

## Web 上のテキスト情報を用いた人物評価手法

緒方 進<sup>†</sup> 池田 真司<sup>‡</sup> 牟田 高信<sup>†</sup> 木本 勝敏<sup>†</sup>

<sup>†</sup>株式会社 NTT ドコモ 〒239-0847 神奈川県横須賀市光の丘 3-5

<sup>‡</sup>NTT ソフトウェア株式会社 〒231-8551 神奈川県横浜市中区山下町 209

E-mail: <sup>‡</sup>{ogatas, muta, kimotom}@nttdocomo.co.jp, <sup>†</sup>nikeda@po.ntts.co.jp

**あらまし** 現在のインターネットコミュニティでは、不特定多数の人とコミュニケーションを行うため、安心でかつ信頼できる相手かどうかを評価・判断することが重要である。しかし現在、尺度を統一した自動的な人の評価手法は確立されていない。本稿では Web 上の書き込み記事から人物の特徴を評価することを目的とし、まず人の感情を評価する指標の一つである EQ (Emotional Intelligence Quotient) を用いて評価の尺度を統一し、人物評価にテキストマイニング技術を適用するための技術手法、及び人単位にてデータを評価するための人物評価手法を提案する。また本評価手法の精度に関する検証を行い、その検証結果を報告する。

**キーワード** 人物評価、テキストマイニング、スコアリング、Blog、掲示板

## Evaluation Method of Person Using the Text Information on the Web

Susumu OGATA<sup>†</sup> Shinji IKEDA<sup>‡</sup> Takanobu MUTA<sup>†</sup> and Masatoshi KIMOTO<sup>†</sup>

<sup>†</sup>NTT DoCoMo, Inc. 3-5 Hikarino-oka, Yokosuka-shi, Kanagawa, 239-8536, Japan

<sup>‡</sup>NTT Software Corporation 209, Yamashita-cho, Naka-ku, Yokohama-shi, Kanagawa, 231-8551, Japan

E-mail: <sup>†</sup>{ogatas, muta, kimotom}@nttdocomo.co.jp, <sup>‡</sup>nikeda@po.ntts.co.jp

**Abstract** With the dynamic growth of the Internet community, there is more chance to interact with anonymous people. Hence there are great demands for judging the anonymous person's credibility. Although, there is no current standardized method to evaluate the anonymous person over the network automatically.

In this paper, we intend to evaluate the feature of the anonymous person from the written messages from the web. For the purpose, we apply Emotional Intelligence and Quotient(EQ) for basic criterion for the evaluation. This paper reports our proposals of technical method for applying text mining technology to evaluation of the feature of the anonymous person, and evaluation method of the anonymous person for evaluating the data by a person. And this paper reports the result of verification of accuracy of our proposal method.

**Keyword** Evaluation of person, Text-mining, Scoring, Blog, Bulletin Board

### 1. はじめに

近年、インターネット利用者の増大により、インターネット上には様々な情報・知識を共有する場であるインターネットコミュニティが多く構築されている。このインターネットコミュニティ内では、多くの人がコミュニケーションサービス(掲示板、チャット、Blog等のコミュニケーションを実現するために提供されるサービス)を利用しているため、人は不特定多数の人とコミュニケーションを行うこととなる。そのため人

が安心・安全にコミュニケーションを行うためには、コミュニケーション相手が安心でかつ信頼できるかを評価・判断する(相手の人の評価情報を取得する)仕組みが必要である。

人の評価情報を必要とするサービスの中でも、オークションのように取引の実績情報を基に評価を実施するサービスにおいては、人の評価情報[1]を収集/流通する仕組み[2]は既に実現されている。

一方、コミュニケーションサービスにおいてコミュ

ニケーション相手の評価手法として存在するものは、人が主観で入力した値の累積結果による評価手法であり、手間がかかり必ず入力されるとは限らない、主観なので正確性に欠ける(絶対的な尺度を定義できない)などの問題点[3]があり、評価情報の利用範囲が限定されてしまう。このようにコミュニケーションサービスにおいては尺度を統一した自動的な人の評価手法は確立されていない。

