

# 演劇台本における登場人物間の親密度推定手法<sup>†</sup>

松本 和幸<sup>\*1</sup>・秋田 恭佑<sup>\*2</sup>・任 福継<sup>\*1</sup>・吉田 稔<sup>\*1</sup>・北 研二<sup>\*1</sup>

近年、音声アシスタント機能を搭載した携帯型端末が普及し、より使い手に配慮した対話システムが求められている。従来型の対話システムの問題点として、雑談のような非タスク型対話への対応が未熟な点があげられる。非タスク型対話においては、ユーザとの会話を円滑かつ柔軟にするための工夫が必要となる。たとえば、ユーザの現実世界での人間関係を考慮することによって、ユーザと親しい人物に関する話題の提供を積極的に行ったり、ユーザと親しくない（仲が良くない）人物に関する話題の提供を避けたりすることができる。本論文では、演劇台本を題材に、対話中の2者間の人間関係を「親密度」という尺度により表現する。親密度に関わると考えられる要素として、発話の応答回数や発話中の態度などがある。本論文では、その中でも発話中の感情状態に着目することで、高精度な親密度推定の実現を試みる。評価実験の結果、発話役割に基づく従来手法を上回る高精度な親密度推定を実現することが出来た。

キーワード：親密度、シナリオ対話、感情状態

## 1. はじめに

近年、スマートフォンをはじめとする携帯型情報端末が爆発的に普及し、既に多くの人々が日常的に使用している。これらの情報端末には様々な機能が備わっており、その1つに、音声アシスタント機能がある。これは、話し言葉による問いかけや要求に対し、適切な回答や動作を行う、音声認識と自然言語処理技術を用いたタスク型対話システムといえる。このように、人間と機械との対話は、音声インタフェースや、大規模データを用いて学習する人工知能技術を採用することで、より高精度かつ身近な存在となってきた。しかし、肝心の会話内容は、雑談のような非タスク型対話になると、一辺倒（変化のない）な受け答えや、ランダムな応答となってしまう、人間的な対話を再現できていないと難しい。

対話破綻検出チャレンジ [1] では、雑談対話システムの応答文が破綻していないかどうかを検出するタスクが設定され、様々なアプローチによる破綻検出の試みがなされている。こうしたことから、今後、雑談対話システムの技術的な発展が期待できる。しかしながら、対話の破綻が検出でき、正しい対話が可能になったとしても、ユーザを飽きさせない、雑談相手となるような対話システムが実現できるとは限らない。雑談において重要なのは、ユーザの属性や興味、さらには人間関係を考慮した話題提供である。しかし、属性や興味は、そう毎日変動するものではないため、提供する話題の方向性がワンパターン化してしまうといった問題もある。

本論文では、人同士の対話から、その仲の良さの度合い（親密度）を計算する方法を提案する。たとえば、SNS 上の2者間

の対話履歴を収集して仲の良さを解析することも可能であると思われる。しかしながら、SNS 上や現実世界において人間関係が劇的に変化することは稀であり、SNS やチャットを解析する場合、長期的に大規模なデータを蓄積し、分析することが必要になる。本研究では、仲の良さの変化や、人間関係に影響するイベントについての情報を比較的容易に得ることができる対象として、演劇台本に着目する。演劇台本は、台詞に対する発話者が明記されている。また、周囲の状況を示す書き情報も含まれる。このような情報と、親密度の推定結果とを組み合わせることにより、複合的な観点からの分析が可能と考える。

馬場ら [2] は、小説テキストを対象として小説の登場人物の情報を抽出する手法を提案している。この研究で提案されている手法では、小説テキストを入力し、そこから登場人物名を抽出し、関連度の計算を行うことにより、登場人物間の関係性から物語をモデル化している。しかし、小説テキストは演劇台本のように台詞主体で構成されているわけではないことや、発言者が明示されていなかったり、人物名の表記ゆれが存在するため、抽出自体が難しいといった問題がある。また、この手法では、小説中の重要人物とそれに関連する人物といった相関図を得ることが可能であるが、人物間の人間関係（親子関係、友好関係、敵対関係）などはわからない。小説の場合は、人物間の関係（友人関係など）を文章で明示することで表現することも多い。しかし、現実の対話においては、親しさを明示的に表すことはほとんどない。以上のことから、小説テキストは本研究の分析対象としない。

提案する手法では、現実世界でのメールや電子チャットでのやり取りと同様、親しく、交流が頻繁であることを推測可能な手がかりとして、発話に対する返答回数を用いる。また、親しくない（交流が無い）という状態以外に、交流はあるがあまり仲が良くない（口論が絶えない）関係をも考慮するため、発話における発話者に生起しているポジティブな感情、ネガティブな感情といった感情状態を考慮する。対話における発話者の感情状態を考慮するためには、発話文に対し、発話者の感情を推測し、付与する必要がある。Matsumoto ら [3] は、対話におけ

<sup>†</sup> Intimacy Estimation of the Characters in Drama Scenario  
Kazuyuki MATSUMOTO, Kyosuke AKITA, Fuji REN,  
Minoru YOSHIDA, and Kenji KITA

\*1 徳島大学理工学部  
Faculty of Science and Technology, Tokushima University

\*2 株式会社日立システムズ  
Hitachi Systems, Ltd.

る感情推定を目的として演劇台本データに感情タグを含む各種タグを付与したシナリオコーパスを構築している。本論文では、このシナリオコーパスをもとに、登場人物の感情状態を考慮した人物間の親密度推定手法を提案する。

## 2. 関連研究

### 2.1 人物間の親密度推定

関連研究として、西原ら [4] や阿部ら [5] が人物間の親密度 (仲の良さ) を推定する手法を提案している。西原らは、会話の発話役割の種類数が多いほど親しいと仮定し、発話役割の種類数を用いて 2 者間の親密度を推定している。発話文から抽出した特徴から発話役割を自動同定し、その結果を親密度計算に用いている。この研究では、発話役割の自動同定において、発話文から抽出する特徴として、助詞・助動詞の組み合わせを用いている。発話役割の同定実験の結果、対象となる発話役割の種類によって再現率・適合率に違いが見られ、同定の困難なもの易しいものが存在すると述べている。しかし、西原らは、話者の発話の傾向が同定できれば良いと考え、こうした違いは誤差になると判断している。

また、同定した結果を用いて、親密度の推定を行っている。式 (1) に親密度算出の式を示す。式中、 $F(S_a, S_b)$  は親密度、 $k(S_a), k(S_b)$  は発話役割の種類数、 $l(S_a), l(S_b)$  は発話数、 $S_a$  と  $S_b$  は話者を示している。

$$F(S_a, S_b) = \frac{k(S_a) \times k(S_b)}{l(S_a) \times l(S_b)} \quad (1)$$

この研究における親密度推定の評価には、親密度に関するアンケートと、親密度推定結果の相関を用いている。この評価の結果、一人の発話者との発話における相関の平均値は 0.564 となり、様々な話者間での発話における相関の平均値は 0.467 となった。西原らの手法では、親密度推定の過程で発話役割を用いているため、会話における発話内容から発話に込められた意図を認識しているといえる。しかし、西原らは、発話役割に基づく手法では、テーマの無い会話においての適切な人間関係の推定は困難であると述べている。つまり、チャットや音声対話による雑談 (非タスク会話) に対してどの程度有効であるかは分からない。

阿部らは、メールの送受信履歴を用いて、対象となる人物から見た主観的な仲の良さの順位付けを行っている。そして、推定した仲の良さランキングと、被験者のアンケート結果から得た仲の良さの順位をレーベンシュタイン距離を用いて比較し精度を測っている。送信や受信などの動作 (行動) にそれぞれ重みを設定し、以下の式 (2) によって順位が計算される。

$$i = \sum_{k=1}^l m_k \times x_k \quad (2)$$

$i$  は親密度を示し、 $l$  は送受信メールの総数、それぞれのメールの送り方・受け取り方におけるメール数は  $m_k$ 、重みを  $x_k$  と表現している。阿部らの手法では、メールの内容までは考慮せずにメールの回数という簡単な特徴のみで親密度の推定を行っている。そのため、会話内容が複雑でも推定に影響しないという利点がある。しかし、この手法では、一時的に業務など、仲

の良さとは無関係なメールを交わしている場合には、誤った推定結果となってしまう。

本研究では、親密度推定の手がかりの 1 つとして話者の発話時の感情を用いる。一般に、親しい間柄であれば、対話においてネガティブな感情よりポジティブな感情のほうが多く出現すると考えられる。そのため、親密度推定においても、発話者の感情が特徴として有効であると考えた。

また、Gustarini ら [6] は、円滑な情報共有を目的として、携帯電話から収集したコンテキストデータから日常における人の親密さの認知を推定する手法を提案している。しかし、携帯電話の通話時間などの情報のみからでは、発話内容や発話時の感情状態は推測できないことから、より詳しい情報を得るためには発話そのものを取得すべきと考えた。

また、近年の SNS の普及から、SNS におけるつながりをもとに人物間の親密度を分析した研究も存在する [7]。この研究では、SNS 上の投稿頻度と親密度の関係を調べるために、SNS の統計的分析調査を行い、投稿数が多いユーザとそうでないユーザとの間で、明確な親密度の違いを示すことができた。対話における発話の返答回数を調べることは、ある記事に対する投稿数 (どれだけコメントのやりとりがあるか) を調べることに、同等の意味を持つと考えた。このことから、本研究では、親密度推定の手がかりの 1 つとして、対話における発話の返答回数を用いる。

