

テキストマイニングシステム STM の感情・意思分析機能の組み込みを中心とした機能拡充

石田涼[†] 山西和広[†] 早川紘代^{††}
酒匂佳織^{††} 後藤良太^{††} 原田実^{††}

今日のテキストマイニングツールは、限られた同義語辞書による語の同義性を元にテキスト分類を行っているので、多様な表現を持つ意見の同義性を判定できず、その限界が問われている。この解決のため、原田研究室では 2006 年度より、意味解析に基づいて深層的な文の同義性による分類を可能とするテキストマイニングシステム STM の開発を行っている。本研究では、STM に、文節の主辞と副主辞の同一性によって文節を分類する主題分析機能、分類された文節と係り受け関係にある文節を分類する主題関連分析機能、他の属性毎の頻度分析を行うクロス分析機能を追加する。さらに、意見に現れる感情を 10 分類、意志を 8 分類し、個々の意見がどの感情や意志を持つかを分析する機能を追加した。

Enhancement of text mining system STM centering on feelings and intention analysis

RYO ISHIDA[†] KAZUHIRO YAMANISHI[†]
HIROYO HAYAKAWA^{††} KAORI SAKOU^{††}
RYOUTA GOTO^{††} MINORU HARADA^{††}

Because today's text mining tool classifies the text based on the synonymy of the word by the limited synonym dictionary, the similarity of the opinion with various expressions cannot be judged, and the limit is asked. Text mining system STM which enables classification by the similarity of deep meaning of a sentence based on the semantic analysis has been developed in the Harada laboratory since 2006. In this research, the subject analysis function to classify clauses by the identity of the head word and sub-head word of the clause, the subject relation analysis function to classify the clauses which depend on a clause classified by the subject analysis, and the cross analysis function to

analyze the frequency depending on other attributes are added to STM. In addition, we classified feelings and intension which appeared in the opinion into 10 and 8 classes, respectively, and added the function to analyze the feelings and intension of each individual opinion.

1. 序論

1.1 研究背景

近年、人々の生活はコンピュータに依存するようになり、個人が Web 上に情報を配信する機会が多くなっている。また、高速ネットワーク技術と安価な大容量記憶装置の発達も伴い、膨大なテキストデータが Web や企業に蓄積されるようになってきた。テキストデータとは、アンケートの回答やコンタクトセンターの問い合わせなどの顧客の意見が入ったデータである。数値データや属性データは、データマイニングという手法で幅広く活用されているが、企業が蓄積しているデータは一般に、数値データと属性データが 2 割、テキストデータが 8 割と言われている。このため近年、多くの企業がテキストデータの重要性に気付き、蓄積されたテキストデータから有益な情報を取り出して経営に活かそうとしているが、大量のテキストデータを人手で分類・分析を行うには、多くの時間と人材が必要になる。こうしたテキストを分析することで、社会においてどのような出来事が問題になっているかを把握することができる。このような構造化されていないテキストデータを分析し、有益な情報を得るのがテキストマイニングである。しかし、今日のテキストマイニングでは、語の表層的な同一性あるいは限られた同義語辞書による語の同義性を元にマイニングの基礎である分類を行っているので、多様な表現を持つ意見の同義性を判定できずその限界が問われている。この解決のため、原田研究室では 2006 年度より意味解析に基づいて深層的な文の同義性による分類を可能とするテキストマイニングシステム STM[16][17]の開発を行っている。

1.2 研究目的

従来の STM では SAGE[18][19][20][33][34]による意味解析機能を用いて、アンケート文を意味グラフに展開し、2 つの意味グラフの対応する節同士の概念的な類似度や節間の深層格の類似度をベースに、類似部分グラフの大きさで 2 文の類似度を計測している[15][27]。これにより、表現が異なっても同様な趣旨をもつ文を同意見として集約し分類できる。一方、最近、意見に現れる感情や意志を元にした分類の信憑性が問われている。そこで、本年度我々は STM の分析機能の向上を目指し、感情・意志分析を中心とした新たな分析機能の追加を行うことにした。

[†] 青山学院大学大学院 理工学研究科 理工学専攻 知能情報コース

^{††} 青山学院大学 理工学部 情報テクノロジー学科

2. 感情辞書作成

我々は感情分析を行うにあたり、まず感情カテゴリーの分類と感情辞書の作成を行った。

2.1 感情カテゴリ

感情カテゴリーの分類は、本研究では他の研究で比較的多く用いられていた感情表現辞典[13]を参考することとし、こちらで採用されている基本感情 10 種類{ 喜・怒・哀・怖・恥・好・厭・昂・安・驚 }を感情カテゴリとして表 1 のように分類することとした。

