

バイオ燃料をめぐる言説について

ネット上の記述を対象にしたテキストマイニング分析

松上 伸丈 (和光大学) 伊藤武彦 (和光大学)

キーワード: バイオ燃料, テキストマイニング, ブログ分析

問 題

環境問題の一つに地球温暖化防止がある。地球温暖化防止京都会議(1997)にて、気候変動枠組条約京都議定書(以後、京都議定書)の採択がおこなわれた。京都議定書において地球温暖化防止のために削減を求められた温室効果ガスのひとつが二酸化炭素である。日本は国会において2012年までに温室効果ガスを1990年の排出量と比較してマイナス6%の量に抑える京都議定書に批准した。そこで、京都議定書で温室効果ガス排出量としてカウントされないバイオ燃料の導入が計られ、2007年4月27日に「バイオガソリン(バイオエタノール換算3%)」の試験販売が始まった。

バイオ燃料が自動車用燃料に使われる理由は、環境省エコ燃料利用推進会議(2005)によると、「日本の石油エネルギーの5割は乗用車やトラックなどの運輸部門で消費している。また、運輸部門のエネルギーのほとんどが石油燃料に頼っている。」とある。自動車燃料の脱石油が求められている。

バイオ燃料とは、広義では植物や動物を原料に発酵または精製を行うことにより、石油燃料の代替物として使えるようにしたものを指す。種類は、トウモロコシやサトウキビといったエネルギー作物をエタノール発酵させたバイオエタノール、廃油をエステル化させたバイオディーゼルフェーゼル(BDF)などがある。現状は、アメリカではトウモロコシ、ブラジルではサトウキビを原料としたバイオエタノールの生産量が多い。その多くは自国で消費されている。ブラジルはガソリンの半額程度の価格で販売されているとの報道がある。バイオ燃料の長所は、温室効果ガスの排出を押さえられる。また、廃食用油や廃棄物を素

材とするものはリサイクルになる。化石燃料の輸入量の多い国では、化石燃料の代替物を自国生産が可能となる。短所として、トウモロコシなど人間や家畜の食料と重複する素材が使われ、食料問題となる。

バイオ燃料について先行研究をみると、市民の意識に耳を傾けたものがみあたらなかった。本研究ではブログを分析することによって市民の意見に注目する。

ブログとは、IT用語辞典によると「Web日記は独自の進化を遂げ、それまでの個人サイトでもない、紙の日記でもない新しいメディアとして台頭した。そうした新しい形式の日記風サイトを指す言葉として「Web」と「Log」(日誌)を一語に綴った「weblog」(ウェブログ)という言葉が誕生した。現在では略して「blog」(ブログ)と呼ばれることが多い。」となっている。ブログの特徴は、私生活の出来事を書く日記や、世相や時事問題、専門的話題に関しての独自の情報や見解を掲載するというものがある。記入者の率直な感想や意見が書かれている。ブログにはこのような特徴があるので、バイオ燃料についても率直な意見が示されていると考える。

ブログにおける環境問題の議論は、「環境問題」をGoogleブログ検索で検索すると、404,524件(2008/05/09検索)となった。書かれている内容を関連性が高いもの上位からみると「里山のカエルの環境」、「環境省に積極性を求める」衆議院議員のページ、「ISO14001」など多岐にわたっている。同様にバイオ燃料をみると、「バイオ燃料」を検索すると、41,518件(2008/05/09検索)となった。環境問題の検索結果と比較すると約1割とまだ少数であることが読み取れる。とはいえ、記入内容の分析をするには十分な書き込み量であると考えられる。

分析方法は自然言語解析技術の1つテキストマイニングを使用する。

テキストマイニングとは、Consumer Generated Media分析（消費者生成メディア分析）に代表される自然言語解析技術の一つである。IT情報マネジメント用語辞典によると、「形式化されていないテキストデータ（通常は自然文）を単語などに分割し、その出現頻度や相関関係などをデータマイニングの手法を使って解析することで、一定の知見や発想を得るテキストデータ分析手法の総称」となっている。

テキストマイニングの仕組みは、日本語の場合は文章上で言葉間の区切り位置が明確ではないため区切り位置の解明を最初に行う。言葉の語形変化を直し、言葉レベルに直したデータから、共起や係り受けの結果を集計出力し、言葉と言葉のつながり関係から文章全体でどの様な単語が潜在的に重要かを読み解いていく。

