Machine improvisation through generalized transition probability graphs

Mattia Barbaresi[1,*], Andrea Roli[1,2]

[1]Department of Computer Science and Engineering, Campus of Cesena, Università di Bologna
[2]European Centre for Living Technology, Venezia, Italy

Abstract
Improvisation plays a cardinal role in the arts and is acknowledged to be a typical manifestation of creativity. In performing arts, an impromptu consists in playing extemporaneous sequences of actions, i.e. notes or movements, in accordance to some rules and constraints. Typically, a good improviser masters those constraints and can produce meaningful paths in the feasible space of allowed actions and can also explore some areas in the adjacencies of this space. From a computational perspective, one of the possible ways to capture this creative production is to make use of statistical learning mechanisms, which are also believed to be involved in human musical improvisation. At the basis of statistical learning are transitional probabilities between segments of a sequence and their following segments of symbols. In this paper we present preliminary results of a statistical learning model in which a transitional probability graph is computed from a set of sample pieces of music. This graph is subsequently generalized by applying a node similarity mechanism. This generalized graph is used for generating melodies that resemble improvisations in a given musical style.

1. Introduction

According to the *Grove Music Online* [1], improvisation is "The creation of a musical work, or the final form of a musical work, as it is being performed. It may involve the work's immediate composition by its performers, or the elaboration or adjustment of an existing framework, or anything in between. To some extent every performance involves elements of improvisation, although its degree varies according to period and place, and to some extent every improvisation rests on a series of conventions or implicit rules.". We emphasize here that the notion of improvisation involves the extemporaneous creation of sequences of notes (i.e., pitches and durations, including dynamic and agogic expressions) performed according to shared, implicit and explicit, conventions and rules. Another important property that characterizes improvisation is *risk*, i.e. " the need to make musical decisions on the spur of the moment, or moving into unexplored musical territory with the knowledge that some form of melodic, harmonic, or ensemble closure will be required." [1]. Therefore, the act of improvising requires the capability of balancing the adherence to the rules that have been learned and an ingenious exploration outside their boundaries.

CREAI2022: Workshop on Artificial Intelligence and Creativity
*Corresponding author.
mattia.barbaresi@unibo.it (M. Barbaresi); andrea.roli@unibo.it (A. Roli)
0000-0003-2684-5311 (M. Barbaresi); 0000-0001-9891-5441 (A. Roli)

Recent works address musical improvisation in the context of statistical learning [2, 3]. Generally speaking, Statistical Learning (SL) is the ability of the brain to grasp regularities of the environment in an autonomous and unsupervised way, often without awareness [4], and it is considered a cornerstone of cognition [5]. Moreover, besides language and music, SL is ubiquitous over modalities and species [6]. It mainly involves the detection of transitional probabilities (TPs): seminal experiments exploring this phenomenon in the acquisition of spoken language showed that infants are sensitive to TPs of syllables in a continuous speech stream [7].

Inspired by the SL literature, especially regarding computational approaches like [8], we are developing a model for emulating implicit sequential learning and creativity. Here, we take the opportunity to show, in particular, the effects of generalization on produced sequences. In this work, we illustrate a SL mechanism that creates melodic improvisations by performing a stochastic walk on a generalized graph of TPs. The use of the generalized graph makes it possible to combine both the adherence to a given set of implicitly learned rules and a cautious exploration outside those conventions. In Section 2 we describe the model and the creative algorithm, while results are illustrated in Section 3. We conclude with discussing further improvements and future perspectives of this approach.

2. Model and algorithm

In implicit sequence learning, such as in language, music or movements, initial acquisition of implicit sequences may arise from SL [9]. In addition, previous studies suggested that musical creativity in some measure depends on SL [3], and that implicit knowledge governs music acquisition [10]. Drawing upon these perspectives, we wanted to grasp the implicit aspects of a creative process in a minimal model capable of learning and generating (musical) sequences. Hence, the basic idea is to exploit the implicitly learned knowledge to produce novel musical strings. In addition, we also provided a generalization step from this implicit knowledge, to acquire structured information from the context.