表 1 感情カテゴリ

感情 カテ ゴリ	喜(よろこび)	例:喜ぶ、わくわく、晴れやか
	怒(いかり)	例:怒る、腹立たしい、憤る
	哀(かなしみ)	例:悲しい、傷付く、嘆く
	怖(きょうふ)	例:気味悪い、怖い、悲鳴
	恥(はじ)	例:恥ずかしい、照れる、こそばゆい
	好(すき)	例:友情、慕う、愛する
	厭(いや)	例:厭がる、むかつく、不快
	昂(たかぶり)	例:焦る、気が急ぐ、やきもき
	安(やすらぎ)	例:ほっと、安心、気楽
	驚(おどろき)	例:驚き、ショック、思いも寄らず

2.2 感情辞書構築

感情の分析を行うにあたって、感情語を抽出した感情辞書の構築を行う。今回作成する感情辞書は、(1)最も基本的な単語での感情表現を収めた感情語辞書、(2)「腹が立つ」などの個々の単語のみでは感情を表さないが複数の単語の組み合わせによって感情を表す表現を集めた慣用句感情辞書、(3)ブログやレビュー等に頻出するインターネット特有の感情表現である顔文字を集めた顔文字辞書、の三点である。

2.2.1 感情語辞書

国語辞典では、一般的に見出しの言葉をより平易な言葉で説明した語意説明が記述されている。つまり、基本的な感情表現(悲しい、笑う、怒りなど)を種表現として、語意説明中にその種表現を含む見出しを新たに加えていくことで、最終的には複雑な感情表現も比較的容易に抽出する事が出来る。そのため本研究では、感情語辞書の構築に、①感情表現辞典を用いて基本的な語句(種表現)を収録し、②それらを元に EDR

辞書の語意を用いて、より網羅的な辞書の構築を目指す。感情表現辞典は感情表現単体のみを記載してある語句編を使用し、その中でも単語それ単体で感情を表すものだけを手作業で抜き出して種表現とする。具体的には下記の手順で行う。

- ① 抽出した種表現を、感情語辞書の形式 {EDR 辞書見出し、読み、EDR 大分類品詞、EDR 詳細品詞、不変化部(活用語幹)、概念 ID、感情カテゴリ(10 感情)、用例、語意} に変換する。変換した種表現を人手で確認し、正しく解析・分類されていることを確認後、辞書登録する。
- ② EDR 単語辞書の各語の各語意の語意説明の中に、種表現の単語を含む単語について①と同様の形式で感情語意を抽出する。

以上の手順を行い、重複を避けるため抽出した言葉の中で同一の概念 ID をもち、なおかつ同一の感情カテゴリを有するものを除外して感情語辞書とした。結果として、種表現としては感情表現辞典の語句編 2337 語を元に 1513 語を抽出し、種表現をもとに EDR 辞書から抽出した語意と合わせて、合計 2083 語意を感情語意辞書として登録した。

2.2.2 慣用句辞書

本研究では、慣用句感情辞書の構築には、のべ 17998 個の慣用句とその意味が掲載されているくろご式慣用句辞典[31]を用いた。そこから慣用句、読み、感情カテゴリ、キー、意味、類義語、反対語、例文を抽出した。本研究で必要なのは感情表現を含む慣用句だけなので、抽出された各慣用句の意味から感情表現を含むかどうかを人手で判断し、含むと判断された慣用句のみを{慣用句見出し、読み、感情カテゴリ、キー、意味、類義語、反対語、例文、意味グラフ}の形式でのべ 3119 個登録し、慣用感情句辞書を構築した。

2.2.3 顔文字辞書(EDR 辞書への追加)

顔文字を用いた感情分析は SAGE による解析で得られた概念 ID を用いるので、SAGE で用いている概念見出し辞書に 10 感情分の概念 ID(語意)を新たに登録した。その後、よく使われている感情を表す顔文字や記号をインターネットのブログ・レビューサイトの書き込みから人手で探し、合計 214 個の顔文字と記号を SAGE で用いられている日本語単語辞書に登録した。

