

修士論文

雰囲気推定を用いた
作業通話マッチング支援サービスの提案

公立はこだて未来大学大学院 システム情報科学研究科
高度 ICT 領域

立花 虎太郎

指導教員 奥野 拓

提出日 2023 年 1 月 12 日

Master's Thesis

Proposal of a Work Call Matching Support System
Using Atmosphere Estimation

by

Tachibana Kotraro

Graduate School of Systems Information Science, Future University Hakodate
Advanced ICT Field

Supervisor: Okuno Taku

Submitted on January 12, 2023

Abstract—

In recent years, “work calls” , in which multiple people work on individual tasks while sharing time via a call tool, have become widely used. The purpose of this study is to support casual matching between users who have the same working preference but have weak relationship with each other in a work call inviting participants on SNS. This study proposes a visualization method of how to proceed with a work call focusing on the dialogue atmosphere and a participant recruitment system using the visualization method. This study uses machine learning to estimate the dialogue atmosphere. The model is constructed using speech state temporal features that can provide highly accurate estimation of dialogue atmosphere while ensuring the privacy of participants. This study constructs and verifies a model for estimating the atmosphere of a working conversation held with 2 to 4 speakers.

Keywords: Work Call, Dialogue Atmosphere Estimation, Communication Support

概要：

近年、複数人が通話ツールを介して時間を共有しながら、各々の個人作業に取り組む活動である「作業通話」が広く行われている。本研究では、SNS を利用し参加者を募る作業通話の募集者に対して、作業の進め方の好みは一致しているが関係は希薄なユーザとの気軽なマッチングの支援を行う。そのために対話雰囲気推定を用いた作業通話の進め方の可視化手法と、それを用いた参加者募集システムの提案を行う。本研究では機械学習を用いた対話雰囲気推定を行う。モデルの構築には、参加者のプライバシーを担保しながら高精度の対話雰囲気推定を行える発話状態時間特徴を用いる。話者数 2 ～ 4 名で開催される作業通話の対話雰囲気を推定するモデルの構築及び検証を行なった。

キーワード： 作業通話, 対話雰囲気推定, コミュニケーション支援

目次

第 1 章	序論	1
1.1	背景	1
1.2	目的	2
1.3	章構成	3
第 2 章	アプローチ	4
2.1	自身と近い作業嗜好を持つ人の募集を探すことが困難	4
2.2	参加の意思表示の心理的負担が大きい	5
第 3 章	関連研究	7
3.1	対話雰囲気推定に関する研究	7
3.2	SNS 上で行われるコミュニケーションに関する研究	8
第 4 章	提案システム	9
4.1	概要	9
4.2	対話雰囲気推定モジュール	9
4.3	マッチング支援モジュール	12
第 5 章	対話雰囲気推定モデルの構築	14
5.1	対話データの収集とラベル付け	14
5.2	発話状態集合の抽出	15
5.3	特徴量の抽出	16
5.4	学習	20
5.5	特徴量選択	21
第 6 章	学習結果と考察	22
6.1	評価手法	22
6.2	特徴量選択を用いない対話雰囲気推定モデルの学習結果	22
6.3	ベースモデルの決定	25
6.4	特徴量選択を用いた対話雰囲気推定モデルの学習結果	26

6.5	学習結果の考察	30
第 7 章	結論	32
参考文献		35

第 1 章

序論

1.1 背景

近年,「作業通話」と呼ばれる活動が広く行われている。作業通話とは,複数人が Discord[1]等の通話ツールを介して時間を共有しながら,各々の個人作業に取り組む活動を指す。作業通話にはいくつかの別称があり,コーディングを行うエンジニアの間では「オンラインもくもく会」,イラスト制作や小説の執筆を行うクリエイターや,勉強に取り組む学生の間では「さぎょいぶ」としても浸透している。

作業通話の開催には大きく二つの利点がある。一つ目の利点として作業効率の向上が挙げられる。作業通話は他者と通話をしながら作業を行うことから,常に他者の存在を感じながら作業を進めることができるという特性を持つ。それにより,個人の経験や性格によって個人差があるものの,適度に他者の存在を感じる環境にすることで観衆効果の発生が期待できる。一般に観衆効果は作業効率が向上することが報告されている [2][3][4]。また,参加者同士で会話を行うことでリフレッシュ効果が期待できるほか,作業内容が類似している場合は作業中に生まれた疑問を即座に他の参加者に聞くことができるなど作業効率の向上が見込める点が多くある。二つ目の利点として他者との関係構築の機会の増加が挙げられる。作業通話の参加者は常に他者と通話回線がつながっており,声を発するだけで会話が始められる状態である。また,参加者には作業を進めるという目的があることから数十分から数時間に及ぶ作業時間になることが多い。これらの特性から自然に会話が発生し,お互いの理解が深まる。また,並行して作業も行なっていることから無理に会話を続ける必要もなく,コミュニケーションにおける心理的な負担も小さい。加えて,SNS 上で知り合った関係が希薄もしくは皆無な人(以下,「無縁ユーザ」)と作業通話を開催する文化もあり,作業通話はコミュニケーション媒体の一種として機能している。

一方で,作業通話の参加者には各々進め方の好み(以下,「作業嗜好」)があり,それが一致する参加者を募ることが難しいという問題がある。作業通話には開催する時間帯,作業の類似性,利用するコミュニケーションツールなどの要素があり,参加者には会話量,性格などの要素がある。作業嗜好が合わない参加者と長時間の作業通話は,前述した作業通話の利

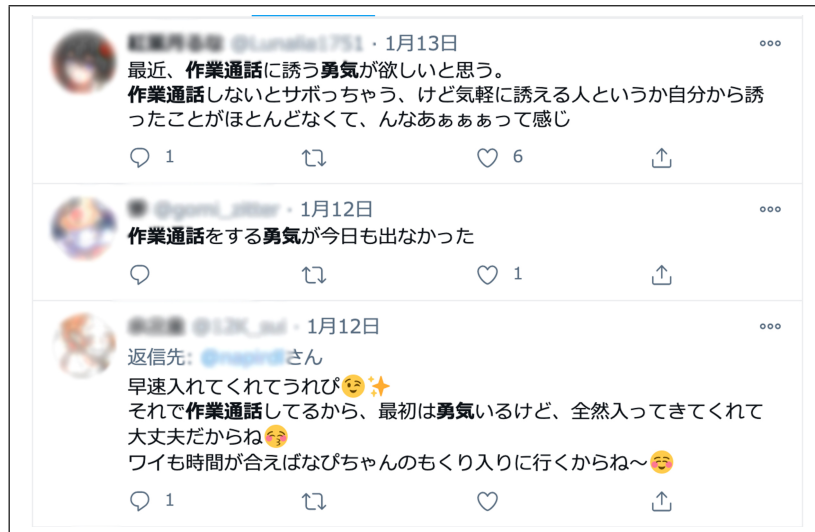


図 1.1 Twitter で「作業通話 勇気」で検索した結果

点を十分に活かすことができない場合がある．このことから，より効果的な作業通話の実現には作業嗜好の合う参加者を募ることが重要であるといえる．作業通話の参加者を募る方法は大きく二つある．一つは友人など既に所属しているコミュニティで参加者を募る方法（以下，「コミュニティ募集」）である．コミュニティ募集は，身近な人を気軽に招待，依頼できる反面，募集対象者数が少ないため各々の都合や作業嗜好の違いによって十分に参加者が集まらない場合がある．もう一つの募集方法として，Twitter[5] 等の SNS を通じて不特定多数の人を対象として参加者を募る方法（以下，「SNS 募集」）がある．SNS 募集は募集対象者数が多いため，その中に作業嗜好の合う人が存在する可能性が高い一方で，実際に利用されることは少ない．これには大きく二つの要因がある．一つ目は募集対象者が数ある募集の中からどれが自身の作業嗜好と合うものか判断することが困難であるという点である．二つ目に募集者が無縁ユーザに対して参加者を募ることや，募集対象者が募集に対して参加意思を表明することの心理的負担が大きいという点がある．事実，Twitter 上には「作業通話したいけど勇気がでない」といった投稿が多く見られる（図 1.1）．

本研究では，これら二つの要因を解消することで，作業嗜好の合う無縁ユーザとの気軽なマッチングを実現し，より効果的な作業通話の開催機会の増加を目指す．

1.2 目的

本研究の目的は，作業嗜好の合う無縁ユーザの気軽なマッチングを実現し，より効果的な作業通話の開催機会の増加を目指すことである．そのために，作業通話の音声やメタデータから作業嗜好を推定し，可視化及び無縁ユーザのマッチングを行うシステムを提案する．特に，本研究では対話雰囲気に着目し機械学習を用いた対話雰囲気の推定を行うことで作業嗜好の効果的な可視化を目指す．

1.3 章構成

本論文は全7章で構成されている。第1章では、本研究の背景と研究目的について述べた。第2章では、本研究のアプローチについて述べる。第3章では、本研究に関連する研究について述べる。第4章では、本研究で提案するシステムについて述べる。第5章では、本研究で構築した対話雰囲気推定モデルの構築について述べる。第6章では、本研究で構築した対話雰囲気推定モデルの評価と考察について述べる。第7章では、まとめと今後の展望について述べる。

第2章

アプローチ

本章では第1章で述べた問題に対する本研究のアプローチについて述べる。第1章では以下の問題について述べた。

- i. 自身と近い作業嗜好を持つ人の募集を探すことが困難
- ii. 参加の意思表示の心理的負担が大きい

2.1 自身と近い作業嗜好を持つ人の募集を探すことが困難

本問題の原因の一つに作業通話の募集文に作業嗜好を判断する情報が少ないことが挙げられる。特に開催時間や募集人数、開催場所などの情報が欠落していることが多い（図2.1）。コミュニティ募集の場合、募集対象者は募集者の人柄や作業の進め方などは事前に把握していることが多いためこの問題は無視されることが多い。一方、SNS募集の場合、募集者の人柄から自身の作業嗜好との相性を推察することができないため募集対象者の参加意欲を減衰させている可能性が高い。

