社団法人 情報処理学会 研究報告
IPSJ SIG Technical Report

2006－DBS－139 (8)
2006－ＦＩ－ 83 (8)
2006／5／31

Blogを用いた世論調査に関する研究

鎗田雄一† 成田祐一‡

†日本大学大学院工学研究科 ‡日本大学工学部

あらまし Blogを用いてコミュニティごとに世論調査をするシステムの開発をしている。本論文では、Blogから意見の判別と評価について述べる。本システムでは評価表現抽出を利用したテキスト解析により世論調査を行う。本システムでは、Blogサイトの RDF Site Summary(RSS)ファイルを定期的に解析しpingサーバに投稿されたBlogを収集し、ノイズとなるBlogをフィルタリングする。その後、ユーザにキーワードと対比できる2種類の評価表現辞書を入力してもらい、Blogから世論調査を行う。評価実験は2006年1月1日から3月31日までのデータを使い検証した。その結果、入力したキーワードを含むBlogから26%のBlogの意見を判別できた。そのうち、51%は正しく判別できていた。
キーワード: Blog、RSS、評判情報検索

A study of public opinion surveys using weblogs

Yuichi Shimota† Yuichi Narita‡
†Graduate School of Engineering, Nihon University
‡College of Engineering, Nihon University

Abstract We develop a system that tries to survey the public opinion by community from weblogs. This paper describes the discrimination of opinions from weblogs and the estimation. This system surveys the public opinion using extraction of reputation expressions and analysis of HTML documents. This system uses a RSS file and collects contributed weblog articles in a ping server of weblog sites. This system also filters noise weblogs. Users input keywords and make two comparable dictionaries including reputation expressions for a survey of public opinion. Estimation is done using collected data between January and March 2006. As a result, the rate of weblogs including reputation expressions in weblogs including keywords is 26%. The rate of the correct discrimination is 51%.
Key words: Blog , RSS , opinion information retrieval

1．はじめに

近年、一般ユーザの情報発信ツールとしてBlogと呼ばれる日記型のサービスが一般ユーザに爆発的に普及してきている。2005年5月に発表された総務省の資料[1]によると、2005年3月末時点の国内Blog利用者数は335万人となっている。そして、2007年3月末のBlog利用者数は782万人に達すると予測されている。このBlogに書かれている情報を活用しようという研究が盛んに行われている。Blogを対象とした検索エンジンとして、BlogWatcher[2]やテクノラティ[3]等が登場してきている。これらは、興味のあるBlogを検索できる。また、ホットキーワードの表示や評判情報検索を行えるものもある。

Blogの特徴は、HTMLを知らないユーザでも文章を打ち込むだけで簡単に情報発信ができることである。また、一般ユーザがコンテンツを供給するので書き込みの内容が社会的な出来事を反映した内容になる傾向がある。そして、トラックバック[4]という独特の相互リンク機能を有する。これらの特徴から次のことが言える。

まず、Blogは世論がリアルタイムに反映される情報源として捉えられる。また、Blogは情報を発信するだけでなくトラックバックを用いてコミュニケーションを行えるツールであるといえる。

現在、世の中の動向を知る上で我々の指標となるものに世論調査がある。現在の世論調査は、人手による作業であるため集計するまでに時間とコストがかかるという問題がある。今後、情報化が進むに連れて世論調査もリアルタイムに集計することが重要になってくると考えられる。そこで、本研究ではBlogから世論調査を行うことに焦点をあてる。

世論調査は正当な統計であるということを立証できるだけの論拠を提示しなければならないと考える。WEB上の掲示板において世論調査をする場合には、匿名性があるために性別や年齢別で母集合を取ることは不可能である。Blog上では、一般ユーザがBlogを書く際には、匿名ではあるがブログサービスを行っているサイトにユーザ登録をするために1ユーザを特定できる。そして、あるユーザの書いたBlogのコンテンツを追うことで、そのユーザがどのような分野に興味を持ちどのようなコミュニティに属するかを推定できる。そこで、本研究ではコミュニティごとに世論調査をするシステムの開発を目指す。本稿では、その礎としてBlogからの意見の判別と評価について述べる。

以下第2章では、関連研究について述べる。3章では本手法について詳しく述べ、4章では試作したシステムの概要を説明する。5章で評価実験について述べ、6章では考察、7章でまとめを記す。

