

作詞者情報量がBERTに基づく作詞者 分類性能に与える影響についての評価

名古屋工業大学

森田光紀，菊地真人，大園忠親

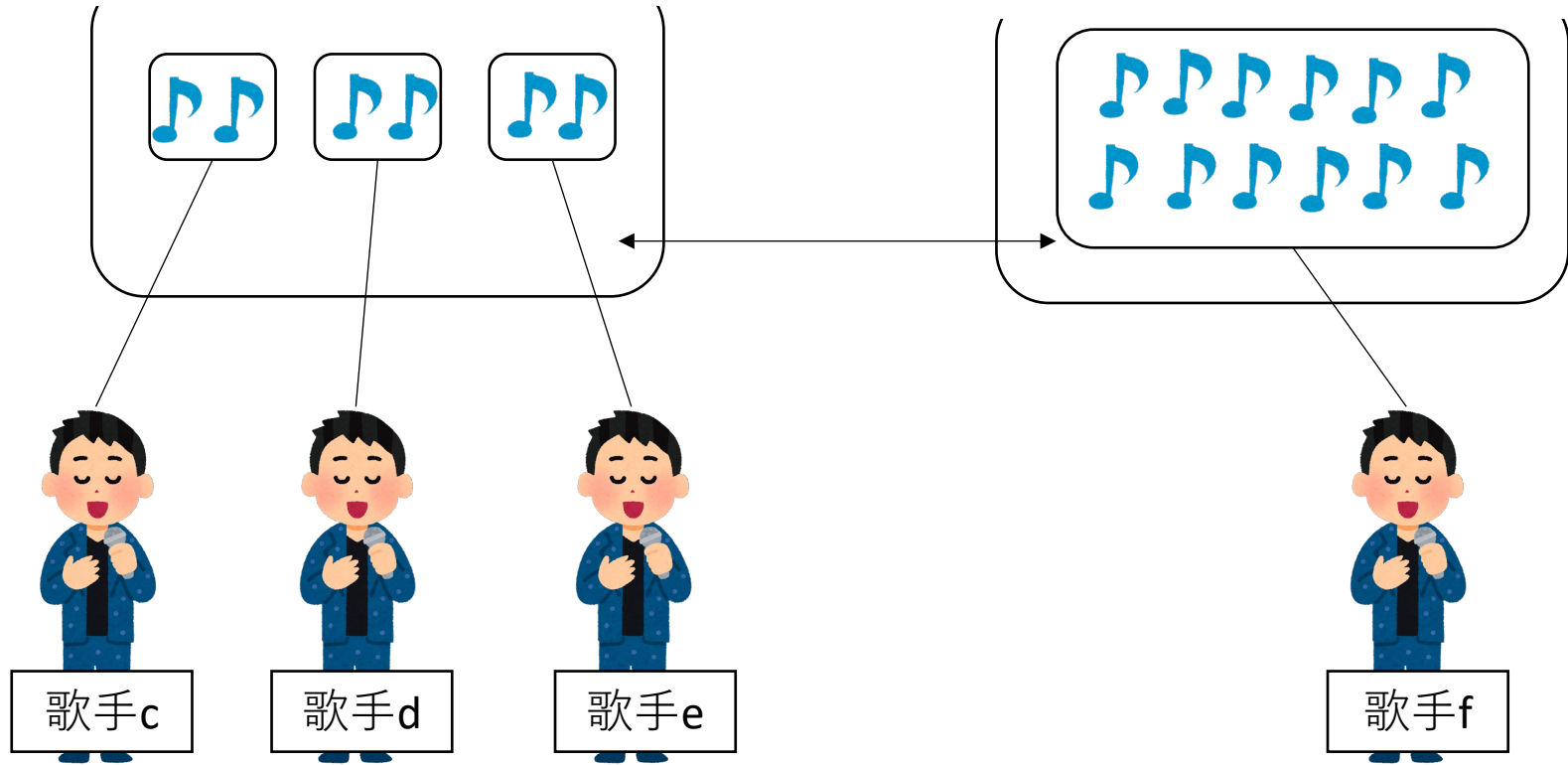
- 研究背景
 - 歌詞を作詞者へ分類する研究は少ない
 - 歌手と作詞者の個性が混在しているのでは
 - 歌詞は特殊な自然言語の文章である
 - 他人が発表するために作詞する
- 研究目的
 - 作詞者の楽曲提供の傾向が分類性能に与える影響を調べる

