Hyperbolic Embeddings for Music Taxonomy

Maria Astefanoaei[*]
IT University of Copenhagen
`msia@itu.dk`

Nicolas Collignon[*]
University of Edinburgh
`nccollignon@gmail.com`

Abstract

Musical genres are inherently ambiguous and difficult to define. Even more so is the task of establishing how genres relate to one another. Yet, genre is perhaps the most common and effective way of describing musical experience. The number of possible genre classifications (e.g. Spotify has over 4000 genre tags, LastFM over 500,000 tags) has made the idea of manually creating music taxonomies obsolete. We propose to use hyperbolic embeddings to learn a general music genre taxonomy by inferring continuous hierarchies directly from the co-occurrence of music genres from a large dataset. We evaluate our learned taxonomy against human expert taxonomies and folksonomies. Our results show that hyperbolic embeddings significantly outperform their Euclidean counterparts (Word2Vec), and also capture hierarchical structure better than various centrality measures in graphs.

1 Introduction

Music genre is the most popular way to describe music - whether in the context of describing one's listening preferences, or when organising music libraries for efficient user access, exploration and discovery of new music. As is often the case with human categories, genres are richly textured and can be difficult to define explicitly. The rules defining them are ambiguous, complex and dependent on historical, genealogical or geographical factors (Aucouturier and Pachet, 2003). Representing the genre space is therefore particularly challenging. Genre taxonomies are important structures that allow us to represent the relationship between different forms of music. Knowing that *be bop* is a sub-genre of *jazz* that originated from *swing*, and later led to the development of *hard bop* can help better understand and explore the relationships between artists and their music.

For long, musicologists held the responsibility of labelling and organising genres. With the digitalisation of music and the rise of internet music consumption, online communities have shown impressive crowd-sourcing efforts in labelling and organising music by sharing structured music knowledge (e.g. DBpedia, LastFM). Beyond the practical use of a structured organisation of music, musical genres also carry an intuitive psychological reality - studies have shown it can take only a quarter of a second for a person to identify the genre of a particular track (Gjerdingen and Perrott, 2008). However, reaching an agreement on what should be considered a genre and how genres are organised amongst each other remains a difficult task. With the ever growing number of music genre labels (e.g. Spotify has over 4000 genre tags, LastFM over 500,000 tags), the effort of manually defining a complete music taxonomy is daunting, and makes capturing the full spectrum of rich interactions between genres beyond reach (Sordo et al., 2008; Pachet et al., 2000).

Historically, the most common approach to represent a music genre space has been by defining trees that capture the hierarchical structure present in genre data sets (see Figure 1).
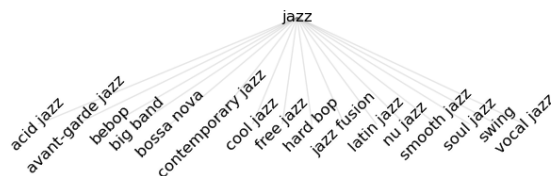

Figure 1: Tree for a subset of the subgenres of jazz.

For example, Schreiber (2015) exploited the asymmetry between main genres and sub genres to construct hierarchies, e.g. alternative is repeatedly associated with rock, whereas rock is associated with a large number of other genres. This asymmetry puts forward rock as a more general genre. Epure et al. (2019) made direct use of the DBpe-

---

[*] The two authors contributed equally.

dia music genre ontology to create a translation of music labels across taxonomies.

Embeddings are an alternative approach to capture semantic relatedness by constructing a continuous similarity space. Both approaches face shortcomings: tree-taxonomies fail to address the complexity of the genre space as genres are restricted to one category and measures of similarity between genres of different families are lost. While Euclidean embeddings allow for continuous similarity judgements between music genres, they are unable to capture latent hierarchical information. The goal of this paper is to construct a music genre taxonomy using embeddings in hyperbolic space, which offer a way of representing concepts in a continuous space while preserving hierarchy. We infer this general music genre taxonomy directly from the co-occurrence patterns of music genres across a wide and representative sample of artists.

In the first part of this paper, we introduce hyperbolic embeddings as a powerful method for inferring continuous concept hierarchies. We detail how we created the dataset of genre co-occurrence, our evaluation methods and results.

