Vowel classification using machine learning: Random forest analysis

Daiki Hashimoto (the University of Tokyo)

Artificial Intelligence (AI) plays an important role in our society these days. Machine learning is a part of AI, and it is used in a variety of contexts. For example, you may check your smartphone when you want to go to a station that you have never been to before. Using a smartphone teaches you the way to get to the station quickly or cheaply. Here, machine learning comes in. A statistical model is constructed on the basis of a tremendous amount of information with regards to transportation, and the model returns the best way to the station with regards to time or cost.

As this technology attracts attention from a larger number people ever, linguistic studies also start using a technology such as regression analyses and classification analyses. <u>The aim of this study is to apply a statistical analysis called random forest (Breiman 2001) to the classification of vowels in Japanese.</u> Random forest has been developed to overcome main flaws of classification trees such as instability and noise susceptibility. More specifically, it builds multiple classification trees with some restrictions on splitting variables, and combines the prediction of the set of trees by taking the majority of votes (see Stroble et al. 2009).

The data come from Corpus of Spontaneous Japanese (CSJ) developed by Maekawa et al. (2000). Using a Praat script (Boersma and Weenink 2019) and an R script (R Core Team 2019), we explored academic speech by randomly selected ten speakers (five male speakers and five female speakers), and every short vowel token in the speech was measured with regards to their phonetic values such as intensity, duration, and formant values. Devoiced tokens were removed from our dataset. Phonetic values around 50% points of each vowel were measured. In all, our dataset includes 31,967 short vowel tokens. Random forest analyses were performed on this data set using the package ranger (Wright and Ziegler 2017).

The current study demonstrates that random forest analyses can successfully capture a variety of vowel realizations in the corpus, and classify them into the five phonemes /a, e, i, o, u/. This demonstration tells us <u>(1) that a random forest model needs to learn about one third of the whole dataset in order to deduce robust prediction, and (2) that even building ten trees results in a robust result (about 78% accuracy).</u>

<u>References:</u>

Boersma, Paul, & David Weenink. (2019). Praat: Doing phonetics by computer (version 6.0.49). www.praat.org.

Breiman, Leo. (2001). Random forests. *Machine Learning*, 45. 5-32.

Maekawa, K., Koiso, H., Furui, S., & Isahara, H. (2000). Spontaneous speech corpus of Japanese. Paper presented at the Proceedings of LREC2000 (Second International Conference on Language Resources and Evaluation).

Strobl, Carolin, James Malley, and Gerhard Tutz. (2009). An introduction to recursive partitioning: Rationale, application, and characteristics of classification and regression trees, bagging, and random forests. *Psychological methods*, 14. 323-348.

R Core Team. (2019). R (version 3.5.3). https://www.r-project.org/

Wright, Marvin, N., & Andreas Ziegler. (2017). ranger: A Fast Implementation of Random Forests for High Dimensional Data in C++ and R. *Journal of Statistical Software*, 77(1), 1-17. doi:10.18637/jss.v077.i01