SASAKILAB shirakawa

会話ログ分析による パーソナライズされた話題生成手法の提案

数理工学科４年　佐々木研究室
白川桃子

1 研究目的・期待される成果

2 本研究の流れ

・興味のあるトピック抽出
・話題テーマ決定
・トークテーマの生成

3 今後の目標

本研究の目的

会話ログから興味のあるトピックを抽出し
ユーザー同士の共通トピックから話題を生成する

期待される成果

ユーザーの興味関心を分析し
円滑なコミュニケーションを実現する

本研究の流れ

ユーザーの会話ログを分析

興味のあるトピックを抽出

データ取得・前処理
Zero-shot-learning
トピック抽出

ユーザー同士の共通トピックから

話題テーマを決定

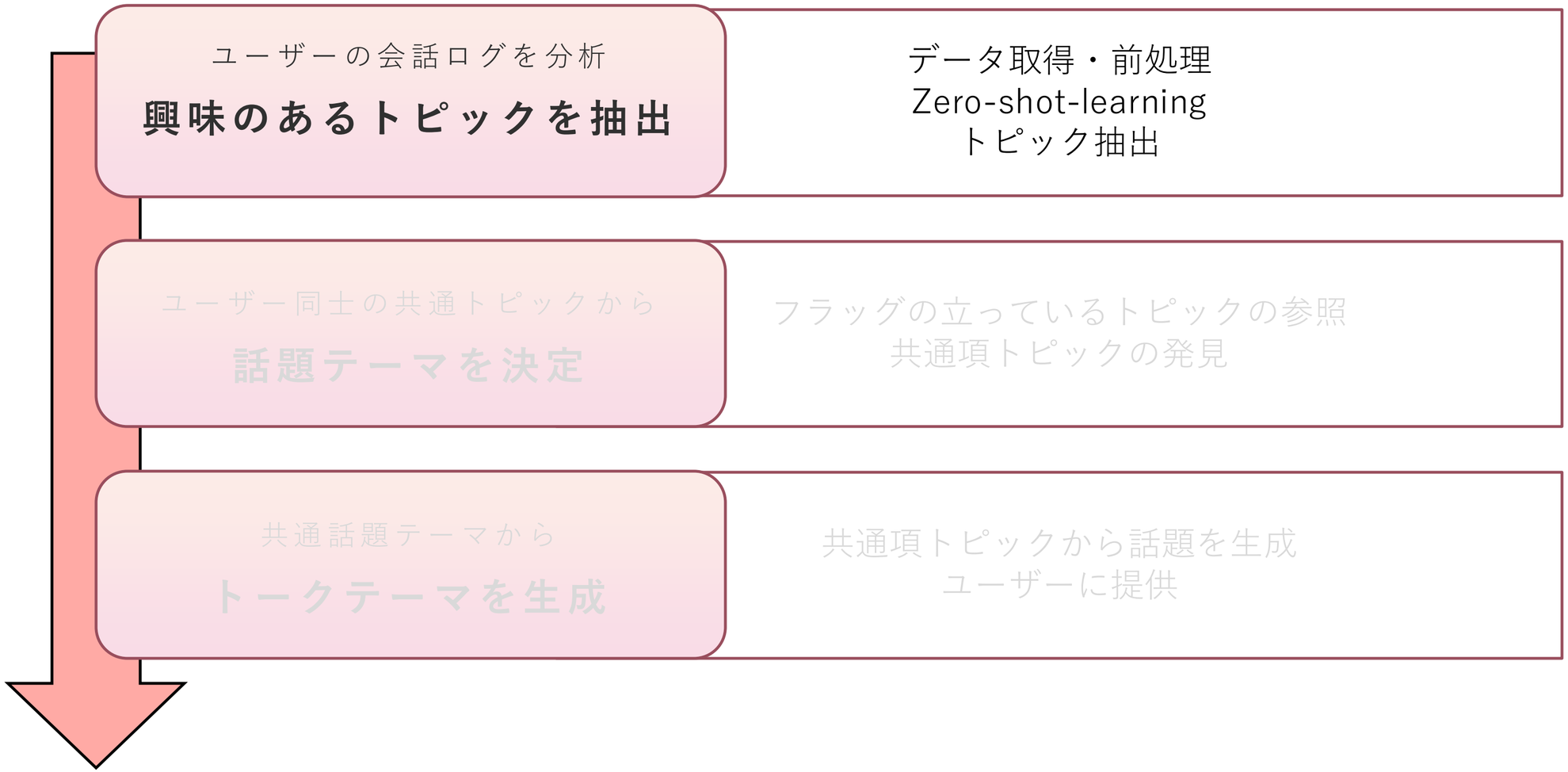
フラッグの立っているトピックの参照
共通項トピックの発見

共通話題テーマから

トークテーマを生成

共通項トピックから話題を生成
ユーザーに提供

本研究の流れ


ユーザーの会話ログを分析
興味のあるトピックを抽出
データ取得・前処理
Zero-shot-learning
トピック抽出
ユーザー同士の共通トピックから
話題テーマを決定
フラッグの立っているトピックの参照
共通項トピックの発見
共通話題テーマから
トークテーマを生成
共通項トピックから話題を生成
ユーザーに提供

活用するデータ・前処理

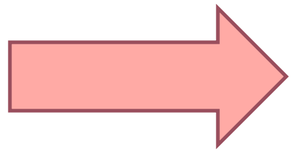
名古屋大学の会話コーパスを活用
(形式は図1を参照)

→不必要な情報を削除し，DataFrameでデータ整理

@データn番（会話時間）
@収集年月日：xxxx年xx月xx日
@場所：会話場所
@参加者ナンバー：性別，年齢，出身，所在
.
.
@参加者の関係：関係性
参加者ナンバー:会話内容
.
.