本稿では、尺度を統一するため人の評価軸を選定し、Web上の書き込み記事から人物の特徴を自動的に評価することを目的とし、人物評価にテキストマイニング手法を適用するための技術手法、及び人単位にてデータを評価するための人物評価手法を提案する。また本評価手法の精度に関する検証を行う。

以下、第2章では人の評価情報生成に関連する技術手法について述べる。第3章ではテキストマイニング技術を用いて、テキストデータから人の内面的特徴を抽出/評価することを目的とした技術手法の概要について述べ、第4章では第3章の手法の詳細内容について述べる。第5章では第3、4章の手法に関する評価精度を検証した結果、考察について述べる。最後に第6章で本稿にて検討、検証した内容についてまとめる。

## 2. 人の評価情報生成に関する技術手法

本章では人の評価情報生成に関連する技術手法について述べる。本稿における評価対象は、現実存在している「人」ではなく、インターネットコミュニティに参加している仮想的な「人」である。このようなインターネットコミュニティ上の人を評価するために用いられる評価対象データは、図1のような分類で表すことができる。

ットオークションのように取引相手が匿名に近い状況で取引を円滑化するため、取引時の満足度等の主観的評価情報を投票・活用する「投票方式」がある。またアンケートや人気投票、サイトへのアクセス回数といった第三者から得られる情報を基にした人の評価でも、アクセス情報や投票結果の情報を利用する「投票方式」が用いられる。これらは既に実用化された既存技術である。

次にチャットや掲示板の情報の中で、コミュニケーション上のリンク情報を基にした人の評価技術としては、オピニオンリーダーやオーソリティを発見する「リンク解析」がある。この技術は情報発信者間やウェブサイト間におけるコミュニケーション上の繋がりを明らかにすることで、情報発信元(オピニオンリーダー)の情報、影響度等を獲得することができるが、実際の実用例は少なく、まだ研究段階の技術である。

一方チャットや掲示板の文章や第三者からの投稿などのようなテキストデータを基にした人の評価については、テキスト中に現れる特徴を抽出・評価する「テキストマイニング技術」を適用することが考えられる。一般的な「テキストマイニング技術」では、テキストデータから文書検索や傾向・関係を抽出する事を目的として利用され、アンケート分析・問い合わせ分析等、様々なテキスト分析としての実用例がある。このテキストマイニング技術を人単位で適用させて「人物評価」を行うためには、一般的なテキストマイニング手法を用いてデータ単位で評価するだけでなく、人単位にてデータを評価するための人物評価手法の検討が必要となる。

本検討では、評価情報生成手法としてテキストマイニング技術を用いて、テキストデータから人の内面的特徴を抽出/評価するために必要な人物評価手法の検討を行うこととする。

## 3. テキストマイニング技術を用いた人物評価手法

本章では、テキストマイニング技術を用いて、テキストデータから人の内面的特徴を抽出/評価する技術手法の概要に関して述べる。

図2に本手法の概要を示す。従来のテキストマイニングでは、テキスト中に現れる話題を抽出/評価することを目的としているが、本検討ではテキストデータから人の内面的特徴を抽出/評価することを目的とするため、従来手法でのデータ単位の評価に加えて、人単位の評価を実施する必要がある。

人単位の評価においては同一の尺度が必要であり、人の内面的特徴の評価軸を明確にすることが重要となる。人の感情を評価するためのモデルは多く存在する

評価対象	評価対象データ	評価結果	情報提供効果	技術手法	
人 (インターネット上のコミュニティ)	取引などの実績情報	B2C取引実績	ブライオ対など	与信情報の活用	
		C2C取引実績	投票結果(高い○件、悪い△件など)	投票方式	
	第三者からの評価情報	アンケート、人気投票、サイトへのアクセス回数	人の評価把握に困るコミュニケーションサービスの円滑・活性化		
	チャット・掲示板内の会話などのテキスト情報	コミュニティへの書き込み(返信)の件数情報	人単位の評価情報(ポジティブ、趣味・嗜好など)	人の評価把握に困るコミュニケーションサービスの円滑・活性化	本稿検討範囲
		リンク情報(リンクの数、URLの情報、伝播したテキスト情報など)	オピニオンリーダー情報	オピニオンリーダーの情報流通に関するコミュニケーションサービスの円滑・活性化	リンク解析

