Data Mining / Live Scoring – A Live Performance of a Computer-Aided Composition Based on Twitter

Alexandros Drymonitis[1] and Nicoleta Chatzopoulou[2]

[1] PhD Candidate Royal Birmingham Conservatoire
Birmingham City University
[2] PhD Candidate Technical University of Crete (TUC)
School of Architecture
alexdrymonitis@gmail.com

Abstract. *Data Mining / Live Scoring* is a project for an algorithmic composition for a six-member acoustic ensemble aided by a computer. This performance is based on Twitter, where tweets are being downloaded and analyzed during the performance. The analysis of the tweets determines various parameters of the algorithmic composition which is being carried out in parts during the show. The resulting scores are being projected on screens for the ensemble members to sight read. The score projection includes a blinking cursor controlled by the main computer which facilitates coordination between the performers and highlights the bar currently being played. The performance is accompanied by a light installation part of which is used to communicate the sentiment of the score to the musicians.

Keywords: Data Mining, Live Scoring, Sentiment Analysis, Twitter, Algorithmic Composition, Computer Aided Composition

1 Introduction

Data Mining / Live Scoring is an algorithmic composition project based on input from Twitter. It is a collaboration between the authors of this paper and the ARTéfacts Ensemble, a six-member acoustic ensemble[3]. It is a project for sight-reading scores which are being created on-the-fly based on the sentiment analysis of the tweets downloaded during the performance.

The main idea behind *Data Mining / Live Scoring* was to enable the participation of the audience in the creation of the score, which was to be sight-read by the ensemble performers. At the same time, we wanted to include social media and a random input of some sort. Twitter seemed like the perfect source of input because it makes it possible to combine the three attributes mentioned above. Twitter is a micro-blogging social network with a very big and active database. By prompting users to tweet during the performance we were able to incorporate

[3] http://artefactsensemble.gr/

audience input. At the beginning of every performance we chose three trending hashtags along with the *#dataminingliveścoring* hashtag we created. This way we used tweets from all Twitter users tweeting under the selected hashtags. By applying sentiment analysis to the downloaded tweets, which determined various aspects of the resulting score, we introduced randomness to the performance since we could not tell beforehand what the overall sentiment of the tweets would be.

By loosely relating the resulting music to soundtracks for film which intensifies the experience of the spectator (Thompson, Russo, & Sinclair, 1994), the aim was to sentimentally reinforce the tweets which were used to construct the scores. These tweets were being projected on the walls of the performance space one by one, where the projection of each tweet lasted for the duration of the music created by it. Thus, forming a connection between the source and the result.

The authors of this paper and the ARTéfacts Ensemble were commissioned by the Onassis Cultural Centre to create this work. It was performed twice in April 2019 in Athens, Greece. Excerpts are available online. [4]

2 Related Work

2.1 Computer Aided Composition

Long before the use of computers by composers to determine various parameters of the compositional process, there have been examples of works involving systems which employ chance (Roads, 1996, p. 821). An important, pioneering work in the field of computer-aided music composition is the Illiac suite for string quartet, composed in 1956 (Funk, 2018). It was a collaboration between L.M. Hiller and L.M. Isaacson, who programmed the Illiac computer to generate music based on the mapping of random integers to musical pitches through certain mathematical operations, or compositional rules (Hiller & Isaacson, 1958). In particular, the fourth movement of the piece employs Markov chains and stochastic processes (Sandred, Laurson, & Kuuskankare, 2009). There were numerous notable composers around the time of Hiller or later, who used a computer to generate music, such as Herbert Bruen and John Myhill, James Tenney, Pierre Barbaud, Michel Phillipot, Iannis Xenakis and G. M. Koenig (Roads, 1996, p. 831).

2.2 Generative Soundtracks and Soundscapes

There are quite a few examples of generative soundtrack algorithms. Knees et al. produce a synaesthetic piano composition based on the view of the window of a - moving - train (Knees, Pohle, & Widmer, 2008). They achieve this by analyzing the image captured from a camera and split it in four horizontal regions, representing four octaves on the piano keys. By analyzing the color in the pixels at a vertical line in the center of the image, they create melodies and harmonies

[4] `https://vimeo.com/369534737`

by mapping this analysis to notes based on the synaesthetic scale by Alexander Scriabin. The audio is produced via a MIDI piano sound.

