

歌詞提供傾向を考慮した歌詞からの
作詞者分類に関する一考察

名古屋工業大学 B4

森田 光紀，菊地 真人，大冢 忠親

はじめに

2

- 研究背景

- 歌詞の特徴分析は，類似した作詞者推薦への応用が期待される
 - 特徴分析は困難である
 - 作詞者分類を高性能に行うことで特徴を捉えていることを示す

- 研究目的

- 作詞者の歌詞提供傾向が分類性能に与える影響を調べる
 - 作詞者情報量として定量化

歌詞提供傾向と作詞者情報量

3

作詞者Aの楽曲

例:清水依与吏
→back number



歌手c

作詞者Bの楽曲

例:つんく
→ モーニング娘。
Berryz工房
...



歌手d



歌手e



歌手f

	作詞者A	作詞者B
歌詞提供傾向	偏りが大きい	偏りが小さい
作詞者情報量	小さい	大きい

作詞者情報量

4

- 作詞者*i*の作詞者情報量 H_i

$$H_i = - \sum_{j \in J_i} p_{ij} \log_2 p_{ij}$$

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*に提供した割合 : $p_{ij} = \frac{L_{ij}}{L_i}$

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

評価実験

5

- 作詞者情報量と分類性能の関係を評価するための実験を行った
 - 評価実験 1 :
 - 作詞者群の作詞者情報量が上位と下位の作詞者の楽曲でデータセットを作成
 - 評価実験 2 :
 - 作詞者情報量ごとにデータセットを作成した
- 手順
 1. 基本データベースを構築
 2. 学習データセットを作成
 3. 分類性能を評価

