Taller de Aprendizaje Automático

Segundo Proyecto

Freesound AudioTagging 2019

Instituto de Ingeniería Eléctrica
Facultad de Ingeniería

UNIVERSIDAD
DE LA REPÚBLICA
URUGUAY

Montevideo, 2024

Motivación

- En la actualidad se generan inmensas cantidades de grabaciones
- Necesidad de estructurar los datos

kaggle

Search

Create

Home

Competitions

Datasets

Models

Code

Discussions

Learn

More

Your Work

VIEWED

EDITED

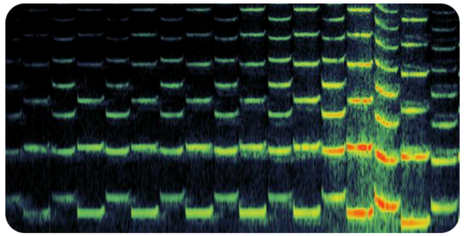
View Active Events

FREESOUND · RESEARCH CODE COMPETITION · 5 YEARS AGO

Late Submission

Freesound Audio Tagging 2019

Automatically recognize sounds and apply tags of varying natures



Overview Data Code Models Discussion Leaderboard Rules Team Submissions

Overview

Start
Apr 4, 2019

Close
Jun 17, 2019

Merger & Entry

Description

One year ago, Freesound and Google's Machine Perception hosted an audio tagging competition challenging Kagglers to build a general-purpose auto tagging system. This year they're back and taking the challenge to the next level with multi-label audio tagging, doubled number of audio categories, and a *noisier than ever* training set. If you like raising your ML game, this challenge is for you.

Competition Host
Freesound

Prizes & Awards
$5,000
Awards Points & Medals

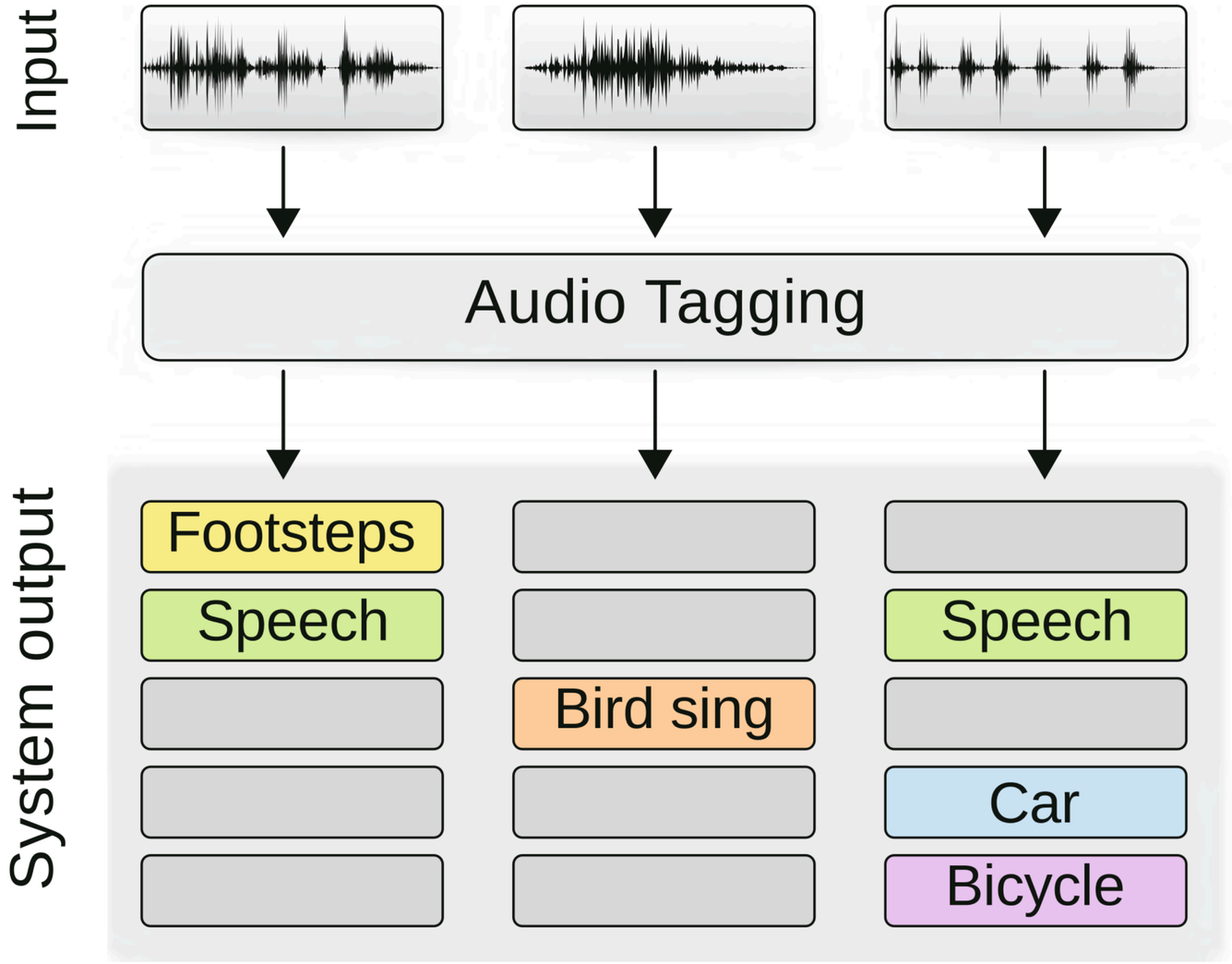
Participation
5,039 Entrants
520 Participants
880 Teams
677 Submissions

