

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

徳久良子^{†1,†2} 乾健太郎^{†2} 松本裕治^{†2}

本論では、ユーザ発話から、嬉しい、楽しい、悲しいなど 10 種類の感情を推定する手法を提案する。まず準備として我々は、Web 上の明示的な感情表現を利用して人が感情を生起する要因を自動獲得した。その結果、約 130 万件の感情生起要因が獲得された。次に、これらの事例を利用して、下記の 2 段階で感情を推定する。まず、第 1 段階として、入力されたユーザ発話の感情極性を推定する。次に、第 2 段階として、あらかじめ推定した感情極性値と大量の感情生起要因とを利用して、ユーザ発話の感情を推定する。評価実験の結果、2 段階の感情推定手法（感情極性を推定した後に感情を推定する手法）は、1 段階の感情推定手法（直接感情を推定する手法）に比べて精度が高いことが分かった。特に、2 段階の感情推定により、対話システムで致命的な誤応答の原因となる感情極性推定の誤りが有意に低減されることが分かった。また、感情語との相互情報量を用いて感情を推定する従来手法と比較した結果、我々の提案手法が有意に高い精度であることが確認された。

Emotion Classification Using Massive Examples Extracted from the Web

RYOKO TOKUHISA,^{†1,†2} KENTARO INUI^{†2}
and YUJI MATSUMOTO^{†2}

In this paper, we propose a data-oriented method for inferring the emotion of a speaker conversing with a dialog system from the semantic content of his/her utterance. We first fully automatically obtain a huge collection of emotion-provoking event instances from the Web. About 1.3 million emotion provoking event instances are extracted using an emotion lexicon and lexical patterns. We then decompose the emotion classification task into two sub-steps: sentiment polarity classification (coarse-grained emotion classification), and emotion classification (fine-grained emotion classification). For each subtask, the collection of emotion-provoking event instances is used as labelled examples to train a classifier. The results of our experiments indicate that our method significantly outperforms the baseline method. We also find that compared with the single-

step model, which applies the emotion classifier directly to inputs, our two-step model significantly reduces sentiment polarity errors, which are considered fatal errors in real dialog applications.

1. はじめに

1.1 背景と目的

従来の対話研究の多くは、対話を通して何らかのタスクを遂行する課題遂行型の対話を対象としてきた¹⁾。しかし、近年、課題遂行型ではなく話すこと自体が目的であるような対話システムが開発され、このような対話では感情的な応答が重要であることが指摘されている^{2),3)}。たとえば、1 人暮らしの高齢者などの話し相手となる対話システムでは、ユーザの感情を理解し共感的な応答を生成することが重要となる。以下に具体例で説明する。下記の 3 つの発話は、我々が構築しているプロトタイプシステムへの入力である。

- a) 遠くのレストランに行ったら定休日だった
- b) 大切にしていたハムスターが死んでしまったの
- c) 家の鍵を掛け忘れたかもしれない

これらのユーザ発話に対して、対話システムがユーザ発話の背景にあるユーザの感情を推定して応答すると、a) に対しては「それは残念でしたね」、b) に対しては「それは寂しいでしょう」、c) に対しては「それは心配ですね」のように応答することができる。

音声情報や表情を用いてユーザの感情を推定する手法は数多く提案されているが、現在の技術では、ユーザは通常より誇張して感情を表出しなければシステムはユーザの感情を正しく推定できない^{4),5)}。しかし、ユーザの発話内容自体を感情推定に相補的に利用できれば、ユーザが明確に感情を表現しない場合でもユーザの感情を適切に推定できる可能性がある。

本論では、対話に応用することを目的として、ユーザ発話の内容から感情を推定する手法を提案する。

1.2 基本的なアイデア

ユーザ発話から感情を推定する問題は、ユーザ発話のある感情クラスに分類するという、

^{†1} 豊田中央研究所
TOYOTA CENTRAL R&D LABS., INC.

^{†2} 奈良先端科学技術大学院大学
NARA INSTITUTE of SCIENCE and TECHNOLOGY

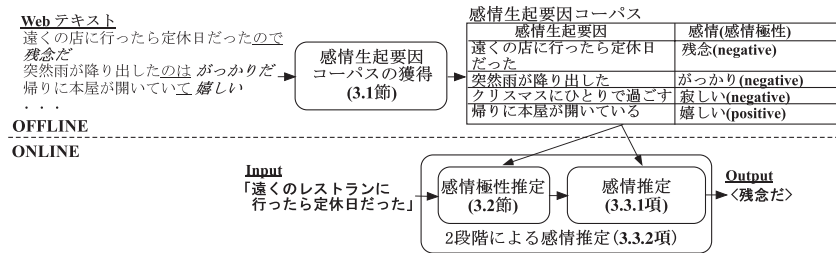


図 1 感情推定の基本的なアプローチ

Fig. 1 Overview of our approach to emotion classification.

いわゆる分類問題ととらえられる．本タスクを分類問題と考えると，仮に十分な量のラベルつき学習データが存在すれば，機械学習により感情クラスに分類できると期待される．そこで我々は，人がどのようなときに嬉しいと感じるか，どのようなときに悲しいと感じるかといった感情生起に関する事例を Web から抽出し，それらを用いてユーザ発話の感情を推定する．図 1 を用いて，我々の提案する感情推定の基本的なアプローチを説明する．

まず，図 1 の上半分は，ある感情が生起する要因を獲得するプロセスを示す．たとえば「遠くの店に行ったら定休日だったので残念だ」という文から，〈残念だ〉という感情が生起する要因として { 遠くの店に行ったら定休日だった } が獲得できる（本論ではこの事象の集合を感情生起要因コーパスと呼ぶ）．

