Mini licenta

Cota Ionas-Calin

January 6, 2019

1 Related Work

Automatic music transcription (AMT) has been attempted since the 1970s and polyphonic music transcription dates to the 1990s [1]

1.1 State of the art in AMT

A model used in [2] uses 87 Support Vector Machine (SVM) classifiers to perform frame-level classification with the advantage of simplicity, and then a Hidden Markov Model (HMM) post-processing was adopted to smooth the results. On top of it, Deep Belief Network(DBN) was added to learn higher layer representation of features in [3]. Since none of the approaches has reached the same level of accuracy as human experts, most music transcription work is completed by musicians. With the development of deep learning in recent years, many researchers were inspired to apply networks to accomplish AMT. A model based on Convolutional Neural Networks (CNN) was proposed in [4]. More models adopted Reccurent Neural Networks (RNN) or Long Short-Term Memory (LSTM) due to its capability of dealing with sequential data [1] [5] [6]. In [7], 5 models were compared and the ConvNet model was reported as resulting in the best performance.

The first majort AMT work is Smaragdis et al.[8]. This approach uses Non-Negative Matrix Factorization (NMF). This is the main methodology employed in software for automatic transcription, but it has it's limitations. For example, it needs to know how many individual notes are desired for the transcription (information that is not always available).

The next work worth mentioning is Emiya et al.[9], not because of their transcription system (as it was out-performed in the same year), but because of the dataset they created that has become the standard in evaluating any multi-pitch estimation system. They created the MIDI-Aligned Piano Sounds (MAPS) data set composed of around 10,000 piano sounds either recorded by using an upright Disklavier piano or generated by several virtual piano software products based on sampled sounds. The dataset consists of audio and corresponding annotations for isolated sounds, chords, and complete pieces of piano music. For our purpose we use only the isolated sounds (daca nu gasesc ceva mai bun).

Sigtia et al.[10] built the first AMT system using CNN, outperforming the state-of-the-art approaches using NMF. Convolutional Neural Networks are a discriminative approach to AMT, which has been found to be a viable alternative to spectrogram factorization techniques. Discriminative approaches aim to directly classify features extracted from frames of audio to the output pitches. This approach uses complex classifiers that are trained using large amounts of training data to capture the variability in the inputs, instead of constructing an instrument specific model.

1.2 Products

In this section, we present products that use deep learning for AMT.

1.2.1 Melodyne

Melodyne is a popular plugin used for Music Transcription and Pitch Correction. It costs up to $700.

The Melodic and Polyphonic algorithms offer you, in the case of vocals as well as both mono- and polyphonic instruments, full access to the notes of which the sound is composed as well as to their musical parameters.

1.2.2 AnthemScore

AnthemScore is a product used for Music Transcription that uses CNN.

They approach note detection as an image recognition problem by creating spectrograms of the audio. They show how the spectrum or frequency content changes over time. The method used for creating the spectrograms is the constant Q transform instead of the more common Short Time Fourier Transform (STFT) method.

References

[1] D. G. Morin, *Deep neural networks for piano music transcription.* 2017.

[2] G. E. Poliner and D. P. Ellis, *A discriminative model for polyphonic piano transcription.* EURASIP Journal on Applied Signal Processing, vol. 2007, no. 1, pp. 154, 2007.

[3] J. N. J. Nam, H. Lee, and M. Slaney, *A classification-based polyphonic piano transcription approach using learned feature representations.* " Proceedings of the 12th International Society for Music Information Retrieval Conference, pp. 16-180, 2011.

[4] K. Ullrich and E. van der Wel, *Music transcription with convolutional sequence-to-sequence models.* " International Society for Music Information Retrieval, 2017.

[5] J. F. S. B. L. Sturm, O. Ben-Tal, and I. Korsunova, *Music transcription modelling and composition using deep learning.* arXiv preprint arXiv:1604.08723, 2016.

[6] S. Sigtia, E. Benetos, N. B.-L. T. Weyde, A. S. d'Avila Garcez, and S. Dixon, *A hybrid recurrent neural network for music transcription.* IEEE International Conference on Acoustics, Speech, and Signal Processing, pp. 2061-2065, 2015.

[7] E. B. S. Sigtia and S. Dixon, *An end-to-end neural network for polyphonic piano music transcription.* IEEE/ACM Trans. Audio Speech Lang. Process., 24, 927–939, 2016.

[8] P. Smaragdis and J. C. Brown., *Non-negative matrix factorization for polyphonic music transcription.* In Applications of Signal Processing to Audio and Acoustics, 2003 IEEE Workshop on., pages 177–180. IEEE,, 2003.

[9] R. B. V. Emiya and B. David., *Multipitch estimation of piano sounds using a new probabilistic spectral smoothness principle.* IEEE Transactions on Audio, Speech, and Language Processing, 18(6):1643–1654, 2010.

[10] J. Sleep, *AUTOMATIC MUSIC TRANSCRIPTION WITH CONVOLUTIONAL NEURAL NETWORKS USING INTUITIVE FILTER SHAPES.* A Thesis presented to the Faculty of California Polytechnic State University, San Luis Obispo, 2017.