2 Embedding methods and hyperbolic space

Continuous word representations (Sordo et al., 2008; Levy and Sandler, 2007) are a widely used method to organise words/concepts, and carry many benefits for natural language processing tasks. Typically, the objective of embedding methods is to organize symbolic objects so their similarity in the embedding space reflects their semantic or functional similarity. Word embeddings such as Word2Vec (Mikolov et al., 2013), GloVe (Pennington et al., 2014) and Fasttext (Bojanowski et al., 2017) are widely used for tasks ranging from machine translation to sentiment analysis (Nickel and Kiela, 2017).

Hierarchical structures are a core feature of knowledge representations in many domains. They are used in biology to categorise and represent the relationship between animals of different species, in sociology to understand the organisational structures of groups and communities, in linguistics to describe the origin of languages etc. (Nickel and Kiela, 2018). While explicit hierarchical relations are mostly absent from large datasets, recent developments have focused on inferring this latent hierarchy directly from the data. Nickel and Kiela (2017) introduced the idea of Poincaré embeddings to learn continuous representations of hierarchies. Their model exploits the geometrical properties of hyperbolic space to capture two aspects of the relationships between embedded items: relatedness and generality; an entity is a parent of another entity if they are related and if the parent is more general than the child. These two aspects can be separated to infer concept hierarchies through hyperbolic embeddings. Relatedness is captured by the distance in the hyperbolic space, while generality is measured through the norm (smaller norm means higher generality). Due to these properties, hyperbolic spaces are particularly suited for embedding discrete trees and graphs with tree-like structure. While the idea of using hyperbolic space to represent hierarchically structured data is not new (see e.g. Lamping et al., 1995; Sarkar, 2011; Kleinberg, 2007), only recently have hyperbolic embeddings attracted the attention of the machine learning community (e.g. Nickel and Kiela, 2017; De Sa et al., 2018; Chamberlain et al., 2017; Liu et al., 2019).

Hyperbolic embeddings have been recently applied to related Music Information Retrieval tasks with success. Schmeier et al. (2019) looked at hyperbolic embeddings for music recommender systems and showed that this led to a significant increase in performance when compared to its Euclidean counterpart. Gunel et al. showed that hyperbolic embeddings could capture the artist and album relationships between tracks. To our knowledge, this paper presents the first attempt at applying hyperbolic embeddings to learn a general music genre taxonomy.

Hyperbolic geometry is a non-Euclidean geometry that emerges from relaxing Euclid's parallel postulate: for a point not on a line there can pass infinitely many lines parallel to that line. This results in a space that is in some sense larger than the Euclidean counterpart, allowing for low-dimensional embeddings with lower distortion (Sarkar, 2011). A small distortion means that most of the information in the original data is preserved in the embedding. Tree graphs are particularly suited to be represented in hyperbolic space. Sarkar (2011) shows that hyperbolic embeddings of weighted trees can preserve not only the topology of the trees but the tree metric induced by the length of the edges.

There exist multiple, equivalent models of hyperbolic space, such as the Poincaré model and the

Lorentz model. Both have specific strengths: the Poincaré model provides a very intuitive method for visualizing and interpreting hyperbolic embeddings, while he Lorentz model is well-suited for Riemannian optimization, which is particularly useful when considering large datasets. Nickel and Kiela (2018) propose a method that exploits the individual strengths of both these models by building embeddings using the Lorentz model and mapping them into the Poincaré ball. The embeddings are based on large scale, unstructured similarity scores. We use this approach to create hyperbolic embeddings of genres based on their co-occurrence scores, our proposed measure of similarity. In the next section, we explain how we constructed our dataset and the evaluation methods.