この問題の解決には作業嗜好の可視化が効果的であるといえる。作業通話の募集概要（募集対象者や開催時間など）を募集文に明示することで、募集対象者は参加するかの判断に必要な情報の多くを把握できる。しかし、作業嗜好の要素として雰囲気のお好みは挙げられるが、これについては募集概要から推察することは困難である。作業通話における雰囲気（対話雰囲気）は参加者の気構えや心情が深く関わっており、作業の進め方や参加者の満足感に強く影響を与える要素である。したがって、本研究では対話雰囲気が作業嗜好の大きな要素になると仮定し、対話雰囲気を推定し可視化するシステムの構築を行う。また、本研究では対話雰囲気を「盛り上がり」、「真面目さ」など複数人からなる対話における場の空気感と定義する。

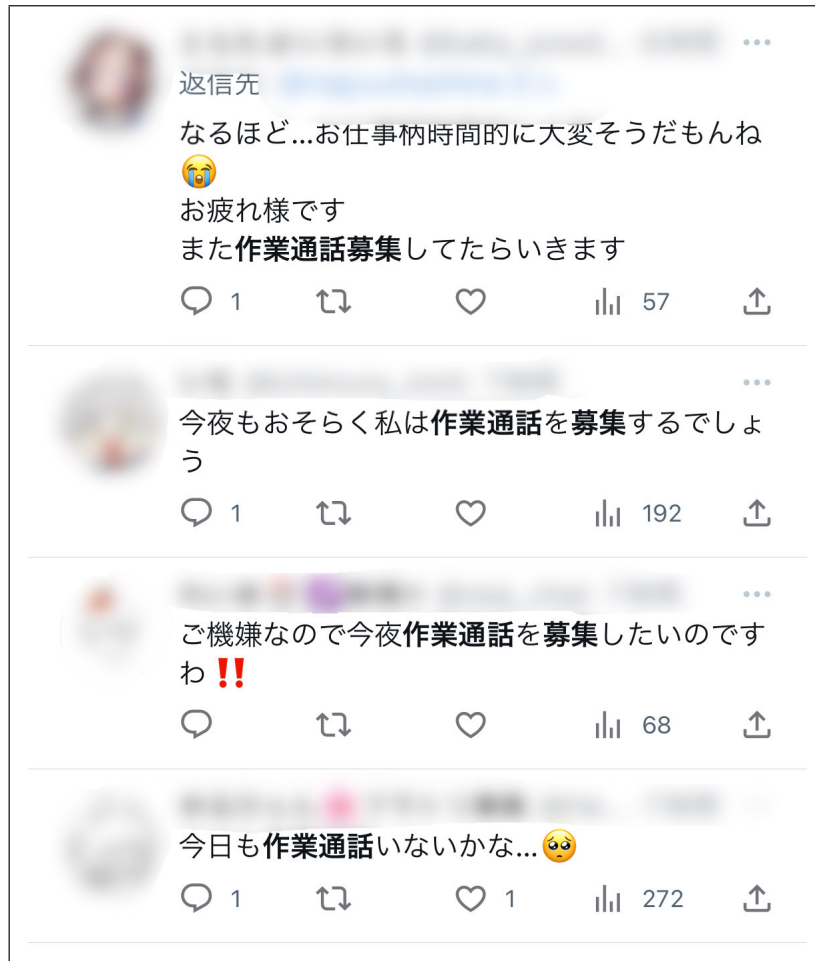


図 2.1 Twitter で「作業通話 募集」で検索した結果

2.2 参加の意思表示の心理的負担が大きい

本問題の原因として、無縁ユーザとの作業通話という文化が十分に浸透していないことが挙げられる。加えて、募集は募集以外の投稿と混在しながらフランクに投稿されている場合も多い。また、第 2.1 節の問題も加わり募集文を見ただけではその募集がコミュニティ募集と SNS 募集どちらなのかを判断することが難しいという問題もある。これらのことから、募集対象者は興味のある募集を発見しても参加の意思表示をする心理的負担が大きいといえる。

そこで本研究では、第 2.1 節で述べたアプローチに加え、参加の意思表示の心理的負担の軽減を目的に SNS 募集の枠組みを提案する。具体的には、SNS 募集における募集文（ツイート）を生成できるシステムの提案を行う。またこのシステムは、SNS 募集対象者が自身に向けて募集が行われていることや、そのような募集方法が存在することを明確に認知すること

で参加の意思表示の心理的負担の軽減を目指している。このシステムは、システムを通したコミュニケーションは受け入れやすいということが報告されていることから一定の効果が期待できる [6][7][8][9]。このシステムを用いて作成された募集文には、第 2.1 節で述べた募集概要や対話雰囲気も併せて掲載することで、作業の透明性の向上を図りさらなる心理的負担の軽減を目指す。

第3章

関連研究

本研究では、遠隔で複数人によって行われる対話の雰囲気の推定と、その結果を用いた SNS 上で行われるコミュニケーションの支援を目指している。本章では、その上で本研究に関する研究について述べる。はじめに、対話雰囲気推定を行っている研究について述べる。次に、SNS 上で行われるコミュニケーションを支援している研究について述べる。

3.1 対話雰囲気推定に関する研究

近年、人間と適切な会話を行う対話エージェントを開発するために、対話の雰囲気を推定する研究が数多く行われている。

人間同士の対話を対象とし、対話雰囲気推定を行った研究として徳久ら [10] の研究がある。徳久らは対話内容に対して Dialogue Act (DA) と Rhetorical Relation (RR) を用いた発話分類を行い、対話の盛り上がりとの関係を分析している。DA とは、各対話内容が「質問」や「要求」、「提案」などのどのようなカテゴリに属するかをラベル付けする手法である。DA が発話にラベル付けする手法に対し、RR は修辞構造を表す手法である。結果として、言及される内容の主観性や、「嬉しい」や「つらい」などの感情に関する発話が盛り上がりとの関係が深いことが明らかになっている。しかし、言語情報を実際の対話から抽出し、会話内容に基づいたラベル付けをし利用することは難しいことも示唆された。

仕草や表情等の非言語情報を用いた対話雰囲気推定を行った研究として伊藤ら [11] の研究がある。伊藤らは、二者間の対話における盛り上がりに関与する要素を話者の動作や発話から分析している。結果として、対話の盛り上がりを判断する要素として、発話の重なりやうなずき、笑い等の要素の重要度が高いことが明らかになっている。また、頭部の動作特徴が場の雰囲気表現に有効であることが示唆された。しかし、表情認識や動作認識は高いコストを必要とするため、リアルタイムの人間の会話に介入することは難しいという問題がある。また、遠隔で行われる対話においても、これらの非言語情報が失われるため推定が困難である。

容易に抽出可能な特徴を用いて、対話の雰囲気を推定する研究として豊田ら [12] の研究

がある。豊田らは、発話時間特徴と主観評価を用いた機械学習によって対話雰囲気の推定を試みている。対話の雰囲気を推定する学習済みモデル（以下、「対話雰囲気推定モデル」）は Tree-Augmented Naive Bayes 分類器 (TAN) と遺伝的アルゴリズムを用いた特徴量選択が採用されている。モデル構築の過程では、二者の発話全体における同時発話時間の占有率が比較的高い場合、対話が盛り上がっていると推定される傾向があることが明らかになっている。また、対話をリードしている話者の発話の分散に対して、追従している話者の発話の分散が比較的高い場合にも、対話が盛り上がっていると推定されることが示されている。構築した対話雰囲気推定モデルを 5 分割検証法によって分析した結果、「盛り上がり」、「真面目さ」、「親密さ」の推定において 80% を超える全体正答率が示されている。豊田らの手法は、遠隔で行う対話の雰囲気推定も容易に行える反面、二者対話のみを想定しているため複数人での対話における対話雰囲気推定モデルの精度が確認されていない。一方で、十分な精度で対話雰囲気の推定を行えていることから、本研究では豊田らのモデル構築手法を参考に対話雰囲気モデルの構築を行う。

3.2 SNS 上で行われるコミュニケーションに関する研究

SNS 上における繋がり構築を支援する研究として北村ら [13] や田村ら [14]、久米ら [15] の研究がある。北村らは、関与ベースに基づく Twitter ユーザの推薦を行い手法の有効性を示している。田村らは、Twitter ユーザにおける他者とのコミュニケーション量に基づいた会話しやすいユーザの推薦手法を提案している。久米らは、コンテンツベースの推薦手法にカテゴリ情報を付加し、特徴語抽出手法である TF-IDF 法に各カテゴリへの興味度合いを示す「興味領域」という指標を加えたユーザ推薦手法を提案している。これらの手法は、ユーザと相性の良い他のユーザを見つける上では有用だが、実際に繋がり構築する上でユーザは推薦されたユーザに対して接触を図る必要があり、その際の心理的負担は考慮されていない。

SNS 上での募集行為を支援する研究として樋川 [16] の研究がある。樋川は、LINE [17] 上で実験協力者の募集を、Bot を通して行うことができるシステムの構築・評価を行っている。Bot を用いることで適切な距離感の被験者候補に対して募集を行うことができたほか、希薄な関係の人同士におけるシステムを介したコミュニケーションの有効性が示唆された。

第4章

提案システム

本章では、本研究で提案するシステムについて述べる。

4.1 概要

本研究で提案するシステムの目的は、作業通話における作業嗜好の透明性の向上と、作業通話への募集・参加における心理的負担の低下である。

本システムの構成を図 4.1 に示す。本システムは、対話雰囲気推定モジュールとそれを用いたマッチング支援モジュールから構築される。図 4.1 において、データベースを境にそれまでのフローが対話雰囲気推定モジュール、それ以降のフローがマッチング支援モジュールである。本システムは、はじめに対話雰囲気推定モジュールで作業通話の音声から対話雰囲気の推定及び、メタデータやアカウント情報の収集を行いデータベースに記録する。続いて、それらのデータを基にマッチング支援モジュールでは二つの機能を提供する。一つ目の機能は、ユーザの作業記録や作業傾向の可視化を行う機能である。本機能では、それまでに集めたユーザ自身の作業嗜好を表す活動雰囲気のデータの統計値をはじめ、作業時間や発話量などの傾向を算出し表示する。二つ目の機能は、ユーザの作業記録や作業傾向を基にした募集文の生成を行う機能である。本機能では、参加者を募りたい作業通話の概要を入力することで、その概要と一つ目の機能で可視化された自身のデータを掲載された募集文（ツイート）を生成し投稿する。