考慮する傾向

3

作詞者Aの楽曲

作詞者Bの楽曲



多数の歌手に
幅広く楽曲を提供

特定の歌手に
偏って楽曲を提供

作詞者情報量

4

- 作詞者*i*の作詞者情報量 H_i

$$H_i = - \sum_{j \in J_i} p_{ij} \log_2 p_{ij}, \quad p_{ij} = \frac{L_{ij}}{L_i}$$

歌手 $j = 1, 2, 3, \dots, m$, 歌手集合 $J_i = \{j | i \text{ が } j \text{ に 楽曲 提供}\}$

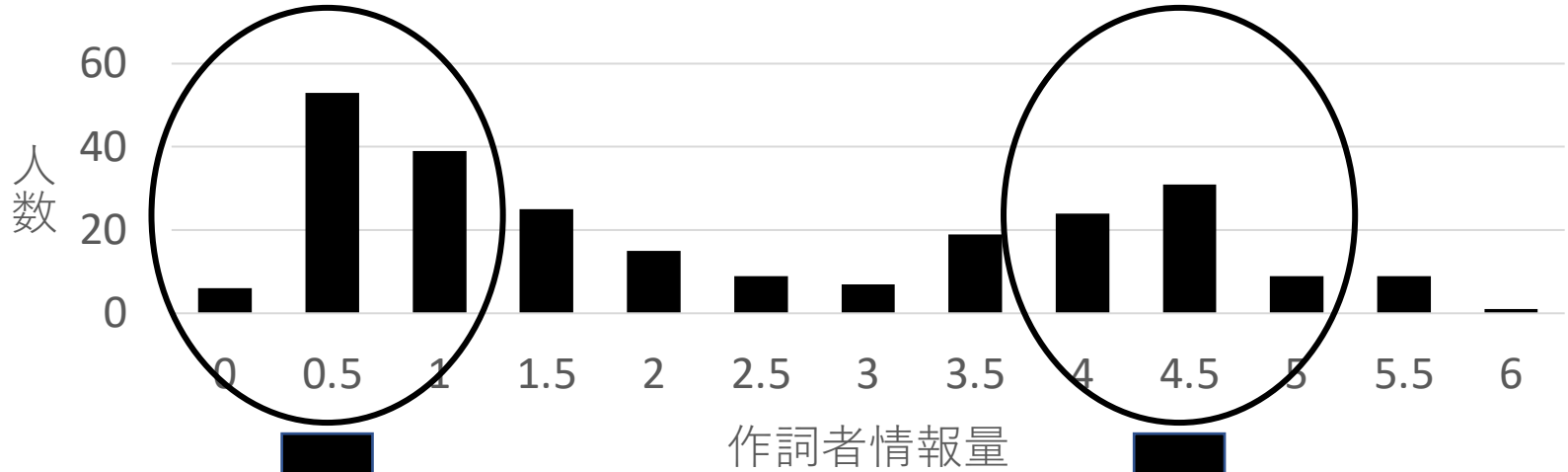
p_{ij} : 作詞者*i*が歌手*j*に楽曲提供している割合

L_i : 作詞者*i*の楽曲数

L_{ij} : 作詞者*i*が歌手*j*に提供した楽曲数

- $H_{i_1} < H_{i_2}$ のとき, 作詞者*i*₂は作詞者*i*₁より楽曲提供に多様性がある

仮説：作詞者情報量と分類性能 5



作詞者情報量が小さい
：分類が容易である

作詞者情報量が大きい
：分類が困難である

評価実験 概要

6

- 目的：
 - 作詞者情報量が分類性能に与える影響の調査
- 手順：
 1. 基本データベースの構築
 2. 作詞者情報量に基づく作詞者の選択
 3. 分類器を用いた分類性能評価