%com:　補足情報
()　相槌
<笑いor間>　自分の笑い，ある程度の沈黙

図1：名古屋大学会話コーパス構成



収集年月日	場所	追加	時間	参加者	会話内容
2001-10-16	ファミリーレストラン	None	35	M023/F107/F023/F128	F107/＊＊＊の町というのはちいちゃくって、城壁がこう町全体をぐるっと回ってて、それが城壁...
2001-10-16	ファミリーレストランガスト	None	60	F107/F023/F128	F107/今度はーイギリスにもアメリカと同様のテロが起こるだろうって言ったんだってよ。F1...
2001-10-23	車中（某大から所属大学への帰り道。運転者F033）	None	43	F056/F033	F033/倒れちゃう。F056/いきなり倒れた。F033/どうしよう。あっ、この間に。...
2001-10-23	車中（知立駅より西尾市まで。運転者M018）	None	35	M018/F128	F128/いや、別にいいよ。ローソンでいいやろ。ちょっと倒していい、これ。どうよ、調子は。...
2001-10-23	M023の自宅	None	55	F116/M026/M023/F128	F128/来てたときによく貸してもらったやつだ。M023/そう、そんな感じのとこ。F12...

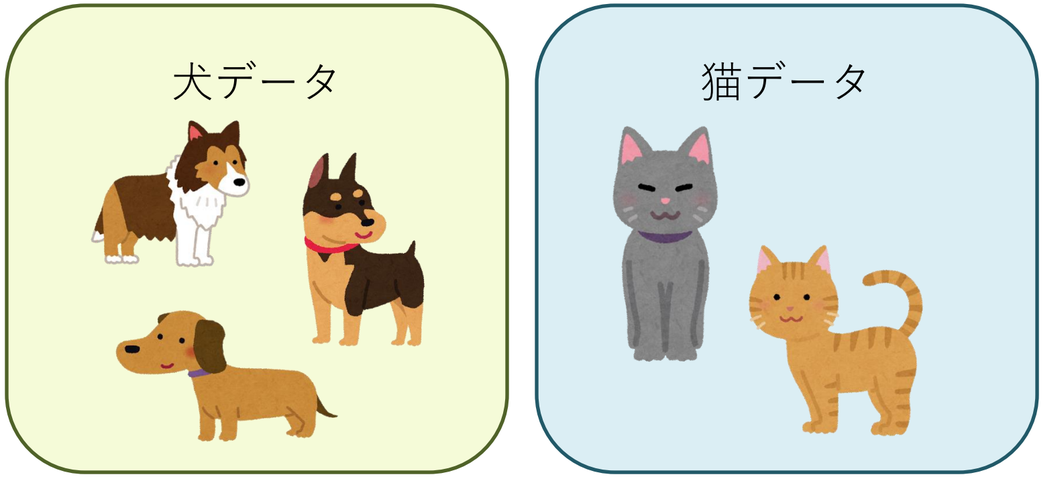
図2：前処理後の名古屋大学会話コーパス

データが全部で**129会話**しかない
→全データをユーザーとして活用したい

Zero-shot-learning とは

学習データが全く存在しないものに対して
事前に特徴を学習する形で識別することが可能

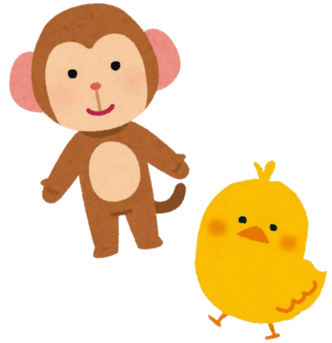
【trainデータ】



【testデータ】

一般的な学習方法の場合…
trainデータと
同じ項目について予測

Zero-shot-learning…
tainデータとは違う
新たな項目について予測



Zero-shot-learning とは

学習データが全く存在しないものに対して
事前に特徴を学習する形で識別することが可能

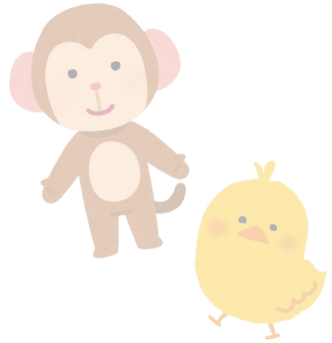
【train

犬データ

場合…

測

Zero-shot-learning…
tainデータとは違う
新たな項目について予測



【 解決へのアプローチ方法 】

ラベルの代わりに文章を使う
膨大な量のデータで事前学習をする

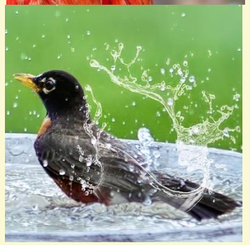