: 既存技術(実用化済)
  : 研究段階(一部実用化済)

図1. 人の評価に用いる評価対象データ

B2C 取引実績情報を基にした人の評価については、サービス利用者の取引実績に基づいて評価情報を蓄積・活用する「与信情報の活用」がある。またC2C取引実績情報を基にした人の評価としては、インターネ

が、本検討では現在実践的な枠組みとして扱われ、評価手法が確立されている EQ (Emotional Intelligence Quotient) [4] のモデルを評価軸として取り上げる。EQ は「心の知能指数」といわれ、心内知性 (自己認識、ストレス共生、気力創出)、対人関係知性 (自己表現、アサーション、対人関係)、状況判断知性 (対人受容、共感) など 8 つの因子から構成されており、個人の総合的な能力に関する指標となる。この EQ モデルで定義される因子を評価軸として選定することで、尺度を統一した人の評価が可能となる。本検討では例として EQ モデルの 1 因子である「共感因子 (思いやり)」を人物評価の評価軸として設定した。

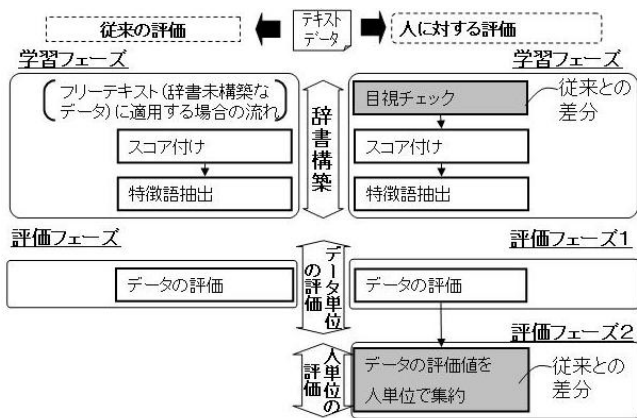


図 2. 人物評価手法の概要

#### 4. 辞書構築と人物評価

本章では、第 3 章にて提示した手法の具体的な内容である、辞書構築と人物評価について述べる。

##### 4.1. 学習フェーズによる辞書構築

学習フェーズでは、評価フェーズで特徴度合いを数値化する際に必要となる辞書を構築する。

辞書構築では、事前準備として学習用データ内の各記事に共感因子が含まれているかどうかの判定が必要となる。本検討では目視によって共感因子がデータ中に含まれるかチェックを実施したが、目視のため判断を行う側の主観が強く働く可能性が高い。従って評価の精度をあげるには複数人でチェックを実施する必要がある。本検討では学習用データの各記事毎の「本文」の内容について、共感因子を感じるかどうか 3 人でチェックを行い、2 人以上が思いやりを感じた場合には共感因子「有り」、そうでない場合には共感因子「無し」とした。また本検討では、各人の判断ミスを極力減らすため、判断は以下の観点で行うよう、意識合わせを行った。

- ・ 相手の発言に対し、発言意図・状況を正確に汲み取っている事が文章から読み取れること

- ・ 上記を踏まえて、相手の立場にたった発言が行えていること

以上のように目視チェックを完了した学習用データを用いて、以下の手順で辞書の構築を行う。

1. 共感因子「有り」、「無し」の記事を分離し、2 つのデータを生成。
2. 共感因子「有り」のデータ中の「本文」のテキストデータに対して、形態素解析により品詞分解を行い単語を抽出する。共感因子「無し」のデータについても同様の作業を実施する。
3. 共感因子「有り」のデータから抽出された各単語を対象として、その出現頻度などから、「有り」のデータ中における出現のし易さを数値化する (各単語のスコア付け)。共感因子「無し」のデータについても同様の作業を実施する。
4. 共感因子「有り」のデータから抽出された各単語とそのスコアを参考にして、共感因子の特徴語となる単語を抽出し、共感因子「有り」用の辞書を生成する。共感因子「無し」のデータについても同様の作業を実施する。