3 Creating the dataset

To construct the similarity matrix, we use genre playlists from Every Noise At Once (ENAO)[1] - a Spotify project that attempts to capture the musical genre space. Our dataset consisted of 1368 ENAO playlists, each corresponding to a unique genre, which all tracks in a playlist have in common. Across these playlists, 4309 different genres were represented (see Johnston, 2018, for an idea of how these genres are decided). The ENAO playlists are generated algorithmically and take into account audio features of tracks, user listening patterns and artist relatedness (McDonald, 2013). Spotify attaches genre labels at the artist level, on average 4.9 different genres per artist ($SD = 3.7$). We analyse the artist genre labels within each playlist. The median length of a playlist is 164 tracks ($SD = 181.0$) with at least two different genre tags, and an average of 65.4 different genres ($SD = 54.2$). We compute the similarity matrix by looking at genre co-occurrences across the artists from the tracks in the ENAO genre playlists, following the assumption that these are a fairly accurate representation of the musical genre space. In this context, the list of genres associated to artists can be seen as 'sentences' in a more typical word embedding sense. Rather than computing embeddings on the full similarity matrix ($\sim 4000 \times 4000$), we restrain our analysis to the genres matching the ones present in our evaluation datasets.

4 Evaluating Genre Taxonomies

To evaluate our embedding, we compare it to taxonomies defined explicitly by music experts and taxonomies inferred from user annotations (also referred to as *folksonomies*). Four of the five datasets used in the experiments were based on the datasets used in the 2018 AcousticBrainz Genre Task, part of the MediaEval benchmarking initiative (Bogdanov et al., 2017). The dataset in its original form was aimed at testing the automatic genre annotation from content-based features of musical items. These include the All Music dataset, Discogs, LastFM and Tagtraum. The fifth dataset we used was the FMA dataset (Defferrard et al., 2017). Each dataset consists of a set of trees representing different music genres (Figure 1 shows an example from the LastFM dataset).

Dataset	Trees	Genres	Depth	Annotation
FMA	11	68	2	Expert
Allmusic	16	322	2	Expert
Discogs	9	204	1	Expert
LastFM	15	211	1	User
Tagtraum	17	167	1	User

Table 1: Description of the evaluation datasets

We normalise the genre labels across datasets following the rules proposed by Schreiber (2015) and Geleijnse et al. (2007), which included capitalization, spelling standardization, tokenization, and concatenation of the strings. After normalising the genre labels, we find the closest match for each genre from the test set to the genres in the Everynoise playlists. The string matching is based on the Levenshtein distance, which counts the number of edits needed to transform one string to another[2]. We keep genre labels that are above a 0.90 similarity threshold, leaving us with 503 unique genres. Of those, 94% were exact matches.

We use Word2Vec, a popular word embedding method, as baseline (Mikolov et al., 2013). It creates vector representations based on word contexts using shallow neural networks. In our implementation, we consider the genres associated with each track as a sentence, giving us a context for a genre.

5 Results

Using our proposed co-occurrence counts as a pairwise similarity measure between genres, we compute hyperbolic embeddings based on the Lorentz model, following the method proposed by Nickel and Kiela (2018). An example of a 2D embedding mapped to the Poincaré disc is shown in Figure 3. We highlight the genres present in the LastFM trees dataset. Even in 2d, we find the

---

[1] http://everynoise.com/

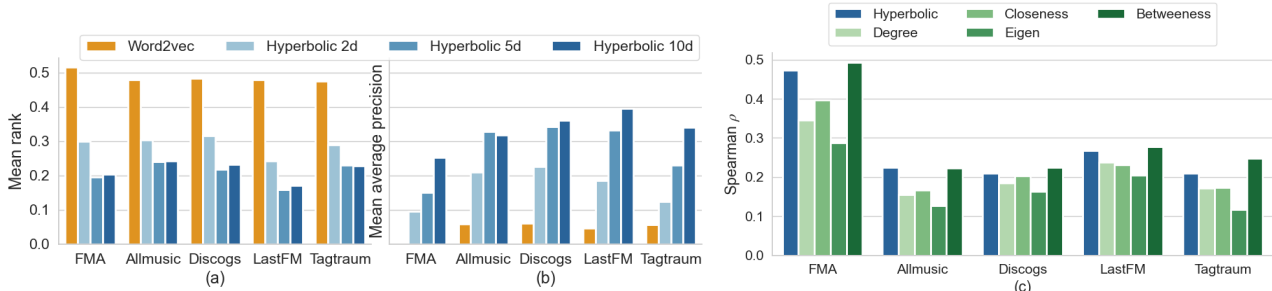[2] We use the FuzzyWuzzy Python package (Cohen, 2011)

