

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第4677563号
(P4677563)

(45) 発行日 平成23年4月27日(2011.4.27)

(24) 登録日 平成23年2月10日(2011.2.10)

(51) Int.Cl. F 1
G06F 17/30 (2006.01) G06F 17/30 220Z
G06Q 50/00 (2006.01) G06F 17/60 150
G06F 19/00 (2011.01) G06F 19/00 130
 G06F 17/30 180A

請求項の数 4 (全 31 頁)

(21) 出願番号	特願2006-66715 (P2006-66715)	(73) 特許権者	504171134
(22) 出願日	平成18年3月10日(2006.3.10)		国立大学法人 筑波大学
(65) 公開番号	特開2007-241901 (P2007-241901A)		茨城県つくば市天王台一丁目1番1
(43) 公開日	平成19年9月20日(2007.9.20)	(74) 代理人	100098589
審査請求日	平成19年3月5日(2007.3.5)		弁理士 西山 善章
前置審査		(72) 発明者	藤井 敦
			茨城県つくば市天王台一丁目1番1 国立 大学法人筑波大学内
		(72) 発明者	石川 徹也
			茨城県つくば市天王台一丁目1番1 国立 大学法人筑波大学内
		(72) 発明者	佐々木 千晴
			茨城県つくば市天王台一丁目1番1 国立 大学法人筑波大学内

最終頁に続く

(54) 【発明の名称】 意思決定支援システム及び意思決定支援方法

(57) 【特許請求の範囲】

【請求項1】

インターネットに接続されたマルチメディア上のデータを検索するブラウザを有するウェブ検索サーバと前記検索されたデータを収集格納するデータベースサーバとから成るデータベースと、前記データベースにアクセスして入力されたテーマについてテキストマイニングする装置と、により構成されたテキストマイニングによる意思決定支援システムであって、

前記テキストマイニングする装置は、

前記データベースにアクセスして、入力されたテーマに関連する肯定又は否定の意見に予め分類されている前記データベース内の意見テキスト群を収集し、当該意見テキスト群の中に現れる名詞句及び動詞句を前記テーマに関連する複数の論点として抽出する論点抽出手段と、

前記複数の論点毎に、当該論点を含んでいる前記意見テキスト群の中から当該論点に対する肯定又は否定の何れか一方の立場を有する意見の割合を表す固有度を算出する固有度算出手段と、

前記複数の論点毎に、前記意見テキスト群において前記肯定又は否定の何れか一方の立場の出現頻度の中で、前記一方の立場での前記論点の出現頻度の割合を求めることにより算出される前記論点に対する前記意見の重要度を算出する重要度算出手段と、

前記論点を含んでいる意見テキスト群の中から重複して出現する名詞、動詞及び形容詞を含む内容語を前記論点に関連する関連語として抽出する関連語抽出手段と、

10

20

前記複数の論点毎に肯定又は否定の複数の代表意見を前記内容語の出現頻度を計算することにより選定し、当該選定され複数の代表意見の出現頻度をスコアリングし、当該スコアリングのポイントの高い順に前記複数の代表意見を順位付けして出力する代表意見選択手段と、

前記固有度と前記重要度の2次元面上に、前記複数の論点、当該論点毎の前記固有度、前記重要度、前記関連語及び前記代表意見をプロットした画像データを出力するインタフェース手段と、

の各手段を有することを特徴とするテキストマイニングによる意思決定支援システム。

【請求項2】

前記データベースは、インターネットに接続されたマルチメディア上のデータを検索するブラウザを有するウェブ検索サーバと、前記検索されたデータを収集格納するデータベースサーバと、を含むことを特徴とする請求項1に記載の意思決定支援システム。

10

【請求項3】

前記固有度算出手段により算出される前記固有度は、個々の前記論点に対する肯定の立場を有する意見の確率を求めることにより算出されることを特徴とする請求項1に記載の意思決定支援システム。

【請求項4】

前記代表意見選択手段は、前記論点の賛成と反対の立場毎に当該論点を含む前記意見テキスト群の全てを形態素解析する手段を含むことを特徴とする請求項1に記載の意思決定支援システム。

20

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、人間が所与のテーマについて意思決定する際に関連する有用な情報を収集し提示して人間の意思決定を支援する意思決定支援システムに関する。

【背景技術】

【0002】

世の中を取り巻く環境やシステムの高度化及び複雑化の進展に伴って、人間が何らかの事項について合理的な判断を下すためには、多くの場合種々の観点からの評価基準に基づいて、より多くの情報を収集し整理してこれらを総合的に考慮した上で判断することが重要である。このため、従来からコンピュータとデータベースを利用した種々の意思決定支援システムが提案されている。

30

【0003】

特開平08-83182号公報は、意思決定すべき対象に関する情報を入力し、指定された1または複数のソフトウェアによって入力対象情報を評価し、その評価結果または評価結果に基づく代替案を出力する代替案選択処理手段を備えた意思決定支援システムであって、前記代替案選択処理手段によって意思決定を行った事例を学習し、その入力対象情報および処理手順の情報を含む事例情報を抽出し、当該抽出した事例情報を蓄積し保存する事例蓄積手段を備えて、新たな意思決定のための対象に関する情報の入力に対し、入力した対象情報と前記事例蓄積手段に蓄積している事例ごとの入力対象情報とを比較し、対象情報が類似する事例を選び出して類似した事例の評価結果を、新たな意思決定のための対象に関する評価結果として出力するようにした意思決定支援システムを開示している。

40

【0004】

また、特開2004-185614号公報は、意思の決定を行う際の複数の条件を要素として含む入力情報を受けて、所定の複数の意味要素のそれぞれと入力情報の各要素との相関及び相関強度を決定し、入力情報の各要素の相関及び相関強度によって表される目標に応じて、所定の複数の意味要素の相関及び相関強度を最適化することによって、最適化された結果を意思の決定を行う際の意思決定支援情報として出力するようにした意思決定支援装置を開示している。

【特許文献1】特開平08-83182号公報

50

【特許文献2】特開2004-185614号公報

【発明の開示】

【発明が解決しようとする課題】

【0005】

しかし、これらの従来の意思決定支援システムは何れも、既に蓄積された事例情報及び入手可能な大量の情報の中から当該事例に有する類似事例を抽出して当該事例に関連する情報を整理した状態で提示するものであったり、さらには、当該事例と先行事例及び類似事例における相関強度の傾向又は法則を発見しそれに基づいて目的遂行のための最適化された結果を示そうとするものであった。

【0006】

このため、従来の意思決定支援システムにおいては、意思決定の対象である事例そのものの関連情報（同一事例及び類似事例）の中から当該事例を構成する要素毎の価値判断をしようとするものであって、世の中の多くの人の考えや意見等を広く参照したものではなかったことから、当該事例に係わる多くの世人の考えや意見から遊離した情報提示や提案を行う結果となることが多かったのである。

【0007】

一方、近年のインターネットの目覚ましい普及及び発展に伴って、所定のテーマについてインターネットウェブサーバに掲載されているアンケート結果、意見、レビュー、感想、ブログ等閲覧することにより、当該テーマに関連する多くの客観情報及び主観情報を得ることが可能となっている。

【0008】

インターネット等で得た大量の情報の中からあるテーマに関する事項の内容を発掘して何らかの傾向等を抽出する手法をテキストマイニングと言うが、従来のテキストマイニング手法はその多くが、所与のテーマに関する客観情報と当該テーマに対する賛成又は反対等の主観情報を分類して提示するものであったことから、それを利用して意思決定のツールと使用とした場合その有用性は極めて低かったのである。

【0009】

本発明は、テキストマイニング手法を有効活用し、所与のテーマに関連する多くの世人の意見に基づいての複数の論点を参照することにより合理的且つ妥当性のある決定支援のための判断情報を提供することが可能で、操作者の操作手法に左右されずに有用性に優れた意思決定支援システムの提供を目的とする。

【課題を解決するための手段】

【0010】

このため、本発明は、インターネットに接続されたマルチメディア上のデータを検索するブラウザを有するウェブ検索サーバと前記検索されたデータを収集格納するデータベースサーバとから成るデータベースと、前記データベースにアクセスして入力されたテーマについてテキストマイニングする装置と、により構成されたテキストマイニングによる意思決定支援システムであって、前記テキストマイニングする装置は、前記データベースにアクセスして、入力されたテーマに関連する肯定又は否定の意見に予め分類されている前記データベース内の意見テキスト群を収集し、当該意見テキスト群の中に現れる名詞句及び動詞句を前記テーマに関連する複数の論点として抽出する論点抽出手段と、前記複数の論点毎に、当該論点を含んでいる前記意見テキスト群の中から当該論点に対する肯定又は否定の何れか一方の立場を有する意見の割合を表す固有度を算出する固有度算出手段と、前記複数の論点毎に、前記意見テキスト群において前記肯定又は否定の何れか一方の立場の出現頻度の中で、前記一方の立場での前記論点の出現頻度の割合を求めることにより算出される前記論点に対する前記意見の重要度を算出する重要度算出手段と、前記論点を含んでいる意見テキスト群の中から重複して出現する名詞、動詞及び形容詞を含む内容語を前記論点に関連する関連語として抽出する関連語抽出手段と、前記複数の論点毎に肯定又は否定の複数の代表意見を前記内容語の出現頻度を計算することにより選定し、当該選定され複数の代表意見の出現頻度をスコアリングし、当該スコアリングのポイントの高い順

10

20

30

40

50

に前記複数の代表意見を順位付けして出力する代表意見選択手段と、前記固有度と前記重要度の2次元面上に、前記複数の論点、当該論点毎の前記固有度、前記重要度、前記関連語及び前記代表意見をプロットした画像データを出力するインタフェース手段と、の各手段を有することを特徴とするテキストマイニングによる意思決定支援システムを提供するものである。

【0011】

ここで、前記データベースは、インターネットに接続されたマルチメディア上のデータを検索するブラウザを有するウェブ検索サーバと、前記検索されたデータを収集格納するデータベースサーバとを含む。