基本データベースの構築

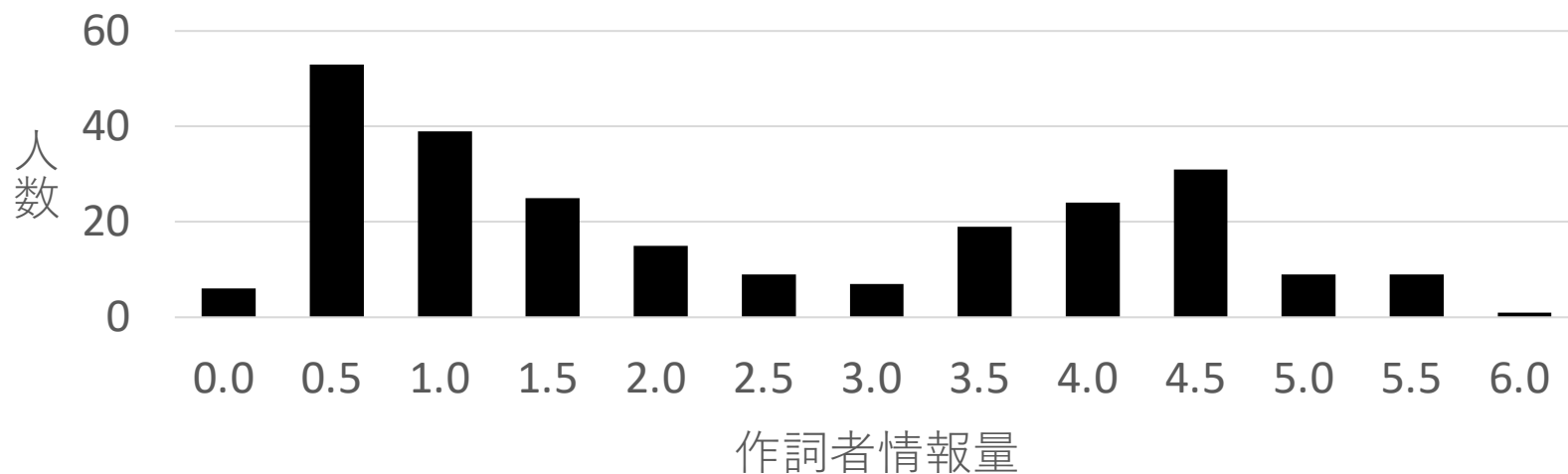
6

- 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

7

- 総作詞者数：247名
- 総楽曲数：56,107曲
- 歌手数：7,055名



作詞者分類器

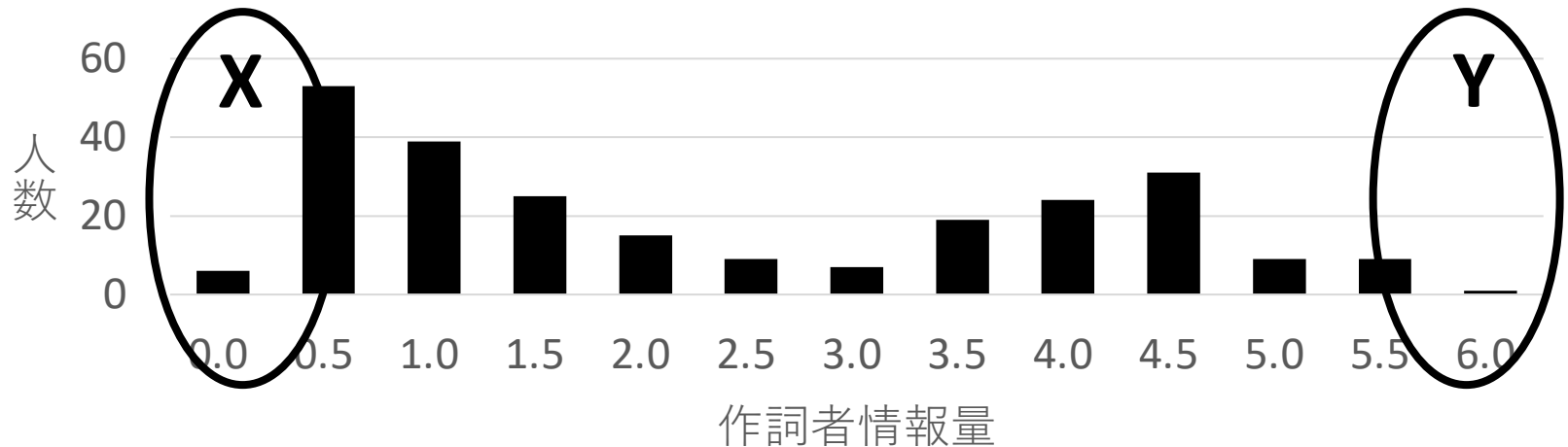
8

- 分類先の作詞者を設定して作成
