

# 対話の初期段階を支援するための 特定人物模倣型チャットボット PMCBの開発

学籍番号：s1922027

氏 名：萩本 新平

指導教員：中村亮太 上林憲行

初めて  
中村先生や上林先生、他の先生と  
話したとき

就職して初めて上司・先輩と話するとき

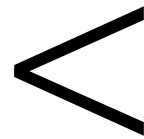
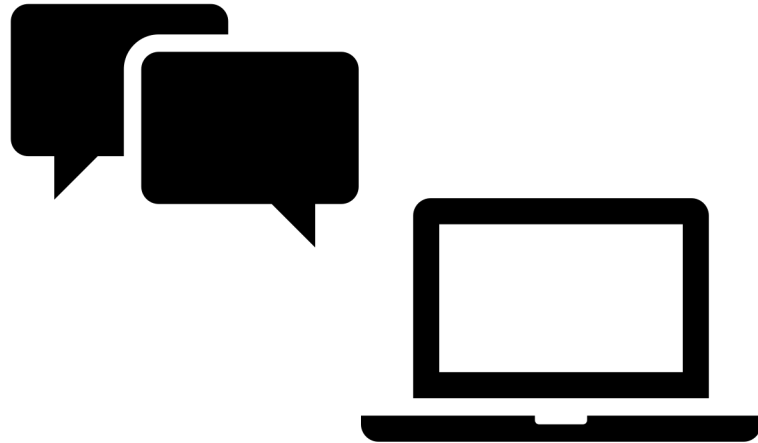
！！！！話しかけたい！！！！  
！！！！緊張せずに話したい！！！！

## 特定人物模倣型チャットボットPMCB（Personal-Mimic Chat Bot）

事前にPMCBを使うことで  
初対面時のコミュニケーションを  
円滑に進められるようにする

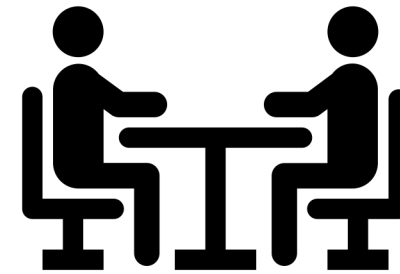
# 従来研究

コンピュータ・コミュニケーション



緊張感

対面コミュニケーション

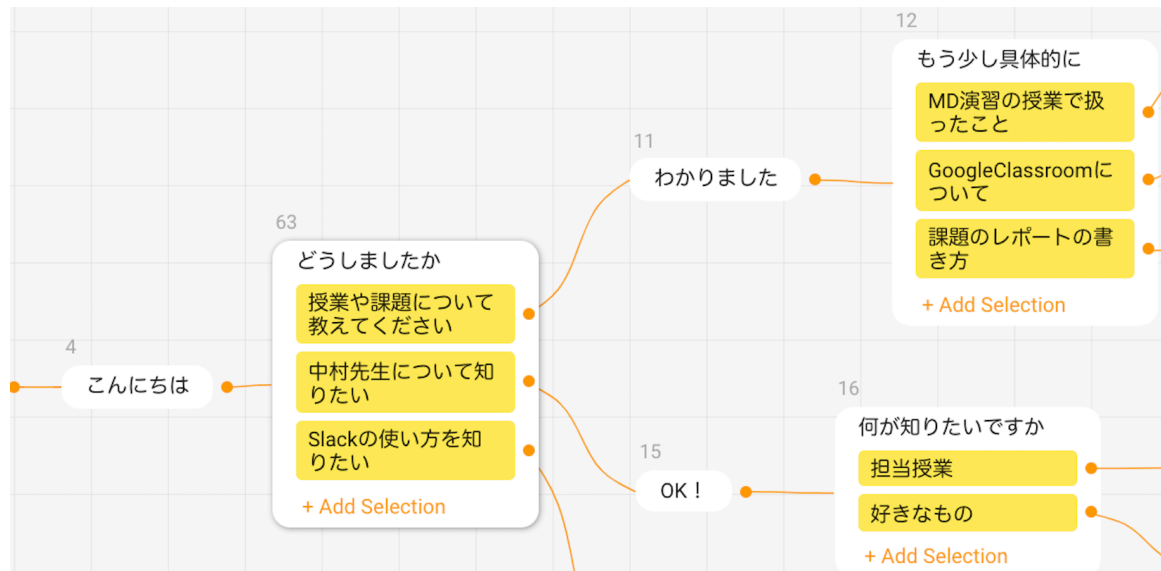


本人みたいなボットと話す練習ができれば  
話しかけられるようになる！  
緊張せず話せるようになる！

木村，都築：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

# ボットの作成は Bot Editor

簡単にシナリオを書くことができ即実行可能



模倣型チャットボットのシナリオ作成には  
その人物の特徴を分析する必要がある

Ryota Nakamura (Bot)

こんにちは

Ryota Nakamura (Bot)

どうしましたか

Guest

授業や課題について教えてください

Ryota Nakamura (Bot)

わかりました

Ryota Nakamura (Bot)

もう少し具体的に

Guest

GoogleClassroomについて

Ryota Nakamura (Bot)

OK !

