1I2-5

発話時間長に着目した Tree-Augmented Naive Bayes による対話雰囲気推定

Dialogue mood estimation focusing on utterance intervals by using Tree-Augmented Naive Bayes

豊田 薫 Kaoru Toyoda　宮越 喜浩 Yoshihiro Miyakoshi　山西 良典 Ryosuke Yamanishi　加藤 昇平 Shohei Kato

名古屋工業大学工学研究科情報工学専攻
Department of Computer Science and Engineering Graduate School of Engineering Nagoya Institute of Technology

In the field of the communication robots, many recent studies are focused on robots supporting dialogue communications. It seems important for the robots to estimate dialogue moods. We believe that as the robots estimate dialogue moods and behave suitably, communicating among humans with the robots will be smoothly. Therefore, this paper focuses on dialogues between two persons, and proposes an estimation method for the dialogue moods observed by the third party. Because we believe that the dialogue moods are influenced by utterance time, this method focuses and uses the utterance intervals features to estimate the dialogue moods, e.g., solitary utterance intervals, simultaneous utterance intervals, and silent intervals between two speakers. With these utterance intervals features and the subjective evaluations of spoken dialogue corpus, we constructed the estimation system for the dialogue moods by using Tree-Augmented Naive Bayes. In this paper, we confirmed the availability of the estimation method for dialogue mood, especially for dialogue mood "excitement".

1. はじめに

近年，人間とのコミュニケーションを目的としたロボットの研究が盛んに行われ，人間とコミュニケーションを取りながら施設を案内するロボットや高齢者とコミュニケーションを行うロボットなど，様々な場面でのロボットの活躍が期待されている．特に，ロボットとの自然な会話によって，幸福感や安心感といった心的作用を与えることを目的とした会話コミュニケーションロボットの開発が注目されている [竹内 07, 伊藤 09]．これらのロボットは多くの場合，人間対ロボットの一対一コミュニケーション，つまり人間とロボットとの対話を想定しており，人間同士の会話にロボットが介入を行うものは少ない．

一方，人間同士のコミュニケーションでは，一対一対話だけではなく，多人数での会話を行う機会が多い．そこで会話に参加し，相槌や話題提供を行うことで会話をより楽しく，豊かなものに支援するロボットの開発が期待されている．本研究では，会話に介入を行うことで円滑な会話コミュニケーションを創発する会話コミュニケーション支援ロボットの構築を目指す．

図 1 に，会話コミュニケーション概要を示す．円滑なコミュニケーションは，話者 A と話者 B の対話を話者 O が聴取し，2 名の対話の雰囲気を把握して適切な介入を行うことで創発されると考える．話者 A と話者 B の対話雰囲気が「盛り上がっている」場面と「まじめな」場面ではそれぞ話者 O が取るべき介入は異なる．そのため，会話コミュニケーションを円滑にするには,「会話雰囲気の推定」と「雰囲気に応じた介入の選択」の二つが重要であると考える．本稿では，会話コミュニケーション支援ロボットの要素研究として，二者間対話における対話雰囲気の推定手法を提案する．

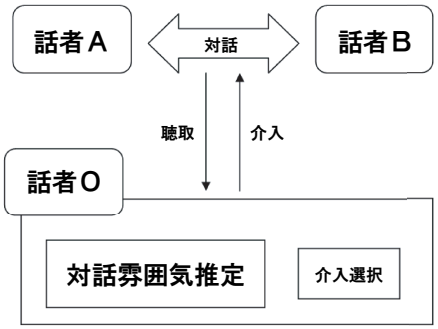

図 1: 会話コミュニケーション概要

2. 関連研究

人間とロボットのコミュニケーションの円滑化を目指した研究は多く，これまでにロボットが発話タイミングを変化させた場合に人間が受ける印象を調査 [Hayashi 09, 高杉 10] した研究や，対話音声の韻律情報と感情との関係を調査した研究 [若松 01] がされている．しかし，会話コミュニケーション支援ロボットにはそれらの能力とは別に人間同士の対話雰囲気を推定し，適切な介入行動を選択する能力が必要だと考える．

本稿では，話者 2 名で行われる対話にロボットが介入する場面を想定し，話者 2 名の対話雰囲気を推定するシステムを提案する．これまでに対話雰囲気を推定する研究として，テキスト情報 [稲葉 11] や音声韻律情報 [渋谷 07] など様々な特徴量が検討されている．本稿は，発話時間長に着目した特徴を対話雰囲気推定に用いる．これは対話の雰囲気が，対話の内容やイントネーション等だけでなく，対話における「間」が強く関係していると考えたためである．また本稿で用いる発話時間特徴は，話者認識の技術 [松井 96] を利用することで実際の対話から容易に抽出できる特徴であり，システムを現実に用いる場合に求められるリアルタイムな特徴抽出が可能と考える．

