

人の価値観を表すカテゴリを対象にした 複数カテゴリへの自動分類の試み

石田 栄美*・An-Shou Cheng**・Douglas W. Oard**
Kenneth R. Fleischmann**

【要旨】 内容分析の自動化を試みた。ネットの中立に関する公聴会での 28 証言に対し、人の価値観を表すカテゴリ Schwartz Values Inventory を付与したテストコレクションを作成した。これを用いて自動分類実験を行った。テストコレクションには複数のカテゴリが付与されているデータも含まれていたため、分類器のための学習用データ表現方法 3 種を検討した。分類結果として最終的にいくつかのカテゴリを選択するかという問題には、正解と同数を選択する方法と閾値を用いて選択するカテゴリ数を推定する方法を実験した。学習用データの表現方法は、複数のカテゴリが付与されていた場合に各カテゴリの学習用データとして用いる複製手法の F 値が 0.324 であり最も有効であった。カテゴリ数の決定には閾値を用いた方法の F 値が 0.303 であり、正解と同数のカテゴリを選択した場合の F 値 0.324 と比べて、ほぼ同じ性能が得られた。

【キーワード】 自動分類・内容分析・価値観・Schwartz Values Inventory

1. はじめに

1.1 内容分析自動化の試み

近年、情報技術 (IT) 分野の発展はめざましく、様々な技術が開発され、それらを用いた製品が広く普及している。IT の普及は、技術開発が支えているだけでなく、政治的、社会的背景などの要素も、影響を及ぼしていると考えられる。その社会的背景には、人々の IT に関する関心、企業や人が IT に関して持っている価値観なども反映されているといえる。PopIT¹⁾ プロジェクトでは、新聞記事や雑誌記事などを対象に、IT に関する用語の出現傾向や IT 用語と企業や人名などの関係性を分析することによって、IT 分野の構造や現状を把握することを目指している。本研究は、このプロジェクトの中で IT における政策と価値観との相互作用に着目し分析を進めている。たとえば、2006 年と 2008 年に米

国の公聴会で行われたネットの中立に関する議論の証言をもとに、それぞれのステークホルダーが持つ価値観を内容分析により抽出した²⁾。その結果、ネットの中立に関する議論は時間が経つにつれて変化してきていることがわかった。ネットの中立に関する議論が始まった 2006 年では、ネットの中立を支持している人たちは“平等”という価値観を持ち、反対している人たちは“自分たちの目標を選ぶこと”という価値観を持っていた。2008 年になると、支持者たちはコンテンツプロバイダーやネットの利用者が求める内容やサービスが制限されずに提供されるような“自由”が重要だと考え、支持しないものはサービスプロバイダーが有益なものであり社会的に公正であること、何らかの利益を得られることなど“豊かさ”が重要だと考えていた。

このように、内容分析は、発言や文書から人がどのような価値観をもっているか、世の中の価値観が

* 駿河台大学メディア情報学部

** University of Maryland, USA

どのように変化しているかなどを調べるための重要な研究手法の一つである³⁾といえる。

内容分析は、問題の認識、テキストの選択、概念化、実際のコーディングなど、様々な段階がある。その中で、問題の認識、概念化、テキストの選択、コーディング結果の分析などは質的な作業が必要であり、人手によって行わなければ難しいものである。しかしながら、コーディングを、同じカテゴリを用いて異なるテキストに繰り返し行う場合や大量のテキストを対象にしたい場合には、これを自動的に行うことができる可能性がある。本研究では、人の価値観というカテゴリを対象に、コーディングを自動的に行うことを提案しており⁴⁾、本稿ではその自動化の実験に焦点をあてた。コーディングが自動的に行えるようになれば、ニュースやブログ、スピーチなどさまざまなタイプの大量のテキストに適用することができ、異なるタイプのテキスト間での価値観を比較したり、価値観の変化を長期的に観察したりすることが可能になる。

本稿では、これまで人手によって行われてきたコーディングを自動的に行うことを目的とする。人が行ったコーディング結果を、コンピュータを用いてどこまで再現できるかに着目し、様々な実験を行った。人がどのような価値観を持っているかを発言や文章から分析することを大きな目的としているため、対象となるカテゴリには人の価値観を表すカテゴリセットを用いた。

1.2 自動分類実験の概要

コーディングは、対象とするテキストに、あらかじめ設定されたカテゴリの中から、そのテキストが表す適切なカテゴリを付与していく行為であるといえる。このような観点からみると、コーディングは、テキストにカテゴリを付与するテキスト自動分類（カテゴリライゼーション）と同じ手法を適用することができる。本研究では、テキストに対してカテゴリを付与するコーディングを、テキストをカテゴリに分類するという自動分類の問題として捉えた。

テキストの自動分類研究は、主題カテゴリを対象

にしたものが多かったが、近年になって、情報源のオーソリティ、意見分析、リーディングレベルなど非主題のカテゴリを対象にした分類が注目を集めている。人の価値観を表すカテゴリを対象にした分類も、非主題のカテゴリを対象にした自動分類の範囲を広げたものにとらえることができるであろう。

本研究で行った実験の概要を図1に示した。自動分類実験は、まず、学習用データで分類器を学習させ、次に、分類対象となるテキスト（評価用データ）を用いて実際に分類器が分類を行い、分類器が出力した分類先が、どの程度正確であるかを評価するという流れで行われる。学習用データと評価用データは各テキストにあらかじめカテゴリが付与されているものを用いるが、これをテストコレクションと呼ぶ。本研究では、テキストに対して内容分析を行い、テストコレクションを作成した。自動分類は、どのような分類器を用いるかによって、性能に大きな影響を与えるが、すでに数多くの性能の高い分類器が提案されており、本実験ではそれらの分類器を用いることにした。本研究で注目したのは、ひとつのテキストに対して複数のカテゴリが付与されている場合の自動分類である。複数カテゴリへの自動分類の場合、主に2つの問題がある。一つは、分類器を効果的に学習させるための学習用データの表現方法であり、もう一つは、分類器が出力したカテゴリの中からいくつのカテゴリを最終的な分類結果とするかという分類先の決定手法である。本実験では、はじめに、テストコレクションを作成し、次に、自動分類実験として、学習用データの表現方法3種、分類先の決定方法にかかわる2種の実験をした。

以下では、2章においてテキストの自動分類に関する研究について述べる。3章では、テストコレクションの作成について述べる。4章では、実験で用いた手法について述べ、実験結果は5章に示す。6章では、実験結果の個々の事例について検討し、7章でまとめを行う。

