

テキストに現れる感情，コミュニケーション， 動作タイプの推定に基づく顔文字の推薦

江村 優花^{†1} 関 洋平^{†2}

現在，電子メールや Twitter で，顔文字が使用されている．顔文字はユーザの感情を表すのに便利であるが，反面，その種類は膨大であり，適切な一つを選ぶことは難しい．そこで，本研究では，ユーザの顔文字選択支援を目的とし，ユーザが入力したテキストに現れる感情，コミュニケーション，動作のタイプ推定を行い，顔文字を推薦する方法を提案する．感情，コミュニケーション，動作のタイプの推定は k -NN に基づき実装し，推定精度は，マクロ平均で 53.1% となった．また，システムが推薦する顔文字がユーザの意図にどの程度適合しているか実験した結果，66.7% の顔文字が適切に推定されており，感情カテゴリのみを用いて推薦された結果と比べて，提案手法の顔文字推薦の精度が有意に向上していることがわかった．

Facemark Recommendation based on Emotion, Communication, and Motion Type Estimation in Text

YUKA EMURA^{†1} and YOHEI SEKI^{†2}

Many users used facemarks in recent computer mediated communication environments such as e-mail, Twitter or facebook. Facemarks are useful to express the emotion or communication functions beyond texts. However, many users feel difficult to choose the right one from lots of candidates. We propose a method to recommend facemarks based on the estimation of emotions, communication, or motion types in texts written by users. We implemented emotion, communication, or motion type estimation system with k -NN, and the accuracy of estimation is 53.1% on macro average. We also estimated the relevance of recommended facemarks for user intention, and found that 66.7% of facemarks were recommended properly, which improved significantly over the recommendation results only from emotions categories.

1. はじめに

現在，電子メールやチャット，Twitter など，文字ベースのコミュニケーションが頻繁に利用されている．これらのコミュニケーションにみられる特徴の一つとして，顔文字があげられる³⁾．顔文字の特徴としては，オンラインコミュニケーション中の感情を表現しやすくすることや，その使用に際して社会的な地位，相手との関係などが影響をおよぼすことが指摘されている^{1),7)}．

顔文字とは“(^ - ^)”のように，記号や文字を組みあわせて表情を表現したもので，テキスト中で表現された感情を強調・補足できる，という利点がある反面，その種類は数多く，増加の一途をたどっている．その中から，ユーザが文章で伝えたい感情に適切な顔文字を，ただひとつだけ選択するのは困難である．また，顔文字入力の主な方法である顔文字辞書や予測変換機能では，感情以外の意味での使用を目的とした顔文字を入力することは難しく，予測変換機能では，単語単位を対象としてしか顔文字を提示できない．そのほかの手段として，他のテキストからのコピーアンドペーストやユーザ自身による直接入力があるが，これらは操作数が多く，効率的ではない．

そこで，本研究では，ユーザによる適切な顔文字選択の支援を目的とし，ユーザの入力文章から，感情カテゴリやコミュニケーションや動作を反映したカテゴリを推定し，顔文字を推薦するシステムの構築を目指す．

2. 関連研究

顔文字を考慮した感情推定システムとして，篠山ら¹¹⁾の研究がある．このシステムでは，文章と顔文字それぞれの感情を推定し，両方の結果をもとに全体の感情を推定している．しかし，彼らの研究では，対象とする顔文字が少なく，幅広い顔文字推薦への応用は難しい．より大規模なものとしては，Ptaszynski ら²⁾らの CAO システムがあげられる．このシステムでは，顔文字の抽出と感情解析を行い，90%以上の精度を上げている．また，村上ら⁶⁾は，顔文字に関する調査や考察を行っている．彼らの調査の結果，顔文字は，単純な極性にはわけられず，複雑な感情にわかれることや，顔文字自身の極性と文の極性が異なる場合が

†1 筑波大学 情報学群 知識情報・図書館学類

College of Knowledge and Library Sciences, School of Informatics, University of Tsukuba

†2 筑波大学 図書館情報メディア研究系

Faculty of Library, Information and Media Science, University of Tsukuba

あることが明らかとなった。川上ら¹⁰⁾の研究では、顔文字の表す感情について調査しており、1つの顔文字は複数感情を表す場合があることを示している。これらの研究では、研究対象を感情カテゴリに絞っており、感情以外のカテゴリは対象としていない。