【0015】

また、前記代表意見選択手段は、前記論点の賛成と反対の立場毎に当該論点を含む前記意見テキスト群の全てを形態素解析する手段を含む。

【発明の効果】

【0018】

このように、本発明においては、テキストマイニング手法を有効活用し、入力された所与のテーマに関連する意見テキスト群を収集し、この意見テキスト群の中から前記テーマに関連する複数の論点を抽出してその論点毎の肯定又は否定等の立場を有する意見の割合を表す固有度と、その論点毎の立場を含んでいる意見の重要度を算出することにより、当該テーマに関連する多くの世人の意見に基づいての複数の論点を参照することにより合理的且つ妥当性のある決定支援情報を提供することを可能にすると共に、操作者の操作手法に左右されずに有用性に優れた意思決定支援システムを提供することができたのである。

【0019】

また、本意思決定支援システムは、最新の時事問題に係る論点を含むテーマについて特に有効であり、最新の意見テキストを収集するので収集する情報内容が固定化又は陳腐化することがなく、当該テーマに関連する多くの人の意見のトレンドに応じた決定支援のための判断情報の提供を可能としているのである。

【発明を実施するための最良の形態】

【0020】

以下、本発明に係る意思決定支援システム及び意思決定支援方法の詳細について説明する。

【0021】

図1は、本発明の意思決定支援システム10の構成機能ブロック図を示すものである。図1に示すように、本発明の意思決定支援システム10は、データベースサーバ118、ウェブ検索サーバ119及び複数の記憶装置(ディスク装置)22乃至24とにより構成されるデータベースを備える。

【0022】

このように、本意思決定支援システム10は、インターネットに接続されたマルチメディア上のデータを検索するブラウザを利用して、ウェブ検索サーバ119と検索されたデータを収集するデータベースサーバ118と所定のデータを格納しておく記憶手段22乃至24とから成るデータベースを有し、所与のテーマに関するテキストマイニングを行うことにより意思決定を支援するのである。

【0023】

データベースサーバ118は、データベース管理システム(DBMS)を有し、当該DBMSの制御下において、所定の検索言語(例えばSQL)によりデータ要素が格納され、検索され加工されて出力される。尚、図1に示したファイヤーウォール21は、WWWインターネット25からの外部からの不正な侵入を防ぐためのものである。

【0024】

本発明の意思決定支援システム10は、上記したデータベースにアクセスして、入力手段11により入力されたテーマに関連する意見テキスト群を収集し、この意見テキスト群の中から前記テーマに関連する複数の論点を抽出する論点抽出手段12を有する。ここで

10

20

30

40

50

、この論点収集手段12は、入力されたテーマに関連する意見テキスト群の中に現れる名詞句及び動詞句を抽出することにより前記テーマに関連する複数の論点を抽出するようにしている。これにより、所与のテーマに関連する多くの世人の意見に基づいての複数の論点を参照するようにしているので、結果的に合理的且つ妥当性のある決定支援情報を提供することを可能としている。

【0025】

そして、本発明の特に主要な構成として、論点抽出手段12により収集された複数の論点毎に、これを含んでいる意見テキスト群の中から当該論点に関して是認又は否定等の何れかの立場を有する意見の割合を表す固有度を算出する固有度算出手段13と、その複数の論点毎に、是認又は否定等の何れかの立場を含んでいる意見の重要度を算出する重要度算出手段14とを有するのである。

10

【0026】

ここで、この固有度算出手段13により算出される固有度は、個々の前記論点に対する肯定の立場を有する意見の確率を求めることにより算出され、また、重要度算出手段14による前記重要度は、全意見テキストにおいて肯定又は否定の何れか一方の立場における全論点の出現頻度の中で、当該一方の立場での論点の出現頻度の割合を求めることにより算出されるのである。このようにして、論点毎の意見の傾向とその重み付けがなされるのである。

【0027】

そして、本意思決定支援システムは、さらに、複数の論点に関連する関連語を抽出する関連語抽出手段15と、当該複数の論点毎に対する肯定又は否定の意見における複数の代表意見を選択する代表意見選択手段16を備える。ここで、関連語抽出手段15により抽出される関連語は、前記した複数の論点を含んでいる意見テキストの中から重複して出現する名詞、動詞及び形容詞等の内容語の中から抽出するようにしている。また、代表意見選択手段16は、ある論点を含む意見テキストの全てを形態素解析し、前記内容語の出現頻度を計算することにより前記複数の代表意見を選定し、当該選定され複数の代表意見の出現頻度をスコアリングすることにより複数の代表意見を順位付けして出力するようにしたのである。

20

【0028】

そして、インタフェース手段17は、表示装置の画面上において、前記固有度と前記重要度の2次元面上に前記複数の論点、前記関連語又は前記代表的意見をプロットした画像データを出力することにより、操作者に対して、意思決定支援情報を一目で視認し易い形態で表示する。これによって、操作者は、本意思決定支援システムを利用することにより、入力したテーマに関して、その操作手法に左右されずに、合理的且つ妥当性のある決定支援のための判断情報を得ることができるのである。

30

【0029】

図2は、本意思決定支援システムにおいて、収集された意見テキスト群の中から主観情報を抽出してから、意思決定を支援する情報画面を出力に至るまでのフローを模式的に表した図である。図2において、破線で囲まれている部分は、自動処理によって事前に行う。本システムは、あるテーマに対する意見を入力する。入力する意見は、対象のテーマに対して賛成か反対に分類されているものとする。入力した意見から「論点」を抽出し、それぞれの論点について、固有度と重要度を計算する。固有度は「賛成意見と反対意見のどちらで多く論じられているか」を表す尺度である。

40

【0030】

重要度は「どれだけ多くの人に論じられているか」を表す尺度である。また、抽出された論点それぞれについて、関連語と代表的な意見を求める。インタフェースによって、論点の分布を可視化する。

【0031】

図3は「株式会社による病院経営への参入」というテーマに対して論点の分布を可視化した例である。重要度を縦軸、固有度を横軸として2次元の平面上に論点が表示されてい

50

る。

【0032】

図3の中央付近にある「患者」、「診療」などの論点は、賛成派でも反対派でも論じられていることを表している。「競争」、「健康保険」、「改善」などの論点は、固有度が賛成派に寄っているので、賛成派の論拠となっている論点であることを表している。「利益」、「企業」、「医療法人」などの論点は、固有度が反対派に寄っているので、反対派の論拠となっている論点であることを表している。一方の立場に偏って論じられている論点は、逆の立場にとっては不利な論点であるか、もしくは議論が不十分なために一方の立場では未だ言及されていない論点である可能性がある。固有度を求めることで、それぞれの立場で何が論拠となっているのかがわかると同時に、議論が不十分な可能性が高い論点を発見することができる。また、図3中の反対派の論点を見てみると、「医療」の重要度が最も高く、「利益」、「企業」、「営利」と続いている。これは、反対派の人々には、これらの論点を重要と考えている人が多いことを表している。重要度を求めることで、多くの人々が何を重要と考え議論しているのかを発見することができる。

10

【0033】

インタフェース上で固有度と重要度に基づいて可視化した論点を選択すると、その論点の関連語と代表的な意見が表示される。「株式会社の病院経営への参入」というテーマにおける「情報」という論点の関連語の表示例を図4に、代表的な意見の表示例を図5に示す。図4では、「情報」の周りに「営利」や「医療」などの関連語が表示されている。「医療」や「企業」といった、関連する論点も表示されている。図5では、「情報」について論じている意見の中で代表的なものが、賛成派と反対派から選択され表示されている。そこで、ユーザは、「情報」という論点が実際にどのように論じられているかを知ることができる。

20

【0034】

Web上の掲示板などで、議論に途中から参加したユーザが既に議論された点について発言することがある。それに対して、「その点は既に検討済み」や「月日の投稿を見よ」などの反応が返ってくることもある。過去に議論された論点を把握し、重要な論点に対する代表的な意見を拾い読みすることができれば、前述のような無駄なやり取りを回避することができる。

【0035】

以下、A～Eで、「論点の抽出」、「固有度の計算」、「重要度の計算」、「関連語の抽出」、「代表的な意見の選択」の各処理について説明し、Fでインタフェースについて説明する。Gで多段階評価への応用について説明する。

30

【0036】

A. 論点の抽出

(1) 概要

システムの実装に先立ち、論点がどのような形式で表現されているのかを確認するために、ewomanとBSディベートに掲載された意見を人手で分析した。その結果、論点となる範囲は、「単語」、「フレーズ」、「係り受け(文)」、「文章」など、様々であることがわかった。本システムで扱う論点のモデルを複雑にすると、表現能力は増す。しかし、解析精度は低下してしまう。そこで、現在の自然言語処理技術で誤りを少なく抽出できる「単語」、「フレーズ」、「係り受け」で表現される論点を抽出対象とする。具体的には、意見中の名詞句と動詞句を抽出する。実際に、「株式会社の病院経営への参入」というテーマに対する反対意見の1つを本システムに入力した場合に抽出された論点の例を図6に示す。ここで、動詞句は助詞を削除して、「名詞_動詞」の形式に統一している。詳細は(3)で説明する。

40

【0037】

(2) 名詞句の抽出

名詞句を抽出するために、Chasen (<http://chasen.naist.jp/hiki/ChaSen>) によって意見テキストを形態素解析して、品詞情報を利用する。Chasenは、入力した文

50

を形態素解析し、形態素、読み、原形、品詞情報を出力するツールである。図7に、「情報を公開することは重要だ」と入力した場合のChasenの出力例を示す。「情報」、「公開」、「重要」の品詞として、それぞれ「名詞-一般」、「名詞-サ変接続」、「名詞-形容動詞語幹」が付された。このように、名詞に対して、さらに細かい分類まで出力される。これらの品詞情報を利用して名詞句を構成し、論点として抽出する。

【0038】

具体的には、表1に挙げたパターンが意見テキストに出現したら、名詞句としてまとめる。「/」は形態素の境界を表す。表1の(a)～(f)は単純に連結させるパターンであり、(g)は言い換えを行うパターンである。(g)は、表記が異なっても同じ意味である表現を統一するための規則である。