2．関連研究

近年、文書の意見を分類する研究が盛んに行われており、Blogを扱った研究事例も増加している。

2．1　評判情報検索に関する研究

評判情報検索の先行研究として、立石ら[5] [6]の研究がある。この研究では、ユーザが入力したクエリとあらかじめ作成された評価表現辞書の単語を近接演算する方法を用いてWEBページから意見抽出を行っている。この手法ではさらに近接演算をして抽出した文書のノイズを減らす為に係り受け解析[7]により構文解析でフィルタリングを行っている。しかし、評価表現は話題によって大きく異なり、分野別の辞書の作成が容易ではないという問題があるが立石らは、複数の小規模な初期辞書からブートストラッピング的に増やすという手法で解決している。しかし、係り受け解析を行う場合は解析器に精度が依存し、未知語や日本語の文法に正確ではない文以外で精度が落ちてしまう。近接演算を行う点では同じだが、我々はこれらにも対応できるように、キーワード入力によるテキストマッチングを利用した近接演算を利用している点で方法が異なる。

2．2　意見抽出に関する研究

佐藤ら[8]によって、Peiら[9]が提案するPrefixSpanを改良し、テキストデータからシーケンシャルパターンマイニングにより感情表現抽出をする研究が行われており、アイテム間の距離(アイテム数)を30に制約すれば従来手法よりも1.56倍ものパターンを抽出できるという成果を得ている。我々も距離に注目してテキストデータの意見を取り出す点で同じだが、アイテム数ではなく1文章内のテキストデータに着目して意見を抽出・判別する点で方法が異なる。また、藤村ら[10]によって、肯定的な評判と否定的な評判をコーパスとし、コーパスから統計的に評価表現を抽出する研究が行われている。予め作成されたコーパスを基に意見を抽出するのではなく、動的に意見を抽出する点で方向性が異なる。

2．3　文書分類に関する研究

村上ら[11]によって語間の意味関係を2重のクラスタリングにより考慮し、自由記述アンケートの自動分類するシステムを開発している。村上らの研究は、文の意見を分類するのに対して我々は、文章の意見を分類する点で方向性が異なる。

2．4　Blog検索に関する研究

Blogを扱い肯定と否定の評判抽出の研究として、奥村ら[12]によってBlogWatcherが提案されている。これは、Blogを定期的に収集、解析している評判情報検索システムである。BlogWatcherでは、藤木ら[13]によってKleinbergら[14]が提案するBurst度を拡張した方式を用いてキーワードのBurst度を提示できる機能を備えている。藤木らの研究では、肯定と否定に着目して分類しているのに対して我々は様々な評価表現を扱う点で方法が異なる。また、検索システムであるため一つの文章を探し出すのが目的であるが、我々は世論調査という目的のためBlog全体でキーワードがどのように評価されているかを調べる点で方向性が異なる。

2．5　Blogの信頼度に関する研究

竹原ら[15]によってBlogの信頼度をコンテン

ツの信頼度と書き手の熟知度から算出し、入力されたキーワードと内容の深い Blog を検索エンジンの上位に表示させる研究が行われている。Blog の内容を調べるという点では同じだが、我々は Blog から有用な情報を検索するのではなく、Blog から意見を抽出・判別することが目的であり方向性が異なる。

3．提案手法について

本手法の特徴は、検索エンジンと同様にキーワードを入力しただけで即時に Blog から世論調査を行うことを目指すことである。また、本手法では Blog から意見を取り出すために使う評価表現辞書をユーザ自身に作成させることで動的に世論調査する。また、Blog の投稿期間を指定して意見を取り出すことで検索エンジンのように有益な 1 文章を探し出すのではなく、Blog 全体で入力したキーワードがどのように評価されているかを調査する。

本章では、Blog から意見を抽出する手法について触れる。まず評判情報検索と RSS のモジュール拡張[16]について書く。その後、Blog 収集とフィルタリングについて述べる。次に、本提案に用いる辞書の位置付けについて説明し、最後に Blog の意見の判別方法を記す。

3．1　評判情報検索について

評判情報検索とは、文章からある事柄に関する評判を検索する技術のことである。評判情報検索の一般的なモデルは図 1 に示すような収集部・解析部・辞書部の 3 部構成をしている。どのような文章を解析対象として集めるかという収集部。集めた文章からどのような語やパターンを評判情報として取り出すかという辞書部。そして、キーワードの入力を受けて、実際にテキスト解析で評判を結果として取り出す解析部。我々が作成するシステムもこれに準拠した形式を持つ。