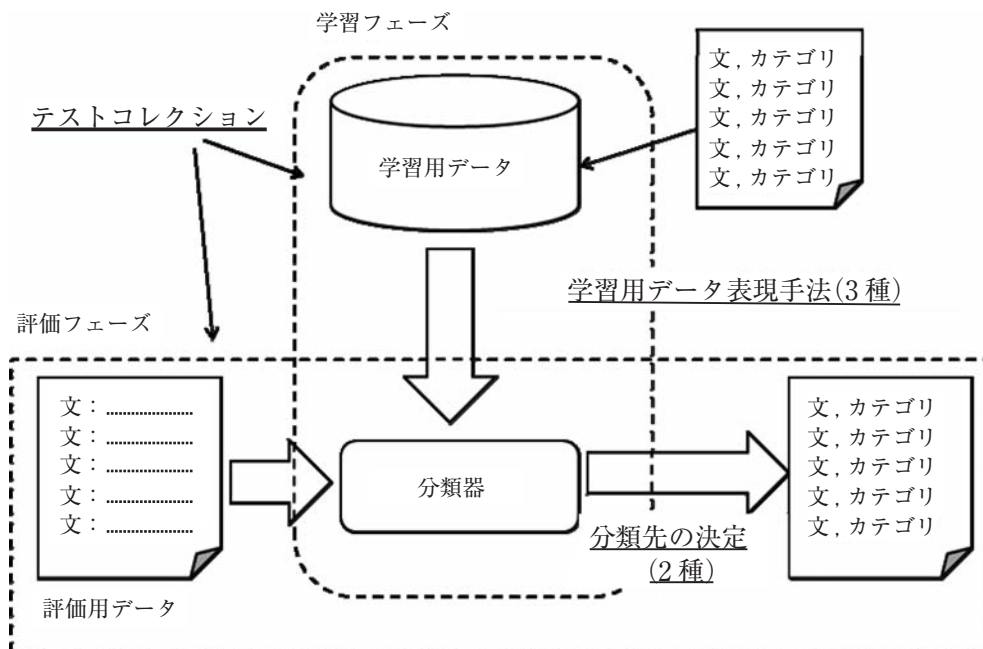


図1 自動分類実験の概要

2. テキスト自動分類研究

テキストの自動分類研究の初期では、人が作成したルールを自動的に判別できるようにし分類するというアプローチがとられていたが、近年の自動分類研究では機械学習手法を用いたものになってきている。機械学習を用いた自動分類研究では、あらかじめカテゴリに分類されたデータを用いて分類器を学習することが必要である。近年になって、大規模なテキストが利用可能になり分類器を学習するための十分なデータ量が得られるようになったこと、コンピュータの性能向上により複雑な計算ができるようになったことなどにより注目されるようになった。

Breeseらによれば、分類器には、主にモデルベースとメモリベースと呼ばれる二つのアプローチがある⁵⁾。Support Vector Machine (SVM) は、モデルベースの中で最も性能が高いといわれている手法である⁶⁾。SVMの基本的なアイデアは、正のサンプルと負のサンプルと呼ばれる2つの値を特徴ベクト

ルで表現し、それらを平面上に配置したときに、この2つの値を持つサンプルが最もよく分かれる境界を分類器に学習させることである。SVMはもともと2つの値を分類するのに適した分類器であるが、最近では、複数の値を持つ事例にも適用できるようになっている。メモリベースの分類器として、よく知られているのは、k近傍法(k-Nearest Neighbor、以下ではkNNとする)である⁷⁾。kNNの基本的なアイデアは、ベクトル空間上に、それぞれの学習用データのベクトルとカテゴリを記憶しておき、分類対象データのベクトルが入力されたときに、それに近い学習用データサンプルのカテゴリを、その分類対象データの候補のカテゴリとして投票するというものである。最終的には投票結果によって、最も多く投票されたカテゴリがその分類対象サンプルのカテゴリになる。いくつの学習用データサンプルを用いるかは、kの値によって指定する。投票の時点で、すべて同じ重みで投票することも可能であるが、類似度などを用いた重みづけを用いた投票のほうが、性能が高い場合が多いといわれてい

る。

特徴ベクトルに用いる特徴には、語の出現回数を用いるなど語の特性を用いる場合と、文の長さなど語の特性を用いない場合がある。本研究では、語の特性を用いた。特徴ベクトルはそれぞれの語の重みで表現されることが多いが、カテゴリの特徴をより表現する語に高い重みが付与されることが多い。語の重みづけ手法としては、TF・IDFが最も用いられている手法である。これは、テキスト中に該当する語が出現すればするほど (TF)、かつ、その語が他のテキストに出現しなければしないほど (IDF)、重みが高くなる重みづけ手法である。本研究では、プレ実験の際に、TF・IDFを用いた場合とテキスト中に出現するかしないか (1, 0) の2値を用いた場合を比較したところ、性能が同じであったため、本実験では2値 (1, 0) の重みを用いることにした。

SVMもkNNの両分類器とも、単独のカテゴリしか付与されていないテキストを用いた自動分類をベースに設計されている。本研究では、複数カテゴリへの分類を行う。これまでに複数カテゴリのための分類手法はいくつか提案されている⁸⁾⁹⁾¹⁰⁾が、本研究では、分類器そのものは既存のものを用いて、学習用データの表現手法、最終的な分類カテゴリの決定手法に着目した。実際に用いた手法は、4章で詳しく述べる。

3. テストコレクション

本章では、実験に用いるために新しく作成したテストコレクションについて述べる。

3.1 コーパス

本研究では価値観と政策の相互作用に着目しているため、公聴会における発言を用いた。これらの発言からは、ステークホルダーの価値観やそれを反映した公的な立場、政策などをとらえることができる。

テキストには、2006年2月7日に米国上院商務・科学・運輸委員会によっておこなわれた“ネットの

中立性 (net neutrality)”に関する公聴会¹¹⁾と2008年4月17日連邦通信委員会が開催したブロードバンドネットワークに関する公聴会¹²⁾における証言を用いた。各委員会のウェブサイトでは、これらの証言を文字に起こしたものが提供されており、ダウンロードすることができる。2つの公聴会から28の証言を得ることができた。

3.2 カテゴリとコーディング

分類対象のカテゴリとしたのは、人の価値観を表すカテゴリセットであるSchwartz Values Inventory (以下、SVIとする)を用いた。これは三階層構造になっており、第一階層 (“value dimensions”)には4つのカテゴリがあり、第二階層 (“value types”)の10のカテゴリがある。第三階層は56の “basic human values”があり、“Social Power”、“Successful”、“Equality”、“Politeness”、“Social Order”などのカテゴリが含まれている。

SVIは、文化、言語、地理、宗教、人種などの分野の分析に広く適用されている¹³⁾¹⁴⁾¹⁵⁾。たとえば、心理学分野における行動と価値観との関係¹³⁾¹⁴⁾や価値観と所属政党の関係の調査¹⁴⁾¹⁵⁾などの例があり、公聴会での発言に対しても適用できると考えた。

28の証言に対して、第二著者 (第一コーダー) が、文単位で、コーディングを行った。価値観を表している表現には、複数の文で一つの価値観を表している場合や文の一部だけに価値観が表れている場合もあった。本来は、価値観が表れている部分だけを対象にすることが望ましいが、本研究では文単位でコーディングを行った。文単位という制限をなくすことは、今後の課題である。

コーディングにはATLAS.tiを利用した。これは、社会科学の分野で手作業による内容分析をサポートするために広く用いられているソフトウェアである。コーディングの結果は、ATLAS.tiからxml形式で出力した。さらに、それらのファイルから、本実験で用いた分類器ソフトウェアであるWeka¹⁶⁾に入力するために、文とカテゴリを組み合わせたフォーマットに整形した。

28 証言の全てに含まれる文 2,294 に対し、コーディングを行った結果、2,005 文に 3,162 の SVI カテゴリが付与された。文には、カテゴリが付与されていないものもあれば、複数カテゴリが付与されているものもあった。文に対し最大で 7 カテゴリが付与されており、付与されているカテゴリ数の平均は 1.58、中央値は 1 であった。実験では、何らかのカテゴリが付与されていた 2,005 の文を、テストコレクションとして用いた。