3. テキストマイニングシステムの概要

3.1 システム概要

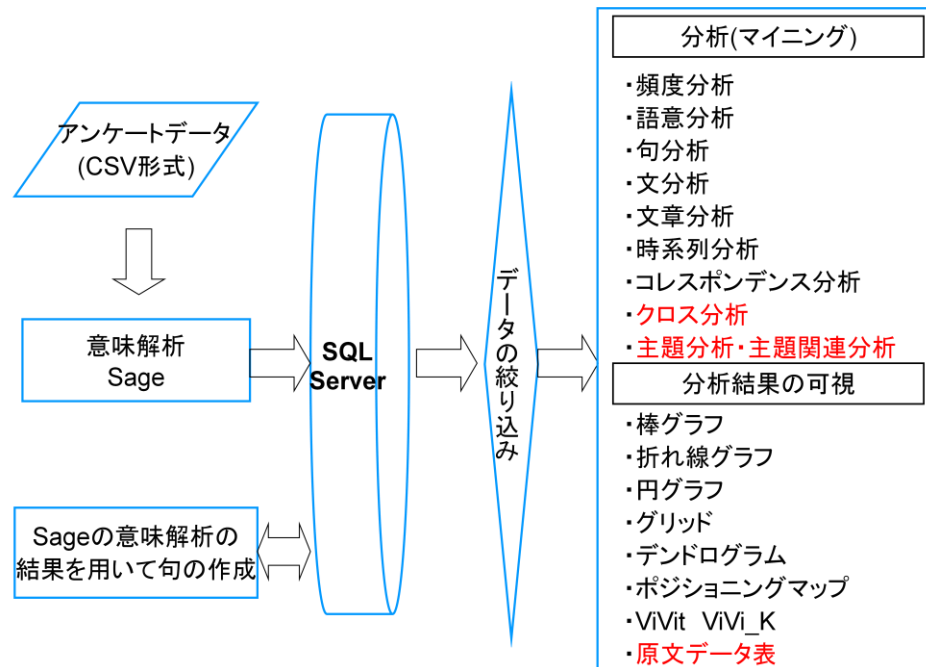


図 1 STM のシステム構成

STM では分析を行う前に、図 1 に示すようにアンケートデータをもとにデータベースを作成する。データベースの作成では、初めにアンケートデータの自由記述部分に対して SAGE を用いた意味解析を行い、その解析結果と分割された解析結果 (形態素や文節など) をデータベースに格納する。そして、データベースに保存された解析結果を用いて句の作成を行い、作成された句をデータベースに格納する。分析を始める際には、まず属性による分析対象の絞り込みを行い、必要なデータをデータベースから取り出す。取り出されたデータを文節、句、文レベルで個々の要素間の類似性を計算し (Metis) [6]、この類似度を基にそれらをクラスタリング (AQUA) し、結果を可視化する。

STM2008 年度までに作成されていた機能としては、意見中に現れる語を表記の同一性でクラスタリングし出現回数をカウントする頻度分析、意見中に現れる句、文、文章を意味的な類似性でクラスタリングする句分析、文分析、文章分析がある。さらに、

これらの句や文のクラスタリング結果を、それらの意見の発生時刻で分割して各時刻における各クラスターの頻度を時系列表示する時系列分析と、離散値を持つ各属性値毎に各クラスターに属する要素数をカウントしたクロス集計表を元に類似した意見同士を近くに配置するコレスポネンス分析の機能も開発されていた。そこで、これらに加え今年度は、文節の主辞と副主辞の同一性によって文節を分類し、さらに、分類された文節ごとにこれと係り受け関係にある文節を分類する主題分析・主題関連分析機能を開発する。さらに、これらの句や文のクラスタリング結果を、それらの意見の他の離散値属性ごとに各クラスターの頻度を表示するクロス分析と感情・意志分析の機能も追加する。