テキストマイニングは大量の文章の特徴、とくに形式的特徴を比較的短時間に量的に分析してくれるという特長をもつ。

目 的

バイオ燃料に対する一般市民の意識は、それを書いているブログに反映されていると考え3つの目的を設定した。1つめは、バイオ燃料またはバイオエタノールなどの言葉が検索できるブログで、どのように論じられているかを明らかにする。2つめは、「バイオガソリン」が首都圏50店舗において試験販売が開始された2007年4月以降の言説と、2007年3月以前の言説を比較し試験販売前後の言説の変化を明らかにする。3つめは、時系列にみると話題の盛衰がある、その理由を明らかにする。

方 法

バイオ燃料について論じている日本語ブログを分析対象とした。検索語は「バイオ燃料」「バイオエタノール」「バイオディーゼル」「バイオガソリン」とした。それぞれの検索語で、「Google ブログ検索」により検索した。検索期間は、2000

年1月1日～2008年3月31日とした。検索結果として約8万件のブログを確認し、検索結果を各月ごとに等間にサンプリングし1,335件のデータを作成した。

1,335件のテキストをExcelにて目視でスパムブログ（意味を成さないブログ）と重複検索を除き、改行区切りのブログには句点と句読点をいれ1,179件のデータを作成した。1,179件を有効テキストとし、分析対象とした。有効テキストの期間は、2006年4月1日～2008年3月31日となった。

Text Mining Studio ver3.01にて「ディーゼル」「ターボ」「エンジン」などと分かれてしまった言葉を「ディーゼルトーボエンジン」とまとめる作業や、「CO₂」「c o 2」「二酸化炭素」「温室効果ガス」「温暖化ガス」を「CO₂」に類義語としてまとめる作業をおこなった。

結 果

分析はText Mining Studio ver3.01を用いた。

全体分析

基本情報（表1.）は、テキスト数（サンプル数）は1,179件、平均行長（平均サンプル長）は420.7文字、総文数は19,778文、平均文長は25.1文字、述べ言葉数は、153,735単語となった。1テキストあたり言葉数は130.4単語であり、1文あたりの言葉数は7.8単語となった。

表1. 基本情報

項目	値
サンプル数	1,179 (テキスト数)
平均サンプル長(文字数)	420.7 (文字/テキスト数)
総文数	19,778 (文)
平均文長(文字数)	25.1 (文字/総文数)
延べ単語数	153,735 (単語)
単語異なり度数	31,027 (単語)

述べ言葉数の単語頻度解析（表2.）は、最多は「バイオエタノール」で1,827回、延べ言葉数の1.2%となった。次いで「作る」(1,567回, 1.0%)、「バイオ燃料」(1,389回, 0.9%)、「利用」(1,202回, 0.8%)、「上がる」(1,084回, 0.7%)となった。

表2. 延べ言葉数(言葉総数153,735単語中)

順位	単語	頻度	(%)
1	バイオエタノール	1827	(1.2)
2	作る	1567	(1.0)
3	バイオ燃料	1389	(0.9)
4	利用	1202	(0.8)
5	上がる	1084	(0.7)
6	CO2	848	(0.6)
7	トウモロコシ	804	(0.5)
8	日本	798	(0.5)
9	原料	790	(0.5)
10	燃料	752	(0.5)

言葉の出現テキスト数の単語頻度解析(表3.)では、「作る」を使ったテキストが633件で最多となった。次いで「バイオエタノール」(631件)「バイオ燃料」(582件)と検索語が上位となった。上位10位までの言葉を述べ単語数と比較すると、「自然環境」が10位以内に入り「CO2」が10位から外れた。それ以外は、順位に変動があるものの共通した言葉となった。

表3. 言葉の出現テキスト数
(1,179テキストで言及された数中)

順位	単語	件数
1	作る	633
2	バイオエタノール	631
3	バイオ燃料	582
4	利用	567
5	原料	456
6	燃料	429
7	上がる	412
8	トウモロコシ	404
9	日本	373
10	自然環境	336