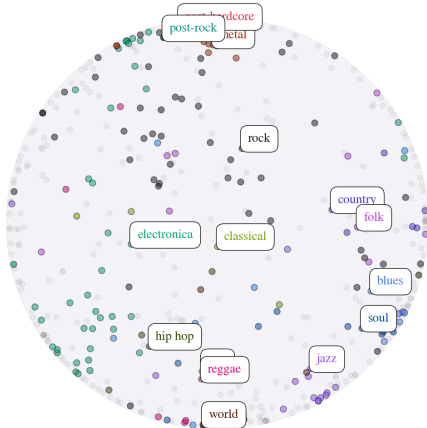Figure 2: Comparison with Word2vec embeddings and graph centrality measures.

Figure 3: Embedding of the 2d hyperbolic coordinates for genres in the LastFM dataset. All nodes in one tree have the same color. We highlight the the parent genre.

embedding captures the similarity between genres well and highlights meaningful clusters. Parent nodes are typically closer to the center of the disc than their children, suggesting that the embedding successfully represents hierarchical relations. Moreover, we observe natural transitions between clusters of genres, such as $jazz \rightarrow soul \rightarrow blues \rightarrow folk \rightarrow country$; these types of transitions cannot easily be captured with disjoint tree representations. Next, we show that hyperbolic embeddings outperform Word2Vec embeddings significantly when evaluated against human expert taxonomies and folksonomies.

To measure the quality of the embedding, for each observed edge (u, v) we compute the corresponding distance $d(u, v)$ in the embedding and rank it among the distances of all unobserved edges for u, i.e., among $\{d(u, v') : (u, v') \notin \mathcal{D}\}$. We then report the mean rank (MR) and mean average precision (MAP) of this ranking. We compare 2D, 5D and 10D hyperbolic embeddings[3] to Word2vec (100D)[4]. Hyperbolic embeddings perform better on both metrics regardless of dimension, with the 10d embeddings giving up to 8 fold

---

[3]We rely on the following implementation: github.com/theSage21/lorentz-embeddings

[4]We use the *gensim* (Řehůřek and Sojka, 2010) package

precision improvement and 2 fold rank improvement when compared to Word2Vec (Figure 2a,b).

To evaluate how hierarchy is captured by the hyperbolic embedding, we compare the embedding norms with commonly used centrality measures on graph: degree, closenesss centrality, eigenvalue centrality and betweeness centrality. Specifically, we evaluate how they correlate to the ranks in the evaluation trees. Across all datasets, we find that the norm of the hyperbolic embeddings is on par with betweeness centrality and consistently outperforms the other centrality measures (Figure 2c). The trees across evaluation datasets are shallow (mostly depth 1, and exceptionally 2) while the embedding gives more granularity of the hierarchy. The shallowness may explain the generally low correlation scores across measures. As further evaluation, we compute the ratio of parent-child relations in trees that were preserved in the embedding (i.e. the parent norm is lower than the child norm). We find that for 2D, 5D, and 10D embeddings, at least 80% of the parent-child were correctly preserved.

6 Conclusion

In this paper, we inferred a general musical genre taxonomy from a large dataset of playlists. We used Lorentz embeddings to learn continuous hierarchies directly from the co-occurrence patterns of genres across tracks. We evaluated our learned taxonomy against human expert taxonomies and folksonomies and found that hyperbolic embeddings significantly outperform their Euclidean counterparts, while also capturing hierarchy better than a number of centrality measures in graphs. Beyond their direct usefulness for computational studies of music, and domains such as music recommendation or genre classification, these results present hyperbolic embeddings as a powerful tool to study other human classification systems, and perhaps for the study of their corresponding psychological representations.

References

Jean-Julien Aucouturier and Francois Pachet. 2003. Representing musical genre: A state of the art. *Journal of New Music Research*, 32(1):83–93.

Dmitry Bogdanov, Alastair Porter, Julián Urbano, and Hendrik Schreiber. 2017. The mediaeval 2017 acousticbrainz genre task: Content-based music genre recognition from multiple sources. CEUR Workshop Proceedings.

