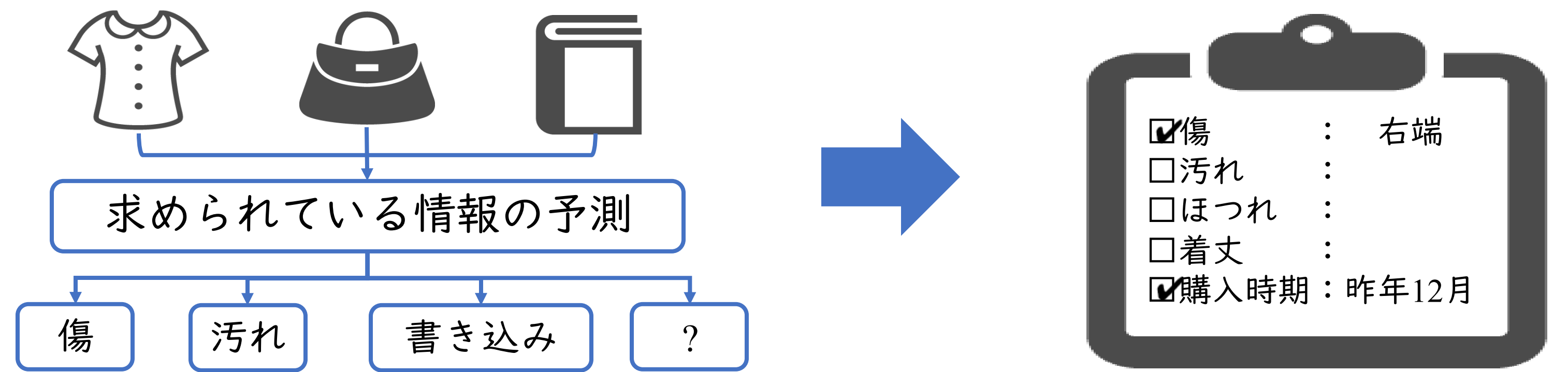


## 1. 背景と目的

- **背景**
    - フリマサイトでは**商品の状態**に関する情報が重要<sup>[1][2]</sup>
    - 購入者が求める情報を網羅した商品説明文の作成は**困難**
  - **問題**
    - 商品ごとに求められる情報(属性)は異なる
- ➡ 出品時に商品特性に合わせた必要な情報の入力を補助

### ● 目的

- 商品カテゴリ**特有**の商品属性の抽出
- 抽出した商品属性と対話システムを用いた商品説明文の**作成補助**



[1] Yang, X., Ma, Y., & Han, J. (2023). Product information the more the better? A moderated mediation model to explain consumer purchasing intention in cross-border e-commerce. *Asia Pacific Journal of Marketing and Logistics*, 35(11), 2601-2620.  
 [2] Shen, Y., Choi, P., Li, J., Zhang, X., & Han, J. (2019). How Online Descriptions of Used Goods Affect Quality Assessment and Product Preferences: A Conjoint Study (*Journal of Applied Business and Economics*, 21(2)). <https://doi.org/10.33423/jabe.v21i1.1457>