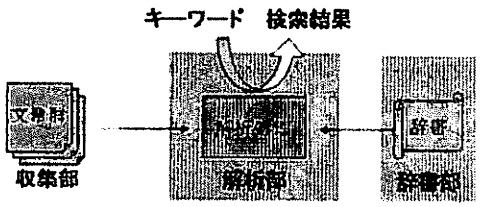

図 1. 評判情報検索システムの基本モデル

3．2　RSS のモジュール拡張

RSS は、WEB サイトの要約などの情報をメタデータとして記述した XML データのことである。日本で一般的に使われている RSS のバージョンは RSS1.0 である。RSS1.0 では、基本的なサマリー提供機能をコアな RSS として定義している。しかし、RSS のコアの要素だけでは単純であり、サイトの更新時間などの応用上重要な情報を記述する場所がないため、RSS1.0 では、モジュールという仕組みを使って公開する情報の種類を拡張している。拡張モジュールの中でよく使用される DublinCore モジュールの頻出要素を表 1 に示す。DublinCore モジュールは更新日時を記述する際に有益である。

表 1. DublinCore モジュールの頻出要素

要素名	内容
Date	時刻。フォーマットは W3CDTF の時刻表記「YYYY-MM-DDhh:mm:ss+TF」
Creator	筆者名など
Publisher	記事を提供している母体
Subject	記事のサブジェクト
rights	権利情報
language	利用されている言語。日本語は「ja」

3．3　Blog 収集とフィルタリング(収集部)

我々は Blog を解析対象とする。多くの Blog は、RSS 生成機能を備え、更新情報を URL やタイトル、要約をメタデータとして公開している。そのため、RSS を定期的に解析することで Blog の情報を得ることが出来る。解析する RSS は、先に説明した DoblineCore モジュールの頻出要素で Date 要素・Creator 要素・Publisher 要素を含む RSS を利用しているココログ[17]を解析対象とした。

作成したクローラにより、ココログの RSS ファイルを定期的に解析し ping サーバに投稿された Blog と RSS の内容をデータベースに登録する。しかし、日本における Blog には、WEB 上で書き綴る日記という側面と個人が触れた興味深いニュースや商品の感想等を書き綴るという側面がある。後者の Blog は世論調査の情報源として扱えるもが多いが前者は本研究においてノイズとなるものが多い。また、収集した Blog の HTML テキストをそのまま解析するとノイズが多く入

ってしまうことがある。そこで、本手法では収集したBlogに対して、本文の切り出しと更新回数、本文の長さによるフィルタリングを行う。

3．3．1　Blog本文の切出し

RSSを解析して収集したBlogを対象に、キーワードを抽出しようとすると、ノイズが多く入ってしまう。その原因は、ほとんどのBlogは、図2のような構造をしており本文以外にも広告や過去のBlogのタイトルやコメント等が書き込まれているためである。そこで、正規表現によりBlogの本文部分を切出す。これにより、Blog本文以外の余計なテキストを排除する。

図2. Blogの構造

3．3．2　ノイズBlogのフィルタリング

世論調査を行うにあたり意見を含んでいないBlogはノイズとなる。ノイズとなるBlogを判定するためには、Blogに書かれている情報にどれだけ信頼があるか判定する必要がある。しかし、意見を持つBlogの信頼度を数字で表すのは困難であり明らかなノイズを排除する方式を採る。

意見を含んでいないBlogとして、新聞記事等を引用しただけのBlogと広告だけを目的としたBlogが挙げられる。新聞記事等を引用しただけのBlogは客観的な情報であり抽出対象から外れる。また、引用することでBlogが長くなる傾向がある。広告だけを目的としたBlogは更新回数が多い傾向がある。これらの傾向から、ある社会的な出来事に対して何らかの意見を述べているBlogは、ある範囲内の長さで書かれていて、且つ更新回数がさほど多くない傾向がある。そこで本手法では、Blog本文の長さと同一筆者による更新頻度をフィルタリングの要素とした。Blog本文の長さは、式(1)で表される範囲のBlog以外を排除する。

$$500 \leq (Blog\ の文字数) < 3000 \quad (式\ 1)$$

また、RSSからデータベースに収集したCreator要素とpublisher要素から同一筆者の投稿回数を割り出し、1日に5回以上投稿されたBlogを排除した。

3．4　辞書の位置付け(辞書部)

