原著論文

発話内容と口調の関係に基づく発話者の嗜好情報推定[†]

目良 和也*・青山 正人*・黒澤 義明*・竹澤 寿幸*

ヒューマンコンピュータインタラクションシステムがユーザの気持ちを考慮したりユーザの希望に沿った対応を行うためには、各ユーザがどのような事物を好きか嫌いかという嗜好情報が重要となる。しかし、一般的な嗜好情報データベースはあくまで普遍的な嗜好傾向を表しているにすぎず、個人差や経験による変化を考慮できない。しかし人間は会話を通じてリアルタイムに相手の嗜好情報を推定することができる。本研究では、『Xに良い事があった』と喜んでいるならば、きっとXの事が好きなんだろう』のようなヒューリスティックに基づき、発話内容と口調から発話者の嗜好情報を推定する手法を提案する。本研究では情緒計算手法に基づいて発話内容の構文構造と話者感情の関係を表すことにより、上記のヒューリスティックを実現する。発話者の感情状態は、聞き手による感情推定結果を用いる。実験の結果、精度 0.76、再現率 0.88 という結果が得られた。

キーワード：嗜好情報、発話音声の口調、話者感情、情緒計算手法

1. はじめに

現在、AI スピーカーをはじめとして、人間とコミュニケーションを行う様々なコンピュータ対話システムが社会実装されつつある。このようなシステムが文字通り機械的な対応ではなく、人間のようにユーザの気持ちを考慮したり、ユーザの希望に沿った対応を行うためには、各ユーザがどのような事物を好きか嫌いかという嗜好情報が重要となる。

嗜好情報を扱っているデータベースとしては、SentiWord-Net [1] のように人手で作成したものや、日本語評価極性辞書 [2, 3] のように大規模 Web テキストコーパスから統計的に自動で構築したものなどがある。しかし図 1 に示すように、事物に対する好き嫌いの印象は人によって異なることもあり、また同一人物でも後に獲得した知識や経験によって変化することもある。不特定多数の人間が発信した情報を統計的に処理した場合、普遍的な多数派の嗜好傾向は獲得できるが、図 1 のような個人差や経験による変化には対応できない。

それでも人間は、相手との対話を通じて相手が何を好きで何を嫌いかといった嗜好情報を獲得することができる。例えばある人が「私は広島カープが好きです」と言ったならば、「話者は広島カープのことが好き」であることは自明である。また、前提知識無しに「広島カープが勝ったので嬉しい」と言ったとしても、『Xに良い事があった』と喜んでいるならば、きっとXの事が好きなんだろう』というヒューリスティックに基づいて、「話者は広島カープのことが好き」であることが推論できる。さらに、図 2 左のように「(嬉しそうに) 広島カープが勝ったよ」と自分の感情を直接「～なので嬉しい」のように発話していなくても、嬉しそうな話者の様子から「話者は広島カープ

図 1 嗜好情報の個人差と経験による変化

図 2 同じ発話でも推定される好感度が異なる例

図 3 同じ快感情の発話でも推定される好感度が異なる例

[†] Estimating Personal Preference Information on the Basis of Relationship Between Content and Tone of Utterance
Kazuya MERA, Masahito AOYAMA, Yoshiaki KUROSAWA,
and Toshiyuki TAKEZAWA

* 広島市立大学大学院情報科学研究科
Graduate School of Information Sciences, Hiroshima City University

のことが好き”であることが推測できる。

しかし、全く同じ文言を発話したとしても、その時の発話者の様子によって推定される好感度が異なる場合がある。図 2 右のように「(悔しそうに) 広島カープが勝ったよ」と言った場合だと、全く同じ文字列からなる発話にも関わらず「話者は広

島カーブの事が嫌い”と推測できる。

とはいえ、【快感を伴った発話に含まれる事物のことを発話者は好きである】のように話者感情のみから単純に事物の好感度を推定できるわけではない。図3のように、同じ快感を伴った発話であっても、意味構造によっては好感度の推定結果が逆転することもある。

そこで本論文は、【『Xに良い事があった』と喜んでいるならば、きっとXの事が好きなんだろう】のような**事象と反応に基づくヒューリスティックと発話の口調から推定した話者感情**を用いて、**発話内容に含まれる事物に対して話者が感じている好感度**を推定する手法を提案する。本手法では、情緒計算手法[4]に基づいたヒューリスティックを用いる。

以下、関連研究のまとめと本研究の位置づけについて2章で述べ、提案手法である発話内の事物の好感度を話者感情から推定する手法について3章で述べる。そして4章で評価実験とその結果、5章で考察について述べる。6章でまとめと今後の課題について述べる。

2. 関連研究のまとめと本研究の位置づけ

現在、テキストから評価情報を発見、抽出することを目的とした様々な研究が行われている[5]。小林[2]は、「〈対象〉の〈属性〉は〈評価値〉」のような典型的な文型パターンを手がかりとして用いることで、〈対象、属性、評価値〉という3つ組情報を抽出する手法を提案している。平野[6]はパターンマイニング手法を質問回答ペアに拡張し、〈述語項構造、エンティティ、人物属性、トピック〉の構造の情報抽出を行っている。また、縄手[7]は対話中の興味推定対象語の前後の単語ベクトル系列をニューラルネットワークに入力することで対象語に関するユーザの興味の有無を推定している。これらの手法では「広島カーブが好きです」や「広島カーブが勝ったよ」のように語の評価極性がそのまま対象語の評価に影響するため、図2右のような“「広島カーブが勝った」ことに対して怒っているので広島カーブのことが嫌い”といった状況には対処できない。

このような状況を把握するには、発話文字列で表している出来事とその時の話者の心情の両方の情報が必要となる。徳久[8]は、「...なので嬉しい」や「...だけど嬉しい」のように因果関係の前後の節における感情生起の関係を関係表現知識として定義し、対象節に対する感情を対象節の外から推定する方法を提案している。また、小林[9]は、感情極性語の係り受け関係に基づいて対象語の感情極性を推定している。しかしこれらの手法もテキストベースの処理であるため、「...なので嬉しい」のように言葉で表された感情しか捉えられず、声が弾んでいたたり笑顔だったといったノンバーバルな情報を考慮することはできない。