Tags

Audio

Weighted Label Ranking Average Precision

Problema a resolver



Esquema de un problema de Audio Tagging [1]

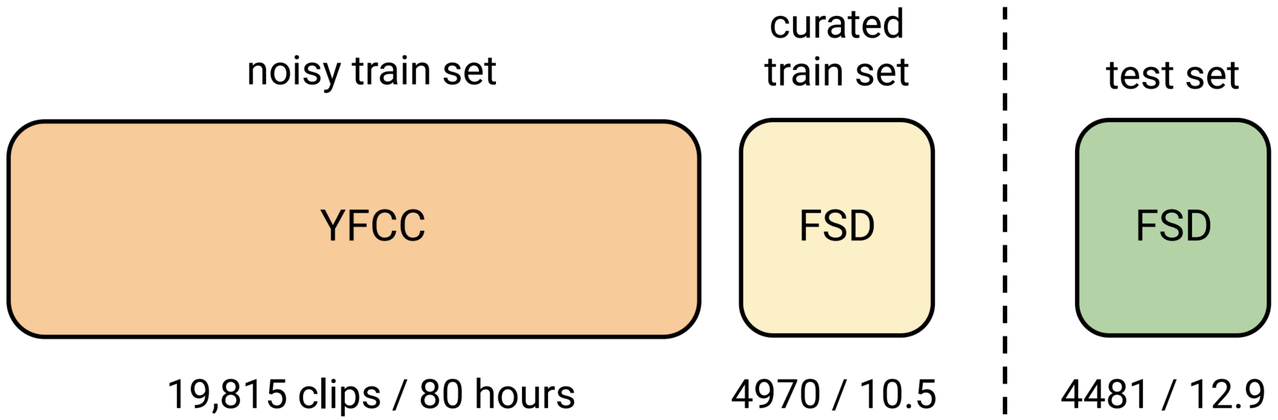
Objetivos del Desafío

1. Construir un modelo de Audio Tagging capaz de reconocer eventos sonoros de naturaleza diversa.

2. Aprovechar subconjuntos de datos de entrenamiento con anotaciones de fiabilidad variable.

Datos del desafío

- Los datos provienen de Freesound y Flickr



Aspect	curated train	noisy train	test
Clips/class	∼75*	300	∼ 50 - 150
Total clips	4970	19,815	4481
Labels/clip	1.2	1.2	1.4
Clip length	∼0.3 - 30s	∼15s	∼0.3 - 30s
Total duration	∼10.5h	∼80h	∼12.9h
Labelling	correct (inexhaustive)	noisy	exhaustive

Distribución de los datos [1]

- Datos etiquetados con 80 clases posibles.