また、豊田ら [8] の研究では、発話状態時間長に着目して対話における雰囲気推定している。この研究では、二者間の対話の雰囲気を「盛り上がり」、「まじめさ」、「噛み合い」、「明るさ」、「親密さ」、「対等さ」でラベル付けしている。本研究では、人物間の親密度の推定を重視しているため、豊田らの研究で定義される「親密」な雰囲気であっても実際には、人物間の仲が良いわけではない場合をも考慮する必要がある。演劇においては、演者の顔の表情が読み取りづらいため、動作、声色、台詞などを用いて人物の感情を表現することで観客に伝わりやすくしていると考えられる。そのため、演劇の台詞から、人物間の対話における感情の生起を検出し、その感情の生起状態に基づき人物間の親密度を判断できると考えた。本研究では、感情の生起の程度や感情の種類 (ポジティブ・ネガティブ) と発話の返答回数を組み合わせた親密度推定手法を提案する。

### 2.2 対話における人物の感情状態推定

従来から、人と人の対話における感情状態の推定を目指した研究がいくつか存在する。松本ら [9] は、対話シナリオにおける人物の感情状態を、感情状態の継続性を考慮して推定する手法を提案した。

また、Matsumoto ら [3] は、感情の変化のパターンを用いて、発話相手の感情状態を予測するアプローチをとっている。これらの研究では、主に対話中の発話文や話の流れをもとに感情を推定しようとしている。しかし、発話文中の話者の内的な状態を明確に定義することは難しく、表層的な感情表現や感情の変化パターンのみから感情状態を推定することは難しいと考えられる。感情状態をより高精度に推定するためには、対話相手との関係や、周囲の状況、発話役割等、様々な手がかりを利用す

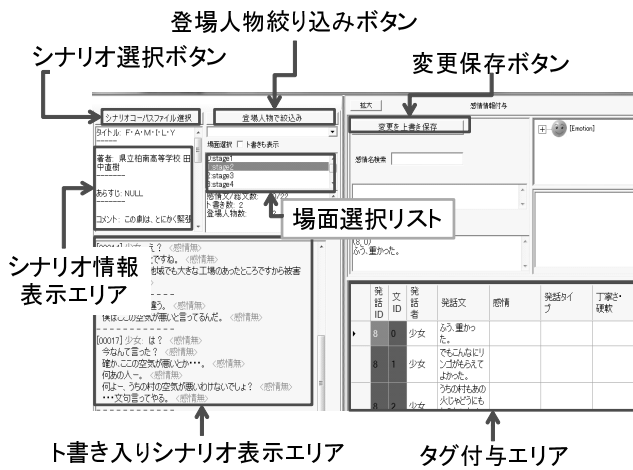


図1 シナリオコーパスエディタのインタフェース画面

必要がある。人物の感情は発話文脈によって様々に変化し、人物の性格等に依存するとも考えられ、様々な研究者が、表情・音声・言語などからの対話における感情推定の研究に取り組んでいるが、多くの研究が実世界の人間とロボットとの対話を対象としている。人間が会話をするとき、表情や音声から感情を読み取ったり、表現する傾向があり、言語のみからの感情推定では不十分となる。

本研究では、登場人物間の親密度推定が目的であり、その手掛かりとするのが台詞であるため、感情を考慮する場合、言語情報のみから推測する必要がある。しかし、既存研究の結果から、言語情報のみからの感情推定の精度はあまり高いとはいえないため、親密度推定の対象となる発話文において、話者の感情状態が既知の状態を想定し、あらかじめ発話文単位で付与された感情タグを用い、感情状態を推定して用いた場合との比較を行うことで、推定された感情による親密度推定の有効性を検証する。

### 3. 親密度推定手法

われわれは日々、意識せずに互いの親しさを感じ取りながら生活している。この親しさを第三者の視点から推定するために、発話文の特徴を大まかに2種類に分け、それらを用いた手法を提案する。本節では、まず、発話文の特徴について述べ、つぎに、それらの特徴を用いた提案手法について説明する。

#### 3.1 発話に含まれる特徴

表層的特徴や発話内容から得られる特徴を用いて、演劇の登場人物間の親密度を推定する手法を提案する。前章でも述べたように、発話文から得られる特徴は、大まかには以下のように分けられる。

- 表層的特徴
  - ・ 発話回数
  - ・ 返答回数
- 発話内容から得られる特徴
  - ・ 感情
  - ・ 発話役割

表1 シナリオコーパスに収録されている対話例

発話文	話者	感情	発話役割
あの、どちらからいらっしたのですか？	少女	平静	質問
ここもか	旅人	絶望	価値判断
え？	少女	驚き	質問
ああ、戦争のことですね	少女	平静	理解

表2 感情一覧

p/n/0	感情カテゴリ	感情タグ
p	喜び (Joy)	喜び (joy), 興奮 (excitement), 充足 (sufficiency), 満足 (pride), 歓喜 (ecstasy), 安心 (relief)
p	愛 (Love)	愛 (love), 情熱 (passion), 愛慕 (adoration)
0	平静 (Neutral)	平静 (neutral)
n	恐れ (Fear)	恐れ (fear), 不安 (anxiety)
n	怒り (Anger)	怒り (anger), 憤り (resentment), 失望 (frustration), 悪意 (despite), 妬み (jealousy), つらさ (bitter)
n	悲しみ (Sorrow)	悲しみ (sorrow), 悲嘆 (heartbreak), 失望 (disappointment), 恥辱 (shame), 哀しみ (pity), 絶望 (despair)
0	驚き (Surprise)	驚き (surprise)

前者の表層的特徴とは、会話文を見たときに現れる文字数であったり、会話数や返答回数など発話の内容に影響されない特徴のことである。2.1節で述べたように、阿部らは、Eメールを対象にメール回数のみを用いて親密度の推定を行った。結果的に、被験者にとって最も親しい人物の親密度は高いという結果が得られていることから、会話内容に関わらず、親密さとメールやり取りの回数はある程度関係性が見られると考えられる。後者の会話内容から得られる特徴とは、発話文に含まれる感情や意図(発話役割)や態度のことであり、一般に、形態素解析により単語単位に分割した後、辞書等を用いて感情や意図の推定が行われている。

#### 3.2 提案手法

##### 3.2.1 シナリオコーパス

まず提案手法で用いるシナリオコーパスについて説明する。Matsumotoら[3]が構築したシナリオコーパスは、発話文に話者・感情・発話役割・文体のタグが人手により付与される。発話文間の区切りとして、句読点「。」「…」「!」「?」を用いる。

図1に、コーパスへのタグ付与用に使ったツールのインタフェース画面の一例を示す。今回、このコーパスのうち、最も発話文数が多く、主人公が明確であり、2者間の会話が比較的多くのシーンで交わされるシナリオを、実験対象とする<sup>1</sup>。表1にシナリオコーパスに登録されている発話文の例を示す。また、表2に収録されている感情の一覧、表3に発話役割の一覧、表4に、文体の一覧を示す。表3の発話役割の中に、「応答」というタグが存在するが、本研究では、この「応答」は、何

1 各シーンの区切りはシナリオに明記されている。

表3 発話役割一覧

発話役割カテゴリ	発話役割タグ
挨拶	感謝, 謝罪, あいさつ
応答	同意, 理解, 否定, 反論, 呼びかけ, 納得
質問・問いかけ	質問, 確認, 自問
価値判断	価値判断
知識	知識提供, 知識獲得
独り言・叫び	独り言, 叫び
命令	命令
依頼・要求	依頼, 行動要求受入, 発言要求受入
第三者を対象とした発話	第三者対象
意味を持たない発話	引用・曖昧な発話
事実	事実
予定	予定, 勧誘
思ったこと・説明など	思ったこと・説明など, 補足
提案	提案
その他	復唱, 禁止, 気付き

表4 文体一覧

文体カテゴリ	文体タグ
1	とても硬くて丁寧
2	どちらかといえば硬くて丁寧
3-1	どちらかといえば軟らかくて丁寧
3-2	どちらかといえば軟らかくてくだけている
4-1	とても軟らかくて丁寧
4-2	とても軟らかくてくだけている

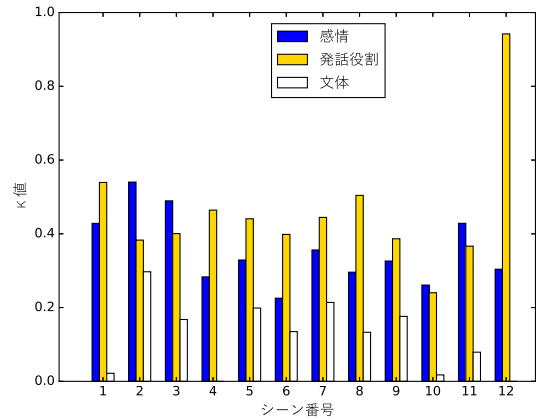
かの問いかけに対する応答を意味し、発話における「返答」とは異なるものとして考える。表2の、p/n/0の列において、「p」はポジティブ感情、「n」はネガティブ感情、「0」はそれ以外のニュートラル感情を表している。ポジティブ感情とネガティブ感情は、フィッシャーの提唱する感情系統図 [10] を参考に決定した。

本コーパスは、5名の評定者によりタグが付与された。評定者間のシーンごとの一致率（κ値）[11]の平均値を調べると、図2に示すとおりとなった。図中の横軸は、シナリオのシーン番号を示している。感情タグの一致率は平均しても0.3程度であり、あまり高くない結果となった。しかし、感情タグの一致率が0.41を超える（適度に一致していることを示す）シーンも存在し、これらのシーンは、ある程度の信頼性があると判断できる。文体は人物間の関係を見る際に有効な特徴と考えられるが、今回の調査により、一致率が低いため、被験者による判断の個人差がそのほかのタグよりも大きいと考えられる。したがって、文体タグは本論文では親密度推定に用いる特徴量としては用いないことにした。