表情や口調などノンバーバルな情報も加えたマルチモーダルな興味推定手法として、Schuller[10]やWang[11]は、発話のテキストと音響的特徴を用いた興味推定手法を提案している。また、西本[12]はテキストと音響的特徴に加え顔画像を用いている。また倉野[13]は、モバイルデバイスの三軸加速度センサの情報も加えた手法を提案している。このような情報を使うことで、言葉として発していない話者の心情を知ることがで

きる。しかし、これらの手法はトピック全体に対する興味推定として処理を行っているため、図3右のようなシニカルな感情表出に対しても、シンプルに対象語である広島カーブのことが好きと誤推定してしまう。

提案手法では、“表出感情＝発話内対象語への嗜好”のようなシンプルな関係ではなく、【『Xに良い事があった』と喜んでいるならば、きっとXの事が好きなんだろう】のようなヒューリスティックに基づいて発話文字列の構文構造と表出感情の関係を処理する。また、発話の音響的特徴から話者感情を推定する手法を組み合わせることで、話者感情を言葉として発していなくても、好感度推定処理を行うことができる。

3. 話者感情を用いた語の好感度推定手法

3.1 用語の定義

本論文では、“発話の口調から推定した話者感情に基づいて、話者が対象の事物に対して持っている好き嫌いのイメージ（好感度）を推定する”手法を提案している。本節では、本論文で用いる各用語の定義について説明する。

- **話者感情**：ある発話を行った発話者が、その発話内容が表している事象に対して生起している感情。提案手法では、“発話時に表出している話者感情”を“発話内容が表す事象に対して生起した感情”とみなしている。提案手法では、Russellの感情円環モデル[14]など感情を多次元空間で表現する心理モデルにおいて最も多く採用されている valence 軸に基づき、{快, 不快, neither (快でも不快でもない)}の3クラスによって話者感情を表現する
- **事象**：『〇〇がxxを～した』のように事物の関係からなる一つの出来事。深層格構造によって記述される
- **事物**：事象のように特定の出来事を表すのではなく、物や事の概念に相当する
- **好感度**：ある人が、ある事物に対して持っているまたは感じている好き嫌いの度合。日本語評価極性辞書[2, 3]などの関連研究では評価極性という語が事物に対する印象を表すためによく用いられているが、評価という言葉は好き嫌いだけを対象としているわけではなく、「物の善悪・美醜などを考え、価値を定めること」という意味がある。本研究では、対象の事物に対して好ましく思っているか／嫌いだと思っているかという観点に注目して扱うため、評価極性ではなく好感度という語を用いる。本論文では、事物に対する好感度を{positive, neutral, negative}という3クラスで表現することによって、話者感情のクラスと区別する。

3.2 発話に付随する話者感情が好感度推定に与える影響

本論文では、発話の口調から推定される話者感情に基づいて話者が事物に対して持つ好感度を推定するという手法を提案しているため、はじめに本手法の前提となる“口調の違いが聞き手による好感度推定処理に影響を与えているか否か”について調査を行った。調査方法としては、同一の文字列からなる発話を異なる感情を表出しているように発声し、その音声を聞いて聞き手が推定した対象事物の好感度の推定傾向に違いがあるか

表 1 発話文字列と好感度推定対象事物

No.	発話文字列	好感度推定対象事物	対象の感情極性 [2, 3]
1	「『あなたはファンキーだ』と言われました」	ファンキー	positive
2	「今度スケートに行くことになりました」	スケート	neutral
3	「私は華奢な体型です」	華奢	negative

表 2 話者感情による好感度推定結果の違い

No.	音声の感情	好感度推定結果の回答数			p
		positive	neutral	negative	
1	快	10	5	0	p<0.01
	不快	0	0	15	
2	快	4	11	0	p<0.01
	不快	0	0	15	
3	快	5	10	0	p<0.01
	不快	0	0	15	

を分析した。

本実験では、3種類の発話文字列についてそれぞれ“快”と“不快”の感情を込めた発話音声を実験参加者に提示し、その発話文字列内の特定の語に対して話者がどのような好感度を持っているかを推定させた。表1に実験に用いた発話文字列と好感度推定対象事物を示す。好感度推定対象事物は、日本語評価極性辞書 [2, 3] に記載されている語から選択した。日本語評価極性辞書中で各語に割り振られている感情極性は、それぞれ「ファンキー (positive)」、「スケート (neutral)」、「華奢 (negative)」となっている。

実験で用いる感情のこもった音声は、音声創作ソフトウェア CeVIO Creative Studio S [15] を用いて作成した。CeVIO では合成音声作成時に、元気／怒り／哀しみ／平静の4感情の比率および、大きさ、速さ、高さ、声質、抑揚のパラメータを指定できる。さらに、各音素について声の大小、高低、長さを微調整できる。本実験では、CeVIOでの快音声作成時に感情の比率を{元気, 怒り, 哀しみ, 平静}={100, 0, 0, 0}と設定した。さらに Russell の感情円環モデル [14] における delight や glad を想定し、より arousal (活性度) が高く聞こえるように【高さ=50→60, 抑揚=50→80】とパラメータ調整した。一方、不快音声作成時には感情の比率を{元気, 怒り, 哀しみ, 平静}={0, 0, 100, 0}と設定した。さらに [14] における gloomy や depressed を想定し、より arousal が低く聞こえるようにパラメータを【高さ=50→40】に、また発話の最後の音の長さを1.8倍となるように調整した。

実験参加者は22歳～24歳の男女15名（男性11名、女性4名）である。実験参加者は、合成発話音声を No.1 (快), No.1 (不快), No.2 (快), No.2 (不快), No.3 (快), No.3 (不快) の順で聞き、各発話において話者が好感度推定対象事物に対して感じている好き嫌いの度合を、{positive, neutral, negative, 判断できない}の4つから1つ選んで回答した。回答結果を表2に示す。なお、本実験において“判断できない”という回答はどの事例においても1つもなかった。