Hazzard et al. created a sound walk for the Yorkshire Sculpture Park in the UK (Hazzard, Benford, & Burnett, 2014). Based on GPS data, according to the location of the listener, the mobile application created for this project would either change between various short phrases which were looping until a new one was triggered, or change the orchestration.

Music Paste is a system which chooses certain music clips from an audio collection which the system considers appropriate, and concatenates them seamlessly by creating transition segments to compensate for changes in tempo, loudness, and other parameters (Lin, Lin, Tien, & Wu, 2009). AutoFoley is a generator of Foley audio tracks based on the analysis of the image and the sound of movie clips (Ghose & Prevost, 2020). They achieve this by incorporating deep neural networks for predicting various sound classes, like footsteps, typing, cars, and others.

Manzelli et al. combine symbolic with raw audio in order to create music which both captures emotions, mood etc. but also does not sound like improvisation (Manzelli, Thakkar, Siahkamari, & Kulis, 2018). Their approach utilizes the WaveNet model by van den Oord et al. (van den Oord et al., 2016). They produce symbolic melodies as in MIDI which are then treated as a local conditioning of a WaveNet model. Koutsomichalis and Gambäck produce audio mashups and synthetic soundscapes by downloading and analyzing audio over the Internet (Koutsomichalis & Gambäck, 2018). They achieve this by performing onset analysis and spectral feature extraction of the downloaded audio, and temporal scheduling and spatializing. Their work was aimed at a sound installation.

2.3 Twitter Sonification

Twitter is a tool which has been used widely in audio works. The techniques used in these projects vary. Ash analyzed the sentiment of tweets on trending musicians on Twitter and applied the results to an additive synthesis program, mapping the sentiment to the frequencies and the amount of harmonics (Ash, 2012). Boren et al. use geolocation of tweets which determines pitch changes in a granular synthesis program, based on the location's distance from a focal point (Boren, Musick, Grossman, & Roginska, 2014). Hermann et al. also focus on geolocation calculating distance from a focal point. The result is mapped to reverberation. They also take into account the number of followers of people tweeting which determines timbre parameters, and the longitude values determine panning (Hermann, Nehls, Eitel, Barri, & Gammel, 2012).

Other examples include TweetDreams by Dahl et al., which creates nodes and sub-nodes which are provided with a randomly generated melody. These nodes are created based on similarity between the downloaded tweets (Dahl, Herrera, & Wilkerson, 2011). Tweet Harp by Endo et al. is an Arduino-based laser harp which recites tweets with Text-To-Speech (Endo, Moriyama, & Kuhara, 2012).

A rather different example is MMODM by Tome et al. This is an online drum machine where Twitter users can jam collectively by sending tweets to hashtags created by the users via the project's website (Tome, Haddad, Machover, & Paradiso, 2015). This project is different in that the users are aware of the fact that their tweets are being analyzed, plus it utilizes a specific syntax in order for the drum machine to comprehend what the user wants. An example is this:

riff on this [a-a-a-abc-cc] with me on #mmodm

where the square brackets contain the text to be analyzed. The letters in this text represent various pre-selected instruments and the hyphen characters represent rests.

3 Tweet Analysis and Parameter Mapping

This project was based on the sentiment analysis of the tweets which were downloaded during the performance in an effort to create a composition that would relate to the text which was projected for the audience to read. Sentiment analysis is widely used in various fields where consumer opinion is important (Liu & Zhang, 2012). There are mainly two approaches to it, the semantic approach and the learning-based approach. The semantic approach is based on the use of lexicons with words or expressions paralleled with a polarity (positive – neutral – negative). The learning-based approach utilizes machine learning with a variety of supervised learning algorithms like Support Vector Machines (SVM), Naive Bayes, and K-Nearest Neighbor (KNN) (Poonguzhali, Vinothini, Waldiya, & Livisha, 2018).