次に，感情生起要因コーパスを利用して，ユーザ発話の感情を推定する．図 1 の下半分は感情生起要因コーパスを用いた感情推定のプロセスを示す．図 1 に示すとおり，本論では入力発話の感情極性（ポジティブ/ネガティブ/ニュートラル）をあらかじめ推定してから，ポジティブまたはネガティブの事例に対してより詳細な感情を推定するという 2 段階による感情推定手法を提案する．この 2 段階の感情推定には次の 2 つの利点がある．まず 1 つめは分類器としての利点である．一般に，多クラスの分類問題では上位クラスを分類した後に下位クラスを分類した方が分類精度が良いことが知られている．したがって，あらかじめ感情極性（粒度の粗い感情）を分類した後に，粒度の細かい感情を分類すると，良い精度が得られると期待される．2 つめは対話システムとしての利点である．対話システムがユーザの感情極性の推定を誤ると，対話システムは非常に致命的な誤応答を生成してしまう．具体的には，「大切にしていたハムスターが死んでしまったの」というユーザ発話に対して，対話システムが「それは愉快だね」と応答することを考えれば，その誤りの重大さは容易に想像できるであろう．あらかじめ高い精度で感情極性を推定し，その後に粒度の細かい感情を

推定することで，このような致命的な誤応答を避けられると期待する．

本論では，上記で述べた方法を実現し，以下の 3 つの課題を明らかにする．

- (1) 感情生起要因コーパスの規模と精度
- (2) 感情生起要因コーパスを用いた感情極性推定の手法と精度
- (3) 感情生起要因コーパスを用いた感情推定の手法と精度

まず，(1) では，実際に Web テキストから感情生起要因を獲得し，どのような規模の感情生起要因コーパスがどのような精度で獲得できるかを調べる．感情生起要因コーパスの構築方法を 3.1 節で，評価結果を 4.1 節で述べる．次に，(2) については，感情生起要因コーパスを利用した感情極性の推定方法を 3.2 節で，評価結果を 4.2 節で述べる．特に，感情極性推定に有効な特徴量，および，コーパスの規模との関係について調べる．さらに，(3) については，感情生起要因コーパスを用いて感情極性より粒度の細かい感情を推定する手法を提案する．3.3.1 項で基本的な感情推定手法を，3.3.2 項で感情極性推定の後に感情推定を行うという 2 段階による感情推定を提案し，4.3 節で評価結果を述べる．

2. 関連研究

近年，ユーザの感情が推定できる対話システムがいくつか報告されている⁵⁾⁻⁷⁾．ITSPOKE は，音声情報や言語情報などを利用して，チュータリング対話における学習者の感情を推定する．しかし，感情を学習者の理解度の指標として用いているため，彼らの対象としている感情は *Uncertain* および *Certain* の 2 値である．しかし，我々の目指すユーザに共感する対話システムへの応用を考えると，さらに細かい粒度で感情を推定する必要がある．

対話以外の研究に目を向けると，近年，感情極性（ポジティブ/ネガティブ/ニュートラル）の推定手法が数多く提案されている．しかし，感情極性より粒度の細かい感情推定の研究についてはまだ始まったばかりである⁸⁾⁻¹²⁾．感情推定のアプローチを大きく分類すると，ルールベースのアプローチと統計的なアプローチの 2 つに分けられる．Masum らおよび Chaumartin はルールベースにより感情を推定する手法を提案している^{9),12)}．Chaumartin の感情推定システムは，SentiWordNet や WordNet Affect などの言語リソースを利用して単語自体の感情を推定したうえで，構文的なルールに従って文全体の感情を推定している．しかし，言語リソースの規模が十分でないため感情推定できる文のカバレッジは十分でない．一方，統計的なアプローチでは，Kozareva らが Turney の感情極性推定手法を応用し，6 種類の感情語（e.g. joy, fear）との相互情報量から新聞の見出しの感情を推定する手法を提案している¹⁰⁾．これは，ある単語が高い確率で感情語と共起する場合（e.g. birthday は

joy と高頻度で共起, war は fear と高頻度で共起) は, 当該単語は当該感情を意味するという仮定に基づいたものである. しかし, 本論で後述するように, 彼らの手法では十分な精度で感情を推定することができない. 我々は, Kozareva らの手法より有意に優れた精度で感情を推定する統計的手法を提案する.

3. 感情推定手法

3.1 感情生起要因コーパスの構築

感情の種類に関しては古くから多くの議論がある¹³⁾⁻¹⁵⁾. たとえば, Plutchik は心理学的観点から感情を {joy, acceptance, curiosity, anger, fear, sadness, disgust, surprise} の 8 種類に分類している¹³⁾. また, 中村は言語表現の観点から感情を {喜, 好, 安, 怒, 哀, 怖, 厭, 恥, 昂, 驚} の 10 種類に分類した¹⁵⁾. 一方, 1 人暮らしの高齢者との対話では {残念, 心配, 寂しい} などの感情を理解することが必要といわれている¹⁶⁾. これらの先行知見を考慮し, 本論では {嬉しい, 楽しい, 安心, 恐い, 悲しい, 残念, 嫌, 寂しい, 心配, 腹立たしい}, および, <neutral>^{*1} の 11 種類の感情を推定する.

まずはじめに, <neutral> 以外の 10 種類の感情を表現する下位語を人手で収集した. 感情表現の収集には, 寺村の定義¹⁷⁾ を参考とした下記の基準を用いた.

X = 感情主, Y = 対象, Z = 当該語のとき, 「X は Y を Z」「X は Y に Z」「X は Y が Z」のいずれかが表現できれば, Z は感情表現である^{*2}.

上記の定義に従い, 小林らの評価値表現辞書¹⁸⁾,^{*3} から感情表現を抽出した結果, 349 語の感情表現を得た. 表 1 に抽出した感情ごとの感情表現の数と例を示す.

次に, 図 2 に示す言語パターンを用いて Web テキストから自動的に感情生起要因を獲得する. 獲得の手がかりとする感情表現には表 1 の 349 語を, 接続表現には右記の 8 種類 (ので, から, ため, て, のは, のが, ことは, ことが) を用いる. たとえば「突然雨が降り出した のは がっかりだ」という文からは, <がっかり> が生起する要因として {突然雨が降り出した} を獲得する.