5名の評定者が付与したタグをカテゴリに変換したのち、発話文ごとに集計し、この結果、各発話文ごとに最も多く出現したカテゴリを今回用いるカテゴリとした。最も多く出現したカテゴリが複数ある場合は、それらのカテゴリをすべて用いた。

3.2.2 表層的特徴を用いた親密度推定手法

表層的特徴を用いる大きなメリットの1つは、特徴を得る



シーン番号	1	2	3	4	5	6	7	8	9	10	11	12	平均値
感情	0.428	0.540	0.489	0.283	0.329	0.225	0.356	0.296	0.326	0.261	0.428	0.304	0.356
発話役割	0.538	0.383	0.400	0.464	0.441	0.398	0.445	0.504	0.387	0.240	0.366	0.942	0.459
文体	0.022	0.297	0.168	0.000	0.199	0.135	0.214	0.133	0.176	0.017	0.079	0.000	0.129

図2 タグ付与における評定者間のκ値のシーン別平均値

表5 会話例

2者間の会話	
発話者	発話文
A	何言ってるんだ？
B	あれ？
B	違うの？
A	この村で一番広い平野は？
B	平野？
B	平野って程じゃないけど昔畑だった原っぱがあるわよ。
A	そこに僕を連れて行ってくれ。
B	いいけど、何すんの？
3者間の会話	
発話者	発話文
C	その人、お姉ちゃんの恋人？
D	ちっ、ちがうわよっ。
C	あー赤くなってる。
C	やっぱりそうなんだ。
D	こらっ。
E	あっ、さっきのチョウチョ。

ことが容易な点である。たとえば、日本語の言語処理の研究では、形態素解析を行い、発話文を単語に分割してからさまざまな分析を行うことが一般的であるが、表層的特徴から分析する場合、単語に分割しなくてもよい。そのため、口語や俗語、未知語などが用いられやすい日常会話の発話文に対して、形態素解析による分割誤りの影響を受けない推定ができる。本研究では、発話回数と返答回数に着目し、返答率という特徴を算出し、これを用いて親密度推定を行う。返答回数のカウント方法は、会話から話者が交替した箇所を抽出し、前の話者に対して返答したもとしてカウントする。表5に会話例を示す。一般に、話者交替の箇所での発話文の組は、発話の順番交替の隣接ペア [12] と考えられる。そのため、隣接ペアにおいては必ずしも実際の返答ではない場合もあるが、本研究で対象としているシナリオコーパス中では、3人以上の登場人物が出現するシーンであっても3者以上が参加している対話は少なく、ほぼ2者間の

対話となっていた。さらに、3者以上の対話において、ほとんどの返答が直前発話に対するものであったため、この返答回数のカウント方法を採用することによる実際の返答回数との誤差は少なく、影響がほとんど無いと考えた。

この例で説明する場合、会話を上から下へと順に読んでいくと、まず、話者AとBの2者間の会話について、話者がAからB、またはBからAに交替していることが見て取れる。話者の交替時にBはA、またはAはBに対して返答したと仮定する。このように考えると、この対話例において、A、Bそれぞれの返答回数は、AからBに対しては2回、BからAには3回となる。また、発話回数は、Aが3回、Bが5回となる。

次に、話者C、D、Eの3者間の会話例について、話者がC、D、Eが交替しているが、返答回数は、DからCに対しては2回、CからDに対しては1回となる。話者Eは、2者間の会話からは外れているが、Eから直前のDへの返答をしていると見なして返答回数をカウントする。

一般に、「会話はキャッチボール」などといわれるように、一方的な会話より、双方の発話量が同程度となるほうが、親密さが高くなると考えられる。実際、小川[13]によって2者間の発話量が均衡している会話には好印象を抱くことが確認されている。本研究では、発話回数や返答回数が均衡していると、親しさに好影響を及ぼすと仮定する。式(3)に、返答率の算出計算式を、式(4)に親密度  $ID_{RR}$  の計算式を示す。RRは返答率を表している。

$$RR(S_a, S_b) = \frac{r(S_a, S_b)}{f(S_a)} \quad (3)$$

$$ID_{RR}(S_a, S_b) = 1 - |RR(S_a, S_b) - RR(S_b, S_a)| \quad (4)$$

ここで、 $S_a, S_b$  は登場人物を示し、 $r$  は返答回数、 $f$  は発話回数を示す。 $ID_{RR}$  は親密度を示す。たとえば、 $r(S_a, S_b)$  はAがBに対して返答した数を示し、 $f(S_a)$  はAの発話回数を表すことになる。ここで、 $r(S_a, S_b)$  と  $r(S_b, S_a)$  を別々に計算する理由について述べる。基本的には、 $S_a$  と  $S_b$  の二者間の対話を想定しているが、三者以上の人数で対話が行われることもある。もし、話者  $S_c$  という話者が対話に途中参加した場合、たとえば、次のような順： $S_a \rightarrow S_b \rightarrow S_c \rightarrow S_a \rightarrow S_b$  で対話が進んだ場合、 $r(S_a, S_b) = 0$ 、 $r(S_b, S_a) = 2$  というように計算される。実際の対話では、誰が誰に対して返答したのかは、順序だけでは分からないが、本研究では、直前の自分以外の発話者に対して返答していることとしてカウントする。

### 3.2.3 感情情報を用いた親密度推定手法

発話文を解析して得られる特徴についてのメリットは、表層的特徴より多くの情報が得られるという点である。表層的特徴のみでは、発話の内容までは考慮していないため、発話に込められた思いや意味を得ることは困難である。しかし、発話に込められた感情や役割といった情報は、日常会話においても重要な情報であるため、親しさを推定する特徴の1つとして有効だと考えられる。

本研究では、感情を用いた親密度推定手法について提案する。親密度の推定には、ポジティブ感情出現率 (PP) という特

徴を算出し、これを用いることで親密度を推定する。式(5)に、ポジティブ感情出現率の計算式を示す。

$$PP(S_a, S_b) = \frac{\sum_{e_i \in E_{a,b}} p(S_a, S_b, e_i) \times w(e_i)}{\sum_{e_i \in E_{a,b}} p(S_a, S_b, e_i)} \quad (5)$$

PPはポジティブ感情出現率を表し、 $S_a, S_b$  は登場人物、 $e_i$  は感情の種類を示す。また、 $p(S_a, S_b, e_i)$  は、 $S_a$  から  $S_b$  への発話に込められた感情  $e_i$  の出現回数を示す。 $w$  は、各感情に対する重みを示している。 $E_{a,b}$  は、 $S_a$  から  $S_b$  への発話において出現した感情の集合を表す。本手法では、表2に示したように、各感情をネガティブな感情とそれ以外の感情に分け、ネガティブな感情には低い重み、それ以外には高い重みを設定した。この理由として、ネガティブな感情は、親密度を低下させる要因となり、それ以外の感情 (ポジティブ、ニュートラル) は、親密度の低下には影響しないと考えたからである。

続いて、式(6)に感情情報を用いた親密度推定手法について示す。 $ID_{PP}$  は親密度を示している。

$$ID_{PP}(S_a, S_b) = \frac{PP(S_a, S_b) + PP(S_b, S_a)}{2} \quad (6)$$

### 3.2.4 返答率と感情情報を組み合わせた手法

会話における雰囲気は、話者間の仲の良さにも影響されやすいため、前節で述べたポジティブ感情返答率が親密度推定に有効であることが予想される。しかし、仲が良いほど相互に同情や共感が起こりやすくなると考えられるため、親密な人物間の会話において、ネガティブな感情 (「悲しみ」など) が頻出することも想定される。そのため、感情返答率だけでなく、返答率に基づく親密度推定手法を組み合わせることで、前述の「同情や共感によるネガティブ感情の頻出」の影響を軽減できると考える。

式(7)に、返答率に基づく親密度計算とポジティブ返答率に基づく親密度計算を組み合わせた場合の親密度計算の式を示す。 $ID_{(RR,PP)}$  は親密度を表している。

$$ID_{(RR,PP)}(S_a, S_b) = ID_{RR}(S_a, S_b) \times ID_{PP}(S_a, S_b) \quad (7)$$

この計算式により、2者間の返答率の差が少なく、ポジティブ返答率の平均値が高いほど、2者間の親密度が高くなる。

## 3.3 感情情報の自動付与

### 3.3.1 感情語辞書に基づく自動付与

発話文に対し、感情情報を人手により付与したシナリオコーパスの場合、その感情情報をもとに、親密度推定を行うことは可能である。しかし、親密度の分析を行うために、すべての演劇台本の台詞に対して感情情報を人手により付与することはコスト面から考えて現実的ではない。そのため、本節では、シナリオコーパスへの感情タグの自動付与を試みる。ここでは、単語に感情情報を付与した辞書 (以下、感情語辞書と呼ぶ) とのマッチングによって行う。感情推定手法 [9] を参考にし、感情語辞書を用いて感情推定手法による感情情報の自動付与実験を行う。感情の推定には、各感情語辞書を参照して推定した感情