実際に付与されたカテゴリは、第一階層では 4 カテゴリ、第二階層では 10 カテゴリ、第三階層では 46 カテゴリ（付与されなかったカテゴリもあった）であった。階層構造を明確にするために、第一階層は A、B、C、D とラベル付けし、第二階層は第一階層にあわせて A0、A1、A2 のようにラベル付けした。第三階層は、もともとのカテゴリ名に加えて第二階層のラベルをカテゴリ名の前に追加した。

図 2 に、テストコレクション全体で 20 回以上出現しているカテゴリとその回数を示した。表 1 には、第二階層レベルでのカテゴリの出現回数を示し

た。なお、A2 と B0 にそれぞれ Hedonism と同じ名前のカテゴリがあるが、第三階層のカテゴリは異なるため、別に扱っている。20 回以上出現しているのは 23 カテゴリであり、出現している 46 カテゴリの半数である。また、出現回数が多い上位 5 カテゴリ（A0-Wealth、B2-Freedom、A1-Capable、A1-

表 1 第二階層レベルでのカテゴリの出現回数

第一階層カテゴリ	第二階層カテゴリ	出現回数
Self-enhancement (A)	Power (A0)	817
	Achievement (A1)	745
	Hedonism (A2)	1
Openness to change (B)	Hedonism (B0)	2
	Stimulation (B1)	104
	Self-direction (B2)	537
Self-transcendence (C)	Universalism (C0)	438
	Benevolence (C1)	266
Conservation (D)	Tradition (D0)	33
	Conformity (D1)	78
	Security (D2)	141

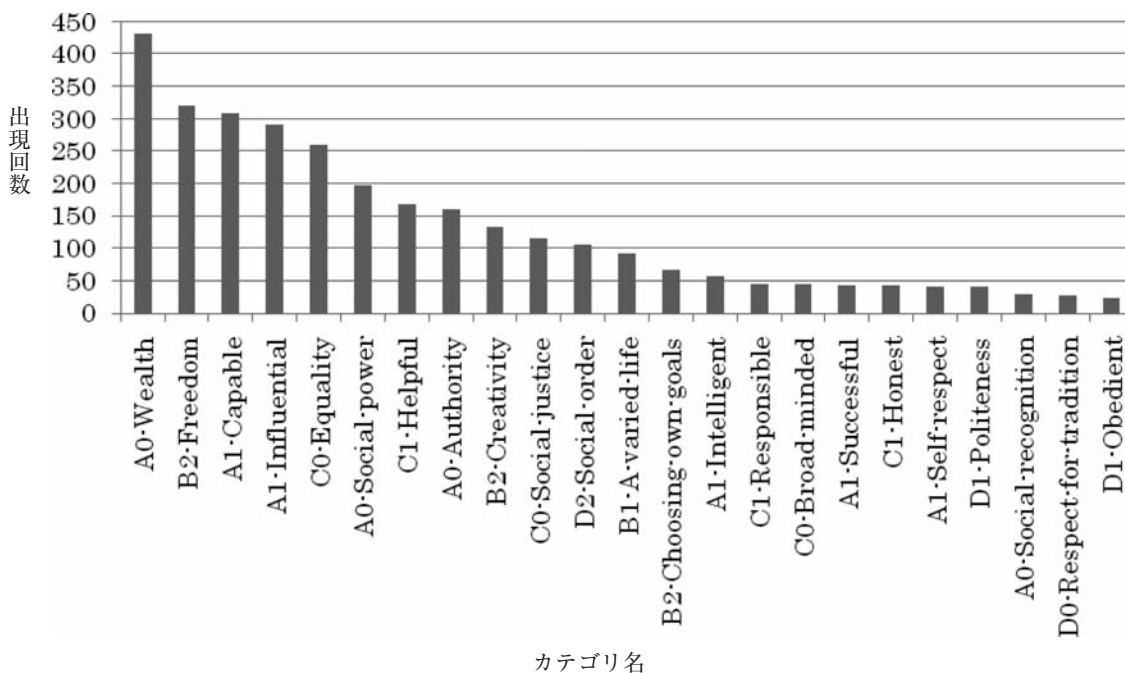


図 2 20 回以上出現する SVI カテゴリ

Influential、C0-Equality) の出現頻度を合計すると全体の50%を占めていた。このことから、カテゴリによって出現に偏りのあるデータであることがわかる。

作成したテストコレクションから All、Pruned10、Pruned20 という3つのデータセットを作成した。Allにはカテゴリが付与されたすべての文が含まれている。Pruned10は、Allの中で少なくとも10回以上出現しているカテゴリが付与されている文のみを用いている。同様に、Pruned20は Allの中で20回以上出現するカテゴリが対象である。表2に、それぞれのデータセットにおける、各階層でのカテゴリ数およびそれらが付与されていた文の数を示した。

表2 データセット

データセット名	カテゴリ数			文の数
	第一	第二	第三	
All	4	11	46	2,005
Pruned10	4	9	29	1,987
Pruned20	4	9	23	1,959

3.3 コーダー間の一致率

テストコレクション中から4つの証言を選択し、第二コーダー（第四著者）が内容分析を行い、第一コーダーの結果と比較した。同じ文に対し、一人はカテゴリを付与していないが、もう一方は一つ以上のカテゴリを付与している例や、どちらのコーダーもカテゴリを付与しているが、付与されたカテゴリがそれぞれ異なる例などがあった。

第一コーダーの結果を正解とみなし、第二コーダーとの一致率をみたところ、第二コーダーの再現率は0.420、精度は0.359、F値は0.387であった。さらに、コーダー間の一致率を示す尺度である Cohen's Kappa 値を求めた。Cohen's Kappa 値は、一つの文に対して、それぞれのコーダーが付与したカテゴリが、どの程度、一致していたかをみるための指標であり、カテゴリ全体の一致率を計算する。本研究では、コーダーが一つの文に対して付与したカ

テゴリ数が異なるため、Cohen's Kappa 値をそのまま用いることはできない。そこで、すべての文を対象に、カテゴリが付与されていたか否かを基準に、カテゴリごとの Kappa 値を求めた。この場合、ある文に対して、両コーダーともカテゴリを付与している、もしくは両方とも付与していない場合を一致したとみなし、どちらか一方が付与していれば不一致とみなした。コーパスの中で両コーダーが2回以上付与した17カテゴリを対象に、以上のような方法でカテゴリごとの Cohen's Kappa 値を求め、平均したところ0.383であった。これは、Cohen's Kappa 値の基準からいえば適正 (fair agreement) であるが、決して一致率が高いとは言えない。本実験で試みようとしているタスクは、人手で行ったとしてもその結果を一致させることは難しいことを示している。

このようにコーダー間による一致の問題があるので、本来は複数のコーダーによる結果を利用することが望ましいが、現時点では一人分の結果しか入手できないため、本実験では一人分のデータのみを使った。

4. 複数カテゴリへの自動分類

本実験では、先に述べたように、分類器は研究対象とせず、ひとつのデータ（本研究では、文にあたる）に複数のカテゴリが付与されている場合に、それらを学習用データとしてどのように表現するかという学習用データの表現方法の問題と、分類器から出力された分類先の候補となるカテゴリの中から最終的にいくつのカテゴリを選択するかという分類先決定手法について、検討した。以下では、それぞれで実際に用いた手法について述べる。

4.1 学習用データの表現方法

学習用データとはカテゴリとデータのセットである。分類器は、データ中に出現する語の出現回数などをもとに、カテゴリと語の関係を学習し、分類対象となるデータが入力されたときに、そこに含まれ