世論を抽出する上で、評判情報検索に用いる辞書はシステムの精度を左右する大きな要因の一つとなる。評判情報検索に用いる辞書を本研究では評価表現辞書と呼ぶことにする。評価表現辞書作成における既存の研究での問題点は、一般的な辞書を作成するあまりにユーザの意図する意見を抽出できない点と分野別の辞書の作成が容易ではない点である。

そこで、我々は評価表現辞書をユーザ自身に作成してもらうことでこれらの問題を解決する。Blogから興味対象の評判情報検索を行うユーザは、検索を行うキーワードに対してある程度の基礎知識があると仮定できる。この仮定に基づきユーザに評価表現辞書を作成してもらうことで、自ずとユーザが意図する意見を表現する決定的な言葉が選ばれるはずであり、分野別の評価表現辞書にも対応できることが出来る。そして、世論調査という目的の為に「肯定」「否定」のように、対比させることが出来る辞書を2つ作成してもらう。これらを用いてBlogの意見を判別する。

3．5　Blogの意見の判別方法(解析部)

Blogの意見をテキスト解析により判別し世論調査を行う。作成したインターフェースから、2種類の評価表現と調べたい事柄に関するキーワードを入力してもらい、結果を解析しユーザに提示する。結果の解析には、Blogの絞込み、評価値の算出と判別の2ステップで行う。

3．5．1　Blogの絞込み

蓄積したデータベースの内容にクエリを発行して具体的に解析するBlogを絞り込み、月別に集計した。

3．5．2　評価値の算出方法と判別方法

評価値の算出と判別を行う。絞り込まれたBlogと入力されたキーワードと2つの評価表現辞書から評価表現辞書毎に評価値を算出する。評価値を算出するときに次の3点を考慮する。

一点目は、ユーザの入力したキーワードと検索されたBlogの主題が異ならないようにBlog本文のキーワードの一致数を評価値の要素とする。二点目は、評価表現辞書は決定的な言葉で作られている前提であり、評価表現辞書との一致数が多いほど意見を持つBlogといえるので評価表現辞書との一致数を評価値に加味する。三点目は、キーワードと評価表現の距離が近いほど、キーワードに対して評価しているBlogであるといえる。そこで、キーワードと評価表現の近接演算を評価

値の要素とする。単純に、キーワードと評価表現を近接演算すると、評価表現が実際にキーワードを評価しているか分からない。本来であれば係り受け解析を行い評価表現がキーワードに係っていることを確認するほうがより正確に判断できる。しかし、検索時に係り受け解析を行うと検索時間が大幅に増えてしまうと考え今回は使用しない。その代わり、キーワードと評価表現の近接演算に制約をつける。その制約は、1文の中にキーワードと評価表現が存在しているものだけを近接演算の評価値として扱うことである。これにより、1文中にキーワードと評価表現があるものが高い評価値を得る。

キーワードの出現回数を TF、1つの評価表現辞書の評価表現の出現回数を RTF、キーワードと評価表現の距離を Dis_{ij}、パラメータとして α、β、γ を与えたとき、ある Blog のある評価表現辞書に対する評価値 Score を以下の式で算出する。n は Blog に含まれるキーワードの数、m は Blog に含まれる1評価表現辞書の評価表現の数である。

$$Score = \alpha\, TF + \beta\, RTF + \sum_{i=1}^{n}\sum_{j=1}^{m}\left(\frac{\gamma}{Dis_{ij}}\right) \qquad (式2)$$

$$(Dis_{ij} > 0) \qquad (式3)$$

今回は式(2)の α を 20、β を 10、γ を 1000 として評価値を算出した。式(3)の Dis_{ij} が 0 より大きいというのは文章中でキーワードより評価表現が前方にあるものを扱うということである。

そして、入力された2つの評価表現辞書のうちどちらの評価表現辞書に適した意見を持っているかを評価値で判断する。ここで、RTF が 0 の場合はキーワードに対して意図する意見を持っていない Blog とみなし、判定を行わないものとする。また、入力した2つの評価表現辞書の評価値が同じ Blog のはどちらの意見にも数えないことにした。