連絡先: 加藤昇平，名古屋工業大学，愛知県名古屋市昭和区御器所町，052-735-5625，shohey@juno.ics.nitech.ac.jp

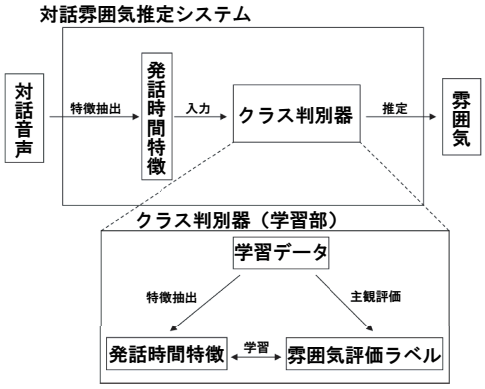

図 2: 対話雰囲気推定システム

3. 対話雰囲気推定システム

図 2 に本稿で提案する対話雰囲気推定システムの概要を示す．提案手法では対話雰囲気推定の特徴量として，対話全体の「無音時間」や，両話者の発話が重なった場合に生じる「同時発話時間」といった「発話時間」に着目した「発話時間特徴」を用いる．本システムでは，まず対話音声を入力として与えることで，発話時間特徴を抽出する．そして事前の学習で構築されたクラス判別器に抽出した特徴量を入力し．推論することで，対話雰囲気の推定を行う．本稿ではクラス判別器に学習データを用いて特徴量間の因果関係を構築し推論を行う Tree-Augmented Naive Bayes（TAN）を採用した．学習データから発話時間特徴を抽出し，同時に各学習データについて主観評価実験を行うことにより，対話雰囲気のラベル付けを行った．そして対話雰囲気および特徴量間の事前事後確率に基づいて TAN の学習を行う．

4. 発話時間特徴

本稿では，対話雰囲気を推定するために「発話時間」に着目する．「発話時間」として話者交替時に生じる「無音時間」や，話者が同時に発話を行う「同時発話時間」，一人の話者が単独で発話し続けている「単独発話時間」を対話雰囲気推定の特徴量に用いる．

4.1 発話状態集合の生成

本研究では対話者の発話状態 st を次のように定義する．

1. A 単独発話状態
2. B 単独発話状態
3. 同時発話状態
4. 無音状態

このとき，合計発話時間が長い話者を A 話者，短い話者を B 話者とする．次に合計 n 個の対話データをそれぞれ上に示した 4 つの発話状態に分け，各状態の開始時刻と終了時刻の差からそれぞれの発話状態の時間長を算出する．そして，算出した時間長を要素とする状態毎の多重集合 S_{st} を生成する．図 3 に対話インデックス d ($D = \{1, 2, \ldots, n\}$　$d \in D$) の対話における発話状態概要を示す．また例として，この場合の A 単独発話の時間長算出手法を示す．発話状態 $st = 1$ 時の発話状態集合は多重集合 $S_1^d = \{3, 1, 1\}$ と示すことができる．

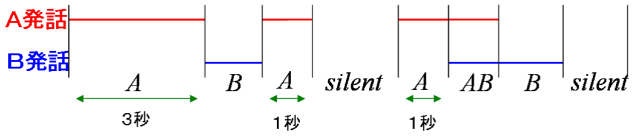

図 3: 発話状態概要

4.2 発話時間特徴算出

表 1 に発話状態集合から抽出する合計 152 個の発話時間特徴の例を示す．特徴番号 1 から 32 の計 32 個が発話時間統計特徴であり，特徴番号 33 から 152 までの計 120 個が発話時間割合特徴である．発話時間統計特徴については 4.2.1 節で説明を行い，発話時間割合特徴については 4.2.2 節で説明を行う．表 1 において，関数 stat(S_{st}^d) とは 4.2.1 節に示す関数のいずれかを適用することをあらわす．

4.2.1 発話時間統計特徴

第 4.1 章で作成した発話状態集合 S_{st}^d を基に 8 個の統計量を以下に示す 8 関数によりそれぞれ算出する．

- mean(S_{st}^d)　:発話状態集合 S_{st}^d の平均
- var(S_{st}^d)　:発話状態集合 S_{st}^d の分散
- min(S_{st}^d)　:発話状態集合 S_{st}^d の最小値
- 25p_tile(S_{st}^d)　:発話状態集合 S_{st}^d の 25 パーセンタイル点
- median(S_{st}^d)　:発話状態集合 S_{st}^d の中央値
- 75p_tile(S_{st}^d)　:発話状態集合 S_{st}^d の 75 パーセンタイル点
- max(S_{st}^d)　:発話状態集合 S_{st}^d の最大値
- count(S_{st}^d)　:発話状態集合 S_{st}^d の要素数