3.2 感情極性の推定

3.1 節では, <neutral> 以外の 10 種類の感情の生起要因の獲得方法について述べた. これ

表 1 感情生起要因コーパスの規模と例

Table 1 Distribution of the emotion expressions and examples.

感情極性	10 感情	感情表現 (349 語)	
		計	例
Positive	嬉しい	90	嬉しい, 狂喜, 喜ぶ, 歡ぶ
	楽しい	7	楽しい, 楽しむ, 楽しめる
	安心	5	安心, ほんと
Negative	恐い	22	恐い, 怖い, 恐ろしい
	悲しい	21	悲しい, 哀しい, 悲しむ
	残念	15	がっかり, がっくり
	嫌	109	嫌, 嫌がる, 嫌い
	寂しい	15	寂しい, 淋しい, わびしい
	心配	17	不安, 心配, 気がかり
	腹立たしい	48	腹立たしい, 腹立つ, 立腹

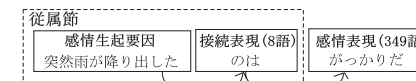


図 2 感情生起要因を獲得するための言語パターン
Fig. 2 An example of a lexico-syntactic pattern.

らは, ポジティブおよびネガティブの感情極性を持つ事態と考えることができる^{*4}. これらに加えて, ニュートラルの事態も大量に入手できれば, 機械学習により感情極性 (ポジティブ/ネガティブ/ニュートラル) を分類する感情極性推定モデルが構築できる. 感情生起要因コーパスを構築した際は, 「嬉しい, 楽しい」など 349 語の感情表現と言語パターンを用いて感情生起要因を獲得したが, どのような表現がニュートラルを表す語かが自明でないことから, 感情生起要因の収集と同じ方法でニュートラルの事例を抽出するのは難しい. しかし, 仮に Web 上に出現するテキストのほとんどがニュートラルの場合は, 単純にランダムにテキストをサンプルすることでニュートラルの文を大量に収集できる. そこで, 実際に Web テキストから 1,000 文をランダムに抽出し, これらの感情極性の頻度を調べた. 結果を表 2 に示す. 表 2 から, 実際の Web テキストではニュートラルの事例は少ないことが分かる. このことから, ニュートラルの事例を大量に収集することは容易ではないと考えられ

*1 <neutral> はユーザ発話が感情的な意味を持たないことを表す.

*2 Z がサ変動詞や助動詞に接続して述語表現になる場合は「狂喜スル」「嫌ダ」のように動詞や助動詞を補って判断する.

*3 評価値表現辞書は下記からダウンロードできる. http://www.syncha.org/evaluative_expressions.html

*4 たとえば, 図 2 の「突然雨が降り出した」はネガティブの感情極性を持つ. 10 種類の感情の感情極性は表 1 に示す.

表 2 web テキストにおける感情極性の割合

Table 2 Distribution of the Sentiment polarity of sentences randomly sampled from the Web.

感情極性	数	割合
ポジティブ	650	65.0%
ネガティブ	153	15.3%
ニュートラル	117	11.7%
文脈依存	80	8.0%

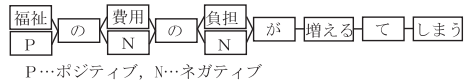


図 3 感情極性付きのラティスの例
Fig. 3 An example of a word-polarity lattice.

る．そこで我々は、感情生起要因コーパスに含まれるポジティブおよびネガティブの事例を用いて感情極性推定モデルを構築し、ユーザ発話の感情極性推定結果が感情極性推定モデルの分離面に近い場合にニュートラルと推定するモデルを考える．

感情極性の推定に関しては多くの手法が提案されている¹⁹⁾⁻²¹⁾．これらを参考にして、我々は、{ 単語, 単語自体の感情極性, 係り受け } の情報を特徴量として文の感情極性を推定する．図 3 に「福祉の費用の負担が増えるてしまう」という文を、単語の感情極性つきで記述した例を示す．ここでは「福祉」はポジティブ、「費用、負担」はネガティブの感情極性を持つとする．図 3 のラティスに対して、3-gram 以下の列を展開すると、「福祉, P, 福祉の, P の, 福祉の費用, P の費用, ...」などが得られる．これらを素性として SVM^{*1}で学習し感情極性推定モデルを構築する．

3.3 感情推定

3.3.1 kNN による感情推定

粒度の細かい感情推定のため、我々は、感情生起要因コーパスを用いた k-nearest-neighbor (kNN) による感情推定手法を提案する．

まず、入力ユーザ発話が与えられると、ユーザ発話と類似する上位 k 個の事例を感情生起要因コーパスから検索する．たとえば、図 1 の「遠くのレストランに行ったら定休日

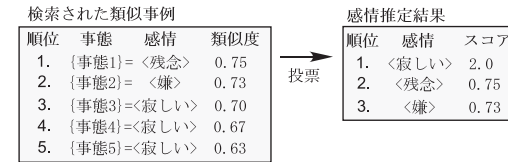


図 4 kNN による感情推定 (k=5)
Fig. 4 Emotion Classification by kNN (k=5).

だった」というユーザ発話に対しては、「遠くの店に行ったら定休日だった」などが検索される．類似度の計算には、入力 I と感情生起要因 EP の単語のコサイン類似度を用いる； $sim(I, EP) = \frac{I \cdot EP}{|I||EP|}$ ．さらに、抽出した上位 k 個の類似度の感情クラスごとの和をとり、スコアが最大となる感情を入力 I に対する感情推定結果とする．仮に、ある入力 I に類似する事態を検索した結果、図 4 が得られたとする．図 4 では入力文と最も類似する事態の感情は〈残念〉、次に類似する事態の感情は〈嫌〉である．ここで、それぞれの感情について類似度の和を求め、最もスコアが高い感情を入力発話に対する感情として出力する．たとえば、図 4 の例では〈寂しい〉のスコアが最も高いので、入力文 I の感情推定結果は〈寂しい〉となる．