In this project we used a lexicon-based approach utilizing the Opposing Polarity Phrases lexicon (Kiritchenko & Mohammad, 2016) which can be found online. This lexicon consists of 1,178 terms with a valence value for every entry between -1 for negative and 1 for positive. By summing the valence of all entries in a tweet we could derive the overall sentiment of the downloaded tweets. The resulting composition was realized in blocks which were constructed by analyzing up two thirty tweets for each block. These blocks were a set of elements derived from a musical library composed for this project by the second author of this paper. This library contained nine distinct forms structures with various elements, analyzed in the next section. These forms were composed intuitively with tweet sentiment valence and density –which was determined by the number of tweets downloaded against the time it took to download them– as criteria. An SVM classifier was trained with summed sentiment values and density of tweets mapped to these nine forms. Listing 1 shows an example of the classifier training data. Once the values for all the tweets for a block were derived they were fed to the classifier which predicted the form for this block. Each block lasted approximately three minutes.

The next steps were to determine the rhythmic phrases for the block. This was done either by analyzing the prosody of the tweets, using the Prosodic module[5], or by choosing short rhythmic phrases from a tree structure by doing a

[5] https://pypi.org/project/prosodic/

random walk from its root through its branches. Even though prosody is determined by sonic elements of spoken language (Carlson, 2009), the Prosodic module developers state that the module performs with a high percentage of accuracy when compared to prosodic analysis of text entries from litery humans. Whether the prosody or the trees were used was hard-coded in the structure of the form which was predicted by the SVM classifier. When all the rhythmic phrases were collected the algorithm would create melodies for the six instruments of the ensemble. The melodies were also created from tree structures. Lastly, various techniques were inserted in random points of the notes. These techniques included slap tongue, staccato, glissando, tremolo, Bartok pizzicato, and others. Dynamics were also inserted at this last stage. The various techniques and the dynamic ranges were also hard-coded into the form structures and were chosen randomly by the algorithm. Figure 1 shows a flowchart of the whole process.

Listing 1

```
sents_and_dens = [[2.5, 0.2], [2.0, 0.18], [4.6, 0.13]]
form_indexes = [0, 1, 2]
clf.fit(sents_and_dens, form_indexes)
```

(Example of a training set for the sklearn SVM classifier from the SciPy Python module. The sentiment valence is mapped to a range between 1 and 5 to fit the five different music scales used in the music library. The density values are normalized to a scale between 0 and 1 where 1 is thirty tweets downloaded before the previous form block had ended.)

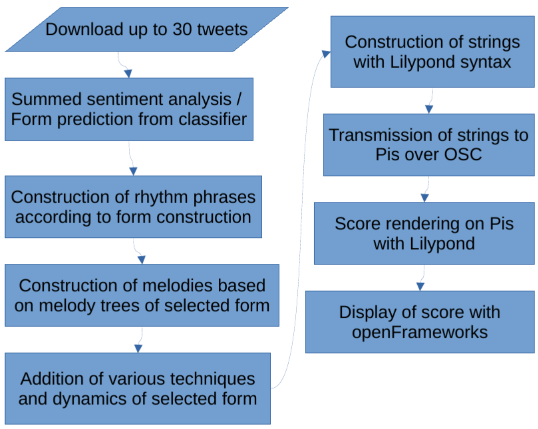

Fig. 1. Flowchart of the process of Data Mining / Live Scoring from the first stage of downloading the tweets up to displaying of the resulting score.

3.1 Music library and compositional decisions (rules)

Having as a main goal to create a music composition based on tweets, a set of initial compositional decisions were made. The first decision was that the algorithm would draw pitches from music scales. The chosen scales were a whole-tone, an Iwato, a Pelog and 2 mixed-mode scales. Various possibilities for pitch successions sprang out from tree structures within each scale. Within a scale, every pitch was used as the root of a tree. From each root, two branches were created corresponding to one pitch each, and consequently two new branches for each branch, leading to a total of four melodic sequences. Additionally, for each scale, three possible roles were assigned, entitled as solo, mixed and accompaniment. For each role, trees were created in the way already described. A maximum pitch variety for a solo melody was used and a minimum pitch variety for a melody which would be used as accompaniment by the algorithm. These structures provided not only a plethora of melodic possibilities, but also a basis for contrapuntal formations. The second decision regarded rhythm, which was determined in two ways; tree structures and the prosody of the tweets. The trees for the rhythm were constructed in an analogous way to those for the pitch. The root and the first branch of each rhythm tree was in 4/4 metre, whereas the consequent branches were in 2/2 metre. For each tree, there were four different options for the root and four for the first branching in order to have more variety in the melodic structures forming. Similarly, there were tree structures for all three roles as previously described. In the case of the prosody, the algorithm used a collection of predetermined material to form the prosody based on the strong syllables of a tweet, where one syllable was mapped to one rhythmic symbol. A table was then constructed with rhythmic patterns consisting of 2 - 9 rhythmic symbols, for six different meters (2/4, 3/4, 4/4, 5/8, 6/8, 7/8) and for the three different roles.

