MIDI REPRESENTATION WITH GRAPH EMBEDDINGS

Pasquale Lisena
EURECOM, France
`lisena@eurecom.fr`

Albert Meroño-Peñuela
King's College London, UK
`albert.merono@kcl.ac.uk`

Raphaël Troncy
EURECOM, France
`troncy@eurecom.fr`

ABSTRACT

In MIR, feature extraction has been extensively used for learning models of MIDI. We propose an alternative approach that relies on the extraction of latent features from a graph of connected nodes. We show that our MIDI2vec approach has good performance in metadata prediction.

1. INTRODUCTION AND METHOD

MIDI libraries require metadata in order to be effectively searched and browsed by the final user. Automatically generating these metadata at scale is a challenging task. It has been traditionally addressed by analysing the content of the symbolic notation, i.e. identifying symbolic patterns in melody, harmony, rhythm, structure, etc. From these extracted features, different classification models can be obtained e.g. for predicting the genre [1, 2], emotions, or even the composer of a musical work [3], mostly using supervised machine learning techniques. However, these techniques rely on manual feature engineering and on an appropriate selection of the the features to be extracted. Hence, relevant characteristics of the music may not be taken into account, having an impact on the final results.

In this paper, we propose to use graph embeddings techniques, which have been successfully applied to represent graph information [4]. To do so, we use MIDI2vec [5], a method for representing MIDI data as vector-space embeddings for automated metadata classification. The approach relies on *node2vec* [6], an algorithm which embeds the graph by computing random walks and using these walks as sentences in a word embedding algorithm (*word2vec*, which it extends).

The MIDI2vec approach is composed of 2 steps:

1. A preliminary conversion of the MIDI event sequence into a graph. As shown in Figure 1, a *MIDI node* (the circle) represents the MIDI file that will be connected to nodes representing different parts of the MIDI content: tempo, programs, time signature, notes. These elements are discretised. Simultaneous notes are grouped together in a *group of notes*, to which an identifier based on the content is assigned. A MIDI node can be linked to one or more nodes for each type.

2. We use *node2vec* to compute a 100-dimensions embedding vector for each node of the input graph. Each dimension of the vector cannot be attributed to a specific feature – e.g. the tempo – but it rather represents a latent feature learned by the embedding algorithm.

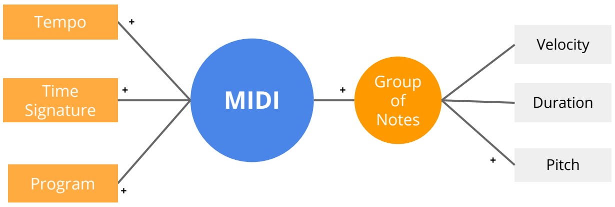


Figure 1: Schematic view of the graph generated from MIDI. The **+** indicates edges representing connections of type many-to-many.