4.2 対話雰囲気推定モジュール

本モジュールは Discord で開催される作業通話を対象としており、作業通話の音声を Bot によって収集し対話雰囲気の推定を行う。

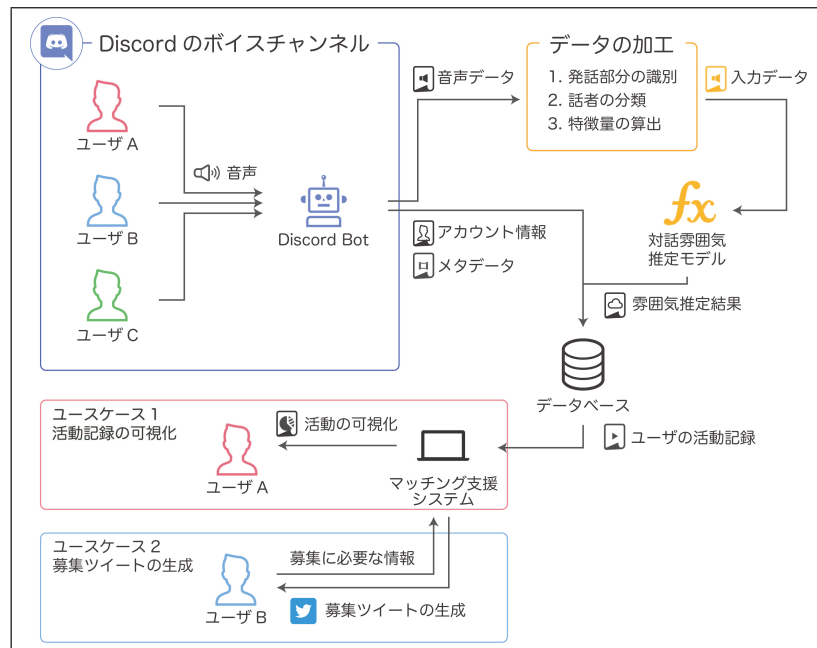


図 4.1 システムイメージ

4.2.1 対話雰囲気推定モジュールで収集するデータ

本モジュールは、対話雰囲気推定のために音声データの収集を行う。その際には、第3章で述べた豊田らの手法を参考に、ユーザのプライバシーの担保や第5章で述べる対話雰囲気推定モデルの構築における学習コストの低減を目的に限られた非言語情報のみを対象とする。具体的には、どのユーザごとの発話時間情報（ユーザがいつどのくらいの時間発話を行ったかの情報）と作業通話のメタデータのみを用いる。作業通話のメタデータとしては、参加者のユーザ情報、開催時間情報が収集対象である。また、収集した発話時間情報は特徴量として変換され対話雰囲気推定モデルの入力データとして対話雰囲気の推定に利用する。

4.2.2 対話雰囲気推定モジュールで推定する雰囲気

本モジュールで推定を行う対話雰囲気は「盛り上がり」、「真面目さ」、「明るさ」、「くつろぎ」の四つである（表 4.1）。本研究のそれぞれの定義について述べる。

「盛り上がり」は対話の活発性を表現する雰囲気を指す。「盛り上がり」が肯定の状態、つまり盛り上がっている雰囲気とは参加者同士のコミュニケーションが頻繁に行われており、相互行為が活発に行われている状態を指す。反対に、「盛り上がり」が否定の状態、盛り下がっている、沈んでいる雰囲気とは参加者同士の会話や相互行為が少ない状態を指す。

「真面目さ」は対話の生産性や参加者が対話に熱心に取り組んでいるかを表現する雰囲気を指す。「真面目さ」が肯定の状態、つまり真面目な雰囲気とは参加者が作業や議題に集中して

表 4.1 推定を行う雰囲気

名称	定義	肯定	否定
盛り上がり	対話の活発性を表現	盛り上がっている	沈んでいる
真面目さ	対話の生産性や参加者が対話に熱心に取り組んでいるかを表現	真面目である	怠けている
明るさ	話題の内容や参加者の態度を表現	明るい	暗い
くつろぎ	参加への心理的安全性や対話への満足感を表現	くつろいでいる	緊張している

おり、生産性が高い状態を指す。反対に、「真面目さ」が否定の状態、真面目でない、怠けている雰囲気とは参加者が雑談をしていたり、主目的とは別の言動を行なっている状態を指す。

「明るさ」は話題の内容や参加者の態度を表現する雰囲気を指す。「明るさ」が肯定の状態、つまり明るい雰囲気とは話題がポジティブで、参加者が興奮している状態を指す。反対に、「明るさ」が否定の状態、暗い、落ち着いた雰囲気とは話題がネガティブで、参加者が物事に対して否定的な状態を指す。

「くつろぎ」は参加への心理的安全性や対話への満足感を表現する雰囲気を指す。「くつろぎ」が肯定の状態、つまりくつろいでいる雰囲気とは参加者同士の心理的安全性が高く、リラックスして参加できている状態を指す。反対に、「くつろぎ」が否定の状態、気遣っている、緊張している雰囲気とは参加者がお互いの距離感や接し方がわからない状態や、物事に注視して集中している状態を指す。

これらの対話雰囲気は本研究独自で定義したものであり、参考になっている豊田らの定義した雰囲気とは異なる。これは、豊田らは対面で行われる二者間の対話を対象としていることに対して、本研究では遠隔で行われる複数名での作業通話中の対話を対象としているためである。

これらの対話雰囲気の推定には機械学習を用いる。具体的には対話雰囲気推定モデルを構築し、そのモデルによって入力された一つの対話における各雰囲気について肯定か否定の2値で推定を行う。対話雰囲気推定モデルについては第5章で詳細に述べる。

4.2.3 対話雰囲気推定モジュールの利用方法

対話雰囲気推定モジュールのユーザが操作を行う箇所は Discord Bot のみである。本節では Discord Bot の利用方法の想定を示す。まず、Discord Bot はユーザが作業通話を開催するサーバに招待されることで利用可能状態となる。ユーザは作業通話を開始した際に開始用のコマンドを Discord Bot に対して入力することで録音を開始する。また、ユーザは作業通話が終了した際に終了用のコマンドを入力。終了用のコマンドを受信した Discord Bot は録音の停止と対話雰囲気の推定結果の出力をテキストチャンネル上に行う。ユーザは Discord

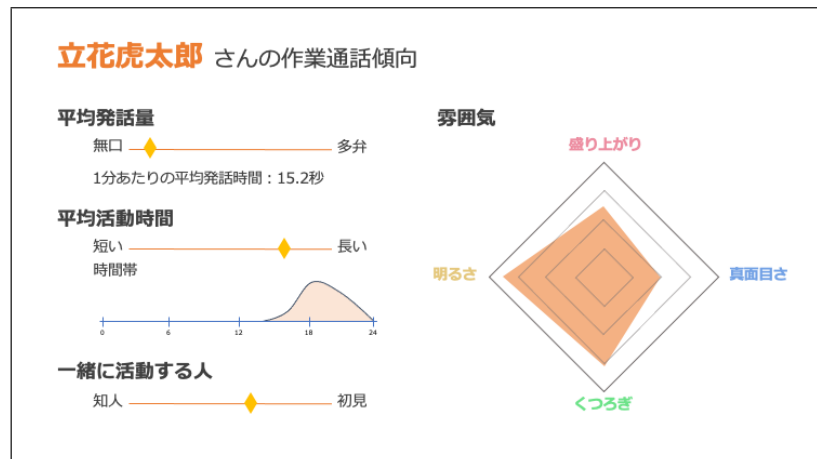


図 4.2 作業嗜好の可視化イメージ

Bot により出力された対話雰囲気の推定結果の確認や、主観的に正しいと感じた雰囲気への修正を行える。それらのデータはデータベースに記録される。記録したデータは再び学習データとしてモデルの精度向上に利用されることを想定している。

4.3 マッチング支援モジュール

本モジュールは、第 4.2 節のモジュールによってデータベースに記録されたデータを用いた二つのユースケースを想定した機能の実装をしている。

一つ目は、自身の活動記録や作業嗜好を客観的に確認したいというユースケースに向けた活動記録や作業嗜好の可視化機能である。本機能は、ユーザの今まで開催した作業通話に関するデータを表示する（図 4.2）。表示する情報は、平均発話量、平均活動時間、一緒に活動する人、雰囲気それぞれについての傾向値である。平均発話量の傾向値は、発話回数と発話時間を基に対象ユーザが無口であるか多弁であるかをスライダーで示す。また、1 分あたりの平均発話時間も併せて記載する。平均活動時間の傾向値は、メタデータを基に平均活動時間長が短いか長いかをスライダーで示す。また、作業通話を開催している時間帯の傾向についても併せて記載する。一緒に活動する人の傾向値は、メタデータを基に作業通話を知人と初見どちらの人と開催することが多いかをスライダーで示す。ここにおける知人とは Discord 上でフレンドになっているか、一度以上過去にともに作業通話を行なった記録がある人を指す。雰囲気の傾向値とは盛り上がっている雰囲気の対話が多いことなどをレーダーチャートにて表現した値を示す。

二つ目は、効果的な SNS 募集を行いたいというユースケースに向けた募集文（ツイート）の生成機能である。本機能は、参加者を募りたい作業通話の概要を入力することで、その概要と一つ目の機能で可視化された自身のデータを掲載された募集文（ツイート）を生成し投稿する（図 4.3）。本機能の対象は、Twitter で行われる SNS 募集である。本機能でユーザが