Zero-shot-learning の仕組み

トレインデータ

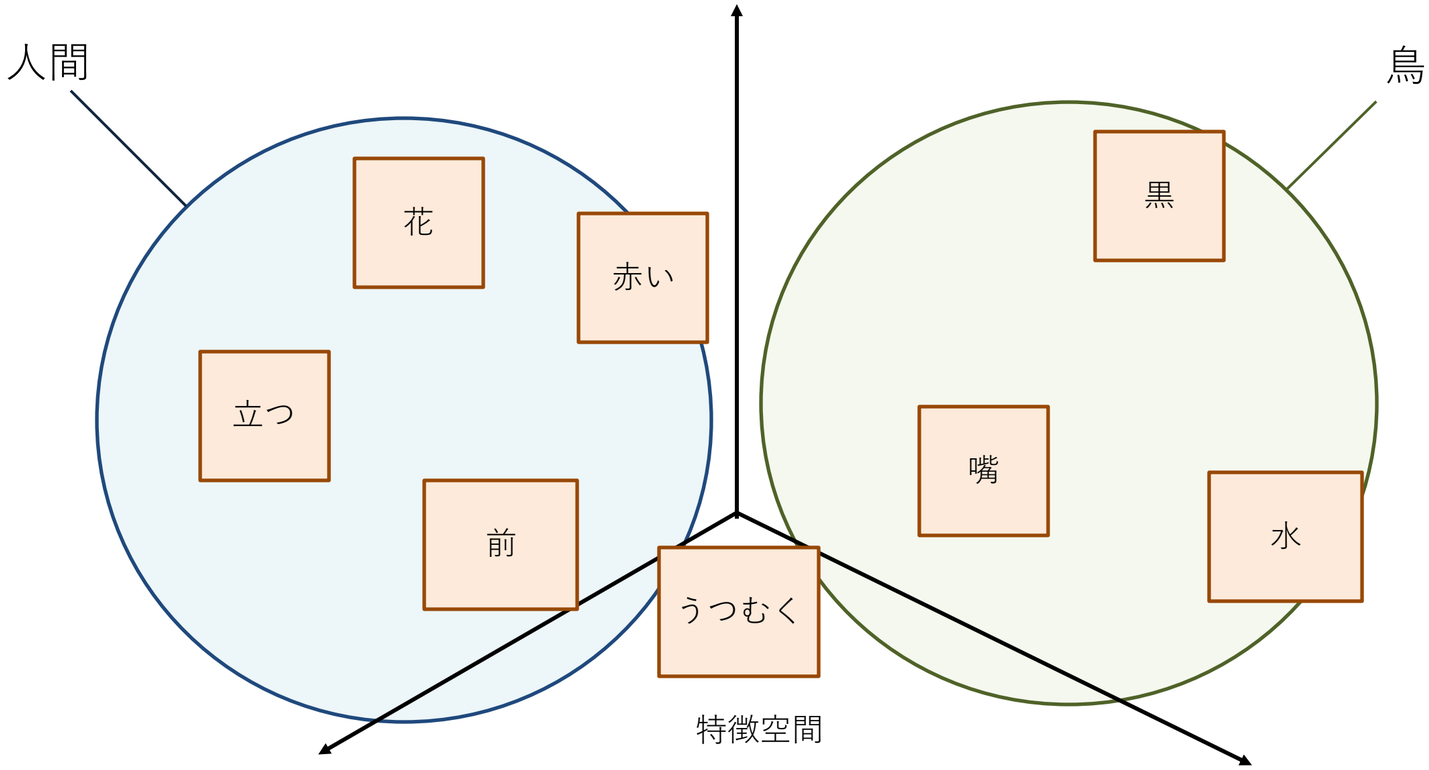
赤いうちわを持って
花の前に立つ女性

黄色い花の前で
うつむく女性

赤い羽と黒い目を
持った鳥がうつむく



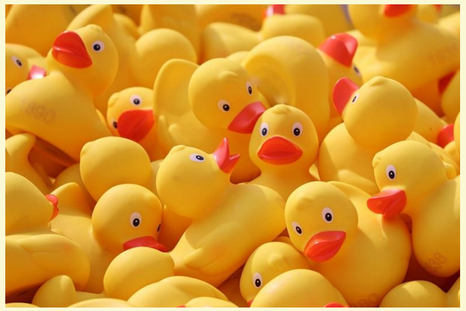
水の上で遊ぶ黒い鳥



細かい情報まで特徴空間に表現可能

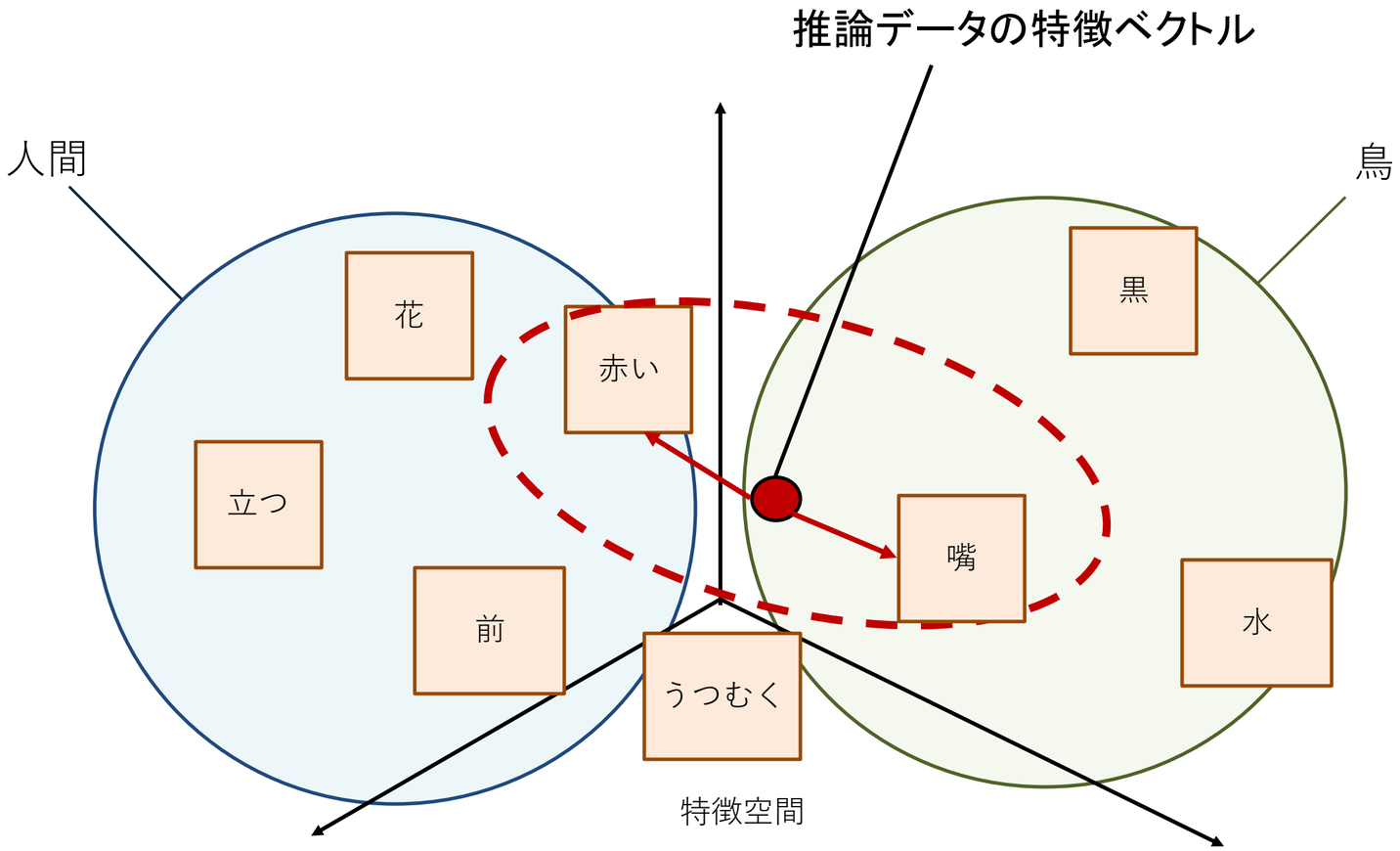
Zero-shot-learning の仕組み

テストデータ



推論データに近い特徴を
使い文章を出力

赤い嘴の鳥



データ量が増えると
特徴数を増やすことができる

今回のtrainデータ,testデータ

ラベル候補の中からそれぞれの会話に**最適なラベル**を予測したい

【trainデータ】
日本語SNLI(JSNLI)データセット　約550,000データ
各行が「ラベル　前提　仮説」で構成されている.
(e.g.)　entertainment　自転車で2人の男性がレースで競います。　人々は自転車に乗っています。

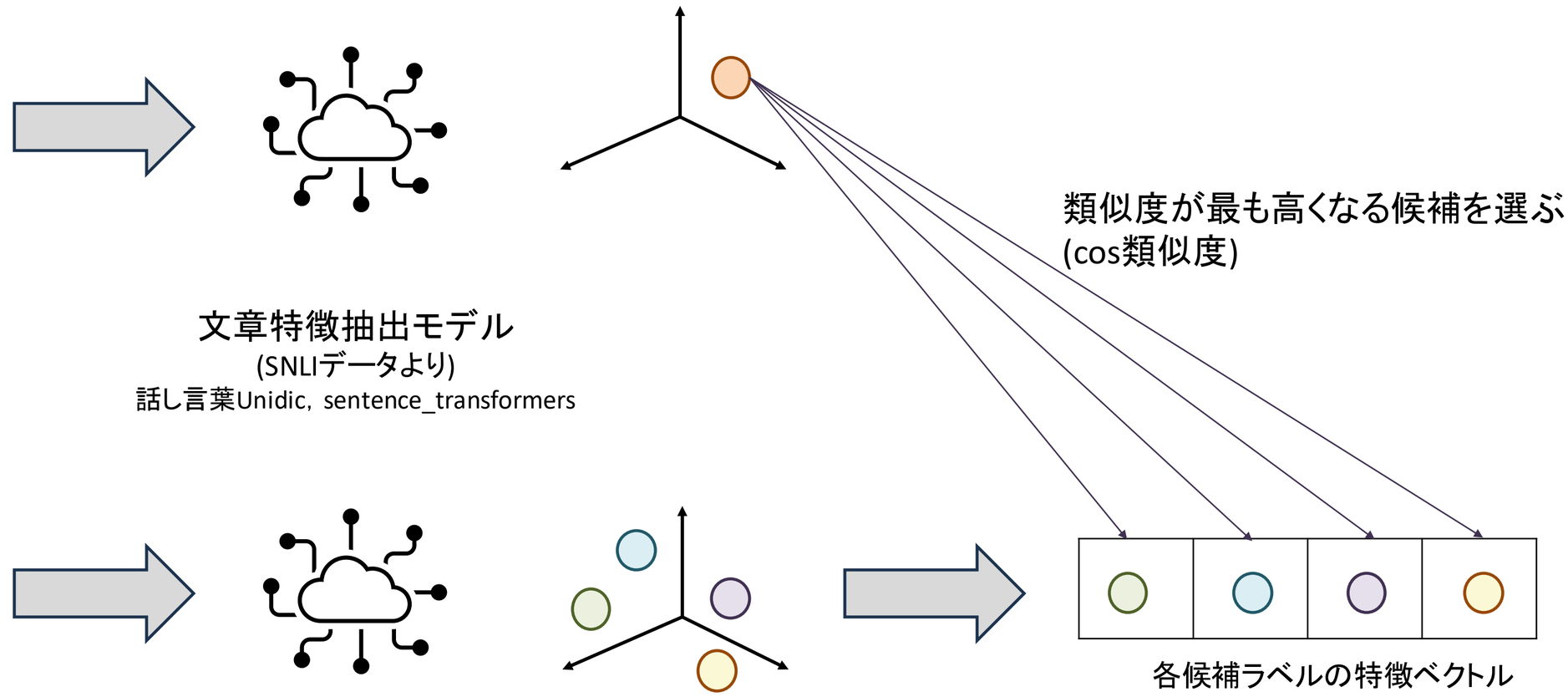
【testデータ】
・ラベル候補
note(メディアプラットフォーム)のカテゴリ