The instrumentation was clarinet in Bflat, soprano or baritone saxophone, violin, viola and percussion (six different percussion groups for two performers). A great deal of aesthetic and practical choices went into the formation of the percussion groups. Each of the four main groups had a combination of skin, wood and metal instruments, however each with a distinct character with respect to pitch possibilities and volume of sound. There were also two complementary groups, labeled as junk percussion, which were chosen for their timbral and gestural contribution to the instrumentation and performance. For all instruments, there was a detailed study of the techniques and articulations that they could perform, taking into account the fact that the performers would have to sight read the score. Additionally, this study was done having in mind the relationship of articulations and techniques to the resulting dynamics. A table with five columns was formed, which corresponded to articulations with dynamic gradations ranging from *fff-ff* to *pp-ppp*. The information contained in this table was hard-coded in the sub-blocks of the form. In total, there were nine different blocks (options) for the form. The criteria for choosing a form depended on the sentiment and density of the tweets. Each form followed a distinct order of orchestrational combinations (as a percentage of the whole), choices of tempo, dynamics and articulations (also as percentages within each sub-block). So, in fact there was a random aspect of the composition, based on the sentiment and prosody of the tweets, but also a deterministic aspect regarding the hard-coded rules built into each form block.

4 The Score Rendering System

The system for the realization of the scores utilized Lilypond and openFrameworks. Once all the elements for the full score of a block were collected by the main algorithm, strings for each Lilypond score were created which were sent over OSC to the six Raspberry Pi computers which were displaying the scores. When a Raspberry Pi had received all the Lilypond strings it created two .png files for each score.

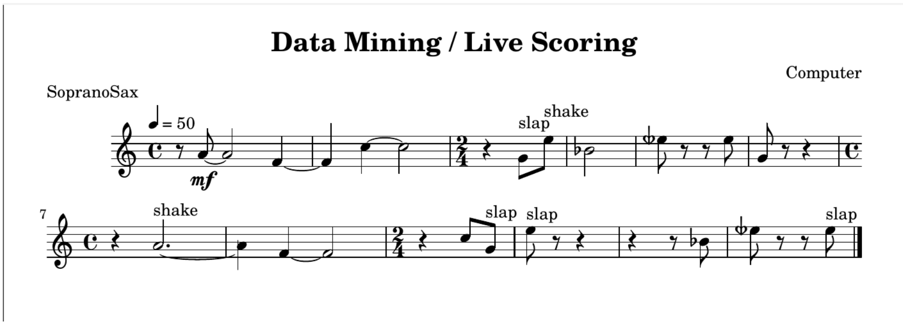

Fig. 2. A score sample with visible rests and transparent additional notes. In the beginning of the second bar the note stem is longer than usual because there is a C note above the displayed F which is rendered transparent.

These two files had a small difference at the beginning of each bar. In case a bar started with a rest, one .png file would render the score properly, while the other file would render the rest transparent. In case a bar started with a note, the score included an additional C one octave above middle C (in case of a G clef). This additional note was rendered transparent in the first .png file while the file would render it properly. This way the beginning of each bar had this slight difference between the two files. Figure 2 is a sample of the first file, which was the file displayed in the performer's monitor, and figure 3 is a sample of the same score containing the transparent rests and the visible extra notes.