実際にキーワードが「牛肉」評価表現辞書の評価表現が「食べる」として評価値を算出する例を図3に示す。

文章1：私は食べる理由は、牛肉が好きだから。

TF:1 RTF:1 キーワードと評価表現の距離:なし

Score=20*1+10*1+0=30

文章2：私は牛肉が危険でも食べる。

TF:1 RTF:1 キーワードと評価表現の距離:5

Score=20*1+10*1+1000/5=230

図3．評価値算出の例

4．試作したシステムの概要

本システムは、3章で述べた手法を用いて世論調査を行うシステムである。本システムの構成図を図4に示し流れを説明する。また、図5に実装図を示す。

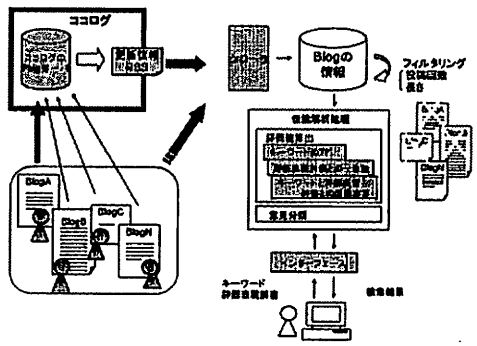

図4．システムの概要図

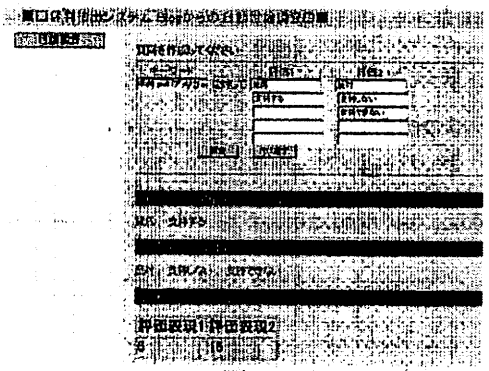

図5．システムの実装図

① RSS ファイルを定期的に解析し ping サーバに投稿された全 Blog の本文と RSS の内容をデータベースに蓄積する。

② ノイズとなる Blog をフィルタリングする。

③ ユーザにキーワードと2種類の評価表現を入力してもらう。

④ ユーザの入力と蓄積したデータを基に式(2)によりBlogを判定し、結果を表示する。

5. 評価実験

前節までに説明したシステムを用いて評価をした。評価は、ココログのRSSファイルを用いて2006年1月1日から3月31日まで収集したBlogを対象として評価実験を行った。

5．1　世論調査を行う話題について

本システムは、Blogから意見を抽出・判別して世論調査をするシステムである。評価に用いるキーワードはBlog上で話題になっているものを選ぶ。2006年に入ってからトリノオリンピック等の大きなイベントや品格を問われる凶悪事件が多く発生した。しかし、オリンピックや事件等は世論調査という項目には該当しない。やはり、政治的な問題に対して調査をするべきだと考え、2006年1月20日に発生した米国産牛肉特定危険部位混入問題(以下、牛肉問題)と、2006年2月16日の衆院予算委員会から物議をかもし出した民主党の偽メール問題(以下、偽メール問題)について本システムを用いて世論調査を行う。牛肉問題と偽メール問題に対して用いたキーワードと評価表現を図6、図7に示す。評価表現はそれぞれ2系統ずつ用意した。

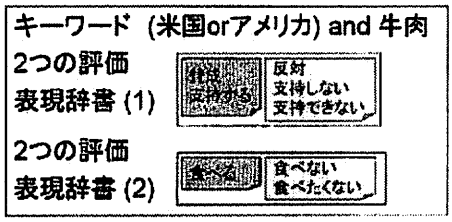

図6．牛肉問題に用いたキーワードと評価表現

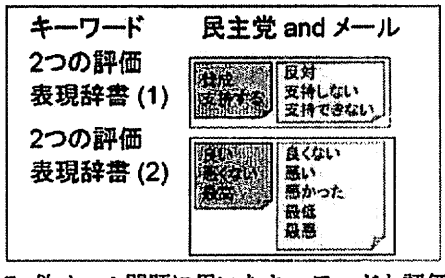

図7．偽メール問題に用いたキーワードと評価表現

5．2　評価方法

評価は、システムによる評価と人手による評価を行った。

5．2．1　システムによる評価

システムによる評価は、2006年1月から3月まで月別に調べ、図6、図7の内容をそれぞれ入力して行う。この結果を表2、表3に示す。表2、表3の抽出という項目は、データベースにクエリを発行して取り出せたBlogの数である。