## 2. 提案手法

### ● LLMを用いたデータの前処理



コメントデータから不要な情報を取り除く必要

### ● BERTopicを用いた商品属性の抽出

- 前処理を行ったコメントデータから文単位でのクラスタリングを実施
- 各クラスタからLLMで商品属性を決定

### ● 手順

- コメントデータから商品情報以外の文を除外
- 文単位でクラスタリング, 重み付けによってカテゴリ特有の属性を抽出
- 対話システムによる属性フィリング

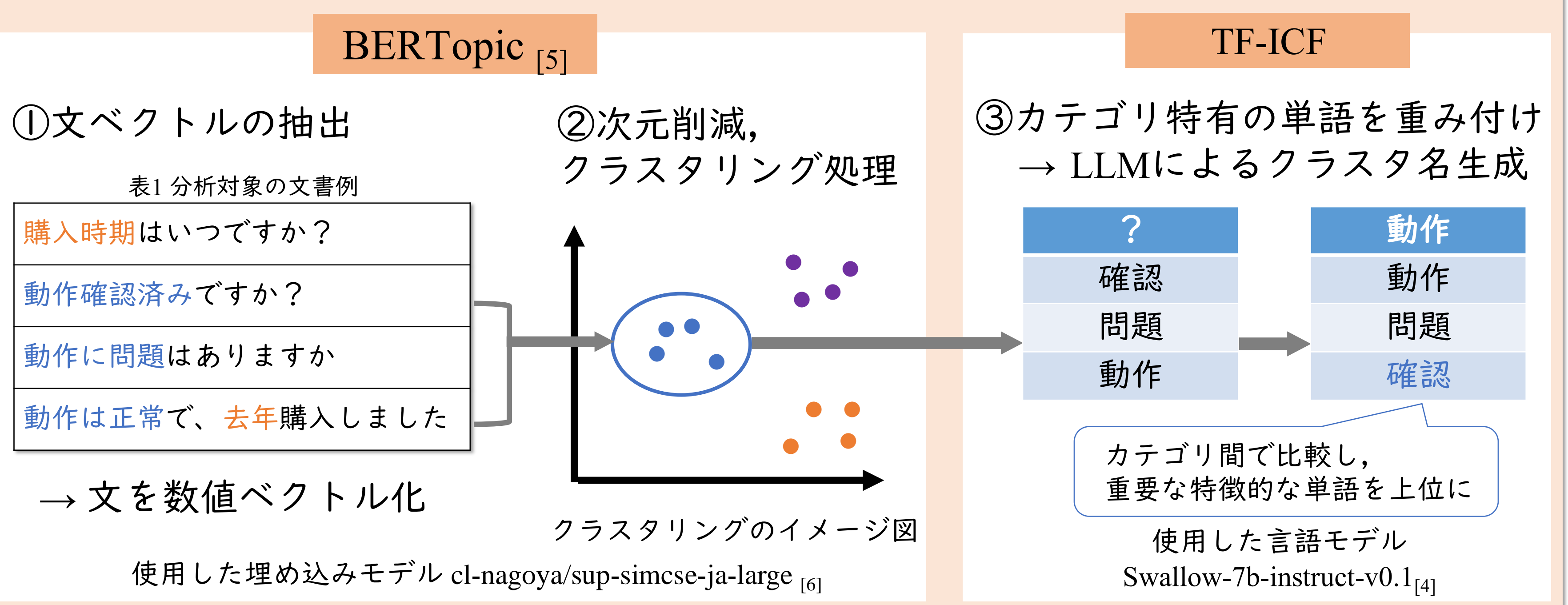
[3] NII, 情報学データリポジトリ, メルカリデータセット <https://www.nii.ac.jp/dsc/idr/mercari> (閲覧: 2024/12/06)  
 [4] Llama-3.1-Swallow-8B-Instruct-v0.1 <https://huggingface.co/tokyotech-llm/Llama-3.1-Swallow-8B-Instruct-v0.1>  
 [5] Grootendorst, M. (2022). BERTopic: Neural topic modeling with a class-based TF-IDF procedure. *arXiv preprint arXiv:2203.05794*.  
 [6] sup-simcse-ja-large <https://huggingface.co/cl-nagoya/sup-simcse-ja-large>

### 1. データの前処理

LLMによる文分類 → 商品自体とは関係のない文を取り除く (Few-shotを使用)

