ESANN 2025 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 23-25 April 2025, i6doc.com publ., ISBN 9782875870933. Available from http://www.i6doc.com/en/.

Motif-augmented classical music synthesis via recurrent neural networks

Alexandru-Ion Marinescu

Babeş-Bolyai University - Department of Mathematics and Computer Science
Str. Mihail Kogălniceanu, nr. 1, Cluj-Napoca 400084 - Romania

Abstract. We propose a motif-augmented approach to classical music synthesis using LSTM-based recurrent neural networks trained on J. S. Bach violin compositions. By combining motif augmentation with temperature-based sampling, we improve the entropy alignment between generated sequences and ground-truth data. Our experiments show that motif augmentation significantly reduces entropy deviation and enhances sequence coherence, as confirmed by statistical analysis. This method advances generative music modeling, offering potential applications in music composition and sequence prediction tasks.

1 Introduction and research outline

Since the dawn of mankind, we have taken inspiration from nature and adopted and made use of what nature had to offer us. Almost, if not all, of our discoveries are mere efforts to understand the inner workings of our surroundings. Music is by no means an exception, having its origin in the way animals communicate, be it with their offspring, with each other, or as a way to impress a potential partner.

In the contents of this paper we wish to explore the feasibility of treating musical compositions as a form of natural language expression. It serves as a direct continuation of our research in this field [1] and focuses on the possibility of training an LSTM-type recurrent neural network on a composition data set by J. S. Bach and then improving the quality of the predictions by introducing a technique which we have called *motif sampling*. Repetition and subtle alteration are key factors in writing a musical piece. By incorporating these so-called *motifs* (small sequences of notes with a high occurrence count in the composer's musical corpus), we successfully keep in line with the composer's original work. The quality measure that we will employ during our experiments is the *entropy* of the sequence of notes that make up the musical composition.

Judging from a sample musical sheet, the bare minimum necessary for mapping a musical composition to natural language is that a musical unit must be described by 2 attributes: *unit duration* (whole, half, quarter and subdivisions at least up to a sixteenth of a note) and *unit position* on the musical sheet (also known as pitch, with support for flat or sharp note alterations). Thankfully, there exist well-established open source libraries for extracting this information from MIDI files (`https://pypi.org/project/music21/`), which the authors have used extensively in their research (`https://github.com/Zerseu/bach21`).

ESANN 2025 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 23-25 April 2025, i6doc.com publ., ISBN 9782875870933. Available from http://www.i6doc.com/en/.

2 State of the art in literature

In contemporary music synthesis, deep learning architectures have advanced the integration of musical motifs, recurring thematic elements that provide coherence and identity to compositions. These developments focus on effectively representing, generating, and manipulating motifs to enhance the structural and aesthetic quality of synthesized music.

A significant challenge in computational music is to capture the implicit relationships between the motifs and their variations. Traditional methods often struggle with the diversity and subtlety of motif transformations. To address this, researchers have employed representation learning techniques, such as Siamese networks combined with regularization-based methods such as Variance-Invariance-Covariance Regularization (VICReg). This approach enables models to learn nuanced motif representations, facilitating tasks such as motif retrieval and providing a foundation for motif-based music generation [2].

Deep learning models have been developed to generate music with long-term coherent structures focusing on hierarchical representations. For instance, frameworks like MusicFrameworks decompose music generation into multiple stages, including high-level structural planning and detailed content creation. This method allows for the generation of melodies guided by motifs, rhythmic patterns, and harmonic progressions, resulting in compositions that exhibit both local and global coherence [3].

An emerging trend involves breaking down the music generation process into subtasks such as extraction, variation, and the development of the motif. This decomposition enables models to handle complex musical structures more effectively. By integrating musical knowledge and neurosymbolic methods, these approaches facilitate the generation of music that reflects human-like thematic development and structural integrity [4].

Given the repetitive nature of motifs, some models utilize discrete neural representations to generate musical loops. By learning discrete latent codes, these models can produce loops that capture the essence of motifs, allowing the creation of music with consistent thematic elements. This technique improves both the fidelity and diversity of the music generated, particularly in genres that rely heavily on repetitive structures [5].