実験参加者の回答分布について Fisher's exact test [16] を行っ

たところ、3つの事例全てについて統計的有意差があることが確認できた ($p < 0.01$)。特に、不快の感情を込めた音声において、実験参加者全員の回答が一致していた。さらに、日本語評価極性辞書で positive の感情極性を持つとされている語でも話し方によって negative と推定されたり、逆に negative の感情極性を持つとされている語も話し方によって positive と推定されたりすることがあった。これらの実験結果から“全く同じ発話文字列でも、話し方によって語の好感度推定結果は有意に変化する”という結論を導くことができる。

このことから、今回は特定のソフトウェアで作成した合成音声であるものの、同じ事物に対しても話者ごとに異なる好感度を発話音声の口調から推定する手法（提案手法）が意味を持つと考えられる。

3.3 情緒計算手法

感情を計算する研究はこれまで数多く行われているが、それらの多くは音響情報や表情、身体情報などのノンバーバル情報を使ったものであり、バーバル情報を用いた手法でも、発話文中に含まれる単語の感情極性を bag-of-words [17]、すなわち構文構造や語順を考慮せず単語の有無にのみ注目して処理を行っているものが大半である。しかし、bag-of-words 方式では、否定表現や係り受け関係の情報は欠落してしまう。一方、情緒計算手法 (Emotion Generating Calculations method: EGC) [4] では深層格フレーム構造を用いることにより各語句の構文構造を考慮した感情の計算が可能である。そこで本論文では EGC をもとにした好感度推定手法を提案する。

EGC では、Arnold が提唱している『《快／不快》は《好きな／嫌いな対象》の《接近／回避》により生起する』という思想 [18] に基づき、主体の好感度、対象の好感度、行動タイプ（接近／回避）の3要素の関係にのみ注目して快／不快を判定する。なお、EGC では《接近》を「ある事物を獲得すること」、「ある場所に移動すること」、「ある属性や状態になること」などの事象と定義している。一方、《回避》を「ある事物を失うこと」、「ある場所から移動すること」、「ある属性や状態でなくなる」などの事象と定義している。例えば、「プレゼントをもらう」、「同窓会に行く」、「病気になる」は《接近》事象、「ゴミを捨てる」、「刑務所から出所する」、「病気が治る」は《回避》事象となる。

EGC では、岡田 [19] で提唱されている述語の事象タイプごとに感情ベクトル合成に必要な3つの要素を定義している（表3参照）。岡田 [19] で用いられている必須格要素を以下に示す。

- S (Subject) 主体
- O (Object) 客体
- OF (Object-From) 出発点または源
- OT (Object-To) 目標
- OM (Object-Mutual) 相互作用の相手
- OS (Object-Source) 拠り所

それぞれの要素の取り得る値の組み合わせと生起する快／不快感情は表4の関係になる。なお行動については Arnold [18]

表3 事象タイプと快不快計算に用いる格要素

事象タイプ	主体	対象	行動
V(S)	主体	なし	述語
V(S,OF) V(S,OT)	主体	目標と出発点の 好感度の差分	述語
V(S,OM)	主体	相互作用の相手	述語
V(S,OS)	主体と拠り所の 好感度の差分	なし	述語
V(S,O)	主体	客体	述語
V(S,O,OF) V(S,O,OT)	客体	目標と出発点の 好感度の差分	述語
V(S,O,I)	客体	道具	述語
V(S,O,OC)	客体	なし	属性の 補足

表4 好感度および行動と生起感情の関係

主体の好感度	対象の好感度	行動タイプ	感情
positive	positive	接近	快
positive	positive	回避	不快
positive	negative	接近	不快
positive	negative	回避	快
negative	positive	接近	不快
negative	positive	回避	快
negative	negative	接近	快
negative	negative	回避	不快

に基づき、好き嫌いの特徴量ではなく“接近／回避”を特徴量として感情の計算を行っている。ただし、述語には様々なタイプがあるため、タイプによって処理は一部異なる。具体的には「死ぬ」や「優勝する」など対象を必須要素としないタイプの動詞では、対象の好感度値の代わりに、その行為によって結果的に好き／嫌いな状態が接近するか回避するかを考慮した値を動詞に付与して計算を行う。また形容詞や形容動詞も対象を必須要素としないタイプの動詞と同様に述語自体の好感度を感情計算に用いている。

3.4 EGCを使った好感度推定手法（提案手法）

EGCは、入力事象に含まれる事物の好感度を用いて事象に対して人が生起する快／不快感情を計算する。このことから逆に、人が生起している快／不快が明らかになっており、事象の主体あるいは対象の好感度が分かっている場合、残り要素の好感度を逆算することが可能である。そこで本論文では、EGCを使った好感度推定手法を提案する。

提案手法の処理手順を図4に示す。

1. 発話者の感情を推定する
2. 発話者が快あるいは不快を生起していれば、(3)以降の手順で好感度推定手法を適用する
3. 発話文字列から発話事象の主体、対象、行動の3要素を抽出する
4. 主体の好感度、対象の好感度、行動のタイプ（接近／回避）をデータベースから取得する
5. 主体の好感度と対象の好感度のいずれかが不明ならば、そ

図4 提案手法の処理の流れ

の不明な好感度を推定するために手法を適用する

6. 主体の好感度、対象の好感度、行動のタイプ、発話者の感情の関係（表4）に基づき、不明な好感度の値を逆算する

なおEGCでは“快あるいは不快が生起した状況”における要素と生起感情の関係に注目するため、“快あるいは不快が生起した状況”にのみ提案手法を適用し、“快でも不快でもない(neither)”という状況は扱わない。

以下に「(嬉しそうに) 昨日、寿司を食べたよ」という発話に対する好感度推定処理の例を示す。まず、話者が(嬉しそう)であることから話者感情＝快と判断する。次に発話文字列「昨日、寿司を食べたよ」に対して構文解析および係り受け解析を適用し、主体＝「自分」、対象＝「寿司」、行動＝「食べる」を抽出する。そして話者の嗜好情報を格納している好感度データベースを参照し、各事物の好感度を得る。その結果、「自分」の好感度＝positive、「寿司」の好感度＝未知、「食べる」の行動タイプ＝接近であったならば、「寿司」の好感度を推定するために本手法を適用する。表4より“主体の好感度＝positiveかつ行動タイプ＝接近かつ話者感情＝快”のときの対象の好感度は“positive”であるため、最終的に対象である「寿司」の好感度は話者にとってpositiveであると算出される。

