Piper Macke

DATS 2101

Identifying and Compiling the Data

Audioset is a data archive consisting of hundreds of sound classifications and a developing ontology for ten second YouTube audio clips. It was created by a team of Google researchers by individually listening to the clips and attributing each of them to a sound category.

The researchers compiling this data work on the Sound and Video Understanding teams at Google and are actively using Audioset in machine perception research. They are also currently revising the dataset by rerating any clips that have high confusability. As a result, this human-verified dataset with over two million YouTube audio clips may easily be used alongside advanced machine learning models to construct artificial intelligence with the purpose of categorizing and creating audio on its own.

Organization

The organization of Audioset is laid out with four main criteria: the unique YouTube video ID, the timestamp of the start and end of the segment, and the alphanumeric code that corresponds to the type of audio identified. One video may be attributed to several different sound types. For example, an audio with a man speaking may be categorized into "Human Sounds," "Speech," and "Speech - Man Speaking." The branches of each larger sound category are what develop the ontology and relationships between the data.

Currently, the team of researchers is entirely responsible for categorizing the clips into their respective groups. The potential for human error to affect the classification process of the clips is a contextual facet of the data that must be considered. In Staffan Müller-Wille's *Making and Unmaking*

Populations, he insists on the importance of understanding the complexities that come with classification in data collection. Müller-Wille examined two case studies involving meticulous data collection, both of which support his claim, "statistics was thus not simply concerned with numbers, but with numbers relating to things classified," (p. 605). With Audioset's most likely use being for machine learning, it is important to consider that any wrongful classification of a sound could significantly distort a program that bases everything it knows on how its example data has been categorized.

Purpose and Questions

Audioset could be used to answer practical questions like how common human sounds are in YouTube videos or what animal sound is the most accurately detectable, but its purpose extends beyond just data visualization and analysis. Audioset is not designed to answer a data-reliant question, but to act as an archive of human-classified sounds to later be used as a means for machine learning and audio detection.

In this case, many new questions arise about the capabilities of a machine that uses Audioset as its sole foundation. In Daniel Rosenberg's *Data as Word*, he explains how the word "data" was derived from the Latin verb meaning "to give," supporting how data was always understood as given fact (p. 559). While people are now more educated on the underlying complexities of data collection, machines are not. Artificial intelligence created on faulty data will only be as accurate as the data it is being fed. Perhaps this means that an entirely accurate AI model could never exist, or that it could only exist without any human intervention. These are the questions researchers like those working with Audioset seek to answer.

Obvious, Natural, and Ethical Uses

Any dataset created with the intention to be used in machine learning means it will ultimately be used with some type of algorithm. While navigating the ethics of AI audio recognition is a challenging process, there are plenty of algorithms that already exist with purposes as innocent as recognizing bird calls. For example, the Merlin Bird ID mobile application allows a user to identify any bird they see by one of four ways: answering a series of questions, uploading a photo, recording a singing bird, or looking at birds native to the region. Merlin Bird ID's description explains that a bird is identified by deep learning algorithms that reference "training sets" of millions of photos and sounds collected by "birders" at eBird.org. In other words, the data that allows the app to function are collected and designed exactly as Audioset is, with the same intention – only at a much larger scale.

An important aspect of these types of datasets is that they are to be used as a starting point. As Soraya de Chadarevian and Theodore Porter examine in their piece *Scrutinizing the Data World,* the function of machine learning algorithms is "to nudge future behaviors on the basis of all the numbers generated by previous actions and choices," (p. 550). Essentially, the training data fed to machine learning models are what generate their functionality. Audioset, along with countless other datasets that exist with the same purpose, are most likely to be used in this fashion.

Modifying Obvious Uses

With any type of AI model, malicious usage must be understood as a possibility. A program developed from training data like Audioset has the potential to digitally produce sounds that are indistinguishable from real ones. A problem that already exists as a result of recent AI advancement is deepfake content, or images, video, and audio that is either created originally or is made to imitate a real person with striking similarity. For instance, a recent deepfake video of Elon Musk describing his new investment platform gained considerable attention online, with Musk saying in the video

that it "will help users increase their income exponentially, without active participation" (Collier, 2023). Of course, every thing and person in the video is AI generated, and the application does not actually exist. The most concerning part about deepfake content is how easy it is to make, find, and most importantly, believe.

Considerations for a Code of Conduct

There is a possibility that, as data like Audioset fosters the creation of more advanced AI, more laws arise concerning copyright protection, ownership requirements, or maliciously using one's image and likeness. Already, victims of deepfake scams have warned about receiving distressing or fraudulent calls using another person's voice, sounding exactly like a family member, close friend, or business partner (Flitter and Cowley, 2023).

Modern data science is more reliant now than ever on training data like Audioset and its potential to be used in advanced machine learning algorithms. It is crucial to remember that human-like AI exists because of its millions of underlying data points that direct its behavior. Truly understanding this data is the first step in contributing to the modern artificial intelligence revolution.

References

Chadarevian, S., & Porter , T. (2018). Introduction: Scrutinizing the Data World. Historical Studies in the Natural Sciences, 48(5), 549–556. https://doi.org/https://doi.org/10.1525/hsns.2018.48.5.549

Collier, K. (2023, August 29). *Deepfake scams have arrived: Fake videos spread on Facebook, Tiktok and YouTube*. NBCNews.com. https://www.nbcnews.com/tech/tech-news/deepfake-scams-arrived-fake-videos-spread-facebook-tiktok-youtube-rcna101415

Cornell University. (2013, December 11). *Merlin Bird ID by Cornell Lab*. App Store. https://apps.apple.com/us/app/merlin-bird-id-by-cornell-lab/id773457673

Flitter, E., & Cowley, S. (2023, August 30). *Voice deepfakes are coming for your bank balance*. The New York Times. https://www.nytimes.com/2023/08/30/business/voice-deepfakes-bank-scams.html

Google. (n.d.). AudioSet. Google. https://research.google.com/audioset/index.html\

Müller-Wille, S. (2018). Making and Unmaking Populations. Historical Studies in the Natural Sciences, 48(5), 604–615. https://doi.org/10.1525/hsns.2018.48.5.604

Rosenberg, D. (2018). Data as Word. *Historical Studies in the Natural Sciences*, *48*(5), 557–567. https://doi.org/10.1525/hsns.2018.48.5.557