基本データベースの構築

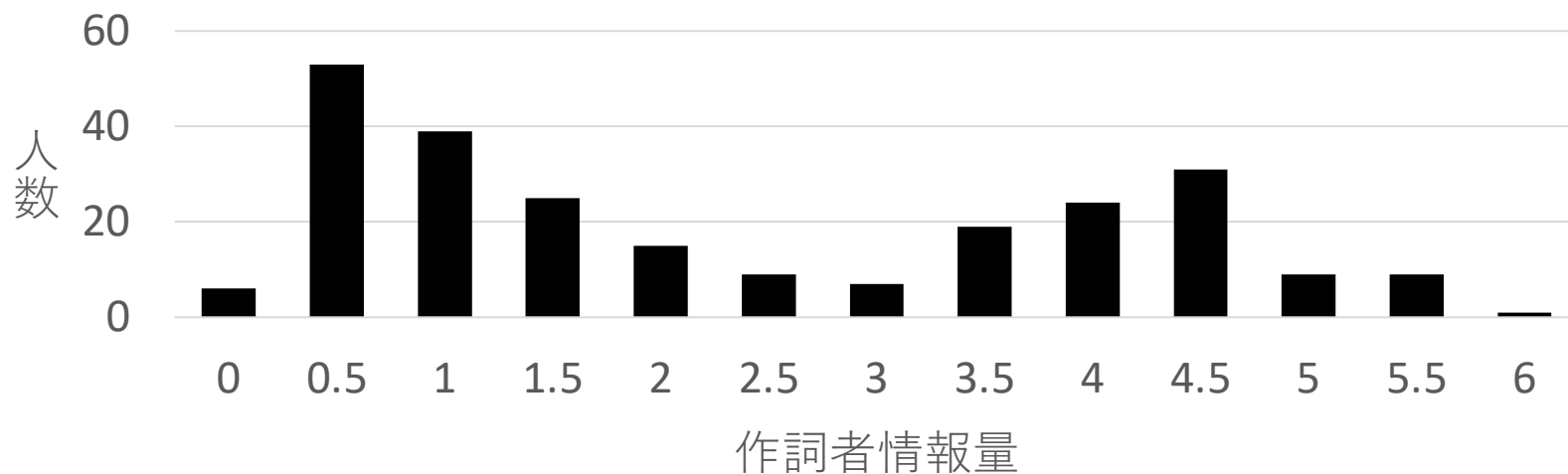
7

- 基本データベース
 - 曲名, 作詞者, 歌手
- データ源: 歌詞検索サービス歌ネット
<https://www.uta-net.com>
- 手順
 1. アーティスト総合ランキングから歌手を
列挙トップ100
 2. 1に含まれる歌手に楽曲を提供した作詞
者とその楽曲を列挙
 - カバー曲を除外
 3. 楽曲数が100曲以上の作詞者を選択