しかし表4の関係が成り立たないケースも存在する。例えば「私は《対象》から逃げてきたんですよ」と不快な口調で話していた場合、表4では2行目に該当するため《対象》の好感度はpositiveと推定される。しかし、「私は《忌まわしいストーカー》から逃げてきたんですよ」という発話だった場合、《忌まわしいストーカー》に対する好感度がpositiveとは考えにくい。これは、《対象》に対する強い嫌悪から《対象》について話

すことや《対象》のことを考えること自体に不快を感じているために不快な口調での発話となったものと考えられる。提案手法では出来事自体に対する悲しみや怒りも発話中の事物に対する嫌悪も全て発話が表す出来事に対して表出された不快感情として処理を行っているため、このような状況を区別して処理できない。このような状況に対応するための方法については、5章で考察する。

3.5 好感度推定手法で用いるデータベースの構築

提案手法では、主体の好感度、対象の好感度、行動のタイプ（接近／回避）をデータベースから取得する。

主体および対象の好感度は、基本的に個人別の好感度データベースを参照して取得する。この個人別好感度データベースには、事前に取得した個人の嗜好情報および本手法によって随時算出された嗜好情報を格納する。しかし世の中全ての事物に対して各個人が持つ好感度を事前にもれなく登録しておくことは不可能である。人間は相手の嗜好が分からない場合、一般的な嗜好傾向に基づいて相手の嗜好を推測する。そこで提案手法では、個人別好感度データベースで網羅できなかった事物の好感度は日本語評価極性辞書 [2, 3] に基づいて構築した普遍的な好感度データベースによって対応する。日本語評価極性辞書には、評価極性を持つ（複合）名詞約 8,500 表現および用言約 5,000 件に対して評価極性情報が付与されている。そして事物の好感度を検索する際は、①個人別好感度データベース、②日本語評価極性辞書から構築した好感度データベース、の順で検索を行う。

行動のタイプ（接近／回避）を参照するデータベースは、岡田 [19] に掲載されている“事象概念の分類データ”の表に基づき 2,371 件の用言を登録している。語義が複数ある用言では複数の行動タイプを持つことがあるが、語義の判別には構文・述語項構造解析システム KNP [20] のような格構造を考慮した処理が必要となる。今回提案するシステムでは話し言葉を対象としていることと、文脈を考慮しない単文での処理を想定しており、省略された格要素の補完が困難であるため、多義性を持つ用言については参照データベースの第一語義に基づいて処理を行うこととする。

4. 評価実験

4.1 評価実験の構成

本節では、“口調から推定した話者感情に対して EGC に基づくヒューリスティックを適用することで発話内の語の好感度を推定する”という提案手法について、“人間が発話音声聞いて語の好感度を直接推定した結果”と比較することで、提案手法の妥当性について評価する。

図 5 に評価実験の構成を示す。提案手法は“話者感情の推定処理”と“話者感情を用いた好感度推定処理”の 2 つの処理から構成される。そこでまず、評価実験用に収集した音声発話データに対して話者感情の推定処理を行う。そして推定した話者感情を入力として、3.4 節で提案した好感度推定手法を適用する。そして、推定された好感度を、実験参加者が発話音声から直接推定した好感度と比較することで、提案手法の妥当性を

図5 評価実験の構成

図6 提案手法の評価方法

評価する。このとき、本来一つの発話において発話者一人の好感度のみが存在するはずのところ、図6に示すように、聞き手によって異なる好感度が推定される状況が起こりうる。これは、「話し手は快である」と感じた聞き手と「話し手は不快である」と感じた聞き手がそれぞれの解釈に基づいて異なる好感度推定をしている状況を意味する。

以下、4.2 節では評価実験用の発話音声の収集方法について説明する。4.3 節では、提案手法を適用する際の入力となる“実験参加者が推定した話者感情”の収集手順および本実験に必要なデータベースの構築方法について説明する。4.4 節では提案手法の出力の妥当性を評価する際の指標となる“実験参加者が発話音声聞いて直接推定した好感度”の収集手順について説明する。4.5 節で実験結果について述べる。

4.2 評価実験用発話音声の収集

まず、人間が発話内に含まれる語の好感度を推定することが可能と思われる発話事例を、テレビアニメ 490 分（一話約 190 発話 × 20 話 = 約 3,800 発話）およびテレビドラマ 965 分（一話約 870 発話 × 22 話 = 約 19,000 発話）から、「はい」や「ありがとう」のような間投詞文や好感度が未定の語を 2 つ以上含む文などを除外することで 3,652 発話収集した。評価用発話データは、収集した 3,652 発話のうち提案手法の適用に必須となる「格要素が直前の発話あるいは話者情報から補完可能」かつ「必須格要素が節構造になっていない」という条件を考慮しながら評価作業の負担にならない分量として人手で無作為に選

定した 61 件である。

なお話者情報による補完とは、補完対象が話者自身あるいは話している相手であることを示す。今回は実験の前処理として人手で格要素の補完処理を行ったが、ゼロ代名詞を補完する研究としては東山 [21] や飯田 [22] などがあるため、将来的には自動で補完を行うことを想定している。

抽出した音声発話はサンプリングレート 44.1 kHz で録音された。なお音声データはテレビアニメと海外ドラマ（日本語吹き替え）から収集しており、61 発話の発話者（声優）数は男性 11 名、女性 14 名の計 25 名、1 声優あたりの発話数は 1～8 発話である。

次に、作業者 1 名により好感度推定対象事物を 1 つ事前に選定した。選定は、提示する発話の主節中の主体、客体、目標、is-a 関係にある語から行われている。

例：「ハヤテ君全然学校に来ないから」

→「ハヤテ君」を好感度推定対象事物と設定

4.3 提案手法の適用準備

本実験において好感度推定処理を適用する際には、**実験参加者それぞれの推定話者感情を入力として好感度推定処理を適用し、推定された好感度を同じ実験参加者が発話音声から直接推定した好感度と比較**することで提案手法の評価を行う（図 5）。本節では、提案手法を適用する際の入力となる“実験参加者が推定した話者感情”の収集手順および実験事例に提案手法を適用するために必要となる“好感度推定対象事物以外の好感度”の設定手順について述べる。