Ryota Nakamura (Bot)

Classroomもいくつか機能があります

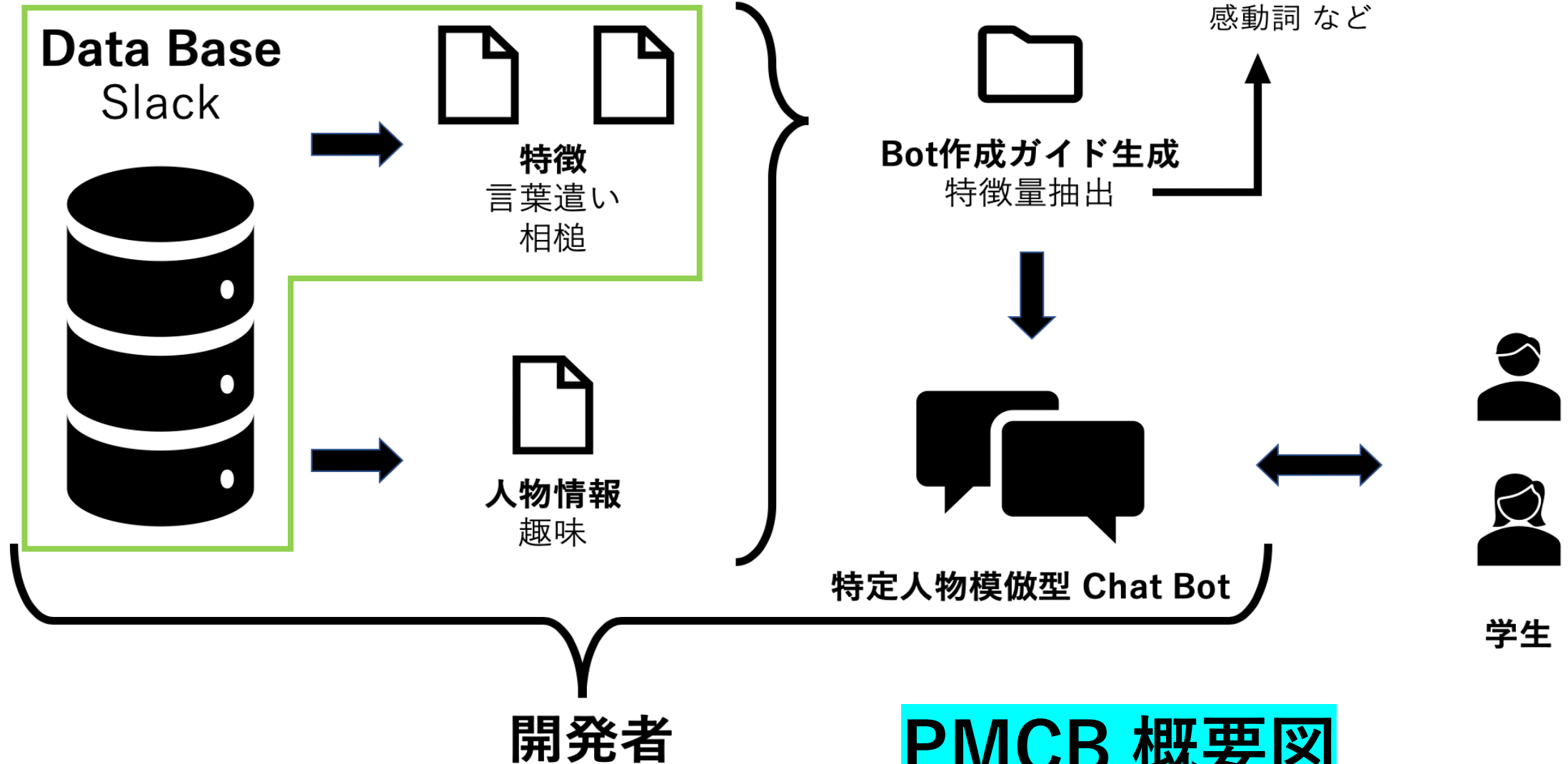
Type text here.