【0039】

【表1】

パターン	例
(a) 名詞の連続	情報／公開、独占／禁止／法
(b) 名詞＋の＋名詞	医療／の／質、顧客／の／ニーズ
(c) 名詞＋(と や ・)＋名詞	水／と／油、クリーム／・／スキミング
(d) 形容動詞語幹＋な＋名詞	高度／な／診療、重要／な／情報
(e) 形容詞＋名詞	高い／モラル、軽い／怪我
(f) 動詞＋名詞	急ぐ／人、調べる／方法
(g) 名詞＋(が を は の も)＋サ変動詞語幹	情報／の／公開、情報／を／公開、情報／が／公開

【0040】

名詞とサ変動詞語幹の名詞との間にある助詞を省くことにより、「情報を公開」、「情報の公開」、「情報が公開」といった表現を、全て「情報公開」に統一する。(a)～(g)のいずれかに合致するパターンが連続している場合は、各パターンを個別に抽出する。例えば、「高い医療費を削減」という文字列からは、(a)(e)(g)によって、「医療費」、「高い医療費」、「医療費を削減」を網羅的に抽出する。これは、形容詞、形容動詞、サ変動詞を含む名詞句は、論点として抽出する範囲の判断が難しいためである。

【0041】

(a)のパターンで構成される名詞句が3語以上の名詞で構成されている場合は、最長の名詞句だけを抽出する。例えば、「独占／禁止／法」の場合、「独占禁止」と「禁止法」は抽出せず、「独占禁止法」だけを抽出する。これは、「独占禁止法」について論じる人は常に「独占禁止法」という表現を用い、「独占禁止」や「禁止法」という表現は用いないと考えるためである。テーマに含まれる名詞句は、論点でなくても意見テキストに高頻度で出現する傾向にある。そこで、テーマに含まれる名詞句は論点として抽出しない。

【0042】

また、複数の語によって構成される名詞句の他に、単独の名詞も論点として扱う。ただし、一般的な語を除くため、2文字以上の名詞のみを対象とし、賛成か反対の立場内で1回しか出現しない名詞は論点として抽出しない。

【0043】

また、Chasenで「数詞」、「代名詞」、「接尾」、「非自立」、「副詞可能」、「ナイ形容動詞語幹」と解析された名詞は削除する。

【0044】

さらに、どのテーマにおいても「論点にはならない」語を経験的に集めてストップワードとした。ストップワードは単独では論点にならない。しかし、他の単語と共に名詞句を構成する場合には、論点として抽出されることがある。ストップワードを表2に示す。

【 0 0 4 5 】

【表 2】

分類	語
意見を述べるときによく使用される名詞	賛成、反対、肯定、否定、意見、理由、考え、主張、立場、思い、観点、視点、問題、ポイント、疑問、課題、利点、欠点、メリット、デメリット、気持ち、前提、不満、議論
議論に参加している人を指す名詞	皆さん、皆様、自分
副詞的に使用される名詞	大変、万が一、お互い、思いのほか、本当
接尾語的に使用される名詞	主義、部門、ついで、イズム、本人、自身、自体、時代、世代、単位、タイム、事業、レベル、特区
程度を表す名詞	大半、全面、最大、一部、程度、度合、過度、唯一、最大限
時間を表す名詞	日々、常々、最近、近年、先日、長期、年間、現在、最初
それ自体では何も表さない名詞	それなり、そこそこ、一つ、そのもの、自体、云々、諸々、方々、個々、人達
特定の語と共に使用される名詞	否応、大差、下火、一翼
どんな内容なのかが重要な名詞	状況、状態、内容、原因、方法、効果、主旨、趣旨、影響、条件、違い、実情、体制、機関、説明、事態、関係、制度、水準、行為、対象、形態、原理
単独の名詞としての頻度は低い形容動詞語幹の名詞	様々、色々、余計、もつとも、非常、同様、明らか、確か、必要、十分
主語または目的語が重要なサ変動詞語幹の名詞	追求、不足、利用、存在、確保、導入、公開、選択、優先、指摘
その他	人々、国々、一般、一言、本気、社会、公社、個々人、現行、全国、電電、期待、民間、意味、理解、個人、具体、自国、ページ、周り、同時、相手、行動、根本、現実、人間、若年、物質、暮らし、基本、法人、少子、崖っぷち、問い

10

20

30

【 0 0 4 6 】

(3) 動詞句の抽出

「株式会社による病院経営への参入」というテーマについて、「利益」という名詞だけを見ても、「利益を求める」なのか「利益を無視する」のかがわからないと、論点として適切でないことがある。そこで、C a b o C h a (<http://chasen.org/~taku/software/cabocho/>) を用いて意見テキストの係り受け解析を行い、解析結果に基づいて動詞句を抽出する。

40

【 0 0 4 7 】

図 8 に「医療は利益追求になじまない」という文を C a b o C h a で解析した例を示す。「*」の後にある数字が文節の番号を表し、その次の番号が、その文節に係る文節の番号を表している。各文節は形態素解析され、それぞれの形態素の品詞情報が出力される。「医療は利益追求になじまない」の場合は、「医療は」と「利益追求に」が「なじまない」に係っている。これらの情報を利用して、名詞と助詞で構成される文節が動詞に係っている表現を論点として抽出する。

【 0 0 4 8 】

ただし、主語または目的語となる名詞が「代名詞」か「非自立」である場合は抽出しない。また、助詞と動詞が「と+する」の場合や、抽出した動詞句が「問題がある」「気に

50

する」のような高頻度表現の場合は、論点として抽出しない。

【0049】

抽出した動詞句は、「お金を払う」と「お金は払う」のように助詞が異なっても本質的に同じ表現を統一するため、助詞を削除して「_」に置き換える。

【0050】

(2)の手法では、サ変動詞を名詞句として扱うので、動詞句として扱うのはサ変動詞以外の動詞とした。例えば、「無駄を省き経費を削減する」という文からは、「無駄を省く」と「経費を削減する」という動詞句が抽出される。しかし、「経費を削減する」は、(2)の手法により「経費削減」という名詞句として抽出される。したがって、この文からは、「無駄を省く」から助詞を省いた「無駄_省く」という論点だけが抽出される。

10

【0051】

B. 固有度の計算

論点Aの固有度とは、「論点Aがどちらの立場で多く論じられているか」を表す尺度である。論点Aの固有度を、「意見を1つ選んだとき、その意見が論点Aについて論じており、かつ賛成派の意見である」という条件付き確率と、「意見を1つ選んだとき、その意見が論点Aについて論じており、かつ反対派の意見である」という条件付き確率の差で計算する。具体的には、論点Aの固有度を数1で計算する。

【0052】

【数1】

$$P(\text{pro}|A) - P(\text{con}|A)$$

20

数1中の条件付き確率は、数2を用いて計算する。

【0053】

【数2】

$$P(X|A) = \frac{\text{論点Aを含んでいる立場Xの意見数}}{\text{論点Aを含んでいる全ての意見数}}$$

ここで、数2中のXは賛成(pro)か反対(con)のどちらかである。数1は確率の差であるから、-1以上1以下の値をとる。そのため、賛成派だけが論じている論点の固有度は1、反対派だけが論じている論点の固有度は-1となる。両方の立場で均等に論じられている論点の固有度は0に近い値をとる。

30

【0054】

表3に、「株式会社の病院経営への参入」というテーマから抽出された論点と固有度の例を示す。固有度が1に近い「健康保険」や「顧客ニーズ」などは主に賛成派の論拠であり、固有度が-1に近い「利益追求」や「医療法人」などは主に反対派の論拠である。「美容整形」の固有度は0であるので、賛成側派と反対派の両方で論じられている中立的な論点である。

【0055】

【表 3】

論点	固有度
健康保険	0.60
顧客ニーズ	0.50
情報公開	0.33
改善	0.27
美容整形	0.00
医療費	-0.60
医療機関	-0.64
医療法人	-0.69
利益追求	-0.78

10

【0056】

C. 重要度の計算

論点の重要度は、賛成派と反対派によって異なる値を持つ。立場 X における論点 A の重要度とは、「論点 A が立場 X でどれだけ多くの人に論じられているか」を表す尺度である。

20

【0057】

論じている人が多い論点ほど、その立場で重要な論点であると考えられる。

具体的には、数 3 を用いて計算する。

【0058】

【数 3】

$$\frac{\text{立場 X における論点 A の出現頻度}}{\text{立場 X における全論点の出現頻度}}$$

【0059】

ただし、まずは数 1 を用いて論点 A の固有度を計算し、固有度が大きい立場に論点 A を固定する。そして、その立場についての重要度を計算する。しかし、固有度が 0 の論点は、賛成と反対の立場ごとに重要度を求め、それらを平均して最終的な重要度とする。

30

【0060】

1 つのテーマ内で、重要度が他の論点と比べて非常に高い論点は、そのテーマによく出現する一般名詞であることが多い。そこで、経験的に、重要度が 20.0×10^{-3} 以上の名詞は論点として抽出しない。

【0061】

表 4 に、「株式会社の病院経営への参入」というテーマから抽出された論点と重要度の例を示す。「改善」から「顧客ニーズ」までの 4 件は賛成派寄りの論点であり、「医療法人」から「利益追求」までの 4 件は反対派寄りの論点である。「美容整形」は固有度が 0 であるため、賛成と反対の立場で計算した重要度の平均を重要度とする。賛成派は「改善」について主に論じており、反対派は「医療法人」について主に論じていることがわかる。

40

【0062】

【表 4】

論点	重要度
改善	9.2×10^{-3}
情報公開	7.9×10^{-3}
健康保険	5.3×10^{-3}
顧客ニーズ	3.9×10^{-3}
医療法人	4.4×10^{-3}
医療機関	3.6×10^{-3}
医療費	3.2×10^{-3}
利益追求	3.2×10^{-3}
美容整形	2.6×10^{-3}

10

【0063】

D. 関連語の抽出

1つの言葉だけを見ても意味がわからない場合や、1つの言葉から様々な連想を行う場合に、「関連語」が役に立つことがある。そこで、各論点の関連語を意見中から抽出して表示する機能を持たせる。本意思決定支援システムでは、「関連する語」と「関連する論点」の2つを関連語として抽出する。賛成派と反対派の両方で論じられている論点の場合は、固有度によって立場を固定した上で関連語を抽出する。

