Making Sense of Sounds: Can machines learn to categorise sounds like humans?

Most of us take it for granted that we can automatically categorise everyday sounds. For example, we may instantly recognise the sound of running water, or a piano, or an engine. Enthusiasts might even be able to tell you what type of vehicle and which model features that engine. However, giving computers the ability to accurately classify specific types of sound data is a complicated and cross-disciplinary problem. The Making Sense of Sounds (MSoS) project was set up to address this, and is currently underway in the university's Acoustics Research Group, in partnership with the Centre For Vision, Speech, and Signal Processing at the University of Surrey. With their knowledge of machine learning, and our expertise in acoustics, we are working together to develop systems that may be useful in many real-world applications. These include automatically generating metadata for effective audio archiving, and triggering actions within intelligent home networks.

One of the key project outputs from the Salford side has been developing a collection of audio files that has been verified for quality, and accurately assigned a label that describes the type of sound. Each file was first assigned a category, such as 'Human', and then a more specific event type, such as 'Laughing'. This was a very time-consuming process which required specialised skills such as psychoacoustics and audio signal processing. Sourcing robust datasets is a significant issue in the growing field of data science, so our collection of accurately-labelled sound data could be of great value to 'machine hearing' engineers who aren't experts in acoustics, but need reliable data to train their sound recognition systems. Our audio dataset was therefore an ideal candidate for open publishing.

At this point it should be mentioned that funding bodies are increasingly trying to ensure that everyone can access the knowledge coming out of research they pay for. The MSoS project received nearly £1.3m from the Engineering and Physical Sciences Research Council (EPSRC), and making results open access was a requirement of the grant award. We were therefore obliged to publish our data at the end of the project, but our main motivation came much sooner when our project partners proposed an ideal way to make use of the information we'd created: a sound classification data challenge. Developers would use our audio dataset to try and create systems that could categorise sounds automatically; the winning systems would be the ones that matched our human-assigned labels as closely as possible.

A major concern was how best to host the dataset so that people from anywhere in the world could access it easily. We decided that figshare was an ideal platform for this. All we needed to do was add a link to the challenge webpage, and users would be navigated straight to the data where it was available for immediate download. Figshare offered a number of other advantages too. We could add comprehensive metadata such as a detailed description of the fileset, and tags to help it show up in searches. The data was automatically assigned a DOI and clear citation information, ideal for making sure the correct credit is given by anyone reusing the content. Submissions are version-controlled; this made it easy to manage phase two of the challenge where we had to add another set of audio files for competitors to access. Figshare also displays the number of views and downloads, helping the challenge organisers to monitor engagement. Finally, figshare automatically generates a preview of the fileset, even if it's a zip folder with nested subfolders. This means that users can preview data before downloading it - a very helpful feature for large submissions. Our

dataset contains 2000 files with a total size just over 600 MB; that's small in terms of typical machine learning datasets, but not insignificant if your broadband connection is slow!

As well as supporting us in navigating the features of figshare, the open research team in the library talked us through other issues associated with publishing data, specifically our obligations regarding creative commons (CC) licences. We discovered that usage terms are a really important aspect of open data, and anyone sharing content needs to understand the implications of the permissions they assign. A key consideration in the MSoS project was that we were curators rather than creators of the audio content in our research dataset. We collected these recordings from other repositories online, such as freesound (https://freesound.org/), then assessed, selected/ rejected, and processed them. We therefore had to ensure that we were complying with the licence terms set by the original owners of those recordings. The open research team guided us through the different levels of CC licence; they then advised us on best practice for writing a statement to accompany our data so that people using it would know what they're allowed to do with the files, as well as giving credit to the original creators of the audio content we used.

In the first stage of the machine learning challenge our dataset was downloaded 115 times from figshare, and that's just from sharing the link with specific people that we thought might like to participate. We've also had interest from another researcher who would like to use the files for experiments in a completely different project. The MSoS audio dataset is therefore a perfect example of how making your data open may not only be convenient within your own project, but allows it to be reused for other purposes that you'd never envisioned. As well as giving you the benefit of extra citations for no extra work, this saves other researchers time and effort so that knowledge is gained faster and funding goes further.

The current audio data challenge is open until the end of October, and winners will be announced in November. We're looking forward to finding out whether machines really can learn to identify sounds like humans!

For more information you can visit the challenge page of the MSoS website here http://cvssp.org/projects/making_sense_of_sounds/site/challenge/ and check out the data set on Figshare here https://doi.org/10.17866/rd.salford.6901475.v2