人間による発話音声からの話者感情推定手順

4.2 節で収集した発話音声データに対して、以下の手順で実験参加者 5 名（男子大学生および男子大学院生 4 名、女子大学生 1 名）による話者感情推定の結果を収集した。

- (1) 対象となる発話音声を再生し、実験参加者に提示する。好感度推定対象事物の部分は 1000 Hz のサイン波でマスクした。
例：「(ビー) 全然学校に来ないから」
- (2) この発話音声を聞いて、“話者はどのような感情で発話したと思うか”を“快”、“不快”、“快でも不快でもない (neither) ”、“判断できない”の 4 項目から 1 つを選択する。このとき「判断に悩むものは“判断できない”を選択する」よう教示した。

実験参加者 5 名による発話音声からの話者感情推定の結果を表 5 に示す。表 5 より、全員が快あるいは不快と判断した発話は 61 発話中 28 発話、4 名が同じ回答をした発話は 17 発話であった。一方、同じ発話音声に対して快と不快両方の回答があった発話が 10 発話、2 名以上が neither あるいは“判断できない”と回答した発話が 6 発話あった。

評価実験用データベースの構築

提案手法を適用するにあたって必要となる“好感度推定対象

表 5 人間による話者感情推定結果

実験参加者が推定した話者感情	発話数	
	快	不快
5名全員が同じ回答（快あるいは不快）	15	13
4名が同じ回答&1名が“判断できない”	8	5
4名が同じ回答&1名が“neither”	1	3
快と不快両方の回答あり	10	
3名が同じ回答&2名が“判断できない”	1	1
3名が同じ回答&1名が“判断できない”&1名が“neither”	0	2
2名が同じ回答&1名が“neither”&2名が“判断できない”	0	1
1名が“neither”&4名が“判断できない”	1	

表 6 質問紙調査によって設定した語の好感度

一致回答数	語の好感度
4人一致	positive：優しい、きれい、満ちあふれる、すごい、カッコいい、魅力的、守る negative：粗い、ひどい、怖い、イカれる、モルモット（にする）、丸つぶれ、泣く
3人一致	negative：重い、おしまい

表 7 設定した動詞の行動タイプ

接近	回避
来る、行く、居る、通う、抱く、会う、（薬を）打つ、戻る、作る	消える

事物以外の発話内の事物の好感度”については、日本語評価極性辞書 [2, 3] を参照した。日本語評価極性辞書に記載が無かった語については、実験参加者 4 名（男子大学生および男子大学院生 3 名、女子大学生 1 名）に質問紙調査を行った結果に基づいて値を設定した。表 6 に質問紙調査に基づいて好感度を設定した語と値を示す。なお 3 人一致の事例において残り一人の回答はいずれも neutral であった。

また動詞の行動タイプは、岡田 [19] の“事象概念の分類データ”の表に基づき設定した。表に無いもの、または、語義により複数の行動タイプが設定されているものについては、発話文における各語の語義を考慮して人手により表 7 のように設定した。

4.4 人間による発話音声からの好感度推定手順

4.2 節で収集した発話音声データに対して、4.3 節と同じ実験参加者 5 名によって対象事物の好感度を直接推定してもらった。実験参加者は発話音声を聞いて好感度推定対象事物の好感度を回答した。回答は“positive”、“neutral”、“negative”、“判断できない”の 4 項目から 1 つを選択した。

収集した 305 件（= 61 発話 × 5 名）の回答全てを対象に、話者感情推定結果と対象事物の好感度推定結果の組み合わせを分析した結果を表 8 に示す。組み合わせは 9 つあり、A～I で示す。表 8 より、発話音声から話者感情を快か不快のいずれかと

推定した回答 (A, B, C) は 305 件中 263 件であった。一方、推定話者感情が **neither** あるいは判断不可の場合の評定結果 (D, E, F, G, H, I) の回答傾向は実験参加者によって大きく異なっていた。

4.5 実験結果

実験結果を表 9 に示す。3.4 節で述べたように、提案する好感度推定手法では話者感情の推定結果が快あるいは不快であるときのみ手法を適用するため、入力となる推定話者感情が **neither** あるいは判断不可であった 42 件 (= 8 件 + 34 件) については提案手法を適用していない。

提案手法において対象事物の好感度が **positive** あるいは **negative** と推定された 263 件 (= 193 件 + 70 件) から推定好感度が判断不可と回答された 12 件 (= 《pos-判断不可》9 件 + 《neg-判断不可》3 件) を除外した 251 件のうち、191 件 (= 《pos-pos》141 件 + 《neg-neg》50 件) の好感度推定結果が人間による推定好感度の回答と一致した (精度 0.76)。また、人間により **positive** あるいは **negative** と判定された 218 件 (= 《pos-pos》141 件 + 《neg-pos》8 件 + 《pos-neg》19 件 + 《neg-neg》50 件) のうち、191 件 (= 《pos-pos》141 件 + 《neg-neg》50 件) の好感度が正しく推定された (再現率 0.88)。

なお提案手法が **positive** と **negative** を人間の好感度推定結果と逆に推定した回答 27 件 (= 《pos-neg》19 件 + 《neg-pos》8 件) のうち 15 件は、実験参加者 5 名全員が同じ好感度と判断しているにもかかわらず、提案手法が逆の好感度と推定した 3 事例によるものである。また、実験参加者全員が話者感情を快あるいは不快と判定しているにもかかわらず、好感度推定結果で 3 人以上の実験参加者が **neutral** と回答した発話事例も 2 件存在した。これらの発話事例については 5 章で考察する。

5. 考察

提案手法による好感度推定が失敗した発話事例について分析を行った。まず、提案手法による好感度推定結果と人間による好感度推定結果の極性が反転していた発話事例について以下に示す。