20

【0064】

(1) 関連する内容語の抽出

「株式会社の病院経営への参入」というテーマでは、「改善」が論点として抽出される。

このテーマにおいて「改善」の対象として論じられているのは、「医療」や「医者」の現状である。したがって、「改善」という論点と「医療」や「医者」は関連が高く、同じ意見内に共出現することが多い。そこで、論点Aが論じられている意見の中によく出現する内容語を、論点Aの関連語として抽出する。内容語とは、本意思決定支援システムでは名詞、動詞、形容詞とする。

30

【0065】

ただし、一般的な語を除くため、表2に挙げた名詞と、「する」や「なる」などの使用頻度が高い動詞は手でストップワードとして定義し、関連語から排除する。

【0066】

論点Aと内容語Bが同じ意見内に出現する頻度を単純に計上しただけでは、正確な関連度とはならない。一方の出現頻度だけが高い場合に、AとBに関連があるとは限らないからである。そこで、AとBの共出現頻度も考慮する必要がある。本手法では、論点Aと内容語Bの関連度は、数4に示すダイス係数によって計算する。

【0067】

【数4】

$$\frac{2 \times F(A, B)}{F(A) + F(B)}$$

40

【0068】

$F(A)$ は論点Aが出現する立場Xの意見数、 $F(B)$ は内容語Bが出現する立場Xの意見数、 $F(A, B)$ は論点Aと内容語Bが共出現する立場Xの意見数を表す。数4の値が大きいほど、AとBの関連度が高いと見なす。

【0069】

表5に、「株式会社の病院経営への参入」というテーマから抽出された論点と関連語の

50

例を示す。

【 0 0 7 0 】

【表 5】

論点	関連語
改善	医療、健康、保障、医者
情報公開	医療、営利、企業、消費
健康保険	セツ、安い、資本、回収
顧客ニーズ	時代、人間、ブランド、メーカー
医療法人	利益、配当、企業、追求
医療機関	国民、利益、日本、健康
医療費	保険、利益、営利、企業
利益追求	医療、患者、営利、日本
美容整形	お金、医療、病気、特区

10

【 0 0 7 1 】

(2) 関連する論点の抽出

「株式会社の病院経営への参入」というテーマの場合、「顧客ニーズ」という論点と「美容整形」という論点は、「特に美容整形では顧客ニーズに応じた医療が展開されている」といった論調で、同じ意見内で共に論じられることが多い。このような、同じ意見の中で共に論じられることが多い論点の組を抽出する。

20

論点 A と論点 B の関連度は、数 4 と同様に D i c e 係数を用いて計算する。

【 0 0 7 2 】

表 6 に、「株式会社の病院経営への参入」というテーマから抽出された論点と関連する論点の例を示す。1つの論点に対して、「関連する語」と「関連する論点」の両方として抽出される語がある。

【 0 0 7 3 】

【表 6】

論点	関連語
改善	規制、情報公開、競争、サービス
情報公開	改善、情報、アメリカ、企業
健康保険	競争、国民、サービス、医療費
顧客ニーズ	美容整形、整形
医療法人	利益、医療、会社、保険
医療機関	国民、企業、日本、利益
医療費	保険、医療、国民、営利企業
利益追求	会社、患者、営利、医療法人
美容整形	ニーズ、顧客、商売、顧客ニーズ

30

40

【 0 0 7 4 】

E . 代表的な意見の選択

ユーザが、ある論点が実際にどのような文脈で論じられているのか知りたいときに、その論点を含む意見を全て読むのは効率が悪い。そこで、ユーザが指定した論点を含んでいる意見の中から代表的な意見を選択する。

【 0 0 7 5 】

50

本意思決定支援システムでは、指定した論点を含む意見テキスト群の中で出現頻度が高い語を多く含む意見ほど、意見群全体の傾向を反映しており、代表的であると考えられる。

【0076】

論点Aを含む意見の中から代表的な意見を選択する処理は、以下の手順で行う。

1. 賛成と反対の立場ごとに、論点Aを含む意見テキスト全てを形態素解析し、内容語の出現頻度を調べる。

【0077】

2. 数5を用いて意見Xのスコアを計算する。このスコアが高いほど重要な意見であることを表す。

【0078】

【数5】

Xに含まれる内容語の頻度の和

(Xに含まれる内容語の数)×(Xに含まれる形態素の数)

長い意見ほど内容語を多く含んでいるので、有利になってしまう。そこで、その意見が含んでいる形態素の数によって正規化することで、意見の長さによる差を吸収する。

【0079】

3. スコアが高い順に意見をソートし、順位付きリストとして出力する。

この処理を賛成と反対の立場でそれぞれ行い、「賛成の立場で代表的な意見」と「反対側の立場で代表的な意見」を表示する。そこで、ユーザは同じ論点について賛成派と反対派の意見を比較することができる。

【0080】

数5において、逆の立場に含まれる内容語の重みを下げることによって、代表的な意見の選択精度を高められる可能性がある。しかし、さらなる検討は今後の研究課題である。

【0081】

F. インタフェース

A～Eで説明した手法によって、入力した意見テキスト群から、「論点」、「各論点の重要度」、「各論点の固有度」、「関連語」、「代表的な意見」が得られる。これらを集約して表示するインタフェースをFlashによって実装した。

【0082】

図3は、「株式会社の病院経営への参入」というテーマについて、固有度を横軸、重要度を縦軸として論点を表示した図である。ここで表示されている論点は、入力した意見から抽出された論点のうち、各立場における重要度が決められた閾値よりも高い論点である。

【0083】

右側の操作パネルから論点の重要度の下限を指定することで、表示する論点の個数を調節することができる。

【0084】

図4は、論点の関連語と関連する論点を表示させた結果である。論点をクリックすると、その周囲に関連度が高い語の上位4件が表示される。図4では、「情報」の関連語として、「営利」、「医療」、「企業」、「公開」が表示されている。関連する語は、複数の論点について同時に表示させ、比較することが可能である。右側の操作パネルには、クリックした論点の固有度と重要度の値が表示される。操作パネルから「関連する論点を表示する」を選択すると、関連度が高い論点のうち、その場に表示されているものの色が変わる。図4では、「情報」と関連する論点として、「情報公開」、「医療」、「営利」、「企業」という論点が表示されている。

【0085】

図5は、指定した論点について論じている意見の中から、代表的な意見を表示した結果である。詳しく調べたい論点をクリックし、操作パネルから「代表的な意見を見る」を選択すると、その論点について論じている意見の中から、Eの手法で得られる順位付きリス

10

20

30

40

50

トの上位3件が表示される。指定した論点が賛成でも反対でも論じられていれば、図5のように、賛成の立場で代表的な意見と、反対の立場で代表的な意見が同時に表示される。ユーザは、指定した論点がそれぞれの立場で実際にどのように論じられているのかを比較することができる。

【0086】

G. 多段階評価への応用

「星5つ」のような多段階評価のレビューを星の数によって賛成派と反対派に分類することで、本システムに入力することが可能である。図9は、映画「スター・ウォーズ エピソード3」のレビュー (<http://moviessearch.yahoo.co.jp/detail?ty=mv&id=321602>) のうち、星の数が1~3個のレビューを「反対」、4~5個のレビューを「賛成」とみなして入力した結果である。「アナキン」や「オビワン」といった登場人物の名前や、「戦闘シーン」、「ストーリー展開」などが論点として抽出された。

10

【0087】

H. 評価実験

(1) 概要

システムの評価方法には、「ブラックボックステスト」と「ガラスボックステスト」と呼ばれる2つの方法がある。ブラックボックステストは、入力出力のみでシステムの性能を評価する方法である。しかし、システムが複数の要素で構成されている場合は、各要素の貢献度を個別に評価する必要がある。そこで、ガラスボックステストは、システムを構成する各要素の性能を個別に評価する。システムの評価を行う場合には、これら2つの評価を行うことが必要である。

20

【0088】

本意思決定支援システムで作成したシステムは、人間の意思決定支援を目的としている。そのため、本システムに対してブラックボックステストを行う場合、システムを被験者に使用してもらって、意思決定にどのくらい役立ったかを評価する必要がある。しかし、被験者が個人かグループかによって、意思決定がユーザに及ぼす影響は異なる。また、テーマによっても得られる結果が異なる可能性がある。すなわち、「不確定な要素」が多い。不確定要素が多い評価実験を行う例として、TREC (<http://trec.nist.gov/>) や NTCIR (<http://research.nii.ac.jp/index-ja.html>) などの評価ワークショップがある。これらのワークショップでは、1年から1年半の期間をかけて、情報検索や自然言語処理の評価データや評価手法を開拓していく。言い換えれば、評価実験の方法を確立すること自体が1つの研究テーマである。そこで、本意思決定支援システムはブラックボックステストは行わずに、ガラスボックステストだけを行った。具体的には、「論点の抽出」、「名詞句の言い換え」、「固有度の計算」、「代表的な意見の選択」を評価した。

30

【0089】

評価用のデータには、時事問題に対する意見テキストを使用した。いずれも、不特定多数のユーザが投稿した意見情報を公開しているウェブサイトから人手で収集した。表7に、使用したテーマ、意見数、出典を示す。なお、「郵政民営化に賛成ですか」については、ewomanに掲載されていた意見30件に、首相官邸のホームページに掲載されていた意見18件を合わせて評価対象とした。また、以降の説明では、簡単のために表7の「略記」に示したキーワードで各テーマを区別する。評価の客観性を高めるために、各テーマにつき2名の判定者に評価を依頼した。

40

【0090】

【表 7】

テーマ(略記)	意見数		出典
	賛成	反対	
企業の成果主義に賛成ですか (成果主義)	57	29	BS デイベート
株式会社の病院経営への参入に賛成ですか (病院経営)	27	44	BS デイベート
サマータイム制を導入すべきだと思いますか (サマータイム)	14	17	ewoman
郵政民営化に賛成ですか (郵政民営化)	28	20	ewoman + 首相官邸第 2 回政策アンケートの結果について
エスカレーターの片側歩行 (エスカレーター)	29	42	YOMIURI ONLINE 生活フォーラム

10

【 0 0 9 1 】

(2) 「論点の抽出」の評価

a) 実験方法

20

本意思決定支援システムは、意見中の名詞句と動詞句を論点として抽出し、論点の分布を 2 次元の平面上に表すことで、意見が対立する構図を可視化する。そのため、論点の抽出は、システム全体の性能に及ぼす影響が大きい。具体的には、以下の観点について評価する。

