A Fuzzy Rule Model for High Level Musical Features on Automated Composition Systems

Iván Paz, Àngela Nebot, Francisco Mugica and Enrique Romero

Departament de Llenguatges i Sistemes Informtics, Soft Computing research group
Technical University of Catalonia
ivnpaz@gmail.com
http://www.semimuticas.org

Abstract. Algorithmic composition systems are now well-understood. However, its capacity for consistently produce specific sensations in the listener, referred as high level musical features (like coherence, emotion or personality), is still object of discussion. Furthermore, whether or not the machine learning processes, used for its modeling, are really capable to capture such features to a great degree remains as an open question. As a consequence of this lack, not all legal outputs of those systems are considered as "good enough" for the composers. Then, even though the number of valid outputs is huge, many times the selection is performed manually, either using expertise in the algorithmic model, by means of sampling techniques, or some times by chance. The present work discusses a possible approach, that combines expert's opinion and a fuzzy methodology for rule extraction, for the modeling of high level features. An early implementation able to explore the universe of outputs of a particular algorithm by means of the extracted rules is discussed. The rules search for objects similar to those having a desired and pre-identified feature. In this sense, the model can be seen as a finder of found objects.

Keywords: Algorithmic composition, musical representation, musical features.

1 Exposition

Algorithmic composition systems (ACs), that create music by means of formalizable methods, are now well-understood, and documented (Nierhaus 2009, Fernández 2013). Examples have followed the trend of the formalization of thought and technology development, starting around 1000 AC with the Guido D'Arezzo method for the generation of melodies from text material, based simply on the mapping of vowels over pitches. Later in the "musical dice game" of the 18th century the outputs were generated, based on combinatorial structure. Players rolled a dice to choose, for every temporal unit, a bar from a table until complete a short piece. Although playing was trivial, the system required expertise in music composition for consider the harmonic and voice-leading aspects

The final publication is available at Springer via https://doi.org/10.1007/978-3-319-47337-6_25

of the combinations. In 1956 the Illiac suite was the first completely computer-generated composition. The piece was composed by Lejaren Hiller and Leonard Isaacson using a Monte Carlo method which outputs were musical notes. Over the 20th century, compositional methods, like twelve tone technique (that assured all 12 notes of the chromatic scale sounded as often as one another, and so, preventing the emphasis of any one note), and serialism (that uses a series of values to manipulate musical elements) used structural parameters and logical conditions. Xenakis' theory of sieves, is an application of logical combinations for the generation of musical structure. With the development of techniques of machine learning composition systems are becoming more and more complex. For example, "Autocousmatic" (Collins 2012) is an algorithmic system for electroacoustic music composition that incorporates machine-listening processes within the design cycle. In this way, the system is able to assess the "worthiness" of intermediate files from which the final output will be created. The formal structure is drawn from models based on the analysis of an exemplar corpus of pieces. A similarity-measure is used to decide between candidates for the final mix.

Despite the extensive work developed on ACs, its capacity for accurately produce high level musical features (like coherence, emotion or personality), is still object of discussion. Until this moment, the representations obtained for high level musical features were side effects of the machine learning research. This has motivated the emergence of new paths for explicitly work on this subject. For example, "affective algorithmic composition", which seeks to develop systems for the selective adjustment of emotional responses using parameterizations of musical features (Duncan et al, 2013). In his work "Musical form and algorithmic composition" Collins (2009) argue that "algorithmic music often seems stuck in a static moment form, able to abruptly jump between composed sections but unable to demonstrate much real dramatic direction." Models for musical form that takes into account the psychological perspective of the listener to a great degree have not been well explored. Then, whilst it is possible to generate sections of a composition by means of the different ACs, to achieve greater expressiveness, it is still necessary to implement strategies to capture the relations between sections, hierarchical layers (possibly by relating low and high level features) and evolution of the tension.