このように、上位 k 個の投票により感情を推定することで、感情生起要因コーパスにノイズが含まれる場合も頑健に感情推定できると期待する．

3.3.2 2段階による感情推定

1.2 節で述べたとおり、分類器および対話システムへの応用面の利点から、本論では、感情極性をあらかじめ推定した後に感情を推定するという 2 段階による感情推定手法を提案する．

具体的には、3.2 節で述べた感情極性推定モデルに基づき感情極性を推定した後、ポジティブおよびネガティブと推定されたユーザ発話に対して感情を推定する．基本的な感情推定手法は 3.3.1 項で述べたとおりだが、kNN で類似事例を検索する際は、感情生起要因コーパス全体から類似事例を検索するのではなく、あらかじめ推定した感情極性と同一の感情極性を持つ事例のみを対象として類似事例を検索する．具体的には、図 1 の「遠くのレストランに行ったら定休日だった」の例で感情極性推定結果がネガティブとすると、感情生起要因コーパスのネガティブ(怖い, 悲しい, 残念, 嫌, 寂しい, 心配, 腹立たしい)の感情生起要因を対象として類似事例を検索し、感情を推定する．

*1 SVM の学習には TinySVM を使用した．下記から取得できる．
<http://chasen.org/~taku/software/TinySVM/>

4. 実験

4.1 感情生起要因コーパスの評価

3.1 節で述べた方法で, Kawahara らの Web 5 億文コーパス²²⁾ から実際に感情生起の要因を獲得した. その結果, 約 130 万件の感情生起要因が獲得された. 表 3 に 10 種類の感情ごとの感情生起要因の獲得数を示す. 我々の知る限り, これほど大規模な感情生起要因のコーパスはこれまで存在しない.

感情生起要因コーパスからランダムに 2,000 事例を抽出し, 獲得した感情生起要因の精度を調べた. 評価は, 感情生起要因コーパスの構築に関わっていない作業者が 1 名で行った. 表 4 に評価結果を, 表 5 に具体例を示す. 表 4 および表 5 の「感情極性」は獲得した事態

表 3 感情生起要因コーパスの規模

Table 3 Number of emotion-provoking events.

10 感情	感情生起要因	10 感情	感情生起要因
嬉しい	387,275	残念	106,284
楽しい	209,682	嫌	396,002
安心	46,228	寂しい	26,493
恐い	49,516	心配	45,018
悲しい	31,369	腹立たしい	8,478

表 4 感情生起要因コーパスの評価結果

Table 4 Correctness of samples from the EP corpus.

	感情極性	感情
正例	1,140 (57.0%)	988 (49.4%)
文脈依存	678 (33.9%)	489 (24.5%)
負例	182 (9.1%)	523 (26.2%)

表 5 感情生起要因コーパスの評価の例

Table 5 Examples from in the EP corpus.

感情生起要因コーパス			評価結果	
事態	感情表現 (349 語)	10 感情 (P/N)	感情極性	感情
花持ちが悪い	残念だ	残念 (N)	正例	正例
元写真がもうない	悲しい	悲しい (N)	正例	文脈依存
ちんげん菜が多い	嬉しい	嬉しい (P)	文脈依存	文脈依存
会社員になる	嫌だ	嫌 (N)	文脈依存	文脈依存
ジュースが飲みたい	大変だ	嫌 (N)	負例	負例

と感情との感情極性が一致するかどうかを評価した結果を, 「感情」は獲得した事態が当該感情の生起要因かどうかを評価した結果を示す. また, 「正例」は「正例」, 「文脈依存」は「文脈によっては正例である事例」, 「負例」は「負例」を示す. たとえば, 表 5 の「元写真がもうない」については, 作業者は「元写真がもうない=ネガティブ」は文脈によらずに成立すると判断し「感情極性」は正例, 「元写真がもうない=悲しい」が成立するかどうかは文脈に依存すると判断し「感情」は文脈依存としている. なお, 「ジュースが飲みたい」が「大変だ」の感情生起要因として獲得されているが, これは, 本来は係り受け関係にない従属節と感情表現が係り受け解析誤りによって獲得された例である. このように実際には当該感情の生起要因としてふさわしくない事例の評価結果は「負例」とする.

表 4 に示すとおり, 感情極性の評価については, 57.0%が正例, さらに, 文脈によって正例となる事態も加えると 90.9%の事態が正例であった. また, 感情の成立性に関しては 49.4%が正例, 文脈によって正例となる事態を加えると 73.9%が正例であった. 感情生起要因コーパスの精度が十分かどうかは, 感情生起要因コーパスを用いた感情推定精度を評価して初めて言及できる. 4.3 節で述べるとおり, 感情生起要因コーパスを用いることで, 従来手法より有意に高い精度で感情推定できたことから, 有用な感情生起要因コーパスが構築できたと考える.

4.2 感情極性推定実験

4.2.1 感情極性推定実験方法

3.2 節で述べた感情極性推定手法について, 感情極性に有効な特徴量を明らかにするため, 以下の 3 つのセッティングにより感情極性を推定する.

a) 単語 n-gram

単語の {1-gram, 2-gram, 3-gram} を特徴量とする.

b) 単語 n-gram + 単語の感情極性

単語と感情極性のラティス (図 3) から構築した単語および単語自体の感情極性の {1-gram, 2-gram, 3-gram} を特徴量とする. なお, ここで用いる単語列は係り受け構造を考慮したものでなく単語の出現順の列を用いる.

c) 単語 n-gram + 単語の感情極性 + 係り受け

単語と感情極性のラティス (図 3) から構築した単語および単語自体の感情極性の {1-gram, 2-gram, 3-gram} を特徴量とする. なお, ここで用いる単語列は係り受け関係を考慮し, 係り受け構造の単語と感情極性のラティスを用いる.

なお, b) および c) の単語の感情極性には, 高村らが作成した単語感情極性対応表²³⁾ で

表 6 ポジティブ/ネガティブに対する感情極性推定結果
Table 6 F-values of sentiment polarity classification (positive/negative).