る語をもとに最も適切なカテゴリに分類する。学習用データはカテゴリと語の関係を表す重要な情報になる。一つのデータに一つのカテゴリが付与されている場合には、該当するカテゴリの学習用データとしてそのまま用いればよい。しかしながら、本テストコレクションは、表3に示したように一つのデータに複数カテゴリが付与されているものも含まれている。この場合、学習用データの表現方法としては、それぞれのデータを付与されているすべてのカテゴリの学習用データとして用いる方法や、一つのカテゴリのみの学習用データとして用いる方法など、いくつかの表現方法が考えられる。どのような方法が有効であるかは、対象とするカテゴリや、学習用データの量、複数カテゴリが付与されているデータの状況などによって異なるため、それぞれの事例においてどの手法が適切であるかを検討する必要がある。Tsoumakas と Katakis¹⁷⁾ は、複数ラベルが付与されている学習用データの表現方法として5つの手法を提案しており、本実験ではこの中から3つの手法を実験した。

表3 複数カテゴリが付与されているデータの例

カテゴリ	文
A1-Influential, B1-A-varied-life	As a result high speed access to the Internet is revolutionizing the way we work learn seek medical advice gather our news engage in public discourse interface with government socialize and almost every aspect of the way we live.
A0-Wealth, A1-Capable, B2-Freedom	To compete in the world we need a simple inexpensive and open network not a costly complex and balkanized one.
C1-Helpful, D2-Healthy	A Veterans Administration study showed you could cut hospital stays in half for many patients - and yet monitor and watch over them for longer periods of time.

4.1.1 複製手法

この表現方法は、複数カテゴリが付与されている

場合でも、一つのカテゴリが付与されていた場合と同様に、データを複製し、各々のカテゴリの学習用データとするという方法である。本研究では、これを複製手法と呼ぶ。この手法は、それぞれのカテゴリが同じデータで学習するため分類器の判別能力が低くなる可能性はあるが、学習用データを簡単に作ることができるという利点がある。また、Tsoumakasらは、すべての学習用データを用いることができるため、データ量に制限がある場合には、この手法を用いることを勧めている。表3で示したデータに対し、複製手法を適用した結果を表4に示す。この例では、“As a result” から始まる文は、A1-Influential と B1-A-varied-life の学習用データとなる。

表4 複製手法を適用した学習用データの例

カテゴリ	文
A1-Influential	As a result high speed access ...
B1-A-varied-life	As a result high speed access ...
A0-Wealth	To compete in the world we ...
A1-Capable	To compete in the world we ...
B2-Freedom	To compete in the world we ...
C1-Helpful	A Veterans Administration ...
D2-Healthy	A Veterans Administration ...

4.1.2 選択手法

複製手法は、出現回数が多いカテゴリに対してより多くのデータを用いて学習することになり、少ないカテゴリに対しては十分な学習をすることができないという恐れがある。Tsoumakasらは、他の方法として、複数カテゴリが付与されている場合にその中から適切な一つのカテゴリを選択し、そのカテゴリのみの学習用データとするという方法も提案している。どのようなカテゴリを選択するかは、ランダムに選択する、アルファベット順に選択する、出現回数によって選択する方法などが考えられる。本実験では、付与されているカテゴリの中でもっとも出現回数が少ないカテゴリを選択した。この手法を用いることで、データの偏りの問題が解消されると予想する。表3の例に選択手法を適用した結果を表

5に示した。学習用データ中で A1-Influentian、B1-A-varied-life、A0-Wealth の出現を数えたところ、B1-A-varied-life の出現が少なかったため、“As a result” から始まる文は B1-A-varied-life のための学習用データになる。この手法の場合、カテゴリ間で同じデータで学習しないため、分類器の識別能力が高くなる可能性があるが、反対に学習用データが小さくなってしまうという恐れがある。

表5 選択手法を適用した学習用データの例

カテゴリ	文
B1-A-varied-life	As a result high speed access ...
A1-Capable	To compete in the world we ...
C1-Helpful	A Veterans Administration ...

4.1.3 組み合わせ手法

Tsoumakas らが提案している3つめの手法は、複数カテゴリが付与されていた場合に、その組み合わせを一つのカテゴリとみなす方法である。これを組み合わせ手法と呼ぶ。この手法は、一つの文がカテゴリの組み合わせで表現されている場合、それを忠実に学習することができるため、カテゴリとテキストの関係を最も適切に表すことができる。しかしながら、カテゴリ数が多い場合、または、その組み合わせが多い場合には、カテゴリ数が多くなってしまい、学習するためのデータが十分に確保できないということが起こる。表6に、表3で示した例に組み合わせ手法を適用した結果を示した。一つの文に対し、A1-Influentian、B1-A-varied-life、A0-Wealth が付与されていた場合、カテゴリは“A1-

表6 組み合わせ手法を適用した学習用データの例

カテゴリ	文
“A1-Influentian+B1-A-varied-life”	As a result high speed access ...
“A0-Wealth +A1-Capable+B2-Freedom”	To compete in the world we ...
“C1-Helpful+D2-Healthy”	A Veterans Administration ...

Influentian+B1-A-varied-life+A0-Wealth” となる。

3.1.4 その他の手法

その他に、Tsoumakas らは2つの手法を提案している。一つ目は、学習用データの中で、一つのカテゴリしか付与されていないデータを用いるという方法である。この手法は、的確に一つのカテゴリを表しているデータだけを用いるということができるという利点がある。しかしながら、複数のカテゴリの組み合わせで表されるような文に対する分類には向かないともいえる。また、学習用データを選択するため、学習用データが少なくなってしまう。なお、プレ実験において、複製手法と選択手法のほうが、この手法よりも高い性能を示したため、本実験では行わなかった。

残りの手法は、各々の文は、各カテゴリに対して、正のサンプル（そのカテゴリに分類されている）、負のサンプル（分類されていない）とみなし、分類器を学習するというものである。たとえば、文sがカテゴリc1、c2、c3に分類されていた場合、文sはカテゴリc1、c2、c3の正のサンプルとみなされ、他の全てのカテゴリに対しては負のサンプルとみなされる。この手法は、2値の分類タスクにおいてもっとも有効な手法として知られている Support vector machine(SVM) で用いることが適切である。本実験では、分類器として2値分類ではないkNNのみを用いているため、本実験ではこの手法は実験しなかった。

4.2 分類手法

分類対象データを一つのカテゴリだけに分類する場合は、分類器が結果として出力したカテゴリをそのまま用いる、スコアを伴って出力される場合はもっとも高いスコアのカテゴリを選択すればよい。しかしながら、複数のカテゴリが付与されている場合には、いくつかのカテゴリを最終的な分類カテゴリとして選択すればよいのかが、問題となる。

本研究では、分類対象データに対し、各カテゴリへ分類される確率を伴って結果を出力する分類器を用いて、その確率をもとにカテゴリの選択を行っ

た。具体的には、正解と同数のカテゴリを選択する方法と、閾値を用いて選択する最適なカテゴリ数を推定する方法の2つを検討した。ここでは、これらの手法を分類手法と呼ぶ。それぞれの分類手法の説明のまえに、本実験で用いた分類器について述べる。

4.2.1 用いた分類器

分類器として性能が良いと報告されているkNN ($k=5, 10, 15, 20, 25, 30, 35, 40, 45, 50$)を用いた。kNNはWeka¹⁶⁾で提供されているものを用いた。プレ実験では、ポーターのアルゴリズムを用いて語幹処理を行い、4回以下の出現回数の語は削除した場合の分類性能が高かったため、本実験でも同様の処理を行った。