一方、河野¹²⁾は、年代や性別による顔文字の使用理由について調査を行い、文字だけではそっけない画面をにぎやかにするため、自分の伝えたい感情の強調、文章の深刻さを和らげる、などがあることを明らかにした。加藤ら⁷⁾は、顔文字が感情を表すことを感情表現機能、顔文字を使うことで言葉での表現が減ることをメール本文代替機能と呼んでいる。彼らは、顔文字を使用することでメール本文の文字数が減ること、親しい間柄では感情表現機能、メール本文代替機能がより使用されることがわかっている。

加藤らの研究から、文の感情を顔文字が代替する機能、つまり文の感情を強調する機能を持っていることが明らかとなった一方、河野らの研究からは、画面をにぎやかにする、文章の感情を和らげる、といった機能を持つことがわかった。このことから、顔文字を推薦するためには、単純に文章の感情のみを手がかりとするだけでなく、文章に現れる特定の表現なども考慮したルールを設定しなければならない。本研究では、これらの傾向をより詳細な形で明らかにするため、顔文字を含む文章を分析し顔文字の使われ方を調べる。そしてその結果から、顔文字推薦を行う上で感情以外に必要な、コミュニケーションや動作を反映したカテゴリがあることを明らかにし、顔文字推薦システムを構築する。

また、顔文字ではないが、Twitterの文章に絵文字を推薦する研究として橋本の研究⁸⁾がある。橋本の研究では、まず入力文章を形態素解析し、それを単語3グラムに分割する。そして絵文字入りコーパスを用い、単語3グラムと類似するコーパス中の文で用いられている絵文字を文章に挿入している。また、ポジネガ推定を行ってから感情推定の2段階推定にすることで、精度が向上することが徳久らの研究¹³⁾から明らかになっている。

本研究では、Twitterの文章を収集し、コーパスを作成する。また、本研究では単語3グラムを用いるのではなく、入力文に含まれる感情語や特定の表現などを手がかりにコーパスから類似文を見つけ、感情やコミュニケーションのカテゴリ推定を行う。カテゴリ推定については、徳久ら¹³⁾にならい、2段階での推定を行う。その結果を用いて、作成した顔文字データベースから顔文字を推薦する。

3. 顔文字推薦システムに必要なデータの作成

本研究で提案する顔文字推薦システムは、大きく分けて感情推定と顔文字推薦の2種類の処理から構成される。本章では、それぞれの処理を行う際に必要となる情報が何かを分析

し、その結果必要となる基本情報を定義する。また、それぞれの処理に必要なデータについて説明を行う。

3.1 感情推定処理に用いるカテゴリの定義

顔文字推薦のための感情推定を行うにあたって、どのような感情カテゴリ、また、コミュニケーションや動作を反映したカテゴリが必要なかを明らかにするため、調査を行った。調査データには、きざしラボ¹⁴⁾が提供する顔文字92個を含むTwitterのつぶやき1,722件を収集し使用した。これらのデータに、極性(ポジティブ、ネガティブ、なし)の3種類、感情(中村⁵⁾「喜」「怒」「哀」「恥」「怖」「好」「厭」「昂」「安」「驚」の10種類)を分類した。また、10種類のカテゴリに分類することができず、よく現れるような表現の場合は、新たにカテゴリを定義した。

調査の結果、わかったことの例を、以下に示す。

- (1) 感情の種類についてわかったこと
 - (a) Twitterのつぶやきには、期待を表す表現や、不快を表す表現などを含む文章が多く見られた。しかし、中村の感情分類にはそれらを表す感情カテゴリはないため、新たに「期待」「不快」という感情カテゴリを定義した。
- (2) 感情以外の特徴についてわかったこと
 - (a) “m(_ _)m”は、「ごめんなさい」や「申し訳ない」などの謝罪表現を含む文に付与されることが多かった。しかし、謝罪を表すような感情カテゴリは中村の分類には含まれておらず、また、謝罪は感情とは言えない。そこで、感情分類とは別のカテゴリとして、新たに「謝罪」カテゴリを定義した。このような、誰かとのやり取りや、人に伝えることを前提としたカテゴリが多く存在したため、コミュニケーションタイプとして、新たに定義した。
 - (b) “(-_-)”や“(*_*)”は、睡眠を表す文章で使用されていた。睡眠は動作であり、感情ではない。感情カテゴリに当てはめることはできないので、これも新たに「睡眠」カテゴリを定義した。この動作を表すカテゴリを、動作タイプとして新たに定義した。