In summary, the current state of the art in applying musical motifs to augment music synthesis using deep learning architectures involves sophisticated representation learning, hierarchical modeling, subtask decomposition, and the use of discrete neural representations. These approaches collectively contribute to the generation of music that exhibits coherent and engaging thematic development.

3 Experimental setup and numerical results

The entropy of a signal consisting of a sequence of numbers measures the amount of uncertainty or randomness present in that sequence. In information theory,

ESANN 2025 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 23-25 April 2025, i6doc.com publ., ISBN 9782875870933. Available from http://www.i6doc.com/en/.

entropy quantifies the average amount of information produced by a stochastic data source [6, 7]. For a discrete signal, the entropy $H(X)$ is defined using the probabilities of occurrence of each possible value in the signal sequence. If X is a random variable representing the possible values of the signal and $p(x_i)$ is the probability of the value x_i occurring, the entropy is given by the following:

$$H(X) = -\sum_i p(x_i) \log_2 p(x_i) \tag{1}$$

Thus, a high entropy indicates high randomness and less predictable signal, whereas a low entropy suggests low randomness and greater predictability of the signal [8]. The entropy of a sequence of musical pitches can be a compelling indicator of the qualitative characteristics of a composition for several reasons [9, 10].

Our experiment investigates the impact of motif augmentation combined with Andrej Karpathy's temperature sampling mechanism [11, 12] on the entropy of sequences generated by an LSTM architecture (Fig. 1 and Alg. 1). The temperature parameter is applied to the output probabilities of the model before sampling the next token. A higher temperature increases the likelihood of picking less probable and more surprising tokens, while a lower temperature favors the most likely tokens and thus more predictable output. The goal is to determine whether motif augmentation improves the alignment of the generated sequence entropy with the theoretically expected entropy, derived from ground truth data (Fig. 2).

The data set comprises note pitch sequences extracted from MIDI [13] files from J. S. Bach compositions, specifically focusing on the violin parts. Using the music21 library, musical units were reduced to symbolic representations of pitch, discarding attributes such as note duration and dynamics. This preprocessing produces a sequence of standard pitch notations (that is, "C5" for the C note in the fifth octave), preserving the temporal order and structural patterns inherent to Bach's work. For training, an LSTM-based recurrent neural network was used with sequences of 16 consecutive pitches as input (number of steps = 16). The model, with a hidden layer of 256 units, was optimized using mini-batches of size 8 and trained over 50 epochs. Using this data set, the LSTM is trained to predict the next pitch in a sequence, effectively learning the compositional rules embedded in Bach's music. The motifs themselves were *precomputed* by querying the entire corpus with all possible subsequences of pitches of lengths 4 to 8 (using regular expressions for performance considerations). A subsequence is considered a motif if it appears *more than once*.

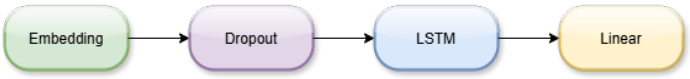

Fig. 1: Architecture of the recurrent LSTM model under scrutiny.

ESANN 2025 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 23-25 April 2025, i6doc.com publ., ISBN 9782875870933. Available from http://www.i6doc.com/en/.

Algorithm 1 Temperature sampling ([11]) and motif augmentation pseudo-code for note sequence prediction. Based on our Python implementation, using Torch as back-end. The motif dictionary can be precomputed once for the composer's entire corpus and serialized for later use.

Input:
  $pred$ ▷ Prediction tensor from the model
  $temp$ ▷ Temperature for scaling logits
  $model$ ▷ Trained sequence generation model
  seq ▷ Input sequence (list of token indices)
  $motifs$ ▷ Dictionary of motifs for augmentation
  $motif\_threshold$ ▷ Threshold for motif augmentation
  $motif\_augmentation$ ▷ Boolean flag for augmentation

function TempSample($pred$, $temp$)
  $pred \leftarrow \text{detach}(pred)$ ▷ Detach tensor from computation graph
  $pred \leftarrow \exp(pred/temp)$ ▷ Scale logits by temperature and exponentiate
  $pred \leftarrow pred/\text{sum}(pred)$ ▷ Normalize to obtain probabilities
  $prob \leftarrow$ random multinomial sample from $pred$
  return $(\max(pred), \text{argmax}(prob))$ ▷ Return max probability and its index
end function