値の平均を用いた。辞書 1 は日本語アブレイザル評価表現辞書 [14]，辞書 2 は感情表現辞典 [15]，辞書 3 は独自に作成したもので，日本語アブレイザル評価表現辞書に収録されている表現のうち，感情表現辞典に収録されている単語とコーパスにおいて共起する頻度をもとに感情カテゴリの重みを登録した辞書である。辞書 4 は高村ら [16] の構築した感情極性対応表（日本語版）である。その結果，ポジティブ/ネガティブ/ニュートラルの分類正答率は 28% を示した。個別に見ると，ネガティブが 93%，ポジティブが 7% の正答率であった。このことから，ネガティブな感情は，単語からの感情推定が比較的容易であり，ポジティブな感情は単語からの感情推定が困難であるということが分かる。この原因として，ポジティブな感情を単語のみで表現することがシナリオコーパスの対話においては少ないことや，辞書に登録されている語彙のうち，ネガティブな感情を示す語彙が多いことがあげられる。ここで，全発話文中で感情語辞書に収録されている単語が含まれている発話文数と各感情語辞書の収録単語数を表 6 に示す。

コーパス内の総発話文数は 898 文であった。そのうち，辞書に収録されている単語数が最も多い辞書 4 の場合で，コーパス中の発話文に含まれる語と一致したものは 513 文であり，感情の一致数はその半分以下の 245 文であった。これは，シナリオコーパスにおける発話文は当然ながら話し言葉で書かれたものであるため，驚きの声やあいづちなどといった，辞書に登録されていない言葉が多く含まれていることが原因と考えられる。表 7 にその例を示す。表中，感情語辞書中の語とマッチした単語に下線を引いている。

以上の結果から，シナリオのような連続する発話に対しては，従来の感情語辞書を用いた単語のみからの感情推定があまり有効でないことが分かる。

### 3.3.2 条件付確率場に基づく感情情報自動付与

松本ら [3] の提案する感情推定手法では，感情状態が対話において一定時間継続することを考慮し，感情語辞書や文型パターン辞書では推定不可能なタイプの発話文に対する感情推定を実現させている。しかし，対話の状況によっては，感情状態が急激に変化することも考えられ，また，発話者の性格などといった個別の要素を考えると，一定の感情状態の減衰は有効ではないと考えられる。

任 [17-19] の提案した心的状態遷移ネットワークは，被験者に対する実験により，ある感情状態から別の感情状態への遷移確率を，外的刺激が作用していない状態で求めたものである。しかし，対話における感情推定においては，外的刺激が必ず存在するため，モデルをそのまま適用することは難しい。

Matsumoto ら [3] が，現在の発話の直前までに，自身の発話および対話相手の発話において出現した感情状態を考慮して，感情を推定する手法を提案している。この手法では，感情状態の変化をポジティブ，ネガティブ感情の増減および変化なしの 5 種類に大別し，それぞれの変化がどの感情の出現によって誘発されるかを，それらの変化の回数と感情の出現回数をもとにコーパスから求めている。この手法の評価実験の結果から，提案手法による感情認識の精度はベースライン手法を大きく上

表 6 各辞書の語が含まれる発話文数と感情の一致した文数，各辞書の収録語数

	感情語辞書			
	1	2	3	4
発話文数	224	47	224	513
感情一致数	117	25	112	245
収録語数	8545	2342	3658	55125

表 7 会話例（相槌など）

話者	発話文
少女	あたしもやっ <b>て</b> いい？
旅人	え？
少女	だから，あたしも種を植 <b>え</b> るのを手伝 <b>っ</b> てもいい？
少女	っ <b>て</b> きいてるの
旅人	あ，あ <b>あ</b>
少女	ほん <b>と</b> ？
少女	やっ <b>た</b>

表 8 感情状態のパターン

ptn	条件
p	$(f_p - f_n > 1) \text{ AND } (f_p - f_z > 1)$
n	$(f_n - f_p > 1) \text{ AND } (f_n - f_z > 1)$
0	$(f_z - f_p > 1) \text{ AND } (f_z - f_n > 1)$
pn0	$( f_p - f_n  \leq 1) \text{ AND } ( f_n - f_z  \leq 1) \text{ AND } (f_p * f_n * f_z > 0)$
pn	$( f_p - f_n  \leq 1) \text{ AND } (f_n - f_z > 1) \text{ AND } (f_p * f_n > 0)$
p0	$( f_p - f_z  \leq 1) \text{ AND } (f_z - f_n > 1) \text{ AND } (f_p * f_z > 0)$
n0	$( f_n - f_z  \leq 1) \text{ AND } (f_z - f_p > 1) \text{ AND } (f_n * f_z > 0)$

回った。しかし，この手法においては，相手および自分の直前発話における感情が既知の場合を想定している。そのため，自己感情が未知の状態からの感情推定は不可能という点で，問題がある。

本研究では，直前の自己感情が未知の状態から感情推定を行う。また，感情状態のパターンを，ポジティブ/ネガティブ/ニュートラルといった単純な 3 種類ではなく，その強さの大小関係と組合せを考慮したパターンで表すことにする。また，直前発話を 2 つまで考慮することにより，推定精度を高める手法を用いる。

感情状態 ポジティブ (p)，ネガティブ (n)，ニュートラル (0) の 3 つの状態のパターンを，1 つの発話内におけるそれぞれの強度の違いで，表 8 のように 7 つのパターンに場合分けする。f<sub>p</sub>, f<sub>n</sub>, f<sub>z</sub> は，それぞれ発話におけるポジティブ感情，ネガティブ感情，ニュートラルの出現頻度を示す。今回，出現頻度は，複数の作業員によりコーパスに付与された感情タグの付与頻度を用いる。

この感情状態のパターンの遷移を，系列の学習を行うことができるアルゴリズムである条件付確率場 (Conditional Random Fields: CRF) により学習させ，直前の幾つかの発話における感情状態および現在の発話における感情状態から，次の感情状態を推定するモデルを構築する。本実験では，CRF の学習および計算ツールとして，CRF++ ver.0.58 [20] を，デフォルトのパラメータ設定で用いた。訓練データとして，Matsumoto ら

[3]の構築した感情タグが付与されている別のシナリオコーパス3種類(S-2, S-3, S-4)を用いて学習させた。シナリオコーパスの概要を表9に示す。

また、感情状態だけでなく、コーパスに付与済みの、発話役割、文体のデータも素性として用いる(図3)。今回、素性の組合せを複数試した。素性の種類について表10に示す。各素性パターン別の実験結果を表11に示す。素性は、 $x_{(i,j)}$ と表している場合に、現在の発話から*i*発話前の、素性ID *j*を用いることを示す。 $i=0$ は、現在の発話を示す。同様に、Bigram素性は、それらの組合せを素性として用いることを示す。提案手法は、直前の2発話までを感情推定の手がかりとして用いるため、 $-2 \leq i \leq 0$ であり、 $i=0$ のとき、素性ID  $j=4$ の感情カテゴリは未知であるため、素性として用いない。

この結果、直前の2つまでの(話者交替フラグ+発話役割+感情カテゴリ)をそれぞれ素性として用いた(素性パターンNo.6)場合に最も高い0.482の正答率が得られた。次に正答率が高かったのは、直前の2つまでの素性パターンNo.4(話者交替フラグ+発話役割+文体+感情カテゴリ)を用いた場合で、0.444の正答率が得られた。また、感情カテゴリを素性として用いない場合でも、直前の発話から得た素性を用いることで、最高で0.319の正答率(素性パターンNo.2)が得られた。また、素性パターンNo.3を用いた場合、正答率が0.293と最も低い結果が得られた。文体を素性に含めた素性パターンNo.4よりも、文体を除く素性パターンNo.6のほうが正答率が高かったことと、コーパスに付与されたタグのκ値を計算した際に、文体タグが最も低い一致率であったことから、このコーパスに付与されている文体タグは信頼性が低く、感情状態推定の素性としてはあまり有用でないと考えられる。

この結果より、CRFを用いた感情状態推定手法は、感情表現が含まれないような発話文の場合でも、用いる素性によっては約0.444の精度で感情推定を行えることがわかる。しかし、CRFを用いた感情状態推定手法では、直前の発話文における感情状態が既知でなければ得られない特徴も用いるため、発話文中に含まれる表現からの感情推定手法との併用が望ましい。CRFによる感情推定手法と、感情語辞書による感情推定手法とを組み合わせることで、感情タグが付与されていない状態からの推定も可能になる。しかし、感情語辞書により感情推定できるかどうかは、発話文の内容の難しさに依存する。このため、辞書以外の、たとえば感情タグが付与されたコーパスをもとに機械学習による感情タグ付け手法を組み合わせるなどの工夫が必要と考えられる。

### 3.4 提案手法の流れ

以下に提案手法の概要の流れ図で示す(図4)。図中、 $U_{A,1}, U_{B,1}, \dots$ は、シナリオコーパスにおける発話文、 $E_{A,1}, E_{B,1}, \dots$ は、それら発話文に付与されている感情タグ、同様に、 $S_{A,1}, S_{B,1}, \dots$ は、発話文に付与されている発話役割タグを示す。

- Step 1 シナリオコーパスから発話文情報を一文ずつ取得
- Step 2 発話文から発話者名、感情タグ、発話役割タグを抽出し、それぞれの情報をデータベースに格納

表9 学習に用いたシナリオコーパスの概要

	シナリオ番号		
	S-2	S-3	S-4
発話文数	239	411	344
文字数	2644	5247	3717
登場人物数	5	8	7
話者交替回数	148	225	196
シーン数	4	5	2

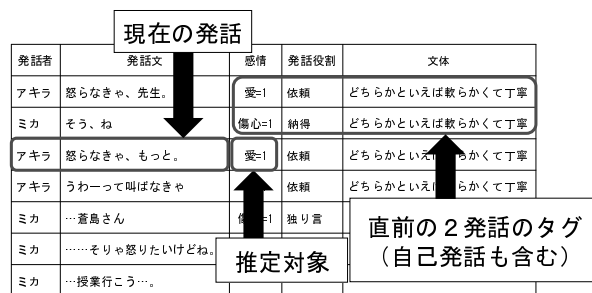


図3 直前の情報を用いた感情状態の推定

表10 素性の種類と例

素性ID	素性の種類	素性値の例
1	話者交替フラグ	0: 交替なし, 1: 交替あり, 2: 先頭の発話
2	発話役割	挨拶, 応答, 質問・問いかけ, etc.
3	文体	とても硬くて丁寧, どちらかといえば硬くて丁寧, etc.
4	感情カテゴリ	喜び, 悲しみ, 怒り, etc.