以上の作業により、共感因子「有り」「無し」用の辞書 2 種類を生成することができる。

##### 4.2. スコア付け

共感因子に関して、データを共感因子「有り」「無し」で分類したが、それぞれの分類に含まれる特徴語のスコアを生成することがポイントとなる。特に、共感因子「有り」でのスコアと「無し」でのスコアで差異が発生し、特徴的であればそのスコアが高く、特徴的でなければそのスコアが低くなるようなスコア付けが必要である。このスコアの差が、特徴語を含むデータが共感因子を含むか否か判定する基準となる。スコア付けの手法としては、SOM [5]、SC/ESC [6]、SVM [7]、tfidf 等が代表的であるが、本検討は人の評価情報の抽出が行えるか基本的な可否検証を目的とするため、スコア付け手法として代表的な tfidf を用いることとした。その数式を式(1)、(2)に示す。

$$tfidf_{ik} = tf_{ik} \times idf_i$$

$$tf_{ik} = w_{ik} / \sum_j w_{jk} \quad \dots \dots (1)$$

$$idf_i = \log \frac{M}{v_i} + 1 \quad \dots \dots (2)$$

式(1)において、 $w_{ik}$  を分野  $k$  (評価軸有り/無し) における単語  $i$  の総出現回数とすると、分母  $\sum_j w_{jk}$  は分野  $k$  における総特徴語数、 $tf_{ik}$  は分野  $k$  における単語  $i$  の出現頻度を表し、式(2)において  $M$  を学習用データの記事総数、 $v_i$  を単語  $i$  を含む記事数とすると、 $idf_i$  は全記事から見た単語  $i$  の重要度を表す。

抽出された特徴語に対して、tfidf による数値化の例

を図3に示す。

特徴語「旦那」は「有り」「無し」データ共に出現するが、それぞれの出現頻度の偏りにより、「有り」と「無し」の辞書のスコアがそれぞれ異なることもある。



図3. tfidfによる数値化

### 4.3. 特徴語抽出

共感因子の判定において、抽出された全ての単語がその特徴語となるわけではなく、判定に必要な無い(判定に悪影響を及ぼす)単語も含まれる。そのような単語をふるい分けることによって、特徴語のみで構成される辞書を生成する。本検討では以下の観点にて分類される単語を非特徴語として処理する。

- ・ idf 値が 1.3 以下の単語、すなわちその単語が含まれる記事の数が、全ての記事数の半数以上となる単語は、多くの記事に登場するため、共感因子としての特徴性を持たないと判断する
- ・ 全データ中における出現総数が 5 未満のような出現総数が少ない単語は、限られた人のみが使う特殊語と判断する
- ・ 共感因子として意味の無い記号(「」( )・など)系の単語は、共感因子としての特徴性を持たないと判断する
- ・ 共感因子として意味の無い名詞(数字系、人の名前など)系の単語は、共感因子としての特徴性を持たないと判断する

なお、従来のテキストマイニング技法では、文章の主題などを決定する「名詞」「動詞」を抽出対象の特徴語として扱うため、辞書構築時にその他の品詞をフィルタリングする。一方本検討では人物の内面的特徴を抽出することを目的としているため、名詞、動詞以外の品詞にも着目し、特徴語の抽出を行った。

### 4.4. 評価フェーズ1によるデータ評価

評価フェーズでは、評価対象の人物が書き込んだデータに対して、データ中に共感因子に関連する特徴語を含むか否かを判断する。そしてその特徴語の特徴度合いを数値化(スコアリング)する。