以上の結果に基づき、顔文字推薦に必要な感情、感情以外のカテゴリを決定した。感情の定義を表1に、感情以外の特徴をコミュニケーションタイプ、動作タイプとして定義した結果を表2に示す。

感情カテゴリは全部で27種類ある。中村の感情分類を基礎カテゴリとして用いて、それら27種類の感情を下位カテゴリとして10種類の基礎カテゴリに分類した。また、各下位

カテゴリごとに、極性 (ポジティブ/ネガティブ/なし) の設定を行った。

表 1 感情の定義

基礎感情	基礎感情の数	下位感情	下位感情の極性 (POS = ポジティブ, NEG = ネガティブ, - = なし)	下位感情の数
喜	1	嬉しい	POS	5
		めでたい	POS	
		快い	POS	
		楽しい	POS	
		満足	POS	
怒	1	怒り	NEG	1
哀	1	悲しい	NEG	4
		残念	NEG	
		悔しい	NEG	
		寂しい	NEG	
恥	1	恥ずかしい	NEG	1
怖	1	怖い	NEG	2
		不安	NEG	
好	1	好き	POS	2
		あこがれ	POS	
厭	1	嫌い	NEG	6
		困る	NEG	
		不快	NEG	
		憂鬱	NEG	
		苦しい	NEG	
		疲れ	NEG	
昂ぶり	1	感動	POS	4
		興奮	-	
		期待	POS	
		焦り	NEG	
安	1	安らぎ	-	1
驚	1	驚き	-	1
合計	10			27

コミュニケーション・動作タイプは全部で 10 種類ある。感情には分類できないが、表現としてよく現れるものを分類した。コミュニケーションタイプは、人とのやり取りや他者に物事を伝えることを中心とした表現のタイプであり、人にどのように伝えるかによって「やりとり型 (他者との会話に含まれる表現)」「つぶやき型 (他者に見られることを前提として

いるが、会話にはなっていない文章に含まれる表現)」「不特定多数型 (やりとり、つぶやきのどちらにも使われる表現)」の 3 つに分類しており、それらの分類の下に詳細カテゴリを設定した。また、感情カテゴリ、コミュニケーションタイプに分類できず、文章の中心が動作表現となるものを動作タイプとして定義した。現在、動作タイプとして定義できる程度に頻出した動作表現は「睡眠」だけであるため、動作タイプは 1 種類のみである。

表 2 コミュニケーション・動作タイプの定義

分類		詳細カテゴリ	カテゴリの数
コミュニケーションタイプ	やりとり型	感謝	5
		謝罪	
		心配	
		はげまし	
	不特定多数型	ねぎらい	3
		あいさつ	
ひとりごと型	願望	1	
	報告	1	
動作タイプ	-	意気込み	1
		睡眠	1
合計			10

3.2 必要なデータの作成

顔文字推薦に使用する顔文字は、予備実験で用いた 92 個では各感情を表す顔文字を推薦するのに十分な数ではなかったため、アンケートを実施し追加を行い、163 種類とした。

次に、前述のカテゴリ分類を用いて、感情 (コミュニケーションタイプ、動作タイプ) タグ付きコーパス (以下、感情タグ付きコーパス) を作成した。コーパスの文章は、Twitter の 2011 年 5 月 1 日 ~ 31 日、7 月 1 日 ~ 16 日のパブリックタイムライン上のツイートから、前述の顔文字 163 個を含む 1,369 件 (ツイート数) を収集し、それらの文章に、それぞれ (1) 極性、(2) 感情カテゴリ、コミュニケーションタイプまたは動作タイプのいずれかを付与し、作成した。

作成したコーパスの文章に付与したカテゴリがどの程度安定しているかを調べるため、協力者 1 名にコーパス中からランダムに選んだ 190 件について、(1) 極性と (2) 感情カテゴリ、コミュニケーションタイプまたは動作タイプのいずれかを付与させ一程度 (カッパ係数⁴⁾) を調べた。その結果、極性の一致については、 κ が、0.850 (almost perfect/ほとんど一致)、極性、感情カテゴリ、コミュニケーションタイプまたは動作タイプの一致につい