Step 3 全発話を入力し終えた後、シーンごとに返答率、ポジティブ感情出現率を計算

Step 4 返答率、ポジティブ感情出現率から親密度を推定

以上の流れで、各シーン、各登場人物間ごとに親密度の計算を行う。あるシーン内において、ある登場人物間に会話が存在しない場合には、その親密度の値は0となる。

## 4. 評価実験

本節では、提案手法の評価について述べる。まず、正解データとなる、親密度評価値を得るためのアンケートについて説明する。また、実験手順について述べたあと、実験結果を示し、その内容を分析し、考察する。さらに、話者の感情状態を自動推定した場合の評価もおこなう。

### 4.1 親密度評価値

推定した親密度が正しいかどうかを評価するために、実際に実験で用いる演劇台本に対して登場人物間の親しさに関するアンケートを行った。被験者は、情報工学の研究に携わる20代の大学院生10名(全員男性)とした。アンケートに回答してもらった項目はシーンごとの人物間の親しさであり、親しさを6段階の中から選択することにより回答してもらった。

親密度評価値は、6段階の回答をそれぞれ親密度に変換し、



表 11 CRF を用いた感情状態推定精度

素性 パターン No.	素性		正答率
	Unigram	Bigram	
1	$X_{(-2,1)}, X_{(-1,1)}, X_{(0,1)}$	-	0.301
2	$X_{(-2,1)}, X_{(-1,1)}, X_{(0,1)}, X_{(-2,2)}, X_{(-1,2)}, X_{(0,2)}$	-	0.319
3	$X_{(-2,1)}, X_{(-1,1)}, X_{(0,1)}, X_{(-2,2)}, X_{(-1,2)}, X_{(0,2)}, X_{(-2,3)}, X_{(-1,3)}, X_{(0,3)}$	-	0.293
4	$X_{(-2,1)}, X_{(-1,1)}, X_{(0,1)}, X_{(-2,2)}, X_{(-1,2)}, X_{(0,2)}, X_{(-2,3)}, X_{(-1,3)}, X_{(0,3)}, X_{(-1,4)}, X_{(-2,4)}$	-	0.444
5	$X_{(-2,1)}, X_{(-1,1)}, X_{(0,1)}, X_{(-2,2)}, X_{(-1,2)}, X_{(0,2)}, X_{(-2,3)}, X_{(-1,3)}, X_{(0,3)}, X_{(-1,4)}, X_{(-2,4)}$	$X_{(-2,1)} + X_{(-1,1)}, X_{(-1,1)} + X_{(0,1)}, X_{(-2,2)} + X_{(-1,2)}, X_{(-1,2)} + X_{(0,2)}, X_{(0,1)} + X_{(0,2)}, X_{(0,2)} + X_{(0,3)}, X_{(0,1)} + X_{(0,3)}, X_{(-1,1)} + X_{(-1,2)}, X_{(-1,2)} + X_{(-1,3)}, X_{(-1,1)} + X_{(-1,3)}, X_{(-2,1)} + X_{(-2,2)}, X_{(-2,2)} + X_{(-2,3)}, X_{(-2,1)} + X_{(-2,3)}$	0.402
6	$X_{(-2,1)}, X_{(-1,1)}, X_{(0,1)}, X_{(-2,2)}, X_{(-1,2)}, X_{(0,2)}, X_{(-1,4)}, X_{(-2,4)}$	-	0.482

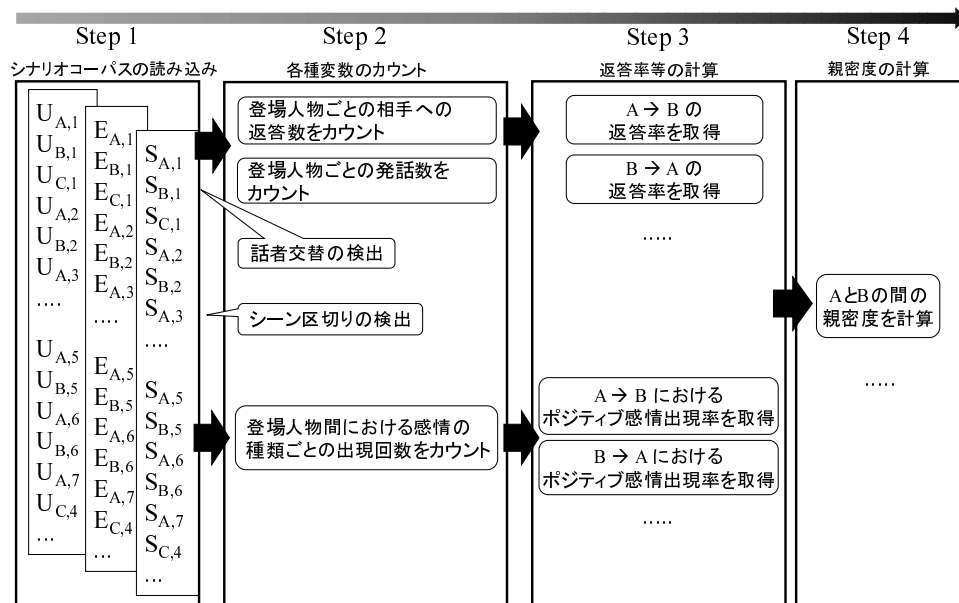


図 4 提案手法の流れ

表 12 アンケートにおける親密度レベルと親密度値

親密度レベル	親しさ	親密度値
5	親しい	1.0
4	まあまあ親しい	0.75
3	普通	0.5
2	あまり親しくない	0.25
1	親しくない	0.0
0	判断不可能	-

計算時に用いやすいよう数値化を行ったものである。表 12 に、親しさの 6 段階のレベルを示す。なお、レベル 0 の「判断不可能」という項目については、演劇台本のみから親しさを分類できるだけの情報が得られない（2 者間の会話が台本に記述されていない）ため判断できない場合とし、このレベルが選択された回答は、分析対象から除外する。

アンケートでは、被験者は演劇台本を一通り読んだあと、回答する。これは、読みながら回答してもらう形式だと、最初に

感じた親しさと読み進めたあとで感じた親しさが同じシーンでも変化する可能性があると考えたためである。

次に、被験者の回答結果の妥当性を重み付き  $\kappa$  係数を用いて評価したところ、結果は、平均値 0.648 を示した<sup>2</sup>。このことから、10 名の被験者の回答結果は、実質的に一致していると判断できる。

評価値は、上記でも説明したように 10 名の被験者の回答を表 12 に当てはめ、数値化したものに対して平均値を計算したものをを用いている。この評価値は、合計 73 通りの登場人物の組み合わせごとに計算した。組み合わせは、シナリオコーパスにおける各シーン内に登場する人物の、12 人についての組み合わせである<sup>3</sup>。なお、シーンに登場しない人物との組み合わせについては、アンケートの親密度の回答項目には含めない。

2 前述のとおり、被験者の回答結果として '0: 判断不可能' が選択されているものは除外した。

3 各シーンで会話があった登場人物の組合せ数の総和。



### 4.2 実験概要

実験で用いるシナリオコーパスの登場人物を表 13 に、シーン別発話文数と総登場人物数を図 5 に示す。

実験で用いる感情出現率の重みを表 14 に示す。ネガティブな感情の重みを 0 や負の値でなく 0.1 にした理由として、本研究で提案する親密度計算は、返答率と感情出現率の掛け合わせたものであり、感情出現率の重みを 0 や負の値にしてしまうと返答率により得られた親密度が打ち消されてしまい、まったく考慮されなくなってしまうためである。また、一般に、ネガティブな感情に関して、「喧嘩をするほど仲が良い」「親密な仲ほど悲しい会話ができる」などといったこと考えられ、「怒り」や「悲しみ」といった感情が発話において出現するからといって、二者がまったく親しくないということにはならないと考えたため、このように重みを設定した。

また、CRF に基づく感情状態の推定結果を用いての親密度推定も評価する。このとき、感情状態のパターンに対しては、感情カテゴリとは別に、表 15 に示すとおり重みを決定した。この重みは、前述の感情カテゴリごとの重みと同様の考え方で、各感情状態パターンにおいて、ポジティブ 'p' は 1.0、ニュートラル '0' は 1.0、ネガティブは '0.1' という重みを与え、各パターンごとに重みの平均値を計算することにより算出したものである。

この重みを用いて、ネガティブ感情が多く出現するほど PP が 0.1 に近づくように設定する。本研究では、以下に示す 5 種類の手法を用いた実験を行い、結果を比較する。

- $ID_{RR}$  を用いた手法 (R)
- $ID_{PP}$  を用いた手法 (E)
- $ID_{(RR,PP)}$  を用いた手法 (R+E)
- 返答率と、推定された感情状態を用いた手法  
( $R+E_{(1),(2),(3),(4),(5),(6)}$ )
- 発話役割を用いた手法 (従来手法) (T)

