 0.15 を用い、記事ベクトルとしては、単語の出現頻度の逆数 (IDF) を用いた。比較結果を表 1 に示す。

表 1 群平均法・最短距離法の単独判定の精度
Table 1 Precision of Single/Non-Single Judge by using GroupWise-Average Method and Single-Link Method

	正解	群平均法	最短距離法
単独記事	512	420 (単独と判定)	397 (単独と判定)
		92 (非単独と判定)	115 (非単独と判定)
非単独記事	292	77 (単独と判定)	48 (単独と判定)
		215 (非単独と判定)	244 (非単独と判定)
合計		804	

表より単独記事が約 64% と多いことが分かる。また、[3] で最も精度のよかった群平均法でも、単独記事を非単独記事と判定したり、非単独記事を単独記事と判定したりしており、改善の余地があることが分かる。

以上から、単独記事をクラスタリング前にフィルタリングすることで精度の向上が得られると期待される。また、単独記事を除外することにより、記事数を減らすことができるため、クラスタリングに必要な計算時間の短縮も可能となる。

3.2 概要

クラスタリングの前に単独記事のフィルタリングを行う手法を提案する。提案手法は大きく分けて二つの手順に分かれる。概要を図 1 に示す。

3.3 単独記事候補群の作成

単独記事の候補群を作成するにあたり、「当該記事だけに含まれる固有名詞の数」(以下 N_{unique}) を利用する。これは、3.1 で用いたコーパスにおいて、単独記事と判断されたものと非単独記事で判断されたもので、当該記事だけに含まれる固有名詞数が大きく異なったためである。具体的には、単独

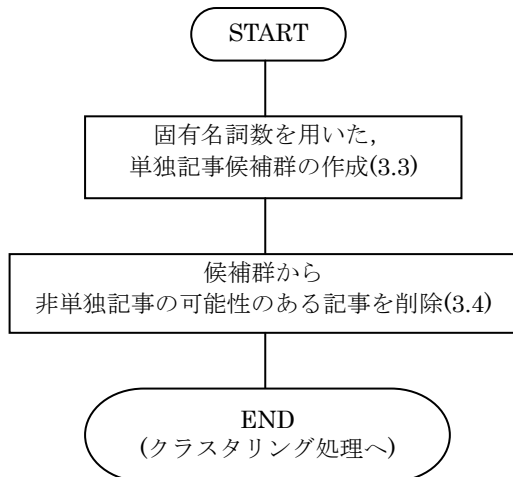


図1 提案手法概要フローチャート
Fig. 1 Overview of the Proposed System

記事の平均 N_{unique} が約 4.03 に対し、非単独記事の平均 N_{unique} は約 1.96 であった。

まず、 $N_{unique} \geq \theta$ の記事集合を単独記事候補群とする。4 節では、 $\theta=3, \theta=4$ について評価実験を行うが、 N_{unique} が多いことと、単独記事らしさの相関を証明するため、「 $N_{unique} \geq \theta$ の記事であることと、記事の単独 or 非単独については何の関係もない」という仮説を $\theta=3, \theta=4$ についてそれぞれ χ^2 検定を用いて棄却した ($p < 0.01$)。よって、 $N_{unique} \geq 3$ であること、 $N_{unique} \geq 4$ であることと、それぞれその記事が単独記事らしいということの関係は、有意基準 1% において有意である。これより、本稿では N_{unique} の情報を単独記事判定に用いる。また、ここで固有有名詞とは、[8]で「外国人姓、外国人名、固名、人姓、人名、人名その他、組織、組織企業、地名、地名行政区分、地名町名」に分類されるものである。

3.4 候補の削除

単独記事候補群から非単独記事の可能性のある記事を以下の 2 ステップにより削除する。そして、削除後に残った単独記事候補群をクラスタリングの対象とする記事群から削除する。

3.4.1 地理的情報に基づく削除

ニュース記事には、イベントがあった場所が記載されることが多い。そしてまた、同じクラスタに入れるべき記事は、同一の場所が記事中に記載されることが多い。このことから、同じ地理的情報が書かれた記事が対象記事の前後 D_{sp} 日以内に存在した場合、それら両方を単独記事候補群から削除する。

