

メールマガジンを利用した注目ニュースの自動抽出

柳瀬 隆史

仲尾 由雄

(株) 富士通研究所 コンピュータシステム研究所

{yanase,nakao}@flab.fujitsu.co.jp

本稿では、メールマガジンサービスから配信されるニュース情報を対象にして、注目ニュースを自動抽出する手法を提案する。同時期に複数のメールマガジンから配信された記事内容を分析した結果、多くの記事で報じられた話題には注目に値するものが多く、同一話題を扱った記事の多くは記事見出しの比較で判定可能なことが分かった。そこで、見出しの類似性に基づいて配信記事群をグループ化し、各グループ中の記事数などをもとに話題性が高いと思われるグループを選別する実験を試みた。グループ化の結果により抽出した注目ニュースと週刊メールマガジンの掲載記事との比較などにより、各グループに含まれる記事数を適切に制御すれば、話題性の高い情報を含んだ注目ニュースを効率的に抽出できる見込みが得られた。

キーワード メールマガジン、ニュース記事、話題抽出、最長一致文字列照合、クラスタリング

Automatic Extraction of Noteworthy Topics Using E-mail Newsletter

Takashi YANASE

Yoshio NAKAO

Computer System Laboratories, Fujitsu Laboratories Ltd.

E-mail: {yanase,nakao}@flab.fujitsu.co.jp

In this paper, we propose a method of automatic extraction of noteworthy topics using news articles delivered by e-mail newsletter services. At first we manually analyzed a set of articles delivered in a week by several e-mail newsletter services, and found that most of the topics reported in many different articles are noteworthy ones and it is possible to judge the semantic identity of articles with comparison of their headlines.

Then we made an experiment of automatic extraction of noteworthy topics. In the experiment we classified delivered articles into groups based on the similarity of their headlines, and choose some groups based on the number of articles in each group.

Keywords e-mail newsletter, news articles, topic extraction, maximum length string match, clustering

1 はじめに

近年のパソコンおよびインターネットの普及により、大量の情報が個人レベルで手軽に入手できるようになった。インターネットを利用した情報の収集手段の一つとして、電子メールを用いた情報配信サービス（以下、本稿では「メールマガジン」と呼ぶ）がある。メールマガジンを利用すると、配信元にあらかじめメールアドレスを登録しておくことにより、自動的に電子メールでニュースや各種情報などの配信を受け、居ながらにして最新情報を購読することができる。WWWとは違っていちいち特定のサイトにアクセスすることなく、手軽に質の高い情報が得られるという利点もあり、今後ますます広く利用されていくサービスであると考えられる。

メールマガジンの配信を受けていて陥りやすい問題として、配信数が多くなりすぎてすべてに目を通すのが困難になるということがある。つまり、利用者がその中から有用な情報を見つけ出すのが困難になってしまうということである。

利用者がたまったメールマガジンの中から有用な情報を探すための手法としては、キーワードなどの入力による検索がある。しかし検索の場合は、利用者は「この話題に関する記事が欲しい」というような、記事内容に対する明確な検索意図を持っている必要がある。従って、例えば「記事の話題分野に関係なく、最近広く話題になっているニュースが知りたい」などという要求には応えることができない。

そこで本稿では、世間で広く話題となりそうなニュースを「注目ニュース」として自動抽出する手法を提案する。まず、広く話題になっているニュースすなわち「注目ニュース」が、実際のメールマガジンでどのように記事として配信されているかを考察する。そして実際にメールマガジンサービスから配信されたメール（以下では「配信メール」と呼ぶ）を用いて注目ニュースを抽出する手法について説明する。また各段階の処理のうち一部の処理については実装を行ったので、それについても報告する。

2 メールマガジンにおける注目ニュース

2.1 注目ニュースとは

メールマガジンの利用者は常に明確な検索意図を持って記事を探しているわけではなく、何か面白い情報はないかとメールを斜め読みしている場合も多

い。また、多くのメールマガジンを購読していて、何日間かメールを読めなかった後で、久々にメールを読もうとすると大量の配信メールがメールボックスにたまっていて、とても全部に目を通すことができないという事態に陥る。この場合、たまったメールの中から有用な情報だけを取り出して読むことができるという利点である。