【 0 0 9 2 】

i) 論点を名詞句と動詞句として抽出する手法が妥当であるか

i i) 人間が抽出した論点をシステムがどれだけ抽出できるか

「論点の抽出」の目的は、人間が抽出した論点を、システムによって正確かつ網羅的に抽出することである。そこで、正確性と網羅性を「精度」と「再現率」でそれぞれ求める。

30

【 0 0 9 3 】

テーマごとに、システムで抽出した論点と判定者が抽出した論点を比較し、数 6 と数 7 によって精度と再現率を計算する。

【 0 0 9 4 】

【数 6】

$$\text{精度} = \frac{\text{システムと判定者が抽出した論点数}}{\text{システムが抽出した論点総数}}$$

【 0 0 9 5 】

【数 7】

40

$$\text{再現率} = \frac{\text{システムと判定者が抽出した論点数}}{\text{判定者が抽出した論点総数}}$$

【 0 0 9 6 】

システムは、抽出した各論点について、まず固有度が大きい方の立場(賛成か反対)に分類する。そして、賛成と反対それぞれの立場に分類された複数の論点群を重要度に基づいてソートして、順位付きリストを作る。最後に、賛成と反対の各リストについて上位 N 件の論点を抽出し、精度と再現率を計算する。なお、N を段階的に増やしながら精度と再現率の変化を観測する。判定者には、表 7 に示したテーマに対する賛成意見と反対意見のリストを渡した。

50

【 0 0 9 7 】

判定者は、各意見を読んで、論点だと思ふ箇所を抜粋した。論点とは、各意見の「賛成または反対の論拠となっている部分」と定義した。抜粋する単位は、語、句、文などを問わない。また、1つの意見に論点だと思ふ箇所が複数ある場合は全て挙げてもらった。

【 0 0 9 8 】

図10に、判定者によって抽出された論点の例を示す。図10は「株式会社の病院経営への参入」というテーマに対する反対意見の1つであり、下線部が判定者によって抽出された論点である。

【 0 0 9 9 】

b) 実験結果

判定者とシステムが抽出した論点数の内訳を表8に示す。「判定者」の欄にある括弧内の数字は、システムが抽出した論点と完全一致した論点数を表す。

【 0 1 0 0 】

【表8】

テーマ	判定者 A	判定者 B	システム
成果主義	194 (113)	101 (45)	1968
病院経営	91 (27)	75 (14)	711
サマータイム	66(33)	194 (118)	1864
郵政民営化	77 (50)	96 (35)	949
エスカレーター	43 (21)	43 (26)	508

これらの数字から、判定者ごとに精度と再現率を求め、判定者を横断して平均した値をそのテーマでの精度と再現率とした。表9に各テーマの精度とその平均を、表10に各テーマの再現率とその平均を示す。さらに、テーマを横断して、精度と再現率をそれぞれ平均して描いた再現率・精度グラフを図11に示す。

【 0 1 0 1 】

【表9】

重要度	成果主義	エスカレーター	病院経営	郵政民営化	サマータイム	平均
0.010～	1.00	0.31	0.00	1.00	0.19	0.50
0.008～	0.50	0.31	0.13	0.50	0.20	0.33
0.006～	0.42	0.24	0.29	0.37	0.21	0.31
0.004～	0.41	0.12	0.23	0.22	0.09	0.21
0.002～	0.19	0.03	0.18	0.22	0.09	0.14
0～	0.04	0.03	0.04	0.04	0.05	0.04

【 0 1 0 2 】

【表 10】

重要度	成果主義	エスカレーター	病院経営	郵政民営化	サマータイム	平均
0.010～	0.01	0.03	0.00	0.01	0.03	0.02
0.008～	0.01	0.03	0.00	0.02	0.05	0.03
0.006～	0.02	0.05	0.02	0.03	0.06	0.04
0.004～	0.03	0.07	0.03	0.07	0.36	0.11
0.002～	0.07	0.16	0.08	0.14	0.41	0.17
0～	0.54	0.25	0.58	0.49	0.55	0.48

10

重要度が上位の論点ほど精度が高く、下位の論点ほど精度が低くなる傾向が出た。また、再現率は重要度が上位であるほど低く、下位にいくほど高くなった。このことから、ユーザがインタフェース上に表示される重要度の下限を調節することにより、精度と再現率のどちらかを優先して表示することが可能であることがわかった。

【0103】

本システムの比較対象として、判定者の精度と再現率を計算した。具体的には、一方の判定者が抽出した論点を正解として、もう一人の判定者がそれをどれだけ正確かつ網羅的に抽出できるかを評価した。表11にテーマごとの内訳を示す。全テーマの平均は、精度と再現率ともに0.10であった。本システムの精度と再現率は、それぞれ0.50と0.48であった。

20

【0104】

【表 1 1】

テーマ:成果主義

	抽出した論点数	一致した論点数	精度	再現率
判定者 A	194	23	0.23	0.12
判定者 B	101		0.12	0.23

テーマ:エスカレーター

	抽出した論点数	一致した論点数	精度	再現率
判定者 A	91	2	0.03	0.02
判定者 B	75		0.02	0.03

テーマ:病院経営

	抽出した論点数	一致した論点数	精度	再現率
判定者 A	66	13	0.07	0.20
判定者 B	194		0.20	0.07

テーマ:郵政民営化

	抽出した論点数	一致した論点数	精度	再現率
判定者 A	47	4	0.04	0.09
判定者 B	96		0.09	0.04

テーマ:サマータイム

	抽出した論点数	一致した論点数	精度	再現率
判定者 A	43	5	0.12	0.12
判定者 B	43		0.12	0.12

図 1 2 ~ 1 3 に、判定者間で抽出する論点が異なった例を示す。図 1 2 と図 1 3 はともに、「株式会社による病院経営への参入」というテーマに対する賛成意見の 1 つである。図 1 2 の下線部は判定者 A、図 1 3 の下線部は判定者 B が抽出した論点を表している。判定者 A、判定者 B とともに、「情報公開」と「モラルの低さ」を論点として抽出した。しかし、「医療ミス」と「信用」は判定者 A だけが抽出し、「高いモラル」は判定者 B だけが抽出した。

【 0 1 0 5 】

判定者によって論点として抽出する単位が異なるため、完全一致する件数が少なくなった。例えば、「苛酷な労働条件が改善される」という意見から、判定者がともに「労働条件」という言葉を中心に論点を抽出していても、抽出した箇所は「苛酷な労働条件」と「労働条件が改善」というように、表記が異なってしまう。これに対し、システムは、「苛酷な労働条件」、「苛酷な労働条件改善」、「労働条件」、「労働条件改善」のように、異なる範囲を網羅的に論点を抽出する。そのため、システムの精度と再現率の方が、判定者よりも高くなった。

【 0 1 0 6 】

また、人間が抽出した論点のうち、78.5%が名詞句、2%が動詞句、19%がそれ以外であった。この結果から、名詞句と動詞句を論点として扱う本手法では、最高で、人間が抽出した論点の約 8 割を網羅することがわかる。そこで、名詞句と動詞句を論点とし

10

20

30

40

50

て扱う本手法は有効であることがわかった。動詞句の割合は2%と少なかった。しかし、人間が抽出する論点の中に含まれていること、意見中に存在する絶対数が少ないことから、抽出しても特に弊害はないと考える。また、「時間が掛かる」のように、動詞句として抽出しないと意味がない論点が存在することから、動詞句の有用性は定性的に明らかである。

【0107】

c) 誤り分析

人間が抽出した論点のうち、システムでも抽出できた論点の割合は、名詞句が60%、動詞句が44.3%であった。人間が抽出した論点のうち、名詞句または動詞句であるにも拘わらずシステムは抽出できなかった論点と、名詞句でも動詞句でもないために抽出できなかった論点の内訳を、「エスカレーターの片側歩行」というテーマについて、表12～14に示す。

【0108】

【表12】

分類	例
(a)意見テキストにおける出現頻度が低い名詞	痴漢、苛立ち、横暴、欧米
(b)名詞の間に複数の助詞がある名詞句	注意書きとは矛盾、デパートなどでは禁止
(c)修飾部が動詞句である名詞句	お子さんを連れた人、障害を持つ者
(d)修飾部が名詞+助詞+(形容詞 形容動詞)である名詞句	1分が惜しい場合、階段歩行が困難な方

【0109】

【表13】

分類	例
(e)動詞	急ぐ
(f)進行形の動詞	急いでいる、並んでいる
(g)進行形の動詞句	海外でもやっている

【0110】

【表14】

分類	例
(h)形容詞	早い、危ない、長い
(i)名詞句+助詞+動詞	ベアリングへの負荷が増す、重大な人身事故が起きる
(j)その他	トラブルを避けるためやむをえず、漫然と運ばれるのでは困る、階段だから歩いて何が悪い

表12～14に挙げた例は、いずれも、判定者は抽出したものの、システムは抽出できなかった論点である。(a)～(j)は、いずれも抽出対象にしていなかったことに起因する誤りである。そのため、これらのパターンを抽出規則に加えることで抽出は可能になる。しかし、副作用として、正解の数よりもノイズが増える可能性がある。特に、(a)、(e)、(h)に対応すると、論点ではない一般的な語が大量に抽出されるため、ストップワードの整備が重要となる。

【0111】

(3)「名詞句の言い換え」の評価

a) 実験方法

本意思決定支援システムで作成したシステムは、論点を抽出する際に、表記が異なっても意味が同じ論点を統一するため、サ変動詞語幹の名詞を含む名詞句を以下のように言い換える。

名詞 + (が | を | は | の) + サ変動詞語幹 名詞 + サ変動詞語幹

この規則により、例えば、「情報を公開」、「情報が公開」、「情報の公開」は全て「情報公開」に統一される。

【0112】

しかし、言い換え規則の対象となる全ての名詞句が、この規則で言い換えることができるとは限らない。例えば、「他国の侵略」という名詞句において、「侵略」はサ変動詞語幹の名詞と解析されるので、本システムでは「他国侵略」と言い換えられる。しかし、言い換え前の「他国の侵略」は、「他国による侵略」という意味であるのに対して、言い換え後の「他国侵略」は「他国への侵略」という意味にも取ることができる。そこで、この規則による言い換えがどの程度妥当であるのかを評価する必要がある。