なお、本論文では以降、返答率を用いた推定手法を手法 R、感情を用いた推定手法を手法 E、両方を用いた推定手法を手法 R+E、推定された感情状態を用いた推定手法を、表 11 に示した素性の組合せごとに手法  $R+E_{(1)}$ ,  $R+E_{(2)}$ ,  $R+E_{(3)}$ ,  $R+E_{(4)}$ ,  $R+E_{(5)}$ ,  $R+E_{(6)}$ 、発話役割を用いた推定手法を手法 T と表記する。

各手法で得られた親密度とアンケートから得た評価値との間の平均誤差や相関を求めることで、各手法の妥当性を評価する。手法  $x$  の親密度推定結果における平均誤差の計算式を、式 (8) に示す。  $ans_i$  は、シーン  $i$  における、ある登場人物間の親密度評価値を示し、  $out_i^x$  は、シーン  $i$  において手法  $x$  により推定された親密度を示す。  $N$  は、親密度評価値が 0 以上となったシーンの総数を示す。

$$Error_{avg.}(x) = \frac{\sum_{i=1}^N |ans_i - out_i^x|}{N} \tag{8}$$

シーンごとの登場人物間の親密度の相関について、相関係数 ( $r_{cc}$ ) は、式 (9)、スピアマンの順位相関係数 (Spearman)

表 13 登場人物一覧

旅人, 少女, 男, 村人1, 村人2, 村人3, 声1, 声2, 声3, 声4, 子供1, 子供2

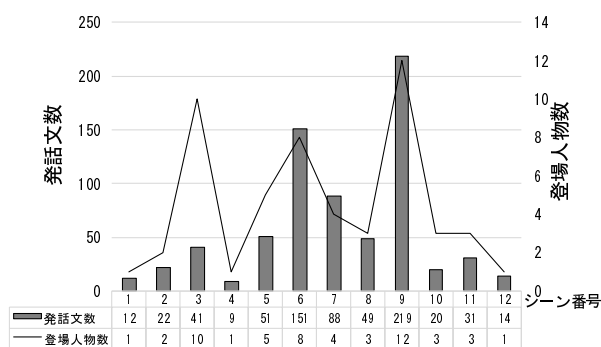


図 5 シーン別発話文数と登場人物数

表 14 各感情の重み

感情	重み
喜び (Joy)	1.0
愛 (Love)	1.0
平静 (Neutral)	1.0
驚き (Surprise)	1.0
悲しみ (Sorrow)	0.1
恐れ (Fear)	0.1
怒り (Anger)	0.1

表 15 各感情状態の重み

感情状態のパターン	重み
p	1.0
n	0.1
0	1.0
pn0	0.7
pn	0.55
p0	0.55
n0	0.55

は、式 (10) で表される式を用いて計算する。式 (9) 中の  $a_i, b_i$  は、各登場人物の、正解となる親密度の順位、推定された親密度の順位をそれぞれ表し、  $\bar{a}, \bar{b}$  は、正解となる親密度の平均順位、推定された親密度の平均順位をそれぞれ示す。  $n$  は、旅人との親密度が推定された登場人物数を示す。また、スピアマンの順位相関係数の計算においては、同順位を考慮し、同順位間での平均順位 (midrank) を用いて順位相関係数を求めた。以下、2 者間の親密度のシーン別の平均誤差に加え、シーンごとの登場人物間の親密度の相関による親密度推定の評価について述べる。

$$r_{cc} = \frac{\frac{1}{n} \sum_{i=1}^n (a_i - \bar{a})(b_i - \bar{b})}{\sqrt{\frac{1}{n} \sum_{i=1}^n (a_i - \bar{a})^2} \sqrt{\frac{1}{n} \sum_{i=1}^n (b_i - \bar{b})^2}} \tag{9}$$

$$Spearman = 1 - \frac{6 \sum_{i=1}^n (a_i - b_i)^2}{n(n^2 - 1)} \quad (10)$$

### 4.3 2者間のシーン別親密度の推定

ここでは、対象となるシナリオにおける主人公である“旅人”と“少女”を対象として、シーン別の親密度の差異について分析する。分析をこの2者に絞った理由は、台詞が他の登場人物と比較して多く、シナリオ全体を通して、ほぼすべてのシーンに登場しているためである。正解（親密度評価値）との平均誤差を式(8)により計算することで、各手法の比較を行う。

実験の結果，“旅人”と“少女”の会話があったシーンは2, 3, 5, 6, 8, 9の6つであった。そのため、本節での以下の結果は、6つの親密度に対して行った比較結果となる。表16に、本実験で得られた各手法ごとの平均誤差を示す。

感情状態推定した結果を用いない手法の中では，“R+E”による親密度推定手法が最も良い結果（最も低い平均誤差）を得た。また，“T”による親密度推定手法では、最も高い平均誤差が得られた。この結果から、提案手法の中で、表層の特徴と感情を組み合わせた手法が最も客観的な親しさの推定が出来ていると判断できる。

また、感情状態推定した結果を用いた手法“R+E<sub>(4),(5)</sub>”は、人手により付与されている感情タグを用いる手法“R+E”よりも平均誤差が低くなる結果となった。しかし、手法“R+E<sub>(1),(2),(3),(6)</sub>”は、“R+E”と同等かより高い平均誤差となった。これは、CRFによる感情状態推定実験において正答率が4割程度であり高くなく、正しく感情状態を推定できない場合が多かったためであると考えられる。しかし、文体タグによる精度低下の影響はあまり見られず、必ずしも、人手によりラベル付けされた感情状態を正確に再現していなくても、返答率や発話役割、ポジティブ感情返答率のみの単純な手法よりも良い結果が得られるということが分かった。CRFによる感情状態推定手法は、人手による感情タグそのものを推定しているわけではなく、ポジティブ、ネガティブ、ニュートラルのそれぞれの感情の出現頻度の大小関係のパターンを推定している。そして、その結果得られた感情状態パターンの重みをポジティブ感情返答率の計算に利用することにより、人手により付与した感情タグをそのまま用いるよりも親密度推定に適した特徴が得られたのではないかと推察する。実際、手法“R+E<sub>(5)</sub>”は、最も誤差が低くなった。ほかの手法と比較すると、用いた特徴量が複雑（Bigram素性も使用）ということもあるが、親密度推定に適した感情状態のパターンを推測するためには、一致率は低いが文体の情報も間接的には有効と考えられる。図6に親密度の推移を表す。縦軸が親密度を、横軸がシーン番号を示しており、“Answer”は、親密度評価値（正解）を示す。続いて、図7に“旅人”、“少女”の2者の感情（*pnvalue*）と親密度の関係を示す。縦軸は、親密度（親密度評価値）および感情のポジティブ/ネガティブの値（*pnvalue*）を示しており、横軸はシーン番号を示している。図中の棒グラフが親密度、折れ線グラフが“旅人”と“少女”の *pnvalue* を示す。シーン内においてポジティブな感情が多いほど、*pnvalue* は1に近くなり、ネガティブな

表 16 親密評価値に対する平均誤差の比較

手法	平均誤差	シーンごとの誤差					
		2	3	5	6	8	9
T	0.41	0.89	0.19	0.45	0.25	0.34	0.38
E	0.27	0.69	0.32	0.15	0.15	0.24	0.08
R	0.32	0.29	0.22	0.35	0.45	0.56	0.08
R+E	0.17	0.19	0.02	0.15	0.15	0.34	0.18
R+E <sub>(1)</sub>	0.25	0.14	0.17	0.32	0.06	0.22	0.61
R+E <sub>(2)</sub>	0.18	0.14	0.08	0.09	0.12	0.29	0.33
R+E <sub>(3)</sub>	0.17	0.24	0.01	0.09	0.06	0.22	0.42
R+E <sub>(4)</sub>	0.15	0.14	0.01	0.14	0.00	0.29	0.31
R+E <sub>(5)</sub>	0.13	0.11	0.02	0.01	0.00	0.22	0.40
R+E <sub>(6)</sub>	0.18	0.20	0.00	0.13	0.15	0.29	0.31

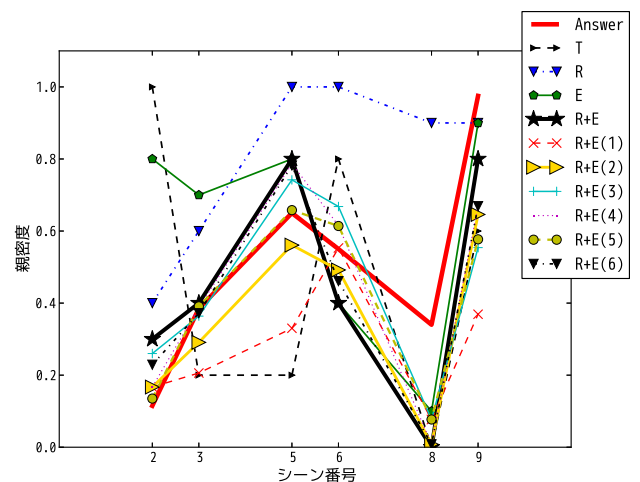


図 6 旅人と少女のシーン別親密度の推移

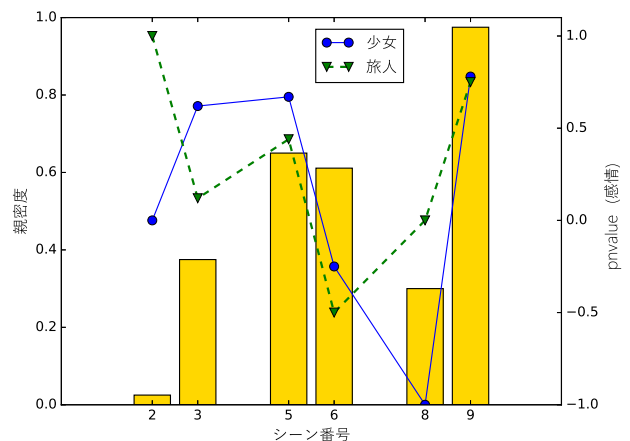


図 7 旅人と少女の感情と親密度

感情が多いほど *pnvalue* は -1 に近づく。*pnvalue* は、各シーンにおける旅人と少女間の会話中の全発話文におけるポジティブ感情、ネガティブ感情の付与頻度 ( $f_p, f_n$ ) をもとに、式(11)により計算した。

$$pnvalue = \frac{f_p - f_n}{f_p + f_n} \quad (11)$$

これを見ると、シーン6では2者ともにネガティブな感情が

多く出現している。また、シーン6の親密度もシーン5よりも低下していることから、感情極性（感情のポジティブ/ネガティブ）は親密度に影響を与えていることが分かる。実際、図6に示すように、両者の感情がネガティブ寄りに変化したタイミングで、2者間の親密度が低下している。