「(不快そうに) X はあまり得意じゃないのだ」
 「(不快そうに) X だけはダメなんです」
 「(不快そうに) X はなんて感じの悪い医者なの？」

いずれの発話でも、主語 X に対して「ダメ」や「感じの悪い」というネガティブな属性が付与されている。提案手法では X にネガティブな評価が下されている状況に対して話者が不快感を示している状況では話者は X に好感を持っていると判定する。しかし、3.4 節で述べたように、本論文では表出された快／不快は全て発話が表す出来事に対して表出された感情として処理を行うこととしているため、**表出された不快が発話が表す出来事に対する悲しみや怒りなのか、それとも発話中の事物 X に対する嫌悪なのかを区別できない**。今後口調から分類できる感情クラスの種類を増やし、悲しみ、怒り、嫌悪を正しく区別できるようになれば、本事例のような失敗を回避できることが期待できる。

表 8 人間による推定話者感情と対象事物の推定好感度

記号	推定話者感情	推定好感度	実験参加者					合計
			P1	P2	P3	P4	P5	
A	快 or 不快	pos or neg	40	40	41	51	46	218
B	快 or 不快	neutral	5	12	5	6	5	33
C	快 or 不快	判断不可	5	0	6	0	1	12
D	neither	pos or neg	0	0	1	0	0	1
E	neither	neutral	0	0	7	0	0	7
F	neither	判断不可	0	0	0	0	0	0
G	判断不可	pos or neg	3	4	0	1	0	8
H	判断不可	neutral	0	5	0	3	4	12
I	判断不可	判断不可	8	0	1	0	5	14

(pos: positive, neg: negative)

表 9 提案手法による好感度推定結果

		人間による好感度推定結果				計
		pos	neu	neg	判断不可	
提案手法による 好感度推定結果	pos	141	24	19	9	193
	neg	8	9	50	3	70
推定話者感情= neither		0	7	1	0	8
推定話者感情= 判断不可		6	12	2	14	34
計		155	52	72	26	305

(pos: positive, neu: neutral, neg: negative)

次に話者感情が判明し対象事物の好感度も推定可能であるにも関わらず 3 人以上の実験参加者が好感度を **neutral** と回答した発話事例を以下に示す。

「(不快そうに) お嬢様は二度と X をしなくなる」
 「(快そうに) X も正常に機能してるわ」

これらの事例はどちらも行動や状態そのものに対する話者の悲しみや喜びが表出されていると考えられる。そのため好感度推定対象である事物 X について『必要なものや重要なものではあるが、必ずしも好きなものであるとは限らない』という判断がなされることもある。このような状況に対応するためには、好感度推定対象事物に関連する概念知識や会話トピックの背景知識による常識推論などの機能が必要である。

また、提案手法では“実験参加者が推定した話者感情”を用いているが、提案手法を社会実装するうえで想定される主な利用状況として“対話システムとの対話音声からのユーザの嗜好情報の推定”が考えられる。機械学習分類器を用いた提案手法の自動化に対する検討については、付録に付す。

6. まとめ

本論文では、対話の中から相手は事物に対して持っている好感度情報を推定するため、発話音声の口調から推定した話者感情と発話内容に基づくヒューリスティックに基づいて発話に含まれる事物の好感度を算出する手法を提案した。【X に良い事があったので喜んでいるならば、X の事が好きである】や【X に悪い事があったので喜んでいるならば、X の事が嫌いである】などといったヒューリスティックを表現するために、EGC をベースとした好感度推定手法を提案した。また、発話時に話者が体験している感情を直接知ることは難しいため、代用とし

て他者によって推定された話者感情を用いる手法を提案した。話者感情として“快”あるいは“不快”を検出し、かつ EGC が適用可能な構文構造の発話を行った場合、発話内の事物に対して話者が感じている好感度を推定することが可能となった。

実験では、収集した発話音声聞いて実験参加者 5 名が推定した話者感情を入力として提案手法を適用した結果と、各実験参加者が直接推定した好感度を比較した。その結果、251 件のうち 191 件の好感度推定結果が人間による好感度推定結果と一致した (精度 0.76)。また、人間が positive あるいは negative と好感度を推定した 218 件のうち、191 件を提案手法によって正しく再現できた (再現率 0.88)。

今後の課題としては、まず扱う話者感情クラスの種類を増やすことがある。本手法では同じ不快として扱われている感情も、より詳しく感情クラスを識別できることで好感度推定手法をそのまま適用可能か否かを判別できることが期待できる。また、「X をしてみたい」や「X のせいで」のように感情が付随しなくても言語表現のみで好感度推定が可能な表現についても検討する予定である。

謝辞

本研究は、国立研究開発法人科学技術振興機構 (JST) の研究成果展開事業「センター・オブ・イノベーション (COI) プログラム」及び JSPS 科研費 26330313 の助成を受けたものである。

参 考 文 献

- [1] S. Baccianella, A. Esuli, and F. Sebastiani: “SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining,” *Proc. of the Int. Conf. on Language Resources and Evaluation (LREC 2010)*, pp. 2200-2204, 2010.
- [2] 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.3, pp. 203-222, 2005.
- [3] 東山昌彦, 乾健太郎, 松本裕治: “述語の選択選好性に着目した名詞評価極性の獲得,” *言語処理学会第 14 回年次大会論文集*, pp. 584-587, 2008.
- [4] 目良和也, 市村匠, 相沢輝昭, 山下利之: “語の好感度に基づく自然言語発話からの情緒生起手法,” *人工知能学会論文誌*, Vol.17, No.3, pp. 186-195, 2002.
- [5] 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.
- [6] 平野徹, 小林のぞみ, 東中竜一郎, 牧野俊朗, 松尾義博: “パーソナライズ可能な対話システムのためのユーザ情報抽出,” *人工知能学会論文誌*, Vol.31, No.1, pp. 1-10, 2016.
- [7] 縄手優矢, 稲葉通将, 高谷智哉, 山田整, 高橋健一: “雑談対話ログを用いた話者の潜在的興味対象の推定,” *2017 年度人工知能学会全国大会論文集*, 1B2OS25b3, 2017.
- [8] 徳久良子, 乾健太郎, 徳久雅人, 岡田直之: “規模とコストを考慮した感情タグつき言語コーパスの作成方法,” *電子情報通信学会総合大会ワークショップ「心を持つロボット, 対話するロボット」講演論文集*, pp. 514-515, 2001.
- [9] 小林峻也, 萩原将文: “ユーザの嗜好や人間関係を考慮する非タスク指向型対話システム,” *人工知能学会論文誌*, Vol.31, No.1, DSF-A_1-10, 2016.
- [10] B. Schuller, N. Köhler, R. Müller, and G. Rigoll: “Recognition of Interest in Human Conversational Speech,” *INTERSPEECH 2006*, pp. 793-796, 2006.
- [11] W. Y. Wang, F. Biadys, A. Rosenberg, and J. Hirschberg: “Automatic detection of speaker state: Lexical, prosodic, and pho-