Wekaに組み込まれているkNNは、サンプルデータの近さをはかる尺度として1) 同じ重み (voting, vote)、2) 逆距離加重法 ($w=1/\text{distance}$, iw)、3) 類似性加重法 ($w=1-\text{distance}$, sw) の3つの重みづけ手法を提供していたため、これら3つの重み付け手法を用いた。以下に示す実験結果では、それぞれの重み付け手法をvote, iw, swとして示す。

Wekaが提供しているkNNは、評価用データにそれぞれのカテゴリに対する確率分布を付与して分類結果を出力することができる。kNNが、ある文に対して出力した結果を、確率分布が高い順に並べた例を図3に示す。すべてのカテゴリの確率分布が示されるが、ここでは確率分布が高い上位6カテゴリを示した。この確率分布が高いほど、そのカテゴリへ分類される度合いが高いことを示している。この確率分布をもとに、以下で述べる2つの手法によるカテゴリ選択を行った。

文ID	カテゴリ	確率分布
1	A0-Wealth	0.158
	A1-Capable	0.105
	C1-Helpful	0.105
	C0-Social-justice	0.053
	B2-Choosing-own-goals	0.053
	A0-Authority	0.053

図3 kNNの確率分布伴った出力例

4.2.2 分類手法1：正解と同数カテゴリを選択

実験環境においては、評価用データに付与されている正解カテゴリ数を知ることができる。分類手法1では、この正解カテゴリとして付与されているカテゴリと同じ数のカテゴリを選択した。具体的には、評価用データsにi個のカテゴリが付与されている場合、kNNの確率分布をもとに上位i番目までのカテゴリを、sに対する最終的な分類カテゴリとした。i番目のカテゴリと同じ確率分布を持っているカテゴリがあった場合は、それらも選択した。

図3の例の場合、A0-Wealthの確率分布は0.158、A1-CapableとC1-Healthの確率分布は0.105である。もし、文ID1に2つのカテゴリが正解として付与されていた場合、上位2のカテゴリであるA0-WealthとA1-Capableが選択されるが、C1-Healthも同じ確率分布を持っているため、このカテゴリも選択される。文ID1に対する最終的なカテゴリは、A0-Wealth、A1-Capable、C1-Healthとなる。

4.2.3 分類手法2：閾値を用いる方法

実世界では、分類対象となるデータにいくつのカテゴリを付与することが適当であるかはわからない。そこで、kNNの確率分布が閾値以上であれば、最終的な分類結果のカテゴリとするという方法を用いて、カテゴリの選択を行った。図3の例では、閾値を0.05とした場合、確率分布が0.05以上をもつカテゴリが選択される。つまり、A0-WealthからB1-Daringまでが分類結果となる。同様に、閾値を0.10とした場合には、A0-Wealth、A-Capable、C1-Helpfulのみが選択される。

この手法では、選択の境界となる適切な確率分布の値を推定することが必要である。まず、テストコレクションを、学習用データ、閾値推定用データ、評価用データと3分割した。それぞれのデータの割合は、80%、10%、10%とした。閾値推定用データは、どの確率分布の値を境界とすればもっとも性能が良いかを推定するためのものである。以下の手順で、閾値の設定とその評価を行った。

- 1) 学習用データを用いて分類器を学習する

- 2) 分類器に、閾値推定用データを入力し、閾値用データの各カテゴリに対する確率分布を得る
- 3) 閾値を設定し、閾値推定用データのカテゴリを選択し、正解と結果を比較する
- 4) 閾値推定用データにおいて、もっともF値が高くなる確率分布の値を閾値とする
- 5) 評価用データを分類器に入力し、データの各カテゴリに対する確率分布を得る
- 6) 4) で得られた閾値をもとに各評価用データの分類先となるカテゴリを選択する
- 7) 評価用データの正解と分類結果を比較する

5. 結果

5.1 評価尺度

複数カテゴリへの自動分類の評価は、様々な観点からの評価尺度を用いることができるが、ここでは、分類されるべきカテゴリに分類されなかった割合、誤って分類された割合に基づいて評価するため、従来から用いられている精度、再現率、およびそれらを組み合わせたF値を用いた。

評価用データにあらかじめ付与されているカテゴリ（人がコーディングした結果）を正解とし、分類器が分類した結果のうち、どの程度、正解と一致しているかの割合を調べた。精度、再現率、F値は、それぞれ以下のように求めた。

精度(P) = 分類器が分類したカテゴリの中で正解と一致した数 / 分類器が分類したカテゴリ数

再現率(R) = 分類器が分類したカテゴリの中で正解と一致した数 / データにあらかじめ付与されているカテゴリ数

F値 = $2 \times \text{精度} \times \text{再現率} / (\text{精度} + \text{再現率})$

5.2以降で示すF値は、10交差検定した結果を平均したものである。ただし、分類手法2の閾値をカテゴリ数の選択手法では、閾値推定用データと学習

用データの間での交差検定は行っていない。

5.2 学習用データの表現方法の比較

表7に、学習用データの学習手法である複製手法と選択手法を用いて実験した場合のF値を示した。10通りのk (k=5、10、15、20、25、30、35、40、45、50) とkNNの3種の重みづけ手法 (vote、iw、sw) のすべての組み合わせで実験し、分類手法には正解カテゴリと同数のカテゴリ数を選択する分類手法1を用いた。表7において、最もF値が高かったのは、複製手法を適用した場合 (k=20、iw) の0.324であった。これは、選択手法を用いた場合に最もF値 (0.307) が高かった組み合わせ (k=40、iw) よりも高い性能を示していた。kの値と同じ重みづけ手法が同じ場合、複製手法と選択手法を比較すると、いずれの場合も複製手法のほうのF値が高かった。

同様に、表8と表9に、それぞれ第二階層、第一階層のカテゴリを基準に分類実験を行った場合のF値を示した。第二階層 (表8) において、もっともF値が高かったのは、複製手法を用いたときの0.478であった (k=15、sw)。kと重みづけ手法に同じ手法を用いた場合に、複製手法と選択手法を比較したところ、k=50の場合と、k=30とswの組み合わせ以外は、複製手法のF値が上回っていた。

第一階層 (表9) では、k=15、重み付け手法iwの場合のF値 (0.651) がもっとも高かった。第三階層と同様、いずれのk、重み付け手法の組み合わせにおいて、複製手法と選択手法を比較すると、複製手法のF値のほうが高かった。

第二階層では、いくつかのケースで選択手法のほうがF値が高いものもあったが、その差は0.001から0.002であり、第三階層、第一階層では、すべてのケースにおいて、複製手法のほうがF値が高かった。よって、学習用データにおける表現手法は、選択手法よりも複製手法のほうが有効であるといえる。

この実験では、10通りのkと3種の重み付け手法をそれぞれ組み合わせたが、特定のk値が高いという傾向や特定の重み付け手法が高いという傾向

表7 第三階層におけるF値（All、分類手法1）

	複製手法			選択手法		
	vote	iw	sw	vote	iw	sw
kNN						
5 NN	0.267	0.289	0.283	0.249	0.267	0.259
10NN	0.301	0.314	0.305	0.276	0.289	0.280
15NN	0.311	0.319	0.314	0.287	0.287	0.289
20NN	0.319	0.324	0.320	0.298	0.298	0.301
25NN	0.320	0.323	0.321	0.296	0.298	0.300
30NN	0.316	0.317	0.317	0.299	0.300	0.301
35NN	0.314	0.313	0.313	0.301	0.304	0.303
40NN	0.311	0.309	0.311	0.299	0.307	0.301
45NN	0.309	0.308	0.310	0.295	0.306	0.298
50NN	0.306	0.306	0.306	0.296	0.305	0.298