ては、 κ が、0.747 (substantial/かなり一致) となった。

また、日常的に顔文字を使用する 20 代の協力者 2 名に、作成したコーパスの各文章について、(1) 極性と (2) 感情カテゴリまたはコミュニケーションタイプまたは動作タイプのいずれかを付与させた。その結果をもとにコーパスの修正と拡張を行い、3,887 件 (各カテゴリ平均 100 件) となった。

次に、前述のカテゴリ分類を用いて、手がかり語辞書を作成した。手がかり語辞書は感情表現辞典⁵⁾、単語感情極性対応表⁹⁾を参考にし、全 1,440 語の項目を設定した。それぞれの語に極性 (ポジティブ、ネガティブ、なしのいずれか) と感情カテゴリ、コミュニケーションタイプまたは動作タイプのいずれかを付与した。以下、感情カテゴリを付与した語を感情語、コミュニケーションタイプを付与した語をコミュニケーション語、動作タイプを付与した語を動作語、とする。

顔文字推薦処理を実装するために、収集した文章に含まれる顔文字の使われ方を分析し、顔文字データベースの作成を行った。作成した感情タグ付きコーパスには顔文字が含まれているため、それを分析することで各顔文字の感情ごとの出現頻度を調べた。

また、単純なコーパス中での出現頻度以外に、同じカテゴリに分類される顔文字でも、文章中の表現の違いで使用のされ方が違う場合には、補足ルールを作成し、補足ルールにおいて設定した下位カテゴリ別の出現頻度も調べた。補足ルールは 2 つあり、以下のとおりである。

(1) 「あいさつ」カテゴリ

「あいさつ」カテゴリでは、同じあいさつに関する表現であっても、どの時間帯のあいさつなのか (例: おはよう “()”, おやすみ “((- -)...zzzzZZ)”, あるいは時間帯に関係ないあいさつ (例: よろしく “(^)”, さようなら “(^)”) なのかによっても、使用頻度の高い顔文字が変わってくる。そこで、「あいさつ」カテゴリでは、あいさつ表現別に 9 つの下位カテゴリを作成した。各顔文字がどの分類に当てはまるのかを分析し、それぞれの分類ごとに出現頻度を計算した。

(2) 「感謝」カテゴリ

「感謝」カテゴリでは、丁寧な口調の表現を含む場合、泣き顔や申し訳なさを表す顔文字 (例: “\ (; ;) /”) がより使用されていた。そこで、「ございます」「ありがとうございます」などの言葉を含む文章について、顔文字ごとの頻度を計算した。

また、入力文章がこれらの補足ルールに一致する語を含んでいるかを判断する必要があるため、補足ルールの下位カテゴリごとに、それらを表す表現を集め、補足ルール辞書として

作成し、30 語を設定した。

データベースには、1 感情と 1 顔文字を組み合わせるとし、それに加え、コーパス中での使用頻度、補足ルール別の使用頻度を 1 レコードとして格納している。

4. 顔文字推薦システムの実装

構築した顔文字推薦システムは、感情推定処理と顔文字推薦処理の 2 つから成り立っている。以下に、顔文字推薦システム全体の流れを示す。

- (1) ユーザが文章を入力する。
- (2) 入力文章を用いて、感情推定を行う (感情推定処理)。
- (3) 感情推定の結果を用いて、顔文字データベースから適切な顔文字を取り出す (顔文字推薦処理)。
- (4) 取り出した 5 件の顔文字を推薦候補として、画面に表示する。
2 つの処理について、以下に示す。

4.1 感情推定処理

感情推定を行うために、分類器を構築した。分類器には k -NN を用い、文章を入力すると、各学習データとの類似度を計算し、類似している上位 k 件の中で最も多い感情を推定結果として返す。

学習データには、3.2 節で説明した、感情タグ付きコーパス 3,887 件を用いた。

分類器による感情推定の手順を以下に示す。

- (1) 極性を推定する。
 - (a) まず、手がかり語辞書に含まれる語の、入力文における出現頻度を計算する。
 - (b) (a) の結果から得られた各手がかり語の頻度を用いて、極性ごとに素性ベクトルを構築する。
 - (c) 学習データ中の各文との平方距離を計算する。
 - (d) 平方距離の近い上位 k 件のうち、学習データにおいて最も多く付与されている極性を推定結果とする。
- (2) 推定した極性の下位カテゴリとして、感情、コミュニケーションタイプ または動作タイプのいずれかを推定する。
 - (a) 素性ベクトルの構築や各学習データとの距離の計算は、1 の (a)-(c) と同様に推定する。
 - (b) 平方距離の近い上位 k 件のうち、学習データにおいて最も多く付与されてい