10

【0113】

判定者が表7の各テーマに対する意見から抽出した全ての論点のうち、本システムにより言い換えが行われた論点のべ58件を評価対象とした。各論点と抽出元の意見を共に判定者に見てもらい、抽出元である意見の文脈において言い換えが適切であるかを、以下の3段階で判定してもらった。

【0114】

20

i) A : 意味が変わらない、違和感がない

ii) B : 違和感があるものの、許容できる

iii) C : 意味が変わってしまう、違和感がある

どの判定者が抽出した論点であるかは問わず、言い換え対象となる全ての論点を、全ての判定者に判定してもらった。また、異なる意見から同じ論点が抽出されている場合や、異なるテーマにおいて同じ論点が抽出されている場合は、抽出元である意見の文脈ごとに判定してもらった。

【0115】

b) 実験結果

判定者2名の判定結果の平均を正解と考えた場合の本手法の正解率を、表15に示す。評価対象全体の正解率は、Aだけを正解と考えた場合が46.6%、AとBを正解と考えた場合が74.2%であった。

30

【0116】

【表15】

正解	助詞が「を」の場合	助詞が「が」の場合	助詞が「の」の場合	全体
A だけ	49.8%	70.0%	40.0%	46.6%
A と B	93.0%	70.0%	63.3%	74.2%

判定された名詞句の内訳を表16～18に示す。括弧内の数字は、その論点が複数の意見から抽出された場合の意見数を表す。今回の評価対象には、名詞間の助詞が「は」であった例は存在しなかった。

40

【0117】

【表 1 6】

言い換え前	言い換え後	判定者 A	判定者 B
利潤を追求(2)	利潤追求	A	B
採算を追求	採算追求	A	C
経営を健全化	経営健全化	A	A
動機を軽視	動機軽視	B	C
人的資源を活性化	人的資源活性化	A	C
利益を追求(6)	利益追求	A	B
サービスを向上	サービス向上	A	A
医療を空洞化	医療空洞化	A	B
利用者を差別	利用者差別	A	A
生活を安定	生活安定	B	C
成果を評価	成果評価	B	C
階段を利用(5)	階段利用	A	B

10

【 0 1 1 8 】

20

【表 1 7】

言い換え前	言い換え後	判定者 A	判定者 B
情報が公開(2)	情報公開	A	A
サービスが充実	サービス充実	A	C
サービスが向上	サービス向上	A	A
報酬がリンク	報酬リンク	C	C

【 0 1 1 9 】

30

【表 18】

言い換え前	言い換え後	判定者 A	判定者 B
社員の活性化	社員活性化	A	C
医療費の削減	医療費削減	A	A
利益の追求	利益追求	A	B
利益率の向上	利益率向上	A	A
利潤の追求	利潤追求	A	B
資産の運用	資産運用	A	A
国民の負担	国民負担	C	A
物流の効率化	物流効率化	A	C
会社組織の硬直化	会社組織硬直化	A	C
医業の経営	医業経営	B	B
医療の荒廃	医療荒廃	B	B
外資の参入	外資参入	A	A
価格の設定	価格設定	A	A
医療機関の機能	医療機関機能	B	B
仕事の評価	仕事評価	C	C
労働者の代謝	労働者代謝	B	C
第三者の評価(2)	第三者評価	C	A
公教育の徹底	公教育徹底	A	C
個人の配慮	個人配慮	C	C
過疎地のカバー	過疎地カバー	B	C
税金の無駄遣い	税金無駄遣い	B	C
病院の経営	病院経営	B	B
経済の活性化	経済活性化	A	A
社会情報の物質化	社会情報物質化	B	C
自己満足の活動	自己満足活動	C	C
個人の評価	個人評価	C	A
チームの評価	チームの評価	B	A
会社の配当	会社の配当	C	C
失敗例の検討	失敗例の検討	B	C

10

20

30

40

異なる意見から同じ論点が抽出されている場合や、異なるテーマにおいて同じ論点が抽出されている場合は、抽出元である意見ごとに判定を行ってもらった。しかし、今回の評価実験では、同じ論点を言い換えた結果の適否が意見によって異なる例はなかった。

【0120】

今回の実験結果からは、「A」と「C」と判定された並びや、助詞による違いから何らかの法則性を発見することはできなかった。

【0121】

現在のシステムは、言い換え規則が適用できるもの全てを言い換えている。しかし、同じテーマの中に「表記は異なっても意味は同じ」論点が存在しない場合は、そもそも言い換える必要がない。言い換えの必要がある場合のみに規則を適用することで、言い換えに

50

よる誤りを減らすことができる。また、言い換え前の名詞句と言い換え後の名詞句の出現頻度をコーパスで調べて、出現頻度が高い表現に統一する方法も有効である。

【0122】

(4) 「固有度の計算」の評価

a) 実験方法

本システムでは、論点Aの出現する確率が賛成派と反対派のどちらに多いかに基づいて、論点Aの固有度を計算する。固有度は-1以上1以下の値をとり、正ならば賛成派に、負ならば反対派に固有な論点であることを示す。

【0123】

しかし、賛成意見の中で論点Aについて論じていても、実際は「賛成だが、Aについては反対である」のように、逆の立場の論拠として記述されることがある。このような場合、本手法で計算する固有度は、その論点が属する「真の立場」と一致しないことがある。

10

【0124】

また、論点Aを含んでいる意見の中には、Aを含んでいるだけで、実際にはAについて論じていない意見もある。本手法では「Aを含んでいる意見」を対象として固有度を計算するため、「Aを論じている意見」を対象とする理想的な固有度と値が異なる場合がある。

【0125】

そこで、本手法で計算する固有度がどの程度妥当であるのかを評価した。

固有度は連続値であり、「賛成もしくは反対の度合い」を表す。しかし、人間が各意見に対して「賛成もしくは反対の度合い」を判定することは難しい。そこで、間接的な評価方法として、人間が判定した論点Aが属する立場と、本手法の固有度によって求めた論点Aが属する立場を比較した。

20

【0126】

判定者には、各意見から論点を抽出してもらう際に、その論点が、抽出元の意見が属する立場と逆の立場の論拠であると思う場合は、その旨を明記して抽出してもらった。例えば、「株式会社による病院経営への参入」というテーマの賛成派に属する「賛成だが、やはり利益追求に走るのではという懸念がある」という意見の場合、「利益追求」は反対派の論拠なので、「利益追求(逆)」という形で抽出してもらった。この場合、「利益追求」に対する真の立場は「反対」である。判定者による「真の立場」と、システムが決定した立場を比較した。

30

【0127】

b) 実験結果

表19に、判定者ごとの判定結果を示す。「システムも抽出した論点数」は、その判定者が抽出した論点のうち、システムが抽出した論点と完全一致した論点の数を表す。「立場が同じ論点数」は、判定者による「真の立場」と、システムが決定した立場が一致した論点の数を表す。「正解率」は、固有度による分類の正解率を表す。

【0128】

【表 19】

テーマ:成果主義

判定者	システムも抽出した論点数	立場が同じ論点数(%)	正解率
A	113	111	98.2%
B	45	44	97.7%

テーマ:エスカレーター

判定者	システムも抽出した論点数	立場が同じ論点数(%)	正解率
A	27	26	96.3%
B	14	14	100.0%

テーマ:病院経営

判定者	システムも抽出した論点数	立場が同じ論点数(%)	正解率
A	33	30	91.0%
B	118	111	94.1%

テーマ:郵政民営化

判定者	システムも抽出した論点数	立場が同じ論点数(%)	正解率
A	50	46	92.0%
B	118	32	91.4%

テーマ:サマータイム

判定者	システムも抽出した論点数	立場が同じ論点数(%)	正解率
A	21	20	95.2%
B	26	26	100.0%

テーマごとに、判定者の判定による「真の立場」とシステムが決定した立場が一致した論点数の割合を平均し、それらをさらにテーマを横断して平均し、システムの正解率を求めた。全テーマを総合した正解率は95.6%であった。

【0129】

(5) 「代表的な意見の選択」の評価

a) 実験方法

本意思決定支援システムで提案するシステムは、インタフェース上で、ユーザが選択した論点について論じている意見の中から代表的な意見を表示するという機能を持っている。「全ての意見を読まなくても議論の要点がわかる」という目的を達成するためには、より重要な意見を優先して表示する必要がある。そこで、人間が「代表的」と判定した意見と、本手法で「代表的」と判断された意見を比較し、本手法で選択する代表的な意見が本当に代表的かどうかを評価する。

【0130】

「論点の抽出」の評価と同じように、判定者には表7に挙げたテーマと、それに対する賛成意見と反対意見のリストを渡した。そして、自分が意見中から抽出した論点Aについて、論点Aについて論じていると思う意見を全て挙げてもらった。このとき、論点Aについて論じている意見が複数ある場合には、その中で最も重要だと思える意見を、「代表的な意見」として選んでもらった。論点Aについて論じている意見が、賛成と反対の両方に複

10

20

30

40

50

数ある場合は、それぞれの立場から重要だと思ふ意見を選択してもらった。また、重要だと思ふ意見が複数ある場合は全て挙げてもらった。

【0131】

表記は異なっても同じ意味を表す論点がある場合は、代表的な意見はそれらの各論点について論じている意見全ての中から選択してもらった。例えば、「情報公開」と「情報開示」が同じ意味であると判断した場合は、「情報公開」について論じている意見が3件、「情報開示」について論じている意見が2件あれば、「情報公開」と「情報開示」から成る論点群の代表的な意見は、以上の5件から選択してもらった。これは、表記の異なりごとに代表的な意見を考えて場合に、その論点について論じている意見が1件しかなく、重要だと考える比較対象がなくなるケースを防ぐためである。このため、評価実験のためのシステムへの入力も、先の「情報公開」と「情報開示」の例では、それについて論じている5件の意見と、単に「情報公開」または「情報開示」を含んでいる意見を入力とする。そして、人間が最も代表的だと判断した意見が、システムでも代表的と判断されるのかどうかを調べた。

10

【0132】

判定者の評価によって、各意見を次のように分類した。

- i) A : その論点について論じている意見の中で代表的な意見
- ii) B : その論点について論じているものの、代表的ではない意見
- iii) C : その論点を含んでいても、その論点について論じられていない意見