Piotr Bojanowski, Edouard Grave, Armand Joulin, and Tomas Mikolov. 2017. Enriching word vectors with subword information. *Transactions of the Association for Computational Linguistics*, 5:135–146.

Benjamin Paul Chamberlain, James Clough, and Marc Peter Deisenroth. 2017. Neural embeddings of graphs in hyperbolic space. *arXiv preprint arXiv:1705.10359*.

Adam Cohen. 2011. Fuzzywuzzy: Fuzzy string matching in python. *ChairNerd Blog*, 22.

Christopher De Sa, Albert Gu, Christopher Ré, and Frederic Sala. 2018. Representation tradeoffs for hyperbolic embeddings. *Proceedings of machine learning research*, 80:4460.

Michaël Defferrard, Kirell Benzi, Pierre Vandergheynst, and Xavier Bresson. 2017. FMA: A dataset for music analysis. In *18th International Society for Music Information Retrieval Conference (ISMIR)*.

Elena V Epure, Anis Khlif, and Romain Hennequin. 2019. Leveraging knowledge bases and parallel annotations for music genre translation. *arXiv preprint arXiv:1907.08698*.

Gijs Geleijnse, Markus Schedl, and Peter Knees. 2007. The quest for ground truth in musical artist tagging in the social web era. In *ISMIR*, pages 525–530. Citeseer.

Robert O. Gjerdingen and David H. Perrott. 2008. Scanning the dial: The rapid recognition of music genres. *Journal of New Music Research*, 37:100 – 93.

Beliz Gunel, Fred Sala, Albert Gu, and Christopher Ré. Hypere: Hyperbolic embeddings for entities.

Maura Johnston. 2018. Trap queen and the data scientist: How a subgenre is born.

Robert Kleinberg. 2007. Geographic routing using hyperbolic space. In *IEEE INFOCOM 2007-26th IEEE International Conference on Computer Communications*, pages 1902–1909. IEEE.

John Lamping, Ramana Rao, and Peter Pirolli. 1995. A focus+ context technique based on hyperbolic geometry for visualizing large hierarchies. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 401–408.

Mark Levy and Mark Sandler. 2007. A semantic space for music derived from social tags. *Austrian Compuer Society*, 1:12.

Qi Liu, Maximilian Nickel, and Douwe Kiela. 2019. Hyperbolic graph neural networks. In *Advances in Neural Information Processing Systems*, pages 8230–8241.

Glenn McDonald. 2013. How we understand music genres.

Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*, pages 3111–3119.

Maximilian Nickel and Douwe Kiela. 2018. Learning continuous hierarchies in the lorentz model of hyperbolic geometry. *arXiv preprint arXiv:1806.03417*.

Maximillian Nickel and Douwe Kiela. 2017. Poincaré embeddings for learning hierarchical representations. In *Advances in neural information processing systems*, pages 6338–6347.

François Pachet, Daniel Cazaly, et al. 2000. A taxonomy of musical genres. In *RIAO*, pages 1238–1245.

Jeffrey Pennington, Richard Socher, and Christopher D Manning. 2014. Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pages 1532–1543.

Radim Řehůřek and Petr Sojka. 2010. Software Framework for Topic Modelling with Large Corpora. In *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*, pages 45–50, Valletta, Malta. ELRA. http://is.muni.cz/publication/884893/en.

Rik Sarkar. 2011. Low distortion delaunay embedding of trees in hyperbolic plane. In *International Symposium on Graph Drawing*, pages 355–366. Springer.

Timothy Schmeier, Joeseph Chisari, Sam Garrett, and Brett Vintch. 2019. Music recommendations in hyperbolic space: an application of empirical bayes and hierarchical poincaré embeddings. In *Proceedings of the 13th ACM Conference on Recommender Systems*, pages 437–441.

Hendrik Schreiber. 2015. Improving genre annotations for the million song dataset. In *ISMIR*, pages 241–247.

Mohamed Sordo, Oscar Celma, Martin Blech, and Enric Guaus. 2008. The quest for musical genres: Do the experts and the wisdom of crowds agree? In *ISMIR*, pages 255–260.