る感情詳細カテゴリ, コミュニケーションタイプまたは動作タイプのいずれかを推定結果とする。

また, 入力文に複数の感情語 (コミュニケーション語, 動作語) が含まれる場合, その位置関係によって間違った結果が推定されてしまう恐れがある。

たとえば「おかえり～今日はありがとね」という文には「おかえり」と「ありがとう」というコミュニケーション語が含まれている。この文の場合では, 文全体のコミュニケーションタイプは「感謝」になるのだが「あいさつ」が推定結果として出力されてしまう可能性がある。各カテゴリの出現頻度を計算して推定すると, どちらも同じ頻度なため, どちらが推定されてもおかしくない。これは, 文章内に現れる感情語 (コミュニケーション語, 動作語) の位置関係を考慮していないためであると考えられる。この対策として, 入力文内に現れる順序によって, 感情語 (コミュニケーション語, 動作語) に重みを付与する, といった処理を追加した。具体的には, 以下のような計算を行う。なお, $word_n$ は入力文内に n 番目に現れる感情語 (コミュニケーション語, 動作語), N は入力文内に現れる感情語 (コミュニケーション語, 動作語) の総数である。

$$weight(word_n) = \left[\frac{n}{N} \right]$$

次に, 構築した分類器がどの程度正確に感情を推定できるのかを確かめるため, また, 重み付け処理を行うことで, 推定の正確さにどの程度差が出るのかを調べるため, 実験を行った。テストデータには 3.2 節で記述した, 一致率の算出に用いた 190 件を使用する。これらのデータには, あらかじめ人手で極性と感情, コミュニケーションタイプまたは動作タイプの正解を付与した。

実験の結果, 重み付け処理を行っていない場合では, 極性のみが正解していたものが 73.6%(140/190 件), 極性とカテゴリが正解していたものが 51.5%(98/190 件) であった。一方, 重みづけ処理を行った場合では, 極性のみが正解していたものが 76.8%(146/190 件), 極性とカテゴリの両方が正解していたものが 53.1%(101/190 件) であり, 精度が向上した。

4.2 顔文字推薦処理

次に, 顔文字推薦処理の実装を行った。顔文字推薦処理では, 感情推定の結果を受け取り, その結果に適切な顔文字を顔文字データベースから取り出し, 最高 5 件を推薦する。また, もし感情推定の結果が補足ルールに一致するカテゴリであれば, 補足ルール用辞書を用いてさらに細かい分類を行い, 適合した顔文字を推薦する。補足ルールに当てはまらないカテゴリであれば, 感情推定の結果だけを用いて, 顔文字データベースから顔文字を取り出す。

顔文字推薦処理には, 3.2 節で説明した, 顔文字データベースと, 補足ルールを適用する

ための補足ルール用辞書を用いた。

5. 顔文字推薦システムの評価

顔文字推薦システムを実際に被験者に使わせ, 推薦する顔文字がどの程度ユーザの入力した文章に対して適切であるか調べるため, 被験者実験を行った。被験者は, 日常的に顔文字を使用する 20～22 歳の男性 2 名, 女性 3 名である。

5.1 実験内容

実験は大きく 2 つの設問から構成される。

実験 1 では, あらかじめこちらで文章を用意し, その文章に対して推薦される顔文字が適切かどうかを回答させた。また, 1 つの文章につき, (a) 感情, コミュニケーションタイプ, 動作タイプのいずれかを推定したカテゴリに対して, 推薦した顔文字 (提案手法), (b) 分類器の学習データを感情カテゴリのみに限定して, 推薦した顔文字 (ベースライン) のそれぞれに回答をさせた。なお, 文章は全部で 91 個あり, 3.2 節で記述した, 一致率の算出に用いた 190 件から, 各感情カテゴリ, コミュニケーションタイプ, 動作タイプにつき 1-3 個の文章をランダムに抜き出し使用した。

実験 2 では, 被験者に実際に顔文字推薦システムを使用させた。被験者自身が, 実際に Twitter や電子メールで使用しそうな文章を考え, システムに入力し, その結果推薦された顔文字が適切かどうかを回答させた。文章は全部で 20 文入力させ, 10 文はポジティブな文章, 10 文はネガティブな文章を考えさせた。