コーパス サイズ	a) 単語				b) 単語+感情極性				c) 単語+感情極性+係り受け			
	TestSet1		TestSet2		TestSet1		TestSet2		TestSet1		TestSet2	
	Pos	Neg	Pos	Neg	Pos	Neg	Pos	Neg	Pos	Neg	Pos	Neg
1/1,000	0.559	0.536	0.487	0.589	0.564	0.530	0.535*	0.572	0.565	0.523	0.537*	0.575
1/100	0.643	0.520	0.553	0.556	0.672	0.569	0.616*	0.573	0.674	0.582	0.617*	0.571
1/10	0.721	0.765	0.660	0.776	0.727	0.766	0.722*	0.790	0.722	0.764	0.712*	0.786
約 130 万	0.839	0.853	0.794	0.842	0.839	0.853	0.808	0.849	0.820	0.841	0.804	0.850

表 7 ポジティブ/ネガティブ/ニュートラルに対する感情極性推定結果
Table 7 F-values of sentiment polarity classification (positive/negative/neutral).

コーパス サイズ	a) 単語				b) 単語+感情極性				c) 単語+感情極性+係り受け			
	TestSet1		TestSet2		TestSet1		TestSet2		TestSet1		TestSet2	
	Pos	Neg	Pos	Neg	Pos	Neg	Pos	Neg	Pos	Neg	Pos	Neg
1/1,000	0.267	0.249	0.400	0.433	0.234	0.234	0.245†	0.461*	0.320	0.295	0.372†	0.575*
1/100	0.484	0.326	0.531	0.458	0.459	0.319	0.447†	0.483**	0.586	0.402	0.602*	0.593*
1/10	0.526	0.562	0.632	0.718	0.534	0.572	0.591†	0.725	0.577	0.623	0.690*	0.803*
約 130 万	0.743	0.758	0.610	0.742	0.769	0.769	0.734*	0.805**	0.769	0.769	0.756*	0.807**

確信度が 0.7 以上^{*1}の 8,215 エントリ (ポジティブ 2,349 エントリ, ネガティブ 5,866 エントリ) を用いた^{*2}. また, 学習の際は, 「嬉しい, 楽しい, 安心」の生起要因をポジティブの訓練データ, 「怖い, 悲しい, 残念, 嫌, 寂しい, 心配, 腹立たしい」の生起要因をネガティブの訓練データとする.

さらに, 上記の特徴量の有効性が学習元のコーパス量に依存するかどうかを調べるため, 構築した感情生起要因コーパスの 1/1,000, 1/100, 1/10, 1/1 のコーパスを用いて感情極性精度を調べる.

評価データには, 以下の 2 つを用いる.

TestSet1: 1 つめは, 弊所の所員 6 名がプロトタイプ対話システムに入力した発話に対して話者自身が感情極性を付与した 90 発話である. これらには, ポジティブ 31 発話, ネガティブ 34 発話, ニュートラル 25 発話が含まれる.

TestSet2: 2 つめは, 表 4 で感情極性が正例の 1,140 事例 (ポジティブ 491 事例, ネガティブ 649 事例), および, あらたに収集したニュートラルの 501 事例である. なお, 感情極性推定モデルはこれらを除いた事例で学習する.

4.2.2 感情極性推定実験結果

表 6 にポジティブおよびネガティブの事例を評価データとした場合の感情極性推定精度を, 表 7 にニュートラルも含んだ評価データに対する感情極性推定精度を示す^{*3}. また, TestSet2 を用いて^{*4}, a と b および c の間で T 検定による有意差検定を行った. 表 6 および表 7 の * は有意水準 1% で a に比べて b や c の精度が高いことを, ** は有意水準 5% で a に比べて b や c の精度が高いことを, † は有意水準 1% で b や c に比べて a の精度が高いことを意味する.

まず, 表 6 の結果から, 興味深い知見が 2 つ得られた. まずはじめに, 表 6 が示すとおり

*1 予備検討として, 確信度が 0.6 以上, 0.7 以上, 0.8 以上, 0.9 以上の 4 種類を用いて感情極性推定実験を行ったが, 有意差は確認できなかった.

*2 単語感情極性対応表は下記からダウンロードできる.

http://www.lr.pi.titech.ac.jp/~takamura/pndic_ja.html

*3 学習元のコーパス量が十分でない場合, コーパスによってモデルの性能に揺れが生じる場合がある. そこで, 1/1,000, 1/100, 1/10 については, それぞれ 5 セットの学習データからモデルを構築して感情極性推定精度を求めた.

*4 TestSet1 はデータ量が少なく有意差検定が行えなかったため, T 検定には TestSet2 を用いた.

り、学習元のコーパス量が少ない場合は a に比べて b や c を特徴量とした推定精度が高く、特に、Positive の推定については有意差が認められた。一方、約 130 万件を学習データとした場合にはすべての特徴量で有意差が認められなかった。従来研究では、単語自体の感情極性や係り受けなどさまざまな特徴量を用いて文の感情極性を推定してきたが、ポジティブおよびネガティブの推定については、十分な量の学習データがあれば単語 n-gram のみを特徴量とした推定で十分であることが示唆された。2 つめに、感情生起要因コーパスの評価（表 4）では、感情極性の正解事例は 57% しかなかったが、約 130 万件のデータを用いて学習した結果、感情極性が十分高い精度で分類できた。このことから、感情生起要因コーパスの評価では文脈依存と評価された事例も感情極性の学習に有効であったと考えられる。

一方、表 7 の結果から、ニュートラルも含めた評価データに対する感情極性推定は、ポジティブおよびネガティブのみに対する感情極性推定に比べてはるかに難しく、十分な量の学習データが必要であることが示唆される。また、約 130 万件を学習データとした場合には、単語のみを特徴量とするのに比べて、単語自体の感情極性や係り受けを特徴量とした推定が有意に精度が高かった。このことから、ニュートラルの事例も含むような複雑な感情極性推定については、単語以外の特徴量が有効であったといえる。