平均，及び分散は全体の傾向を，最小値，25 パーセンタイル点，中央値，75 パーセンタイル点，及び最大値は，多重集合に含まれる要素の値を，要素数は対話 d における発話状態 st の状態観測数をそれぞれ表現するために採用した．以上 8 個の統計量を 4 つの発話状態集合に対してそれぞれ算出し，合計 32 個の発話時間統計特徴とする．

4.2.2 発話時間割合特徴

対話全体に対する無音時間の割合や，A 単独発話と同時発話を足した A 合計発話と A 単独発話の比較といった「割合」は対話雰囲気推定を行う場合に，有用性を持つと考えられる．そこで，それぞれの発話時間統計特徴に対して，加算や除算といった算術演算を行い対話雰囲気と因果関係を持つと考えられる割合を発話時間割合特徴として合計 120 個用意した．

5. 対話雰囲気推定学習部

本稿では，学習データとして音声対話コーパス [NII] を利用した．音声対話コーパスの中にはクロスワードタスク，間違い探しタスク，スケジュール調整タスク，及びテレフォンショッピングタスクといった多岐にわたる対話データが収録されている．学習データには対話の始まりと終わりが無音状態であり，時間長が 20 秒から 30 秒の対話 150 個を任意に選択した．用意した学習データに対して聴取実験を行い，以下に示す 6 個の対話雰囲気を表す形容詞対について肯定，否定の 2 値のラベル付けを行った．

- 盛り上がり　：盛り上がっている - 盛り上がっていない

表 1: 発話時間特徴

特徴番号	特徴量	算出式
1-8	A 発話に関する統計量	$\mathrm{stat}(S_1^d)$
9-16	B 発話に関する統計量	$\mathrm{stat}(S_2^d)$
17-24	同時発話に関する統計量	$\mathrm{stat}(S_3^d)$
25-32	無音時間に関する統計量	$\mathrm{stat}(S_4^d)$
33-64	A 発話と B 発話の比較	$\frac{\mathrm{stat}(S_2^d)}{\mathrm{stat}(S_1^d)}$
65-72	発音状態と無音状態の比較	$\frac{\mathrm{stat}(S_4^d)}{\sum_{i=1}^{3}\mathrm{stat}(S_i^d)}$
73-80	対話全体における A 発話状態の割合	$\frac{\mathrm{stat}(S_1^d)}{\sum_{i=1}^{4}\mathrm{stat}(S_i^d)}$
81-88	対話全体における B 発話状態の割合	$\frac{\mathrm{stat}(S_2^d)}{\sum_{i=1}^{4}\mathrm{stat}(S_i^d)}$
89-96	対話全体における同時発話状態の割合	$\frac{\mathrm{stat}(S_3^d)}{\sum_{i=1}^{4}\mathrm{stat}(S_i^d)}$
97-104	対話全体における無音状態の割合	$\frac{\mathrm{stat}(S_4^d)}{\sum_{i=1}^{4}\mathrm{stat}(S_i^d)}$
105-120	同時発話と単独発話の比較	$\frac{\mathrm{stat}(S_3^d)}{\mathrm{stat}(S_1^d)}$
121-136	合計発話と同時発話の比較	$\frac{\mathrm{stat}(S_3^d)}{\mathrm{stat}(S_1^d)+\mathrm{stat}(S_3^d)}$
137-152	合計発話と単独発話の比較	$\frac{\mathrm{stat}(S_1^d)}{\mathrm{stat}(S_1^d)+\mathrm{stat}(S_3^d)}$

- まじめさ　：まじめな - まじめでない
- 噛み合い　：噛み合っている - 噛み合っていない
- 明るさ　：明るい - 明るくない
- 親密さ　：親密な - 親密でない
- 対等さ　：立場が対等な - 立場が対等でない

6. 対話雰囲気推定実験

本手法では，用意した特徴量に対して変数選択を行うことで対話雰囲気と強い因果関係を持つ特徴量を抽出し，対話雰囲気推定システムを構築した．また構築した対話雰囲気推定システムを用いて推定実験を行い，提案した手法の有効性を検証した．検証方法としては 5 分割交差検定法を用いた．