また, 実験後に, 適切でなかった顔文字とその理由, 自由感想を記入させた。

5.2 実験結果

実験 1, 2 の結果を, 表 3, 表 4 に示す。実験 1 の推薦精度は, カテゴリごとの正解数を分母としている。実験 2 の推薦精度は, 入力したポジティブ, ネガティブの文章数を分母としている。実験 1 では, t -検定を用いた結果, 提案手法は感情カテゴリのみに基づいて推薦したベースラインと比べて, 推薦精度が有意に向上していることが分かった (有意水準 5%, 両側検定)。

5.3 実験の考察

実験 1, 2 の結果に基づく考察を以下に示す。

提案手法は, ベースラインと比べて, コミュニケーションタイプ, 動作タイプのみならず, 感情カテゴリにおいても精度の向上が見られる。これは, コミュニケーションタイプや動作タイプを導入することが顔文字推薦において有用であるだけでなく, 感情に基づく顔文字推

表 3 Twitter を対象とした顔文字推薦の評価

文章のカテゴリ	提案手法	ベースライン
感情カテゴリ	69.3%	67.9%
コミュニケーションタイプ	63.6%	53.3%
動作タイプ	72.0%	64.0%
全体	66.7%*	62.8%

表 4 自由入力に基づく顔文字推薦の評価

	適しているとの回答
ポジティブな文章についての推薦	58.8%
ネガティブな文章についての推薦	78.0%
全体	68.4%

薦についても、より適切な顔文字を推薦できることを示唆している。推薦精度が大きく向上したカテゴリは、コミュニケーションタイプでは、“ねぎらい”、“あいさつ”、“心配”、“感謝”、“謝罪”などであり、感情カテゴリでは、“不安”、“安らぎ”、“期待”などであった。このうち“あいさつ”について内容を見ると、「こんにちは」などは曖昧な表現であり、ポジティブな顔文字であれば、カテゴリ推定が間違っている場合、ユーザの要求に適合する可能性はある。しかし、「よろしく〜」、「おやすみ」などの挨拶表現や、“睡眠”、“謝罪”のように、特定の顔文字が文章にぴったり適合するカテゴリの場合、コミュニケーション・動作タイプを推定し、顔文字を推薦することが有効である。

例として「昼からすみません」という文に対する顔文字を推薦する場合を考える。この文章は“謝罪”を表しているが、感情カテゴリのみを推定するベースラインシステムでは、別の感情（この場合、“興奮”）の推定が行われる。その結果、下記のような顔文字が推薦される。

∖(^o^)/ (*´Д`) (*´д`*) (*´▽`*) (´;ω;`)

この中で2~4番目の顔文字では、謝罪とは全く逆の印象を与えてしまう顔文字が推薦されてしまう。

しかし、コミュニケーションタイプを推定できる提案手法では、“謝罪”を推定することができ、

m(_)_m orz (´;ω;`) (´・ω・`) (^_^;)

のように、5件とも文章の意図に適合した顔文字を推薦している。

適切でないと回答が多くあったのは、システムの推定した感情が、人手で付与したものと異なる場合である。特に正反対の極性のカテゴリが推定されていると、推薦された顔文字候補はまったく役に立たないこととなる。このため、推定がうまくいかなかったコミュニケーションタイプの文章の顔文字推薦では、良い結果が得られなかった。よって、システムの精度を上げるためには、カテゴリ推定の精度向上の必要が考えられる。

また、実験後アンケートには、「ネガティブな文章に対する顔文字推薦の精度がよい」とあった。これは、ポジティブな文章よりもネガティブな文章の方が、直接的な手がかり語が文中に現れやすいため、カテゴリを推薦しやすいことが原因と考えられる。

同じく実験後アンケートに「同じカテゴリの文章に対して、表示される顔文字の種類・順序が一樣である」とあった。現在、手がかり語辞書に収録している、同じカテゴリの各手がかり語には、一意に極性が付与されている。そのため、含まれている語が異なる場合でも、推定結果に違いがなく、同じ顔文字が推薦される。この点については、同じカテゴリの手がかり語でも、極性を考慮して手がかり語辞書を作成すること、また、極性を考慮した顔文字データベースを作成することにより、表示される顔文字の種類・順序を豊富にするだけでなく、顔文字推薦自体の精度向上にもつながると期待できる。