基本データベースの性質

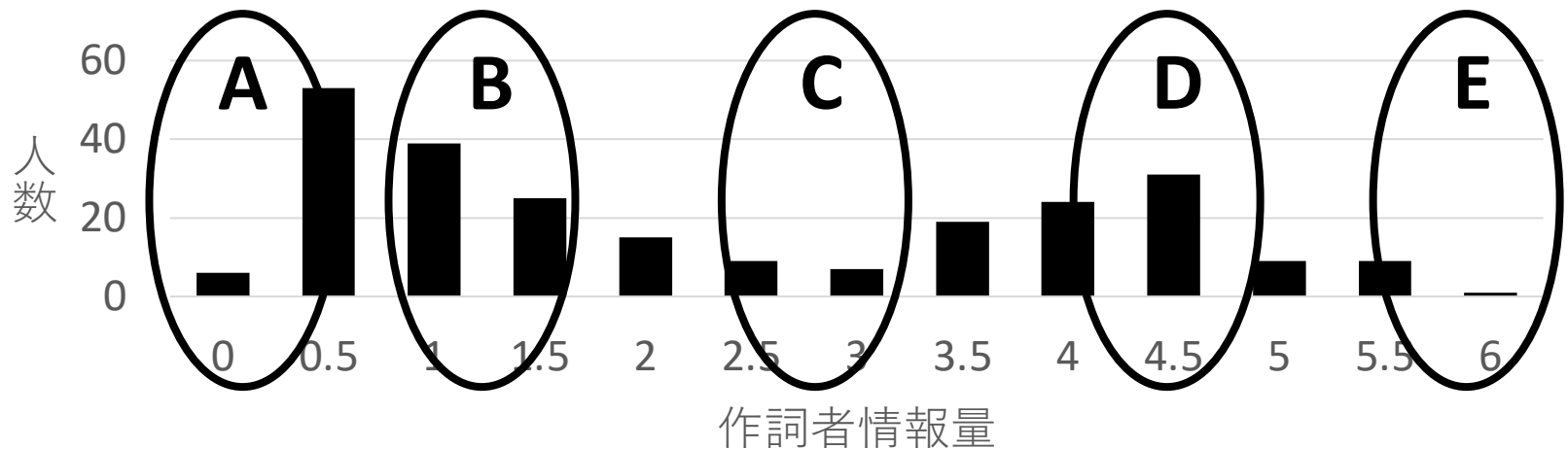
8

- 総作詞者数：247名
- 総楽曲数：96,330曲
- 歌手数：10,676名



作詞者の選択

9

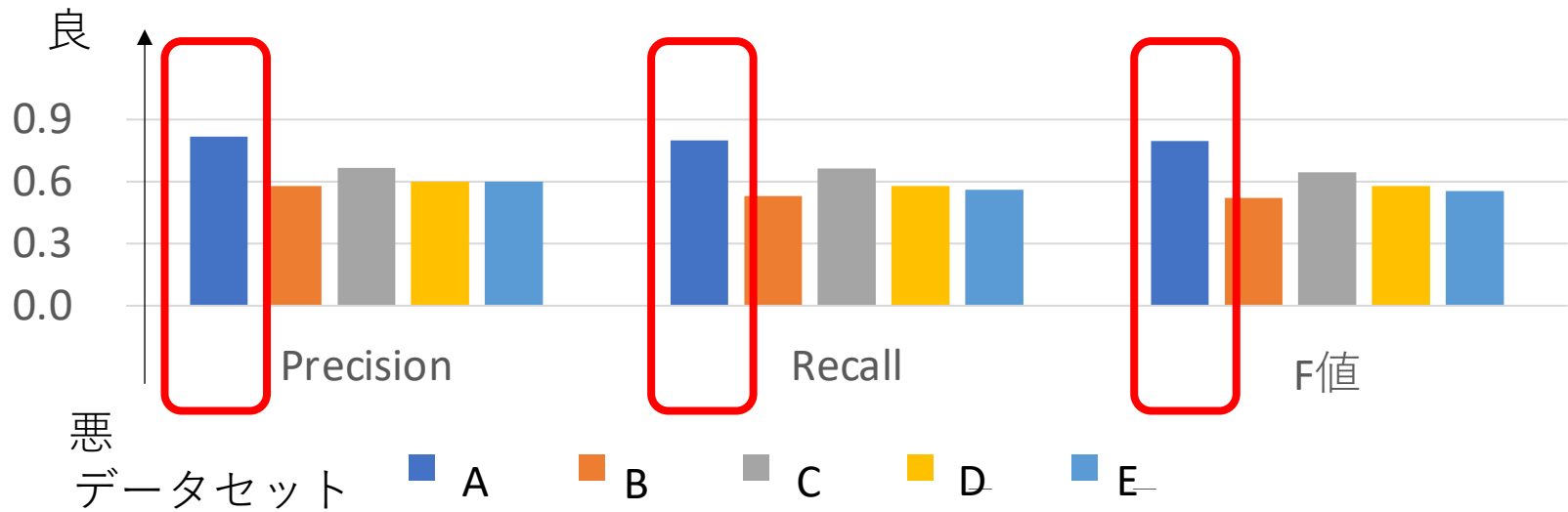


- 学習データセットの作詞者情報量の違いと作詞者分類における分類性能との関係を調査
 - A群:作詞者情報量の最下位10名
 - B,C,D群:作詞者情報量の $\frac{1}{4}$, $\frac{1}{2}$, $\frac{3}{4}$ の周辺10名
 - E群:作詞者情報量の最上位10名

- 作詞者分類器
 - 入力：歌詞全文
 - 出力：作詞者
- 学習データの構築
 1. 作詞者の集合を与える
 2. 各作詞者のランダムな50曲を選択
 3. 3:1:1の割合で訓練データセット，検証データセット，評価データセットに分割
- 学習データは以下のペアの集合
 - 歌詞全文，作詞者（one-hotベクトル）

評価実験結果

11



A群の分類性能が比較的良い

- A群：作詞者情報量の最下位10名
- B,C,D群：作詞者情報量の $\frac{1}{4}$, $\frac{1}{2}$, $\frac{3}{4}$ の周辺10名
- E群：作詞者情報量の最上位10名

- 仮説：作詞者情報量が小さいほど分類性能が良い
- 結果：
 - 作詞者情報量は分類性能に影響を及ぼした
 - 作詞者情報量が最下位群の場合のみ分類性能が良くなった

- 作詞者情報量が分類性能に影響を与える可能性がある
- 作詞者情報量を定義
- 基本データベースの作成
- 作詞者情報量と分類性能の関係を調べた
- 作詞者情報量と歌手分類の関係を調べる