On the search for designing systems capable to consistently produced outputs perceived by the listener with specific characteristics, two approaches have been addressed. On one hand, some systems have followed the explicit rule paradigm, using conditions (many times expressed as rules) given by experts. On the other hand, other system have used machine learning tools to learn patterns and produce compositional models with the learnt parameters. However, the obtained representations do not takes into account musical considerations and then, are more suitable for tasks like pattern recognition. In other words, the fact that machine learning processes have effectively captured high-level features to a great degree is still object of discussion. Moreover, designed systems have not extensively incorporated perception and semantic of the form of generated music, including the psychological sensation of the listener. Attempts to do this often use modules designed for the evaluation and adjustment of the outputs based on pre-established symbolic domain (Collins 2012), like the fitness functions of genetic algorithms to modify the outputs until the desired ones are obtained. As a consequence of this lack, in practical implementations, the composers have to proceed by analyzing the outputs one by one, in order to select the best outputs among all the possible ones. This happens either in automated or assisted composition, in statistical models, and even in algorithms for stochastic synthesis (see for example, Luke 2009). This means that the composer has to create all the possible outputs and then, listen as much as possible in order to select those with the desired expressiveness, style, etc. This process is performed by using expertise on the model or by sampling techniques. However, the universe of all possible valid outputs of such kind systems normally range over thousand or even million, and so, it is impossible for the composer to explore it manually. In the present work we introduce a methodological approach for modeling high level musical features in the outputs produced by ACs. The model takes advantage of fuzzy quantifiers for modeling high level features. The objective is to climb in the semantic level by relating high level features with its representation at beat or raw level data. A remarkable point is that the design of the system requires the whole example to be listened before the evaluation, and so different levels of perception in the time and form domain are taking into account. The rest of the work is structured as follows. Section 2 present and discussed the model. Section 3 presents a general discussion of the results and elucidate possible further work.

2 Development

Given the subjective nature of the high-level features, the use of fuzzy systems to work with them sounds feasible. Fuzzy systems are based of fuzzy logic theory, they allow us to work with objects that are approximate or subjective rather than exact. In the context of the modeling of musical features, fuzzy systems can help to work without been restricted to pre-established structures for the evaluation modules of the outputs. Then, human expertise and its associated psychological perspective can be included within the design (or the evaluation) cycle. This allows the system to extract the musical representation of the expert experience as it is perceived and then, to translate it in terms of combinations of low level variables. We have to point out that the perception of the features constitute a subjective evaluation of the listened outputs, and so it is subjected to context and cultural experiences. Fur this reason it could hardly vary form one listener to other. However, it is possible, for example, to consider different opinions that may be pondered according with the degree of experience. A general structure of the system design is showed in Figure 1.

The system is composed by several modules, each of which can be independently treated. The first module is the algorithmic composer or generative system. It, either belongs to one of the stablished paradigms [1] or a hybrid one.

[1] Different algorithm classes enable specific approaches to musical structure generation. The most popular are Statistical models including Markov models, generative

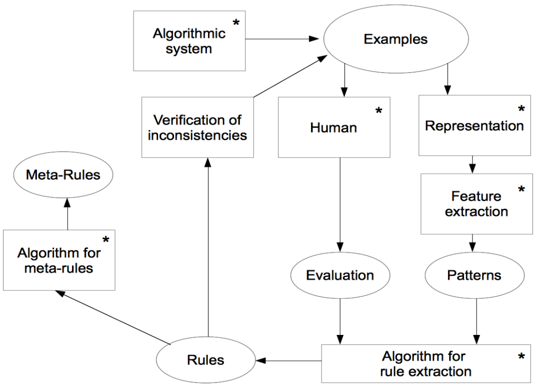

Fig. 1. The system architecture and process flow of the proposed approach. Squares denoted the different modules of the system and ovals denote data.

From the system, a sample of pieces (patterns) are extracted. This sampling process must have to take into account the characteristics of the generative system. For example, if the system belongs to the statistical models paradigm, outputs may be determined by a probability distribution, and so, as the training set may be biased, the extracted rules will describe only a part of the legal outputs. Then, the sample is evaluated by human experts by using linguistic variables in order to categorize if they have or not the desired feature. This process can be performed for short pieces, patterns, phrases or motifs, to explore the system's capacities over different scales of length and hierarchical levels. It has to be noticed that, as the whole output is listened before the evaluation, it takes into account the different levels of organization. For example, if dealing with rhythmic patterns it will consider the different arranges of contrast and repetition as well as its locations present in the whole output. Or, if dealing with melody sequences the evaluation will consider the transitions of adjacent notes, but also relations among notes at different points, as well as its evolution over time. In other words, this kind of evaluation allow us to include the different semantical levels and the subjective human perception of the outputs. The representation module structure the data in the most efficient manner for analysis and feature selection. The feature selection module performs the feature extraction/dimensionality reduction and prepare the data for the algorithm of analysis. The extraction algorithm takes the linguistic evaluation and the set of patterns as input set. The extraction algorithm is the module that analyzes the episodical behavior of the system. i.e the occurred patterns and its association evaluation, and it is able to abstract a model for its description. Algorithms been currently investigated are of the type of logical rules extraction (Paz et al, 2014). The extracted rules are used to explore and classify the universe of legal outputs of the original algorithmic

grammars, transition networks, chaos and self-similarity, genetic algorithms, cellular automata, neural networks, as well as other artificial intelligence methods.

system looking for inconsistencies until a terminal criteria for the quality of the obtained rules is meet. In order to perform a higher level analysis we include an algorithm for the construction of meta-rules, that take the extracted rules as input and perform the extractions of new rules of higher level. It should be noticed that this module does not starts until the iterations of the previous steps finish. Also, the extracted rules used to explore and classify the universe of legal outputs can also be used for the implementation of generative frameworks.

