

ユーザの嗜好を考慮した顔文字推薦システムの提案

Proposal of Facemark Recommendation System Considering User's Preference

松原翔平[†]
Shohei Matsubara

森田和宏[†]
Kazuhiro Morita

泓田正雄[†]
Masao Fuketa

青江順一[†]
Jun-ichi Aoe

1. はじめに

現在、電子メール、TwitterやFacebook等のソーシャルネットワークサービス(SNS)においてテキストベースのコミュニケーションが日常的におこなわれている。

テキストベースのコミュニケーションに見られる特徴として、顔文字の利用が挙げられる。顔文字を利用することで、文章だけでは表現できない微妙なニュアンスを文章に含ませることや、受け手側の感情を促進し、より高度な情報伝達をおこなうことが可能になる[1]。また、顔文字は誰でも容易に作成することが可能であるため、顔文字の種類は現在も増加の一途をたどっており、非常に多くの種類が存在している。例えば顔文字配信サイトの1つである「顔文字辞典 The Facemark Dictionary」[2]では58,130種類の顔文字が登録されている。ユーザはこれらの顔文字の中から、ユーザの好みや伝えたい感情に合ったものを引用し、文章への挿入をおこなう必要がある。

こうしたユーザが顔文字を引用・挿入する際の負担を軽減することを目的とし、本研究ではTwitterに投稿された顔文字を含むツイートを用いて、自動で顔文字データベースを作成し、ユーザの嗜好を考慮した推薦をおこなう。

2. 関連研究

卜部ら[3]は顔文字の持つ感性をアンケート調査で決定し、入力文が持つ感性に最も近い感性をもつ顔文字の推薦をおこなっている。江村ら[4]はツイートに含まれる感性表現を用いて顔文字を感性分類し、入力文の感性に一致する顔文字の推薦をおこなっている。

顔文字の中には同種の感性を持つものが多数存在する。関連研究ではこのような同種の感性を持つ顔文字に対して感性の類似度やコーパス内の出現頻度で推薦における優先度の決定をおこなっている。しかし、ユーザは自分の好みによって顔文字を選択しているため、優先度にはユーザの嗜好を反映させることが有効であると考えられる。本研究では、ユーザの過去の顔文字使用履歴に対して協調フィルタリング[5]を用いることでユーザの嗜好の予測をおこなっている。また、卜部らや江村らの研究では顔文字と顔文字が持つ感性情報を登録した顔文字データベースの作成を人手でおこなっている。そのため、扱う顔文字の種類が多い場合、作成における負担が大きくなると考えられる。そこで、本研究では顔文字データベース作成の自動化をおこなっている。

3. 顔文字データベース

入力文が持つ感性情報に合致した顔文字を推薦するため、顔文字データベースの作成をおこなった。

[†]徳島大学, The University of Tokushima

	\(^o^)/	(;^▽^)	(´ω´)
ユーザ A	2.5	1.5	0.5
ユーザ B	0	0.3	1.5
ユーザ C	0.7	0.3	1.5

図1 評価値行列の例

3.1 ツイートの収集

顔文字を含むツイートの収集をおこなう。本研究ではTwitterに顔文字を含むツイートをおこなっているユーザを対象に、1人あたり200件ずつツイートの収集をおこなった。

3.2 顔文字辞書の作成

3.1節で得られたツイート群に対して、特定の記号や表記パターンで囲まれた部分を検出し、検出された箇所が顔文字であれば顔文字辞書へ登録する。検出には左右で異なる特徴を利用した。左側は“(”や“\”など40種類、右側は“)”や“)”など65種類の特徴を利用した。

3.3 顔文字データベースの作成

3.1節で得られたツイート群、3.2節で作成した顔文字辞書、吉成らが作成した感性理解エンジン[6]を用いて顔文字データベースを作成する。感性理解エンジンでは文書に含まれる感性を解析することができ、マーケティング調査や音声合成に利用されている。本研究では感性理解エンジンで定義される81種類の感性分類を顔文字が持つ感性情報として用いた。顔文字データベースへは顔文字の表記と感性情報を登録する。感性情報は、顔文字の出現する直前の文節を感性理解エンジンで解析する事によって取得する。また、卜部らや江村らの研究から複数の感性を持つ顔文字が存在することが考えられるため、感性の判定結果が複数存在した場合は、すべての感性情報を登録する。