ここで具体的に「地理的情報」とは、[8][10]で「地名行政区分」に分類されるものから、ニュース記事の時系列上局所的にではなく、全期間に渡り常に D_u 日以内に現れる単語と都道府県名を除いたものとした。ここで都道府県名を除くのは、都道府県名が一致しているだけでは、同一のイベントである可能性が高いとは判断できないためである。また、大半の地理的情報が載る記事には、都道府県名以下の地名まで記載されており、都道府県名を除いても問題ないと考えた。

3.4.2 特徴的表現に基づく削除

非単独記事に特有の表現を用いている記事を候補群から削除する。日本語ニュース記事群において、よく用いられる表現に、「～～事件で」「～～問題で」という表現がある。例として、

- ・ 『○○○の土地・建物を舞台にした詐欺事件で』
- ・ 『ファイル交換ソフト「ウィニー」が入ったパソコンから、約 1 万件に上る○○○などがインターネット上に流出した問題で』
- ・ 『私立高校の大学合格実績水増し問題で』
- ・ 『○○○の力士が急死した問題で』
- ・ 『福岡県○○○市で○○○年○月に起きた○○○殺人事件で』

などが挙げられる。

これらの記事は、過去に発生したイベントを参照しているため、非単独記事である可能性が高い。3.1 で用いたコーパスにおいて、「～～事件で」「～～問題で」を含む記事は 115 記事 (14.3%) あり、この中で、単独記事は 37 記事であった。しかし、37 記事中、コーパス取得以前の記事の続報と思われる記事が内 34 記事存在した。つまり、コーパスが長期間になればなるほど、これらの表現を用いている記事の中で単独記事の割合は減ることになる。

以上から、これらの表現を用いている記事を単独記事候補群から削除する。しばしば、一番下の例のように、コーパス外の期間の事件の続報記事が現れるが、コーパス内に当該事件の続報記事が 2 つ以上ないという保証はないことから、文中の時間情報について考慮しないものとする。

3.5 計算時間

提案手法では、従来のクラスタリング手法に比較して、単独記事候補群を作成し、クラスタリング対象となるコーパスからフィルタリングするための計算が必要となる。しかし、単独記事候補群を作成するために必要な処理は、IDF を用いた文書ベクトルを作成する際に同時に行うことができる。また、単独記事候補群作成とその削除にかかる時間は階層型クラスタリングアルゴリズムの実行時間に比べ微少である。

したがって、提案手法の計算時間は、フィルタリング時間を無視した場合、フィルタリングを適用しない場合に比較し、

$$\frac{(\text{全記事数} - \text{単独記事判定された記事数})^2}{(\text{全記事数})^2}$$

となる。つまり、単独記事判定された記事数が多ければ多いほど、計算時間を削減できることがわかる。最後に提案手法の流れを図 2、図 3 に示す。

4. 評価実験

本節では、提案手法を実際のコーパスに適用した後、従来手法との比較を行う。本章で用いるコーパスは、3.1 の事前実験と同じく YOMIURI ONLINE[10]のニュース記事のうち、「社会」ジャンルの記事 (<http://www.yomiuri.co.jp/national/>以下の記事)である。事前実験よりも範囲を拡大し、期間は 2007/7/2～2008/1/2、記事数は 2,136 記事を用いた。記事は時系列に並んでいる。

4.1 評価尺度

本評価では、単に既存の階層型クラスタリング手法を用いた場合と比べ、単独記事を単独記事と判定できるかどうか、という点を評価する。

コーパスから無作為に 500 記事を選び、まず手動でそれらの記事がコーパス内全体で単独か非単独の判定を行った。単独記事と判断された記事は 320 記事であった。これらの記事のうち、階層型クラスタリングのみの場合と、提案手法で単