係り受け分析は、総文数19,778文を基に分析された。

係り受け頻度解析(表4.)は、「自然環境→やさしい」が149回で最多となった。次いで「CO2→吸収」(109回)、「バイオ燃料→利用」(98回)となった。

上位に「CO2」を含む係り受けが3組、「利用」を含む係り受けが7組となっている。検索語の「バイオ燃料」、「バイオエタノール」「バイオディーゼル」「バイオガソリン」を含む係り受けもあった。

話題分析で係り受け関係を図にする(図1.)と11のブロックに分けられた。係り受け言葉数のつながり大きな順にみると、最も大きなブロックは「利用」を主な掛かり先とし、係り元に「バイオエタノール」「バイオ燃料」「バイオディーゼル」を含んだ。次いで「CO2」を掛かり元とし、「吸収」や「植物」「排出」「削減」を掛かり先とするブロック、「バイオガソリン」を掛かり元とし、「試験販売」「始まる」「販売」を係り先としたブロックとなった。4番目に係り受け頻度分析で1位と10位の「自然環境」を掛かり元とし、「やさしい」「良い」を係り先としたブロックがみられた。同じく4位で「影響」を掛かり元とし、「出る」「与える」を係り先としたブロックとなっていた。

表4. 全体係り受け頻度解析(総分数) [上位20]

順位	係り元単語	係り先単語	頻度
1	自然環境	やさしい	149
2	CO2	吸収	109
3	バイオ燃料	利用	98
4	燃料	利用	91
5	バイオエタノール	利用	85
6	CO2	排出	69
7	作る	利用	63
8	CO2	削減	47
8	人	いる	47
10	自然環境	良い	42
11	バイオディーゼル	利用	39
12	バイオ燃料	普及	38
12	車	走る	38
14	バイオガソリン	販売	37
14	原料	利用	37
14	植物	吸収	37
17	化石燃料	利用	36
18	影響	与える	35
18	森林	伐採	35
20	注目	集める	34

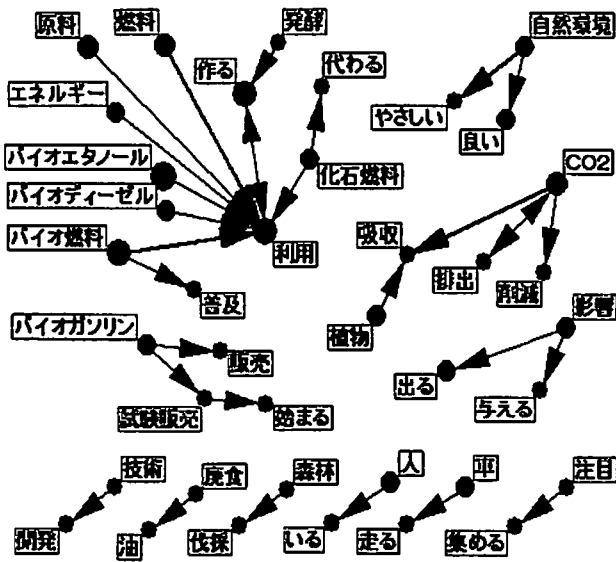


図1. 話題分析 (係り受け関係)

共起分析はテキスト数1,179行を基に分析された。

話題分析で共起関係を分析した(図2.)。分析結果は、3ブロックに分けられた。最も大きなクラスは「バイオエタノール」「作る」「CO₂」が中心となり、「吸収」「京都議定書」と共起しつながっていた。また、このブロックには「地球温暖化対策」や「バイオガソリン」「CO₂排出量」なども共起していた。次いで「小麦粉・価格・上がる」、「自然環境・やさしい」となった。

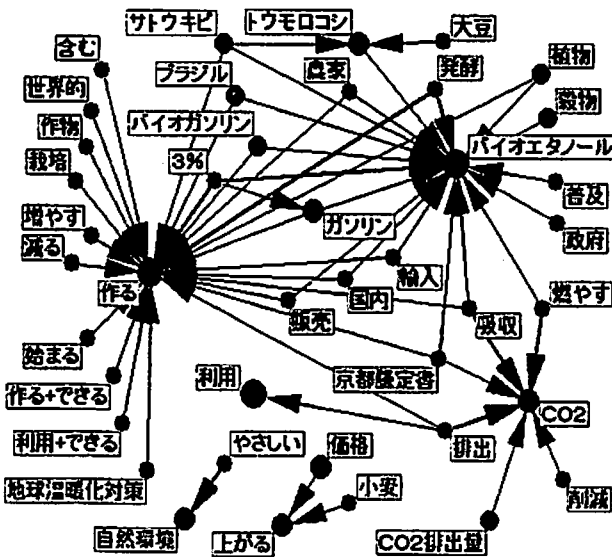


図2. 話題分析 (共起関係)