4. 提案手法

3章で述べた顔文字データベースと協調フィルタリングを用いて顔文字の推薦をおこなう手法について説明する。

以下の式で各ユーザの顔文字 k に対する評価値を算出し、図1のような評価値行列を作成する。

$$\text{顔文字}k\text{の評価値} = \frac{\text{顔文字}k\text{の利用回数}}{\text{全顔文字利用回数の平均値}}$$

次に、推薦を受けるユーザは周辺のユーザとの類似度の算出をおこなう。協調フィルタリングの類似度の計算にはPearson相関を用いる。ユーザ a とユーザ i との類似度 ρ_{ai} の計算式を次に示す。

$$\rho_{ai} = \frac{\sum_{k \in y_{ai}} (s_{ak} - \bar{s}'_a)(s_{ik} - \bar{s}'_i)}{\sqrt{\sum_{k \in y_{ai}} (s_{ak} - \bar{s}'_a)^2} \sqrt{\sum_{k \in y_{ai}} (s_{ik} - \bar{s}'_i)^2}}$$

y_{ai} : a と i が共通で評価している顔文字の集合
 s_{ak} : a の顔文字 k に対する評価値
 s_{ik} : i の顔文字 k に対する評価値
 \bar{s}'_a : a からみた, y_{ai} に対する評価値の平均
 \bar{s}'_i : i からみた, y_{ai} に対する評価値の平均

図 1 の評価値行列で, 推薦を受けるユーザを B とした場合, AB の類似度はおよそ 0.39, BC の類似度は 1 となり, C の方が B と類似しているという予測が得られる.

次に, 評価値が 0 である部分に対して評価値の予測をおこなう. ユーザ a の顔文字 j に対する評価値の予測値 \hat{s}_{aj} は, 以下の式で求める.

$$\hat{s}_{aj} = \bar{s}_a + \frac{\sum_{i \in X_j} \rho_{ai} (s_{ij} - \bar{s}'_i)}{\sum_{i \in X_j} \rho_{ai}}$$

\bar{s}_a : a の評価済み顔文字に対する評価値の平均値
 X_j : 顔文字 j を評価しているユーザの集合

図 1 の評価値行列に対して, 推薦を受けるユーザを B として以上の式を適用すると B の顔文字 “\(^{\circ}\wedge^{\circ}\)” に対する評価値はおよそ 0.34 であり, 平均の利用回数よりも低くなると予測される.

次に, ユーザの入力文に対して形態素解析をおこない, 文節単位に分割する. 各文節に対して感性理解エンジンを用いて, 感性表現を含む文節であれば, 文節の直後を顔文字挿入位置の候補とする. 入力文の解析で得られた感性情報に一致する感性情報を持つ顔文字のリストを顔文字データベースより取得する. ユーザの使用履歴に基づく評価値と, 予測で得られた評価値を基準に並べ替え, 顔文字挿入位置の候補を付与した入力文と共に, ユーザへ提示する.

5. 実験

5.1 実験設定

Twitter に顔文字を含むツイートを投稿しており, 且つ顔文字利用傾向が異なるユーザ 2 人を被験者として選んだ. 顔文字データベースの作成には 2,388 人分のツイートと 3,965 種類の顔文字を使用した. 入力文には顔文字を感性分類した結果, 件数の多かった “楽しい” “悪い” の感性表現を含む文を使用した. また, 収集したツイート群における出現頻度を利用した推薦結果を比較データとして用いる. 実験結果として推薦された顔文字のリストを示す.

5.2 結果と考察

表 1, 表 2 に “楽しい” の推薦結果を示す. 出現頻度を用いる手法では, 推薦結果がどのユーザに対しても同一であったが, 提案手法ではユーザ毎に異なる推薦結果を提示できた. 提案手法によって推薦された顔文字リストにはユーザの利用頻度が高い顔文字が多く含まれており, 出現頻度を用いた手法に比べ, ユーザの嗜好が強く反映されていると考えられる. 表 3 に示した “悪い” の推薦結果ではユーザ A がポジティブな意味で利用している “\(^{\circ}\wedge^{\circ}\)” や “\(^{\circ}\wedge^{\circ}\)” が “悪い” で推薦されてしまっていた. これは少数のユーザが “悪い” で利用していたことが原因だと考えられる. そのため, 感性情報の登録に一定の閾値を設け, 少数派の表現を制限することで, このような推薦を防ぐことができると思われる.