3.2 アンケート形式

入力となる CSV ファイル形式のアンケートデータの例を図 2 に示す。各行のデータの内容は下の通りである。

SampleID AN	回答日時 AD	性別 AS	年齢 AN	年代別 AS	イラク攻撃について日本はどう対応すべきですか QF
1	2006/7/8 0:00:02	女性	28	02.20代	アメリカへの協力は、集団的自衛権の行使であり、憲法に反します。
2	2006/7/8 0:00:02	女性	26	02.20代	平和的解決をゆずすべきである。
3	2006/7/8 0:00:02	女性	40	04.40代	国連決議があっても憲法に反せば、日本は協力できない。
4	2006/7/8 0:00:01	男性	51	05.50代	国連による査察の継続強化が唯一の平和的解決の道である。
5	2006/7/8 0:00:02	女性	67	06.60代	憲法9条をもつ国として、戦争への協力はきっぱり拒否し、平和的解決に向けた外交努力に徹するべきと考えます。
6	2006/7/8 0:00:02	女性	69	06.60代	アメリカのイラク攻撃はあらゆる手立てを以て回避しなければなりません。
7	2006/7/8 0:00:01	男性	56	05.50代	日本がとるべき行動は、アメリカに対し、無法なイラク攻撃を止めるべきであるということを断固として働きかけることです。
8	2006/7/8 0:00:01	男性	60	06.60代	憲法上、軍事行動に対する協力はできないと思います。
9	2006/7/8 0:00:02	女性	42	04.40代	日本は良心的兵役拒否国家であることを、世界に向けて発信すべきです。
10	2006/7/8 0:00:02	女性	50	05.50代	武力攻撃に協力すべきではない。
11	2006/7/8 0:00:02	女性	27	02.20代	憲法で武力による国際紛争解決を否定している限り、武力による解決には、協力すべきでない。
12	2006/7/8 0:00:01	男性	63	06.60代	武力行使を容認する新たな国連決議に至らない努力こそ必要。
13	2006/7/8 0:00:02	女性	52	05.50代	平和的手段で解決をするべき。
14	2006/7/8 0:00:02	女性	62	06.60代	アメリカ・ブッシュ政権のイラク攻撃姿勢は、石油利権を主目的にした明白な侵略行為である。
15	2006/7/8 0:00:02	女性	25	02.20代	イラクの大量破壊兵器の破壊は、査察によって達成されるべき。
16	2006/7/8 0:00:02	女性	45	04.40代	国連を中心に国際社会との関係を構築していくためには新決議があれば協力もやむをえない。
17	2006/7/8 0:00:01	男性	29	02.20代	日本は国連主軸外交に徹すべきであり、国連決議を精査し憲法の許容する範囲で行動するべきだ。
18	2006/7/8 0:00:02	女性	60	06.60代	日本国憲法の国際協調主義原則に則って行動するべき。
19	2006/7/8 0:00:02	女性	43	04.40代	大量破壊兵器があるのなら問題であるが、まずは日本国憲法を考えるべきである。
20	2006/7/8 0:00:01	男性	29	02.20代	攻撃をした場合、一般国民が被害を受けることを忘れてはならない。
21	2006/7/8 0:00:02	女性	22	02.20代	承認決議がある場合は、国際社会の総意、意思と理解せざるを得ない。
22	2006/7/8 0:00:02	女性	57	05.50代	国連中心主義を掲げるわが国としては、応分の協力をすべき。
23	2006/7/8 0:00:01	男性	63	06.60代	日米関係、北朝鮮問題などを考えた時、国連中心主義を貫くことが、国益上もっとも望ましい現実的判断と考える。
24	2006/7/8 0:00:01	男性	46	04.40代	協力は、日本国憲法の枠内の平和的なものに限定される。
25	2006/7/8 0:00:01	男性	41	04.40代	国連の加盟国としての責務として、憲法で認められた範囲内で応分の協力は行うべき。
26	2006/7/8 0:00:02	女性	65	06.60代	協力をせざるをえないと思います。
27	2006/7/8 0:00:01	男性	63	06.60代	武力攻撃は回避すべきだし、協力すべきでない。
28	2006/7/8 0:00:02	女性	50	05.50代	国連決議の内容にもよるが、国際社会が一致するなら協力もやむをえない。
29	2006/7/8 0:00:02	女性	17	01.10代	日本は軍隊を持たず、一人前の独立国家として意見を述べよ。
30	2006/7/8 0:00:01	男性	61	06.60代	攻撃容認決議がある場合でも、武力行使には関係せず人道的協力、復興支援に限るべき。
31	2006/7/8 0:00:02	女性	69	06.60代	攻撃を承認する国連決議があれば国連決議に基づき国連に協力をすべき。
32	2006/7/8 0:00:01	男性	29	02.20代	人道的支援、特に復興支援に積極的に協力することに限定するべき。
33	2006/7/8 0:00:02	女性	36	03.30代	査察が続いている限りは、新しい決議は必要ない。
34	2006/7/8 0:00:01	男性	23	02.20代	できれば攻撃できるための決議を急がず、査察を強化継続すべき。
35	2006/7/8 0:00:02	女性	46	04.40代	仮に新たな決議がされても、支持はしても、資金、自衛隊派遣ではなく、復興に協力すべき。

図 2 アンケートデータの例

- 1 行目：属性名
 - 2 行目：データ種別
- 4 行目以降に入力されるデータの形式を表す。

AN 数値属性

回答者の年齢などの数値データ。

AI 投稿者属性

意味解析を行う必要のない文字列。

QF 自由記述質問

自由記述形式のテキストデータであって意味解析の対象になる。

AD 日時属性

投稿日時などの日時データ。形式は下記の例に限定する。

形式：YYYY/MM/DD HH:MM

AS 選択属性

選択式の回答データ。

形式：選択肢番号.選択肢内容

QC チェック質問

チェックボックスを用いた質問に対する回答データ

形式：0 または 1

- 3 行目：回答条件指定

質問に対して回答を行うための条件を表す。条件のもととなる質問は選択形式の質問のみとする。

形式：質問番号=選択肢番号

- 4 行目以降：回答データ

回答者の属性や質問に対する回答の実際のデータが入力されている。

4. 主題・主題関連分析

4.1 主題・主題関連分析とは

主題分析では絞り込んだ分析対象の文節の主辞と副主辞をデータベースの文節テーブルから抽出し主辞と副主辞を合わせたものをキーとして品詞ごとにクラスタリングしランキングを表示する。これによりアンケート全体でどのような話題が良く出ているのかを大まかに知ることができる。

主題関連分析では、主題分析で出力された主題の一つを選択し、その主題にどのような別の主題がどの程度の頻度で係り受け関係にあるかを分析する。これによりある主題に対する傾向やその主題に対する評価などを詳しく見ることができる。