 - 入力：歌詞全文
 - 出力：作詞者とスコア

- 学習データセット
 - 歌詞全文 + one-hotベクトル

評価実験 1

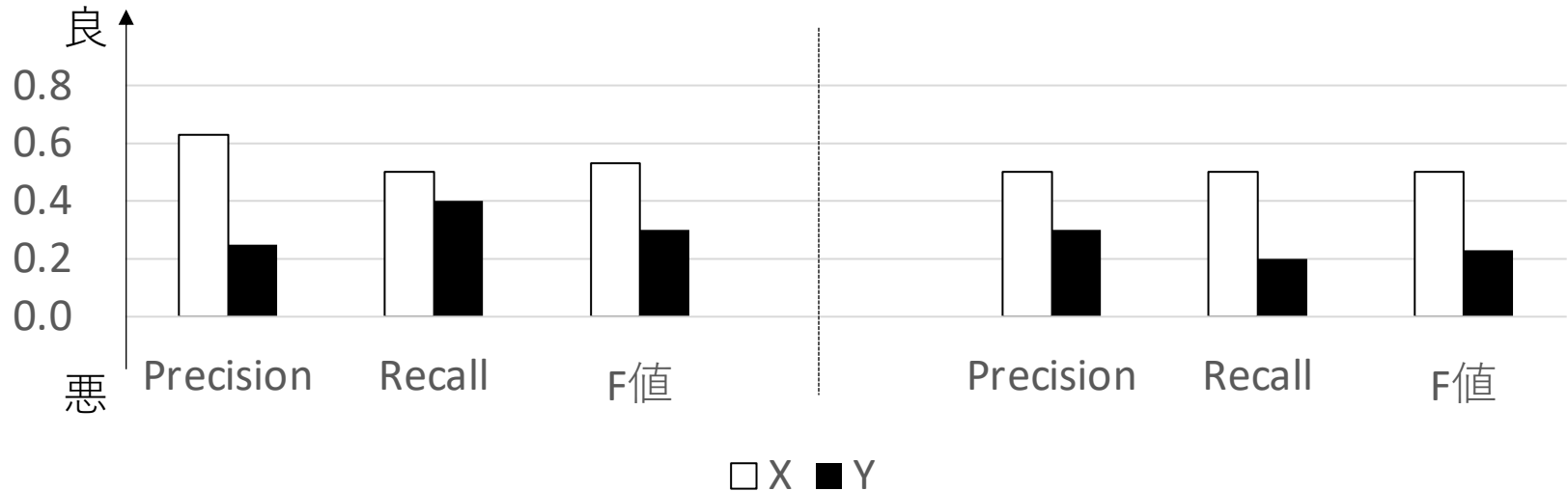
9



- X群：作詞者情報量の小さい20名
- Y群：作詞者情報量の大きい20名
- 以下の要素を変更してデータセットを作成
 - X群の作詞者とY群の作詞者を分割・混合
 - データセットに含む作詞者数や作詞者一人の楽曲数