・予測会話
名古屋大学の会話コーパスの会話ログ

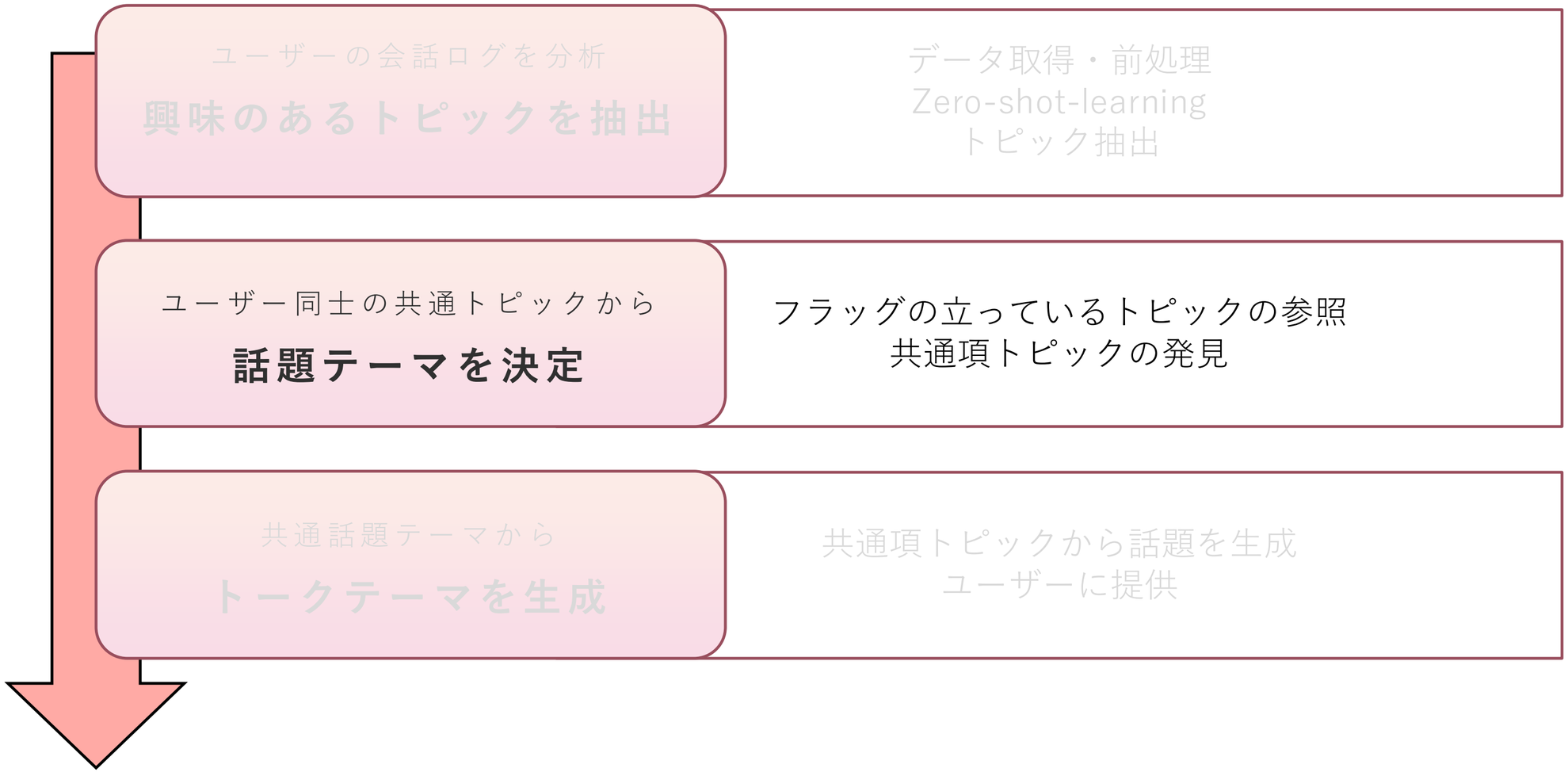
Zero-shot-learning の活用方法

会話データ
(名大会話コーパスより)

候補ラベル
(noteより)

