Multiple Viewpoint Melodic Prediction with Fixed-Context Neural Networks

Srikanth Cherla[1,2], Tillman Weyde[1,2], Artur d'Avila Garcez[2]

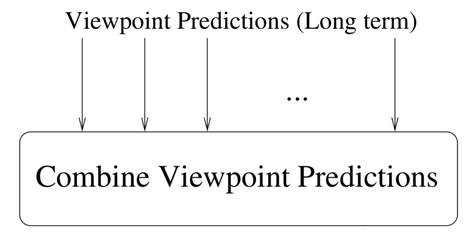
[1]Music Informatics Research Group, City University London
[2]Machine Learning Group, City University London

1. Introduction

- Modelling statistical regularities in melodies.
- A note based approach - Multiple Viewpoint Systems.
- Predict a probability distribution over possible values of pitch of next note given the notes immediately preceding it: $p(s_t|s_{(1,\ldots,t-1)})$.
- **Applications:** Music generation, studying melodic expectation, solo instrument/singing voice transcription.

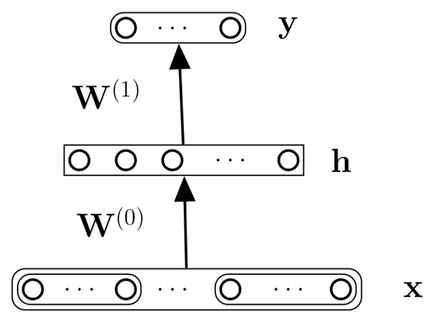
2. Multiple Viewpoints Representation of Melody

Viewpoint	Transformed sequence									
pitch	67	69	71	72	69	72	64	67	72	69
int	⊥	2	2	1	−3	3	−8	3	5	−3
onset	0	2	5	6	9	10	12	15	16	20
ioi	⊥	2	3	1	3	1	2	3	1	4
int ⊗ ioi	⊥	2, 2	2, 3	1, 1	−3, 3	3, 1	−8, 2	3, 3	5, 1	−3, 4

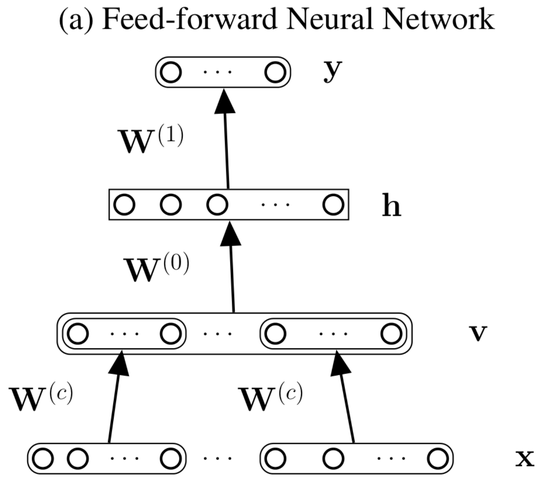


- A framework for analysis and generation of music in symbolic form (MIDI, Kern, etc.) [1].
- Melody broken down into parallel *viewpoint type* sequences.
- Viewpoint types can be *linked* via Cartesian product.
- One n-gram model per viewpoint type to overcome data sparsity.
- Multiple models combined using entropy-weighted mixture- or product-of-experts.
- *Input types:* Types in the context; *Target type:* Type being predicted.