4.3 感情推定実験

4.3.1 感情推定実験方法

本節では、3.3 節で述べた感情推定手法を評価する。なお、ニュートラルは十分な精度で識別できると仮定し、ポジティブとネガティブの事例のみを用いて粒度の細かい感情推定実験を行う。具体的には、以下の 3 種類の手法により感情推定手法を評価する。

a) Baseline

Kozareva らが提案した感情推定手法をベースラインとする¹⁰⁾。彼らの手法では、まず、入力に含まれる単語 cw_j と感情表現 $e \in \{anger, disgust, fear, joy, sadness, surprise\}$ との相互情報量 $PMI(e, cw) = \log \frac{hits(e, cw)}{hits(e)hits(cw)}$ を求める。ここで、 $hits(x)$ は単語 x の検索エンジンでのヒット件数を指す。次に、感情ごとに入力に含まれる単語 cw_j の相互情報量の和を求め、スコアが最大の感情を入力 I の感情とする。

$$score(E_i) = \sum_j PMI(e_i, cw_j)$$

我々の実験では、単語感情表現として 3.1 節で収集した 349 語を用い、10 種類の感情でスコアの平均が最大となる感情を推定結果とした。検索エンジンには Yahoo! を用いた。

表 8 TestSet1 に対する 2 名の作業員による感情付与の例
Table 8 Examples of TestSet1 (2p, best).

	作業員 1	作業員 2
クリスマスにプレゼントをもらった	嬉しい	嬉しい
友達の家遊びに行く	楽しい	楽しい
花見に行ったら突然雨が降り出した	悲しい	悲しい
渋滞でほとんど動かない	嫌	腹立たしい
買い物に行ったら定休日だった	悲しい	残念

b) 感情推定

3.3.1 項で述べた kNN による感情推定手法を評価する。具体的には、感情生起要因コーパス全体から入力のユーザ発話と類似する k 事例を抽出し、感情ごとの類似度の和により感情を推定する。

c) 感情極性 + 感情推定

3.3.2 項で述べた 2 段階による感情推定手法を評価する。具体的には、ユーザ発話の感情極性を推定した後に、推定した感情極性が同一の事例のみを対象として類似する k 事例を抽出し、感情ごとの類似度の和により感情を推定する。

なお、表 3 に示すとおり、感情クラスによって獲得した感情生起要因の数に偏りがあるため、b および c では、感情生起要因コーパスから感情ごとに 5 万事例ずつ^{*1}をランダムに抽出したコーパスを用いた。また、c の感情極性の推定では単語のみを特徴量とする感情極性推定モデルを用いた。

評価データには、以下の 3 つを用いる。

TestSet1 (2p, best): 4.2.1 項の TestSet1 のポジティブとネガティブの 65 事例に対して 2 名の作業員が 10 種類の感情クラスを独立に付与した。1 つの発話に複数の感情が該当する場合には、最も適切と考えられる感情クラスを 1 つ選択して付与するように指示した（表 8 に感情付与の例を示す）。kappa 値を用いて作業員の付与した感情の一致率を評価したところ $\kappa = 0.76$ であった²⁴⁾。作業員 2 名の正解が異なる場合にはどちらかに合致すれば正解とした。

TestSet1 (1p, acceptable): 4.2.1 項の TestSet1 のポジティブとネガティブの 65 事例に対して 1 名の作業員が 10 種類の感情クラスを付与した。ある文に複数の感情が該当

*1 { 安心, 恐い, 悲しい, 寂しい, 心配, 腹立たしい } に関しては、感情生起要因が 5 万事例に満たないため全事例を用いる。

表 9 TestSet1 に対する 1 名の作業員による感情付与の例
Table 9 Examples of TestSet1 (1p, acceptable).

	感情クラス
クリスマスにプレゼントをもらった	嬉しい
友達の家遊びに行く	楽しい, 嬉しい
花見に行ったら突然雨が降り出した	悲しい, 嫌, 残念, 腹立たい
渋滞でほとんど動かない	嫌, 腹立たい
買い物に行ったら定休日だった	悲しい, 残念, 嫌

表 10 TestSet2 に対する 1 名の作業員による感情付与の例
Table 10 Examples of TestSet2 (1p, acceptable).

	感情クラス
みなさんが返事を下さる	嬉しい, 安心
市役所に電話しても係りが分からない	嫌, 腹立たい
夜に地震があると寝付けけない	嫌, 怖い, 心配
その割に地元商店主が少ない	残念, 寂しい

する場合は、該当する感情クラスをすべて付与した。その結果、ポジティブの文に対しては平均 1.48 個、ネガティブの文に対しては平均 2.47 個の感情クラスが付与された (表 9 に例を示す)。感情推定実験では、作業員の付与した感情のうち 1 つを推定できれば正解とした。

TestSet2 (acceptable): 表 4 で感情が正例の 988 事例に対して 1 名の作業員が 10 種類の感情クラスを付与した。ある文に複数の感情が該当する場合は、該当する感情クラスをすべて付与した。その結果、ポジティブの文に対しては平均 1.19 個、ネガティブの文に対しては平均 1.66 個の感情クラスが付与された (表 10 に例を示す)。感情推定実験では、作業員の付与した感情のうち 1 つを推定できれば正解とした。

4.3.2 感情推定実験結果

感情推定の評価結果を図 5 に、感情推定の際の感情極性誤推定率を表 11 に示す。図 5 の Accuracy は、全評価データに対してシステムが正解の感情を出力した割合を示す。つまり、この値は、ユーザ発話が入力された場合にシステムがどのくらいの割合のユーザ発話の感情を正しく推定できるかを示す。なお、図 5 の「1-NN, 3-NN, 5-NN, 10-NN」は、感情ごとのスコアを求める際 (図 4) に上位何個の類似事例を用いたかを示す。

まず、図 5 の結果、ベースライン手法に比べて kNN による感情推定は有意に精度が高いことが分かった。ベースラインである相互情報量による感情推定手法は、感情極性推定では

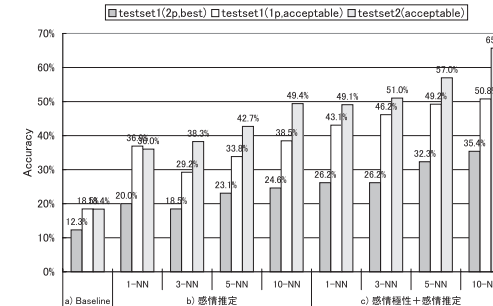


図 5 感情推定の評価結果

Fig. 5 Results of emotion classification.