The algorithm proceeds through three subsequent phases: learning, generalization, and generation (see Figure 1). In the learning phase, we introduced TPs at two specific levels: between symbols, as a cue for segmenting the incoming input into small segments (or chunks), and between these formed chunks. After the learning phase, the graph of TPs between chunks undergoes a generalization phase. This phase draws on the distributional learning hypothesis [11, 12] which argues that people use statistical learning to acquire grammatical categories from the input (i.e., the contextual information surrounding a word). Indeed, by relying solely on distributional information (i.e., contextual information in the graph), this mechanism exploits node similarity (SimRank) to reveal these categories (namely, form classes in language).

Finally, this new generalized graph is employed to generate novel, structured sequences using an ad hoc Monte Carlo creative walk.

The model can be applied both on unsegmented and segmented corpora.

2.1. Learning

The learning phase consists of two mechanisms: tracking the transitional probabilities between symbols (second order TPs) to be used as cues to segment the input into words (or units,

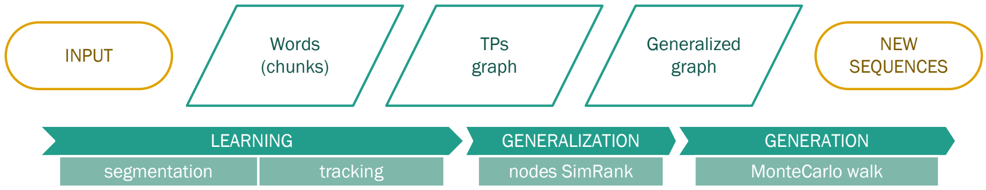

Figure 1: Sketch of the entire process

chunks), and tracking TPs between those words (first order transitions) to form a graph made of transitions between chunks.

At each perception cycle, TPs between the observed symbols are stored. Initially, the algorithm tries to use stored TPs to find a drop in the transitions between symbols that would determine the boundary of a word. On the other hand, if no TPs cue is found, a syllable is perceived (two consecutive symbols). The segmentation strategy used in this work is one of the simplest where a boundary is detected if the transitional probability of the upcoming symbol drops under a certain threshold, so if $TP_i > TP_{i+1} + \epsilon$. In the present study, we used $\epsilon = 0.2$ as an empirically selected threshold. However, various strategies could be exploited (see [13] for an old but detailed analysis): recent studies, for example, suggest the use of backward TPs [14, 15], but this is out of scope here.

After segmentation, TPs between the resulting ordered units are recorded. Note that this *per se* represents an abstraction intended to grasp the dynamics, the transitions, between formed words—not between symbols.

2.2. Generalization

The output of the learning phase is a graph where nodes represent units/words and edges represent transitional probabilities between words. To construct the generalized graph, the procedure first computes the form classes, using similarity between nodes, and then generates some sequences (with the TPs graph) that are parsed to build the higher-level graph. The similarity between nodes is computed using a SimRank [16] measure over inward and outward edges. SimRank is a graph-theoretic measure that says "two objects are considered to be similar if they are referenced by similar objects". In this case, we used a slightly modified version where "two objects are considered to be similar if they are referenced by similar objects . . . and refer to similar objects". That is, nodes are grouped if they have similar inward and outward edges, so if they have a similar neighborhood. Precisely, we group nodes with both inward and outward SimRank greater than a threshold value γ. In this case, we used $\gamma = 0.5$ as, in our experiments, it provided convincing similarity values over known samples. So for each node i we calculate $FC_i = \{I_i \cap O_i\}$ where, for each node j:

$$I_i = \{N_j : SimRank_{IN}(N_j, N_i) \geq \gamma\}$$

$$O_i = \{N_j : SimRank_{OUT}(N_j, N_i) \geq \gamma\}$$

The formed groups represent what in language acquisition is called form classes [17]. Once calculated, the form classes are used to parse some generated sequences (using the TPS graph), and the new generalized graph is then built. Transitional probabilities between formed (form) classes are computed as well, counting transitions over the parsed sequences.

2.3. Generation

The generalized graph is then used to produce novel sequences. In the present experiment, we opted for a simple Monte Carlo choice over the edge probabilities to traverse the graph. At each visited node, as in general it may contain words that can be used in the same position in the construction, a word is picked randomly (the nodes that contain alternative words are called here *choice nodes*). Another possibility for selecting a word is to use the weights (the frequencies) of the words to employ another Monte Carlo choice at each node.