表8 第二階層におけるF値（All、分類手法1）

	複製手法			選択手法		
	vote	iw	sw	vote	iw	sw
5 NN	0.451	0.447	0.447	0.434	0.432	0.434
10NN	0.469	0.465	0.469	0.451	0.450	0.453
15NN	0.476	0.476	0.478	0.451	0.453	0.455
20NN	0.474	0.474	0.475	0.454	0.460	0.458
25NN	0.470	0.472	0.471	0.456	0.460	0.459
30NN	0.462	0.464	0.462	0.459	0.461	0.463
35NN	0.463	0.464	0.464	0.459	0.459	0.461
40NN	0.461	0.462	0.462	0.461	0.461	0.461
45NN	0.461	0.462	0.462	0.461	0.463	0.461
50NN	0.459	0.461	0.461	0.462	0.463	0.462

表9 第一階層におけるF値（All、分類手法1）

	複製手法			選択手法		
	vote	iw	sw	vote	iw	sw
5 NN	0.630	0.628	0.628	0.617	0.614	0.615
10NN	0.642	0.642	0.641	0.639	0.634	0.635
15NN	0.650	0.651	0.649	0.643	0.641	0.641
20NN	0.645	0.648	0.644	0.644	0.645	0.644
25NN	0.645	0.647	0.644	0.638	0.641	0.638
30NN	0.641	0.643	0.640	0.639	0.643	0.640
35NN	0.642	0.646	0.641	0.636	0.641	0.638
40NN	0.638	0.643	0.639	0.636	0.641	0.637
45NN	0.638	0.643	0.638	0.634	0.640	0.634
50NN	0.639	0.644	0.639	0.632	0.638	0.633

は見られず、組み合わせ方や対象とする階層によって異なっていた。また、第三階層、第二階層、第一階層とそれぞれに分けて分類を行ったが、第一階層は4カテゴリのみにも関わらず、高い分類性能が得られたわけではなかった。

組み合わせ手法は、最もF値が高かったケースで0.004（k=40、iw）であり、他の二つの手法に比べて極めて低い値となった。カテゴリごとにF値が高いものをみても、いずれも単一カテゴリであった。学習用データをみても、組み合わせた後のカテゴリ数は407にのぼり、一つのカテゴリに対して平均で4.9文の学習用データを確保することしかできず、学習に十分なデータ量を得ることができなかった。このことが性能が低い理由であると推測できる。本実験では、付与された全てのカテゴリを組み合わせると一つのカテゴリとしたためカテゴリ数が増えてしまったが、出現回数がある一定以上の組み合わせのみを対象にすれば性能があがる可能性がある。

表10にデータセットPruned10を用いた場合、表11にPruned20を用いた場合の、第三階層における複製手法、選択手法の結果を示した。表10において最もF値が高かったのは0.328の（k=20、iw）であり、複製手法を用いたときである。選択手法において、最もF値が高かったのは0.309だった。表11において最もF値が高かったのは0.332であり、複製手法を用いた場合であった（k=20、iw）。Allと同様、同じkと重み付け手法の組み合わせで、それぞれ複製手法と選択手法を比較したところ、すべてにおいて複製手法のF値が高かった。よって、他のデータセットを用いた場合にも、複製手法の性能が高いといえる。

第二階層、第一階層での実験結果は割愛するが、Pruned10を用い、第二階層でF値を比較したところ、k=25、iwのときに選択手法のほうが0.001だけF値が高かった以外は、第一階層、第二階層のすべてのケースにおいて複製手法のほうが高かった。

本実験では3種のデータセットを用いた。Allを用いた場合の最も高いF値は0.324、Pruned10では

0.328、Pruned20では0.332であった。それぞれのデータセットはカテゴリ数がそれぞれ大きく異なるにも関わらず、F値はそれほど差がない。Pruned10やPruned20のデータセットに含まれないカテゴリが、全体の分類性能に大きな影響を及ぼさないことを示している。各カテゴリの十分なデータ量の確保を目的として、3種のデータセットを作成したが、この実験においてはそれらの影響が少なく、それ以上に学習用データが必要であることを示唆している。

5.3 分類手法の比較

表12、13、14に第三階層、第二階層、第一階層

表10 第三階層におけるF値 (Pruned10、分類手法1)

	複製手法			選択手法		
	vote	iw	sw	vote	iw	sw
5 NN	0.282	0.301	0.294	0.261	0.275	0.273
10NN	0.311	0.321	0.316	0.285	0.295	0.290
15NN	0.320	0.324	0.321	0.288	0.293	0.289
20NN	0.325	0.328	0.325	0.295	0.299	0.298
25NN	0.321	0.320	0.320	0.294	0.299	0.296
30NN	0.320	0.320	0.319	0.299	0.303	0.300
35NN	0.315	0.314	0.315	0.302	0.308	0.303
40NN	0.311	0.311	0.311	0.302	0.309	0.302
45NN	0.308	0.312	0.309	0.296	0.304	0.297
50NN	0.309	0.312	0.311	0.300	0.306	0.300

表11 第三階層におけるF値 (Pruned20、分類手法1)

	複製手法			選択手法		
	vote	iw	sw	vote	iw	sw
5 NN	0.298	0.314	0.305	0.265	0.280	0.277
10NN	0.319	0.326	0.320	0.291	0.300	0.295
15NN	0.323	0.330	0.325	0.292	0.295	0.293
20NN	0.329	0.332	0.330	0.299	0.301	0.300
25NN	0.330	0.330	0.331	0.297	0.301	0.299
30NN	0.326	0.327	0.326	0.302	0.306	0.304
35NN	0.322	0.324	0.323	0.304	0.310	0.307
40NN	0.318	0.320	0.317	0.303	0.311	0.305
45NN	0.317	0.319	0.316	0.300	0.308	0.301
50NN	0.313	0.316	0.312	0.300	0.306	0.300

での分類手法1と分類手法2の結果をそれぞれ示す。学習用データの表現手法は複製手法を用いており、データセットはAllの結果である。閾値は、確率分布の値0.07から0.25まで0.01ごとに設定した。閾値は、交差検定で作成したデータセットごとに設定したため、用いられた閾値はデータセットによって異なるが、第三階層では閾値0.10や0.11、第二階層では0.17~0.19、第一階層では0.23~0.24であることが多かった。

分類手法1と分類手法2を最も高い性能を示したケースで比較すると、分類手法1のほうがいずれも性能が高かった。そのF値を比べてみると、第一

表12 第三階層におけるF値 (All、複製手法)

	分類手法1			分類手法2		
	vote	iw	sw	vote	iw	sw
5 NN	0.267	0.289	0.283	0.251	0.257	0.252
10NN	0.301	0.314	0.305	0.285	0.290	0.286
15NN	0.311	0.319	0.314	0.287	0.297	0.287
20NN	0.319	0.324	0.320	0.291	0.301	0.290
25NN	0.320	0.323	0.321	0.298	0.303	0.298
30NN	0.316	0.317	0.317	0.295	0.302	0.296
35NN	0.314	0.313	0.313	0.291	0.296	0.291
40NN	0.311	0.309	0.311	0.288	0.293	0.291
45NN	0.309	0.308	0.310	0.290	0.297	0.290
50NN	0.306	0.306	0.306	0.292	0.300	0.293