表 11 感情推定時の感情極性の誤推定率

Table 11 Fatal error rate in emotion classification experiments.

	a) Baseline	b) 感情推定			c) 感情極性 + 感情推定
		1-NN	3-NN	10-NN	
TestSet1	49.2%	29.2%	26.2%	24.6%	15.4%
TestSet2	41.5%	37.6%	32.8%	30.0%	17.0%

比較よく用いられる手法だが、より詳細な感情を推定する場合に従来の相互情報量の手法をそのまま応用するだけでは難しいことが示唆される。また、手法 b と手法 c を比較した結果、2 段階による感情推定手法の精度が高かった。しかし、T 検定の結果^{*1}、手法 b と手法 c に有意差は認められなかった。次に、表 11 に示す感情推定時の感情極性誤推定率について、手法 b と手法 c の間で T 検定を行ったところ、有意水準 1% で誤推定率が有意に改善されていることが分かった。1.2 節で述べたとおり、ユーザの感情に共感的な応答を生成する対話システムでは、感情極性の誤りは致命的な誤応答を引き起こす。感情推定時の感情極性の誤り率が有意差を持って低減されたことから、対話システムへの応用面で有効な感情推定手法を提案できたと考える。

次に、我々の提案手法における感情ごとの感情推定精度について考察する。表 12 は、TestSet1 (acceptable) および TestSet2 (acceptable) における手法 c の 10-NN の感情ごとの感情推定精度を示す。表 12 の結果から、感情ごとの感情推定精度に偏りがあることが

*1 TestSet1 はデータ量が少なく有意差検定は行えなかったため、T 検定は TestSet2 (acceptable) を用いた。

表 12 c) 感情極性+感情推定 10-NN による感情ごとの感情推定精度
Table 12 Precisions of emotion classification by method c 10-NN.

	嬉しい	楽しい	安心	恐い	悲しい	残念	嫌	寂しい	心配	腹立たしい
TestSet1	92.3%	46.7%	33.3%	20.0%	100%	11.1%	80.0%	25.0%	0%	0%
TestSet2	77.7%	48.9%	45.5%	29.1%	66.7%	77.0%	87.4%	64.3%	68.6%	0%

分かる．特に、〈腹立たしい〉や〈恐い〉については両方の評価データで精度が低かった．〈腹立たしい〉については、獲得した感情生起要因が少ないことが感情推定精度が低い原因の 1 つと考えられる．また、〈恐い〉の誤りを分析した結果、〈恐い〉の感情生起要因には直接的には〈残念〉を生起する事例^{*1}や、〈恐い〉が〈心配〉の意味で使われた文から獲得された生起要因^{*2}が多いことが分かった．これらの課題を解決するためには、今よりさらに量と質の高い感情生起要因コーパスを構築する必要がある．

5. まとめと今後の課題

本論では、ユーザの感情を推定できる対話システムの実現を目指し、感情推定手法を提案した．我々はまず、約 130 万件からなる感情生起要因コーパスを構築した．次に、あらかじめ感情極性を推定したうえで、コーパスから類似事例を検索して 10 クラスの感情を推定する手法を提案した．感情生起要因コーパスを学習データとして感情極性モデルを構築した結果、ポジティブおよびネガティブの推定については単語 n-gram でも十分高い精度が得られることが分かった．次に、k-nearest-neighbor (kNN) による感情推定手法を提案した．評価実験の結果、我々の提案手法は従来手法より有意に精度が高かった．また、この感情推定実験では、感情生起要因コーパス全体から類似事例を検索するより、事前に感情極性を推定したうえで感情極性が同一の事例に対して類似事例を検索し感情を推定する方が精度が高いことが分かった．特に、この 2 段階による感情推定手法は感情極性の誤推定率を有意に改善することが明らかとなった．

*1 たとえば { あの海水浴場は水が汚い } は〈恐い〉の感情生起要因として獲得されているが、直接的には { あの海水浴場は水が汚い } は〈残念〉を生起する事態で、{ (海水浴場の水が汚いの)に泳ぐ } 結果〈恐い〉のように、{ あの海水浴場は水が汚い } は他の事態 { 泳ぐ } を介して間接的に〈恐い〉を生起すると考えられる．また同様に、{ 大量に食べて虫歯になっちゃった } は直接的には〈残念〉を生起し、{ (虫歯になって) 歯医者に行く } ことが〈恐い〉と考えられる．このような事例は、正解が〈残念〉の文と類似度が高くなるため、本来は〈残念〉の事例を〈恐い〉と誤解答する原因となる．

*2 { 壊れた代車を貸すようなところに車を修理に出す }、{ ちょっと間違えと取り返しのつかない事態に発展しかねない } は、恐いが心配 (不安) の意味で使われてた文から獲得された感情生起要因と考えられる．