3. Results

A prominent context for improvisation is of course music. Since the system we developed is mainly focused on sequences of symbols, we opted for melodic pieces of music. Therefore, we provided the system a set of melodies belonging to a given style (e.g. Irish music) upon which the TPs graph and the generalized graph can be built. The latter provides then the basis for the generation of new melodies in the style of the repertoire provided, but with variations and explorations in the implicit boundaries set by the examples. The resulting melodies are characterized by improvisation flavor, as they have not the structure of a complete piece of music, but capture the main stylistic features of the original compositions, like a musician making extemporaneous explorations around a given style.

To test the system we chose two different styles: Irish melodies and the six preludes from solo cello sonatas by J.S. Bach. Irish melodies have been retrieved from Henrik Norbeck's abc tunes [18]. All the 136 melodies in the key of G have been gathered (including variations of the same song) and the abc notation symbols, which encode the music in textual form, have been directly used as sequence symbols. The second repertoire of melodic music, instead, has been retrieved in MIDI format from David J. Grossman's J.S. Bach page [19]; the MIDI files have been converted to an intermediate textual representation by means of PyPianoroll [20] and transposed to the same key, so as to have sequences composed of symbols representing the intervals from a common base note. In both the corpora of examples, a symbol in a sequence represents both pitch and duration.

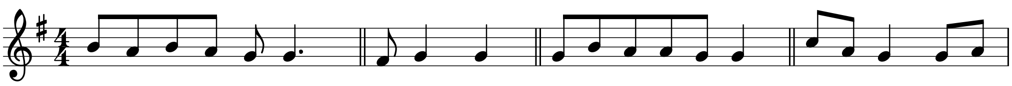

Figure 2: An example of the options for melodic segments in a choice node of the generalized graph.

We are interested here in the features of the generalized graph and the characteristics of the melodies it produces. The number of nodes containing alternative choices and the number of choices estimate the amount of "controlled exploration" around the musical style learned. For example, a typical choice node in the generalized graph of Irish music may have the following alternatives: B A B A G G3 | F G2 G2 | G B A A G G2 | c A G2 G A, represented in Figure 2 in score notation. In general, the segments differ in start and end note, as well as total duration; therefore, they do not represent equivalent alternatives, but rather different sub-paths that can be used to compose a new path which is likely to combine fragments of melodies in an original way, yet keeping the flavor of the melodies in the repertoire. The generalized graph built from Irish music has 151 nodes, of which 19 are choice nodes. The choices in each node are distributed between 2 and 8, with a median of 3. The resulting melodies are similar to the ones in the repertoire, but characterized by a considerable degree of originality. The interested reader can find audio excerpts and score transcriptions at [21].

The generalized graph of Bach preludes for cello solo substantially differs from the one related to Irish music, as it it composed of a greater number of nodes (526) and a lower number of choice nodes (10), all with just 2 choices except one with 4 choices. Another remarkable difference with respect to the previous case is that the melodic segments in each choice node are longer. The musical difference between the two repertoires is wide and this has of course strong impact on the properties of the generalized graph. Irish traditional music is characterized by simple elements: almost all the notes used belong to the scale of G major and the maximal difference in pitch is about two octaves. Moreover, the melodies are often composed of long sequences of notes at small intervals and few large steps (e.g. of an octave or a fifth). Conversely, the preludes for cello solo by Bach span a wider range of pitches and the use of chromatisms is extremely common. In addition, the examples available are much less than the Irish ones, so the probability of overlaps between portions of melodies is much lower. The features of the two graphs reflect the musical properties of the two styles in that the richness of Bach's style and, above all, the hierarchical structure of his compositions limit the adjustable interchangeability of melodic segments which is expressed by the generalized graph. However, the musical result of artificial improvisations in the style of Bach's cello preludes is appreciable (audio and score excerpts are available at [21]).

4. Discussion and future work

We presented a generative model that uses a stochastic walk on a topological generalization of variable Markov Chains (TPs graph), to produce novel musical sequences. The presented work is intended to be a seminal, basic module of a more extensive system conceived for emulating the learning of implicit sequences. It is intentionally domain-general and symbolic since it is intended to model various phenomena: from music and language to movements and social interactions [22]. This learning system assimilates implicit knowledge that becomes the basis for modeling implicit, automatic behaviors. In these regards, we envision adding a short memory module, to model higher-level phenomena such as attention, for example. However, even if in this case the focus was on the learning system, the ultimate goal is, in fact, that of producing creative outputs. In this perspective, the next step will be to use an ad hoc, creative Monte

Carlo walk in place of the simpler stochastic one, that is, to give the model the ability to explore creative paths instead of the most (or the least) probable ones. We believe a model built in this way could also provide a place, an environment, for simulating and studying a variety of behaviors in cognitive science and creativity.

