

研究ノート

# フィラーの頻度を特徴量としたランダム フォレストによる日本語学習者の習熟度 の推定<sup>1</sup>

石山友之

## 要旨：

本研究では機械学習の一手法であるランダムフォレストによる日本語学習者の習熟度の推定を試みた。学習者コーパス I-JAS の中から 327 名の日本語学習者の発話データを使用し、10 種のフィラーの頻度を特徴量として習熟度の推定を行ったところ、判別率は 71.7% であった。誤判別された学習者は「習熟度が低い段階では母音型、特にア類が多く、習熟度が高くなると語彙型、特にアノ類が多くなる」という全体的な日本語学習者の傾向から外れている学習者であった。本研究はフィラーの頻度を特徴量とした機械学習による習熟度の推定がある程度可能であることを示しており、更なる精度の向上のために習熟度の判別に有効なその他の特徴量を今後検討していく必要がある。

**キーワード：**フィラー、ランダムフォレスト、日本語学習者、習熟度推定

## Abstract：

In this study, we attempted to estimate the proficiency level of learners of Japanese using random forests, a machine learning method. Based on the speech data of 327 learners of Japanese in the corpus, we measured the frequency of 10 types of fillers and used them as explanatory variables for random forests, resulting in a 71.7% accuracy rate in estimating proficiency. Misclassified learners deviated from the overall tendency to use fillers, which is that “at the low proficiency level, there are more

‘A’ type, and at the high proficiency level, there are more ‘ANO’ type.” This study shows that machine learning can estimate proficiency to some extent using filler frequencies.

**Keywords:** filler, random forests, learners of Japanese, estimating proficiency