netic approaches to level-of-interest and intoxication classification,” *Computer Speech & Language*, Vol.27, No.1, pp. 168-189, 2013.

- [12] 西本遥人, 駒谷和範: “対話におけるマルチモーダル情報を用いたユーザの興味の有無の推定,” *2018 年度人工知能学会全国大会論文集*, 3C2-OS-14b-04, 2018.
- [13] 倉野大二郎, 松村耕平, 角康之: “マルチモーダルデータによる携帯映像閲覧者の興味推定,” *情報処理学会研究報告ユビキタスコンピューティングシステム (UBI)*, Vol.2013-UBI-37, No.30, pp. 1-8, 2013.
- [14] J. A. Russell: “A Circumplex Model of Affect,” *J. of Personality and Social Psychology*, Vol.39, pp. 1161-1178, 1980.
- [15] CeVIO Official Web: <http://cevio.jp/> [accessed January 15, 2019]
- [16] R. A. Fisher: *Statistical Methods for Research Workers*, Oliver and Boyd, 1954.
- [17] Y. Goldberg: *Neural Network Methods for Natural Language Processing (Synthesis Lectures on Human Language Technologies)*, Morgan & Claypool Publishers, 2017.
- [18] M. B. Arnold: *Emotion and Personality*, Columbia University Press, 1960.
- [19] 岡田直之: 語の概念の表現と蓄積, 電子情報通信学会, 1991.
- [20] 笹野達平, 河原大輔, 黒橋禎夫, 奥村学: “構文・述語項構造解析システム KNP の解析の流れと特徴,” *言語処理学会第 19 回年次大会論文集*, pp. 110-113, 2013.
- [21] 東山英治, 伝康晴: “談話におけるジェスチャー —ゼロ代名詞への同期／非同期事例の比較,” *言語・音声理解と対話処理研究会予稿集*, Vol.59, pp. 43-48, 2010.
- [22] 飯田龍, 乾健太郎, 松本裕治: “文脈的手がかりを考慮した機械学習による日本語ゼロ代名詞の先行詞同定 (自然言語),” *情報処理学会論文誌*, Vol.45, No.3, pp. 906-918, 2004.
- [23] B. Schuller, S. Steidl, and A. Batliner: “The INTERSPEECH 2009 Emotion Challenge,” *INTERSPEECH 2009*, pp. 312-315, 2009.
- [24] F. Eyben, M. Wöllmer, and B. Schuller: “openSMILE —The Munich Versatile and Fast Open-Source Audio Feature Extractor,” *Proc. of the 9th ACM Int. Conf. on Multimedia, MM 2010*, pp. 1459-1462, 2010.
- [25] 丹後俊郎, 山岡和枝, 高木晴良: ロジスティック回帰分析 —SAS を利用した統計解析の実例—, 朝倉書店, pp. 198-200, 1996.
- [26] 有本泰子, 河津宏美: “音声チャットを利用したオンラインゲーム感情音声コーパス,” *日本音響学会 2013 年秋季研究発表会講演論文集*, 1-P-46a, 2013.

(2018 年 11 月 15 日 受付)

(2019 年 4 月 17 日 採録)

付録

A. 音響的特徴からの話者感情推定

本節では、発話音声の音響的特徴量に基づいて話者感情を {快, 不快, neither} のいずれかのクラスと判別する機械学習分類器を用いた話者感情推定手法および自動推定した話者感情を提案手法に適用した結果について述べる。

A.1 機械学習分類器を用いた話者感情の自動推定

本実験では、“The INTERSPEECH 2009 Emotion Challenge”でベースラインとして提示された、発話音声から計算される静的な音響的特徴量を Support Vector Machine (SVM) に学習させる手法 [23] を用いた。機械学習に用いる音響的特徴量としては、オープンソースソフトウェア openSMILE [24] の IS09 特徴量セットで算出される 384 種類の音響的特徴量に話速を表す特徴量 (= モーラ数 ÷ 発話秒数) を加えた 385 種類の特徴量を Forward Stepwise Selection (FSS) 法 [25] によって絞

り込んだものを用いた。特徴量絞り込み用の発話データとしては、“感情評定値付きオンラインゲーム音声チャットコーパス (OGVC)” [26] から、3 人の評価者全員の評価が一致した自発対話音声を用いた。内訳は「喜び」172 発話、「平静」115 発話、「negative」183 発話（悲しみ 48、怒り 59、恐れ 33、嫌悪 43）である。Leave-one-out 交差検定の正解率を基準として FSS 法を適用 (wrapper 法) したところ、特徴量数 26 個で正解率が最大となったため、本実験ではこの 26 個の音響特徴量セットを用いて機械学習分類器を構築した。

SVM を用いた話者感情推定実験では、表 5 において 5 名中 4 名以上の実験参加者が話者感情を **快** と推定した音声 24 件、5 名中 4 名以上の実験参加者が話者感情を **不快** と推定した音声 21 件に、同様の手順で収集した **快でも不快でもない発話音声 30 件** を加えた計 75 件に対して Leave-one-out 方式で SVM (RBF カーネルを使用) による機械学習分類実験を行った。なお、各発話音声から音響的特徴量を計算する際にはマスキング前の音声データを用いている。また音声データは、各変量を平均が 0.0、分散が 1.0 になるように規準化を行った。SVM を用いた話者感情の推定結果を表 10 に示す。表 10 より、快感情、不快感情、neither の推定精度はそれぞれ 0.80、0.68、0.91 であった。また快感情、不快感情、neither の再現率はそれぞれ 0.67、0.71、1.00 であった。そして全体での正解率は 0.81 となった。