係り受け分析と比較すると、係り受け分析で発生した「自然環境・やさしい」以外は言葉が多く

の組み合わせで使われているテキストが多いことがわかった。

注目語分析とは注目語と共起関係をもつ言葉を図にする分析である。テキスト数を元に分析される。

注目語分析(図3.)では、「自然環境」(言葉の出現テキスト数336件, 10位)を注目語とした。共起関係は「やさしい」が131件で最も多く、次いで「メリット」が23件で他の言葉は20件以下と、大きく2群に分かれた。共起した言葉の間に共起関係がなかった。信頼度80、共起数15回以上の設定では「バイオ燃料」や「バイオエタノール」と共起していなかった。

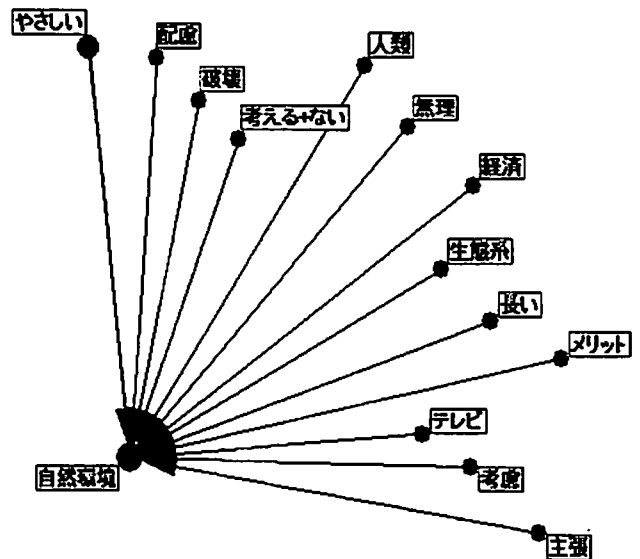


図3. 注目語分析「自然環境」

バイオガソリン発売前後分析

「バイオガソリン」が発売された2008年4月を境とし、その前後にテキストを分割し分析する。話題分析で06年度(n=100)(図4.)と07年度(n=1,079)(図5.)のテキストの言葉の共起関係を図式化した。