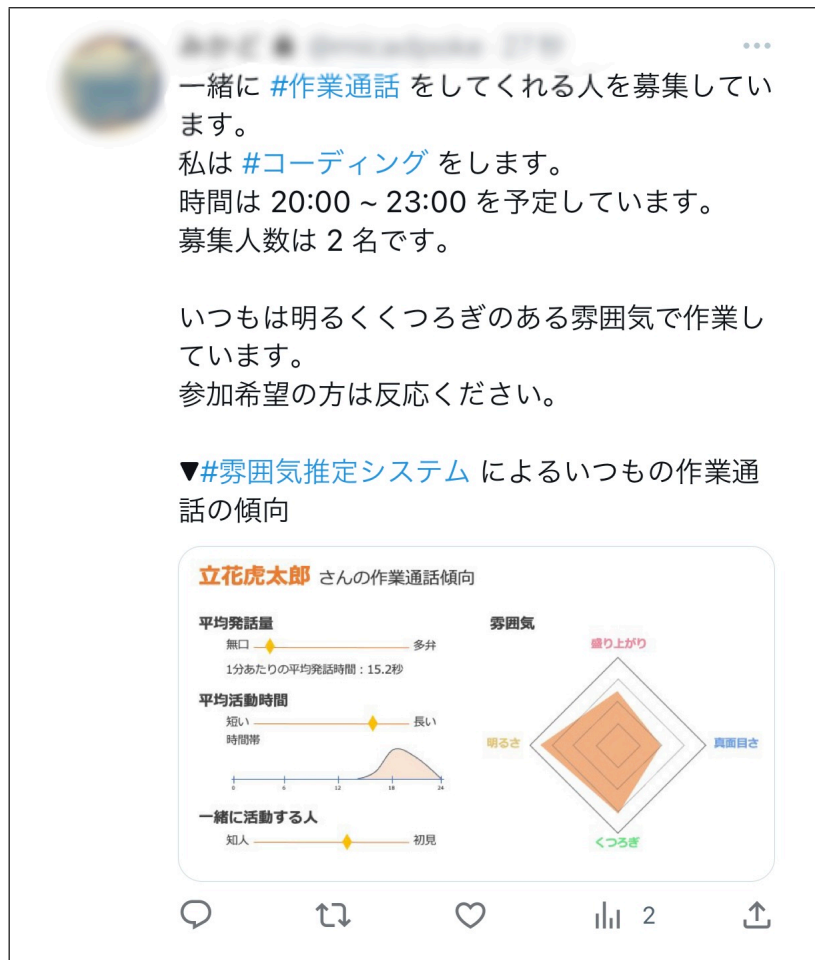


図 4.3 生成する投稿のイメージ

入力する作業通話の概要情報は、作業内容、開催時間、募集人数の三つを想定している。本機能は、一目で無縁ユーザ向けの募集文であることがわかることと、開催したい作業通話の概要がわかることによって明確で効率的な募集形態の実現を目指している。

将来的には、本モジュールを用いて作業嗜好の合うユーザの推薦等無縁ユーザとの繋がり形成の機会を増加させる機能の実装なども目指している。結果として、SNS 募集のデファクトスタンダードとなることで SNS 募集の効率化と心理的負担の低下の実現を行えるモジュールを目指す。

第 5 章

対話雰囲気推定モデルの構築

本研究では、機械学習によって通話中の音声から対話雰囲気を推定する。本研究で構築する対話雰囲気推定モデルは、対話データから抽出した特徴量を入力値とし、第 4.2.2 項で述べた各雰囲気について肯定、否定の 2 値での分類結果を出力値とする。豊田らは二者対話を対象に発話時間特徴を特徴量とする機械学習を用いた対話雰囲気推定を行なっている [12]。豊田らの構築したモデルは「盛り上がり」「まじめさ」「囁み合い」「明るさ」「親密さ」「対等さ」の 6 つの雰囲気を推定対象としており、それぞれ肯定、否定の 2 値で分類している。特に「盛り上がり」「まじめさ」「親密さ」の推定においては全体正答率が 0.8 を超える高い値を示していることから、本研究でもこの手法に基づいて対話雰囲気推定モデルの構築を行う。本章では、本研究における対話雰囲気推定モデルの構築手順について述べる。

5.1 対話データの収集とラベル付け

学習データを構築するために対話データの収集を行う。豊田ら是对話データとして音声コーパス [18] を採用している。一方で、本研究では実際に開催された作業通話の録音データを採用する。これは一般的な対話と作業通話中の対話はそれぞれ性質が異なるためである。例えば、一般的な対話では沈黙は気まずく回避される傾向があるが、作業通話における沈黙は作業に集中していることの現れであり許容される傾向がある。加えて、作業通話中の対話においては突発的な話題の変化が起こりやすく、独り言が行き交うことが多いなどの特徴がある。このような作業通話中の対話の独特な性質が、対話雰囲気の形成に影響を与えていると十分に考えられることから、本研究では実際の作業通話の音声を学習に用いる。

対話データの収集とラベル付けの手順について述べる。まず、対話データとして 1 ～ 4 名の話者からなる計 6 時間弱の作業通話音声の録音を行う。次に全ての音声を 10 ～ 120 秒程度の対話になるように切り出す。続いて、それぞれのデータに対して第 4.2.2 項で述べた各雰囲気（「盛り上がり」、「真面目さ」、「明るさ」、「くつろぎ」）について主観で肯定か否定のラベル付けを行う。ここで付与するラベルはのちの学習での教師ラベルとなる。以上の手順により本研究では、計 130 対話を抽出しラベル付けされた対話データとした。

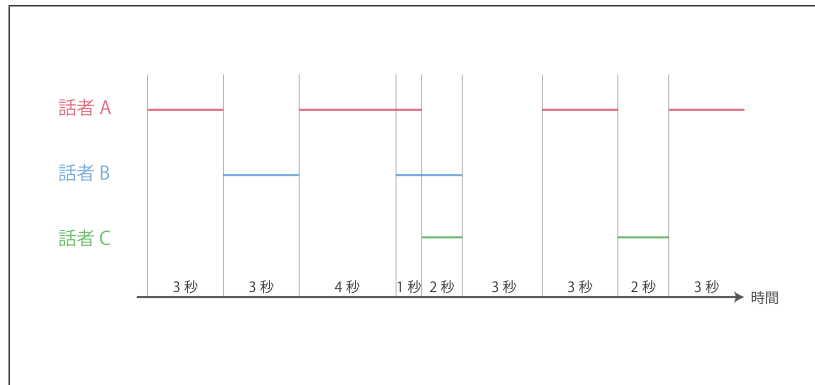


図 5.1 対話データにおける発話の例

5.2 発話状態集合の抽出

特徴量の抽出を行うために、対話データから発話状態集合の抽出を行う。発話状態集合とは、対話データにおける話者ごとの発話時間を発話状態ごとに記録した集合のことを指す。豊田らの手法では、一つの対話データについて各話者の単独発話集合と二者の同時発話集合を抽出している。この手法は二者の対話でのみ有効な抽出方法である。しかし、本研究の手法における対象話者数は2～4名であるため、異なる手法を用いて発話状態集合を抽出する。

はじめに、豊田らの手法は1名の話者のみが発話している状態を単独発話状態としてその集合を抽出しているが、本研究では他の話者の発話に依存せず話者ごとに発話している状態を発話集合として抽出する。図 5.1 を例に述べる。図 5.1 は対話データの一部を例として掲載しており、横軸が時間の流れを表している。また、縦軸の各話者と同じ高さで引かれている線はそれぞれの話者が発話していることを表している。図 5.1 において豊田らの手法の場合、話者 A における単独発話集合は $\{3, 4, 3, 3, \dots\}$ となる。一方本研究の場合、話者 A における発話集合は $\{3, 5, 3, 3, \dots\}$ となる。

次に、豊田らの手法は二名の話者が同時に発話している状態を同時発話状態としてその集合を対話データにつき一つ抽出しているが、本研究では話者ごとに同時発話集合を抽出する。図 5.1 において豊田らの手法の場合、同時発話集合は $\{1, 2, \dots\}$ となる。一方本研究の場合、話者ごとに抽出するため話者 A の同時発話集合は $\{1, \dots\}$ 、話者 B の同時発話集合は $\{1, 2, \dots\}$ 、話者 C の同時発話集合は $\{2, \dots\}$ となる。

以上が豊田らの手法で抽出した発話状態集合との比較である。加えて本研究では、一つの対話データ全体としての傾向が対話雰囲気の形成に影響を与えていると仮定し、新たに三つの発話状態を定義しそれらの集合を抽出した。一つ目は非発話状態である。非発話状態は対話において一人も発話していない状態である。非発話状態の集合である非発話集合は図 5.1 の場合、 $\{3, \dots\}$ となる。二つ目は全発話状態である。全発話状態は全話者の発

表 5.1 特徴量として採用する統計量と算出式

統計量	算出式
合計値	$sum(S^d) = \sum_{i=1}^n S^d_i$
平均値	$mean(S^d) = \frac{sum(S^d)}{n}$
標準偏差	$std(S^d) = \sqrt{\frac{1}{n} \sum_{i=1}^n (S^d_i - mean(S^d))^2}$
最大値	$max(S^d) = \{r \in S^d r \geq \forall a \in S^d\}$
要素数	$size(S^d) = S^d $
占有率	$share(S^d) = \frac{sum(S^d)}{sum(S^d_{as})}$

話状態を集約した発話状態である．全発話状態の集合である全発話集合は図 5.1 の場合， $\{3, 5, 3, 3, 3, 3, 2, 2, \dots\}$ となる．三つ目は全状態である．全状態は全発話集合へ非発話状態を加えた状態である．全状態の集合である全集合は図 5.1 の場合， $\{3, 5, 3, 3, 3, 3, 2, 2, 3, \dots\}$ となる．

以降，対話 d において話者数は n ，任意の発話状態集合は S^d と表現し，その集合における i 番目の要素は S^d_i と表現する．加えて，話者 x についての発話集合は S^d_{stx} ，同時発話集合は S^d_{ssx} と表現する．また，同様に対話 d に対する非発話集合は S^d_{ns} ，全発話集合は S^d_{us} ，全集合は S^d_{as} と表現する．

5.3 特徴量の抽出

抽出した発話状態集合から特徴量の抽出を行う．豊田らは各集合に対する統計量とその統計量を比較した値（以下，「比較量」）を算出し特徴量としている．本研究でも統計量と比較量を特徴量とするが，具体的に算出する値は異なる．

5.3.1 統計量を用いた特徴量の抽出

本研究で特徴量として採用する統計量とそれらの算出式を表 5.1 に示す．本研究では，統計量としてそれぞれの集合に対する合計値，平均値，標準偏差，最大値，要素数，占有率を算出する．それぞれの統計量の採用理由について述べる．はじめに，合計値と平均値，要素数，占有率を計測することで話者の会話への積極性や興味の強さが対話雰囲気の形成に与える影響を測定するために採用した．次に，標準偏差は話者間の発話の偏りが対話雰囲気の形成に与える影響を測定するために採用した．最大値は対話における発話時間の値域が対話雰囲気の形成に与える影響を測定するために採用した．