また、ここで、シーン単位でのコーパスへのタグ付与一致率と、親密度の平均誤差との関係について分析してみる。表16を見ると、手法“R+E”を用いた場合、図2において感情タグの一致率が平均値よりも高い（0.48）のシーン3において、平均誤差が0.02と低くなり、感情タグの一致率が平均値よりも低い（0.296）のシーン8において、平均誤差が0.34と、高くなったことが分かる。このことから、感情タグの一致率は、親密度推定にある程度の影響を与えることが分かる。

#### 4.4 同一シーン内における親密度の推定

本節では、対象を“旅人”に絞り、旅人とその他の登場人物との間の親しさの比較を行う。今回比較するシーンは、旅人と2名以上の登場人物との会話があったシーンの3、6、8、9、10を対象とした。各シーンの旅人と会話があった登場人物数を表17に示す。

上記の5つのシーン内での旅人と他登場人物との親密度を計算し、相関係数およびスピアマンの順位相関係数により、4つの手法による親密度推定結果と評価値との相関の有無を分析し、さらに、シーン内での平均誤差についても分析した。表18と表19、表20に、その結果を示す。

表18を見ると、相関係数の平均値が最も高かったのは手法“E”を用いた場合で、次に手法“R+E”を用いた場合であった。シーン別に見ると、手法“T”は、会話人数の多いシーン3、6において低い相関を示している。また、手法“R”は、シーン3、8において低い相関を示している。手法“R”による親密度推定には返答率の差を用いるため、親密度推定対象となる2者間の会話がほぼない場合でも返答率の差は小さくなり、結果として親密度が高くなってしまふ欠点がある。このことが低い相関を示す原因となったと考えられる。シーン3では、“旅人”はほとんど“少女”と会話していたためその他の登場人物との親密度は低いと一般的には考えられるが、返答率に差があまりないため、その他の登場人物との親密度は高くなってしまい、低い相関を示したと考えられる。一方、手法“E”は、シーン6以外で1に近い相関を示し、有効性を示せたといえる。シーン6における“旅人”と“少女”の会話は、物語が進み、二人がある程度親しくなった状態で、返答率は高くなっているが、ネガティブな話題がメインであるため、発話文の負の感情の影響を強く受けてしまい、親密度推定が上手くいかなかったと考えられる。

また、表19を見ると、手法“R+E”では、シーン8とシーン10において相関が1を示しており、同一シーン内の会話人数が2人ほどの少ない場合では親密度推定の特徴として有効であることが分かった。しかし、前述のとおり、“R”と同様、同一シーン内の会話人数が増えることで相関が低くなり、返答率を考慮しない“E”のほうが全体的に高い相関を示す傾向が見られた。しかし、シーン6、9のように会話人数が多くなる場合でも比較的高い相関が得られている。また、手法“E”が全体的

表17 旅人との会話が2名以上あったシーン

シーン番号	3	6	8	9	10
会話人数	6	6	2	5	2

表18 シーンごとの登場人物間親密度の手法別相関結果（相関係数）

シーン番号	手法			
	R	E	R+E	T
3	0.141	0.919	0.619	0.016
6	0.582	0.367	0.410	0.237
8	0.195	0.996	0.996	0.996
9	0.615	0.967	0.977	0.906
10	0.823	0.983	0.823	0.983
Avg.	0.471	0.846	0.765	0.628

表19 シーンごとの登場人物間親密度の手法別相関結果（スピアマンの順位相関係数）

シーン番号	手法			
	R	E	R+E	T
3	0.452	0.720	0.443	0.736
6	0.916	0.902	0.916	0.823
8	0.991	1.000	1.000	0.898
9	0.759	0.961	0.868	0.807
10	1.000	0.998	1.000	0.998
Avg.	0.824	0.916	0.845	0.852

表20 シーンごとの登場人物間親密度の手法別平均誤差

シーン番号	手法			
	R	E	R+E	T
3	0.368	0.140	0.093	0.126
6	0.315	0.120	0.115	0.260
8	0.089	0.031	0.072	0.041
9	0.188	0.098	0.109	0.141
10	0.058	0.149	0.058	0.149
Avg.	0.204	0.108	0.089	0.143

に高い順位相関が見られ、手法“T”もすべてのシーンにおいて0.7以上の相関が得られた。

また、表20から、手法“R+E”が最も平均誤差が低くなり、一方で手法“R”が最も高い平均誤差になったことから、感情情報の有効性がある程度示せたといえる。手法“R”の平均誤差が高くなったシーン3とシーン6は、ともに会話人数が多い（6人）のシーンであることから、手法“R”の欠点である、「返答率の差を利用しているため会話回数が少ない相手でも親密度が高くなる」ことが大きく影響したと考えられる。

全体として見ると、手法“R+E”が相関係数、順位相関係数、平均誤差のどの指標においてもある程度良い結果を示しており、感情と返答率を組み合わせることによる有効性が示せたといえる。手法“E”も、平均誤差では手法“R+E”には及ばないが、相関係数と順位相関係数において手法“R+E”よりも高い相関が得られ、感情情報を考慮することの有効性が明らかとなった。また同様に、既存手法である発話役割を用いた手法“T”もある程度有効であることが分かった。この結果から、手

法“R+E”と手法“T”を適切に組み合わせることができれば、より高精度な親密度推定が可能になると考えられる。

返答率の差を用いる手法“R”が3つの指標すべてにおいて最も悪い結果となったが、これは、返答率の差を用いることにより、会話回数の少ない人物間と会話回数の多い人物間での親密度に差がつきにくいことが大きく影響した結果と考えられる。これを改善する方法として2者間の返答率の平均値を用いる方法も考えられるが、今回は手法“E”および手法“R+E”において、ポジティブ返答率の平均値を用いたため、これらの手法では、前述の問題点のある程度回避できたのではないかと考える。

## 5. おわりに

本研究では、シナリオコーパスを対象に返答率と感情を用いた親密度推定手法を提案した。提案手法により、シナリオ上の会話から登場人物間の親しさを推定することが可能となった。返答率を用いた手法“R”、感情を用いた手法“E”、それらを組み合わせた“R+E”、そして比較対象となる従来手法である発話役割を用いた“T”の4種類の手法を用いて親密度推定の実験を行った。正解データである親しさアンケートの結果と比較することで、親密度の平均誤差を計算した。この結果、返答率と推定された感情状態を用いた“R+E<sub>(5)</sub>”手法で最も良い結果が得られた。また、文体を素性に含めて感情状態を推定した結果を用いた手法が、それ以外の手法によるものよりも高い性能が得られた。このことから、推定された感情状態が人による感情状態の判断と完全に一致する必要はなく、ある程度の精度で推定できれば有用であるということが分かった。

人手により付与した感情状態を用いた手法も、平均誤差において、手法“R”、“E”、“R+E”、“T”を上回る良好な結果を得ることができた。最も高い平均誤差を示したのは比較手法である“T”手法であった。このことから、対象を2者に限定したシーン別の親密度の推定には、会話回数と感情の組み合わせが特徴として有効であることが分かった。同一シーン内における登場人物間の親密度推定手法の比較実験では、手法“T”が各シーンで0.7以上のスピアマンの順位相関を示し、複数の人物から仲の良さをランキングする場合には発話役割も有効であることが確認できた。この比較実験において、最も高い平均誤差を示したのは手法“R”であった。これは、手法“R”は返答率RRの差を用いていることから、「会話回数が少ない相手でも親密度が高くなる」という欠点があったためと考えられる。

実際の対話中の発話文に人手により感情タグがあらかじめ付与されているような状況は少ない。しかし、CRFを用いた感情状態の推定の評価実験でも明らかになったとおり、直前の発話文における話者の感情状態が既知であれば、ある程度の精度は得られる。そのため、今後は、コーパスや文型パターン辞書に基づく発話文からの感情推定手法に基づき、発話文に感情状態の付与を行い、その感情状態を素性に加えてCRFによる感情状態の推定を行うことで、どの程度親密度推定の精度が向上するかを調査したい。さらに、話者間の発話における文脈情報の関連度を計算することにより、会話における雰囲気も考慮し、会話の流畅度をも考慮した親密度推定手法に改良したいと考え

