Iterative Refinement of Cluster Assignments

Unsupervised Unit Discovery

Step1

Step2

Representation Learning via Masked Prediction

Cross Entropy Loss

Quantized Units $Z_t = [z_1, \ldots, z_T]$

Hidden Units Embedding $e_1\ e_2\ e_2\ e_3\ e_3\ e_2\ e_2\ e_3\ e_1$

Logits

Cosine Similarity

$o'_3\ o'_4\ o'_5$

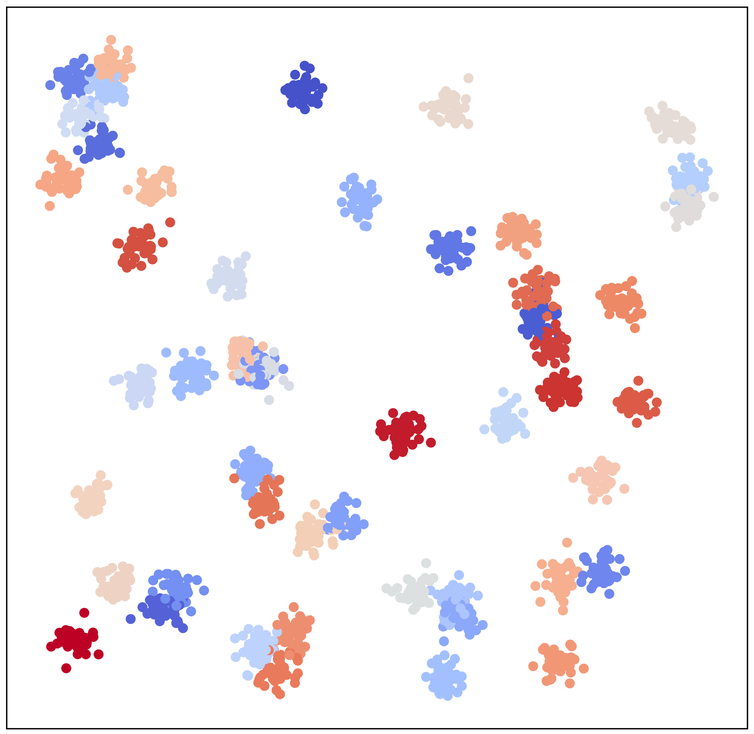
Projection

$o_3\ o_4\ o_5$

Transformer

$x_0\ x_1\ x_2$ MSK MSK MSK $\cdots$ $x_{T-1}\ x_T$

CNN Encoder



K-means Clustering

Use intermediate features as clustering features from second iteration

MFCC $m_0\ m_1\ m_2\ \cdots\ m_t\ \cdots\ \cdots\ \cdots\ m_T$

Clustering Feature Extraction