3. Prediction Models



(a) Feed-forward Neural Network

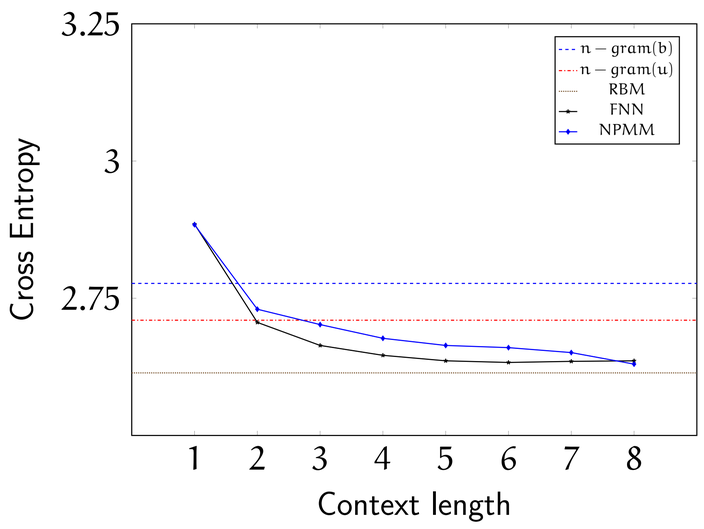


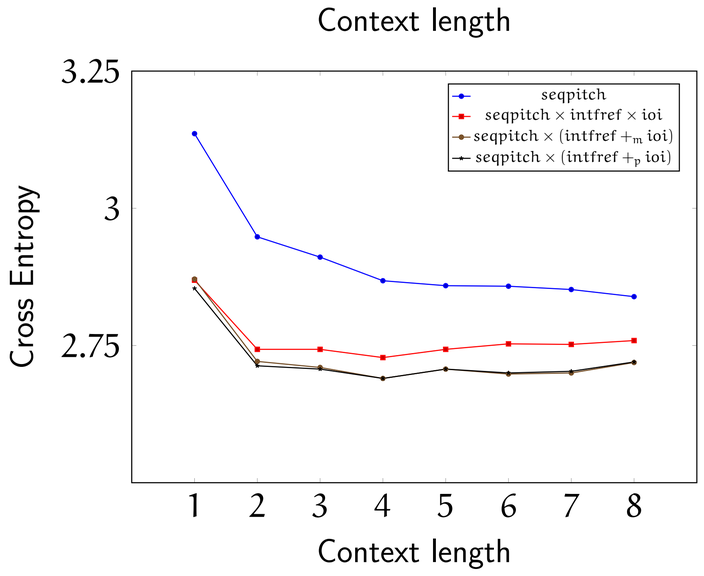
(b) Neural Probabilistic Melody Model

- Two non-recurrent neural network prediction models.
- Context events in one-hot representation given as input.
- Multiple input type one-hot vectors concatenated.
- Networks have softmax output layer to predict $p(y|x)$.
- The feedforward neural network has a single logistic sigmoid hidden layer.
- The neural probabilistic melody model [2] has a hyperbolic tangent hidden layer and an additional linear *embedding* layer.
- Both networks trained using the backpropagation and mini-batch gradient descent [3].

4. Experiments & Evaluation

- **Dataset:** Bach chorales ($9,227$ notes), and Folk melodies of Canada ($8,553$ notes), China ($11,056$ notes) and Germany ($8,393$ notes) from the Essen Folk Song Collection [6].
- **Criterion:** Cross Entropy - A measure of the divergence between model predictions and the true data distribution.
- **Methodology:** Model selection through grid search, with each model evaluated on folds identical to those in [4].





Experiments

- Predict pitch, given a context of...
  - pitch.
  - pitch, inter-onset interval, scale-degree.
- Multiple viewpoint types combined via...
  - a single model.
  - a combination of multiple models.
- Compared with previous n-gram and restricted Boltzmann machine prediction models [4, 5].

Observations

- Predictions improve with context length.
- The respective best cases of both neural network models better than n-grams but worse than RBMs.
- Additional viewpoint types (in the present case) improve predictions.
- A combination of multiple models better than a single model particularly for longer contexts.
- Both MoE and PoE model combinations improve predictions. PoE only slightly better in some cases.

5. Future Work

- Neural network "tricks" and better optimization to improve existing results.
- Recurrent models to limit increase in the input space with context length.
- Online learning to update model parameters while it predicts unseen sequences.
- Application to melody segmentation, voice separation and singing-voice transcription.

6. Acknowledgements

Srikanth Cherla is supported by a PhD studentship from City University London, and his travel to ISMIR has been funded in part by the City Graduate School Conference Attendance Fund.

References

[1] Conklin, Darrell, and Ian H. Witten. "Multiple viewpoint systems for music prediction." *Journal of New Music Research 24.1 (1995): 51-73.*

[2] Bengio, Yoshua, et al. "Neural probabilistic language models." *Innovations in Machine Learning*. Springer Berlin Heidelberg, 2006. 137-186.

[3] Rumelhart, David E., Geoffrey E. Hinton, and Ronald J. Williams. "Learning representations by back-propagating errors." *Cognitive modeling* (1988).

[4] Pearce, Marcus, and Geraint Wiggins. "Improved methods for statistical modelling of monophonic music." *Journal of New Music Research* 33.4 (2004): 367-385.

[5] Cherla, Srikanth, et al. "A Distributed Model For Multiple-Viewpoint Melodic Prediction." *ISMIR*. 2013.

[6] http://www.esac-data.org/ (Last accessed on Oct 15, 2014).