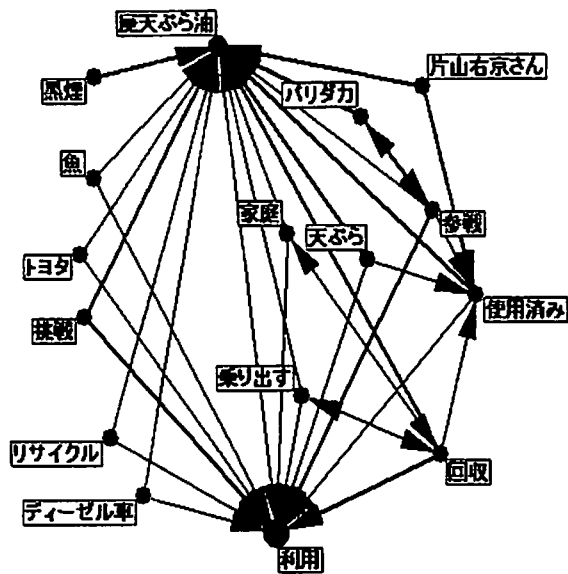


図8. 06年度の注目語分析「廃天ぷら油」

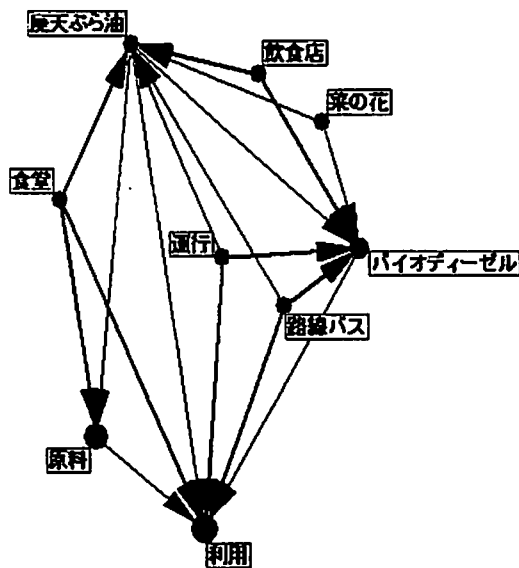


図9. 07年度の注目語分析「廃天ぷら油」

06年度に挙げられた「天ぷら油のリサイクルを象徴する言葉の「廃天ぷら油」に注目し、「廃天ぷら油」を注目語にした注目語分析をおこなった(図8. 図9.)。

「廃天ぷら油」と共起する言葉が06年度は「家庭」であることに対し、07年度は「食堂」「飲食店」があげられていた。これは06年度の意識は「家庭からリサイクルする」という“自発的な”テキストであったものが、07年度には「飲食店や食堂がリサイクルする」といった“他人まかせな”内容のテキストとなっていた。

全体としては類似した傾向の06年度と07年度であったが、細部においては変化がみられた。

時系列分析

時系列分析を行い話題の盛衰をみた。2007年4月と11月にテキストの出現数が増える(図10.)現象を月別に分析する。分析にあたりサンプリング結果(図11.)から重複を解除した合計データも同傾向であることを確認した。

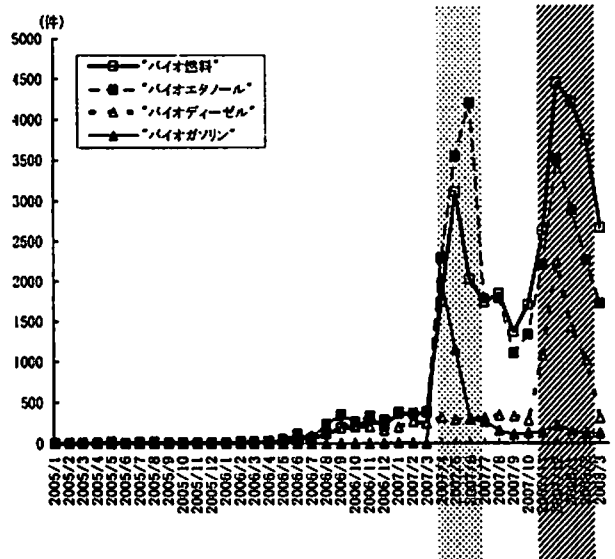


図10. Googleブログ検索による検索結果の時系列推移 (2008/05/11取得)

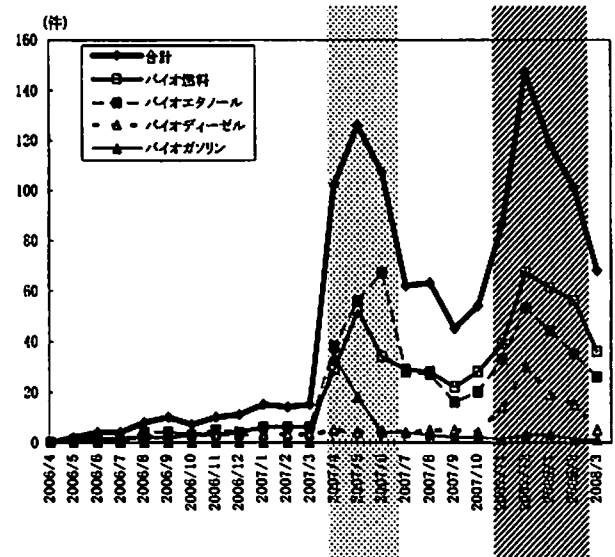


図11. 分析データの時系列推移(検索語別)

各月の話題分析を通してみると、話題のブロックが「CO2」「自然環境」「作物」「バイオ燃料」「その他」「上がる」の6項目に大きく分類されることがわかった。07年3月から08年3月までの各月の出現言葉数を簡易的な図表(図12.)にし、テキスト量の変化とテキスト内容の

