WORD2VEC APPLIED TO THE YALE CORPUS

First Author
Affiliation1
author1@ismir.edu

Second Author
Retain these fake authors in
submission to preserve the formatting

Third Author
Affiliation3
author3@ismir.edu

ABSTRACT

We apply a word embedding model to a large corpus of classical music to learn an embedding space where chords are represented by real-valued vectors. The first two principal components of the embeddings of major triads are arranged on a circle. In music from earlier composers, this circular topology is more evident than in later composers who used less counterpoint. Remarkably, the order in which major triads are arranged on this structure corresponds to their order in the circle of fifths. The emergence of this structure is justified by reasoning about the probabilistic embedding model and the composition of classical music. We situate our results in the context of current statistical research into functional harmony in common practice music. We show how this technique can be used for large-scale, quantitative stylistic analysis of music.

1. INTRODUCTION

1.1 Word embeddings

Probabilistic models such as Latent Dirichlet Allocation [3] are standard tools for analyzing text data. However, such models use bag-of-words representations. Therefore they extract meaning from word co-occurrance counts on the document level, and ignore the sequential nature of language. Syntax, punctuation, and grammar are intrinsically sequential, and a good model of natural language or tree structures should be able to capture both information from co-occurrance counts and information across time. While Latent Dirichlet Allocation has been used to model music (**Q:citation?**), music is inherently sequential. For in-depth analysis of style in music, we need embedding models.

Word embeddings, or real-valued vectors representing words in a vocabulary, were first introduced by [2] but popularized by [10]. Such models typically have a log-bilinear form [11], and are trained using negative sampling with a fixed size context window [10]. This is equivalent to matrix factorization of a shifted pointwise mutual information word-context matrix [8]. To see this, construct a co-occurrance matrix of words, where the rows and columns are the words in the vocabulary. Each entry i, j in the matrix is the count of how many times word i (e.g. ‘dog’) ocurred in the context of word j (e.g. ‘the’). Word embedding models such as the skip-gram model (word2vec) can be viewed as performing singular value decomposition on a transformed version of this matrix. Compositional word embeddings for learning paragraph or document embeddings have also been proposed [4, 7]. However, [9] suggests that much of the success of these types of distributed representations of words is due to the tricks needed to train such models such as noise contrastive estimation.

As useful models of discrete data, word embedding models are starting to be used in domains outside of natural language. For example, [1] embed protein sequences for classification, and [5] develop an embedding model to build a recommendation system (for example, for recommending movies to users). There exists some prior work on applying word embedding models to music. [6] trained an embedding model on a corpus of 200 rock songs for the task of recommendating chords to composers.

1.2 Quantitative stylistic analysis of music

2. DISCUSSION

As shown in Figure **??**, we can see that the circle of fifths emerges from the structure of the learned vector space of chords in classical music. This is not an intuitive result at first. However, to see that this is a reasonable result of applying the skip-gram word2vec model (**or CBOW, whichever you used in gensim**), consider the log-likelihood of the model. Gradients of the log-likelihood with respect to the embeddings are used to train the model. The log-likelihood means the model will maximize the probability of correctly classifying the context given the training example. If the model assigns too high a probability to correct contexts, it will be overconfident on other (incorrect) contexts, and the derivative of the log-likelihood will push the embeddings further apart. But if the model assigns too low a probability to correct contexts, the gradient of the log-likelihood will flip signs and pull the embeddings closer together. The minimal geometric structure that minimizes these constraints is a circle. If the parts of the circle are perturbed (e.g. imagine shifting the values of an embedding in the circle by a large amonut), the above arguments show that it will return to a circular structure by virtue of the gradients of the objective function. We thus expect a circle from the principal components of highly stable embeddings (such as the embeddings of chords in the circle of fifths). To see why the circle of fifths respects the ordering, we consider the context window ((**size 5? in**

our case)). Chords in the circle of fifths occur in each others' contexts, but usually only nearest neighbors (e.g. it is rare to see C major followed by B major). Therefore C major and G major occur in each others' contexts and will be pushed closer together during training. But they will be pushed apart from their non-nearest-neighbors (such as B major). This shows they will respect the ordering apparent in classical music where common practices such as counterpoint result in transitions prevalent on the circle of fifths.

3. PAPER LENGTH & FILE SIZE

Instead of the strict limit of six pages (last used in ISMIR 2014), we adopt a "(6+1)-page policy" from ISMIR 2015 for ISMIR 2016. This means, the paper may have a maximum of 6 pages for technical content including figures and possible references with one additional optional 7th page containing only references. Note that this is a strict requirement. The seventh page (if used at all) must not contain any other material except for references.

Paper should be submitted as PDFs and the file size is limited to 2MB. Please compress images and figures as necessary before submitting.

4. PAGE SIZE

The proceedings will be printed on portrait A4-size paper (21.0cm x 29.7cm). All material on each page should fit within a rectangle of 17.2cm x 25.2cm, centered on the page, beginning 2.0cm from the top of the page and ending with 2.5cm from the bottom. The left and right margins should be 1.9cm. The text should be in two 8.2cm columns with a 0.8cm gutter. All text must be in a two-column format. Text must be fully justified.

5. TYPESET TEXT

5.1 Normal or Body Text

Please use a 10pt (point) Times font. Sans-serif or non-proportional fonts can be used only for special purposes, such as distinguishing source code text.

The first paragraph in each section should not be indented, but all other paragraphs should be.

5.2 Title and Authors

The title is 14pt Times, bold, caps, upper case, centered. Authors' names are omitted when submitting for double-blind reviewing. The following is for making a camera-ready version. Authors' names are centered. The lead author's name is to be listed first (left-most), and the co-authors' names after. If the addresses for all authors are the same, include the address only once, centered. If the authors have different addresses, put the addresses, evenly spaced, under each authors' name. Here's a test of a citation.