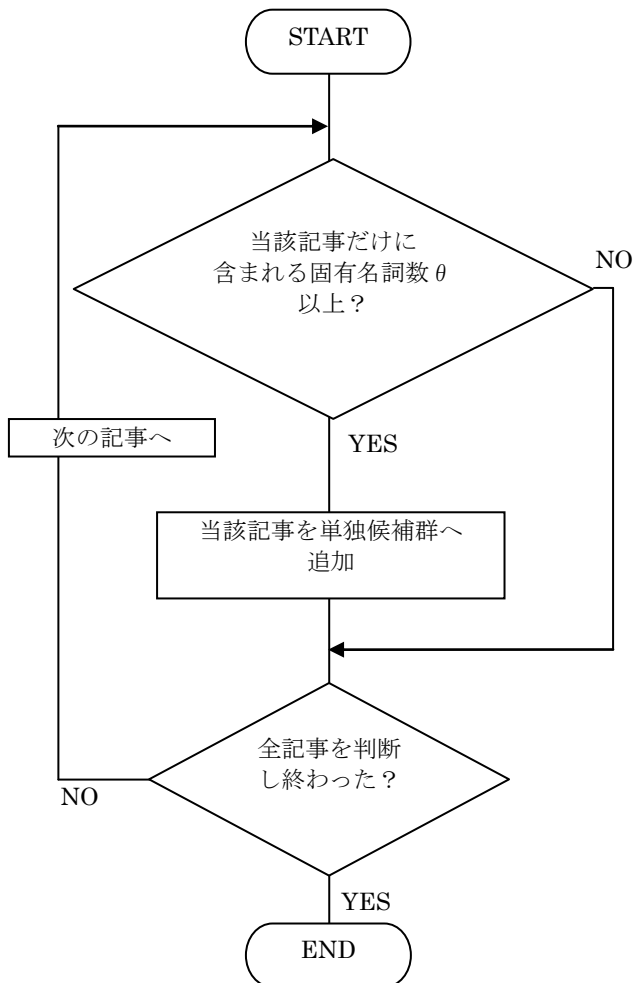


図2 単独候補群作成段階(3.3)フローチャート
Fig.2 Overview of Step 3.3: Lining up the Single Article Candidates

独記事のフィルタリングを行った後に階層型クラスタリングを適用した場合で、正しく単独記事と判断できた記事数を比較し精度で評価する(P_s)。また、手動で非単独記事判定された180記事についても、正しく非単独記事と判断できているか同様に比較し精度で評価する(P_{ns})。

また、2つをまとめた総合評価を P_{all} とおく。表と式で表すと表2のようになる。

さらに、3.5で述べた計算時間の利点を確かめるため、実際に単一記事を除外する時間と、階層型クラスタリングを実行する時間を合わせた計算時間を計測しその比較を示す。なお、ここで用いた計算機はCPUがPentium Core2 Duo 1.66GHz、メモリ2GBのものである。

4.2 実験結果と考察

既存のクラスタリング手法として、最も良い評価値を出した群平均法と、クラスタ間が結合しやすいため単体では単独記事を判定しづらい最短距離法の2つを用いる。共に類似度(ベクトルのコサイン類似度)の閾値は0.15とし、地理的情報の閾値 $D_{sp} = 4$ 、 $D_u = 25$ とした。また、 θ については、3.3に示した通り、単独記事らしさが十分である $\theta = 3, \theta = 4$ について実験を行った。実験結果をまとめたものを表3に示す。表に示すように、計算時間は大幅に短縮することができ、この点で提案手法の有効性が確かめられた。