課題の出し方について

授業のスライドや資料の場所

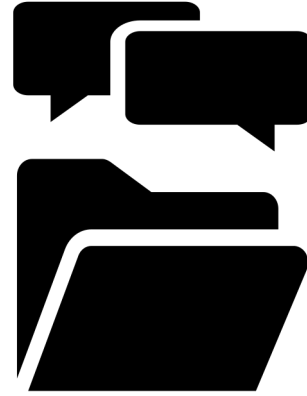
1 つ前に戻る

## 今回行ったこと





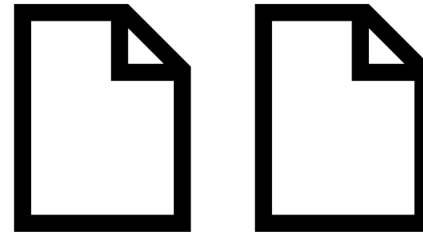
Slack ログ



ログの取得  
SlackのAPIを使用



メッセージテキスト抽出



特徴  
言葉遣い  
相槌



データクレンジング

手順

# 分析\_1

- 1-1. 教員ごとにSlackのメッセージに違い・特徴はあるか
- 1-2. 教員の役割によって違い・特徴はあるか

## 方法

- ログから先生ごとテキストを集計
- データの前処理（右）
- BERTで処理

### <BERT>

- Bidirectional Encoder Representations from Transformers
- 自然言語処理モデル
- ラベルなしデータで事前学習
- 学主婦済みモデルをベースとして、ラベル付きデータでファインチューニング

表1. ラベリング

Feature	Label
中村先生のメッセージ	0
上林先生のメッセージ	1
長谷川先生のメッセージ	2
中西先生のメッセージ	3
林先生のメッセージ	4
佐々木先生のメッセージ	6
福原先生のメッセージ	6

### ➤ 使用したデータ

- Slackの全チャンネルの、先生方のメッセージデータ

### ➤ 前処理

- 先生ごとにラベル付け（右表）
- ノイズ除去  
メンション<@\_\_\_>など

# 結果\_1-1

表2. BERTによる教師全員の分類

	Precision	Recall	F1-score	Support
中村先生	0.65	0.66	0.65	183
上林先生	0.72	0.84	0.77	226
長谷川先生	0.59	0.63	0.61	76
中西先生	0.60	0.47	0.53	19
林先生	0.61	0.35	0.45	54
佐々木先生	0.83	0.62	0.71	24
福原先生	0.20	0.06	0.10	16
Accuracy			0.67	598

比較的高い正答率  
全体で約60%

先生ごとに分類ができている



先生それぞれに  
メッセージに違いがある

→ 違いが特徴と言える

データが比較的少なかったので  
比較的正答率は低い



## 結果\_1-2

表3. BERTによる分類

	Precision	Recall	F1-score	Support
中村先生	0.88	0.85	0.87	80
上林先生	0.89	0.91	0.90	103
Accuracy			0.89	183

正答率  
約 **90 %**



違いがある！！

2人にそれぞれに特徴があるということ

先生ごとにある程度違いがあり、その違いに**役割は依存しない**

# 分析\_1

## 1-3. 期間によって違い・特徴はあるか

### 方法

- 中村先生のテキストを取得
- データの前処理
  - 3月~7月、8月~11月に分割
  - ラベル付け
- BERTで処理

### ➤ 使用したデータ

- 中村先生メッセージデータ

### ➤ 前処理

- 期間でラベル付け（右表）
- ノイズ除去
  - メンション<@\_\_\_>など

表4. ラベリング

Feature	Label
3月	0
・	0
7月	0
8月	1
・	1
11月	1

## 結果\_1-3

表5. BERTによる分類

	Precision	Recall	F1-score	Support
中村先生_前期	0.79	0.87	0.83	62
中村先生_後期	0.72	0.60	0.66	35
Accuracy			0.77	97

正答率  
約 **75 %**



前期・後期で違いがある  
前期・後期を分ける特徴がある

# 分析\_2-1

中村先生のメッセージにはどのような特徴が存在するのか  
高い正答率で分類できた上林先生と比較する

## 方法

- ログから中村先生と上林先生のテキストを集計
- 使用データの前処理
- MeCabで処理
- 残ったノイズ除去
- 品詞ごとに集計、比較

### ➤ MeCab

- 形態素解析エンジン
- 品詞情報を利用した解析・推定ができる
- 日本語の形態素解析でよく使われている

### ➤ 使用したデータ

- Slackの全チャンネルの、中村先生と上林先生のメッセージデータ

### ➤ MeCabでの分類のための前処理

- ノイズになる記号の消去  
    <@UTNDSF5F>、<!channel>
- CSVファイルに出力

### ➤ ノイズ除去

- 残った「」や<>を除去  
    集計しやすいように

# 結果\_2-1

表5. 特徴有りの比較、無しの比較

	副詞				名詞			
	中村先生		上林先生		中村先生		上林先生	
0	よろしく	23	とても	21	授業	75	データ	147
1	まだ	14	まず	16	グループ	65	こと	140
2	どうも	13	既に	13	教室	64	皆さん	115
3	必ず	9	必ず	13	よう	61	グループ	99
4	より	7	よろしく	13	お願い	58	日	75
5	そう	6	実際	11	課題	58	人	75
6	いつも	5	さらに	10	こと	56	時間	71
7	できるだけ	5	再度	8	人	54	よう	64
8	特に	5	特に	8	学生	52	参加	58
9	どんどん	5	しっかり	8	データ	51	学生	57