今後は、kNN 以外のモデルを用いて評価することで我々の手法の有効性を確認する．また、今回提案した感情推定手法を対話に応用し有効性を評価する．

謝辞 Web コーパスを提供してくださった河原大輔氏、評価値表現辞書を作成された小林のぞみ氏、TinySVM を開発された工藤拓氏に深く感謝いたします．

参 考 文 献

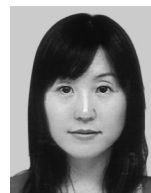
- 1) Allen, J.F., Schubert, L.K., Ferguson, G., Heeman, P., Hwang, C.H., Kato, T., Light, M., Martin, N.G., Miller, B.W., Poesio, M. and Traum, D.R.: The TRAINS Project: A case study in building a conversational planning agent, *Journal of Experimental and Theoretical AI*, Vol.7, pp.7-48 (1995).
- 2) Foster, M.E.: Enhancing Human-Computer Interaction with Embodied Conversational Agents, *Lecture Notes in Computer Science*, Vol.4555, pp.828-837 (2007).
- 3) Tokuhisa, R. and Terashima, R.: Relationship between Utterance and “Enthusiasm” in Non-Task-Oriented Conversational Dialogue, *Proc. 7th SIGdial Workshop on Discourse and Dialogue*, pp.161-167 (2006).
- 4) Pantic, M. and Rothkrantz, L.J.M.: Facial Action Recognition for Facial Expression Analysis From Static Face Images, *IEEE Trans. SMC-B*, Vol.34, No.3, pp.1449-1461 (2004).
- 5) Rotaru, M., Litman, D.J. and Forbes-Riley, K.: Interactions between Speech Recognition Problems and User Emotions, *Proc. 9th European Conference on Speech Communication and Technology*, pp.2481-2484 (2005).
- 6) Batliner, A., Fischer, K., Huber, R., Spilker, J. and Noth, E.: How to find trouble in communication, *Speech Communication*, Vol.40, No.1-2, pp.117-143 (2004).
- 7) Litman, D.J. and Forbes-Riley, K.: Predicting Student Emotions in Computer-Human Tutoring Dialogues, *Proc. 42nd Annual Meeting of the Association for Computational Linguistics*, pp.352-359 (2004).
- 8) Alm, C.O., Roth, D. and Sproat, R.: Emotions from text: machine learning for text-based emotion prediction, *HLT/EMNLP*, pp.579-586 (2005).
- 9) Chaumartin, F.-R.: UPAR7: A knowledge-based system for headline sentiment tagging, *Proc. 4th International Workshop on Semantic Evaluations*, pp.422-425 (2007).

- 10) Kozareva, Z., Navarro, B., Vazquez, S. and Nibtoyo, A.: UA-ZBSA: A Headline Emotion Classification through Web Information, *Proc. 4th International Workshop on Semantic Evaluations*, pp.334–337 (2007).
- 11) Mihalcea, R. and Liu, H.: A corpus-based approach to finding happiness, *Proc. AAAI 2006 Spring Symposia on Computational Approaches to Analyzing Weblogs*, pp.139–144 (2006).
- 12) Masum, S.M.A., Prendinger, H. and Ishizuka, M.: Emotion Sensitive News Agent: An Approach Towards User Centric Emotion Sensing from the News, *Proc. IEEE/WIC/ACM International Conference on Web Intelligence*, pp.614–620 (2007).
- 13) Plutchik, R.: *A general psychoevolutionary theory of emotion*, Vol.1, Academic Press, pp.3–33 (1980).
- 14) Ekman, P.: An Argument for Basic Emotions, *Cognition and Emotion*, Vol.6, pp.169–200 (1992).
- 15) 中村 明: 感情表現辞典, 東京堂出版 (1993).
- 16) ホールファミリーケア協会: 傾聴ボランティアのすすめ, 三省堂 (2004).
- 17) 寺村秀夫: 日本語のシンタクスと意味, くろしお出版 (1982).
- 18) 小林のぞみ, 乾健太郎, 松本裕治, 立石健二, 福島俊一: 意見抽出のための評価表現の収集, 自然言語処理, Vol.12, No.2, pp.203–222 (2005).
- 19) Turney, P.: Thumbs up? thumbs down? semantic orientation applied to unsupervised classification of reviews, *Proc. 40th Annual Meeting of the Association for Computational Linguistics*, pp.417–424 (2002).
- 20) 乾 孝司, 奥村 学: テキストを対象とした評価情報の分析に関する研究動向, 自然言語処理, Vol.13, No.3, pp.201–241 (2006).
- 21) 工藤 拓, 松本裕治: 半構造化テキストの分類のためのブースティングアルゴリズム, 情報処理学会論文誌, Vol.45, No.9, pp.2146–2156 (2004).
- 22) Kawahara, D. and Kurohashi, S.: Case Frame Compilation from the Web using High-Performance Computing, *Proc. 5th International Conference on Language Resources and Evaluation*, pp.1344–1347 (2006).
- 23) 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).
- 24) Cohen, J.: A coefficient of agreement for nominal scales, *Educational and Psycho-*

logical Measurement 20, pp.37–46 (1960).

(平成 20 年 4 月 24 日受付)

(平成 21 年 1 月 7 日採録)



徳久 良子 (正会員)

1999 年九州工業大学知能情報工学科卒業。2001 年同大学大学院修士課程情報科学専攻修了。同年 (株) 豊田中央研究所入所。2007 年奈良先端科学技術大学院大学博士後期課程入学。感情や音声対話に関する研究開発に従事。2000 年度人工知能学会研究会奨励賞受賞, 2005 年度人工知能学会研究会優秀賞受賞。人工知能学会, 言語処理学会各会員。



乾 健太郎 (正会員)

1995 年東京工業大学大学院情報理工学研究科博士課程修了。博士 (工学)。同研究科助手, 九州工業大学情報工学部助教授を経て, 2002 年より奈良先端科学技術大学院大学情報科学研究科助教授。現在同研究科准教授, 情報通信研究機構有期研究員を兼任。自然言語処理の研究に従事。人工知能学会, ACL 各会員。



松本 裕治 (正会員)

1977 年京都大学工学部情報工学科卒業。1979 年同大学大学院工学研究科修士課程情報工学専攻修了。同年電子技術総合研究所入所。1984 ~ 1985 年英国インペリアルカレッジ客員研究員。1985 ~ 1987 年 (財) 新世代コンピュータ技術開発機構に出向。京都大学助教授を経て, 1993 年より奈良先端科学技術大学院大学教授, 現在に至る。工学博士。専門は自然言語処理。人工知能学会, 日本ソフトウェア科学会, 言語処理学会, 認知科学会, AAAI, ACL, ACM 各会員。