ただし、主題関連分析では対象が文節である必要があるため、質問が選択形式、チェックボックス形式であるものは分析することができない。

4.2 分析の流れ

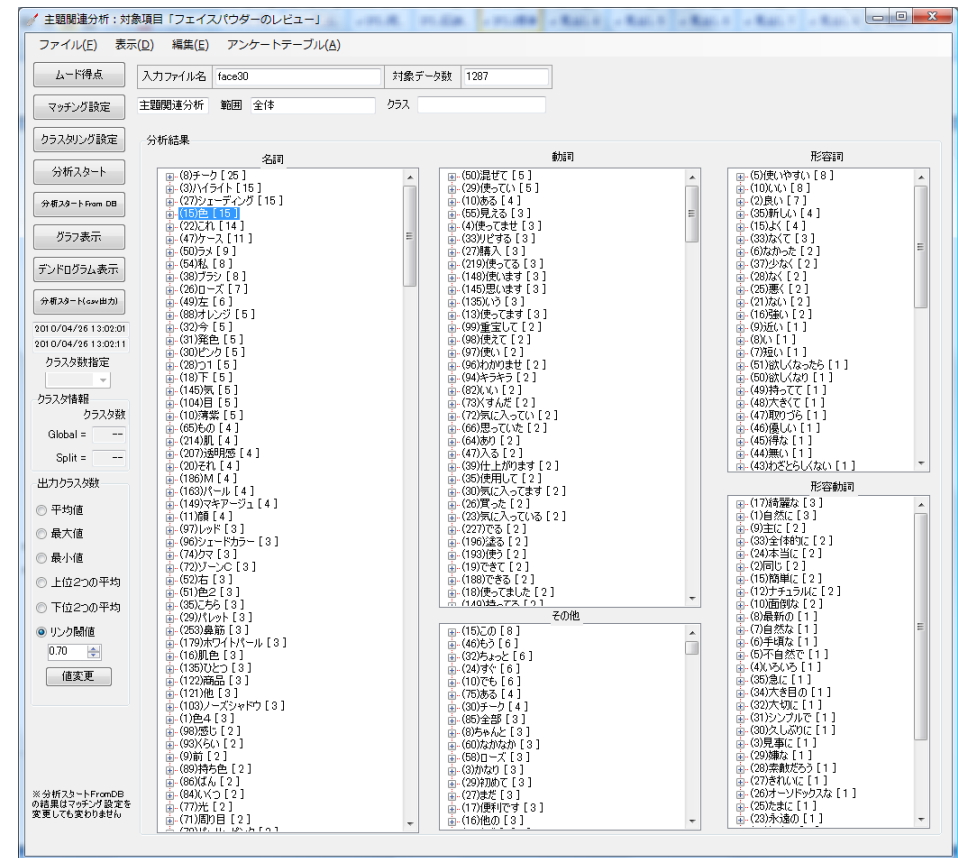


図 3 主題分析画面

具体的には、主題分析では分析スタートをすることで図3のような文節を名詞、動詞、形容詞、形容動詞、その他の品詞に分けてクラスタリングを行った結果をツリー表示する。結果はクラスタの要素数が多い順に表示される。さらにここからある主題を一つ選択し右クリックを行い主題関連分析を行うことで図4のような分析画面が表示され選択した主題に対してどのような話題が関連しているかを調べ、その頻度をランキング表示、グラフ化を行う。