スコアリングの方法としては、従来手法では Dice 係数、Jaccard 係数、余弦、拡張ブーリアン、内積などが用いられるが、本検討では精度比較の基準になる代表的なスコアリング手法である内積を用いてスコアリ

ングを行い、記事毎の評価を行う。具体的には、学習用記事の特徴ベクトル(各単語毎の tfidf 値)と評価用記事の特徴ベクトル(tf 値)の内積を算出して、学習用記事と評価用記事の類似度をスコアリングする。

各記事に対し、共感因子「有り」「無し」辞書の両者で計算を行い、前述の内積によりスコアリングを行う。この計算結果を比較することで、どちらの分野(「有り」「無し」)に評価用記事が近いかを評価することができる。すなわち、「有り」のスコアリング結果が「無し」よりも高ければ、評価用記事は共感因子「有り」の記事とする。また、人の評価値を生成するための基礎値としては、「有り」と「無し」各内積値の差分をその値とした。図4にその例を示す。図4においては、共感因子「有り」の内積値が「無し」より高いため、評価用記事は「共感因子が有る」と判定する。

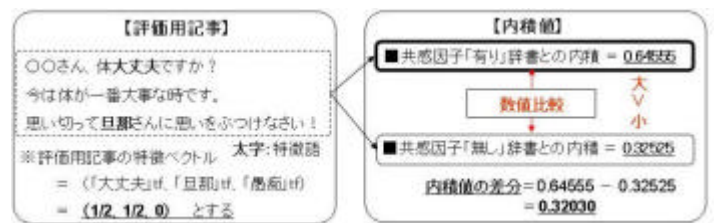


図4. 内積値の比較

### 4.5. 評価フェーズ2による人の評価

4.4 節のスコアリング結果から人の評価を求めるために、この評価値を人単位に再評価する手法が必要である。よって本検討では以下のように人の評価値を求めるための評価式を定式化し、記事単位の評価値を評価式に当てはめることで人単位の評価値を算出する。

1. 評価用記事を発信者毎に分類する(本検討では各記事毎に設けた「ユーザID」をキーに分類を行う)
2. 同一人物の作成記事毎のスコアリング結果を元に、評価式に基づいて人単位の評価値を算出する

人の評価値を求める際に、最も基本となる計算式は、各記事の内積値の総和を求めることである。但し、総和のみで値を決定すると、共感因子のある投稿記事数が多い人程単純に評価値が高くなる。本検討では、投稿記事総数に対する共感記事数の割合を係数として評価し、評価式に組み込んだ。w<sub>ik</sub>を人物kの記事iにおける内積値の差分(スコア)、N<sub>k</sub>を人物kが発信(作成)した共感因子「有り」の記事数、A<sub>k</sub>を人物kが発信(作成)した記事総数とすると、人物kの評価値H<sub>k</sub>は式(3)のように表される。

$$H_k = \sum_i w_{ik} \times \frac{N_k}{A_k} \quad \dots \dots (3)$$



## 5. 評価手法の検証

本章では、第 3、4 章にて述べた人物評価手法の精度に関して検証した結果について述べる。

### 5.1. 検証用データ

今回の検証用データとしては、インターネットコミュニティ（掲示板）上の書込記事を準備した。なお検証用データは複数の記事から構成される。またデータとしては以下のように 2 種類準備し、1 種類は辞書作成時に使用する学習用データ、もう 1 種類は学習用データとは異なる評価用データである。

- ・ 学習用データ：1 トピック（記事数：459）
- ・ 評価用データ：5 トピック（記事数：各々300）

### 5.2. 辞書データの作成

学習フェーズの辞書作成時の目視チェックとして、共感因子「有り」「無し」の 2 種類の記事を分類した。共感因子「無し」の記事については、共感因子が含まれない記事ではなく、共感因子を否定する記事（思いやりが無いことが明らかに分かるような記事）であるべきだが、本検討では共感因子が「有り」の記事と「それ以外」の記事の 2 種類に分類し、それぞれから辞書を作成した。作成した辞書データを図 5 に示す。図 5 は tfidf 値の上位 20 位までの特徴語である。4.3 節において、人物の内面的特徴を抽出することを目的として名詞、動詞以外の品詞にも着目し、特徴語の抽出を行ったことにより、名詞、動詞以外の品詞も特徴語として抽出されている。