6. おわりに

本研究では、ユーザの顔文字選択の支援を目的とし、ユーザの入力文から感情や、感情以外のコミュニケーションを反映したカテゴリを推定し、顔文字を推薦するシステムを構築した。システムを構築するにあたり、新たにコミュニケーションタイプ、動作タイプというカテゴリを定義した。

システムは大きく2つの処理にわかれており、感情推定処理と顔文字推薦処理から構成されている。

感情推定処理については、 k -NN を用いて分類器を構築した。すなわち、文章を入力すると、手がかり語辞書内の語が含まれていないかを調べ、各カテゴリの語の頻度から構築した素性ベクトルを用いて、各学習データとの類似度を計算し、類似度の高い上位 k 件から推定結果を決定する。実装した感情推定処理の精度を実験で確かめたところ、53.1%であった。

顔文字推薦処理では、感情推定の結果を用いて、顔文字データベースから適切な顔文字を取り出し、推薦する。また、コミュニケーションタイプについて、各カテゴリより更に細かい単位で補足ルールを作成し、入力文の推定結果が補足ルールに当てはまる場合は、補足ルールも用いて顔文字推薦を行う。

構築した顔文字推薦システムの推薦がどの程度正確なのか、被験者実験を行った。その結果、66.7%の顔文字が文章に対して適切に推薦されており、感情カテゴリのみの推定に基づく推薦に比べて、有意に向上していることを明らかにした。

今後の課題としては、活用形などの手がかりも考慮したうえで、手がかり語辞書を拡張すること、極性を考慮した顔文字データベース作りを行うことで、精度向上が期待できる。

参 考 文 献

- 1) Derks, D., Bos, A. E.R. and von Grumbkow, J.: Emoticons and Social Interaction on the Internet: The Importance of Social Context, *Computer in Human Behavior*, Vol.23, pp.842-849 (2007).
- 2) Ptaszynski, M., Marciejewski, J., Dybara, P., Rzepka, R. and Araki, K.: CAO: A Fully Automatic Emotion Analysis System, *Proc. of the 24th AAAI Conf. on Artificial Intelligence (AAAI-2010)*, pp.1026-1032 (2010).
- 3) Ptaszynski, M.: 顔文字処理-取るに足らない表現をコンピュータに理解させるに足るには-, *情報処理*, Vol.53, No.3, pp.204-210 (2012).
- 4) Teufel, S. and Moens, M.: Summarizing Scientific Articles: Experiments with Relevance and Rhetorical Status, *Computational Linguistics*, Vol.28, No.4, pp.409-445 (2002).
- 5) 中村 明: 感情表現辞典, 東京堂出版, 初版 edition (1993).
- 6) 村上浩司, 山田 薫, 萩原正人: 顔文字情報と文の評価表現の関連性についての一考察, *言語処理学会 第 17 回年次大会 発表論文集*, pp.1155-1158 (2011).
- 7) 加藤尚吾, 加藤由樹, 島峯ゆり, 柳沢昌義: 携帯メールコミュニケーションにおける顔文字の機能に関する分析-相手との親しさの程度による影響の検討-, *日本教育情報学会学会誌*, Vol.24, No.2, pp.47-55 (2008).
- 8) 橋本泰一: Twitter への絵文字自動挿入システム, *言語処理学会 第 17 回年次大会発表論文集*, pp.1151-1154 (2011).
- 9) 高村大也, 乾 孝司, 奥村 学: スピンモデルによる単語の感情極性抽出, *情報処理学会論文誌*, Vol.47, No.2, pp.627-637 (2006).
- 10) 川上正浩: 顔文字が表す感情と強調に関するデータベース, *大阪樟蔭女子大学人間科学研究紀要*, Vol.7, pp.67-82 (2008).
- 11) 篠山 学, 松尾朋子: 顔文字を考慮した対話テキストの感情推定に関する研究, *香川高等専門学校研究紀要*, Vol.1, pp.51-53 (2010).

- 12) 河野道子: フェイスマークを用いた感情表現におけるコミュニケーション・ギャップに関する研究 (2003). <http://www.sonoda-u.ac.jp/dic/kenkyu/2003/14.pdf>.
- 13) 徳久良子, 乾健太郎, 松本裕治: Web から獲得した感情生起要因コーパスに基づく感情推定, *情報処理学会論文誌*, Vol.50, No.4, pp.1365-1374 (2009).
- 14) 株式会社きざしカンパニー: kizashi.jp:きざしラボ (2011).
<http://kizasi.jp/labolo/lets/wish.html>.