Data Cleansing with Contrastive Learning for Vocal Note Event Annotations

Gabriel Meseguer-Brocal¹, Rachel Bittner², Simon Durand², Brian Brost² 1- Ircam Lab, CNRS, Sorbonne Université Paris, France, Paris. 2- Spotify, USA. gabriel.meseguerbrocal@ircam.fr, rachelbittner@spotify.com, durand@spotify.com, brianbrost@spotify.com

Music data

Most MIR problems are solved using supervised methods \rightarrow labeled data.

Music datasets have label errors:

- Incorrect labels.
- Imprecise event times.
- Missing/extra events.



What can we do about it?

- Getting more human annotations: demanding and costly.
- Automatically correcting the errors: difficult for complex tasks.
- Find the errors in an automatic way.

Data cleansing studies how to mitigate the effects of label noise.

Particularly problematic for datasets deriven from the Internet.

Close-up of the spectrogram



Note annotations \rightarrow errors in time and frequency.

Bad annotations are problematic for *training* and *evaluation*.





We use *contrastive learning* to determinate if an (audio, annotation) pair are a good label or not, exploiting sequential dependencies between labels to predict incorrectly labeled time-frames trained using likely correct labels pairs as positive examples and and local deformations of correct pairs as negative examples

Contrastive learning

The pitch labels are close to the notes annotation we have in the noise dataet. By comparing both, pitch estimations and note annotations, we can select the "likely correct" examples, where the prediction is similar to the label. We distort them to generate the incorrect examples, defining our training set.

clean the noisy dataset.

We take our noisy data where we don't know if the note annotations are good or bad and label it using a model that predict the pitch. This model is trained in another dataset.

2







(Top) The output of the error detection model for a short segment. (Bottom) the corresponding CQT and annotated notes (in white). The error is high at the beginning of the fourth note because it starts late, and at end of the last note because it is too long.

> SORBONNE UNIVERSITÉ

ircam

Pompidou

E Centre

How to validate $g(x, \hat{y}) \rightarrow z$?

a. Directly: no "real" ground truth (only likely correct).

3

- -b. Manually: costly and required expert knowledge.
- c. Data cleansing: identifying incorrect annotations and remove them during training.





Distribution of scores for the three training conditions. The curves are shifted to the right which means we have better results for the models trained after filtering the bad annotations with our error probability function.

Code at: https://github.com/gabolsgabs/contrastive-data-cleansing