NKULULEKO: A TEMPLATE BASED SYSTEM FOR FAST MACHINE LEARNING EXPERIMENTS ON SPEAKER CHARACTERISTICS

Felix Burkhardt

[1]*audEERING GmbH, Germany,* [2]*Technical University of Berlin, Germany*
fburkhardt@audeering.com

Abstract: We present Nkululeko, a template based system that lets users perform machine learning experiments in the speaker characteristics domain. It is based on audformat, a format for speech database metadata description.

1 Background

In the past decades, the research community has been confronted with the tremendous success of approaches to estimate knowledge with artificial neural nets (ANN), predominantly under the label *deep learning*. Especially empirical sciences benefit from the opportunity to test hypotheses with machine learning experiments that are able to analyse statistically very large data quantities.

Many empirical researchers, phoneticians, and linguists, did not study computer science and struggle with the necessary programming skills to set machine learning experiments up.

Nkululeko[1] is being developed preliminary as a tool for a series of machine learning seminars at the institute for speech communication at the Technical University of Berlin to enable students to conduct machine learning experiments with a very flat learning curve by simply filling configuration files. This makes it much easier to use than other high level frameworks for deep learning like Keras, Torch, Google AutoML or end2you [1, 2, 3, 4] while still keeping the flexibility as it is based on Torch and Keras.

2 Method

Nkululeko is open source software written in Python and hosted on github[2]. The data management is based on audformat[3], but a simpler CSV formalism is also supported.

The paper will explain the format of the configuration file[4], give a brief overview on audformat and present case examples of typical experiment setups: comparing the performance of several acoustic features with pre-trained ANN embeddings with a set of typical machine classifiers for speaker emotion (as classification problem) and speaker age (as regression problem).

In Figure 1, an overview on the Nkululeko architecture is given. It consists of a central "experiment" class which can combine a set of acoustic features with machine learning classifiers and regressors.

Figure 2 shows a confusion matrix and the evaluation per epoch for a typical Nkululeko experiment: investigating the performance of ANN embeddings in a cross database experiment for acted basic emotional vocal expressions.

---

[1]On the lookout for a distinctive name for this project we stumbled across an 1980ies punk album title. They tried new things out fastly, so this seemed fitting.

[2]https://github.com/felixbur/nkululeko/

[3]https://github.com/audeering/audformat

[4]https://github.com/felixbur/nkululeko/blob/main/ini_file.md

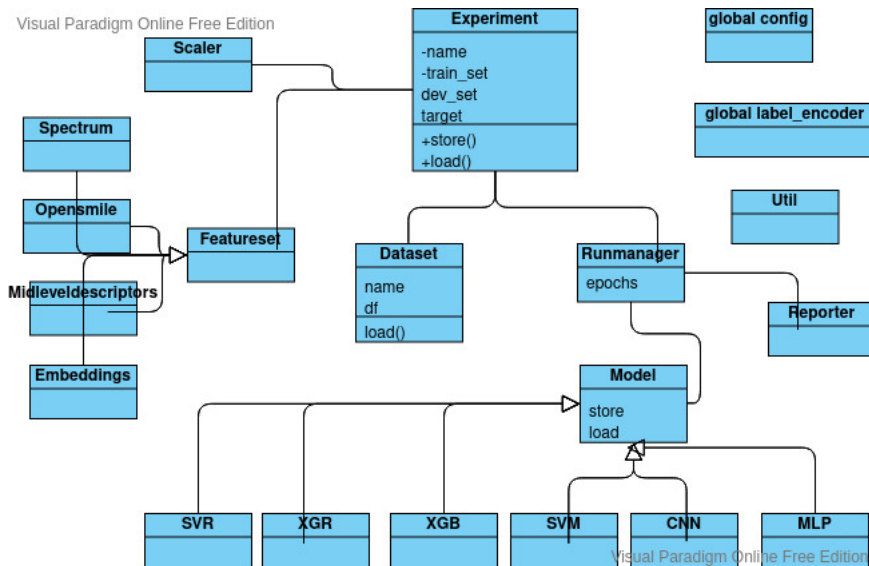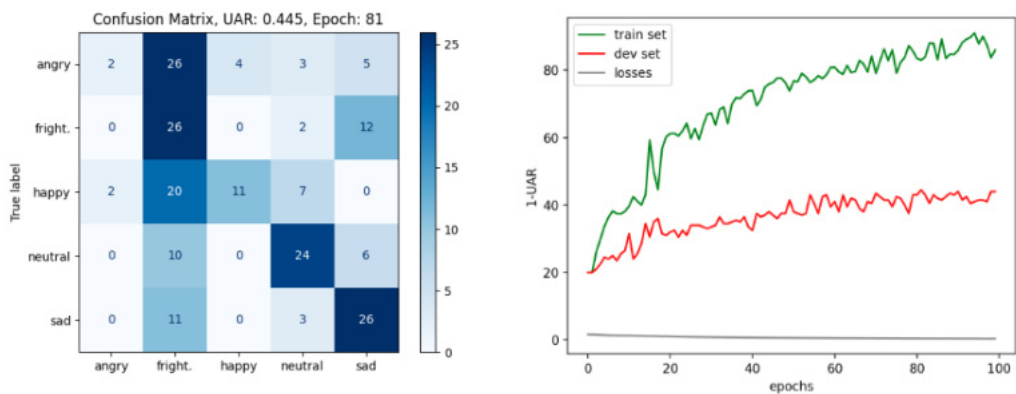**Figure 1** – Class architecture of Nkululeko

Figure 2 – Confusion matrix and performance per epoch for a typical Nkululeko experiment

3   Conclusions

We presented Nkululeko – a free for research new open-source tool to set up machine learning experiments in the speech research domain that can be used without programming skills. Future works will include extension of its functionality.

References

[1] CHOLLET, F.: *Deep learning with python keras*. 2017.

[2] CHAUDHARY, A., K. S. CHOUHAN, J. GAJRANI, and B. SHARMA: *Deep learning with pytorch*. 2020. doi:10.4018/978-1-7998-3095-5.ch003.

[3] TZIRAKIS, P., S. ZAFEIRIOU, and B. W. SCHULLER: *End2you–the imperial toolkit for multimodal profiling by end-to-end learning. arXiv preprint arXiv:1802.01115*, 2018.

[4] BISONG, E.: *Google AutoML: Cloud Vision*, pp. 581–598. 2019. doi:10.1007/978-1-4842-4470-8_42.