Project Report

Aaditya Singh
160002

Fundamental Frequency Estimation

In the paper *Singing Voice Melody Transcription using DNNs*, the logarithmic magnitude spectrogram of the raw audio signal is passed through a neural network to classify the fundamental frequency of the predominant melody at each time frame into a fixed number of classes.

Hidden Markov Model

HMMs are used to model randomly changing systems, where the visible outputs of the system depend on the hidden state of the system. There are two steps involved in training an HMM. Forward Algorithm, Trellis Diagram, and Baum-Welch Algorithm are used for the purpose of Evaluation, Decoding and Learning respectively. HMMs along with CNNs were used in the paper *Note based QBH system using CNNs* to model the posterior probability of each hidden state from the hummed query.

Non Negative Matrix Factorization

Given a non-negative matrix V, NMF decomposes V into the product of two non-negative matrices W and H. The paper *Algorithms for Non-negative Matrix Factorization* presents iterative update rules for the matrices W and H, such that the cost (Euclidean or Kullback-Leibler Divergence) is non-increasing under each update.
The paper *Monoaural sound source separation by NMF* modifies the update rules of NMF by imposing an additional cost for the temporal continuity of the basis vectors.

Probabilistic Latent Component Analysis

If V represents the magnitude spectrogram of a given song snippet, the random variable f represents the frequency index and t represents the time frame. Then PLCA allows us to characterize the joint distribution P(f,t) as

$$P(f,t) = \sum_z P(z)P(f|z)P(t|z) \tag{1}$$

This is a symmetrical decomposition of P(f,t) obtained by considering both f and t dimensions as features. We can instead have a different decomposition by treating the two dimensions differently.

$$P(f,t) = P(t)\sum_z P(f|z)P(z|t) \tag{2}$$

Latent Variable Model

Consider a random process characterized by the probability P(f) of drawing a feature unit f in a given draw. Let the random variable f take values from the set 1, 2,.., F. P(f) is unknown, what we observe instead is the feature counts, that is the number of times f is observed after repeated draws. We can approximate P(f) by using a normalized set of counts.
Assume P(f) comes from K hidden distributions or latent factors, The distributions are selected according to their relative probabilities which remain the same for an experiment. We wish to characterize these hidden distributions.

Let P(f|z) be the probability of observing feature f given z, which represents the index of the hidden distribution being considered. The probability of choosing the z^{th} distribution in the t^{th} experiment is $P_t(z)$

$$P_t(f) = \sum_z P(f|z)P_t(z) \tag{3}$$

Parameter Estimation

Let V_{ft} represents the feature count of f in the t^{th} experiment, which is known. We want to estimate $P_t(z)$ and P(f|z). We use a maximum likelihood formulation of the problem, and the standard procedure of solving it in latent variable models is the Expectation-Maximization Algorithm. For the E-step, we obtain a posteriori probability for the latent variable as

$$P_t(z|f) = \frac{P_t(z)P(f|z)}{\sum_z P_t(z)P(f|z)} \tag{4}$$

For the M-step, we obtain the re-estimation equations

$$P(f|z) = \frac{\sum_t V_{ft}P_t(z|f)}{\sum_f \sum_t V_{ft}P_t(z|f)} \tag{5}$$

and

$$P_t(z) = \frac{\sum_f V_{ft}P_t(z|f)}{\sum_z \sum_f V_{ft}P_t(z|f)} \tag{6}$$

The parameters P(f|z) and $P_t(z)$ are randomly initialized and re-estimated using the above equations until a termination condition is met.

References

[1] Singing Voice Melody Transcription using Deep Neural Networks
https://wp.nyu.edu/ismir2016/wp-content/uploads/sites/2294/2016/07/163_Paper.pdf

[2] A Note Based Query By Humming System using Convolutional Neural Network
https://www.isca-speech.org/archive/Interspeech_2017/pdfs/1590.PDF

[3] Algorithms for Non-negative Matrix Factorization
http://papers.nips.cc/paper/1861-algorithms-for-non-negative-matrix-factorization.pdf

[4] Monaural Sound Source Separation by Nonnegative Matrix Factorization With Temporal Continuity and Sparseness Criteria
https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=4100700