ている。

また、本論文では、演劇台本といった創作された会話の中での親密度推定を対象とした。しかし、親密度推定の最終的な目標は、人とロボットとの対話への応用である。演劇台本における親密度推定手法を実世界の会話に活かすためには、実世界会話との違いも考慮しなければならない。たとえば、実対話では、仲の良い友人との対話において、第三者には理解し難いような(理由の分からないことで盛り上がる)といった場面も出てくると考えられる。こうした対話では、発話内容のみからでは、感情状態の推定は難しいと考えられる。そのため、発話時の音声や顔表情といったマルチモーダルな情報も必要となってくるであろう。さらに、実対話において、初対面の人には丁寧に接することで、こまめな返答が行われる(一時的に返答率が高くなる)ことも考えられる。そのため、「初対面」なのか「旧知の仲」なのか、といったことを判断する手がかりとして、これまでの対話履歴(いつ、どの程度の頻度でどのような会話をしたか)を保持しておき、それを解析して活用する必要もあると考えられる。このように、実対話と演劇台本との違いは多岐にわたるため、今後、提案手法を実対話に応用するにあたり、様々な観点からの検討が必要であろう。次の段階として、実対話応用に向けて、提案手法を用いてSNSを利用した複数ユーザ間の親密度推定を行い、チャットボットとの対話データに応用し、その有効性評価を行いたいと考えている。

## 謝辞

本研究の一部は、科学研究費補助金 15H01712, 15K16077 により行われた。

## 参考文献

- [1] 東中竜一郎, 船越孝太郎: “Project Next NLP 対話タスクにおける雑談対話データの収集と対話破綻アノテーション,” 人工知能学会 言語・音声理解と対話処理研究会 第72回, pp. 45-50, 2014.
- [2] 馬場こづえ, 藤井敦: “小説テキストを対象とした人物情報の抽出と体系化,” 言語処理学会第13回年次大会発表論文集, pp. 574-577, 2007.
- [3] K. Matsumoto, F. Ren, Q. Xiao, M. Yoshida, and K. Kita: “Emotion Predicting Method Based on Emotion State Change of Personae According to the Other’s Utterances,” *Proc. of the CCIS2014*, pp. 427-432, 2014.
- [4] 西原陽子, 砂山渡, 谷内田正彦: “発話テキストからの人間の仲の良さとして上下関係の推定,” 電子情報通信学会論文誌 D, 情報・システム J91-D(1), pp. 78-88, 2008.
- [5] 阿部有希, 川喜田佑介, 今田美幸, 鈴木悦子, 市川晴久: “Eメールの送受信履歴を用いた親密度推定手法,” 電子情報通信学会ソサイエティ大会講演論文集 通信 (2), p. 416, 2010.
- [6] M. Gustarini and K. Wac: “Estimating People Perception of Intimacy in Daily Life from Context Data Collected with Their Mobile Phone,” *Mobile Data Challenge by Nokia Workshop, co-located with the PERSASIVE Conf.*, Newcastle, 2012.
- [7] P. P. Rau, Q. Gao, and Y. Ding: “Relationship between the level of intimacy and lurking in online social network services,” *Computers in Human Behavior* 24, pp. 2757-2770, 2008.
- [8] 豊田薫, 宮越喜浩, 山西良典, 加藤昇平: “発話状態時間長に着目した対話雰囲気推定,” 人工知能学会論文誌, Vol.27, No.2, pp. 16-21, 2012.
- [9] 松本和幸, 三品賢一, 任福継, 黒岩真吾: “感情生起事象文型パターンに基づいた会話文からの感情推定手法,” 自然言語処理, Vol.14, No.3, pp. 239-271, 2007.

- [10] K. W. Fischer, P. Shaver, and P. Carnchan: "A skill approach to emotional development: From basic- to subordinate-category emotions," in W. Damon ed., *Child development today and tomorrow*, pp. 107-136, 1989.
- [11] J. Cohen: "A Coefficient of Agreement for Nominal Scales," *Educational and Psychological Measurement*, Vol.20, No.1, pp. 37-46, 1960.
- [12] 石崎雅人, 伝康晴: "談話と対話 (言語と計算)," 東京大学出版会, 2001.
- [13] 小川一美: "二者間発話量の均衡が観察者が抱く会話者と会話に対する印象に及ぼす効果," *実験社会心理学研究*, Vol.43, No.1, pp. 63-74, 2003.
- [14] 佐野大樹: "アブレイザル理論を基底とした評価表現の分類と辞書の構築," *国立国語研究所論集*, Vol.3, pp. 53-83, 2012.
- [15] 中村明: "感情表現辞典, 東京堂出版, 1993.
- [16] 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.
- [17] 任福継: "言語・表情など外観情報と心的状態遷移に基づく人間感情の認知について," *情報処理学会研究報告 (SLP)*, 音声言語情報処理, Vol.2006, No.73, pp. 43-48, 2006.
- [18] F. Ren: "Affective Information Processing and Recognizing Human Emotion," *Electronic Notes in Theoretical Computer Science*, Vol.225, No.2009, pp. 39-50, 2009.
- [19] F. Ren, C. Quan, and K. Matsumoto: "Enriching Mental Engineering," *Int. J. of Innovative Computing, Information and Control*, Vol.9, No.8, pp. 3271-3286, 2013.
- [20] CRF: Yet Another CRF toolkit: <https://taku910.github.io/crfpp/>  
(2017年10月24日 受付)  
(2018年3月1日 採録)

## [問い合わせ先]

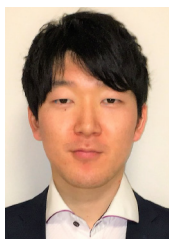
〒770-8506 徳島市南常三島町 2-1  
徳島大学 フロンティア研究センター 2F  
松本 和幸  
TEL: 088-656-7654  
FAX: 088-656-7654  
E-mail: matumoto@is.tokushima-u.ac.jp

## — 著者紹介 —



まつもと かずゆき  
松本 和幸 [非会員]

2008年3月徳島大学大学院工学研究科博士後期課程修了。博士(工学)。2009年10月より、徳島大学大学院ソシオテクノサイエンス研究部助教。2016年4月より現在まで、徳島大学大学院社会産業理工学研究部助教。感情計算、感性情報処理、自然言語処理、対話処理等の研究に従事。情報処理学会、電子情報通信学会、言語処理学会、電気学会、人工知能学会、ヒューマンインタフェース学会各会員。



あきた たけひと  
秋田 恭佑 [非会員]

2015年3月徳島大学大学院先端技術科学教育部システム創生工学専攻修士課程修了。在学中は自然言語処理を用いた俗語分析および対話処理研究に従事。2015年4月、株式会社日立システムズに入社。現在はERPシステムの導入、保守に従事。



にん ぶみ  
任 福継 [非会員]

1991年北海道大学大学院工学研究科博士後期課程修了。博士(工学)。広島市立大学助教授を経て、2001年より徳島大学工学部教授。2016年より現在まで徳島大学理工学部教授。自然言語処理、感性情報処理、人工知能、知能ロボットの研究に従事。日本工学アカデミーメンバー、EU科学アカデミーメンバー、電子情報通信学会フェロー、CAAIフェロー、日本工学会フェロー、情報処理学会、電気学会、人工知能学会、AAMT、IEEE各会員。



よしだ みつひろ  
吉田 稔 [非会員]

1998年東京大学理学部情報科学科卒業。2003年東京大学大学院理学系研究科博士課程修了。博士(理学)。東京大学情報基盤センター助教を経て、2013年より徳島大学講師。テキストマイニングの研究に従事。情報処理学会、言語処理学会、人工知能学会、日本データベース学会各会員。



きた けんじ  
北 研二 [非会員]

1981年早稲田大学理工学部数学科卒業。1983年沖電気工業株式会社入社。1989年カーネギーメロン大学機械翻訳研究所客員研究員。1992年徳島大学工学部講師。1993年同助教授。2000年同教授。2002年同大学高度情報化基盤センター教授。2008年同センター長。2010年より同大学大学院社会産業理工学研究部教授。博士(工学)。言語処理、情報検索、メディア情報学などの研究に従事。情報処理学会、電子情報通信学会各会員。

**Intimacy Estimation of the Characters in Drama Scenario**

by

**Kazuyuki MATSUMOTO, Kyosuke AKITA, Fuji REN, Minoru YOSHIDA, and Kenji KITA****Abstract:**

Recently, a portable digital device equipped with voice guidance has been widely used, increasing the demand for the usability-conscious dialogue system. One of the problems with the existing dialogue system is its immature application to non-task-oriented dialogue. Non-task-oriented dialogue requires some scheme that enables smooth and flexible conversations with a user. For example, it would be possible to provide topics related to a person who is familiar with a user or avoid providing topics related to a person who is not a good friend with the user, by considering relationship with others in real life of the user. In this paper, we focus on the dialogue made by the two characters in a drama scenario, and tried to express their relationship with a scale of "intimacy degree." There will be such various elements related to the intimacy degree as the frequency of response to the utterance and the attitude of a speaker during the dialogue. We focus on the emotional state of the speaker during the utterance and try to realize an intimacy estimation with high accuracy. As the evaluation result, we achieved high accuracy in intimacy estimation than the existing method based on speech role.

**Keywords:** intimacy degree, scenario dialogue, emotional stateContact Address: **Kazuyuki MATSUMOTO***2-1 Minamijosanjimacho, Tokushima 770-8506, Japan**2F Frontier Research Center*

TEL: +81-88-656-7654

FAX: +81-88-656-7654

E-mail: matumoto@is.tokushima-u.ac.jp