Generating music from poems using neural networks

Horia-Ioan Iancu
University Politehnica of Bucharest
313 Splaiul Independenței,
Bucharest, Romania
iancu99horia@yahoo.com

Ștefan Trăușan-Matu
University Politehnica of Bucharest
313 Splaiul Independenței,
Bucharest, Romania
and
Research Institute for Artificial Intelligence
and
Academy of Romanian Scientists
stefan.trausan@cs.pub.ro

ABSTRACT

Sonification is the process of using sound to convey or add a new dimension to information. In particular, textual human-computer interaction is enhanced if sound is added. In order to sonify a poem, a quantified representation of emotions in the form of sentiment dictionaries would be necessary, as well as knowledge of music theory. Diverging from previous research, a method that uses deep neural networks to combat both of these problems is presented. In order to establish an emotional connection between text and music, the general sentiment of a poem is first extracted. An analysis of the poem's musicality, expressed in the form of rhythm, rhyme and punctuation, is then used as the basis for creating a simple rule-based melody. Using this melody as conditioning, a piano accompaniment is then produced using a neural network. Results show that the accompaniments generated adequately reflect the mood of a poem, while also being pleasant to listen to.

Author Keywords

Sonification; poetry; Natural Language Processing; Neural Networks.

ACM Classification Keywords

H.5.5. Information interfaces and presentation (e.g., HCI): Sound and Music Computing.

DOI: 10.37789/rochi.2021.1.1.20

INTRODUCTION

Poetry, the central focus of this paper, is one of the earliest forms of art and closely related to music. Recited or sung, poetry was a medium for orally passing information from one generation to the next. As such, music was used to accompany and enhance the effect of poetry, especially in the case of religious hymns, historical accounts, love songs or chants. Taking advantage of this relationship between music and poems, sonification would enrich the already evocative imagery of poetry, providing a better way of understanding and appreciating this form of art. However, the need of specialized lexicons for sentiment analysis and of extensive knowledge about music theory are major drawbacks that could affect the quality of the outcome. However, more recent advances in the domain of neural networks can address both problems. One more issue is how to create a connection between rhythm, phrasing, and motifs in music and a poem's musicality, expressed through prosody.

Thus, the central aim of this paper is to generate a structured musical performance that can adequately mirror the emotion present in a poem and to enhance the aesthetic experience of reading a poem. This can bring a new dimension, for example, to an interface for reading poems by combining the visual stimulation of text with the auditory stimulation of music. Alternatively, it could be used alongside the text-to-speech function used by the visually impaired in order to mitigate the monotonous voice that may fail to convey a poem's emotional content. Another objective is to perform the sentiment analysis of a poem without relying on a pre-existing sentiment dictionary, using only a small training set of annotated poem lines.

The proposed approach firstly classifies a poem according to the general sentiment present in the text in two broad categories: positive or negative. Then, a musicality analysis is performed on the text by detecting rhyme, rhythm and punctuation, important elements of a poem's expressivity. The data extracted during the previous step is then used to create a melody based on a set of rules. An accompaniment is then generated automatically for the melody using a neural network.

Results show that the mood of a poem, positive or negative, is successfully identified. Various intriguing accompaniments have been produced that present structure, with repeating motifs and phrases. However, as the neural network randomly generates the pieces, some of them are more adequate than others. In order to obtain an accompaniment that is considered to echo a poem's emotions more faithfully, running the application repeatedly on the same poem is recommended.

RELATED WORK

TransProse

Proposed by Davis and Mohammad [1], TransProse is a system for the automatic generation of music from novels. TransProse first generates an emotion profile based on the input text using the NRC Emotion Lexicon [5, 6]. The novel is partitioned into sections and sub-sections and counts for eight emotions are generated for each of these. Then, the ratio of emotion words to the total number of words is calculated – the overall emotion density [1, p.5]. Densities of particular emotions are also calculated: joy density, for example. The key for the piece is chosen according to the ratio of positive words to negative words. The generated musical piece is composed of three melodies. The octaves, number of notes and note pitches for these melodies are chosen according to the emotional densities. The tempo is then generated according to an activity score that depends on the density of active emotions (anger, joy) and passive emotions (sadness).