function TempPredict($model$, seq, $temp$)
  $input \leftarrow \text{reshape}(seq \text{ as } (1, -1))$ ▷ Prepare input for model
  $pred \leftarrow model(input)$ ▷ Forward pass through the model
  return TempSample($pred$, $temp$)
end function

function MotifPredict($motifs$, $model$, seq, $temp$)
  $(prob\_max, prob\_argmax) \leftarrow$ TempPredict($model$, seq, $temp$)

  if $prob\_max > motif\_threshold$ **or** $\neg motif\_augmentation$ **then**
    return $(prob\_max, prob\_argmax)$
  end if

  for all $motif \in \text{keys}(motifs)$ **do**
    $motif \leftarrow \text{map\_direct}(motif)$ ▷ Map motif words to indices
    $length \leftarrow \min(\text{len}(seq), \text{len}(motif) - 1)$
    if last $length$ tokens of seq = last $length$ tokens of $motif$ **then**
      return $(1.0, \text{last token of } motif)$
    end if
  end for

  return $(0.0, prob\_argmax)$
end function

Without motif augmentation, the average difference between the generated sequence entropy and the expected entropy is larger than that for sequences generated with motif augmentation. With motif augmentation, the average difference is consistently smaller, indicating a closer alignment to the expected entropy (Fig. 3). A paired t-test comparing the average entropy values for sequences with and without motif augmentation yielded a statistically significant p-value (below the conventional threshold of 0.05). This demonstrates that the difference in entropy alignment between the two methods is not due to random variation but is likely attributable to the use of motif augmentation. Across the 10 trials, motif-augmented sequences exhibited smaller variations in entropy, suggesting greater stability and robustness in generating sequences closer to the expected entropy. Although not explicitly detailed in this summary, the effect

ESANN 2025 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 23-25 April 2025, i6doc.com publ., ISBN 9782875870933. Available from http://www.i6doc.com/en/.

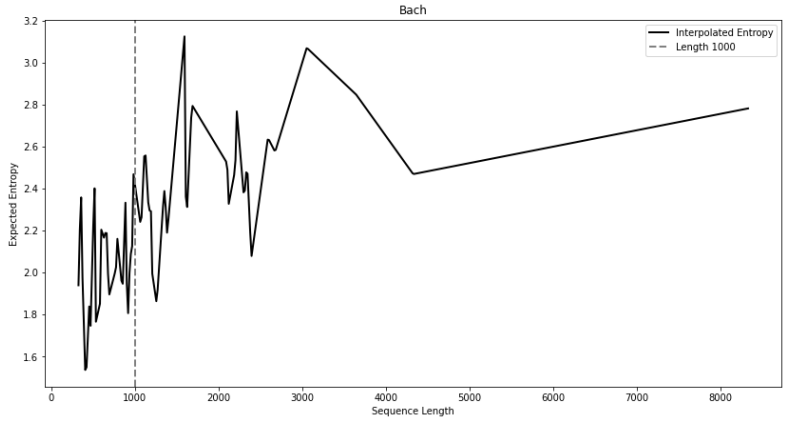

Fig. 2: Plot of linearly interpolated entropy behavior for all considered J. S. Bach compositions (a total of 75 contiguous segments of notes), primary instrument - violin (chosen due to its prevalence over other instruments and it being a favorite of the composer). The **expected entropy** for a synthetic composition of length **1000** evaluates to **2.421** (vertical dashed line).

of different motif thresholds and sampling temperatures appears consistent with the hypothesis: the introduction of motif augmentation provides a corrective mechanism for entropy alignment when the temperature-based method alone struggles to do so.

4 Conclusions and future research directions

The results suggest that motif augmentation, when combined with the temperature sampling mechanism, improves the fidelity of the LSTM-generated sequences to the expected entropy. This improvement is likely due to the additional information provided by the motifs, which act as structured guidance in cases where sampling probabilities deviate from the optimal threshold. These findings highlight the utility of motif-driven enhancements in generative models, particularly in domains where alignment of entropy with ground truth is critical, such as in natural language processing, bioinformatics, and sequence modeling tasks. The statistical significance of the results supports the claim that motif augmentation is a robust technique for reducing the entropy deviation, making it a valuable addition to sequence generation methodologies.