ここでの有用な情報には次の2つが考えられる。

- 利用者が欲しいと思っている情報や、個人の興味に合致した情報
- 個人的な嗜好や興味に関係なく、多くの人の関心を引くような話題性の高い情報

前者は一つは従来からあるキーワード検索や情報フィルタリングなどの技術により、取り出すことが可能な情報である。

これに対し、後者のような情報が本稿でいう「注目ニュース」に相当する。これはつまり、大事件や注目を引く新製品に関するニュースなどである。このような情報に対しては、利用者は明確な検索要求を持つことができないため、キーワード検索などでは取り出すことができない。このような注目ニュースが利用者に提示できると、明確な検索要求を持っていなかった利用者でも、それを手がかりにして利用者個人にとってより有用な情報が得られる可能性もある [2]。

2.2 注目ニュースと配信記事の関係

ここで、注目ニュースと実際の記事の配信状況の関係について考えてみる。

まず、一般の新聞において注目ニュースがどのように扱われるかを考えてみる。例えばある大事件が起こった時、多くの新聞がその事件を報じた記事を掲載する。また一つの新聞をとってみても、場合によっては次の日あるいはそれ以降に続報という形で同じ事件を報じた記事が掲載される。

従って、ある期間における記事のうち、そのニュースを報じた記事総数が多いほど、あるいは記事を掲載した新聞の数が多いほど、それは注目ニュースであると考えることができる。

ニュース情報を配信するメールマガジンに関しても同様のことが言える。注目ニュースに対しては、多くのメールマガジンがそのニュースを報じた記事を掲載する。そして、それらの記事内容は同一のは

AOLとLycosが提携、InstantMessengerサービスをLycosユーザーに提供
 USNews AOLとLycos、InstantMessengerサービスで提携
 AOLとLycosのIM提携はAOLの市場独占宣言？

図 1: 同一話題記事の見出しの例

ずである。以下では、このように同一の事象（事件や新製品の発売など）を報じた記事を「同一話題記事」と呼ぶ。

あるいは、例えば続報という形で、その注目ニュースに関連した別のニュースが記事になることも考えられる（以下では、このような記事を「関連記事」と呼ぶ）。

従って、ある期間に配信された記事のうち、同一話題記事や関連記事の総数が多いほど、あるいはそれらの記事を配信したメールマガジンの種類数が多いほど、そのニュースは注目度の高いものであると考えられる。

同一話題記事では、内容が重複しているはずであるので、見出しや記事本文において多少表現の違いはあっても、表記における類似性が非常に高いと考えられる。図1にそれぞれ異なるメールマガジンで配信した同一話題記事の見出しの例を示す。この場合、どの見出しにも“AOL”や“Lycos”などの同じ固有名詞が現れているなど、見出しが非常に類似していることが分かる。

また関連記事においても、同一話題記事ほどではなくても、一部に同じ固有名詞が共通して現れるなど多少の類似性はあると考えられる。

従って図2に示すように、表記上で類似性のある記事を同一話題記事や関連記事とみなせば、そのような記事が多いものが注目ニュースであると考えられることができる。

2.3 実際のメールマガジンにおける調査

メールマガジンには、時事ニュースを定期的に配信するもの、ある特定の分野に関する情報を配信するもの、企業が自社の顧客を対象に製品情報などを配信するもの、などさまざまな性質のものがある。本研究では、主に新聞社や出版社が運営し、ニュース記事を定期的に配信するメールマガジンを対象にしている。