Tx2Ms (Text-to-Music)

Proposed by Huang et al. [4], Tx2Ms is a system for the automatic generation of music from classical Chinese poetry based on Markov Chains. Tx2Ms relies not only on the text of a poem, but also on a recording of a reading of the poem. Various musical parameters, such as sonority, dynamics and tempo are mapped to characteristics of the poem, such as Chinese tone class and the poetic mood. For example, the rhythm of the composition is determined by the succession of Ping and Ze tones [4, p. 496]. For each line, a Ping/Ze structure is determined. Then, these structures are aggregated, forming a rhythm sequence from which a transition matrix is calculated. The resulting matrix describes a Markov Chain that is used to generate durations for the notes. The composition's scale is chosen from three pentatonic modes: Chinese Pentatonic, Japanese Hirajoshi and Balinese Pelog. The selection is made using data from the recording of the spoken poem. Pitch, duration and dynamics are then chosen for the notes based on the Ping/Ze output of the Markov module.

METHODS

Sentiment analysis

A deep neural network BERT classifier [2] is employed to perform the sentiment analysis of a poem. A BERT model was chosen because the bidirectional representations generated for the input consider both left and right context. This is an important advantage, because figures of speech present in a poem can change the order of the words without changing the meaning of a line. BERT was trained using the dataset from Sheng and Uthus [8], which consists of annotated poem lines. They are split into four categories: "negative", "positive", "no impact", and "mixed", according to their mood. The model trained this way has a resulting accuracy of 85.7%. In order to categorize a poem, it is first separated into its component lines. These are then fed into the BERT model, which outputs a label for each of them. Then, the "negative" and "positive" labels are counted, and the poem is assigned to the majority class. The tempo and key for the rule-based melody are chosen according to the class:

- Adagio (70 BPM) and A minor for negative poems.
- Allegro (120 BPM) and C major for positive poems.

Musicality analysis

The analysis of a poem's musicality implies detecting some of its features that make it sound pleasant to a reader and that can help create a melodic line that is also appropriate and pleasing to a listener's ear. Three elements of a poem are considered here: rhythm, rhyme, and punctuation marks. A rhythm analysis consists of breaking down each line into words and then dividing these into the component syllables in order to determine which are stressed and which are unstressed. This is done using the CMU Pronouncing Dictionary [10], which has entries for more than 134,000 English words and their pronunciation. In the CMU representation, vowels contain a stress marker: '1' for stressed syllables and '0' for unstressed syllables. Thus, a list of stress sequences composed of '1' and '0' that correspond to each line of the poem is generated.

To obtain the rhyme scheme, a rhyme dictionary is built for each word at the end of a line. Then, traversing the end word list, each of them is compared to the ones before. To keep track of the lines that were processed, each line will have an entry in a scheme dictionary. If a rhyme could not be identified for the current end word, a new number is assigned to that line, starting with 0. If a rhyme has been identified, the number of the first line it rhymes with is assigned to the current line. The output of this analysis is a sequence of numbers that describe the rhyme structure of the poem. For example, in the case of a Petrarchan sonnet, a 14-line poem, the function might return the sequence [0, 1, 1, 0, 0, 1, 1, 0, 2, 3, 4, 2, 3, 4]. This means that the first line rhymes with the fourth, the second with the third, and so on.

Accompaniment generation

After the prosodic analysis, the melodic line can be created. The rhythm is first generated based on the list of stresses. Each line of the poem is first tokenized using Spacy [9], then the token list is traversed, using the following set of rules to establish the durations of the notes:

- If the token is a full stop, an ellipsis or a semicolon, the note will be a quarter rest.
- If the token is a comma, the note will be an eighth rest.
- If the token is an exclamation mark, the note will be an eighth note.
- If the token is a question mark, an eighth note will be added, followed by an eighth rest.

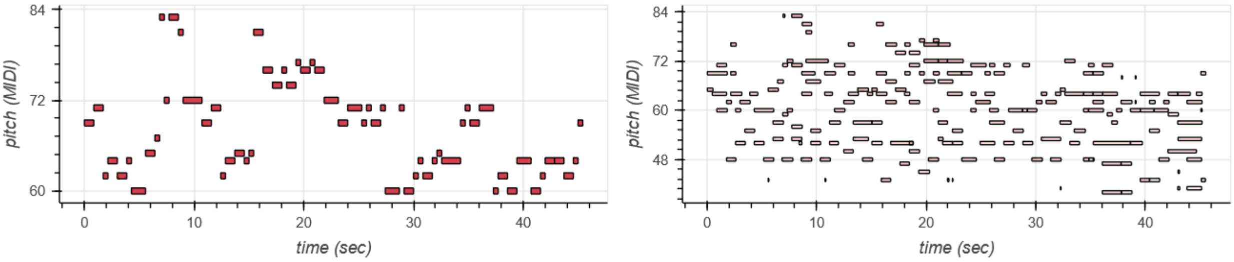

Figure 1: Melody and accompaniment generated for "Music, When Soft Voices Die".