6.1 変数選択

多変量解析では一般的に，用いる特徴量が多くなると特徴空間が大きくなり，学習効率が低下する問題が知られている．また推定する対話雰囲気により，それぞれ因果関係の強い特徴量は異なると考えられる．そこで用意した 152 個の特徴に対して，遺伝的アルゴリズム（GA）による変数選択を行う．

GA におけるエージェントは遺伝子としてビット列を持つ．本手法では一つのビットは一つの特徴量に対応しており，1 は対話雰囲気推定の特徴量として採用，0 は不採用を表す．またエージェントの適応度として対話雰囲気の判別正答率を用いることで，対話雰囲気の推定に適した特徴量の探索を行う．GA の遺伝的操作には，エリート選択，一様交叉，突然変異を採用する．表 3 に使用した GA のパラメータを示す．変数選択により，学習データから主観評価によってラベル付けされた各対話雰囲気と関係の深い特徴量を選び，各形容詞対に対してそれぞれ TAN を用いて学習する．

表 2: 対話雰囲気推定正答率

雰囲気	肯定正答率	否定正答率	全体正答率（%）
盛り上がり	71.1	**90**	**84.1**
まじめさ	70.0	77.3	73.8
噛み合い	**89.2**	45.0	**83.3**
明るさ	**87.1**	55.0	73.8
親密さ	62.2	**92.0**	**82.8**
対等さ	72.9	70	71.7

表 3: GA のパラメータ設定

個体数	150
ステップ数	1000
選択数	5
突然変異率	0.1%

6.2 結果と考察

構築した対話雰囲気推定システムの推定正答率を表 2 に示す．全ての推定システムにおいて，全体正答率は 70%を超えており，各雰囲気について提案する対話雰囲気推定システムによって，正しく推定可能であることが示唆された．また盛り上がり推定，まじめさ推定，親密さ推定，対等さ推定においては，肯定正答率，否定正答率ともに 70%を超えており，高い正答率で対話雰囲気推定が可能であることを確認した．このことから本手法で用意した「発話時間特徴」はこれら 4 つの対話雰囲気推定に有用な特徴量であることが示唆された．

表 4 に，今回構築した各推定システムにおいて選択された特徴量を示す．本論では，最も高い全体正答率を示した盛り上がり推定に着目し，詳細に考察する．盛り上がり推定では，A 発話と B 発話における 25 パーセンタイル点の比較を行う特徴量（44），全体における無音時間の割合を示す特徴量（104），合計発話と同時発話における中央値の比較を行う特徴量（133）が選択された．盛り上がり推定において選択された特徴量のクラス別正規分布図を図 4 - 6 に示す．図 4 より，A 発話の時間と B 発話の時間の比較を行った場合に B 発話の時間が長かった場合，盛り上がっていると推定される傾向があることが確認された．これは，盛り上がっている雰囲気の場合，合計発話時間の短い B 話者が長く発話していることを示しており，話者が 2 名とも同程度に発話していることが示唆される．また図 5 より，全体における無音時間が比較的長い場合には，盛り上がっていないと推定されることが示された．最後に図 6 より，合計発話と同時発話の比較を行った場合に同時発話時間が長かった場合，盛り上がっていると推定される傾向があることが確認された．同時発話は話し手が話者交替を行う場合に相手の発話終了前に自分の発話を始めた場合に起こることが多い．そのため，同時発話の多い対話は活発に意見が交換されており，盛り上がっている対話であることが示唆される．

盛り上がり推定において否定正答率に着目すると，90%の正

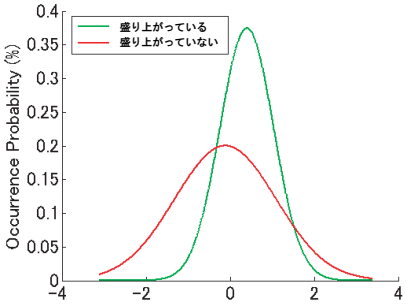

図 4: A 発話と B 発話の比較を行う特徴量

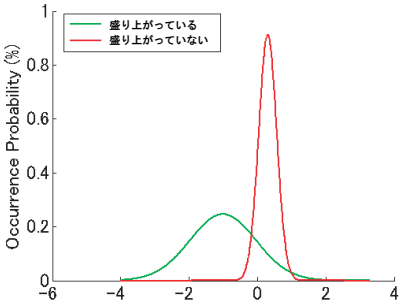

図 5: 全体における無音状態の割合に関する特徴量

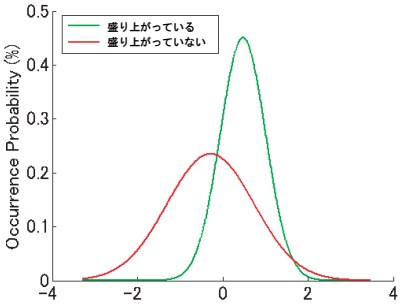

図 6: 合計発話と同時発話の比較を行う特徴量

表 4: 選択された特徴量

雰囲気	選択された特徴番号
盛り上がり	**44,104,133**
まじめさ	17,37,51,99,153
噛み合い	29,35,40,50,66,71,94,107
明るさ	3,7,17,98
親密さ	97,123
対等さ	36,53,82,98,137