ニュース情報を配信するメールマガジンには、発

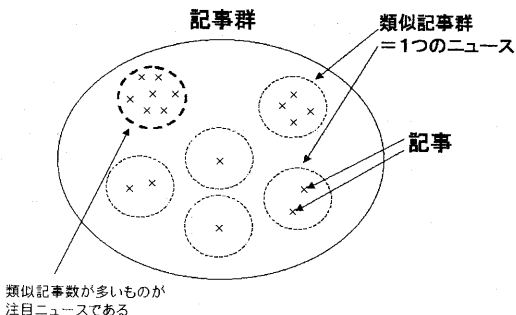


図 2: 配信記事と注目ニュース

■今週のトップニュース:

KDDのNEBER、NTTの定額制IP常時接続サービスへの対応を表明

(中略)

■今週の注目ニュース5本! ■

- ・AMD、Athlon 700MHzを発表[10/6]
- ・マイクロソフト、Windows 98 SE終了時に起こる不具合の修正Patchを配布[10/6]
- ・メモリの高騰続く、アイ・オー、メルコムオープンプライス化[10/6]
- ・ジャストシステム、ATK13を11月5日から単体発売[10/1]
- ・エレクトロニクスショー '99が専攻メッセで開催[10/6]

図 3: 週刊メールマガジン (free-watch) の一部

行間隔が日刊のもの、または週刊のものなどがある。ある週刊メールマガジンの一部を図3に示す。この図から分かるように、週刊のメールマガジンは「今週の注目ニュース」や「今週のトップニュース」などの見出しとともに、その週におけるニュースのうちからいくつかをピックアップして掲載している。これらのニュースはメールマガジンの編集者側が独自に選んだものであるが、選ぶ際の判断基準としては、読者の多くが興味を持つようなものを選んでいられると思われる。従って、週刊メールマガジンで取り上げられているニュースは、その週における注目ニュースであると考えられる。

ここで、実際のメールマガジンを題材にして、配信された記事と注目ニュースとの関係について調査してみた。

対象としたのは、コンピュータやインターネットなどの分野に関するニュースを配信する日刊メー

表 1: 同一話題記事数ごとのニュースの数

| 同一話題記事数 | ニュース数 | うち注目ニュース |
|---------|-------|----------|
| 7 記事以上 | 1 | 1 |
| 6 記事 | 2 | 1 |
| 5 記事 | 5 | 2 |
| 4 記事 | 5 | 0 |
| 3 記事 | 7 | 0 |
| 2 記事 | 12 | 1 |

週刊メールマガジンで注目記事として

取り上げられていたニュース記事数は合わせて 9 記事

ルマガジン 7 誌¹について、ある一週間（1999 年 10 月 4 日～10 月 9 日）に配信されたメール 39 通で、記事の総数は 544 記事である。

これらの記事から、見出しの類似性などを見て、同一話題記事を手作業で抽出し、同一話題記事の記事数別にニュースの数の統計をとった。この結果を表 1 に示す。また、これらのニュースのうち日刊メールマガジンと同時期に配信された別の週刊メールマガジン（図 3 のものと同じ）で注目ニュースとして取り上げられていたニュースの数も示す。

この表によれば、同一話題記事が 5 記事以上のニュースのほぼ半数は、週刊メールマガジンで取り上げられていたことになり、日刊のメールマガジンでの同一話題記事数が多いニュースが、週刊メールマガジンで注目ニュースとして取り上げられたものと、ある程度は一致していると考えられる。

なお、同一話題記事が多いニュースのうちで、週刊メールマガジンで注目ニュースとして取り上げられていないものもあるが、今回は週刊メールマガジンは 1 種類しかなく、そのメールマガジンでたまたま取り上げられなかっただけであるとも考えられる。従って、この週刊メールマガジンで取り上げられなかったニュースが注目ニュースではないとはい切れないと考えている。

3 注目ニュースの抽出手法

ここでは前節までの考察に基づき、本研究で提案するメールマガジンを利用した注目ニュースの抽

¹ pc-watch, internet-watch[5], BizIT, BizTech, SpeedBizTech[6], Mainichi Interactive[7], ZDnet Wire[8] の 7 誌。