表13 第二階層におけるF値 (All、複製手法)

	分類手法1			分類手法2		
	vote	iw	sw	vote	iw	sw
5 NN	0.451	0.447	0.447	0.421	0.428	0.422
10NN	0.469	0.465	0.469	0.441	0.444	0.440
15NN	0.476	0.476	0.478	0.447	0.452	0.446
20NN	0.474	0.474	0.475	0.455	0.460	0.455
25NN	0.470	0.472	0.471	0.456	0.461	0.455
30NN	0.462	0.464	0.462	0.457	0.460	0.457
35NN	0.463	0.464	0.464	0.453	0.460	0.453
40NN	0.461	0.462	0.462	0.453	0.462	0.454
45NN	0.461	0.462	0.462	0.454	0.460	0.454
50NN	0.459	0.461	0.461	0.455	0.460	0.454

階層では0.05程度の開きがあるが、第二階層、第三階層では0.01から0.02の差におさまっている。これらの結果から、分類手法2を用いて正解カテゴリの数を推定したとしても、あらかじめ正解数を知っている場合と比べて遜色ない結果が得られると

いえる。先に述べたように、実験環境においては分類対象テキストの正解数を知ることができるが、実際の場面においては、分類対象テキストに付与されるべきカテゴリ数を事前に知ることはできない。実際の場面での利用を想定した場合には、多少性能が落ちたとしても、閾値を用いる方法で十分に正解のカテゴリ数を推定することができるといえる。

表 14 第一階層における F 値 (All、複製手法)

	分類手法1			分類手法2		
	vote	iw	sw	vote	iw	sw
5 NN	0.630	0.628	0.628	0.580	0.584	0.580
10NN	0.642	0.642	0.641	0.591	0.592	0.590
15NN	0.650	0.651	0.649	0.593	0.599	0.595
20NN	0.645	0.648	0.644	0.598	0.603	0.593
25NN	0.645	0.647	0.644	0.599	0.604	0.599
30NN	0.641	0.643	0.640	0.599	0.601	0.599
35NN	0.642	0.646	0.641	0.596	0.599	0.597
40NN	0.638	0.643	0.639	0.594	0.599	0.596
45NN	0.638	0.643	0.638	0.592	0.595	0.592
50NN	0.639	0.644	0.639	0.592	0.599	0.594

6. カテゴリごとの分析

図4に、第三階層のカテゴリでの人により付与されたカテゴリ数と分類結果として分類されたカテゴリ数を示した。これは、データセットにAllを用いて、第三階層で最も性能が高かった手法(All、k=20、複製手法、分類手法1、iw)の場合であり、20回以上出現するカテゴリの結果のみを示している。軸はログスケールである。このグラフでは、A0-WealthやA1-Influentiaは、人がコーディングした数よりも、自動分類の結果のほうが、多くのこのカ

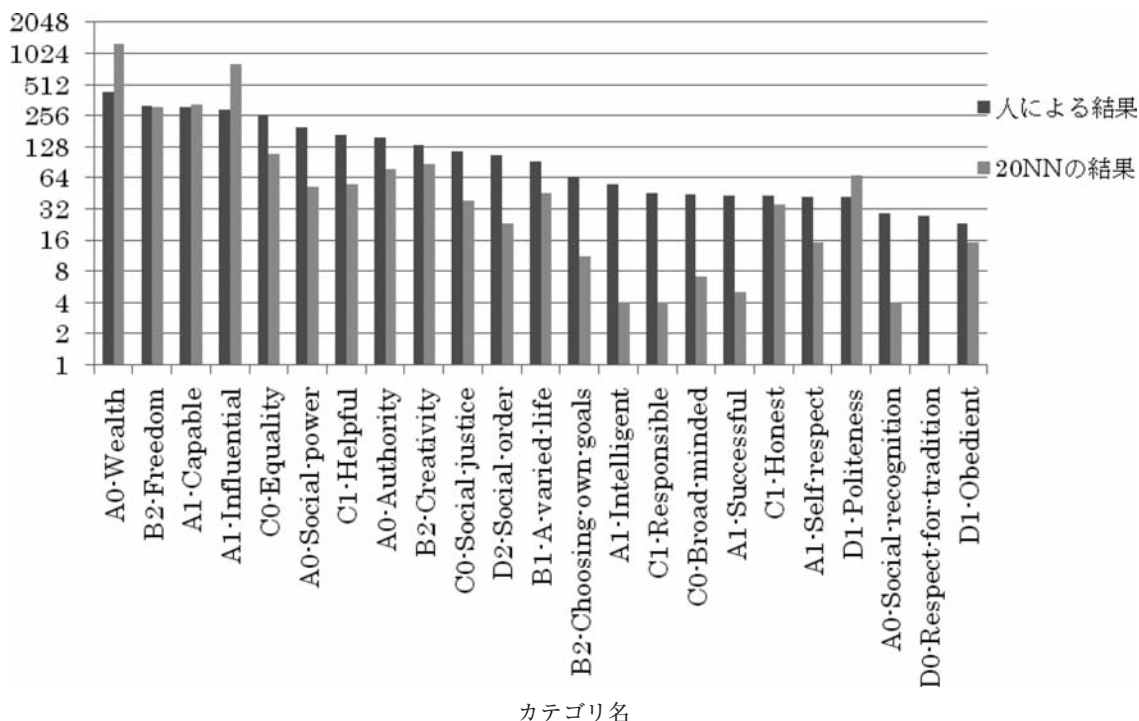


図4 人より付与された数と自動分類により分類された数

テゴリに分類されている割合が高いことを示している。反対に、B2-Choosing-own-goalsやA1-Inteligent、C1-Responsible、A1-Successfulは、自動分類ではあまり分類されていなかった。カテゴリによって、分類器が分類する傾向にあるカテゴリとそうでないカテゴリがあった。これは各カテゴリの学習用データの大きさと比例していない。これの原因として考えられるのは、たとえば、似た特性を持っているカテゴリが複数あり、どちらか一方に分類されてしまっているのではないかということである。今後、カテゴリ間の類似性や関係などをもとに、それぞれのカテゴリに対する効果的な学習用データの表現方法、分類手法を検討していくことが必要である。

個々のカテゴリをみていくと、46のカテゴリのうち20カテゴリにはどの文も分類されていなかった。これらは、学習用データでの出現回数が低いものが多かった(最大で27回の出現回数)。このことから、ある一定以上の学習用データの量の確保は必要だといえる。

つぎに、同じデータを用いて、個別のカテゴリごとに、精度(P)、再現率(R)、F値を計算し、F値が高かった10カテゴリの結果を表15に示した。この表からは、A1-Capableも正解とほぼ同じ数であるが、1/3程度しか正解しておらず、精度、再現率ともに低いこと(0.302、0.318)や、D1-Politenessは正解よりも多くの文が分類されており、再

表15 F値が高い10カテゴリの精度、再現率、F値

カテゴリ	正解	20NN	一致	P	R	F値
B2-Creativity	133	88	65	0.739	0.489	0.588
D1-Politeness	42	68	31	0.456	0.738	0.564
B2-Freedom	319	310	135	0.435	0.423	0.429
A0-Wealth	431	1,254	358	0.285	0.831	0.425
C1-Honest	43	35	16	0.457	0.372	0.410
C0-Equality	260	108	60	0.556	0.231	0.326
A1-Influential	291	818	179	0.219	0.615	0.323
A0-Authority	159	78	38	0.487	0.239	0.321
A1-Capable	308	325	98	0.302	0.318	0.310
A1-Self-respect	42	15	7	0.467	0.167	0.246