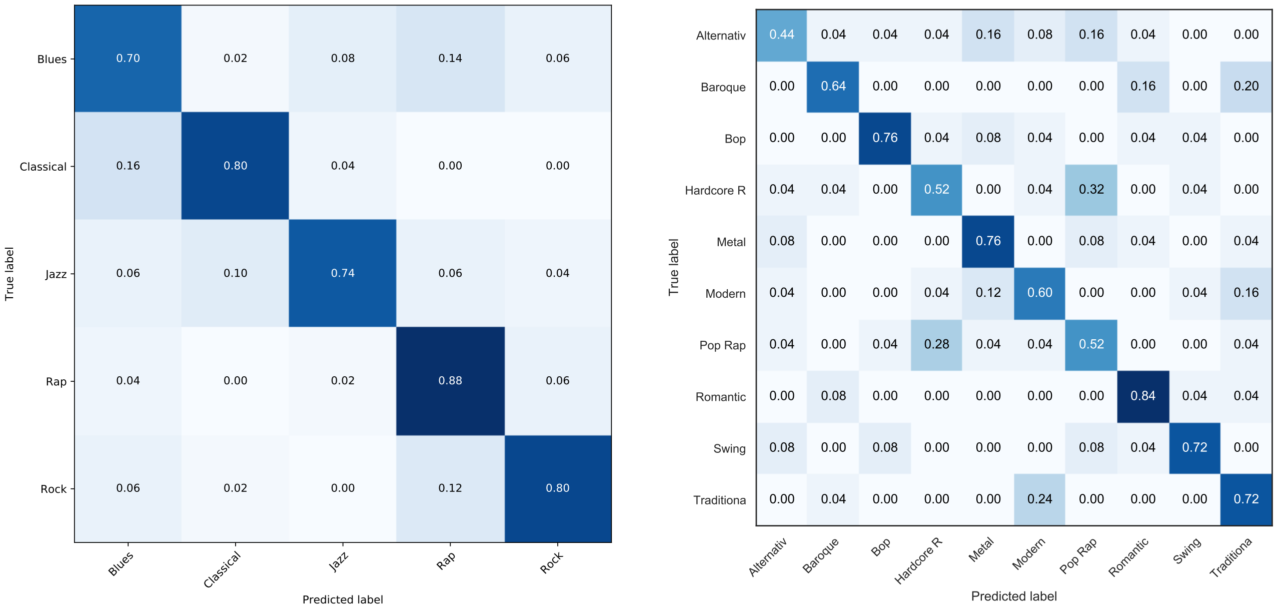
2. GENRE PREDICTION

The obtained vectors from MIDI2vec can be used as inputs to algorithms for tasks such as classification. In [7], a genre classification is performed on a contextually published SLAC Dataset [1], which contains 250 MIDI files classified according to a two-level perfectly balanced taxonomy. The first level includes 5 genre labels (Blues, Classical, Jazz, Rap, Rock), while the second one further specialises each genre by 2 sub-genres, for a total of 10 subgenre labels. The classification has been carried out using the jSymbolic library for extracting 111 features (1021 dimensions). Recently, their work has been extended and improved including, among others, features about chords and simultaneous notes, for a total of 172 features (1497 dimensions) [8].

We perform a 5-class genre classification experiment as well as a 10-class experiment on the SLAC dataset. MIDI embeddings are generated on the dataset using MIDI2vec. A Feed-Forward Neural Network receives the MIDI embeddings as input (100 dimensions) in batches of size 32. The neural network consists of 3 dense layers. The hidden layers count 100 neurons each and use *ReLU* as activation function. The output layer uses a sigmoid as activation

Attribution: P. Lisena, A. Meroño-Peñuela and R. Troncy, "MIDI Representation with Graph Embeddings", in *Extended Abstracts for the Late-Breaking Demo Session of the 22nd Int. Society for Music Information Retrieval Conf.,* Online, 2021.

[1] SLAC dataset: `http://jmir.sourceforge.net/Codaich.html`



(a) Genre prediction

(b) Sub-genre prediction

Figure 2: Confusion matrices of midi2vec predictions for the SLAC dataset.

function and has a number of neurons equal to the dimension of the vocabulary of labels, which is represented with one-hot encoding. We performed 10-fold cross-validation for training the neural network and we provide as final score the average of the accuracy computed on every fold. Please note that only the MIDI embedding vectors are provided in input to the NN, and not the labels to be predicted. In Table 1, we compare our approach with the one of jSymbolic.

Approach		**5 classes**	**10 classes**
jSymbolic 2010 [7]		85%	66%
jSymbolic 2018 [8]		93.2%	77.6%
MIDI2vec + NN	ALL	**86.4%**	**67.2%**
	*N	81.6%	62.4%
	*P	79.6%	61.6%
	*T	27.2%	18.8%
	*TS	25.6%	15.2%
	*300	79.2%	57.2%

Table 1: Accuracy of the genre classification. The reported values are the average of the cross-fold validation.

Our approach slightly outperforms [7], with an accuracy of 86% for 5-classes and 67% for 10-classes prediction. The improvements made in [8] increase these scores by a few percentage points. We believe that the combination of melodic and chords features was crucial in this case and worth investigating in future.

The same Table 1 shows also the accuracy scores obtained with different variations of the complete model (ALL); these variations compute the embeddings on the sole notes nodes (*N), program nodes (*P), tempo nodes (*T), and time signature nodes (*TS). None of these single features reaches the accuracy score of their combination. Finally, we trained the embeddings on all features, but taking into account only the first 300 note groups (*300). The experiment shows that reducing the number of vertexes in the graph causes lower accuracy scores.