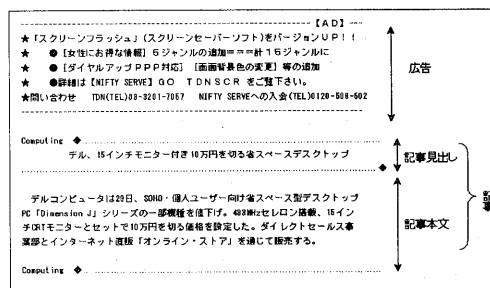


図 4: 配信メールの例と要素タイプ

出手法における各処理について、順を追って説明する。

3.1 配信メールからの記事の切り出し

一般に配信メールには話題の異なる複数の記事が含まれている。従ってまず最初の段階として、メールから個々の記事への切り出しの処理が必要となる。ここでは、実際の配信メールの書式上の特徴を利用した、記事の切り出し処理について説明する。

3.1.1 要素区間と構成要素タイプ

メールマガジンの一般的な配信メールの例を図 4 に示す。まず配信メールは、"_____" や "=====" などの、いくつかの種類の特定の文字列の連続を区切り線として、複数の区間に分割することができる [1]。また、これらの各区間にあるテキストは、記事の見出し、記事本文、広告などメールマガジンにおける様々なタイプの構成要素に対応している。

本研究では、それぞれの区間（以下では「要素区間」と呼ぶ）がもつ構成要素を、以下の 7 種類に分類した。

- 目次
- 広告
- 記事見出し
- 記事本文
- ダイジェスト (記事本文がなく、一つあるいは複数の記事見出しのみで構成される要素区間)

- その他(メールのヘッダ部分や、メールマガジンのタイトル、メールの末尾にあるお知らせなど)

このように分類すると、「記事」とは記事見出しおよび記事本文から構成されるものと考えるのが妥当である。本研究では話題を抽出するためにこの記事を単位としていくため、上に挙げた要素タイプのうち、必要となるものは記事見出しおよび記事本文である。

3.1.2 記事見出し・記事本文の抽出および記事の切り出し

次に、実際に記事を切り出す手法について説明する。

まず、2.3節で用いた数種類のメールマガジンを対象に、前節で述べた要素タイプ別にテキストの表記上の特徴などを分析した。その結果、記事見出しの行や記事見出しおよび記事本文のタイプとなる要素区間には、大まかな傾向として次のような特徴があることが分かった。

- 記事見出しの行頭には ○, ●, □ などのある特定の記号文字が用いられたり、[01] のように通し番号が振られていたりする場合が多い。
- 記事見出しの次に来る要素区間は記事本文である。
- 記事見出しの幅は4行以下で、記事本文の幅は5行以上。

そこで以下に示すような手順で、記事見出し・記事本文の抽出および記事の切り出しを行う。

1. 要素区間の行数、要素区間中のテキストに出現する特定の文字パターンの統計などを手がかりにした要素タイプ認定規則を作成する。これを分割された各要素区間に適用し、それぞれに対して要素タイプを与える。
2. 要素タイプの認定の結果、記事見出しあるいは記事本文と認定された要素区間を取り出す。さらに、一つの記事見出しと、それに続く一つあるいは複数の記事本文を、一つの記事として切り出す。

3.2 類似記事群の抽出

ここでは、各記事から見出しを取り出し、テキストの比較を行うことにより記事間の類似性を記事間類似度として数値化し、この記事間類似度を用いて類似記事群の抽出を行う。

以下で、記事見出しの文字列照合による記事間類似度の計算手法および類似記事のグループ化による類似記事群の抽出について説明する。

3.2.1 記事間類似度

記事見出しは、その記事で述べられている内容を最も端的に表現しているものである [4]。従って、記事見出しのテキストを見比べて、より長い文字列がより多く一致していれば、それらの記事は類似性が高いと考えられる。そこで、以下の手順で記事間類似度の計算を行う。