5.3.2 比較量を用いた特徴量の抽出

統計量をもとに比較量を算出し特徴量とする．本研究における二つの発話状態集合の比較量は、発話状態集合の各統計量について発話合計時間が短い方を長い方で除算することで算出する．よって比較を行う発話状態集合両者に欠損値が存在しない場合は、一回の比較で6つの比較量が得られる．本研究は比較量として話者間発話比較、話者間同時発話比較、他話者比較、話者内発話比較を算出する．これらの比較量は各発話状態集合間で算出する．

はじめに、話者間発話比較は、対話 d における任意の話者二名の発話状態を比較した値である．話者間発話比較を求めることで、話者間の発話時間や発話頻度の比較を行うことができる．これによって、対話における話者間のコミュニケーション量の違いが対話雰囲気の形成に与える影響を測定することができることから話者間発話比較を採用した．

次に、話者間同時発話比較は、対話 d における任意の話者二名の同時発話状態を比較した値である．話者間の発話量の比較について、二者の対話における同時発話状態の発生回数や占有率などの統計量が「盛り上がり」や「真面目さ」の対話雰囲気に影響を与えることが示唆されている [11][12]．本研究では、同時発話集合を話者ごとに抽出しているため同様に同時発話状態の比較が対話雰囲気の推定に有効であると考え話者間同時発話比較を採用した．

続いて、他話者比較は、対話 d における任意の話者一名と、その話者以外の話者の発話集合を結合させた集合を比較した値である．他話者比較を求めることで、特定の話者だけでなく対話全体の中での発話時間や発話頻度を求め対象話者が聞き手であるか、話し手であるかを推定できる．これにより、聞き手（話し手）の発話量が対話雰囲気に与える影響を測定できることから他話者比較を採用した．

話者内発話比較は、対話 d における任意の話者一名の発話状態と同時発話状態を比較した値である．話者内発話比較を求めることで、対話において一人の話者が一方的に発言を続けているかを推定することができる．これにより、話者が他話者の発言をどの程度受け入れているかを測定できることから話者内発話比較を採用した．

5.3.3 抽出する特徴量の一覧

対話 d について、本研究で抽出する特徴量を表 5.2 と表 5.3 に示す．それぞれの表中の x, y は任意の話者を表現しており、 x はより発話時間の合計値が高い話者を指している．本研究で特徴量として算出される統計量、比較量の数は話者数に依存している．抽出される特徴量数は、欠損値を含めると話者数 2 名で 90 個、話者数 3 名で 144 個、話者数 4 名で 210 個である．

表 5.2 統計量を用いた特徴量

特徴量番号	特徴量	算出式
1_x	話者 x の発話集合に関する特徴量	$sum(S_{stx}^d)$
2_x		$mean(S_{stx}^d)$
3_x		$std(S_{stx}^d)$
4_x		$max(S_{stx}^d)$
5_x		$size(S_{stx}^d)$
6_x		$share(S_{stx}^d)$
7_x	話者 x の同時発話集合に関する特徴量	$sum(S_{ssx}^d)$
8_x		$mean(S_{ssx}^d)$
9_x		$std(S_{ssx}^d)$
10_x		$max(S_{ssx}^d)$
11_x		$size(S_{ssx}^d)$
12_x		$share(S_{ssx}^d)$
13	非発話集合に関する特徴量	$sum(S_{ns}^d)$
14		$mean(S_{ns}^d)$
15		$std(S_{ns}^d)$
16		$max(S_{ns}^d)$
17		$size(S_{ns}^d)$
18		$share(S_{ns}^d)$
13	全発話集合に関する特徴量	$sum(S_{us}^d)$
14		$mean(S_{us}^d)$
15		$std(S_{us}^d)$
16		$max(S_{us}^d)$
17		$size(S_{us}^d)$
18		$share(S_{us}^d)$
13	全集合に関する特徴量	$sum(S_{as}^d)$
14		$mean(S_{as}^d)$
15		$std(S_{as}^d)$
16		$max(S_{as}^d)$
17		$size(S_{as}^d)$
18		$share(S_{as}^d)$

表 5.3 比較量を用いた特徴量

特徴量番号	特徴量	算出式
19 _{xy}	話者 x, y の話者間発話比較に関する特徴量	$\frac{sum(S_{sty}^d)}{sum(S_{stx}^d)}$
20 _{xy}		$\frac{mean(S_{sty}^d)}{mean(S_{stx}^d)}$
21 _{xy}		$\frac{std(S_{sty}^d)}{std(S_{stx}^d)}$
22 _{xy}		$\frac{max(S_{sty}^d)}{max(S_{stx}^d)}$
23 _{xy}		$\frac{size(S_{sty}^d)}{size(S_{stx}^d)}$
24 _{xy}		$\frac{share(S_{sty}^d)}{share(S_{stx}^d)}$
25 _{xy}	話者 x, y の話者間同時発話比較に関する特徴量	$\frac{sum(S_{ssy}^d)}{sum(S_{ssx}^d)}$
26 _{xy}		$\frac{mean(S_{ssy}^d)}{mean(S_{ssx}^d)}$
27 _{xy}		$\frac{std(S_{ssy}^d)}{std(S_{ssx}^d)}$
28 _{xy}		$\frac{max(S_{ssy}^d)}{max(S_{ssx}^d)}$
29 _{xy}		$\frac{size(S_{ssy}^d)}{size(S_{ssx}^d)}$
30 _{xy}		$\frac{share(S_{ssy}^d)}{share(S_{ssx}^d)}$
31 _x	話者 x の他話者発話比較に関する特徴量	$\frac{sum(S_{stx}^d)}{sum(S_{stx}^d \cap S_{us}^d)}$
32 _x		$\frac{mean(S_{stx}^d)}{mean(S_{stx}^d \cap S_{us}^d)}$
33 _x		$\frac{std(S_{stx}^d)}{std(S_{stx}^d \cap S_{us}^d)}$
34 _x		$\frac{max(S_{stx}^d)}{max(S_{stx}^d \cap S_{us}^d)}$
35 _x		$\frac{size(S_{stx}^d)}{size(S_{stx}^d \cap S_{us}^d)}$
36 _x		$\frac{share(S_{stx}^d)}{share(S_{stx}^d \cap S_{us}^d)}$
37 _x	話者 x の話者内発話比較に関する特徴量	$\frac{sum(S_{stx}^d)}{sum(S_{ssx}^d)}$
38 _x		$\frac{mean(S_{stx}^d)}{mean(S_{ssx}^d)}$
39 _x		$\frac{std(S_{stx}^d)}{std(S_{ssx}^d)}$
40 _x		$\frac{max(S_{stx}^d)}{max(S_{ssx}^d)}$
41 _x		$\frac{size(S_{stx}^d)}{size(S_{ssx}^d)}$
42 _x		$\frac{share(S_{stx}^d)}{share(S_{ssx}^d)}$

表 5.4 学習の設定

設定項目	範囲・候補
対象雰囲気	盛り上がり
	真面目さ
	明るさ
	くつろぎ
対象話者数	2 ～ 4
学習アルゴリズム	Linear SVC
	k 近傍法
	SVC
	Naive Bayes

5.4 学習

本研究で構築する対話雰囲気推定モデルは、対象とする雰囲気や話者数が豊田らと異なるため一部独自の手法を採用する。本研究では、対象雰囲気、対象話者数ごとに計 12 種の対話雰囲気推定モデルの構築を行う。これは対象となる雰囲気や話者数によって有効な特徴量が異なるという仮説に基づいているためである。

本研究では、学習の前に学習データの前処理を行う。具体的には、学習する対話雰囲気推定モデルの話者数によって欠損値を含む特徴量と学習データの削除を行う。話者数 3 名を対象とする対話雰囲気推定モデルの学習を例とする。まず、特徴量の削減を行う。削減の対象となる特徴量は、話者 3 名を対象としたモデルでは不要となる話者数 4 名の場合に抽出される特徴量である。具体的には、表 5.2 における特徴量番号 1 の 4 者目の特徴量や、表 5.3 における特徴量番号 19 の 4 者目を比較対象とする特徴量などである。続いて、学習データの削除を行う。削除の対象となる学習データは、話者 3 名を対象としたモデルでは不要となる話者数 2 名による学習データである。これは話者数 2 名の学習データは、特徴量番号 1 の 3 者目の特徴量や特徴量番号 19 の 3 者目を比較対象とする特徴量など話者数に依存した特徴量が算出できず欠損値が発生するためである。

前処理を行ったものを最終的な学習データとして学習を行う。学習を行う際に定める必要のある設定を表 5.4 に示す。対象雰囲気は第 4.2.2 項で述べた「盛り上がり」、「真面目さ」、「明るさ」、「くつろぎ」の 4 つから選択する。対象話者数は 2 ～ 4 名から選択する。学習モデルは、どの学習モデルが本研究の学習において有効であるか比較を行うため Linear SVC, k 近傍法, SVC, Naïve Bayes の 4 つを採用する。

表 5.5 遺伝的アルゴリズムの設定

設定項目	範囲・候補
交叉方法	二点交叉 一様交叉
集団の大きさ	100 / 300 / 1000
染色体突然変異率	0.0 ～ 0.5
遺伝子突然変異率	0.0 ～ 0.5
個体評価方法	モデル正答率による単純評価 選択特徴量数とモデル正答率の重み付け和による単純評価 選択特徴量数とモデル正答率の重み付け和による交差検証
選択方法	エリート選択
エリート染色体選択数	10 / 30 / 1000
世代数	100 / 250 / 1000

5.5 特徴量選択

第 5.4 節で述べた設定を行うことで対話雰囲気推定モデルを構築することができる。しかし、一般に多変量解析において解析に用いる特徴量が多くなると特徴空間が広がることで学習効率が低下する問題が知られている。本研究で構築する対話雰囲気推定モデルでは最大 210 個の特徴量を用いるため、数ある特徴量の中から対話雰囲気の推定に有効な特徴量のみを用いて学習することで精度や学習速度の向上が見込める。そこで本研究は特徴量選択を行う。特徴量選択の手法は、豊田らにならい遺伝的アルゴリズム（以下、「GA」）を採用する。GA では特徴量数と同じ長さのビット列からなる染色体を生成し、各特徴量の利用有無を 1：有効，0：無効で表現する。GA の設定を表 5.5 に示す。表中の個体評価における選択特徴量数とモデル正答率の重み付け和による評価は以下の式によって算出する。式中の W は正答率と選択特徴量数の重み付けを表現しており， $[0, 1]$ の値をとる。

$$\text{評価値} = \frac{\text{正答数}}{\text{検証データ数}} W + \left(1 - \frac{\text{選択特徴量数}}{\text{全特徴量数}}\right) (1 - W) \quad (5.1)$$