表2. 牛肉問題の結果

	①賛成	②反対	③食べる	④食べない	小計	抽出
1月	6	16	21	12	55	206
2月	1	3	7	1	12	72
3月	1	7	6	4	18	82
合計	8	26	34	17	85	360

表3. 偽メール問題の結果

	⑤賛成	⑥反対	⑦肯定	⑧否定	小計	抽出
1月	0	1	1	1	3	15
2月	4	10	40	40	94	344
3月	3	10	20	19	52	184
合計	7	21	61	60	149	543

表2、表3より、キーワードにより抽出された903件のBlogから234件(26%)のBlogの意見を判別できた。本手法では、係り受け解析を用いてないために、実際に評価したい内容に対して意見を述べているか分からない。そこで、牛肉問題と偽メール問題で意見を判断できたBlog234件に対して、人手により本当に正しく意見を判別できたかを確かめる。

5．2．2　人手による評価

人手による評価方法は、本システムで抽出できた結果をそれぞれ人手により確認し、意見を判別できて正しいBlog(○)・意見を判別できたが間違っているBlog(×)・分からない(△)という3つのカテゴリに分類して行う。結果が、個人の主観に偏らないように、評価はA、Bという2人で行う。表2、表3の①、③、⑤、⑦のような世論調査においてポジティブな項目に対する意見判別の確認結果を表4に示す。また、②、④、⑥、⑧のよ

うな世論調査においてネガティブな項目に対する意見判別の確認結果を表5、全体の結果を表6に示す。表6では、表4、表5の「A(○) or B(○)」のものを意見が判別でき正しいBlog(○)と数え、「A(×) and B(×)」を意見が判別できたが間違っているBlog(×)と数え、それ以外を判別不能とする。尚、この評価を行う時にデッドリングが3件あった。

また、新聞記事等を引用しただけのBlogと広告だけを目的としたBlogを数えた結果、234件中3件であった。

表4. ポジティブな項目に対する意見判別の確認結果

B＼A	○	×	△
○	1	1	1
×	5	47	10
△	2	29	10

表5. ネガティブな項目に対する意見判別の確認結果

B＼A	○	×	△
○	61	2	4
×	2	0	2
△	38	7	9

表6. 全体の結果

	○	×	判別不能	合計
表4	10 (9%)	47 (44%)	49 (46%)	106
表5	107 (86%)	0 (0%)	18 (14%)	125
全体	117 (51%)	47 (20%)	67 (29%)	231

表6から、全体の○は51%、×は20%、判別不能は29%であった。また、表5のネガティブな項目に対する○は86%の精度で判別することに成功した。

6. 考察

いくつかのキーワードで行った実験結果に対する考察と問題点を以下に述べる。

6. 1 実験の成果について

今回ノイズとなるBlogとして、新聞記事等を引用しただけのBlogと広告だけを目的としたBlogを挙げた。5.5.2節の評価からノイズとなるBlogは3件であった。新聞記事等を引用したBlogはあったが、引用だけでなく意見を含んでいるもが多かった。この結果から、本文の長さと更新回数だけで、かなりのノイズを削除できたといえる。

6. 2 追加する機能について

今後追加している機能について5つ考える。

まず一つ目は、係り受けと今回実装してしたシステムの比較である。日本語文法の解析をせずに行った評価の結果は、表6より×が20%含まれている。これと係り受けを用いた方式とどちらが適しているかを確認する必要がある。また、二つ目として二重否定の処理が挙げられる。表6よりネガティブな項目に対する×は0%であったが、ポジティブな項目に対する×は44%であった。Blogの書き手は牛肉を食べるという意見が多いのではなく、食べないという意見のほうが多いことが分かった。この結果が出た原因は二重否定を上手く処理できていない為に起こっていたことが分かった。三つ目として、評価する期間の指定である。現在は月別に世論調査を行っているが、本来であれば、1ヵ月毎という期間は目安にしかならない。例えば牛肉問題に対して世論を追う時、ニュースが発生した日時を時点(ターニングポイント)としてBlogを解析したほうが話題に添った解析ができると考えられる。そして四つ目は入力補助である。本システムでは、評価表現辞書を含めユーザに多くの項目を入力してもらう形式を取っている。これでは、ユーザに手間を取らせてしまう。そこで、簡単に入力できるようなテンプレート形式等による入力補助機能が必要になってくると考えられる。また、本研究の目的とは異なるが、ユーザが作成した評価表現の履歴を追うことでどのような意見が必要とされているかを知ることができると考えられる。五つ目として、同一ユーザのダブルカウントという問題がある。政治問題に興味のある同一筆者のBlogが何件も抽出・判別され、一人のユーザを何回も数えてしまうという問題が考えられ、対処していく必要があるだろう。