判定者は、その論点について「論じている」と判断した意見の中から代表的な意見を選ぶ。しかし、システムは、その論点を「含んでいる」意見群に重要さの順位を付けるので、含んでいるだけで論じてはいない意見(Cに該当)にも順位を付ける。そこで、システムが出力した順位付きリストの上位10件について、1位から順番に、判定者が「論じている」と判断した意見、「代表的」と判断した意見に対する精度と再現率を調べた。正解は「Aだけ」と「AとB」の2種類を考える。

20

【0133】

順位付きリストのN位までをシステムの出力と見なした場合の精度と再現率は、数8と数9で計算する。

【0134】

【数8】

$$\text{精度} = \frac{\text{上位N件までに含まれる正解数}}{N}$$

30

【0135】

【数9】

$$\text{再現率} = \frac{\text{上位N件までに含まれる正解数}}{\text{正解の件数}}$$

ただし、論じている意見が複数ある論点と意見の対だけを対象とし、論じている意見が1つしかない論点と意見の対は対象としなかった。

40

【0136】

b) 実験結果

テーマごとに判定者間の平均を計算し、さらにテーマを横断して平均した精度と再現率を表20～23に示す。全テーマの平均をとった精度と再現率のグラフを図14に示す。

【0137】

【表 2 0】

	1位	2位	3位	4位	5位	6位	7位	8位	9位	10位
成果主義	0.380117	0.337719	0.256929	0.1875	0.16	0.113333	0.116667	0.0975	0.11	0.1
エスカレーター	0.356818	0.2875	0.144375	0.108631	0.08	0.080667	0.067444	0.076563	0.066786	0.060714
病院経営	0.51061	0.381963	0.251429	0.177083	0.107143	0.099167	0.077	0.0625	0.055	0.05
郵政民営化	0.373626	0.28869	0.251429	0.229167	0.171429	0.1275	0.07	0	0.11	0.1
サマータイム	0.291855	0.343891	0.279231	0.1875	0.15	0.17	0.14	0.125	0.11	
	0.355875	0.303387	0.219159	0.171163	0.131182	0.113286	0.083903	0.066871	0.078631	0.07487

【 0 1 3 8 】

【表 2 1】

	1位	2位	3位	4位	5位	6位	7位	8位	9位	10位
成果主義	0.830065	0.828431	0.733968	0.708333	0.675	0.74875	0.811667	0.91	0.89	0.85
エスカレーター	0.8125	0.810606	0.8145	0.745238	0.733333	0.737545	0.758091	0.75175	0.753056	0.779167
病院経営	0.982143	0.950487	0.895417	0.897727	0.871429	0.831667	0.856	0.916667	0.945	0.875
郵政民営化	0.794872	0.833333	0.817679	0.859375	0.8	0.57	0.4875	0.75	0.78	0.8
サマータイム	0.941176	0.920343	0.878333	0.875	0.9	0.833333	0.715	0.5	0.44	
	0.837904	0.815533	0.767205	0.759471	0.741555	0.689475	0.664432	0.698827	0.691898	0.83

10

【 0 1 3 9 】

【表 2 2】

	1位	2位	3位	4位	5位	6位	7位	8位	9位	10位
成果主義	0.452546	0.641538	0.728571	0.666667	0.8	0.666667	0.833333	0.75	1	1
エスカレーター	0.348485	0.498106	0.425	0.392857	0.308333	0.427273	0.427273	0.55	0.633333	0.472222
病院経営	0.5	0.74697	0.761905	0.708333	0.535714	0.583333	0.55	0.333333	0.5	0.5
郵政民営化	0.338068	0.5	0.75	0.928571	0.9	0.833333	0.666667	0	1	1
サマータイム	0.271635	0.695913	0.791667	0.75	0.75	1	1	1	1	
	0.351789	0.574866	0.646703	0.661071	0.63631	0.663532	0.662879	0.505556	0.744444	0.74

20

【 0 1 4 0 】

【表 2 3】

	1位	2位	3位	4位	5位	6位	7位	8位	9位	10位
成果主義	0.510965	0.713375	0.695079	0.6675	0.7005	0.617958	0.778333	0.8175	0.685	0.73
エスカレーター	0.223977	0.427955	0.477993	0.452857	0.3605	0.42125	0.505	0.564611	0.617125	0.6575
病院経営	0.367744	0.703319	0.734643	0.685833	0.584643	0.75	0.7445	0.61	0.6525	0.6975
郵政民営化	0.262551	0.517706	0.632054	0.758333	0.8835	0.873333	0.85	0.66	0.77	0.88
サマータイム	0.364942	0.697955	0.84525	0.853	0.8225	0.866667	0.875	1	1	
	0.319554	0.562909	0.625398	0.648013	0.649805	0.683063	0.697377	0.675213	0.645563	0.74

30

40

また、ランダムに意見を並べた場合の精度と本システムで順位付けした場合の精度を表 2 4 に示す。表 2 4 は、上位 3 件の意見のうち、各順位までの意見を出力とみなした場合の精度である。

【 0 1 4 1 】

【表 2 4】

正解	手法	1 位	2 位	3 位
A だけ	ランダム	19%	28%	19%
	本システム	38%	32%	23%
A と B	ランダム	81%	83%	75%
	本システム	87%	87%	83%

図 1 4 より、A だけを正解とした場合と、A と B の両方を正解とした場合によらず、上位の意見ほど高い精度が得られた。また、表 2 4 より、ランダムに意見を並べた場合よりも、本システムによって順位付けした場合の方が高い精度が得られた。しかし、再現率は、A だけを正解とした場合と、A と B の両方を正解とした場合でほぼ同じ曲線となった。すなわち、論点 X を含んでいる意見内における A の分布と、B の分布はほぼ同じであることがわかった。A がより上位に分布していることが理想であるため、本手法にはまだ改善の余地がある。

10

【0 1 4 2】

表 2 5 は、各テーマごとに、それぞれの論点を含んでいる意見数の平均と、A と判定された意見の平均順位を示している。それぞれの論点を含んでいる意見数とは、判定者が各テーマから抽出した論点群を含んでいる意見数を平均した値である。例えば、「成果主義」の場合は、論点 X を含んでいる意見数の平均が 6 . 8 2 件であり、そのうち人間が A と判断した意見が平均 3 . 0 4 位であったことを表している。

20

【0 1 4 3】

【表 2 5】

テーマ	含んでいる意見数(中央値)	A の順位
有事法制	6.82 (3.41)	3.04
成果主義	7.88 (3.94)	4.55
エスカレーター	4.44 (2.22)	1.97
病院経営	4.72 (2.36)	2.10
郵政民営化	3.23 (1.62)	2.33
サマータイム	5.43 (2.72)	3.27

30

表 2 5 において、「含んでいる意見数」の括弧内にある数は、含んでいる意見数の中央値を表している。この数字は、本手法による意見の重要さの順位付けを行わず、無作為に出力した結果の平均である。全テーマに対する A の意見の平均順位は 3 . 2 7 であり、無作為の順位である 2 . 7 2 を下回ってしまった。しかし、テーマごとに見ると、5 件中 3 件のテーマ（成果主義、病院経営、郵政民営化）では A の平均順位が高くなった。このことから、本手法が有効に働く場合とそうではない場合があることがわかった。今後は、評価対象のテーマ数を増やして、さらなる分析を行う必要がある。

40

【0 1 4 4】

以上詳しく説明したように、本発明に係る意思決定支援システムは、テキストマイニング手法を有効活用し、入力された所与のテーマに関連する意見テキスト群を収集し、この意見テキスト群の中から前記テーマに関連する複数の論点を抽出してその論点毎の肯定又は否定等の立場を有する意見の割合を表す固有度と、その論点毎の立場を含んでいる意見の重要度を算出することにより、当該テーマに関連する多くの世人の意見に基づいての複数の論点を参照することにより合理的且つ妥当性のある決定支援情報を提供することを可能にしたのである。

【0 1 4 5】

また、本意思決定支援システムは、最新の時事問題に係る論点を含むテーマについて特

50

に有効であり、最新の意見テキストを収集するので収集する情報内容が固定化又は陳腐化することがなく、当該テーマに関連する多くの人の意見のトレンドに応じた決定支援のための判断情報の提供が可能である。

【産業上の利用可能性】

【0146】

本発明は、本発明は、人間が所与のテーマについて意思決定する際に関連する有用な情報を収集し提示して人間の意思決定を支援するコンピュータ及びデータベースを使用した意思決定支援システムに関するものであって、産業上の利用可能性を有する。

【図面の簡単な説明】

【0147】

【図1】本発明の意思決定支援システム10の構成機能ブロック図を示す。

【図2】本意思決定支援システムにおいて、収集された意見テキスト群の中から主観情報を抽出してから、意思決定を支援する情報画面を出力に至るまでのフローを模式的に表す。

【図3】論点分布の表示例を示す。

【図4】関連語の表示例を示す。

【図5】代表的な意見の表示例を示す。

【図6】抽出された論点の例を示す。

【図7】形態素解析の例を示す。

【図8】係り受け解析の例を示す。

【図9】多段階評価のレビューを入力した実行例を示す。

【図10】判定者が抽出した論点の例を示す。

【図11】論点抽出の再現率 - 精度グラフの例を示す。

【図12】判定者Aが抽出した論点の例を示す。

【図13】判定者Bが抽出した論点の例を示す。

【図14】全テーマに対する精度と再現率の例を示す。

【符号の説明】

【0148】

10：本意思決定支援システム

11：入力手段

12：論点抽出手段

13：固定度算出手段

14：重要度算出手段

15：関連語抽出手段

16：代表意見選択手段

17：インタフェース手段

18：データベースサーバ

19：ウェブ検索サーバ

20：ゲートウェイサーバ

21：ファイヤーウォール

22：記憶装置(1)

23：記憶装置(2)

24：記憶装置(3)