1. 双方の記事見出しから、それぞれの内容を示すいくつかの文字列(代表文字列と呼ぶ)を抽出する。具体的には図5に示すように、平仮名、カッコ、句読点など直接的には記事の内容を示す言葉にはなりにくい文字を代表文字列の境界であるとみなして見出しを分割した。なお、分割の結果2文字未満になるものは代表文字列から除外した。
2. 双方の記事見出し A_1, A_2 から抽出された代表文字列 w_i, w_j の任意の組み合わせについて、図6に示すように最長一致部分文字列 W_{ij} を抽出し最長一致部分文字列長を求め、次の(1)式に従って代表文字列間類似度 S_{ij} を求める。

$$S_{ij} = \frac{|W_{ij}|}{\min(|w_i|, |w_j|)} \quad (1)$$

ここで、 $|w_i|, |w_j|, |W_{ij}|$ は、それぞれの代表文字列あるいは最長一致文字列の長さとする。ただし、一致文字列がなかった場合、あるいは最長一致文字列が2文字未満であった場合は、代表文字列間類似度は0とする。

3. 最終的な記事間類似度は、(2)式に示すように、すべての組み合わせについての代表文字列間類似度の総和を、それぞれの代表文字列数 $n(A_1), n(A_2)$ の少ない方で割った値とする。

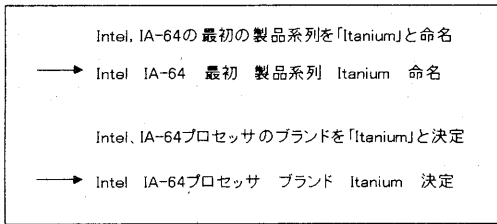


図 5: 代表文字列

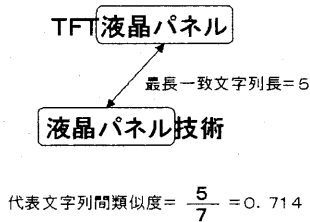


図 6: 代表文字列間類似度

$$S = \frac{\sum_{i,j} S_{ij}}{\min(n(A_1), n(A_2))} \quad (2)$$

3.2.2 類似記事群のグループ化

次に、計算された記事間類似度の結果をもとに、記事をいくつかのグループに分類する。具体的な手法については後で述べるが、基本的には互いに類似度の高い記事どうしを同じグループにする、という手法をとる。

これにより生成されたグループを「類似記事グループ」と呼ぶ。この類似記事グループ内の各記事は、互いに見出しの一部あるいは全部が類似しており、同一内容記事あるいは関連記事であることが期待できる。また、類似度の閾値を高くすれば同一話題記事のみがグループ化され、閾値を下げていくことにより、関連記事も一つのグループにまとめることもできると考えられる。

3.3 注目ニュースの選別

2.2節でも述べたように、注目ニュースにおいては、類似記事の総数や類似記事を配信するメールマ

表 2: 要素タイプの認定精度

| | 記事見出し | 記事本文 |
|-----|-------|-------|
| 全体数 | 405 | 9321 |
| 正解数 | 397 | 928 |
| 正解率 | 98.0% | 99.6% |

ガジンの種類数が多いと考えられる。従って、注目ニュースを提示する際には、類似記事グループ内の記事の総数やメールマガジンの種類数が多いグループを選別するのがよいと考えられる。

4 実際のメールマガジンへの適用

前節までで説明した、注目ニュースの自動抽出機能を実現するための処理のいくつかについて実装を行い、実際のメールマガジンへの適用を行った。ここでは、その際の問題点などについて考察する。

4.1 要素タイプの認定

2.3節で示した7種類のうち、ZDnet Wireを除いた6種類のメールマガジンからそれぞれ10通ずつ、計60通の配信メールを用意した。これらの配信メールについて要素区間への分割を行った後、まず手作業で要素タイプ認定を行い、これを正解とした。次に、今回作成した要素タイプ認定規則を適用して認定を行い、正解と比較した。なお、要素タイプは、広告、目次、記事見出し、記事本文、単一ダイジェスト記事、複数ダイジェスト記事、その他の計7種類である。これらのうち、記事見出しおよび記事本文に関する評価結果を表2に示す。