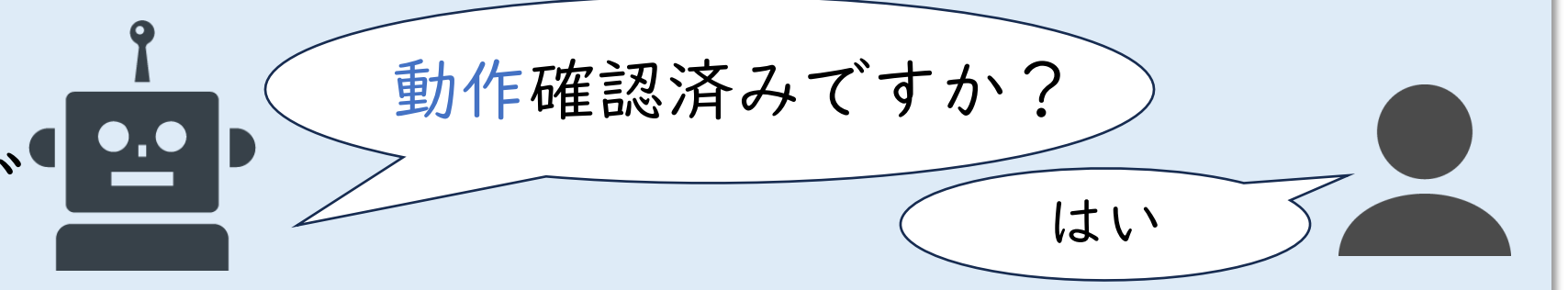


### 2. クラスタリングによる属性抽出



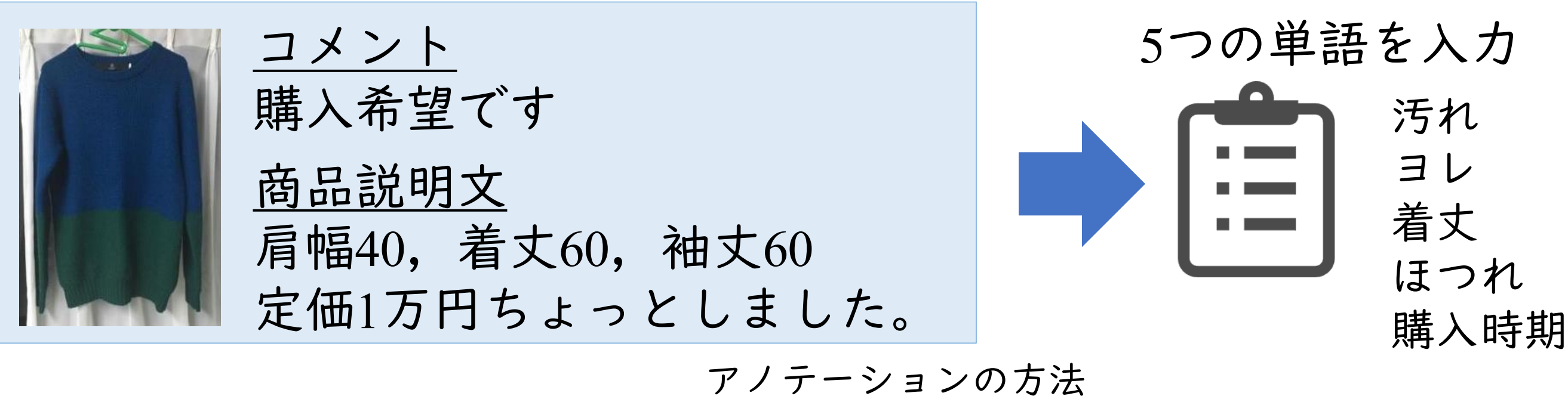
### 3. 対話システムの実装

抽出した属性を用いた属性フィリング



## 3. 評価実験

- **目的**
  - 提案手法で得られた単語が属性として相応しいかの評価
- **使用するデータセット**
  - 株式会社メルカリ提供データ<sup>[3]</sup>



### ● 評価方法

- 1カテゴリにつき2つの商品が無作為抽出し, 購買意思決定に必要な属性を手でアノテーションし, 正解データとして利用
- 抽出された単語が正解データ内に含まれているか確認 → 単語の類似度をもとに再現率, 適合率, F値を計算