Clases del Desafío

Human sounds

- Human voice
- Whistling
- Respiratory sounds
- Human locomotion
- Digestive
- Hands
- Heart sounds, heartbeat
- Otoacoustic emission
- Human group actions

Source-ambiguous sounds

- Generic impact sounds
- Surface contact
- Deformable shell
- Onomatopoeia
- Silence
- Other sourceless

Animal

- Domestic animals, pets
- Livestock, farm animals, working animals
- Wild animals

Sounds of things

- Vehicle
- Engine
- Domestic sounds, home sounds
- Bell
- Alarm
- Mechanisms
- Tools
- Explosion
- Wood
- Glass
- Liquid
- Miscellaneous sources
- Specific impact sounds

Music

- Musical instrument
- Music genre
- Musical concepts
- Music role
- Music mood

Natural sounds

- Wind
- Thunderstorm
- Water
- Fire

Channel, environment and background

- Acoustic environment
- Noise
- Sound reproduction

Ontología de AudioSet [3]

Métrica

- Se utiliza la métrica *label-weighted label-ranking average precision* (lωlrap)

$$Prec(s,c) = \frac{1}{Rank(s,c)} \sum_{r=1}^{Rank(s,c)} \mathbf{1}[Lab(s,r) \in C(s)]$$

Expresión de *label-ranking precision*[1]

- **s** - Indica la muestra con la que estamos trabajando
- **c** - Indica la clase para la cual calculamos la precision
- **C(s)** - Indica una lista de todas etiquetas ground-truth de la muestra **s**
- **Rank(s, c)** - Indica el ranking de la clase **c** en la predicción sobre la muestra **s**
- **Lab(s, r)** - Indica la etiqueta predicha en el ranking **r** para la muestra **s**

Ejemplo de Cálculo

Clases del problema ⟶ `[ dog barking - siren - engine idling ]`

Sea una muestra **s** con las siguientes etiquetas:

C(s) ⟶ `[ siren - engine idling ]`

Predicciones del modelo ⟶ `[ 0.7 - 0.2 - 0.5]`

A partir de las predicciones obtenemos el ranking:

`Rank(s, dog barking) = 1` `Rank(s, siren) = 3` `Rank(s, engine idling) = 2`

A partir de las predicciones obtenemos las etiquetas predichas en cada ranking:

`Lab(s, 1) = dog barking` `Lab(s, 2) = engine idling` `Lab(s, 3) = siren`

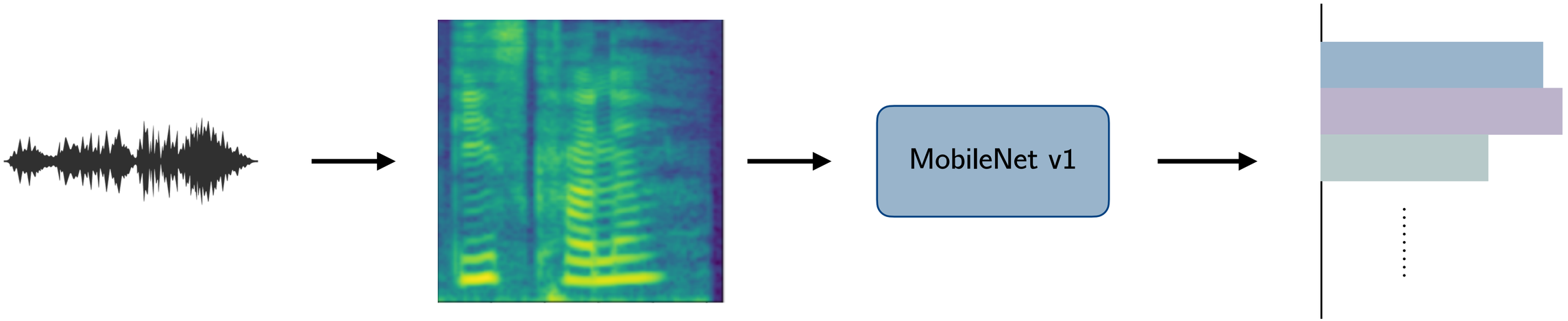
Métrica

- Se utiliza la métrica *label-weighted label-ranking average precision* (lωlrap)