2.1 Implementation

For the first implementation of the methodology we used a generative model, based on probability templates, for beat patterns generation in the context of UK garage proposed by Collins(2003). UK Garage might be characterized by a high tempo, around 140 bpm, with a triplet swing groove, and the renowned '2 step' feel (Collins 2003). In the usual 4/4 metric at the second and fourth beats we found a definite snare strike, and at some point, we can found at the first and third beats a kick. A simplified model of the style can be codify in a regular grid of 16 points per bar (swing variations can be incorporated during performance). This points correspond with each of the sixteenth notes. The beat patterns are modeled in templates as being played of three independent sixteen bits streams, for instruments kick, snare and hihat. These streams codify the information presence/absence for the instruments. For simplify the data representation in the analysis the original algorithm was restricted to considered only cases when one instrument strikes at the same time. Although this can appear as a great simplification, produced patterns are still valid cases of the original, i.e it is a subset of the space of states. The set of probabilities (independent for each strike and taken from 0.0 to 1.0) are:

```
    //kick
[0.7, 0.0, 0.0, 0.1, 0.0, 0.0, 0.0, 0.2, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.3]
    //snare
[0.0, 0.0, 0.0, 0.0, 1.0, 0.0, 0.0, 0.5, 0.0, 0.2, 0.0, 0.0, 1.0, 0.0, 0.0, 0.3]
    //hihat
[0.0, 0.0, 1.0, 0.0, 0.0, 0.0, 1.0, 0.7, 0.0, 0.0, 1.0, 0.0, 0.0, 0.0, 1.0, 0.7]
```

This template can produce 128 valid outputs. From these, we produced short examples composed by 8 bars length. The first 4 bars correspond with a particular output of the system and the following 4 bars with another. We represent each output in an array containing 32 sixteenth notes, 16 from the first pattern and 16 from the second.

```
[1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32]
```

At each entrance for representing the different values that the variables can take we adopt the following convention: kick = 1, snare = 2, hi-hat = 3, and silences = 4. The cardinality of all possible outputs taken in this way is 16384 (128 times 128). It should be notice that all this cases including the extreme cases are valid outputs. From this set we extracted a sample to be evaluated by an expert. However, as different outputs have different probabilities, first all probabilities were calculated, and associated with each one of the 128 possible cases. Then, we create a sample to explore the space. We were interested in being sure that even the rare patterns, those with lower probability (probabilities range between 0.000101 and 0.070560), would represent interesting outputs. After the exploration we decided that all the space was interesting enough. By taking this into account, patterns were selected in a manner that every possible variation (class) in the training data were equally represented. From this set we create all possible combinations (400) among patterns. Then, combinations of each pattern with itself were removed. This was due that, for this experiment, we were interested in patterns constituted by different parts. So, the resulting (380) patterns will constitute our universe. Based on this set we trained and tested our model. Although the size of this set could appear small in comparison with the complete test set (of 16384 patterns), the amount of time needed for listen and evaluate the pieces makes the use of bigger sets impractical. However, the algorithm's restrictions, seeking to emulate a particular musical style, suggest that it is possible to obtain good results with few but well labeled data (as will be shown below).

The training set was evaluated by experts focussing in the perception of the "transition quality" between parts, considered as a high-level feature. We define the transition quality as "how good patterns match together" when they are perceived by a particular listener, in terms of the quality of the individual patterns and by considering the contrasts and repetition sensation between the parts. More precisely, this notion corresponds with a "motif coherence", (or coherence between rhythmic cells) of musical patterns. The evaluation of the transition will depend on the contrast and repetition points between parts A and B and on the moments on which those contrast and repetitions points are situated in the structure. For example, suppose that A and B are rhythmical motifs composed by 4 bars each one, as in our case. In the context of occidental music (at which the UK garage belongs), it could be considered that if the 4th bar of A coincides with the 4th bar of B, it will produce a greater sensation of coherence than if the 2nd bar of A coincides with the 4th bar of B. The evaluation was made by using linguistic variables (fuzzy quantifiers (Kosko, 1986)) defined as low, medium or high. The evaluated training set was processed with the Fuzzy inductive reasoning methodology and the algorithm for logic rules extraction (FIR + LR_FIR). A detailed explanation of both algorithms can be found in (Paz et al, 2014, Castro el al 2011). Each time, this rules were used for the classification of the training set, looking for inconsistencies, and then, for the classification of the universe created with the sample patterns. After the definitive training set were selected, several tests were performed using other training sets. Some of them were chosen randomly, and others by sampling the possible patterns considering its probabilities (e.g those with higher probability). It is remarkable that in the performed tests, the classification of the training set throws no inconsistencies, and that the proportion of classified patterns remains. In the same sense, after processing the rest of the universe (test data), the proportion of patterns that can be classified by rules is also similar among tests. And the order in the number of inconsistencies (in the test set 380 patterns) remains around ten percent.