中村先生と上林先生の  
副詞・名詞  
双方、上位10個で比較

重複した単語  
副詞：**3個**  
名詞：**5個**

名詞より副詞に差がある

→ 副詞に注目すると特徴が  
得られる

## BERTでの分類

先生全員 正答率約60% : 先生ごとに違い・特徴がある  
中村先生と上林先生 正答率約90% : 役割に関係なく違いがある  
前期・後期 正答率約75% : 前期・後期を分けうる特徴がある

## MeCabの解析

中村先生と上林先生の副詞と名詞：  
名詞よりも副詞の重複が少なく、副詞に着目すると特徴を得られる

## まとめ

先生方それぞれSlackメッセージに違い・特徴があり、その違い・特徴はその人物の役割には依存しない。  
また同じ人物でも時期によってメッセージに差がある。

また、名詞より副詞の方が人物によって差があり、副詞に着目すると特徴が得られやすいということが分かった。

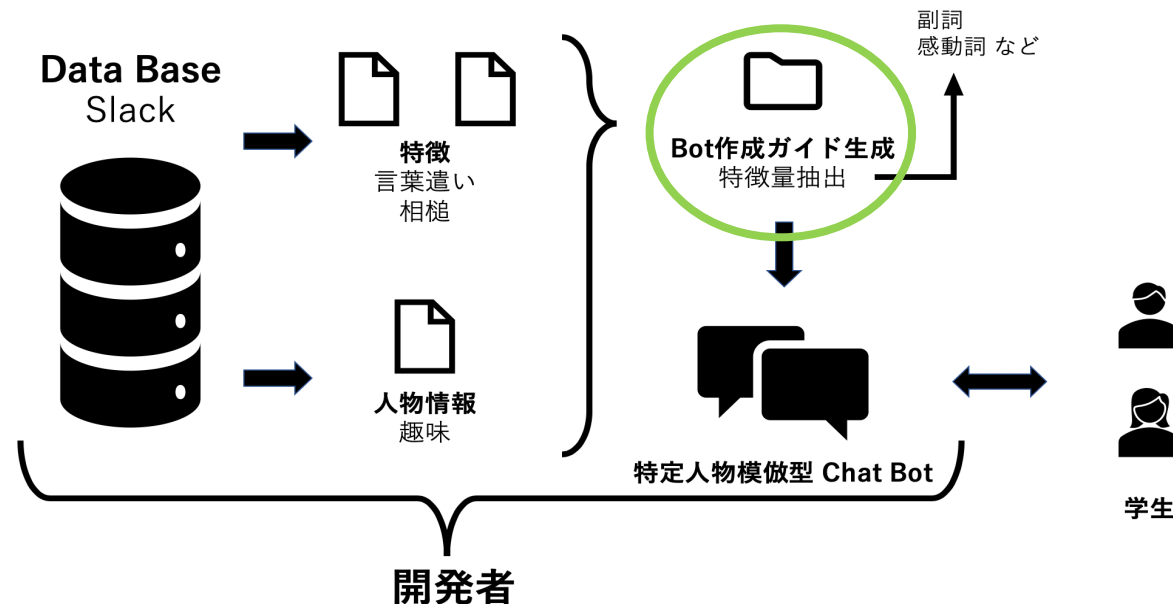
# 今後の展望

## 今後の分析

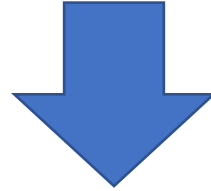
- 副詞以外の品詞（感動詞や接続詞、形容詞など）は特徴になり得るのか
- 前期と後期の（時期による）違いは、常体・敬体が影響していると推測する  
常体・敬体の割合  
対象人物によって変化するのか

## 実装

実装のために、特徴や人物情報などをまとめた**ボット作成のためのガイド**を作成する。



中村先生を模倣するボットの実装が完了



他の先生でも同じ手順でできるか検証

可能：DMなどの対話データを用いた分析に進む  
さらに、slackとボットの連携をする

不可能：手順・方法などの見返し、再度模倣に挑戦



**PMCB完成**





ご静聴

ありがとうございました！