- If the token is a one syllable stop word, the note will be an eighth note.
- For all other cases in which the token is a word, each syllable will receive a duration. If it is a stressed syllable, it will receive a quarter note. If it is an unstressed syllable, it will receive an eighth note.
- The last note of each line, if it does not end with a rest, will be a half note.

The note pitches are chosen from the two previously stated keys: C major if the poem is positive, and A minor if the poem is negative. For each key, there are two octaves: C major starts from the "middle C", or C4, up to B5; A minor starts from A4 and goes up to G6. Choosing the next note that will be played is done randomly, with a 0.7 chance of choosing a higher note. Only notes adjacent to the current one may be selected. Using the same tokenization process, the following rules are used to generate note pitches:

- If the token is an exclamation mark, the last note will be repeated to accentuate it. As stated above, it will also have a shorter duration of an eighth rest.
- If the token is a question mark, an ascending note will be added, in order to mimic the rising tone of a question.
- Then, the rest that follows it mimics waiting for an answer.
- For all other cases in which the token is a word, each syllable of the word receives a note.
- The last note of the line is the third or the fifth note of the key, chosen randomly.
- If a line rhymes with a previous one and the last note is not a rest, the last 3 notes of the current line will be the last 3 notes of the preceding rhyming line.
- Lines are indexed from 0. If the index of a line is even, the melody starts again with the first note of the key.
- If two lines are identical, their notes will be identical.

Combining the duration and note pitches, a MIDI file containing a piano performance is created.

Using the Music Transformer [4], a performance that consists of the melody and an overlayed accompaniment is generated using the output of the previous stage as conditioning. The result is the final piece that represents the sonification of the poem. Using a Transformer confers the piece a quasi-coherent structure and expressivity, such as articulation and dynamics. The final performance will (usually) be equal in length to the rule-based melody.

RESULTS

The polarity of the poems is correctly identified, and interesting piano accompaniments have been produced. However, being randomly generated, some of the pieces are a better fit than others. Running the application repeatedly on the same poem until the desired result is produced is recommended. Sometimes, the performances are too repetitive or too dissonant. Also, in contrast to music written by a human, the piece produced by the Music Transformer does not have a clearly distinguishable time signature and key, so it would be difficult to transcribe it to sheet music. Two examples of the application's output are presented below, one negative poem and one positive poem.

The first example is the poem "Music, When Soft Voices Die", by Percy Bysshe Shelley. The poem speaks about the inevitability of the end, the permanence of the memories of transient things, but also of coming to terms with loss and acceptance. It is one of the most influential of Shelley's works and it has been set to music numerous times. BERT classified it as negative, so the tempo and the key were set to 70 BPM and A minor.

The accompaniment generated for this poem, is the best one so far. It perfectly exhibits the melancholy found in the text, while also having passages that are more hopeful. The dynamics also fit the general feel of the poem, with a softness and delicacy that parallel the first two lines of the poem. The end becomes more and more quiet, its "voice" dying, remaining only a vibration in the memory of the listener. It is a very good example of sonification. Both the melody and the accompaniment are shown in Figure 1 using the "piano roll" representation. The piano roll is a graph which represents notes as colored bars. The horizontal axis displays time position, while the vertical axis displays note pitch. Depending on the duration of a note, a bar can be longer or shorter. Depending on the pitch of a note, a bar can be higher or lower on the graph.