5.3 First Page Copyright Notice

Please include the copyright notice exactly as it appears here in the lower left-hand corner of the page. It is set in 8pt Times.

5.4 Page Numbering, Headers and Footers

Do not include headers, footers or page numbers in your submission. These will be added when the publications are assembled.

6. FIRST LEVEL HEADINGS

First level headings are in Times 10pt bold, centered with 1 line of space above the section head, and 1/2 space below it. For a section header immediately followed by a subsection header, the space should be merged.

6.1 Second Level Headings

Second level headings are in Times 10pt bold, flush left, with 1 line of space above the section head, and 1/2 space below it. The first letter of each significant word is capitalized.

6.1.1 Third and Further Level Headings

Third level headings are in Times 10pt italic, flush left, with 1/2 line of space above the section head, and 1/2 space below it. The first letter of each significant word is capitalized.

Using more than three levels of headings is highly discouraged.

7. FOOTNOTES AND FIGURES

7.1 Footnotes

Indicate footnotes with a number in the text. [1] Use 8pt type for footnotes. Place the footnotes at the bottom of the page on which they appear. Precede the footnote with a 0.5pt horizontal rule.

7.2 Figures, Tables and Captions

All artwork must be centered, neat, clean, and legible. All lines should be very dark for purposes of reproduction and art work should not be hand-drawn. The proceedings are not in color, and therefore all figures must make sense in black-and-white form. Figure and table numbers and captions always appear below the figure. Leave 1 line space between the figure or table and the caption. Each figure or table is numbered consecutively. Captions should be Times 10pt. Place tables/figures in text as close to the reference as possible. References to tables and figures should be capitalized, for example: see Figure 1 and Table 1. Figures and tables may extend across both columns to a maximum width of 17.2cm.

[1] This is a footnote.

String value	Numeric value
Hello ISMIR	2016

Table 1. Table captions should be placed below the table.

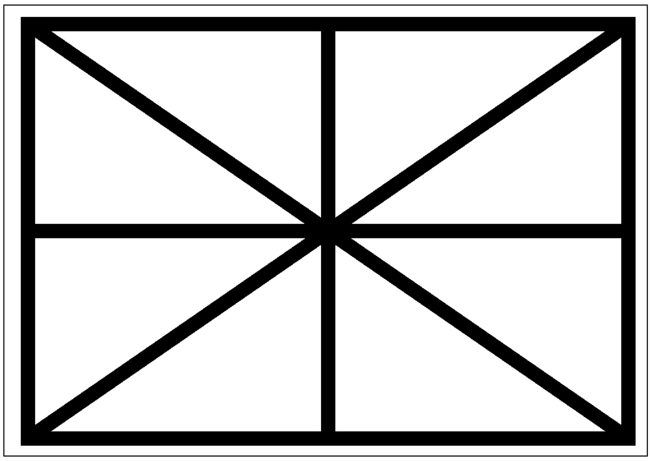

Figure 1. Figure captions should be placed below the figure.

8. CITATIONS

All bibliographical references should be listed at the end, inside a section named "REFERENCES," numbered and in alphabetical order. All references listed should be cited in the text.

9. ACKNOWLEDGMENTS

The authors are grateful to Rajesh Ranganath for helpful discussions (and Dawen Liang? and anyone else who read the paper and gave feedback). Jaan Altosaar expresses gratitude to Prof. David Blei and the Natural Sciences and Engineering Research Council of Canada for research support.

10. REFERENCES

[1] Ehsaneddin Asgari and Mohammad R K Mofrad. Continuous distributed representation of biological sequences for deep proteomics and genomics. *PLoS ONE*, 10(11):1–15, 2015.

[2] Yoshua Bengio, Réjean Ducharme, Pascal Vincent, and Christian Janvin. A Neural Probabilistic Language Model. *The Journal of Machine Learning Research*, 3:1137–1155, 2003.

[3] David M Blei, Andrew Y Ng, and Michael I Jordan. Latent Dirichlet Allocation. *Journal of Machine Learning Research*, 3:993–1022, 2003.

[4] Andrew M. Dai, Christopher Olah, and Quoc V. Le. Document Embedding with Paragraph Vectors. pages 1–8, 2015.

[5] Elie Guardia-Sebaoun, Vincent Guigue, and Patrick Gallinari. Latent Trajectory Modeling : A Light and Efficient Way to Introduce Time in Recommender Systems. *the 2015 ACM conference on Recommender systems, RecSys 2015*, pages 281–284, 2015.

[6] Cheng-Zhi Anna Huang, David Duvenaud, and Krzysztof Z. Gajos. ChordRipple : Recommending Chords to Help Novice Composers Go Beyond Themselves. pages 241–250, 2015.

[7] Qv Le and Tomas Mikolov. Distributed Representations of Sentences and Documents. *ICML*, 32:1188–1196, 2014.

[8] Omer Levy and Yoav Goldberg. Neural Word Embedding as Implicit Matrix Factorization. *NIPS*, pages 1–9, 2014.

[9] Omer Levy, Yoav Goldberg, and Ido Dagan. Improving Distributional Similarity with Lessons Learned from Word Embeddings. *Transactions of the ACL*, 3:211–225, 2015.

[10] Tomas Mikolov, Greg Corrado, Kai Chen, and Jeffrey Dean. Efficient Estimation of Word Representations in Vector Space. *ICLR*, pages 1–12, 2013.

[11] A Mnih and G Hinton. Three new graphical models for statistical language modelling. *ICML*, 62:641–648, 2007.