The rules are expressed in the IF THEN form. Variable's number, for example V24 or V32, correspond to its location at the 32 sixteenth notes. The values in antecedents are 1=kick, 2=snare, 3=hihat, 4=silence. In consequences (system's output V33) are 1 = low transition quality, 2 = medium transition, and 3 = high transition quality. The expressions "Spec" and "Sens" stand for the usual specificity and sensitivity measures for the rules. The set of extracted rules are show below. They describe (or encode) the combination of variables that, according with the model, will produce a transition perceived as "low, medium and high".

```
LOW rules
IF V24 - 4 AND V32 - 1 THEN V33 - 1 Spec= 1 Sens=0.33
IF V16 - 1 AND V24 - 4 THEN V33 - 1 Spec= 1 Sens=0.18
IF V16 - 2 AND V17 - 4 AND V24 - 2 THEN V33 - 1 Spec= 1 Sens=0.078
IF V16 - 1 AND V26 - 2 AND V32 - 2 THEN V33 - 1 Spec= 1 Sens=0.078
IF V8 - 3 AND V24 - 3 THEN V33 - 1 Spec= 1 Sens=0.078
IF V10 - 2 AND V24 - 2 AND V26 - 2 AND V32 - 4 THEN V33 - 1 Spec= 1 Sens=0.078
IF V10 - 2 AND V16 - 2 AND V32 - 4 THEN V33 - 1 Spec= 1 Sens=0.078
IF V8 - 4 AND V26 - 2 AND V32 - 4 THEN V33 - 1 Spec= 1 Sens=0.078
IF V16 - 2 AND V24 - 2 AND V26 - 2 THEN V33 - 1 Spec= 1 Sens=0.059
IF V16 - 3 AND V24 - 3 AND V32 - 1 THEN V33 - 1 Spec= 1 Sens=0.059
IF V8 - 4 AND V20 - 4 AND V32 - 3 THEN V33 - 1 Spec= 1 Sens=0.059
MEDIUM rules
IF V4 - 4 AND V24 - 2 AND V32 - 2- 3 THEN V33 - 2 Spec= 1 Sens=0.2
IF V10 - 4 AND V24 - 2 AND V32 - 2 THEN V33 - 2 Spec= 1 Sens=0.14
IF V10 - 4 AND V17 - 1 AND V32 - 4 THEN V33 - 2 Spec= 1 Sens=0.11
IF V8 - 4 AND V20 - 1 AND V32 - 3 THEN V33 - 2 Spec= 1 Sens=0.11
IF V16 - 1 AND V24 - 2 AND V32 - 3 THEN V33 - 2 Spec= 1 Sens=0.086
IF V16 - 2 AND V24 - 2 AND V32 - 2 THEN V33 - 2 Spec= 1 Sens=0.086
IF V8 - 1 AND V24 - 2 AND V32 - 3 THEN V33 - 2 Spec= 1 Sens=0.086
IF V8 - 1 AND V16 - 1 AND V32 - 4 THEN V33 - 2 Spec= 1 Sens=0.057
IF V16 - 3 AND V24 - 4 AND V32 - 4 THEN V33 - 2 Spec= 1 Sens=0.057
IF V8 - 2 AND V24 - 3 AND V32 - 4 THEN V33 - 2 Spec= 1 Sens=0.057
IF V10 - 4 AND V24 - 4 AND V32 - 4 THEN V33 - 2 Spec= 1 Sens=0.057
IF V16 - 1 AND V26 - 4 AND V32 - 4 THEN V33 - 2 Spec= 1 Sens=0.057
IF V8 - 4 AND V24 - 1 AND V26 - 2 AND V32 - 3 THEN V33 - 2 Spec= 1 Sens=0.057
HIGH rules
IF V16 - 2- 3 AND V24 - 1 AND V32 - 1 THEN V33 - 3 Spec= 1 Sens=0.38
IF V24 - 1 AND V26 - 2 AND V32 - 1 THEN V33 - 3 Spec= 1 Sens=0.34
IF V10 - 4 AND V24 - 1 AND V32 - 1 THEN V33 - 3 Spec= 1 Sens=0.28
IF V16 - 3 AND V24 - 1 AND V26 - 2 THEN V33 - 3 Spec= 1 Sens=0.24
IF V8 - 2 AND V24 - 1 AND V26 - 2 THEN V33 - 3 Spec= 1 Sens=0.17 1
IF V8 - 1 AND V16 - 1 AND V24 - 3 AND V32 - 3 THEN V33 - 3 Spec= 1 Sens=0.069
IF V24 - 1 AND V32 - 2 THEN V33 - 3 Spec= 1 Sens=0.069
```