### ● 比較対象

- LLMによる前処理なしのクラスタリング
- LLMによる属性出力(ChatGPT-4o) → Few-shotとchain-of-thoughtを使用したカテゴリ名から商品属性を出力させるプロンプト

## 4. 結果・考察

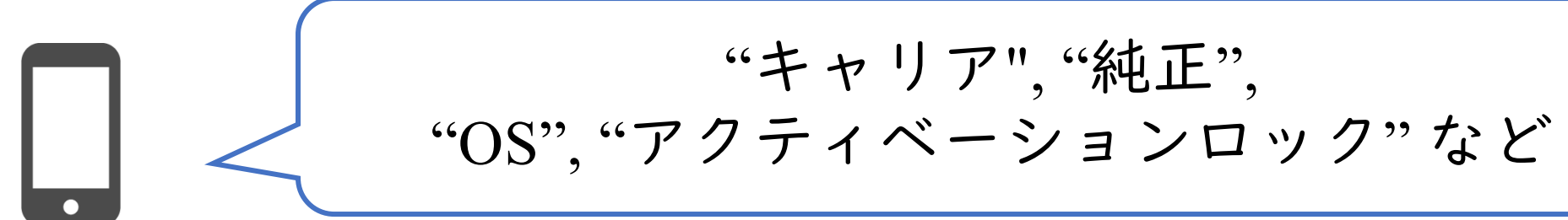
### ● 評価実験結果

|          | 再現率          | 適合率          | F値           |
|----------|--------------|--------------|--------------|
| LLM      | 0.426        | 0.422        | 0.422        |
| BERTopic | 0.425        | 0.360        | 0.389        |
| 提案手法     | <b>0.567</b> | <b>0.460</b> | <b>0.500</b> |

全ての評価項目についてLLM, BERTopicを**超える値**

### ● 具体的な属性の抽出例

一部カテゴリにおいて, その**カテゴリ特有の専門的な単語**が属性名として抽出される  
 <家電・スマホ・カメラ/スマートフォン/携帯電話>



➡ 人手では得られなかった専門的な属性の取得に成功

### ● 属性抽出結果

表2 カテゴリ『家電・スマホ・カメラ/スマートフォン/携帯電話』の例

|            | 属性1 | 属性2     | 属性3 | 属性4  | 属性5  |
|------------|-----|---------|-----|------|------|
| アノテーション    | 傷   | 色       | 汚れ  | 機種   | 画面割れ |
| LLM        | 傷   | バッテリー状態 | 汚れ  | 購入時期 | 付属品  |
| BERTopicのみ | 割引  | スマートフォン | 目標  | 希望   | 郵便   |
| 提案手法       | 傷   | バッテリー   | 動作  | 箱    | sim  |

➡ 専門的な用語を含む場合, それらを**具体的な属性**として提示可能

表3 カテゴリ『ベビー・キッズ/おもちゃ』の例

|            | 属性1 | 属性2  | 属性3  | 属性4   | 属性5 |
|------------|-----|------|------|-------|-----|
| アノテーション    | 色   | 保管方法 | 付属品  | 汚れ    | 匂い  |
| LLM        | 状態  | 素材   | 対象年齢 | 動作確認  | 欠品  |
| BERTopicのみ | 専用  | 価格   | メンバー | 配送    | 写真  |
| 提案手法       | 色   | サイズ  | 部品   | パッケージ | パーツ |

➡ カテゴリ内で商品の形態にばらつきがあると, 全体をカバーする抽出は困難

## 5. まとめ

### ● まとめ

- クラスタリングを用いた商品属性抽出手法を提案
- 評価実験の結果, 提案手法の優位性を確認
- 一部カテゴリにおいて, 人手でも得られなかった専門的な属性を得ることが可能

### ● 今後の課題

- カテゴリ内で商品の形態のばらつきがある場合の属性抽出への工夫
- 全カテゴリへの提案手法の適応
- 対話システムのユーザ評価