この結果から分かるように、記事見出しおよび記事本文に関しては、ほぼ正確に抽出できていることが分かる。

なお今回は、要素区間への分割に関しては定量的な評価を行っていないが、分割を誤る例として、記事中に表が含まれる場合に、表の枠線や罫線を要素区間の区切りを示す線と誤る場合が見られた。このようなテキスト文書における表形式を認定する手法も必要になると考えられる。

また、今回は対象としたメールマガジンの書式における特徴を調査して認定規則を作成したが、新たな書式を持つメールマガジンが現れた場合は、その

都度規則を追加する必要がある。この手間を低減するには、規則を自動学習させるような機能が必要になるであろう。

4.2 類似記事のグループ化と注目ニュースの抽出

次に、配信された記事の記事見出しを用いて、実際に記事間類似度の計算および類似記事のグループ化を行った。対象としたメールは、2.3節で用いたものと同じ7種類のメールマガジンで合計39通である。

まず、これらの配信メールに3.1節で説明した記事への切り出し処理を行い、誤って切り出されたものや記事内容が明らかにニュースではないもの（リンク集やランキング記事など）を取り除いた各記事について、記事見出しを抽出した。記事見出しの総数は544本である。これらの記事見出しに対して記事間類似度を計算した後、以下に示す2通りの方法でグループ化および注目ニュースの抽出を試みた。

4.2.1 類似度の閾値固定によるグループ化

まず、記事間類似度に閾値を設け、その閾値以上の類似度をもつ2つの記事をそれぞれ同じグループとする、という方法でグループ化を行った。また、例えば3つの記事 A_1, A_2, A_3 があって、 A_1 と A_2 、 A_1 と A_3 の類似度がともに閾値を超えていれば、 A_2 と A_3 の類似度が閾値を超えていなくても A_1, A_2, A_3 のすべての記事を同じグループにした。

類似度の閾値は0.7,0.8,0.9の3段階に固定した。また記事数が4記事以上のグループについて、次の3種類のランクの基準で評価した。

- A ランク
グループ中のすべての記事が、同一話題記事あるいは関連記事とみなせる。
- B ランク
一部関連記事ではない記事も含んでいるが、グループ全体ではおおよそ話題が特定できる。
- C ランク
さまざまな話題の記事が混在し、何の話題なのか特定できない。

また、A ランクおよびB ランクのグループの話題を調べ、2.3節で述べた、週刊メールマガジンに

表 3: 閾値固定による注目ニュース抽出の評価

| 閾値 | 0.9 | 0.8 | 0.7 |
|-------------------------|-----|-----|-----|
| A ランク | 0 | 3 | 4 |
| B ランク | 2 | 2 | 4 |
| C ランク | 0 | 2 | 2 |
| 合計 | 2 | 7 | 10 |
| A,B ランクのうち 注目ニュースを含む | 2 | 2 | 0 |

取り上げられた注目ニュースを含んだグループ数についても調査した。これらの結果を表3に示す。

注目ニュースの抽出については、閾値が0.8,0.9のそれぞれにおいて、週刊メールマガジンにおける注目ニュースはそれぞれ2つずつが抽出できているが、閾値を0.7にまで下げると、注目ニュースは抽出できなくなった。これは、2つの注目ニュースが同じ類似記事グループにくっついてしまい、話題の判別ができないCランクのグループになってしまったためである。

また、話題の特定しやすいA ランクおよびB ランクのグループ数は閾値0.9で2つ、閾値0.8で5つしかなく、週刊メールマガジンにおける注目ニュース以外のニュースについては、うまく抽出できているとはいいがたい。

4.2.2 記事数を固定したグループ化

前節の閾値固定による評価で分かったことをまとめると、次のようになる。

1. 類似度の高い記事は同一話題記事あるいは関連記事であることが多い。
 2. 稀に、全く無関係な記事と高い類似度を示してしまう記事があり、これが記事数の非常に多いグループを生成してしまう。
- 2.に示したような記事の割合は少ないので、グループの記事数を10記事程度にまとめれば、多数の記事が共通の話題に関連したグループとなり、見出しを読めば話題の把握に役立つと考えられる。そこで以下に示すような方法で、記事数を制限したグループ化を行った。