With these rules we explored the universe of all possible outputs. The results after the classification were: The total amount of patterns that satisfied at least one rule 10656. From this 5512 one low, 4080 one medium, and 2432 one high. Crossed patterns were: one rule low and one high = 192. One rule low and one medium = 1112. One medium and one high = 88. Finally patterns that satisfy rules low, medium and high = 24. In order to validate the model, we implement a "blind review" process of the classified outputs. Around %60 of the outputs were classified and %14 perceived with the desired property. However, when reviewing the misclassified outputs is is clear that the rules are expressing a partial property of the pattern and that more information is needed. This could be improved by using another level of analysis that can be performed by inspection or by using an algorithm for extraction of meta rules (rules of rules). However, it is important to take into account that as the rules are more specific they tend to describe particular patterns, and so, even though the outputs are the desired ones, the search could be limited.

3 Recapitulation

Although the system is in its primary implementation, the approach sounds promising. The obtained representation classified around 60% of all the possible patterns (with A+B structure) produced by the original system. This is a good result if we consider that the system is labeling instances unlabeled before. Furthermore, the analysis of all possible patterns considered with the same probability, allowed us to find patterns classified as high, that otherwise would have been difficult to obtain with the original model (the smallest associated probability is 0.0001). Remember that the decision of considered all the patterns with the same probability was taken when exploring the less probable patterns with the original probability distribution, patterns considered as "highly" interesting were found. Early programmed versions for pattern generation have been probed in the context of live performances with good results. This implementations were designed by filtering out the low and medium outputs. Other versions oriented to "desk composition" can include other options for selecting the patterns according to rules, for example by choosing those rules supported by great amount of instances for the main parts, and less supported rules for short passages.

Acknowledgments

We kindly thank the support of the Faculty of Sciences of the National Autonomous University of Mexico (UNAM).

References

1. Castro, F., Nebot, A. and Mugica, F., 2011. On the extraction of decision support rules from fuzzy predictive models. Applied Soft Computing, 11 (4), 3463-3475.
2. Collins, N., 2003 Algorithmic Composition Methods for Breakbeat Science ARiADA No.3 May 2003.
3. Collins, N., 2009 Musical Form and Algorithmic Composition. Contemporary Music Review Vol. 28, No. 1, February 2009, pp. 103-114
4. Collins, N., 2012. Automatic Composition of Electroacoustic Art Music Utilizing Machine Listening. Computer Music Journal, 36:3, pp. 8?23, Fall 2012 Massachusetts Institute of Technology.
5. Duncan, W., Kirke, A., Miranda, E,. Etienne B., Roesch, Slawomir J. Nasuto. Towards Affective Algorithmic Composition.(2013) Proceedings of the 3rd International Conference on Music & Emotion (ICME3), Jyväskylä, Finland, 11th - 15th June 2013. Geoff Luck & Olivier Brabant (Eds.)
6. Fernández J, (2013) AI Methods in Algorithmic Composition: A Comprehensive Survey. Journal of Artificial Intelligence Research 48 (2013) 513-582.
7. Kosko B (1986) Fuzzy Cognitive Maps. Man-Machine Studies, 24, 65-75.
8. Luke S, (2009) The Stochastic Synthesis of Iannis Xenakis. LEONARDO MUSIC JOURNAL, Vol. 19, pp. 7784, 2009. edited by Leonardo/the International Society for the Arts, Sciences and Technology, and published by the MIT Press.
9. Nierhaus, G., 2009. Algorithmic Composition. Paradigms of Automated Music Generation. Springer Wien, New York.
10. Paz-Ortiz I, Nebot À, Mugica F, Romero E. (2014) Using the Fuzzy Inductive Reasoning methodology to improve coherence in algorithmically produced musical beat patterns. Artificial Intelligence Research and Development 289 L. Museros et al. (Eds.) IOS Press, 2014.