HHAI2022: Augmenting Human Intellect
S. Schlobach et al. (Eds.)

doi:10.3233/FAIA220225

Annotating Sound Events Through Interactive Design of Interpretable Features

Thiago S. GOUVÊA [a], Ilira TROSHANI [a,b], Marc HERRLICH [a], Daniel SONNTAG [a,b]

[a] *Interactive Machine Learning, DFKI, Germany*
[b] *Applied Artificial Intelligence, Oldenburg University, Germany*

1. Introduction

Professionals of all domains of expertise expect to take part in the benefits of the machine learning (ML) revolution, but realisation is often slowed down by lack of familiarity with ML concepts and tools, as well as low availability of annotated data for supervised methods. Inspired by the problem of assessing the impact of human-generated activity on marine ecosystems through passive acoustic monitoring [1], we are developing Seadash, an interactive tool for event detection and classification in multivariate time series.

2. Concept and Implementation

Seadash is a Python-based application currently implemented with Dash [2]. Its core concept is to offer the user graphical tools—visualisations and input controls—for rapid, iterative design of interpretable [3] data transformations (*features*) that can be used by downstream tasks (e.g., manual annotation, automatic event detection). In the envisioned workflow, a spectrogram is computed once the user loads an audio file. This least-processed (i.e., most faithful to raw) data visualisation is a central UI element, and as such is constantly displayed. Next, the user creates one or more *entities* (e.g., dolphins)—the actors within the phenomenon captured by the data responsible for generating each of the event types. Dedicated input controls are then used to design features that capture events (e.g., dolphin calls) associated with that entity. These controls generate features by applying signal processing (e.g., frequency band selection, Gaussian smoothing) and unsupervised ML (e.g., PCA, embedding in deep autoencoder networks) operations to the spectrogram. The signal processing and ML elements are implemented with SciPy, scikit-learn, and TensorFlow. Features are computed across frequencies, and the temporal dimension is kept unchanged; hence, features can be displayed in alignment with the spectrogram, and their juxtaposition facilitates quick visual curation. When an informative feature is identified, data can be efficiently annotated by setting a threshold on that feature's activation level so as to bring out event occurrences. Such procedure constitutes a graphical implementation of *data programming* [4], a data annotation paradigm that promises higher efficiency compared to classical methods based on manually marking individual event occurrences (e.g., [5,6]). Once data has been annotated, even if partially, supervised ML operations become available as an option for feature design (figure 1).

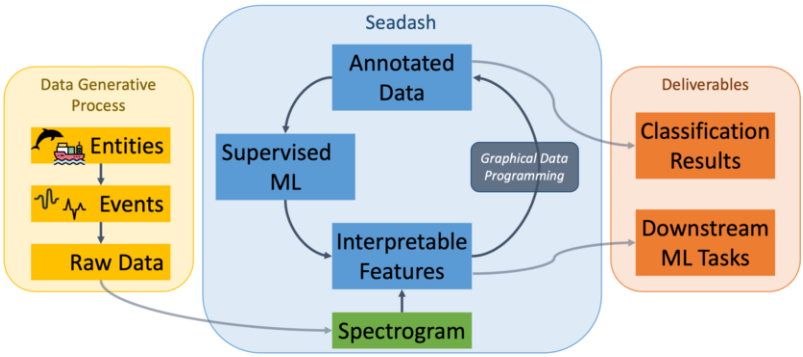

Figure 1. In Seadash, the domain expert-user builds interpretable features from which sound events produced by identifiable entities can be recovered. The data thus embedded can be used in downstream tasks such as annotation through graphical data programming.

3. Conclusion and Future Work

In this poster paper, we present Seadash, an interactive tool for data annotation through interactive, iterative, and graphical design of interpretable features. Building on the concept of data programming [4] and extending it to the graphical domain, Seadash already allows domain experts to efficiently annotate audio datasets. While the tool is inspired by marine acoustic monitoring in the context of ocean research, it will find applications in event detection and classification more generally, both in acoustic (e.g., speaker identity in recorded dialogues) and non-acoustic recordings (e.g., multisensor human-computer interfaces [5,7]). We are working on imbuing it with AI capabilities to 1) improve efficiency of feature design by making use of classification results and 2) output an edge ready inference model for automatic event detection and classification. As a future step, we will systematically study domain experts' experiences and usage.

Acknowledgement TSG acknowledges Gustavo B. M. Mello (OsloMet University, Norway), Luca Tassara (Akvaplan-niva, Norway), and Ehsan Abdi (Cyprus Subsea Consulting and Services) for inspiring ideas and use cases.

References

[1] Duarte CM, Chapuis L, Collin SP, Costa DP, Devassy RP, Eguiluz VM, et al. The soundscape of the Anthropocene ocean. Science. 2021;371(6529):eaba4658.
[2] Dash. Plotly; 2022. Available from: https://dash.plotly.com/.
[3] Rudin C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nature Machine Intelligence. 2019 May;1(5):206-15.
[4] Ratner AJ, De Sa CM, Wu S, Selsam D, Ré C. Data programming: Creating large training sets, quickly. Advances in neural information processing systems. 2016;29.
[5] Barz M, Moniri MM, Weber M, Sonntag D. Multimodal multisensor activity annotation tool. In: Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing, UbiComp Adjunct 2016, Heidelberg, Germany, September 12-16, 2016. ACM; 2016. p. 17-20.
[6] Heartex. Label Studio; 2022. Available from: https://github.com/heartexlabs/label-studio.
[7] Oviatt S, Schuller B, Cohen PR, Sonntag D, Potamianos G, Krüger A, editors. The Handbook of Multimodal-Multisensor Interfaces: Signal Processing, Architectures, and Detection of Emotion and Cognition - Volume 2. vol. 21. Association for Computing Machinery and Morgan amp; Claypool; 2018.