表 4: 記事数の制限による注目ニュース抽出の評価

| | | |
|-------------------------|---|----|
| 記事数制限値 | 5 | 10 |
| A ランク | 5 | 3 |
| B ランク | 4 | 6 |
| C ランク | 6 | 6 |
| A,B ランクのうち 注目ニュースを含む | 3 | 3 |

まず記事間類似度の高い順に、その類似度を持つ2つの記事を同じグループにした。その際、そのグループの記事数が制限値を超える場合は同じグループにはせず、別のグループにした。

次に注目ニュースの選別では、各グループの記事数の上限が制限されているため、単純に記事数の多いグループを選別することはできない。そこで、各グループ中の記事どうしの類似度のうち最大のもので大きいグループを選び、上位の一定数を注目ニュースとした。

表4には、記事の制限数を5および10、提示グループ数を上位15グループとして、前節と同様にランクづけし、週刊メールマガジンにおける注目ニュースを含んだグループ数を調べた結果を示す。

この結果から分かるように、閾値固定によるグループ化に比べて、抽出できた注目ニュースが増えた。また、その他のグループにおいても、AランクおよびBランクのグループ数が比較的多くなり、閾値固定によるグループ化に比べてが話題の抽出性能が向上したと考えられる。

4.3 記事間類似度に関する考察

今回は記事間類似度を求める際に(1)および(2)式を用いたが、本来同一話題記事でも関連記事でもない記事間の類似度が高くなってしまいう問題があった。その理由としては、以下のようなものが考えられる。

- (2) では単に代表文字列の少ない方の数で割っているということもあり、代表文字列が1つや2つなど、つまり長さの短い見出しの記事においては、どんな記事に対しても比較的類似度が高くなってしまふ。
- 今回は代表文字列に対して重みをつけていない

が、例えば「発売」や「発表」などの単語も他の固有名詞などと同じ重みとしたため、全く異なる商品の発売に関する記事どうしの類似度が高くなっている。

この他にも、現時点では同一話題記事と関連記事の分離もなかなかうまくいかないなど問題が多い。今後は、重みづけや見出しの長さによる正規化[3]などを中心に、さらに記事間類似度の計算方法の検討をしていく必要があると考えている。

5 おわりに

本稿では、メールマガジンを用いた注目ニュースの抽出手法を提案し、実際の配信メールを用いて評価を行った。その結果、類似記事のグループ化により注目ニュースが抽出できることが分った。さらに記事間類似度の計算方法を改善していくことにより、より正確な話題抽出ができるようになると考えている。

参考文献

- [1] 大久保雅且, 杉崎正之, 森大二郎, 田中一男, "話題が混在するテキストからの話題セグメントの抽出方式", 情報処理学会第57回全国大会, 1V-5, pp.209-210, 1998.
- [2] 小作浩美, 内山将夫, 村田真樹, 内元清貴, 井佐原均, "対話型ネットニュースにおける「お薦め話題」の自動抽出", 2000年情報学シンポジウム講演論文集, pp.73-80, 2000.
- [3] 角田達彦, 大石巧, 渡辺靖彦, 長尾眞, "キャプションと記事テキストの最長一致文字列照合による報道番組と新聞記事との対応づけの自動化", 情報処理学会研究報告, 98-FI-43-3, pp.17-24, 1996.
- [4] 奥雅博, 鷲崎誠司, 田中智博, "関連記事の判定に関する検討", 言語処理学会第2回年次大会, pp.89-92, 1996.
- [5] <http://www.watch.impress.co.jp/>
- [6] <http://www.nikkeibp.co.jp/>
- [7] <http://dm.mainichi.co.jp/>
- [8] <http://www.zdnet.co.jp/>