$$l\omega lrap = \frac{1}{\sum_s |C(s)|} \sum_s \sum_{c \in C(s)} Prec(s, c)$$

- |C(s)| - Indica la cantidad de etiquetas de la muestra s

Baseline



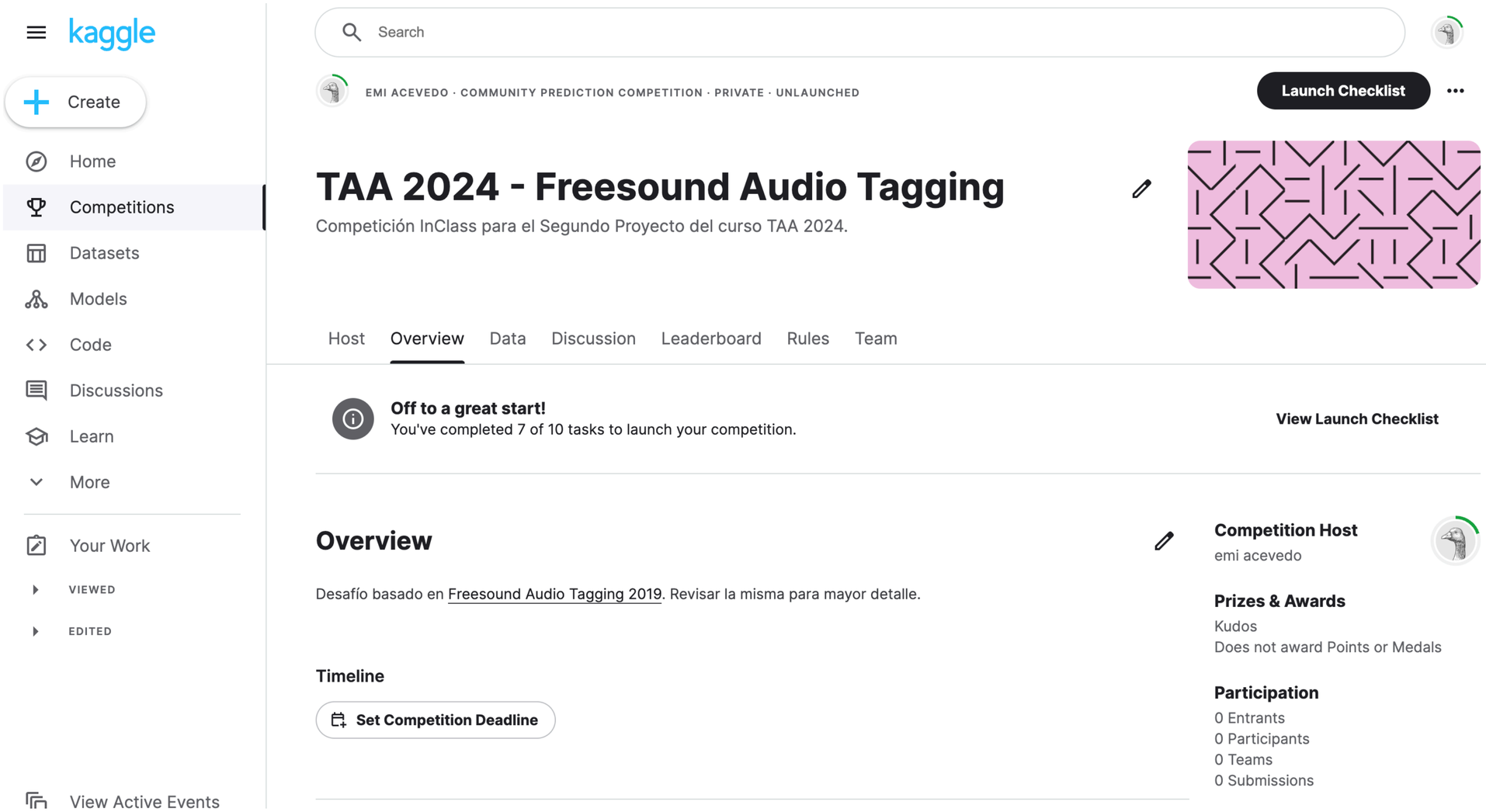
Audios

Espectrograma Log-Mel

Modelo

Score

Datos Disponibles - InClass



Link a la Competencia

Referencias Relevantes

- Documentación del desafío
- Ejemplo de problema de audio en Tensorflow
- Implementación de la Métrica lwlrap

Referencias

[1] Eduardo Fonseca, Manoj Plakal, Frederic Font, Daniel P. W. Ellis, & Xavier Serra. (2020). Audio tagging with noisy labels and minimal supervision.

[2] Virtanen, T., Plumbley, M. D., & Ellis, D. (Eds.). (2018). Computational Analysis of Sound Scenes and Events. Springer.

[3] J. F. Gemmeke et al., "Audio Set: An ontology and human-labeled dataset for audio events," 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), New Orleans, LA, USA, 2017, pp. 776-780,