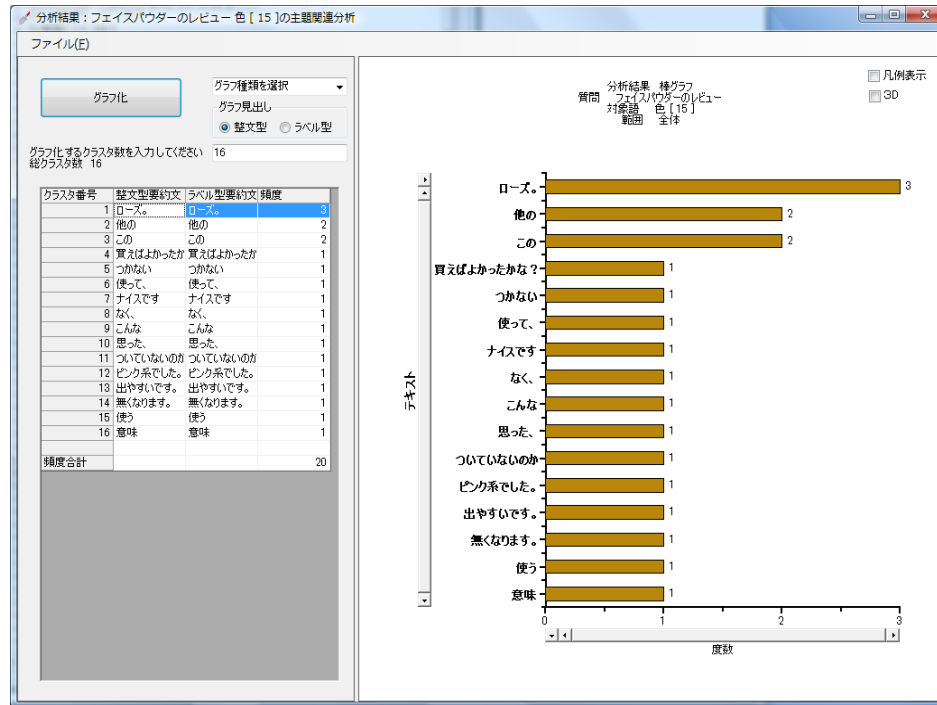


図 4 主題関連分析画面

5. クロス分析

5.1 クロス分析とは

クロス分析とは複数の属性の全組み合わせについて個別にデータを集計した表であるクロス集計表を作り、そこからクロス集計表を元に、グラフ表示するものである。属性別の意見の違いを調査するときに使う。

5.2 分析の流れ

STM のクロス分析では、まず語意、句、文、文章のいずれかのクラスタリング結果からクラスタの要約文と要素数を得る。次に、アンケートの属性リストから選択された属性と各クラスタの要素数を回答者の属性値毎に分類したものをクロス集計表としてまとめ、そのデータをもとに、棒グラフを作成する。

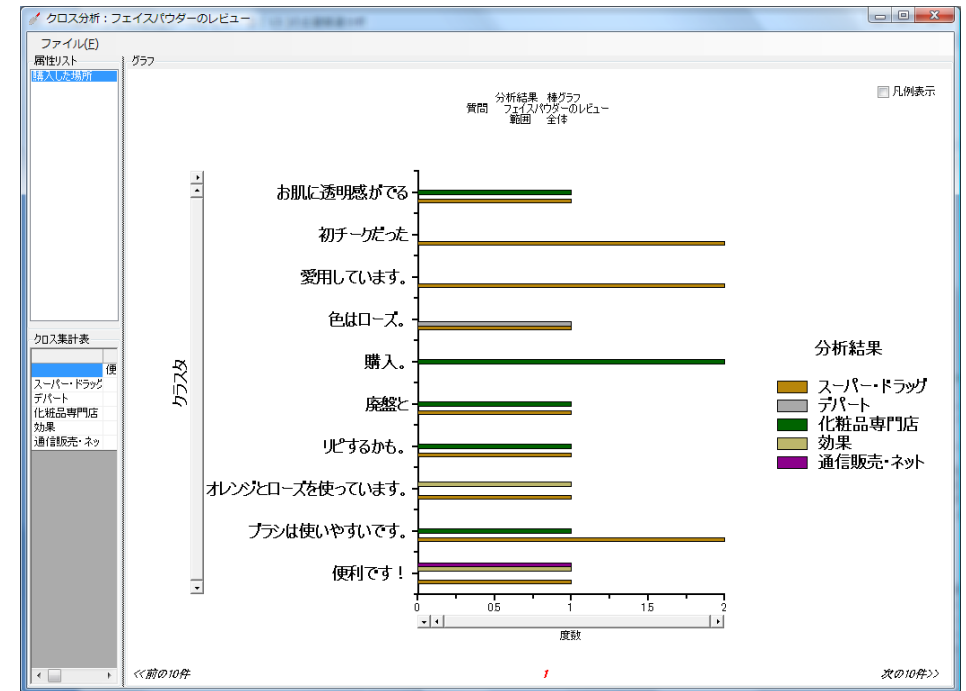


図 5 クロス分析画面

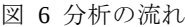
図 5 の左上の属性リストから属性を選択することで、選択した属性と意見とのクロス集計表が左下に作成される。このクロス集計表の行項目は回答者の属性値、列項目は意見（クラスタの要約文）である。回答者の属性値は、属性リストで選択された項目により分類され、それに応じて各クラスタの要素数も各行に振り分けられる。最後に、これを元にグラフ化がおこなわれる。