In terms of future research, we will focus on evaluating the interplay between sampling temperature and motif threshold in finer detail to understand their synergistic effects. Furthermore, we wish to perform some type of task-specific validation by testing the framework in real-world sequence prediction tasks to quantify improvements in task-specific metrics. A straightforward quantitative measure to evaluate the quality of a musical pitch sequence could be the dis-

ESANN 2025 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 23-25 April 2025, i6doc.com publ., ISBN 9782875870933. Available from http://www.i6doc.com/en/.

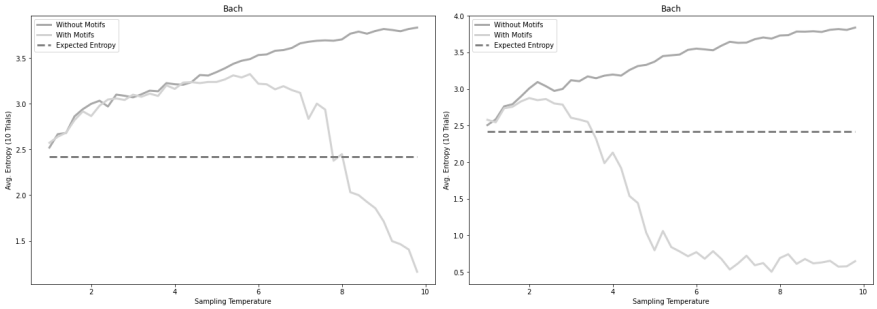

Fig. 3: Effect of using two different motif thresholds (0.10 and 0.25) on the entropy of synthetic sequences without (dark gray) and with (light gray) motif augmentation, as sampling temperature increases from 1 to 10. The horizontal dashed line is the expected entropy of **2.421**.

sonance ratio, defined as the fraction of intervals in the sequence that are considered dissonant (e.g., minor second, major seventh) versus consonant (e.g., perfect fifth, major third).

References

[1] Alexandru-Ion Marinescu. Bach 2.0 - generating classical music using recurrent neural networks. *Procedia Computer Science*, 159:117–124, 2019. Knowledge-Based and Intelligent Information & Engineering Systems: Proceedings of the 23rd International Conference KES2019.

[2] Yuxuan Wu, Roger B. Dannenberg, and Gus Xia. Motif-centric representation learning for symbolic music, 2023.

[3] Shuqi Dai, Zeyu Jin, Celso Gomes, and Roger B. Dannenberg. Controllable deep melody generation via hierarchical music structure representation, 2021.

[4] Keshav Bhandari and Simon Colton. Motifs, phrases, and beyond: The modelling of structure in symbolic music generation, 2024.

[5] Sangjun Han, Hyeongrae Ihm, Moontae Lee, and Woohyung Lim. Symbolic music loop generation with neural discrete representations, 2022.

[6] Claude E Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27:379–423, 623–656, 1948.

[7] Thomas M Cover and Joy A Thomas. *Elements of Information Theory*. Wiley-Interscience, 2006.

[8] Frederick Jelinek. *Probabilistic Information Theory*. McGraw-Hill, 1976.

[9] Edwin T Jaynes. Information theory and statistical mechanics. *Physical Review*, 106:620–630, 1957.

[10] Steven M Pincus. Approximate entropy as a measure of system complexity. *Proceedings of the National Academy of Sciences of the USA*, 88:2297–2301, 1991.

[11] Andrej Karpathy. The unreasonable effectiveness of recurrent neural networks, 2015.

[12] Ari Holtzman, Jan Buys, Li Du, Maxwell Forbes, and Yejin Choi. The curious case of neural text degeneration. In *Proceedings of the 8th International Conference on Learning Representations (ICLR)*. OpenReview.net, 2020.

[13] D. Smith and C. Wood. The 'USI', or Universal Synthesizer Interface. *Journal of The Audio Engineering Society*, 1981.