特徴語	品詞
さん	名詞
お子さん	名詞
大変	名詞
頑張る	動詞
仕事	名詞
ペット	名詞
う	助動詞
大事	名詞
なんて	助詞
一言	名詞
写真	名詞
そういう	連体詞
生まれ	名詞

図 5. 辞書データ

### 5.3. 学習用データの評価結果

5.2 節にて示す辞書を用いて、学習用データに対する目視結果と、辞書を用いた各記事に対するスコアリング結果の誤差を測定した。なお各記事のスコアリング結果（内積値の差分）は、正の値であれば共感因子「有り」、負の値であれば共感因子「無し」とする。図 6 に学習用データの検証結果を示す。目視結果とスコアリング結果の合致率は 83% となった。この結果のうち共感因子「無し」と全員が目視判断した記事において、内積値の差分が大きくなった記事（図 6 ? (左上)）がいくつか存在した。この理由としては、比較的短い文章の記事中に tfidf 値が高い単語が含まれるため、その文章自体のスコアが大きくなることが挙げられる。これを解決するには短文に対する処理として、短文用の辞書と評価式の構築が必要であると考えられる。

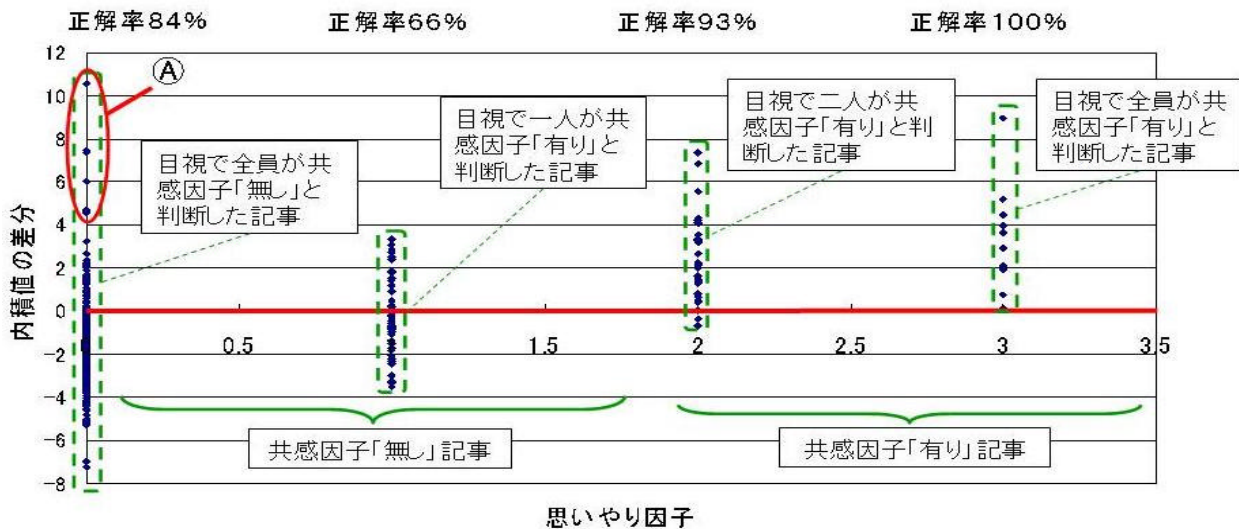


図 6. 学習用データの評価結果

## 5.4. データ単位の評価結果

人物評価用データとして、様々な内容のトピックを5つ準備し、評価を実施した。学習用データと同様に、このスコアリング結果が正しいか否かを判定するために、評価用データに対し目視チェックを行い共感因子が含まれるか否かを手で判定した。その人手での判定結果とスコアリング結果の比較により、合致率を測定した。