A.2 自動推定した話者感情を用いた好感度推定

機械学習分類器を用いた話者感情推定処理では一つの発話音声から推定される話者感情は 1 パターンとなることから、本節の実験では実験参加者 5 人による推定話者感情も推定好感度も完全一致したデータのみを用いて評価を行った。

機械学習分類器で話者感情を推定した 45 件のうち、全実験参加者による推定話者感情も推定好感度も完全に一致している 21 件の発話事例を対象として実験を行った。結果を表 11 に示す。21 件のうち 17 件の好感度が提案手法により正しく推定された (再現率 0.81)。また、提案手法により対象事物の好感度が positive あるいは negative と推定された 19 件のうち 17 件が、人間による好感度推定結果の回答と一致した (精度 0.89)。

本節の実験では実験参加者 5 人による推定話者感情も推定好感度も完全一致したデータのみを用いたため実験データ数が 21 件と少なくなった。そのため実験の信頼性としては十分とはいえないものの、実験結果の数値としては精度 0.89、再現率 0.81 という高い数値が得られた。このことから、今後提案手法を自動化するにあたって機械学習分類器による話者推定を導入する手法は有効であると考えられる。

表 10 SVM を用いた話者感情推定結果

		SVMを用いた 話者感情推定結果			再現率
		快	不快	neither	
人間による 話者感情推定結果	快	16	7	1	0.67
	不快	4	15	2	0.71
	neither	0	0	30	1.00
精度		0.80	0.68	0.91	

表 11 機械学習による感情推定結果からの好感度推定結果

		人間による好感度推定結果	
		positive	negative
機械学習に よる推定感 情を用いた 好感度推定 結果	positive	14	1
	negative	1	3
	推定感情=neither のため提案手法を 適用せず	1	1

[問い合わせ先]

〒731-3194 広島県広島市安佐南区大塚東 3-4-1

広島市立大学 大学院情報科学研究科

目良 和也

TEL: 082-830-1677

E-mail: mera@hiroshima-cu.ac.jp

著者紹介

めら かずや
目良 和也 [正会員]

1996 年九州工業大学大学院情報工学研究科博士前期課程修了。同年、広島市立大学情報科学部助手。2003 年東京都立科学技術大学大学院工学研究科インテリジェントシステム専攻博士課程修了。博士 (学術)。2007 年より広島市立大学情報科学部助教、現在に至る。感情音声分析、マルチモーダル感情推定、感情を考慮したコミュニケーションシステムの研究に従事。人工知能学会、日本音響学会、言語処理学会、情報処理学会、IEEE 各会員。

あおやま まさと
青山 正人 [非会員]

1996 年岡山大学大学院自然科学研究科博士課程修了。博士 (工学)。同年、広島市立大学情報科学部助手。2007 同大学院情報科学研究科講師、2010 同准教授。現在に至る。2000 シカゴ大学放射線科カートロスマン放射線像研究所客員研究員 (文部省在外研究員)。コンピュータ診断支援、コンピュータビジョンなどの研究に従事。電子情報通信学会、情報処理学会、日本医用画像工学会、日本放射線技術学会各会員。

くろさわ よしあき
黒澤 義明 [正会員]

1994 年京都大学大学院人間・環境学研究科修士課程修了。修士 (人間・環境学)。1994 年広島市立大学情報科学部助手。2007 年より広島市立大学情報科学研究科助教、現在に至る。自然言語処理の研究に従事。言語処理学会会員。

たけざわ としゆき
竹澤 寿幸 [非会員]

1984 年早稲田大学理工学部電気工学科卒業。
1989 年同大学大学院理工学研究科博士後期課程修了。工学博士。1987 年早稲田大学情報科学研究教育センター助手。1989 年（株）ATR 自動翻訳電話研究所研究員。音声対話翻訳の研究開発に従事。2007 年より広島市立大学大学院情報科学研究科教授。現在に至る。音声対話や観光情報学の研究と教育に従事。観光情報学会、電子情報通信学会、情報処理学会、人工知能学会、日本音響学会、言語処理学会各会員。

Estimating Personal Preference Information on the Basis of Relationship Between Content and Tone of Utterance

by

Kazuya MERA, Masahito AOYAMA, Yoshiaki KUROSAWA, and Toshiyuki TAKEZAWA

Abstract:

In order to consider user's emotion and feeling on human-computer interaction system, it is important to understand the user's personal preference. However, it is difficult to grasp personal preference information completely because it may differ among people and it can be changed easily by obtained knowledge and experiences. Although there are some database about preference and evaluation, most of them deal with "general" preference information. Nevertheless, people can estimate whether the partner likes an object from his/her utterances. For example, when a person happily says "X won the championship," we can estimate that he/she likes X. On the other hand, when a person gloomily says "X won the championship," we can estimate that he/she does not like X. In this paper, we propose a method to estimate like-dislike polarity for an object in an utterance by using such heuristics on the basis of the case frame structure of the utterance and the speaker's emotion. The heuristics are expressed by the Emotion Generating Calculations method (EGC). The proposed method is applied into the utterance when the speaker expresses pleasure or displeasure. In the experiment, the proposed method calculated like-dislike polarity of a word in the utterance by using the speaker's emotion estimated by a participant and the calculated polarity was compared with the polarity manually estimated by the same participant. The precision and recall of emotion estimation process were 0.76 and 0.88, respectively.

Keywords: personal preference information, tone of utterance, speaker's emotion, emotion generating calculations method

Contact Address: **Kazuya MERA**

Graduate School of Information Sciences, Hiroshima City University

3-4-1 Ozuka-higashi, Asa-minami-ku, Hiroshima 731-3194, Japan

TEL: +81-82-830-1677

E-mail: mera@hiroshima-cu.ac.jp