References

[1] B. Nettl, R. C. Wegman, I. Horsley, M. Collins, S. A. Carter, G. Garden, R. E. Seletsky, R. D. Levin, W. Crutchfield, J. Rink, P. Griffiths, B. Kernfeld, Improvisation, in: Grove Music Online, Oxford University Press, 2001. doi:10.1093/gmo/9781561592630.article.13738, retrieved 3 Oct. 2022.

[2] T. Daikoku, Musical creativity and depth of implicit knowledge: Spectral and temporal individualities in improvisation, Frontiers in Computational Neuroscience 12 (2018) 89.

[3] T. Daikoku, Entropy, uncertainty, and the depth of implicit knowledge on musical creativity: Computational study of improvisation in melody and rhythm, Frontiers in Computational Neuroscience 12 (2018) 97.

[4] B. E. Sherman, K. N. Graves, N. B. Turk-Browne, The prevalence and importance of statistical learning in human cognition and behavior, Current opinion in behavioral sciences 32 (2020) 15–20.

[5] E. S. Isbilen, S. M. McCauley, M. H. Christiansen, Individual differences in artificial and natural language statistical learning, Cognition 225 (2022) 105123.

[6] C. Santolin, J. R. Saffran, Constraints on statistical learning across species, Trends in Cognitive Sciences 22 (2018) 52–63.

[7] J. R. Saffran, R. N. Aslin, E. L. Newport, Statistical learning by 8-month-old infants, Science 274 (1996) 1926–1928.

[8] S. M. McCauley, M. H. Christiansen, Language learning as language use: A cross-linguistic model of child language development., Psychological review 126 (2019) 1.

[9] Y. Du, J. E. Clark, New insights into statistical learning and chunk learning in implicit sequence acquisition, Psychonomic bulletin & review 24 (2017) 1225–1233.

[10] M. Rohrmeier, P. Rebuschat, Implicit learning and acquisition of music, Topics in cognitive science 4 (2012) 525–553.

[11] P. A. Reeder, E. L. Newport, R. N. Aslin, From shared contexts to syntactic categories: The role of distributional information in learning linguistic form-classes, Cognitive psychology 66 (2013) 30–54.

[12] T. H. Mintz, Category induction from distributional cues in an artificial language, Memory & Cognition 30 (2002) 678–686.

[13] D. Dahan, M. R. Brent, On the discovery of novel wordlike units from utterances: an artificial-language study with implications for native-language acquisition., Journal of Experimental Psychology: General 128 (1999) 165.

[14] P. Perruchet, What mechanisms underlie implicit statistical learning? transitional probabilities versus chunks in language learning, Topics in cognitive science 11 (2019) 520–535.

[15] J. R. Saffran, N. Z. Kirkham, Infant statistical learning, Annual review of psychology 69 (2018) 181–203.

[16] G. Jeh, J. Widom, Simrank: a measure of structural-context similarity, in: Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining, 2002, pp. 538–543.

[17] J. F. Schwab, K. D. Schuler, C. M. Stillman, E. L. Newport, J. H. Howard Jr, D. V. Howard, Aging and the statistical learning of grammatical form classes., Psychology and Aging 31 (2016) 481.

[18] H. Norbeck's abc tunes, http://www.norbeck.nu/abc/, last accessed February 2021.

[19] D. J. Grossman's J.S. Bach page, http://www.jsbach.net, last accessed August 2022.

[20] H.-W. Dong, W.-Y. Hsiao, Y.-H. Yang, Pypianoroll: Open source python package for handling multitrack pianorolls, in: Late-Breaking Demos of the 19th International Society for Music Information Retrieval Conference, 2018.

[21] Music excerpts for the paper titled "Machine improvisation through generalized transition probability graphs" by M. Barbaresi and A. Roli, https://tinyurl.com/n4bc6x73, last accessed October 2022.

[22] C. Monroy, M. Meyer, S. Gerson, S. Hunnius, Statistical learning in social action contexts, PloS one 12 (2017) e0177261.