変化を追った。

バイオガソリン試験発売を機に07年4月は「CO2」と「バイオ燃料」でテキストが盛り上がった。5月には「作物」と「(物価が)上がる」話に進み、6月には「CO2」と「作物」の話題が廃れ、下降傾向に入ったとみられる。

07年11月に「自然環境」と「作物」のテキスト数が上昇をはじめ。07年12月は「自然環境」「作物」「バイオ燃料」「上がる」、08年1月は「CO2」「自然環境」「作物」「バイオ燃料」「上がる」、08年2月は「自然環境」「バイオ燃料」「上がる」の話題が多くあった。暖房需要期の原油高による石油製品の高騰と合わせるように食品の値上がりが進んだ時期に、バイオ燃料が食品価格上昇を招いた話題でテキスト数が上り、2月には収束傾向に入った。

07年3月		07年4月		07年5月		07年6月	
CO2	自然環境	CO2	自然環境	CO2	自然環境	CO2	自然環境
△	—	◎	○	◎	○	○	○
作物	バイオ燃料	作物	バイオ燃料	作物	バイオ燃料	作物	バイオ燃料
—	○	○	◎	◎	△	—	△
その他	上がる	その他	上がる	その他	上がる	その他	上がる
◎	○	△	—	—	○	◎	○
07年7月		07年8月		07年9月		07年10月	
CO2	自然環境	CO2	自然環境	CO2	自然環境	CO2	自然環境
△	○	△	—	—	—	△	△
作物	バイオ燃料	作物	バイオ燃料	作物	バイオ燃料	作物	バイオ燃料
—	△	○	△	○	—	—	△
その他	上がる	その他	上がる	その他	上がる	その他	上がる
—	—	—	—	—	○	◎	○
07年11月		07年12月		08年1月		08年2月	
CO2	自然環境	CO2	自然環境	CO2	自然環境	CO2	自然環境
△	○	—	○	○	—	△	—
作物	バイオ燃料	作物	バイオ燃料	作物	バイオ燃料	作物	バイオ燃料
○	—	○	◎	◎	—	—	○
その他	上がる	その他	上がる	その他	上がる	その他	上がる
—	—	—	○	—	○	—	○
08年3月		平均		凡例 —=反応なし △=平均以下 ○=平均の2倍未満 ◎=平均の2倍以上			
CO2	自然環境	CO2	自然環境				
—	○	3.2	1.4				
作物	バイオ燃料	作物	バイオ燃料				
○	—	1.6	5.0				
その他	上がる	その他	上がる				
—	—	2.2	1.3				

図12. 各月のことばネットワーク (共起: 1文単位)比較

考察

(1) 分析から全体的に明らかになったこと

全体分析からバイオ燃料についてのブログは製造過程と流通問題についての内容がおおく、自然環境問題に関連付けて記述しているブログが少なかったのは意外な事実であった。技術的、流通的な問題の論評と環境問題の記述とが乖離し

ていることが明らかにできた。

(2) バイオ燃料への期待の変化

ブログにみられた市民の意見は、地球環境保護・地球温暖化防止・CO2削減のためにバイオ燃料の導入という政府の計画と目的から大きく外れた方向に展開していた。その傾向は時期が経過するにつれますます強まっていった。

「バイオガソリン」発売前後からは、バイオ燃料の問題点が社会的に明らかになるにつれそのイメージは悪化した。

①地域限定的な試験販売では販売店が少数であったため、実際の利用者が少なかった。試験販売以降のブログが他人事な記述に終わったのはその為であろう。

②バイオ燃料の生産が食品の価格上昇を招いたことはブログでも議論されてくるようになった。バイオ燃料の問題点が明らかになると共に善玉イメージから悪玉イメージへの変化が確認できた。

このデータを取り終わった現在、期待はますます落胆に変わっている。これは、原油価格の高騰が沈静化して、高騰期の支出負担感が消失したためである。

③当初は市民参加的な言説であった廃天ぷら油のリサイクルの話題は、後に他人まかせな姿勢へ変化したことを確認できた。

時系列分析からは、後半の時期に起きたガソリン価格の高騰とは関連付けて論じられなかった。かえって食品の値上がり結びついたネガティブな問題として語られたことが明らかになった。

