

スパース学習に基づく万葉短歌の作者の特徴抽出

電気通信大学 大学院情報理工学研究科 川野 秀一
関西大学 文学部 村田 右富実

1. はじめに

万葉研究においては、歌の調子からそれぞれの歌作者の特徴を考えて来た歴史がある。しかし、歌内で多く使用されている（もしくはほとんど使用されていない）単一の音のみに着目した単変量的な解析，あるいは主観的な判断がほとんどであった。

本報告では、統計的手法に基づき複数の音を考慮に入れた多変量的な解析を行うことを考える。具体的には、各歌内で用いられている音から特徴ベクトルを作り出し、得られたデータをスパース判別手法 (Witten and Tibshirani, 2011) に基づき解析する方法について考察する。

2. データ化

万葉歌の中でも短歌 (字余り字足らずを含む) を対象とし、柿本人麻呂, 山上憶良, 大伴旅人の3歌人について考える。サンプルサイズはそれぞれ, 70, 55, 62であり, 総サンプルサイズは $n = 187$ である。特徴量は, 各短歌内で用いられている音素パターンとする。つまり, 「あ」から「を」までの音が一首にどれくらい含まれているかに着目する。これにより, 特徴ベクトルの次元は $p = 67$ となる。

3. 解析手法および結果

上で得られたデータセットに対してスパース判別手法を実行し, 作者の特徴抽出を試みる。図1は, スパース判別手法によりデータを2次元平面上へ射影したものである。図1aは作者を表し, 四角は柿本人麻呂, 丸は山上憶良, 三角は大伴旅人である。また, 図1bは歌番号を表している。得られた結果の詳細については当日報告する。

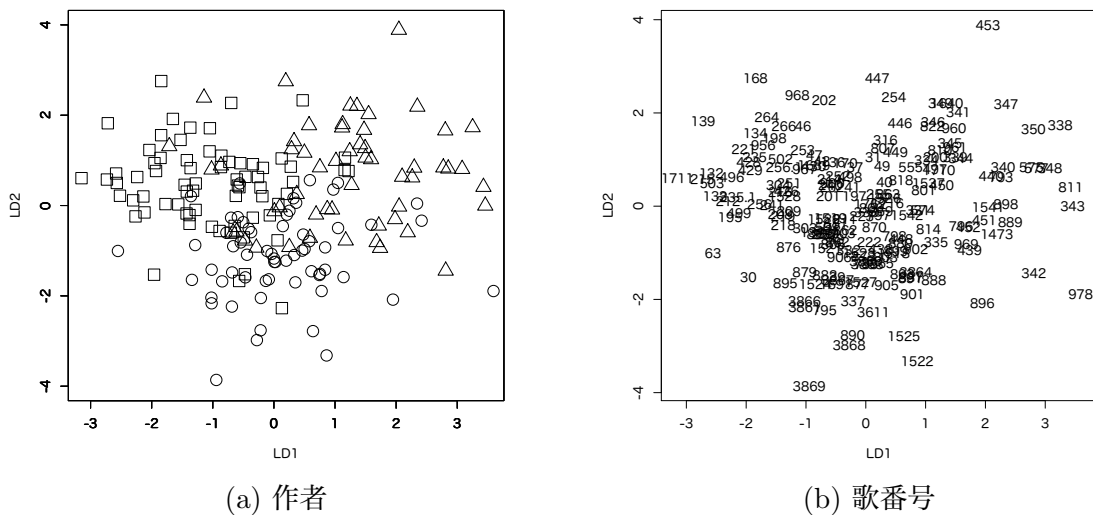


図 1: スパース判別手法によって2次元平面上へ射影した187個のデータ。

参考文献

Witten, D. M. and Tibshirani, R. (2011) Penalized classification using Fisher's linear discriminant. *J. R. Statist. Soc. Ser. B*, **73**, 753–772.