6. 感情・意志分析

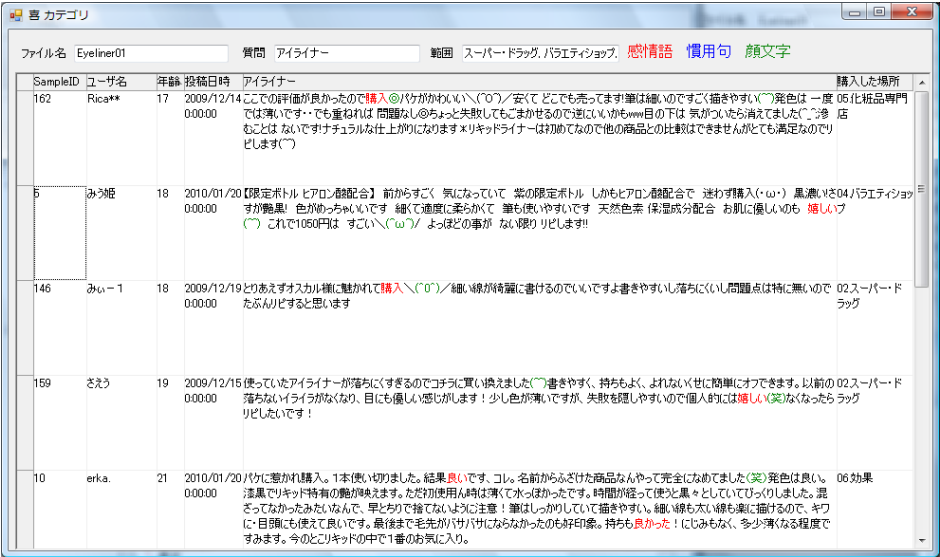
感情・意志分析機能では、文で表現されたアンケート者の意見を SAGE 解析し、文節ごとに感情語辞書、慣用句辞書、顔文字辞書を照合して 辞書に登録された語意が含まれていれば、その語意に割り当てられた感情・意志カテゴリをその文節に付与する。本研究では、SAGE 解析によって得られるモダリティ [3][4] を人手で分類し、意志カテ

表 2 意志カテゴリー

実際の感情・意志分析では図 6 のように、まず分析の前処理として属性を元に分析対象の絞り込みを行い、分析対象のデータを作成し、選ばれた分析方法に従い実際に分析を行う。



STIM を起動し、アンケートデータを読み込んだあと感情分析を行うと図 7 のような画面が表示される。左のカテゴリから分析したい感情・意志カテゴリを選択し分析スタートボタンを押すと、選択された感情・意志に対して分析結果の欄に、ツリー形式でそれぞれの感情・意志を持つ文節を含む文の一覧が表示される。



グラフ表示ボタンを押すと図9のような感情カテゴリごとにそれらのカテゴリを含む文節数の出現頻度をカウントしグラフとして表示できる。これによって視覚的に、どのような感情が多く見られるかがわかる。円グラフなどにすることにより割合を表示することもできる。また感情だけでなく意志カテゴリについてもグラフ化することが可能である。

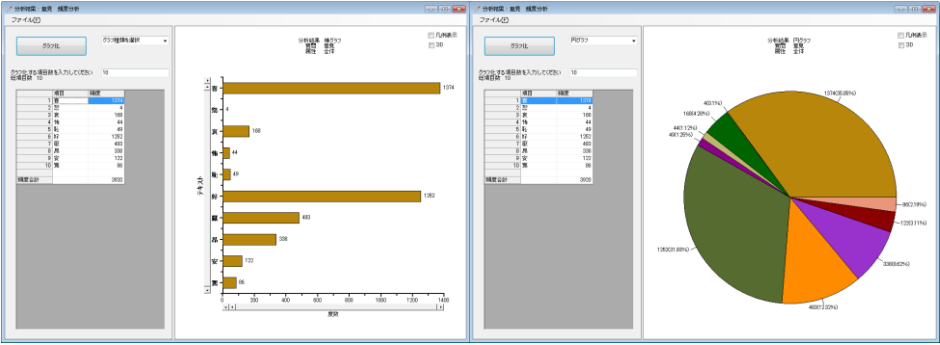


図 9 グラフ表示画面

ここで、分析結果内のツリーのノード（感情・意志カテゴリ上）を右クリックし、原文表示を選択することで図 8 のような画面が現れ、選択されているカテゴリが含まれている原文を表示させる。この画面では感情の起因となったテキストを、感情語（赤）、慣用句（青）、顔文字（緑）色で表示させることができ、どの文節に感情・意志カテゴリが付与されたかがわかる。