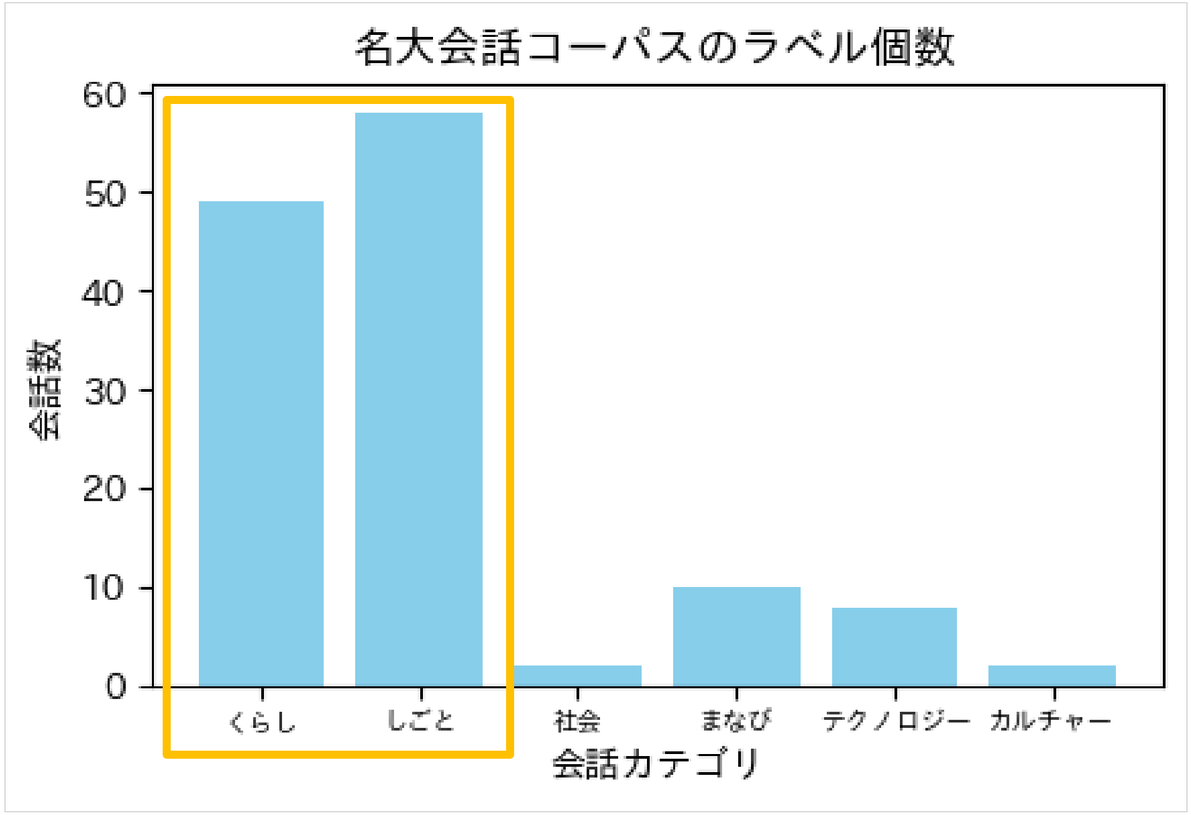


本研究の流れ


ユーザーの会話ログを分析
興味のあるトピックを抽出
データ取得・前処理
Zero-shot-learning
トピック抽出
ユーザー同士の共通トピックから
話題テーマを決定
フラッグの立っているトピックの参照
共通項トピックの発見
共通話題テーマから
トークテーマを生成
共通項トピックから話題を生成
ユーザーに提供