The second example is the poem "Sonnet LXXII. Oft, when my spirit doth spread her bolder wings" from Edmund Spenser's collection of sonnets "Amoretti". "Amoretti" is a

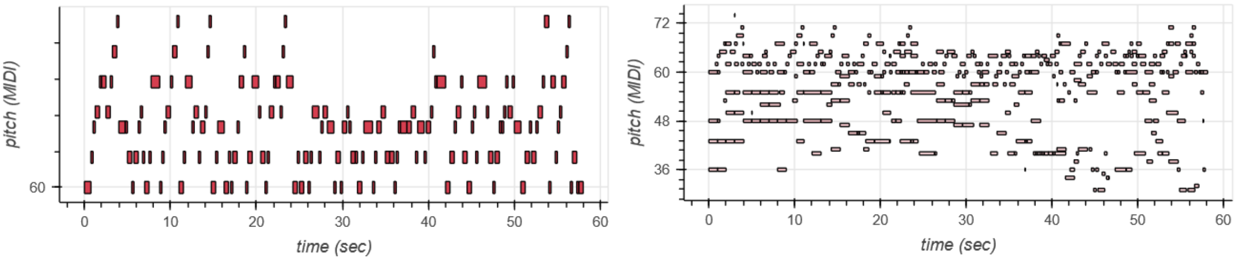

Figure 2: Melody and accompaniment generated for "Sonnet LXXII. Oft, when my spirit doth spread her bolder wings".

cycle of sonnets written by Spenser during the courtship of his second wife. As such, Sonnet LXXII is a poem that exalts his love for her and his desire for her happiness and fulfillment. She becomes his heaven, replacing the celestial one which is unattainable, easing the burden of life that bears heavy weight on his shoulders. BERT classifies this poem as a positive one, setting the tempo and key to 120 BPM and C major. The melody and the accompaniment are shown in Figure 2. While it does transmit positivity, perhaps even happiness, it is relatively unremarkable. It is a bit flat and repetitive, but it is the best one of a series of pieces generated for this poem. For some reason, the Music Transformer had a very difficult time creating an accompaniment that was not strange, with seemingly random notes interspersed among the rest or disjointed sections.

CONCLUSIONS

The system presented here addresses the problem of sonification of poetry: how to convey the information present in a poem using sound. To solve this problem, we have devised a method based on deep neural networks.

A BERT classifier was used to determine the dominant category of emotions present in the text, positive or negative. A rule-based melody is then created from the poem's musical features, such as rhyme, rhythm and punctuation. The melody is generated in three stages: rhythm, the duration of notes, is generated first, then pitch values are assigned to the notes, then a MIDI file containing the melody is created. This melody is then used to condition a Music Transformer model, producing the accompaniment in the form of a MIDI file that can be played in the application. Results have shown that the accompaniments generated adequately inform the reader of a poem's overall sentiment.

For future work, a more detailed sentiment analysis, classifying poems depending on the dominant emotion (joy, fear, anger, etc.), could be implemented in order to make the accompaniment more expressive.

Furthermore, the musicality analysis could be expanded to include more elements of prosody, such as alliteration and assonance, or to detect a poem's foot, such as iamb or trochee.

REFERENCES

1. Davis, H. & Mohammad, S. M. (2014). Generating music from literature. *arXiv preprint arXiv:1403.2124.*
2. Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805.*
3. Huang, C. F., Lu, H. P., & Ren, J. (2012). Algorithmic approach to sonification of classical Chinese poetry. Multimedia Tools and Applications, 61(2), 489-518.
4. Huang, C. Z. A., Vaswani, A., Uszkoreit, J., Shazeer, N., Simon, I., Hawthorne, C., Dai, A. M., Hoffman M. D., Dinculescu M., & Eck, D. (2018). Music transformer. *arXiv preprint arXiv:1809.04281.*
5. Mohammad, S. M., & Turney, P. D. (2013). Crowdsourcing a word–emotion association lexicon. *Computational intelligence*, 29(3), 436-465.
6. Mohammad, S., & Turney, P. (2010, June). Emotions evoked by common words and phrases: Using mechanical turk to create an emotion lexicon. In *Proceedings of the NAACL HLT 2010 workshop on computational approaches to analysis and generation of emotion in text* (pp. 26-34).
7. Schubart, C. F. D. (1839). *Gesammelte Schriften und Schicksale: Ideen zu einer Aesthetik der Tonkunst (Vol. 5).* J. Scheible.
8. Sheng, E., & Uthus, D. (2020). Investigating societal biases in a poetry composition system. *arXiv preprint arXiv:2011.02686.*
9. spaCy. https://github.com/explosion/spaCy
10. The CMU Pronouncing Dictionary. http://www.speech.cs.cmu.edu/cgi-bin/cmudict