第 6 章

学習結果と考察

本章では、第 5 章で述べた手法で構築した対話雰囲気推定モデルの学習結果と考察について述べる。はじめに、本研究における対話雰囲気推定モデルの評価手法について述べる。次に、特徴量選択を用いない対話雰囲気推定モデルの学習結果について述べる。続いて、特徴量選択を用いた学習の比較基となるベースモデルについて述べる。最後に、ベースモデルを中心に特徴量選択を用いた学習の結果について述べる。

6.1 評価手法

本研究における対話雰囲気推定モデルの評価はモデル正答率と選択特徴量数を用いる。これはモデル正答率が高いほど正確な推定が期待でき、選択特徴量数が少ないほど軽量の推定が期待できるためである。また、モデル評価は第 5.5 節で述べた個体評価とは異なり、二つの評価観点を一つの評価値と算出することはしない。これは豊田らの手法においても同様の手法を用いていることと、どちらの観点が重要視されるべきか十分に検討できていないためである。評価手法の検討については考察で詳細を述べる。

6.2 特徴量選択を用いない対話雰囲気推定モデルの学習結果

対話雰囲気推定モデルの学習における特徴量選択の有効性の評価を行うために、特徴量選択を用いない対話雰囲気推定モデルの構築を行う。特徴量選択を用いない対話雰囲気推定モデルの学習結果を表 6.1 ～ 表 6.4 に示す。特徴量選択を用いない対話雰囲気推定モデルは各対象雰囲気、対象話者数、学習モデルを網羅的に組み合わせて構築した。表中のモデル正答率は有効数字 5 桁で、小数点第 5 位を四捨五入したものを掲載している。いずれにおいても特徴量選択は行っていないため、学習に利用した特徴量は話者数に依存した最大値となっている。

学習結果から明らかになったことについて述べる。学習アルゴリズムはいずれの対象雰囲気のモデルにおいても対象話者数が 2 名または 3 名の場合、Naive Bayes のモデル正答率が

表 6.1 「盛り上がり」を対象とした特徴量選択を用いない学習の結果

モデル番号	対象話者数	学習アルゴリズム	モデル正答率	選択特徴量数
1	2	Linear SVC	0.6240	90
2		k 近傍法	0.6400	
3		SVC	0.7040	
4		Naive Bayes	0.7440	
5	3	Linear SVC	0.4699	144
6		k 近傍法	0.5772	
7		SVC	0.6515	
8		Naive Bayes	0.6632	
9	4	Linear SVC	0.3300	210
10		k 近傍法	0.5900	
11		SVC	0.6200	
12		Naive Bayes	0.4800	

表 6.2 「真面目さ」を対象とした特徴量選択を用いない学習の結果

モデル番号	対象話者数	学習アルゴリズム	モデル正答率	選択特徴量数
13	2	Linear SVC	0.6362	90
14		k 近傍法	0.4676	
15		SVC	0.5486	
16		Naive Bayes	0.6610	
17	3	Linear SVC	0.4222	144
18		k 近傍法	0.4417	
19		SVC	0.5583	
20		Naive Bayes	0.6028	
21	4	Linear SVC	0.5000	210
22		k 近傍法	0.7000	
23		SVC	0.5000	
24		Naive Bayes	0.6000	

高いことが明らかになった。一方で、対象話者数が4名のモデルにおけるモデル正答率はモデル番号33の0.3000など極端に低い値であるものも確認できる。また、豊田らの構築した対話雰囲気推定モデルはモデル正答率が0.7000～0.8550であったことから、これらのモデルのモデル正答率が低いことがわかる。

表 6.3 「明るさ」を対象とした特徴量選択を用いない学習の結果

モデル番号	対象話者数	学習アルゴリズム	モデル正答率	選択特徴量数
25	2	Linear SVC	0.5638	90
26		k 近傍法	0.5057	
27		SVC	0.5629	
28		Naive Bayes	0.7181	
29	3	Linear SVC	0.6722	144
30		k 近傍法	0.5306	
31		SVC	0.5389	
32		Naive Bayes	0.7000	
33	4	Linear SVC	0.3000	210
34		k 近傍法	0.7000	
35		SVC	0.6000	
36		Naive Bayes	0.4000	

表 6.4 「くつろぎ」を対象とした特徴量選択を用いない学習の結果

モデル番号	対象話者数	学習アルゴリズム	モデル正答率	選択特徴量数
37	2	Linear SVC	0.6200	90
38		k 近傍法	0.5924	
39		SVC	0.5210	
40		Naive Bayes	0.6648	
41	3	Linear SVC	0.5778	144
42		k 近傍法	0.5806	
43		SVC	0.5583	
44		Naive Bayes	0.7000	
45	4	Linear SVC	0.6000	210
46		k 近傍法	0.5000	
47		SVC	0.6000	
48		Naive Bayes	0.7000	

表 6.5 ベースモデルの設定値

設定項目	範囲・候補
対象雰囲気	盛り上がり
対象話者数	3
学習アルゴリズム	Naive Bayes
交叉方法	一様交叉
集団の大きさ	100
染色体突然変異率	0.05
遺伝子突然変異率	0.10
個体評価方法	選択特徴量数とモデル正答率の重み付け和による交差検証
選択方法	エリート選択
エリート染色体選択数	10
世代数	250

6.3 ベースモデルの決定

本研究では、第 5.4 節と第 5.5 節で示した各設定値を定めた上で学習を行う。そこで、特徴選択を行うにあたって各設定値の学習効率への影響を容易に比較するために、比較基となるベースモデルを定める（表 6.5）。ベースモデルの対象雰囲気は「盛り上がり」とする。これは豊田らの手法と本研究における推定対象の雰囲気を比較した際に、同一の定義を行っていることに加え、豊田らの研究におけるモデル正答率が 0.8550 と高水準であるためである。続いて対象話者数は 3 名とする。これには二つの理由がある。一つ目は、豊田らの手法と異なり三者以上の話者を対象とした対話雰囲気推定モデルを構築していることが本研究の独自性の一つであることから、三者以上の話者を対象としたモデルを比較の中心にすることが妥当であると考えたためである。二つ目は、対象話者数が 4 名の学習データは第 5.4 節で述べた前処理を行った結果、学習に用いることができるデータ数が 3 名に比較し必ず少なくなることから、対象話者数が増加することでモデルの信憑性が低下するためである。学習アルゴリズムは Naive Bayes を採用する。これは特徴量選択を行わない学習では、Naive Bayes が最も良いモデル正答率を持っているためである。それ以降の特徴量選択に関わる設定は、事前に何度か無作為に設定を定めて学習をした上で最もモデル正答率が高く、選択特徴量数が少ないものを採用する。その際の、評価における重み付け W は 0.5 である。以上のベースモデルを用いて各学習設定や特徴量選択の有効性を評価する。

表 6.6 特徴量選択を用いたモデルの学習結果

モデル番号	対象雰囲気	対象話者数	モデル正答率	選択特徴量数
49	盛り上がり	2	0.8000	10
50		3	0.7956	21
51		4	0.8600	45
52	真面目さ	2	0.7752	16
53		3	0.7694	30
54		4	0.8000	38
55	明るさ	2	0.8305	12
56		3	0.8639	33
57		4	0.7000	40
58	くつろぎ	2	0.7905	13
59		3	0.8167	25
60		4	0.9000	40

6.4 特徴量選択を用いた対話雰囲気推定モデルの学習結果

第 6.3 項で述べたベースモデルを基に特徴量選択を用いた対話雰囲気推定モデルを構築する。はじめに、ベースモデルから対象雰囲気、話者数を変更したモデルについて述べ、続いて特徴量選択の各設定を変更したモデルについて述べる。

6.4.1 対象雰囲気、話者数を変更したモデルの学習結果

ベースモデルとベースモデルから対象雰囲気、話者数を変更した全てのモデルの学習結果の一例を表 6.6 に示す。表中のモデル番号 50 がベースモデルを表している。

この結果から明らかになったことについて述べる。

はじめに、ベースモデルに着目する。特徴量選択を行う前のベースモデルと同条件のモデルは表 6.1 のモデル番号 8 である。これらのモデルを比較するとモデル正答率が 0.6632 から 0.7956 に向上し、選択特徴量数が 144 から 21 に減少していることが確認できる。ベースモデルの学習過程を図 6.1 に示す。図 6.1 は横軸が GA の世代数を表しており、オレンジ色の値がモデル正答率、緑色の値が全体に対する選択特徴量数の割合、青色が評価値を表している。図 6.1 から学習が進むにつれてモデル正答率と選択特徴量数が変化していることがわかる。

次に、全体のモデル正答率に着目する。ベースモデル同様、全体としてモデル正答率の向上を確認できる。事実、表 6.1 ～ 表 6.4 のうち Naive Bayes を学習モデルとして採用してい

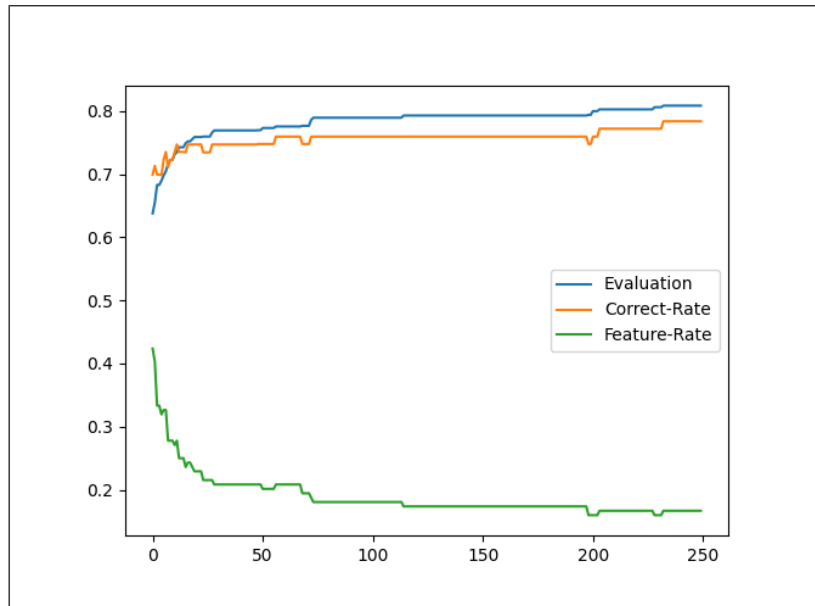


図 6.1 ベースモデルの学習過程

表 6.7 特徴量選択前後のモデル正答率の t 検定

算出項目	算出値
平均値	特徴量選択前: 0.6362
	特徴量選択後: 0.8085
分散値	特徴量選択前: 0.0105
	特徴量選択後: 0.0027
t 値	-6.6579
t 境界値	2.2010
P 値	3.5727×10^{-5}