結果として、平均合致率は72.6%であり、各トピック毎の合致率は、79%、61%、93%、65%、65%となった。3トピックが60%代と低い合致率となったが、共通の要因として、データの文調が学習用データと大きく異なっているため、辞書がうまく機能せず、合致率が低下したことが挙げられる。逆に高合致率の2トピックは、学習データとの文調や参加者構成に類似点が多く、結果として合致率が高くなっている。これらは辞書構築時に学習用データのサンプル数が少なく、語彙数不足、低精度の辞書が構築されたため、学習用データに類似したデータは高合致率となったと考えられる。

## 5.5. 人単位の評価結果

5.4節で実施したデータ単位の評価結果を用いて、人(ユーザID)単位の評価の集約を行った。ここで人単位の集約する際には、投稿記事数に大きくばらつくという問題があり、投稿記事が少ない場合は1つの記事のスコアに影響されるため、高精度の判断が行えない可能性がある。そこで本検討では投稿記事数が5件以上の人を対象に評価を行った。

結果として、目視で共感因子「有り」と判定された人は、スコアリングにおいても全員漏れなく共感因子「有り」と判定された。しかし合致率としては60.2%となった。これは本手法では、本来スコアリングによる判定結果が共感因子「無し」と判定すべき人が「有り」と判定されており、その誤判定が多いため低合致率となっている。判定精度に大きな影響を与える辞書およびデータに関して調査した結果、以下の問題が判定精度に影響を及ぼしていると考えられる。

- ・ 共感因子「無し」として判定されているものが、共感因子を否定する(思いやりが無い)ものでなく、共感因子「有り」ではないものとなっている
- ・ 学習用データのサンプル数が少なく、特徴語のスコアの精度に問題がある
- ・ 記事の長さが非常に短い場合、共感因子にあまり影響を及ぼさない単語のスコアが、スコアリングに大きなウェイトを占めてしまう

今回の結果として、人が目視によって共感因子有りと判定された結果は、スコアリングによる判定結果でも共感因子有りと判断されており、本評価手法は有効

であると考えられる。

## 6. まとめ

本稿では、インターネットコミュニティの参加者がコミュニケーションサービスを利用する際に蓄積されるWeb上の書き込み記事から、人の評価情報を生成可能かどうかについて検討を実施した。具体的には人物評価にテキストマイニング手法を適用するための技術手法、及び人単位にてデータを評価するための人物評価手法を提案した。さらに検証を実施し、精度面で問題があったが人の評価情報を生成することが可能であることを確認した。

今後の課題としては、学習用データのサンプルの充実、特徴語の抽出ロジック(不要な単語の見分けロジック)の精査、特徴語のスコア付けロジックの精査による(辞書の)精度向上、誤判定の誤差を除く手法の検討などが挙げられる。

## 文 献

- [1] Kollock, P. The Production of Trust in Online Markets. Final version to appear in: *Advances in Group Processes* (Vol.16), edited by E.J.Lawler, M. Macy, S. Thyne, and H.A. Walker. Greenwich, CT: JAI Press. 1999.
- [2] G. Akerlof, The Market for 'Lemons': Qualitative Uncertainty and the Market Mechanism, *Quarterly Journal of Economics*, 84, pp488-500, 1970
- [3] Paul Resnick, Richard Zeckhauser, Eric Friedman, and Ko Kuwabara, Reputation systems. *Communications of the ACM*, pp45-48, December 2000.
- [4] Salovey, P., & Mayer, J. D. (1990). Emotional intelligence. *Imagination, Cognition, and Personality*, 9, 185-211.
- [5] 柳田 卓郎, 三浦孝夫, 塩谷勇, SOMによるテキスト分類, データ工学ワークショップ(DEWS), 電子情報通信学会データ工学研究会, 2003
- [6] Hang Li and Kenji Yamanishi, Text Classification Using ESC-based Stochastic Decision Lists, *Information Processing & Management*, 38(3), pp343-361, 2002.
- [7] V. N. Vapnik. *The Nature of Statistical Learning Theory*. Springer-Verlag, second edition, 2000