Applications of word embedding models to a classical music corpus: stylistic analysis and composer classification

Eamonn Bell EPB2125@COLUMBIA.EDU
Columbia University, New York, NY

Jaan Altosaar ALTOSAAR@PRINCETON.EDU
Princeton University, Princeton, NJ

Abstract

We apply a word embedding model to a large symbolic corpus of classical music to learn an embedding space where chords are represented by real-valued vectors. In early classical music, the first two principal components of the embeddings of major triads form a circle. In music from later composers, this circular topology is less evident. The order in which major triads are arranged on this structure corresponds to their order in the circle of fifths. The emergence and perturbation of this structure is justified by reasoning about the probabilistic embedding model and stylistic trends in the composition of classical music. We show how this technique is useful for large-scale, quantitative stylistic analysis of music, and musical document similarity in general, by using our learned embeddings and the word-mover's distance (Kusner et al., 2015) to classify composers.

1. Introduction

1.1. Word embeddings

Probabilistic models such as Latent Dirichlet Allocation (Blei et al., 2003) are standard tools for analyzing text data. However, such models use bag-of-words representations. In contrast, word embeddings (real-valued vectors representing words in a vocabulary) were first introduced by (Bengio et al., 2003) but popularized by Mikolov et al. (2013) under the 'word2vec' moniker. The embedding space returned by a model trained on a sufficiently large and informative corpus models a notion of semantic similarity as cosine distance between word vectors. There exists some prior work on applying word embedding models to music. Huang et al. (2016) trained a word embedding model on a much smaller ($n = 200$) corpus of pop and rock music in service of a novel user interface in computer-aided composition environment.

We train a similar word embedding model on a significantly larger corpus. Models in the word2vec family have shown robust results on natural language reasoning tasks when trained on corpora of the order of a million tokens. At 12M chord tokens, Yale/Classical Archives Corpus (YCAC) introduced by White & Quinn (2014) approaches the size thought to be required for this technique. Despite the fact that Shanahan & Albrecht (2013) show high pitch-class error rates in this corpus, the YCAC remains the most acessible corpus of classical music available at the scale thought to be useful for word embedding models. Our goal is to explore the resulting space for latent structures analogous to those which allow for the usefulness of these models in NLP tasks. We also show how it is possible to use the word-mover's distance (Kusner et al., 2015) in conjunction with learned chord embeddings to classify composers.

1.2. Statistical analysis of harmony

Attempts to determine rules describing the use of harmony by composers of Western art music (hereafter, classical music) have been characteristic of music theory for centuries. Traditional functional harmonic theory classifies chords into three categories: tonic, dominant, and subdominant. These categories can be used to describe prototypical sequences of harmonies. For example, the progression tonic $\rightarrow$ subdominant $\rightarrow$ dominant $\rightarrow$ tonic is normative in classical music. Large symbolic corpora of music allows us to explore these categorization schemes.

Jacoby et al. (2015) characterize theories of harmony as mappings from an annotated corpus to a list of chord categories and provide formal criteria for evaluating novel theories of this kind. Our results suggest that the structure of word embedding spaces trained on musical data captures some intuitions about harmonic function. If so, clustering might be effectively applied to these spaces, which implicitly define chord categorization functions evaluable per Jacoby et al. (2015).

2. Methodology

The Yale/Classical Archives Corpus (YCAC) is a database of pitch-class and time data from MIDI files encoding 8,980 distinct pieces of classical music (White & Quinn, 2014). Each piece is represented by a sequence of time-stamped chord tokens in which a new token is created at every moment a voice enters or leaves the musical texture. We reduce each chord token to a binarized chroma vector. A binarized chroma vector is a twelve-place vector $v \in \{0,1\}^{12}$ in which the i-th component of vector v_i is set to 1 if the i-th pitch class is present in the chord token. We select four comparably-sized subcorpora based on the dates of composition of pieces, in fifty-year chunks from 1700–1899 (around 2M chord tokens each). A word embedding model in the word2vec family was used to learn a number of embedding spaces. First, a space was trained on the entire corpus to evaluate its plausibility. Then, four separate word-embedding spaces were trained on the date-delimited subcorpora described above.[1] We ensure embedding spaces are reasonable by evaluating ranked nearest neighbors to common chords according to cosine similarity.

3. Results & Discussion

3.1. Analysis of time-delimited subcorpora

Figure 1 shows the locations of all twelve major triads in the embedding space trained on the four subcorpora, as represented by the coefficients for the first two principal components. Arrows connect pairs of major triads such that arrows originate from the dominant of its target, in the key of represented by the target triad. The major triads are arranged in a circular topology. The order of major triads in this arrangement respects their order in the circle of fifths. That is, the roots of adjacent triads are related by the interval of a perfect fifth. The arrangement of major triads becomes more deformed as models trained on music from a later period are considered. We hypothesize that the irregularity is due to decreasing use of functional harmony. Briefly, in later music, major chords tend to appear less frequently in the contexts of their fifth-related neighbors.