るモデルと、表 6.6 のモデルのモデル正答率について両側 5 % の t 検定を行った結果、P 値が 3.5727×10^{-5} を示すことから有意差を確認できる（表 6.7）。豊田らの構築した対話雰囲気推定モデルの正答率と比較しても大きな差は見られず同程度の制度を持つモデルであることがわかる。また、対象雰囲気間によるモデル正答率の差は小さいことが確認できる。一方で、話者数の増加に伴ってモデル正答率が不安定になっていることが確認できる。特に、対象話者が 4 名の場合は顕著である（図 6.2 ～ 図 6.5）。

最後に選択特徴量数に着目する。全ての対象雰囲気において対象話者数の増加に伴って、選択特徴量数の増加を確認できる。一方で、同じ対象話者数のモデルを対象雰囲気間で比較しても大きな差は確認できない。また、選択特徴量数とモデル正答率においても本結果からは相関を確認することはできない。

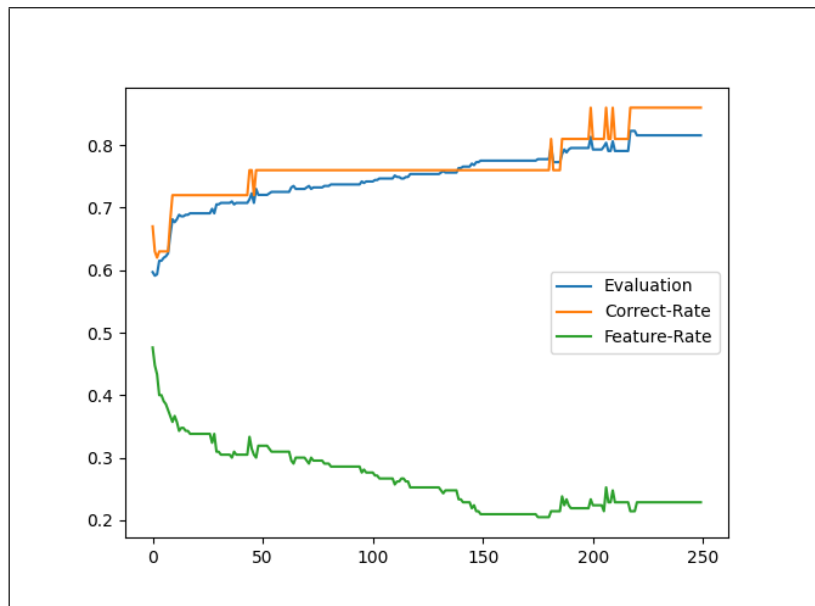


図 6.2 「盛り上がり」話者数 4 名のモデル学習過程

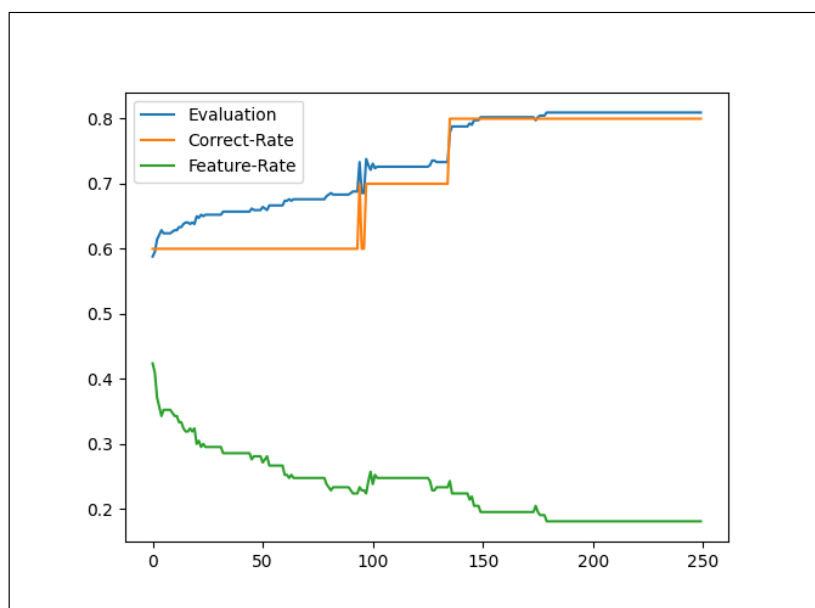


図 6.3 「真面目さ」話者数 4 名のモデル学習過程

6.4.2 特徴量選択の設定を変更したモデルの学習結果

ベースモデルから特徴量選択の各設定を変更した対話雰囲気推定モデルの学習結果の一部を表 6.8 に示す。表中の各モデルについて変更点と得られた学習結果について述べる。

モデル番号 61 は評価における交差検証を排除したモデルである。表中に掲載しているも

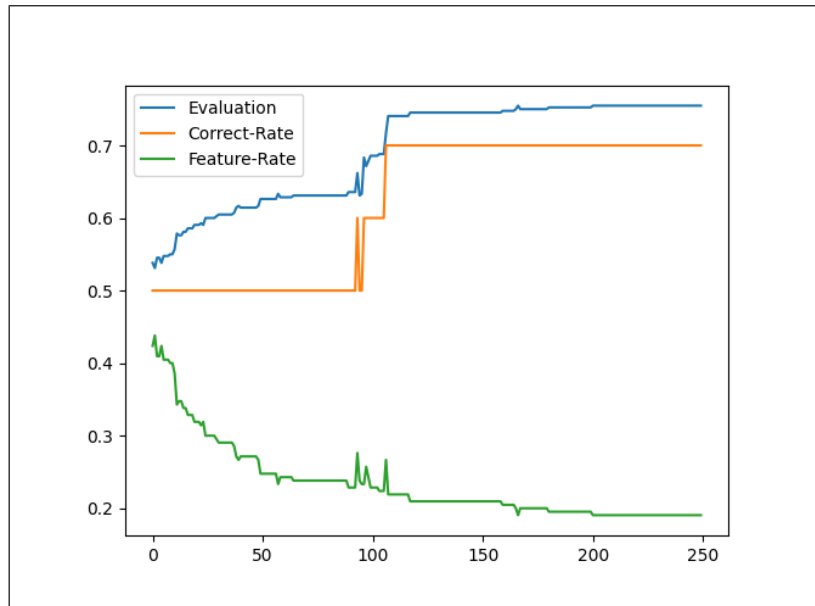


図 6.4 「明るさ」話者数 4 名のモデル学習過程

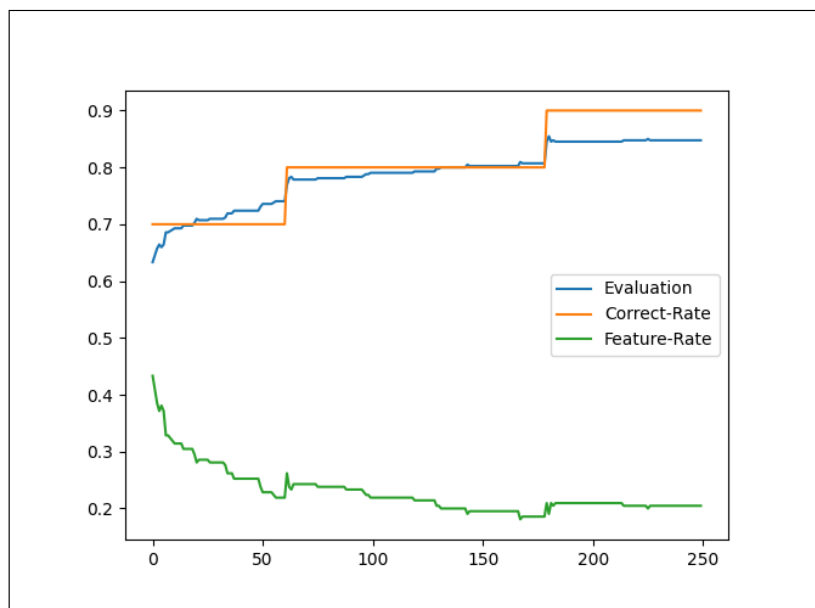


図 6.5 「くつろぎ」話者数 4 名のモデル学習過程

のはベースモデルと比較して高いモデル正答率を示しているが、何度か同条件でモデルを構築した際には 0.6 前後のモデル正答率を示す場合も確認できた。

モデル番号 62 ～ 64 は個体評価における重み付けを変更したモデルである。ベースモデルは重み付け W は 0.5 に設定している。表中のモデルの W は、それぞれモデル番号 62 は 1.0, モデル番号 63 は 0.9, モデル番号 64 は 0.1 に設定している。ベースモデルと各モデルを比較すると、 W が小さくなるに従って選択特徴量数が減少することがわかる。一方で、 W

表 6.8 特徴量選択を用いたモデルの学習結果

モデル番号	モデル概要	モデル正答率	選択特徴量数
50	ベースモデル（再掲載）	0.7956	21
61	交差検証なしモデル	0.8333	21
62	$W : 1.0$ モデル	0.7963	62
63	$W : 0.9$ モデル	0.7869	42
64	$W : 0.1$ モデル	0.6882	14
65	二点交叉モデル	0.7838	24
66	Linear SVC モデル	0.7118	28
67	k 近傍法モデル	0.7103	18
68	SVC モデル	0.8020	19

を大きくしてもモデル正答率に大きな影響がないことが確認できる。

モデル番号 65 は交叉方法を変更し、二点交叉を採用したモデルである。ベースモデルと比較すると、モデル正答率の僅かな減少と選択特徴量数の僅かな増加が確認できる。これらの状況は交叉方法だけでなく、突然変異率や集団の大きさなどを変更した際にも同様の結果が得られる。