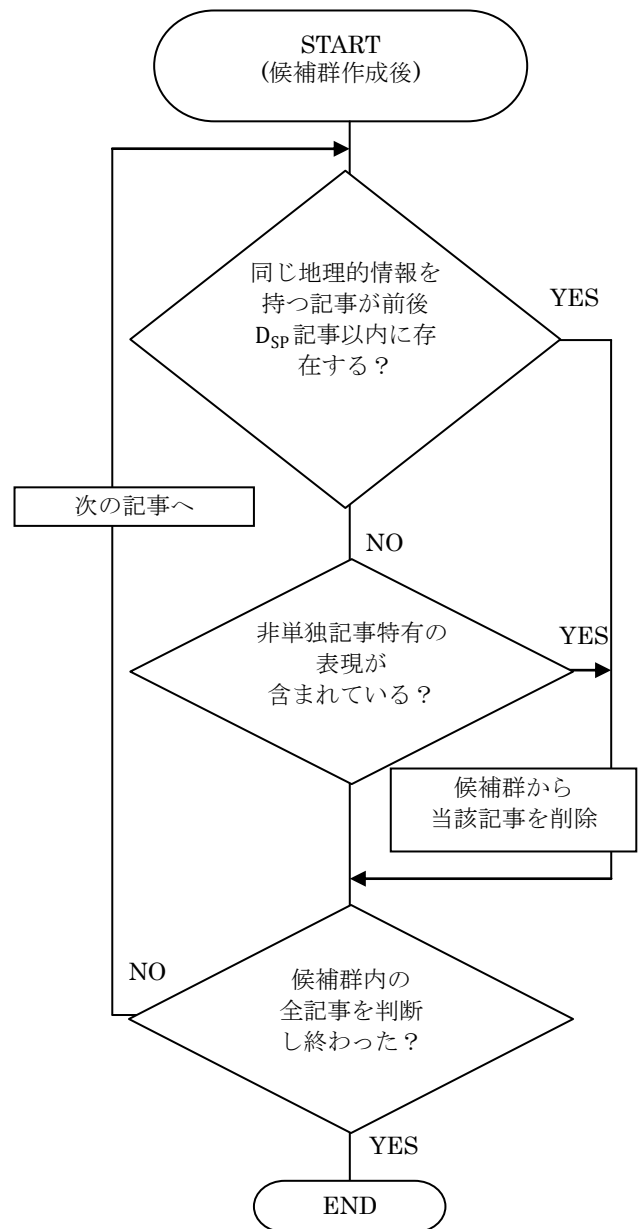


図3 候補削除段階(3.4)フローチャート
Fig.3 Overview of Step 3.4: Refining the Single Article Candidates

表2 評価尺度の説明
Table2 Explanation of Evaluation Measure

システムの出力	手動判定	
	単独	非単独
単独	S_+	N_-
非単独	S_-	N_+
全体	S	$N-S$

$$P_s = \frac{S_+}{S} \quad P_{ns} = \frac{N_+}{N-S}$$

$$P_{all} = P_s \cdot \frac{S}{N} + P_{ns} \cdot \frac{N-S}{N} = \frac{S_+ + N_+}{N}$$

次に、群平均法、最短距離法両方において精度の向上が見られる。特に単独判定のしづらい最短距離法においては、 P_s が約25.1%、 P_{all} が約10.2%上昇した。

表3 実験結果まとめ
Table3 Experimental Result

	P_s	P_{ns}	P_{all}	計算時間
群平均法	0.744	0.894	0.798	1.00
提案手法+ 群平均法 ($\theta=4$)	0.819	0.872	0.838	0.563
提案手法+ 群平均法 ($\theta=3$)	0.853	0.828	0.844	0.313
最短距離法	0.650	0.961	0.762	1.00
提案手法+ 最短距離法 ($\theta=4$)	0.741	0.933	0.810	0.498
提案手法+ 最短距離法 ($\theta=3$)	0.813	0.889	0.840	0.335
提案手法で単独判定された記事数 $\theta=4$441 記事 $\theta=3$678 記事 P_s :単独記事判定の精度 P_{ns} :非単独記事判定の精度 P_{all} :全体の精度 計算時間:階層型手法のみに対する割合				

提案手法でフィルタリングできない単独記事は、「当該記事だけに含まれる固有名詞の数」 $N_{unique} < \theta$ の記事か、候補の削除部分で単独記事にもかかわらず非単独記事の可能性があると削除してしまった記事である。

P_{ns} の減少については、単独記事候補群から非単独記事が完全に削除できていないために起こる。提案手法を用い P_{all} が階層型クラスタリング単体にくらべて減少していることはない、全体としては精度が上がっているが、人名情報、時系列情報などの考慮により P_{ns} の減少をさらに抑えられることが期待できる。

考察を深めるため、既存の階層型クラスタリングアルゴリズムでは単独判定できなかったが、提案手法を加えることで単独判定が可能になった記事についてその実例と N_{unique} 、その内訳を表4に示す。

表4 提案手法で単独判定が可能になった記事例
Table4 Examples of Articles which can be newly Judged Single by the proposed System

記事例	N_{unique}	内訳
有名高校元野球部監督の訃報記事	8	教え子の名前、長男の名前、自宅住所など
千葉でのナイフ刺殺事件の記事	4	自宅住所、容疑者名など
女性が電車にはねられ死亡した記事	1	現場住所

以上から、次の点を確認することができた。

- ・よく起こりうる内容の記事が多い
- ・人名情報、地理的情報が多くを占める
- ・ $N_{unique} < \theta$ であっても、単独判定が可能になることがある
 訃報、殺人事件などは、常に頻出する記事であり、既存のアルゴリズムでは記事ベクトルが近くなりがちで単独判定がしづらい。これらの記事が新しく単独判定できたのは好ましい結果と言える。参考として、184日間の記事(このうち、取得をはじめたばかりの数日間や、取得漏れなど記事数が少ない日も存在する)のうち、訃報記事は135日分の、殺人事件記事は85日分の記事に含まれた。またそれぞれ、提案手法で新しく単独判定できた記事の28.6%、17.9%を占めた。これらの記事に含まれる、「その記事だけに含まれる固有名詞」は人名情報、地理的情報が多くを占め、これらの情報の重要性も確認できた。