参考文献

- [1] 新田義彦：正規表現とテキスト・マイニング，岩波書店（2009）
- [2] 上田隆穂，黒岩祥太，戸谷圭子，豊田裕貴：テキストマイニングによるマーケティング，講談社サイエンティフィク（2006）
- [3] 梅澤俊之，西尾華織，松田源立，原田実："意味解析システム SAGE の精度向上とモダリティの付与と辞書更新支援系の開発"，言語処理学会第 14 回年次大会発表論文集，E3-1，pp. 548-551(2008).
- [4] 梅澤俊之，加藤大知，松田源立，原田実："意味解析システム SAGE の精度向上ーモダリティと副詞節についてー"，情報処理学会 第 191 回自然言語処理研究会，pp. 1-8(2009).
- [5] 大塚裕子，乾孝司，奥村学：意見分析エンジンー計算言語学と社会学の接点ー，コロナ社（2007）
- [6] 加藤裕平，古川勇人，蒲生健輝，韓東力，原田実：WEB検索による知識文の獲得と意味グラフ照合推論による質問応答システム Metis，情報処理学会第 67 回全国大会論文集，1G-06，第 2 分冊 pp.11-12 (2005.3)
- [7] 金井進，堀宣男，神田晴彦，三室克哉，鈴村賢治："顧客の声" 分析・活用術，リックテレコム（2008）
- [8] 神嶋敏弘：データマイニング分野のクラスタリング手法(1)ークラスタリングを使ってみよう！ー，人工知能学会誌，Vol.18, no.1, pp.59-65 (2003).
- [9] 喜田昌樹：テキストマイニング入門，白桃書房（2008）
- [10] 金明哲：テキストデータの統計科学入門，岩波書店（2009）
- [11] 久保田裕章，平塚飛将，吉川ひかる，松田源立，原田実："質問応答システム Metis の回答精度向上-検索フェーズの改良を中心として-"，言語処理学会第 14 回年次大会発表論文集，A5-5，pp. 1017-1020 (2008.3).
- [12] 竹原一彰，安部建助，安田智成，韓東力，原田実："質問応答のための質問文と知識文の間の意味ベースでの精密な照合方式"，情報処理学会第 66 回全国大会論文集，6U-03,第 2 分冊，pp.173-174 (2004.3).
- [13] 中村明："感情表現辞典"，東京堂出版（1993）.
- [14] 那須川哲哉：テキストマイニングを使う技術／作る技術，東京電機大学出版局（2008）
- [15] 西岡晋太郎，久保田裕章，坂東晃文，原田実："意味グラフ照合による質問応答システム Metis の回答精度向上 - 質問文解析フェーズと検索フェーズの改良を中心として -"，情報処理学会研究報告，Vol.2009-NL-191 No.16, pp. 1-8, (2009.5).
- [16] 西脇剛，保立哲志，原田実："意味解析に基づくテキストマイニングシステム STM"情報処理学会第 69 回全国大会論文集，2C-03・第 2 分冊 pp. 89-90. (2007.3).
- [17] 西脇 剛，保立哲志，原田実："意味解析に基づくテキストマイニングシステム STM"情報処理学会第 69 回全国大会論文集，2C-03,第 2 分冊 pp. 89-90. (2007.3).
- [18] 原田実，尾見孝一郎，岩田隆志，水野高宏："日本語文章からの意味フレーム自動生成システム SAGE(Semantic frame Automatic GEnerator). の開発研究"，人工知能学会第 13 回全国大会論文集，pp. 213-216(1999).
- [19] 原田実，水野高宏："EDR を用いた日本語意味解析システム SAGE ".人工知能学会論文誌，16(1)，pp.85-93(2001).
- [20] 原田実，田淵和幸，大野博之："日本語意味解析システム SAGE の高速化・高精度化とコーパスによる精度評価"，情報処理学会論文誌，Vol.43, No.9, pp.2894-2902(2002).
- [21] Minoru Harada, Yuhei Kato, Kazuaki Takehara, Masatsuna Kawamata, Kazunori Sugimura, and Junichi Kawaguchi: "QA System Metis Based on Semantic Graph Matching ",Proc. of the 6th International Conference on NII Test Collection for IR Systems(NTCIR6), Tokyo, Japan, pp.448-459, (2007.5).
- [22] 福原知宏，中川裕志，西田豊明："感情表現と用語のクラスタリングを用いた時系列テキスト集合からの話題検出"，人工知能学会全国大会 2E1-2 (2006)
- [23] 松村真宏，三浦麻子：人文・社会科学のためのテキストマイニング，誠信書房（2009）
- [24] 松本和幸，湊純子，土屋誠司，任福継："日英対訳感情表現コーパスに基づく感情表現抽出手法の提案"，自然言語処理研究会報告 pp.69-75 (2008).
- [25] 松本和幸："会話文からの話者感情推定に関する研究"，博士論文（2008）.
- [26] 三室克哉，鈴村賢治，神田晴彦：顧客の声マネジメント，オーム社（2007）
- [27] 村上裕人：自由記述アンケート文の自動分類システム AQUA の分類精度向上，青山学院大学大学院理工学研究科修士論文（2005）.
- [28] 村田真樹，小木しのぶ，高山泰博，末吉正成，今村誠，洲上美喜：事例で学ぶテキストマイニング，共立出版（2008）
- [29] 安村禎明，坂野大作，上原邦昭："評判情報のレベルを考慮した評価文書の分類と評価情報の信頼性評価への応用"，研究報告「自然言語処理 (NL)」(2007).
- [30] 山崎秀夫，松田潤：顧客を創造するテキストマイニング，日本工業新聞社（2003）
- [31] くろご式慣用句辞典：<http://www.geocities.jp/tomomi965/index2.html>
- [32] @コスメ：<http://www.cosme.net/>
- [33] 京都大学情報学研究科知能情報学専攻メディア講座言語メディア研究室(黒橋研究室)，<http://nlp.kuee.kyoto-u.ac.jp/>
- [34] (株)日本語電子辞書研究所：EDR 電子化辞書仕様説明書(第 2 版)，(株)日本語電子辞書研究所(2002).