表 1 ユーザ A に対する “楽しい” の推薦結果

提案手法		出現頻度	
顔文字	評価値	顔文字	回数
$(*\ ^{\circ}\wedge^{\circ} *)$	6.53	\(^{\circ}\wedge^{\circ}\)	155
\(^{\circ}\wedge^{\circ}\)	4.90	\(((o(*\ \nabla^{\circ} *)o)))	128
\(^{\circ}\wedge^{\circ}\)	4.46	\(*\ (^{\circ}\wedge^{\circ}) *\)	83
\(-^{\circ}\wedge^{\circ}-\)	3.79	\(*\ ^^*\)	74
\(\nabla^{\circ}\wedge^{\circ}\)	3.75	\(*\ ^{\circ}\wedge^{\circ} *\)	58

(網掛け部: ユーザ A の使用履歴がある顔文字)

表 2 ユーザ B に対する “楽しい” の推薦結果

提案手法		出現頻度	
顔文字	評価値	顔文字	回数
\(^{\circ}\wedge^{\circ}\)	9.10	\(^{\circ}\wedge^{\circ}\)	155
\(*\ ^^*\)	6.56	\(((o(*\ \nabla^{\circ} *)o)))	128
\(\nabla^{\circ}\wedge^{\circ}\)	4.01	\(*\ (^{\circ}\wedge^{\circ}) *\)	83
\(\nabla^{\circ}\wedge^{\circ}\)	3.80	\(*\ ^^*\)	74
\(*\ ^{\circ}\wedge^{\circ} *\)	3.69	\(*\ ^{\circ}\wedge^{\circ} *\)	58

(網掛け部: ユーザ B の使用履歴がある顔文字)

表 3 ユーザ A に対する “悪い” の推薦結果

提案手法		出現頻度	
顔文字	評価値	顔文字	回数
\(*\ ^{\circ}\wedge^{\circ} *)	6.53	\(>_<)	246
\(^{\circ}\wedge^{\circ}\)	4.90	\(\cdot\ \cdot\ \cdot\)	198
\(\nabla^{\circ}\wedge^{\circ}\)	4.66	\(^{\circ}\wedge^{\circ}\)	182
\(^{\circ}\wedge^{\circ}\)	4.46	\(\cdot\ \cdot\ \cdot\)	179
\(_{3}\ \nabla^{\circ}\)	3.69	\(\cdot\ \cdot\)	142

(網掛け部: ユーザ A の使用履歴がある顔文字)

6. おわりに

研究では, 協調フィルタリングを用いて, ユーザの嗜好を考慮した顔文字を推薦するシステムの手法を提案した. 実験によりユーザ毎に推薦結果が変化することが確認できた. 今後はユーザの嗜好に合致しているかどうかの評価及び, 推薦のためのインターフェースの開発に取り組んでいきたい.

参考文献

- [1] 竹原卓真, “世代の違いによる顔文字の感情認識効果”, 日本顔学会誌, Vol.7, pp.37-45, (2007)
- [2] <http://kaomoji.sakura.ne.jp/>
- [3] 卜部有記, ジェブカラファウ, 荒木健治, “顔文字の表す感情を用いた顔文字推薦システムの構築”, 言語処理学会第 19 回年次大会, pp.648-651, (2013)
- [4] 江村優花, 関洋平, “マイクロブログにおける感情・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”, 自然言語処理, Vol.19 No.5, (2012)
- [5] 神嶋敏弘, “推薦システムのアルゴリズム”, 人工知能学会誌, Vol.22 No.6 ~ Vol.23 No.2, (2007-2008)
- [6] Tomoko Yoshinari, Atlam EL-Sayed, Kazuhiro Morita, Kumiko Kiyoi and Jun-ichi Aoe : Automatic acquisition for sensibility knowledge using co-occurrence relation, International Journal of Computer Applications in Technology, Vol.33, No.2/3, pp.218-225, (2008)