最後に、フィルタリングによるクラスタ間距離の変動の影響で、 $N_{unique} < \theta$ であっても単独判定が可能になることも明らかになった。

5. おわりに

本稿では、既存のクラスタリングアルゴリズムには、続報・派生記事の出ない単独記事が複数記事のクラスタに取り込まれてしまう欠点があることから、固有名詞情報を用いて新聞記事コーパスから単独記事のフィルタリングを行う手法を提案した。

実験として単純に階層型クラスタリングを用いた場合と、フィルタリングを行った後に階層型クラスタリングを用いた場合で比較を行い、その有効性を確認した。

今後の課題として、固有名詞数以外の情報の利用による単独記事候補群の網羅性向上、人名情報や時系列情報など、地理的情報・特徴表現以外の利用による候補群からの非単独記事削除の強化、コーパスの量や範囲を拡大した時の本手法の結果の確認と改善などが挙げられる。また、本提案はあくまで階層型クラスタリングの前処理である。そのため評価指標として最終的には、クラスタリング自体の精度とする必要があるため、この評価実験も必要である。

【文献】

- [1] J. Allan, J.G. Carbonell, G. Doddington, J. Yarnron and Y. Yang, "Topic Detection and Tracking Pilot Study Final Report," Proc. of DARPA Broadcast News Transcription and Understanding Workshop, pp.194-218, 1998.
- [2] J. Fiscus, G. Doddington, J. Garofolo and A. Martin, "NIST's 1998 Topic Detection and Tracking evaluation(TDT2)," Proc. of the 1999 DARPA Broadcast News Workshop, Herndon, Virginia, pp.19-24, 1999.
- [3] V. Hatzivassilogou, L. Gravano and A. Maganti, "An Investigation of Linguistic Features and Clustering Algorithms for Topic Document Clustering," Proc. of the 23rd Ann. Int'l ACM SIGIR Conf. on Research and Development in Information Retrieval, pp.224-231, 2000.
- [4] 上嶋宏, 三浦孝夫, 塩谷勇, "時系列ニュース記事集合に基づくニュース記事の順序付け," DEWS2004, 1-B-04, 2004.

- [5] 佐藤吉秀, 川島晴美等, “時系列ニュース記事における最新話題語抽出方法,” 自然言語処理, Vol.2005 No.73, pp.1-6, 2005.
- [6] K. McKeown, R. Barzilay, D. Evans, V. Hatzivassilogou, J. L. Klavans, A. Nenkova, C. Sable, B. Schiffman, S. Sigelman, “Tracking and Summarizing News on a Daily Basis with Columbia’s Newsblaster,” Proc. of Human Language Technology Conference 2003, San Diego, CA, pp.280-285, 2002.
- [7] Columbia Newsblaster,
<http://newsblaster.cs.columbia.edu/>
- [8] 日本語形態素解析 Web サービス,
<http://developer.yahoo.co.jp/jlp/MAService/V1/parse.html>
- [9] SlothLib, <http://www.dl.kuis.kyoto-u.ac.jp/slothlib/>
- [10] YOMIURI ONLINE,
<http://www.yomiuri.co.jp/>

中村 智浩 Tomohiro NAKAMURA

早稲田大学大学院基幹理工学研究科修士課程在学中. 日本データベース学会学生会員.

平野 孝佳 Takayoshi HIRANO

早稲田大学大学院基幹理工学研究科修士課程在学中. 情報処理学会, 日本データベース学会学生会員.

平手 勇宇 Yu HIRATE

2008 早稲田大学大学院理工学研究科博士後期課程修了. 博士 (工学). 2006 年より同大学メディアネットワークセンター助手. ACM, 情報処理学会, 日本データベース学会各会員.

山名 早人 Hayato YAMANA

1993 早稲田大学大学院理工学研究科博士後期課程修了. 博士 (工学). 1989-1993 同大学情報科学研究教育センター助手. 1993-2000 電子技術総合研究所. 2000 早稲田大学理工学部助教授. 2005 同大学理工学術院教授, 現在に至る. IEEE, ACM, IEICE, IPSJ, 日本データベース学会各会員.