Once the scores were rendered in .png an openFrameworks program would open these two files and with the aid of the ofxOpenCV addon it would detect these differences in the beginning of each bar. This way the program knew where each bar started. The openFrameworks program was receiving data from the main algorithm via OSC at every rhythm beat. This data would trigger the flicker of a red cursor on top of the beginning of the bar currently being played. This

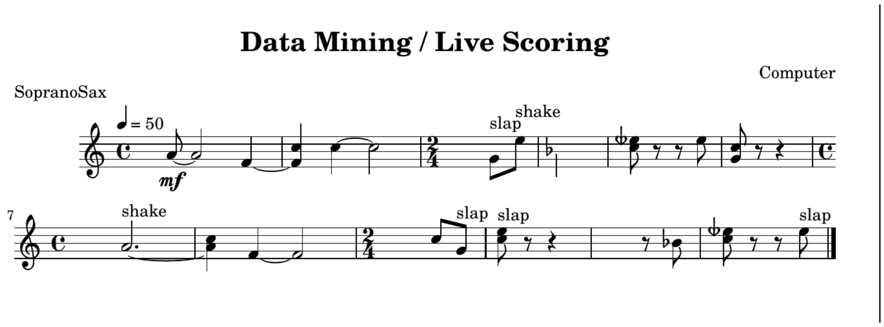

Fig. 3. A score sample with transparent rests and visible additional notes. In the beginning of the fourth bar the note which is a B, is treated as a rest and rendered transparent in order to not cluster two notes.

technique enabled both the synchronization of the ensemble but also ensured the performers would not get lost.

One more information passed to the performers was the sentiment of the tweets by controlling the color of a light bulb which was part of the light installation that accompanied the performance. It was placed in the center of the ensemble (the ensemble was forming a circle) and was connected to the local network over WiFi. Red represented negative sentiment, purple represented neutral, and blue represented positive sentiment.

5 Conclusions

We have presented *Data Mining / Live Scoring*, a performance for acoustic ensemble where music scores created on-the-fly were being sight-read by the performers. Twitter was chosen in order to enable us to utilize input from the audience, incorporate social media and introduce randomness to the entire process. The aim was to create an acoustic composition which would reflect the sentiment of the tweets utilized, and function as a soundtrack. By feeding the sentiment of the tweets to an SVM classifier, the algorithm produced scores for the entire performance in blocks, approximately thirty tweets for three minutes of music. The scores were created by drawing data from a music library which contained nine distinct form structures consisting of various combinations of instrumentation, articulations and dynamics, functioning as intuitive interpretations of the different sentiments. The scores were displayed on monitors and included a blinking cursor above the bar currently being played for synchronization and in order for the performers to keep track of the score. At the time of presentation of the project we did not have any plans on carrying out a research, therefore no survey with audience or ensemble members took place. Such a survey could provide information on how the audience evaluated the performance and how the ensemble members evaluated the efficiency of the system. Nevertheless, this paper provides as much information on the project as possible.

References

Ash, K. (2012). Affective states: Analysis and sonification of twitter music trends. In *Proceedings of the International Conference on Auditory Display (ICAD).* Atlanta, GA, USA.

Boren, B., Musick, M., Grossman, J., & Roginska, A. (2014). I hear NY4D: Hybrid Accoustic and Augmented Auditory Display for Urban Soundscapes. In *Proceedings of the International Conference on Auditory Display (ICAD).* Atlanta, GA, USA.

Carlson, K. (2009, 09). How Prosody Influences Sentence Comprehension. *Language and Linguistics Compass*, *3*, 1188-1200. doi: 10.1111/j.1749-818X.2009.00150.x

Dahl, L., Herrera, J., & Wilkerson, C. (2011). Tweetdreams: Making Music with the Audience and the World Using Real-time Twitter Data. In *Proceedings of the International Conference on New Interfaces for Musical Expression (NIME).* Oslo, Norway.

Endo, A., Moriyama, T., & Kuhara, Y. (2012). Tweet Harp: Laser Harp Generating Voice and Text of Real-time Tweets in Twitter. In *Proceedings of the International Conference on New Interfaces for Musical Expression (NIME).* Ann Arbor, Michigan, USA.