6. 3 コミュニティの抽出について

今後、コミュニティを考慮して世論調査を行っていくにあたっての考察点を述べる。まず、あるユーザの書いたBlogを詳しく解析することによって、そのユーザのコミュニティを特定できると考えられる。また、トラックバックを用いた解析もコミュニティを特定できる重要な要素になるかもしれない。しかし、トラックバックにおけるノイズとしてトラックバックスパムが34%存在するという結果が出ている[18]。コミュニティを推定するには処理方式の検討に加えてノイズ対策も考えていかなければならないだろう。

6．まとめ

本論文では、Blogから意見を判別し世論調査するシステムの実装と評価について述べた。RSSファイルから明らかなノイズを排除しつつBlogを収集する。そして、ユーザからキーワードと2つの評価表現辞書を入力してもらうことでBlogに含まれる意見を判別し世論調査を行った。係り受け解析による日本語分析は行っていないが、Blogの意見を51%の精度で判別することに成功した。今後は、現在分かっている問題点とコミュニティをどう推定していくかという課題が残っている。これらの課題に対して様々な視点から評価と試行錯誤を繰り返し、手法を洗練していく予定である。

参考文献

[1]ブログ・SNSの現状分析及び将来予測
http://www.soumu.go.jp/s-news/2005/pdf/050517_3_1.pdf

[2]BlogWatcher
http://blogwatcher.pi.titech.ac.jp/

[3]テクノラティ
http://www.technorati.jp/home.html

[4]トラックバックの有効な使い方を考える
http://kotonoha.main.jp/2003/12/09trackback.html

[5]立石健二、石黒義英、福島俊一、インターネットからの評判情報検索、人工知能学会誌、Vol19、No.3、pp317-323、(2004)

[6]立石健二 他、Web文書集合からの意見情報抽出と着眼点に基づく要約生成、自然言語処理研究会, 2004-NL-163, pp.1-8, (September 2004)

[7]係り受け解析: CaboCha
http://chasen.org/~taku/software/cabocha/

[8]Jian Pei, Jiawei Han, and et al. Prefixspan: Mining sequential patterns by prefixprojected growth. In *Proc.of International Conference of Data Engineering*, pp.215–224, 2001.

[9]佐藤一誠、平手勇宇、山名早人、距離と属性を制約としたPrefixSpanによる感情表現抽出、DEWS2006、7A-05、(2006)

[10]藤村滋、豊田正史、喜連川優、文書分類を基にしたWeb上の評判抽出に関する一考察、DBSJ LettersVol3、No.2、(2004)

[11]村上祐人、谷澤嘉和、韓東力、原田実、意味解析による自由記述アンケートの自動分類システム AQUA、情報処理学会第 66 回全国大会、6U-2、(2004)

[12] 奥村学、南野朋之、藤木稔明、鈴木泰裕、blog ページの自動収集と監視に基づくテキストマイニング、人工知能学会研究会資料、SIG-SW&ONT-A401-01、(2004)

[13]藤木稔明、南野朋之、鈴木泰祐、奥村学、Document stream における burst の発見、情報処理学会報告、2003-NL-160、pp85-92、2004

[14]J.Kleinbert：Bursty and hierarchical structure in streams,In Proc.of the 8th ACM SIGKDD international Conference on Knowledge Discovery and DataMining,pp.1-25,2002

[15]竹原幹人、中島伸介、角谷和俊、田中克己、Web情報検索のためのBlog情報に基づくトラスト値の算出方式、DBSJ LettersVol3、No.1、(2005)

[16]新納浩幸：入門RSS　株式会社毎日コミュニケーションズ（2004-11）

[17]ココログ
http://www.cocolog-nifty.com/

[18]中島信介、舘村純一、原良憲、田中克己、上村俊亮、ブログ空間におけるトラックバック利用状況の調査および考察 DEWS2006、1B-i6、(2006)