25：WWWインターネット

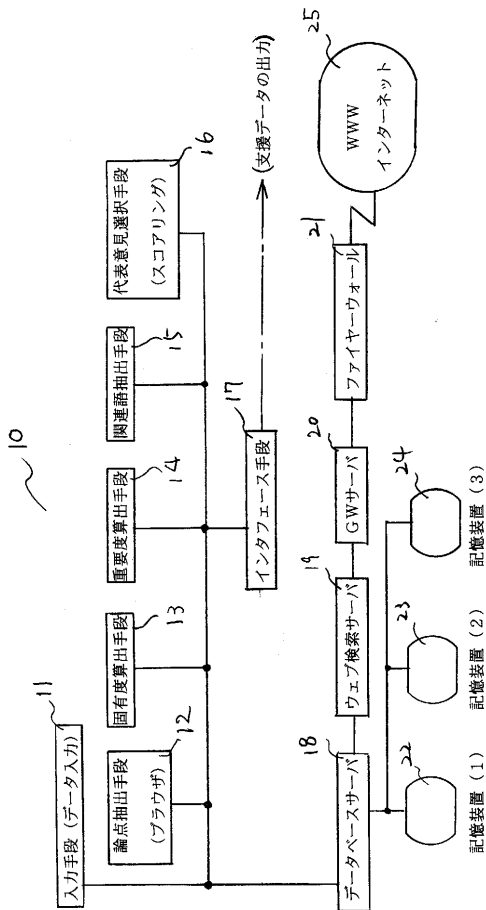
10

20

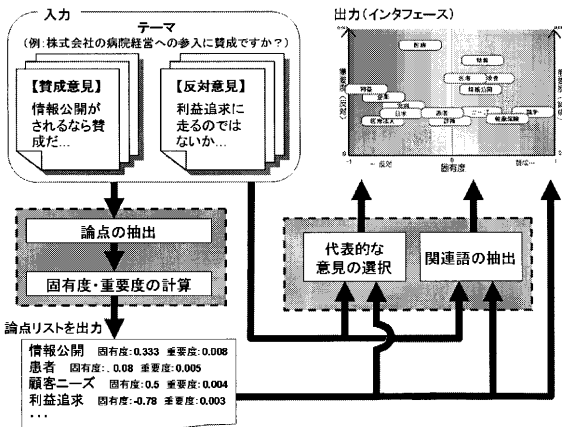
30

40

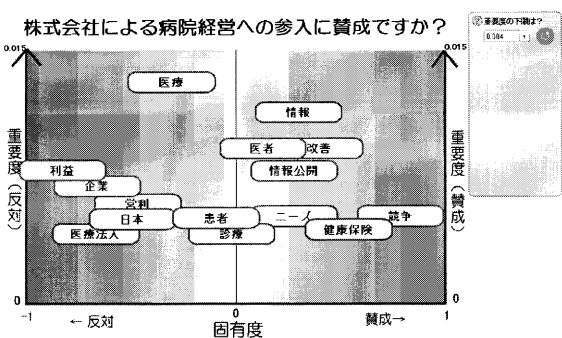
【図1】



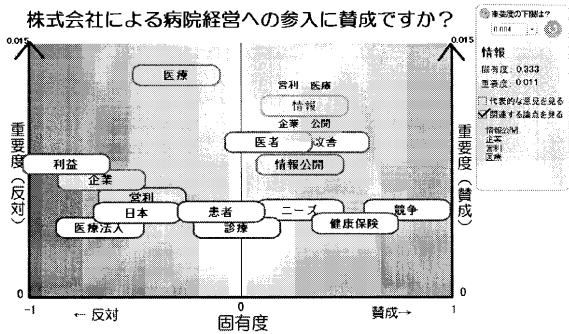
【図2】



【図3】



【図4】



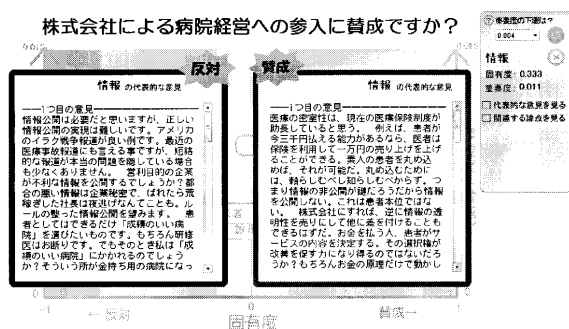
【図6】

意見
医療は利益追求になじまない。株式会社は利益を追求する。営利をあまり考えない、無駄を省き経費削減に努力することは当然、ですむ現在の世界に誇る制度を崩壊させたら、不公平な医療制度になることは明らかだ。

抽出された名詞句
医療、利益、株式会社、営利、無駄、経費、削減、努力、世界、崩壊、不公平、利益追求、経費削減、誇る制度、医療制度

抽出された動詞句
医療なじむ、利益追求なじむ、営利考える、無駄省く、世界誇る、医療制度なる

【図5】



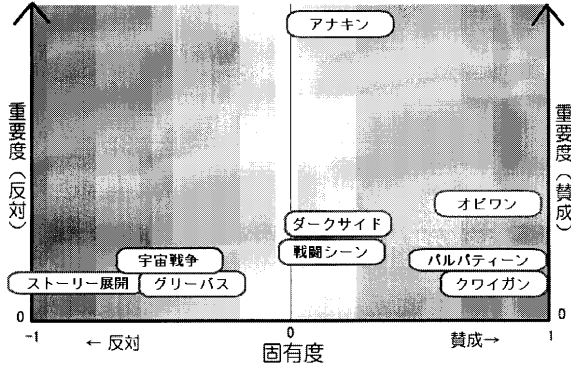
【図7】

形態素	読み	原形	品詞
情報	ジョウホウ	情報	名詞-一般
を	ヲ	を	助詞-格助詞-一般
公開	コウカイ	公開	名詞-サ変接続
する	スル	する	動詞-自立/サ変-スル 基本形
こと	コト	こと	名詞-非自立-一般
は	ハ	は	助詞-係助詞
重要	ジュウヨウ	重要	名詞-形容動詞語幹
だ	ダ	だ	助動詞 特殊-ダ 基本形

【図 8】

* 0 2D 0/1 1.92372506			
医療	イリョウ	医療	名詞-一般
は	ハ	は	助詞-係助詞
* 1 2D 1/2 0.00000000			
利益	リエキ	利益	名詞-一般
追求	ツイキュウ	追求	名詞-サ変接続
に	ニ	に	助詞-格助詞-一般
* 2 -1O 0/1 0.00000000			
なじま	ナジマ	なじむ	動詞-自立 五段・マ行 未然形
ない	ナイ	ない	助動詞 特殊・ナイ 基本形

【図 9】



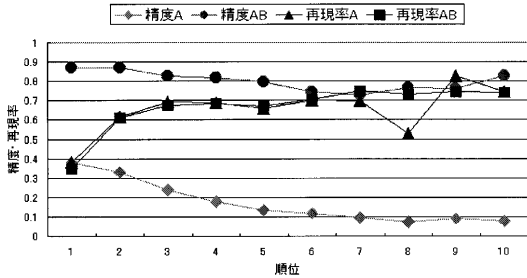
【図 10】

株式会社の病院経営と聞くと、どうしても人の命が「商品化」されてしまうような違和感を感じます。確かに、医療サービスの選択肢が増え、病院にとっての資金調達も多様化するとは思いますが、はたしてそれで本当にいい医療が育っていくのでしょうか？人の命の重さは、平等に守られるのでしょうか？小さな子どもを持つ母親として、とても不安になります。

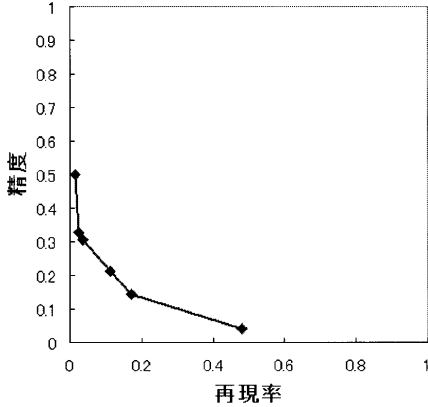
【図 13】

日本の病院は、医師だけでなく看護師までもが、薬剤の量を間違えても報告もしないし、患者の容態がよほど急変しないと引継ぎレポートに書かない状態です。この前の東京女子医大の特集も実は、このモラルの低さが問題なのです。国立大学の附属病院も指導するドクターが皆無の状態です。こんな医師に誰が命を預けられますか？しかし救急で運ばれたら、否応無しに医療ミスの洗札を受けざるを得ません。株式会社がそんなことをすればたちまちマスコミに暴露され、存続が危ぶまれます。ですから、株式会社で医療をやっていくには、余程の知識、経験、技術が要求され、それが信用に繋がっていくのです。アメリカとの比較は議論にもならないくらい日本は遅れていますが、情報公開が出来る体制に少しでも近づけるよう努力し、高いモラルを持つことが重要です。

【図 14】



【図 11】



【図 12】

日本の病院は、医師だけでなく看護師までもが、薬剤の量を間違えても報告もしないし、患者の容態がよほど急変しないと引継ぎレポートに書かない状態です。この前の東京女子医大の特集も実は、このモラルの低さが問題なのです。国立大学の附属病院も指導するドクターが皆無の状態です。こんな医師に誰が命を預けられますか？しかし救急で運ばれたら、否応無しに医療ミスの洗札を受けざるを得ません。株式会社がそんなことをすればたちまちマスコミに暴露され、存続が危ぶまれます。ですから、株式会社で医療をやっていくには、余程の知識、経験、技術が要求され、それが信用に繋がっていくのです。アメリカとの比較は議論にもならないくらい日本は遅れていますが、情報公開が出来る体制に少しでも近づけるよう努力し、高いモラルを持つことが重要です。

フロントページの続き

審査官 岩間 直純

(56)参考文献 特開2004-227343(JP,A)

岡野原 大輔、辻井潤一，評価文に対する二極指標の自動付与，言語処理学会第11回年次大会発表論文集，日本，2005年 3月14日，第664-667頁

乾 裕子、村田 真樹、内元 清貴、井佐原 均，表層表現に着目した自由回答アンケートの意図に基づく自動分類，自然言語処理，日本，言語処理学会，2003年 4月10日，第10巻 第2号，第19-42頁

(58)調査した分野(Int.Cl.，DB名)

G06F 17/30

G06F 19/00

G06Q 50/00

G06F 13/00