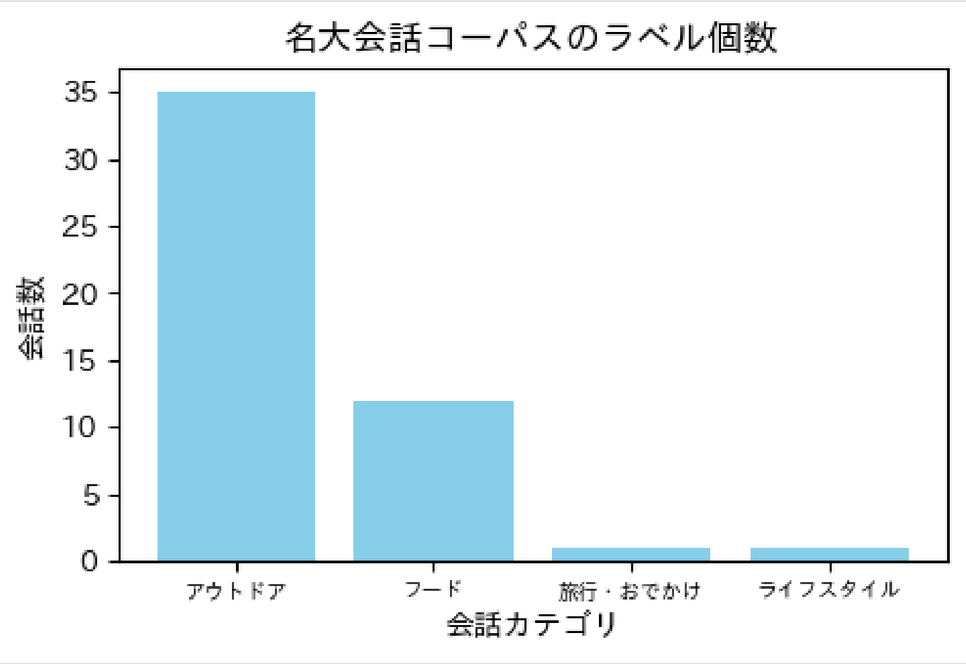
ラベル予測結果（大分類）

各ラベルと文章のマッチ度を計算し
最大値を取るものをその会話のラベルとする

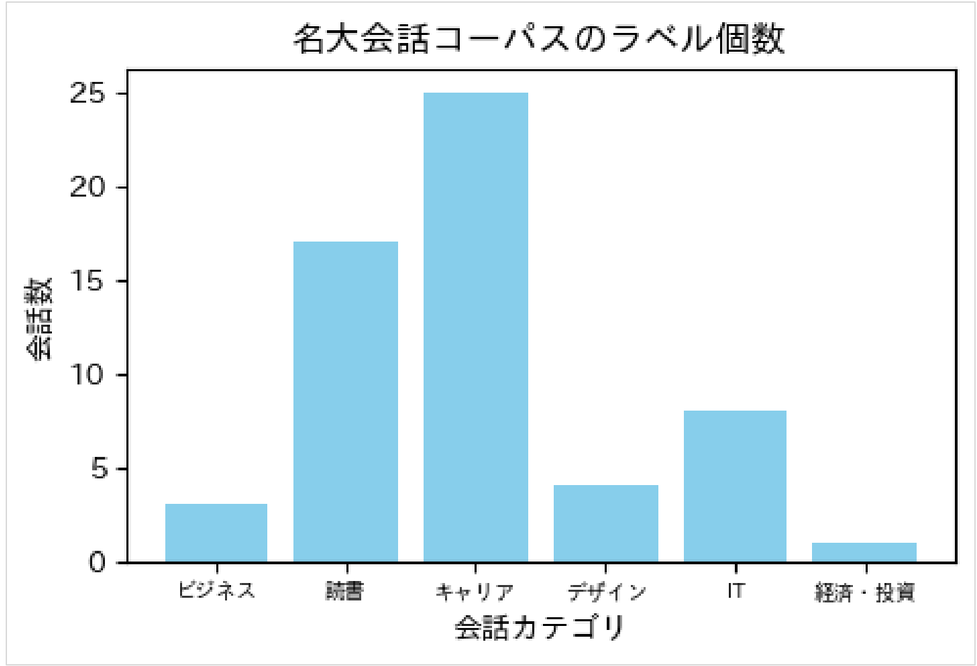


「くらし」と「しごと」だけ分類された会話だけ明らかに多い為この2項目だけ小分類に分けて予測する

ラベル予測結果（小分類）



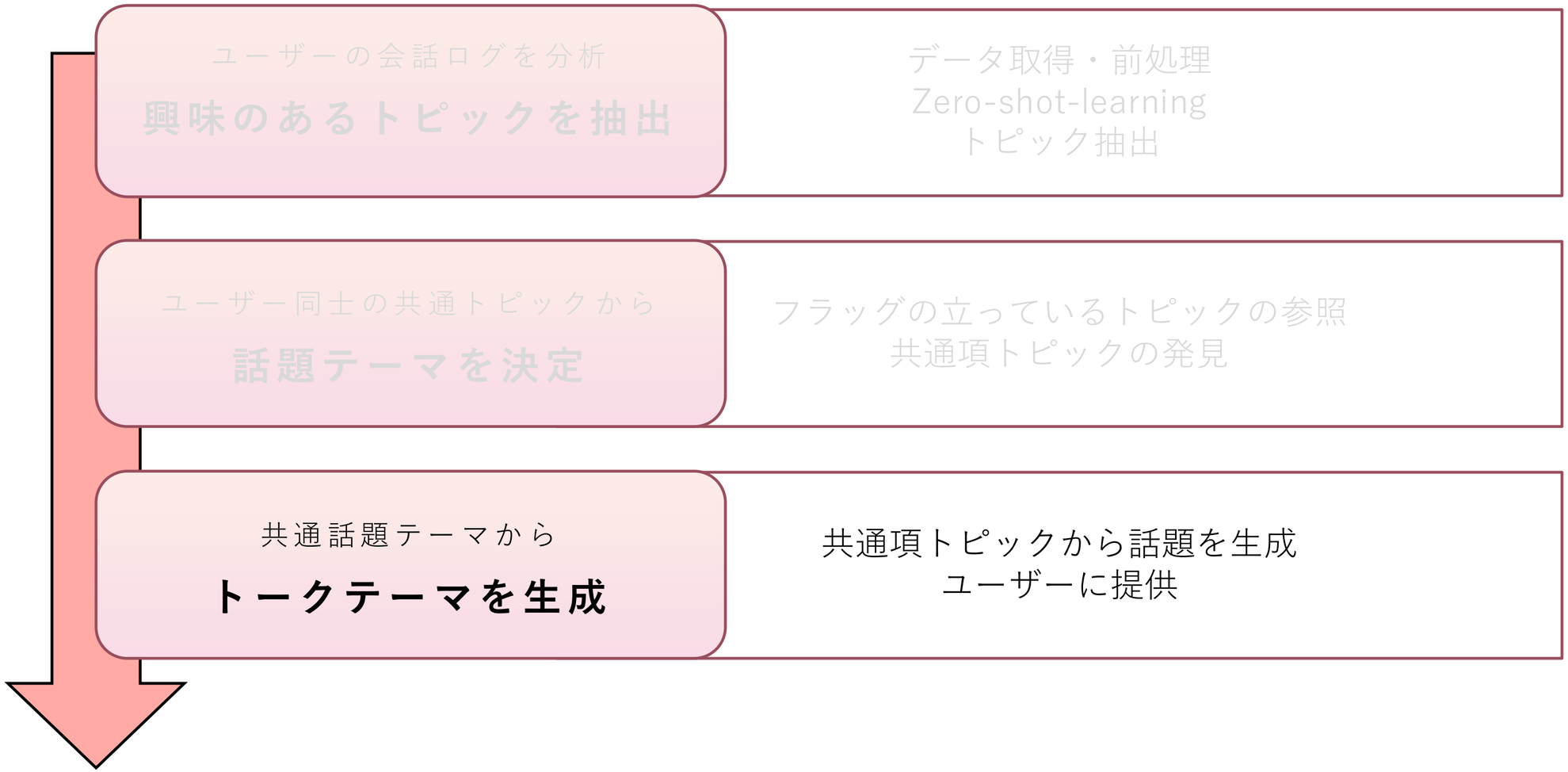
くらし
小分類にしてもなお偏りがある



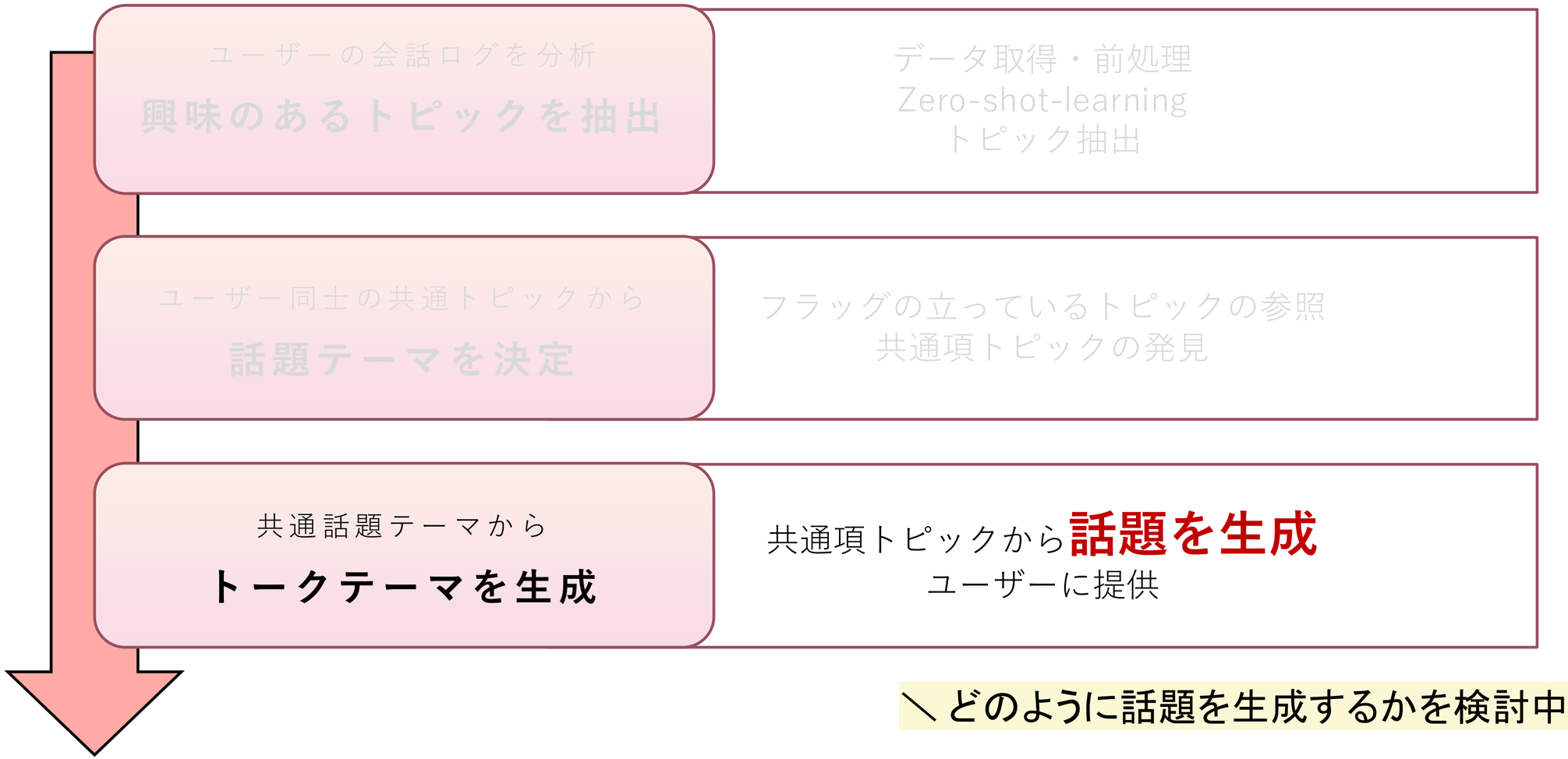
しごと
小分類にするとある程度ばらけた

以上のラベル付けされた会話の参加者にそのラベルを付与する

本研究の流れ


ユーザーの会話ログを分析
興味のあるトピックを抽出
データ取得・前処理
Zero-shot-learning
トピック抽出
ユーザー同士の共通トピックから
話題テーマを決定
フラッグの立っているトピックの参照
共通項トピックの発見
共通話題テーマから
トークテーマを生成
共通項トピックから話題を生成
ユーザーに提供

本研究の流れ



＼どのように話題を生成するかを検討中／

1 研究目的・期待される成果

2 本研究の流れ

・興味のあるトピック抽出
・話題テーマ決定
・トークテーマの生成

3 今後の目標