(3) 分析ツールとしてのテキストマイニング

多数の実験データの自由記述を客観的かつ数量として扱えるため、心理学的研究に有効である。

言葉の量と言葉のつながりの量を視覚に訴える形で出力される特長がある。テキストマイニングは意味を検討した結果の出力がない。そのため分析者の分析対象に対する姿勢と力量を問われる。分析内容に熟知している人とテキストマイニングソフトの操作に熟知している人のチームによる分析が理想であると考えられる。

(4) 本研究の限界と課題

過去にさかのぼってデータを収集したため、情

報発信に対する最初のブログではなく、最後に参照された最新のブログが元データとなった。そのため、情報発信とのタイムラグが大きかったと考える。時代に沿ったリアルな情報を得るためには、ブログを記載された当日に収集する必要がある。また当日に収集することにより参照の多かったメディアの記事についても検証でき、重複して存在するブログ記事をそのままデータとして扱うことが可能である。

テキストマイニングの特徴として、分析データの「言葉」と「言葉のつながり」の数量を出力する。この特徴を生かすには、同程度の量の分析データを複数セット準備し同じ基準でテキストマイニングをおこない結果の相違点に焦点を当てるべきであると考え。本分析においては、1つの分析データを元にテキストマイニングを行ったため、分析結果の比較対象がなく、結果が曖昧になった。また、2年次の比較分析では、06年度と07年度の分析データ量は100件と1,000件と10倍の開きがあり同一の分析条件では比較できなかったため、データ量にあわせた分析条件で結果を出力し比較した。そのため、特徴のみられた言葉についても06年度では1桁のデータであったものもあった。

課題は、複数のデータを扱うこと、日々データ収集を行うことである。

(5) マスコミ報道とブログの語りとの関係

本論文で課題とした「バイオ燃料」についてのブログの語りは、市民活動の報告や日常の実験体験を元にした語りには比政治・産業的話題の語りが多くを占めた。そのため、メディア報道の感想や報道のコピーペーストが多くあった。しかし、その情報源といえる報道の分析を今回は行っていないため、どの程度の語りがメディア報道由来かは明らかになってない。メディア報道もブログ分析と同様にテキストマイニングを用いて分析をおこなうことが可能である。その結果をブログ分析と比較することにより、ブログで使用された言葉の変化やブログの記載内容の変化の元となったメディア報道を明らかにできるのではないかと考える。D4DRのセミナーによると「商品やニュースの検証材料として使われているプロ

グやネット掲示板・商品評価サイトの分析である。」(D4DR, 2008b)とあるが、逆にブログ分析を主とすることにより、市民の意見や考えの変化の元となった報道を探ることが可能である。

[参考・引用文献]

- D4DR(2008a)「マーケティング可視化時代におけるCGM分析」株式会社数理システムCGM分析セミナー資料(2008/03/14開催)
- D4DR(2008b)「ブログからつかみとる消費者の生の声」株式会社数理システムCGM分析セミナー資料(2008/03/14開催)
- D4DR(2008c)「ブログ分析の実際」株式会社数理システムCGM分析セミナー資料(2008/03/14開催)
- Googl ブログ検索
<http://blogsearch.google.co.jp/?utm_source=ja-wh>(2008/04/15)
- IT情報マネジメント用語辞典
<<http://www.atmarkit.co.jp/aig/04biz/textmining.html>>(2008/03/11取得)
- IT用語辞典e-Words <<http://e-words.jp/>>(2008/05/08取得)
- 環境省(1997a)地球温暖化防止京都会議
<<http://www.env.go.jp/earth/cop3/>>(2008/06/23取得)
- 環境省(1997b)京都議定書の主なポイント(平成9年12月11日:報道発表資料より)
<<http://www.env.go.jp/>>
1997.12.11<<http://www.env.go.jp/earth/cop3/kaigi/pointg.html>>(2008/06/23取得)
- 環境省エコ燃料利用推進会議(2005)『資料4 原油価格の高騰とその影響について』
<<http://www.env.go.jp/>>2005.12.13<http://www.env.go.jp/earth/ondanka/conf_ecofuel/01/mat04.pdf>(2006/06/25取得)
- 数理システム(2008a)Text Mining Studio ver3.0 操作マニュアル 数理システム
- 数理システム(2008b)Text Mining Studio ver3.0 チュートリアル 数理システム
- 数理システム(2008c)Text Mining Studio ver3.0 技術資料 数理システム