Funk, T. (2018, 12). A Musical Suite Composed by an Electronic Brain: Reexamining the Illiac Suite and the Legacy of Lejaren A. Hiller Jr. *Leonardo Music Journal*, *28*, 19-24. Retrieved from `https://doi.org/10.1162/lmj\_a\_01037` doi: 10.1162/lmj_a_01037

Ghose, S., & Prevost, J. J. (2020). Autofoley: Artificial Synthesis of Synchronized Sound Tracks for Silent Videos with Deep Learning. *IEEE Transactions on Multimedia*, 1-1. doi: 10.1109/TMM.2020.3005033

Hazzard, A., Benford, S., & Burnett, G. (2014). You'll Never Walk Alone: Designing a Location-Based Soundtrack. In *Proceedings of the International Conference on New Interfaces for Musical Expression (NIME)* (p. 411-414). London, UK.

Hermann, T., Nehls, A. V., Eitel, F., Barri, T., & Gammel, M. (2012). Tweetscapes - Real-time Sonification of Twitter Data Streams for Radio Broadcasting. In *Proceedings of the International Conference on Auditory Display (ICAD).* Atlanta, GA, USA.

Hiller, L. A., & Isaacson, L. M. (1958). Musical Composition with a High-Speed Digital Computer. *J. Audio Eng. Soc*, *6*(3), 154–160. Retrieved from `http://www.aes.org/e-lib/browse.cfm?elib=231`

Kiritchenko, S., & Mohammad, S. M. (2016). Happy Accident: A Sentiment Composition Lexicon for Opposing Polarities Phrases. In *Proceedings of the 10th edition of the Language Resources and Evaluation Conference (LREC)* (pp. 1157–1164). Portorož, Slovenia.

Knees, P., Pohle, T., & Widmer, G. (2008). sound/tracks: Real-Time Synaesthetic Sonification of Train Journeys. In *Proceedings of the 16th ACM international conference on Multimedia (MM '08)* (p. 1117-1118). New York, USA.

Koutsomichalis, M., & Gambäck, B. (2018). Algorithmic Audio Mashups and Synthetic Soundscapes Employing Evolvable Media Repositories. In *Proceedings of the International Conference on Computational Creativity.* Salamance, Spain.

Lin, H.-Y., Lin, Y.-T., Tien, M.-C., & Wu, J.-L. (2009, oct). Music Paste: Concatenating Music Clips Based on Chroma and Rhythm Features. In *Proceedings of the 10th International Society for Music Information Retrieval Conference* (p. 213-218). Kobe, Japan: ISMIR. Retrieved from `https://doi.org/10.5281/zenodo.1415758` doi: 10.5281/zenodo.1415758

Liu, B., & Zhang, L. (2012). A Survey of Opinion Mining and Sentiment Analysis. In C. Aggarwal & C. Zhai (Eds.), *Mining Text Data.* Springer. doi: https://doi.org/10.1007/978-1-4614-3223-4_13

Manzelli, R., Thakkar, V., Siahkamari, A., & Kulis, B. (2018). An End to End Model for Automatic Music Generation: Combining Deep Raw and Symbolic Audio Networks. In *Proceedings of the International Conference on Computational Creativity.* Salamance, Spain.

Poonguzhali, R., Vinothini, S., Waldiya, V., & Livisha, K. (2018, April). Sentiment Analysis on Linkedin Comments. *International Journal of Engineering Research & Technology IJERT (ICONNECT)*, *6*(7), 415.

Roads, C. (1996). *The Computer Tutorial.* London, UK: MIT Press.

Sandred, O., Laurson, M., & Kuuskankare, M. (2009, 01). Revisiting the Illiac Suite - A rule-based approach to stochastic processes. *Sonic Ideas/Ideas Sonicas*, *2*, 42-46.

Thompson, W. F., Russo, F. A., & Sinclair, D. (1994). Effects of Underscoring on the Perception of Closure in Filmed Events. *Psychomusicology Music Mind and Brain*, *13*(1-2), 9.

Tome, B., Haddad, D. D., Machover, T., & Paradiso, J. A. (2015). MMODM: Massively Multiplayer Online Drum Machine. In *Proceedings of the International Conference on New Interfaces for Musical Expression (NIME)* (pp. 285–288). Vancouver, BC, Canada.

van den Oord, A., Dieleman, S., Zen, H., Simonyan, K., Vinyals, O., Graves, A., ... Kavukcuoglu, K. (2016, 09). Wavenet: A Generative Model for Raw Audio.