答率で推定可能であることが確認できる．このことから，本手法を用いることで的確に対話の盛り上がっていない雰囲気を把握し，話題を提供するといった介入を行うロボットの実現可能性が示唆された．

7. おわりに

本研究では対話の発話時間に着目した発話時間特徴を用意し，TAN を用いた雰囲気推定手法を提案した．その結果，全ての推定システムで 70%以上と高い全体正答率で対話雰囲気を推定可能であることを確認した．またその中でも盛り上がり推定，まじめさ推定，親密さ推定，対等さ推定においては 70%を超える肯定正答率，否定正答率を共に示し，これらの対話雰囲気推定には発話時間特徴が有効であることが示唆された．

本稿では，20 秒から 30 秒と日常行われる対話を考慮すると比較的短い時間での対話のみを扱った．そこで今後は，今回提案した手法を対話全体に対して時間的に遷移しながら行うことで，本手法で扱う時間より長い時間の対話雰囲気をより正確に推定する手法を検討する．

また会話支援を行う研究として，会話の観察によりネットワーク構造のデータベースを構築することで話題提供を行う研究 [倉林 02] がされている．会話コミュニケーション支援システム実現のため，本稿で提案した対話雰囲気推定と話題提供の技術を共に用いることで，適切なタイミングで，適切な話題提供を行うことが可能となる．このような対話雰囲気に応じた介入選択手法の検討を行い，会話コミュニケーション支援システムの完成を目指す．

謝辞

本研究は，一部，文部科学省科学研究費補助金（課題番号 20700199）の助成のもと行われた。

参考文献

[Hayashi 09] Takanori Hayashi, Shohei Kato, and Hidenori Itoh: A Synchronous Model of Mental Rhythm Using Paralanguage for Communication Robots. Proceedings of the 12th International Conference on Principles of Practice in Multi-Agent Systems, pp. 376-388, (2009)

[高杉 10] 高杉 将司，吉田 祥平，沖津 健吾，横山 正典，山本 知仁：コミュニケーションロボットの対話における交替潜時長と頷き先行時間長の影響評価, 計測自動制御学会論文集，Vol.46，No.1， pp.72-81（2010）

[若松 01] 若松 良久，近藤 敏之，伊藤 宏司：コミュニケーションロボットにおける音声に基づいた感情形成モデルの構築, 第 40 回計測自動制御学会学術講演会,（2001）

[稲葉 11] 稲葉 通将，鳥海 不二夫，石井 健一郎：語の共起情報を用いた対話における盛り上がりの自動判定，電子情報通信学会論文誌 D Vol.J94-D No.1 pp.59-67（2011）

[渋谷 07] 渋谷 貴紀，益永 祐吾，川端 豪：語の共起情報を用いた対話における盛り上がりの自動判定，情報処理学会研究報告．SLP，音声言語情報処理（2007）

[松井 96] 松井 知子：HMM による話者認識，電子情報通信学会技術研究報告．SP，音声（1996）

[倉林 02] 倉林 則之，山崎 達也，湯浅 太一，蓮池 和夫：ネットワークコミュニティにおける関心の類似性に基づいた知識共有の促進，情報処理学会論文誌　43，pp.3559-3570,（2002）

[竹内 07] 竹内 将吾，酒井 あゆみ，加藤 昇平，伊藤 英則：対話者好感度に基づく感性会話ロボットの感情生成モデル，日本ロボット学会誌　Vol. 25，No. 7，pp. 1125-1133,（2007）

[伊藤 09] 伊藤 千加，加藤 昇平，伊藤 英則：感性会話ロボットの性格付けとその心理評価，日本感性工学会論文誌，Vol. 8，No. 3，pp. 899-906,（2009）

[NII] 京都大学工学研究科，文部省科研費重点領域研究 ,「音声・言語・概念の統合的処理による対話の理解と生成に関する研究」NII 音声資源コンソーシアム， 対話音声コーパス (PASD),（1993-1996）