Figure 2 shows the confusion matrix between the real and the predicted values (configuration ALL). Even if there are no strong patterns, we can state that *Blues* is the genre that attracts more negative predictions. Figure 2b confirms that sub-genres belonging to the same parent genre are easier to be confused.

3. CONCLUSION AND FUTURE WORK

With MIDI2vec – a method to represent MIDI content as a graph and, subsequently, in a vector space through learning graph embeddings – we demonstrated that symbolic music content in MIDI files, and its embedding representation in vector space are a powerful tool for automated metadata classification. MIDI2vec embeddings obtained comparable performances to state-of-the-art methods based on feature extraction, with the added advantages of scalability, automating feature engineering, and reducing the required dimensions by one order of magnitude.

We also release an open-source library for producing MIDI embeddings at `https://git.io/midi2vec`. The experiments are available as notebooks at `https://git.io/midi-embs`.

We plan on improving this work in various ways, in particular including time information using sequence embeddings [9] or temporal graphs [10], using it in combination with other feature extraction techniques in an ensemble system, applying it to a proper MIDI ontology [11].

Acknowledgements. This work is partly supported by the CLARIAH project funded by the Dutch Research Council NWO. This work is part of a project that has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101004746 (Polifonia: a digital harmoniser for musical heritage knowledge).

4. REFERENCES

[1] C. McKay and I. Fujinaga, "Automatic Genre Classification Using Large High-Level Musical Feature Sets," in 5th *International Conference on Music Information Retrieval (ISMIR)*, Barcelona, Spain, 2004. [Online]. Available: http://ismir2004.ismir.net/proceedings/p095-page-525-paper240.pdf

[2] Z. Cataltepe, Y. Yaslan, and A. Sonmez, "Music Genre Classification Using MIDI and Audio Features," *EURASIP Journal on Advances in Signal Processing*, vol. 2007, no. 1, p. 036409, 2007.

[3] Z. Fu, G. Lu, K. M. Ting, and D. Zhang, "A Survey of Audio-Based Music Classification and Annotation," *IEEE Transactions on Multimedia*, vol. 13, no. 2, pp. 303–319, 2011.

[4] P. Ristoski, J. Rosati, T. Di Noia, R. De Leone, and H. Paulheim, "RDF2Vec: RDF Graph Embeddings and Their Applications," *Semantic Web Journal*, vol. 10, no. 4, pp. 721–752, 2019.

[5] P. Lisena, A. Meroño-Peñuela, and R. Troncy, "MIDI2vec: Learning MIDI Embeddings for Reliable Prediction of Symbolic Music Metadata," *Semantic Web Journal - Special Issue on Deep Learning for Knowledge Graphs*, 2021, to appear.

[6] A. Grover and J. Leskovec, "node2vec: Scalable Feature Learning for Networks," in *22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, San Francisco, CA, USA, 2016.

[7] C. McKay, J. Burgoyne, J. Hockman, J. B. L. Smith, G. Vigliensoni, and I. Fujinaga, "Evaluating the Genre Classification Performance of Lyrical Features Relative to Audio, Symbolic and Cultural Features," in *11th International Society for Music Information Retrieval Conference (ISMIR)*, Utrecht, The Netherlands, 2010.

[8] C. McKay, J. E. Cumming, and I. Fujinaga, "jSymbolic 2.2: Extracting Features from Symbolic Music for use in Musicological and MIR Research," in 19th *International Conference on Music Information Retrieval (ISMIR)*, Paris, France, 2018. [Online]. Available: http://ismir2018.ircam.fr/doc/pdfs/26_Paper.pdf

[9] C. Ranjan, S. Ebrahimi, and K. Paynabar, "Sequence graph transform (sgt): A feature extraction function for sequence data mining," *arXiv preprint arXiv:1608.03533*, 2016.

[10] U. Singer, I. Guy, and K. Radinsky, "Node Embedding over Temporal Graphs," in *28th International Joint Conference on Artificial Intelligence, (IJCAI)*. IJCAI Organization, 7 2019, pp. 4605–4612.

[11] A. Meroño-Peñuela, M. Daquino, and E. Daga, "A Large-Scale Semantic Library of MIDI Linked Data," in *5th International Conference on Digital Libraries for Musicology (DLfM)*, Paris, France, 2018.