評価実験 1 実験結果

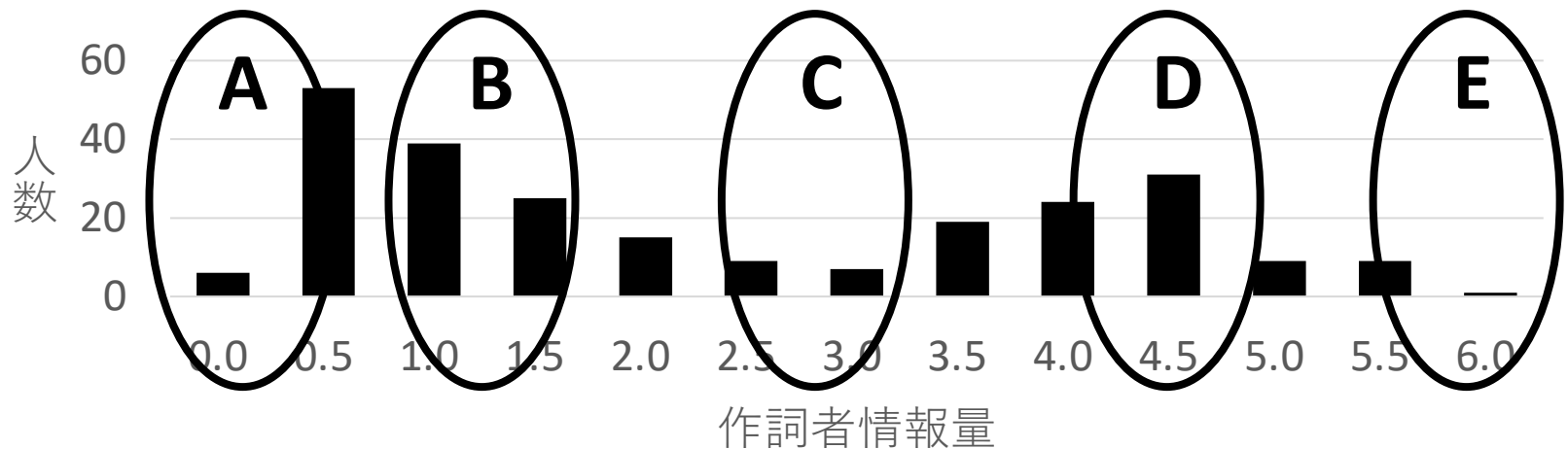
10



- いずれの場合もX群の作詞者の方が分類性能がよいという結果が得られた
 - X群：作詞者情報量の小さい20名
 - Y群：作詞者情報量の大きい20名

評価実験 2

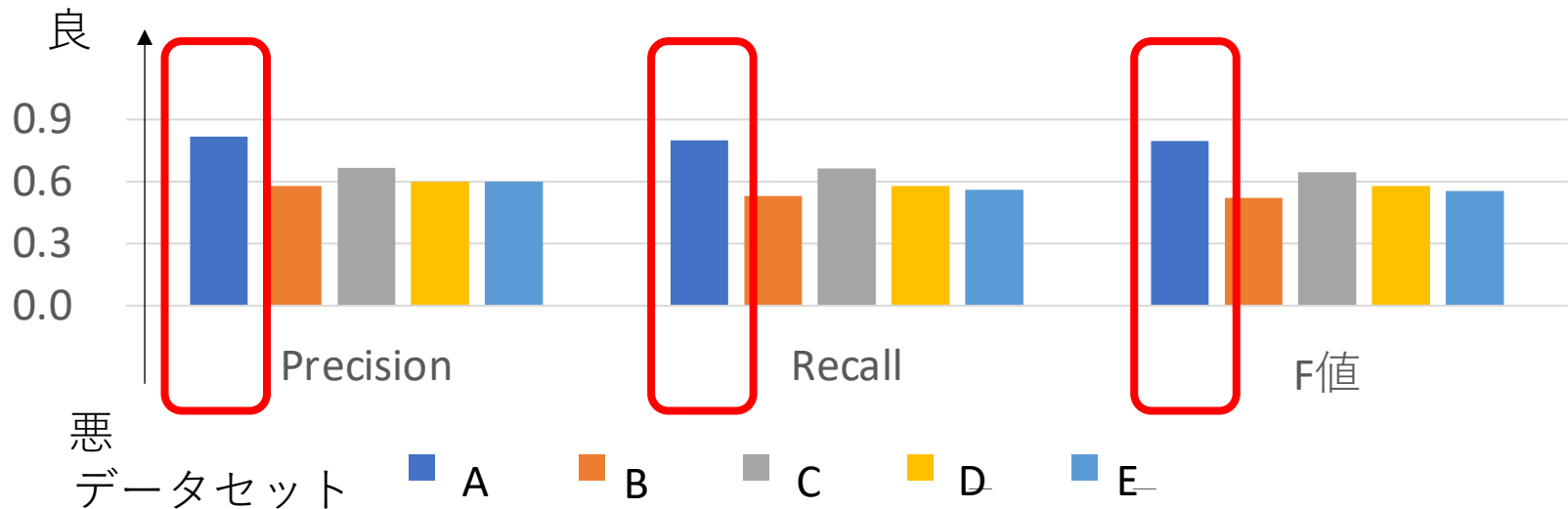
11



- 以下の作詞者についてデータセットを作成
 - 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

評価実験結果

12



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

- 歌詞提供傾向が極端に偏る場合のみ分類性能が良くなった
 - 歌手が一致している場合に分類性能が良くなる
- 歌詞の個性は，歌手によって決まる可能性がある
- 歌手がばらばらの場合も一定以上の性能を発揮した
 - 作詞者の個性が全くないわけではない

- 作詞者が楽曲を提供する傾向について考慮し、作詞者分類を行った
 - 歌詞提供傾向は分類性能に影響を与えていた
 - 分類性能が大きくなったのは、作詞者情報量が最小の場合のみ
- 歌詞の特徴は作詞者の経験ではなく、作詞者と歌手の個性による可能性
 - 歌手と作詞者の両方を考慮して分類を行う