モデル番号 66 ～ 68 は学習アルゴリズムを変更したモデルである。Linear SVC と k 近傍法を採用したモデルは、特徴量選択を用いないモデルと同様のモデル正答率が確認できる。また、ベースモデルと比較するとモデル正答率の減少、選択特徴量数の増加が確認できる。SVC を採用したモデルはベースモデルと同様の学習結果であることが確認できる。

6.5 学習結果の考察

第 6.2 節、第 6.4 節で述べた学習結果に対する考察について述べる。

はじめに、特徴量選択の有用性について述べる。本研究では、特徴量選択を用いない対話雰囲気推定モデルと用いた対話雰囲気推定モデルをそれぞれ構築している。それぞれの比較を行なった結果、多くのモデルでモデル正答率の向上と選択特徴量数の減少を確認できる。このことから、本研究で構築する対話雰囲気推定モデルにおいても特徴量選択が有効であるといえる。特徴量選択の最適な設定については、第 6.4.2 項で示した結果などからベースモデルが最適であることが示唆された。個体評価においては、重み付け和を採用することの有用性は示唆されているものの、最適な重み付け W を一意に定めることはできなかった。これは、GA を採用していることから学習の過程で乱数が発生し、それにより結果が変動するためである。よって、今後の課題として乱数を加味した上で最適な個体評価を行う手法の検討が挙げられる。しかし、本研究で構築したベースモデルは正答率、選択特徴量数ともに十分な値であることから、本研究では W は 0.5 を現状の最適値とし採用する。

次に、話者数の増加に伴う選択特徴量数の増加とモデル正答率の不安定化について述べる。第 6.4.1 項で述べた通り、モデルの対象話者数の増加により選択特徴量数が増加している。一般に選択特徴量数は少ない方が望ましく、豊田らの手法では 2 ～ 5 個にまで選択数を削減している。しかし、本研究では第 5.3 節で述べた通り最大 210 個の特徴量を抽出しており、母数からの削減割合でみると大きく削減に成功していることがわかる。また、話者数の増加に伴う選択特徴量数の増加についても、話者数が増えるに従って母数となる選択特徴量が多くなるためである。したがって、話者数の増加に伴う選択特徴量数の増加についてはモデルの評価へは影響が少ないと考えられる。一方で、話者数の増加に伴ってモデル正答率が不安定になっている現象への解決は大きな課題である。この問題の原因は学習データが少ないことにある。本研究のモデル構築手法は学習を行う直前に前処理として、対象話者数に依存した欠損値を含む学習データの削減を行なっている。それに伴い、話者数が増加するに従って学習データが不足し、結果としてモデル正答率が不安定になっている。よって、原因の解決としては学習データの拡充が挙げられる。しかし、本研究における学習データの収集、構築方法は複雑であり、一つの学習データを作成するために 30 分以上の時間を有する場合もある。そこで、学習データの自動収集システムを構築することが解決手段の一つとして挙げられる。具体的には、人間の主観による判断が必要のない全ての作業をシステムに代替し、主観による判断が問われる学習データへのラベル付けだけを行うシステムを構築することで問題の解決が期待できる。

しかし、学習データを自動で収集するシステムの構築には対話の検知を行わなくてはいけないという課題が挙げられる。第 5.1 節で述べた手法のうち、対話データの切り分けは人間の主観によって話題を判断し、その切り替わりと同時に切り分けを行なっている。システムによって、自動化を行う際には話題の切り替わりをシステムによって判断する必要がある。また、第 4.2.1 項で述べた通り本研究では参加者のプライバシーの担保と、学習コストの低減を目的に言語情報の収集を行わない。つまり、非言語情報である発話時間情報のみを用いて話題の切り替わりを判断する必要がある。その方法として対話中の無音時間による検知を提案する。話題の切り替わり時には誰も喋らず無音の時間が発生することが多い。特に 0.5 ～ 2 秒程度の無音時間が続いた場合には話題の切り替わりが頻繁に発生していた。また、話題の切り替わりが発生した際に途端に対話雰囲気に変化することは稀であり、対話雰囲気は自然に移行する性質も確認できている。よって、厳密な話題の切り替わりを判断する必要性は薄く、本手法で十分に対応が可能であるといえる。また、収集した学習データのうち無作為に抽出したいくつかを対話雰囲気推定結果とともにユーザへ示し、ユーザが修正できるような仕組みを導入することで適切な学習データの収集やモデル構築が期待できる。

第 7 章

結論

本論文では対話雰囲気推定を用いた作業通話の進め方の可視化手法と，それを用いた参加者募集システムの提案及び提案システムで用いる対話雰囲気推定モデルの構築を行なった．対話雰囲気推定モデルは従来手法を参考により作業通話に適したモデルを構築することができた．対話雰囲気推定モデルの構築に関する今後の課題として，より効果的な個体評価手法の検討や学習データの拡充が明らかとなった．特に，学習データの拡充が行うことでより学習効率の良い対話雰囲気推定モデル構築手法の発見や，新たな課題の発見が期待できるため早急な解決が求められている．また，本研究で提案したシステムは提案に留まっており，構築と検証も求められている．その際には，より多くの人が利用しやすい UI や UX も踏まえてシステムを検討していく必要がある．

将来的には作業通話だけでなく，オンラインゲームの通話を始めとした複数人による通話全般の活動雰囲気の推定等に応用できる対話雰囲気推定モデルやシステムの構築を期待している．

謝辞

本研究を進めるにあたり，お忙しい中ご指導いただいた奥野拓教授に深くお礼申し上げます．また，様々な助言をくださった研究室の後輩方や同期の学生にも深くお礼申し上げます．

発表・採録実績

発表等

- [1] 立花虎太郎・奥野拓, 雰囲気推定を用いた作業通話マッチング支援システムの提案, 第118回グループウェアとネットワークサービス・第36回コンシューマ・デバイス&システム・第33回デジタルコンテンツクリエーション合同研究発表会, 2023-01-24, 発表予定.

参考文献

- [1] “Discord” . <https://discord.com/>, (参照 2023-01-12).
- [2] Zajonc.R.B, Social facilitation, Science, 149, p.269-274, 1965.
- [3] 松本芳之, 観察者効果に関するフィード研究: 技能水準, 課題難度, 遂行状況の関係について, 実験社会心理学研究, 第 26 巻, 第 2 号, p.115-123, 1987.
- [4] 宮本正一, 観察者の存在による選択反応時間の抑制, 心理学研究, 第 58 巻, 第 4 号, p.240-246, 1987.
- [5] “Twitter” . <https://twitter.com/>, (参照 2023-01-12).
- [6] 原田悦子, 人の視点から見た人工物研究 (認知科学モノグラフ 6) , 共立出版, 1997.
- [7] 木村泰之・都築誉史, 集団意思決定とコミュニケーション・モードコンピュータ・コミュニケーション条件と対面コミュニケーション条件の差異に関する実験社会心理学的検討, 実験社会心理学研究, 38 巻, p.183-192, 1998.
- [8] 西村洋一, 対人不安、インターネット利用、およびインターネットにおける対人関係, 社会心理学研究, 19 巻, p.124-134, 2003.
- [9] 都築誉史・木村泰之, 大学生におけるメディア・コミュニケーションの心理的特性に関する分析一対面、携帯電話、携帯メール、電子メール条件の比較, 立教大学応用社会学研究, 42 巻, p.15-24, 2000.
- [10] 徳久良子・寺嶋立太, 雑談における発話のやりとりと盛り上がりの関連, 人工知能学会論文誌, 21 巻, 2 号 A, p.133-142, 2006.
- [11] 伊藤秀樹・重野真也・西本卓也・荒木雅弘・新美康永, 対話における雰囲気分析, 情報処理学会研究報告 SLP, 音声情報処理, No.10, p.103-108, 2002.
- [12] 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, 27 巻, 2 号 SP-B, p.16-21, 2012.
- [13] 北村太一・小川祐樹・諏訪博彦・太田敏澄, コミュニケーションに着目した Twitter フォロワーユーザ推薦, 人工知能学会全国大会論文集, 3E1-R-6-5, p.1-4, 2012.
- [14] 田村政人・小林亜樹, Twitter における会話しやすいユーザの推薦手法, 情報処理学会第 75 回全国大会講演論文集, 2013(1), p.605-607, 2013.
- [15] 久米雄介・打矢隆弘・内匠逸, 興味領域を考慮した Twitter ユーザ推薦手法の提案と評価, 情報処理学会研報告 SIG, Vol.2015-ICS-179, No.1, p.1-8, 2015.

- [16] 樋川一幸, 適切な距離の学生に依頼可能な Bot を利用した実験協力者募集手法の研究, 明治大学大学院先端数理科学研究科修士 (工学), 2020.
- [17] “LINE” , <https://line.me/ja/>, (参照 2023-01-12).
- [18] NII 音声資源コンソーシアム：対話音声コーパス (PASD) (1993-1996) .

図目次

1.1	Twitter で「作業通話 勇気」で検索した結果	2
2.1	Twitter で「作業通話 募集」で検索した結果	5
4.1	システムイメージ	10
4.2	作業嗜好の可視化イメージ	12
4.3	生成する投稿のイメージ	13
5.1	対話データにおける発話の例	15
6.1	ベースモデルの学習過程	27
6.2	「盛り上がり」話者数 4 名のモデル学習過程	28
6.3	「真面目さ」話者数 4 名のモデル学習過程	28
6.4	「明るさ」話者数 4 名のモデル学習過程	29
6.5	「くつろぎ」話者数 4 名のモデル学習過程	29

表目次

4.1	推定を行う雰囲気	11
5.1	特徴量として採用する統計量と算出式	16
5.2	統計量を用いた特徴量	18
5.3	比較量を用いた特徴量	19
5.4	学習の設定	20
5.5	遺伝的アルゴリズムの設定	21
6.1	「盛り上がり」を対象とした特徴量選択を用いない学習の結果	23
6.2	「真面目さ」を対象とした特徴量選択を用いない学習の結果	23
6.3	「明るさ」を対象とした特徴量選択を用いない学習の結果	24
6.4	「くつろぎ」を対象とした特徴量選択を用いない学習の結果	24
6.5	ベースモデルの設定値	25
6.6	特徴量選択を用いたモデルの学習結果	26
6.7	特徴量選択前後のモデル正答率の t 検定	27
6.8	特徴量選択を用いたモデルの学習結果	30