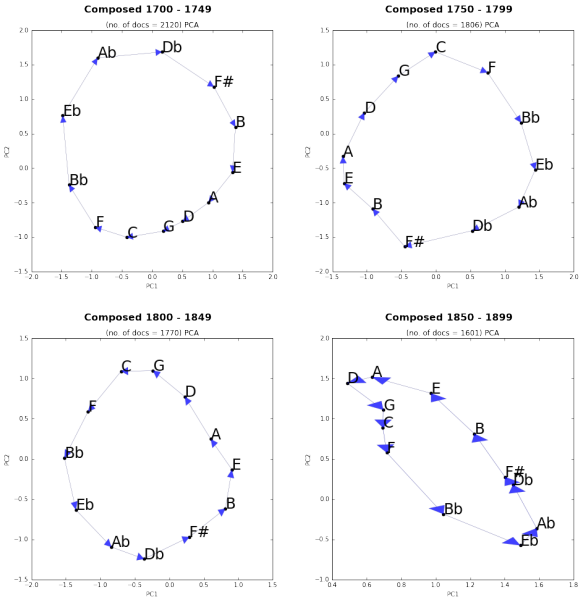

Figure 1. Clockwise from top left: the first two principal components of embeddings of major chords, for music composed in years 1700–1749, in years 1750–1799, in years 1850–1899, and in years 1800–1849. The pitch class of the root of each major chord is shown.

3.2. Improving classifer performance

Using the word-mover's distance (Kusner et al., 2015), we evaluate the accuracy of a supervised composer classifier, using a distance metric designed to capture semantic similarity between text documents. We selected binarized chroma representations of pieces ($n = 2000$) by five composers (Mozart, Bach, Beethoven, Haydn, Telemann). We trained a skip-gram word embedding model ($\mathcal{W}$) on these pieces. The binarized chroma vector representations for each piece were transformed into normalized bag-of-words (BOW) vectors, resulting in a feature space equal in size to the effective vocabulary of the examples (1501).

We trained two 1-nearest-neighbor classifers on a training set of 1000 examples, where class labels indicated the true composer of the piece. The first (baseline) classifer used Euclidean distance between BOW vectors. The second classifer used the word-mover's distance (WMD) between BOW vectors, calculated with respect to the embedding model $\mathcal{W}$. A test set of 500 examples was used to score the the classifiers. The baseline classifer had an accuracy of 54.6%. The WMD metric improves classification accuracy to 61.6%. Improved accuracy in classification tasks with the use of word-embedding models augurs well for more complex tasks such as stylistic analysis and content-based recommendation.

[1] Code to reproduce at https://github.com/eamonnbell/music-mining.

References

Bengio, Yoshua, Ducharme, Réjean, Vincent, Pascal, and Janvin, Christian. A Neural Probabilistic Language Model. *The Journal of Machine Learning Research*, 3:1137–1155, 2003. ISSN 15324435. doi: 10.1162/153244303322533223. URL http://jmlr.csail.mit.edu/papers/volume3/bengio03a/bengio03a.pdf.

Blei, David M, Ng, Andrew Y, and Jordan, Michael I. Latent Dirichlet Allocation. *Journal of Machine Learning Research*, 3:993–1022, 2003. URL http://www.jmlr.org/papers/volume3/blei03a/blei03a.pdf.

Huang, Cheng-Zhi Anna, Duvenaud, David, and Gajos, Krzysztof Z. Chordripple: Recommending chords to help novice composers go beyond the ordinary. In *Proceedings of the 21st International Conference on Intelligent User Interfaces*, IUI '16, pp. 241–250, New York, NY, USA, 2016. ACM. ISBN 978-1-4503-4137-0. doi: 10.1145/2856767.2856792. URL http://doi.acm.org/10.1145/2856767.2856792.

Jacoby, Nori, Tishby, Naftali, and Tymoczko, Dmitri. An Information Theoretic Approach to Chord Categorization and Functional Harmony. *Journal of New Music Research*, 44(3):219–244, July 2015. ISSN 0929-8215, 1744-5027. doi: 10.1080/09298215.2015.1036888.

Kusner, Matt, Sun, Yu, Kolkin, Nicholas, and Weinberger, Kilian Q. From word embeddings to document distances. In *Proceedings of the 32nd International Conference on Machine Learning (ICML-15)*, pp. 957–966, 2015.

Mikolov, Tomas, Corrado, Greg, Chen, Kai, and Dean, Jeffrey. Efficient Estimation of Word Representations in Vector Space. *ICLR*, pp. 1–12, 2013. URL http://arxiv.org/pdf/1301.3781v3.pdf.

Shanahan, Daniel and Albrecht, Joshua. The acquisition and validation of large web-based corpora. Conference for the Society for Music Perception and Cognition, Toronto, Canada, 2013.

White, Christopher Wm. and Quinn, Ian. The Yale-Classical Archives corpus. 2014. URL http://ycac.yale.edu.