現率が高くなっている(0.738)ことなどがわかる。また、B2-Creativityは分類結果は正解よりも少ないが、一致率が高いため精度が高い(0.739)。以上のように、カテゴリによって精度が高いものや再現率が高いものもあった。今後、どのような場合に再現率が高くなるのか、精度が高くなるのかなど、それぞれのカテゴリの精度や再現率をみていき、特徴を分析することが必要である。自動分類は、一つの要素だけでなく、学習用データの量、学習用データの表現方法、カテゴリ間の関係など様々な要素が、分類性能にかかわってくる。今後は、これらのことに注目しながら、より適切な分類手法を提案していく予定である。

7. おわりに

本研究では、人の価値観を表すカテゴリを付与したテストコレクションを新たに作成し、それを用いて自動分類実験を行った。実験の結果、第三階層で最も高いF値は0.324であり、閾値を用いて正解カテゴリ数を推定しても0.303のF値を得ることができた。

3.3で示したように、4証言を用いた場合のコーダー間のF値は0.387であった。4証言と全データを用いた自動分類結果を一概に比べることはできないが、コーダー間で行った場合でもそれほど高い一致率を得ることができないタスクに対して、コンピュータで行った結果は、これらを自動的に行える可能性を示しているのではないかといえる。

実験で検討した手法は、学習用データの表現手法と分類手法のみであり、基本的な手法の検討しか行っておらず、今後、様々な点から性能向上を見込むことができる。事例分析で明らかになったように、カテゴリの中には類似した特徴を持つ識別しにくいカテゴリが存在する可能性がある。これに関しては、言語モデルなどを取り入れて性能向上をはかる予定である。

その他に、今後は、本実験では用いなかったSVMや他の分類器で実験する予定である。また、

本研究では、出現している語のみを対象にしたが、出現している語の属性や同義語の利用などを検討する予定である。さらに、語の特性だけでなく、文の長さや著者の属性情報などを考慮することもできる。

本稿では、人の価値観を表すカテゴリへの自動分類の第一段階として、自動分類実験に用いるテストコレクションの作成と、人が行ったコーディングの結果をどこまでコンピュータが再現できるか、基本的な分類手法を用いて検討した。その結果、コンピュータで人の結果を、ある程度、再現できることがわかった。今度は、対象となるカテゴリやテキストの特徴に合わせた分類手法を提案し、性能向上を目指す予定である。

謝辞

本研究は科研費（19700232）の助成をうけたものである。また、「2008年度駿河台大学在外研究派遣プログラム」により、2008年8月から2009年8月までの一年間、米国メリーランド大学にて研究を行った成果の一部である。在外研究の機会を与えてくださった駿河台大学に心より感謝する。メリーランド大学 College of Information studies および PopIT プロジェクトのメンバからは、研究しやすい環境と適切なアドバイスをいただいた。ここで感謝の意をささげたい。

引用文献

- 1) PopIT, Ping Wang [http://terpconnect.umd.edu/~pwang/PopIT/contact.htm] 2010.01.02
- 2) Cheng, A.-S., Fleischmann, K.R., Wang, P., Ishita, E., and Oard, D.W. (2010). Values of Stakeholders in the Net Neutrality Debate: Applying Content Analysis to Telecommunications Policy, 43rd Hawaii International Conference on System Sciences, Kauai, HI, USA
- 3) Morris, R. (1994). Computerized content analysis in management research: A demonstration of

advantages and limitations. *Journal of Management*, 20 (4), 903-931.

4) Fleischmann, K. R., Oard, D. W., Cheng, A.-S., Wang, P., and Ishita, E. (2009). Automatic Classification of Human Values: Applying Computational Thinking to Information Ethics. Proceedings of the 72nd Annual Meeting of the American Society for Information Science and Technology (ASIS&T), Vancouver, Canada.

5) Breese, J. S., Heckerman, D. and Kadie, C. (1998). Empirical analysis of predictive algorithms for collaborative filtering. In Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence (UAI-98), pp. 43-52

6) T. Joachims (1998). Text categorization with Support Vector Machines: Learning with many relevant features. Proceedings of ECML-98, 10th European Conference on Machine learning, No.1398, pp. 137-142

7) Cover, T. M. and Hart, P. (1967) Nearest neighbor pattern classification, *IEEE Transactions on Information Theory*, Vol.13, No.1, pp. 21-27.

8) A. Wiczorkowska and P. Synak (2006) "Quality Assessment of k-NN Multi-Label Classification for Music Data" ISMIS 2006, 389-398

9) Zhang, M. and Zhou, Z. (2005), "A k-Nearest Neighbor Based Algorithm for Multi-label Classification." *IEEE International Conference on Granular Computing*, Vol. 2, pp. 718-721

10) X. Shen, M. Boutell, J. Luo and C. Brown (2004), "Multi-label Machine Learning and Its Application to Semantic Scene Classification", *International Symposium on Electronic Imaging*

11) U.S. Senate. 2006, Feb 7. Senate Committee on Commerce, Science and Transportation Hearing on Network Neutrality.

12) Fed. Comm. Commission. 2008, Apr 17. Broadband Network Management Practices Public Hearing. Palo Alto.

- 13) Schwartz, S. (1992). Universals in the Content and Structure of Values. In M. Zanna, ed. *Advances in Experimental Social Psychology*, Academic Press, 25, 1-66.
- 14) Schwartz, S.H. (1996). Value priorities and behavior: Applying a theory of integrated value systems. In C. Seligman, J.M. Olson, & M.P. Zanna (Eds.), *The psychology of values: The Ontario Symposium*, Vol. 8 (pp. 1-24). Hillsdale, NJ: Erlbaum.
- 15) Caprara, G. V., Schwartz, S. H., Cabaña, C., Vacine, M., & Barbaranelli, C. (2005). Personality and politics: Values, traits, and political choice. *Political Psychology*, 27 (1), 1-28
- 16) WEKA, [<http://www.cs.waikato.ac.nz/ml/weka/>] 2010.01.02
- 17) G. Tsoumakas, I. Katakis, (2007). "Multi Label Classification: An Overview", *International Journal of Data Warehousing and Mining*, David Taniar (Ed.), Idea Group Publishing, 3 (3), pp. 1-13

Automating multiple-label classification for human values

ISHITA Emi*, **An-Shou Cheng****, **Douglas W. Oard****, **Kenneth R. Fleischmann****

[Abstract] This paper describes the development and use of a new test collection for automatic labeling of sentences with human values. A total of 2,005 sentences from 28 prepared testimonies presented before public hearings on "Net Neutrality" before the U.S. Congress or Federal Communications Commission between 2006 and 2008 were manually annotated using Schwartz Values Inventory categories (a set of 56 basic human values that is widely used in social science research). Some sentences were annotated with multiple values. We employed three methods for selecting training instances when multiple labels are present. We trained three weighted kNN classifiers on a training partition, evaluating each on a disjoint test partition. A balanced F measure of 0.303 (mean across the 46 annotated categories) was achieved using a threshold learned on held-out data.

[Key words] Automatic Classification, Content Analysis, Human Values, Schwartz Values Inventory

* Surugadai University, Faculty of Media and Information Resources

** University of Maryland, USA