A Fitness Function for Creativity in Jazz Improvisation and Beyond

Anna Jordanous

Creative Systems Lab / Music Informatics Research Centre,
School of Informatics, University of Sussex, UK
a.k.jordanous at sussex.ac.uk

Abstract. Can a computer evolve creative entities based on how creative they are? Taking the domain of jazz improvisation, this ongoing work investigates how creativity can be evolved and evaluated by a computational system. The aim is for the system to work with minimal human assistance, as autonomously as possible. The system employs a genetic algorithm to evolve musical parameters for algorithmic jazz music improvisation. For each set of parameters, several improvisations are generated. The fitness function of the genetic algorithm implements a set of criteria for creativity proposed by Graeme Ritchie. The evolution of the improvisation parameters is directed by the creativity demonstrated in the generated improvisations. From preliminary findings, whilst Ritchie's criteria does guide the system towards producing more acceptably pleasing and typical jazz music, the criteria (in their current form) rely too heavily on human intervention to be practically useful for computational evaluation of creativity. In pursuing more autonomous creativity assessment, however, this system is a promising testbed for examining alternative theories about how creativity could be evaluated computationally.

1 Introduction

The motivation for this work is to move towards achieving the goal of autonomous evaluation of creativity. It is initially intended as a test scenario in which to evaluate the usefulness of existing proposals for creativity assessment [1, 2]. Longer term, it provides a environment in which to implement and assess theories about how best to evaluate creativity computationally.

The computational system presented in this paper is designed to develop increasingly more creative behaviour over time. This behaviour is in the domain of jazz improvisation: evolving jazz improvisors based on maximising the level of creativity exhibited by the improvisor.

Ritchie [1, 2] has proposed a set of 18 formal criteria with which to evaluate the level of creativity exhibited by a creative system, using the artefacts which the system produces. Each criterion formally states a condition to be met by the products of the creative system, based on ratings of (at least one of) how typical the set of products are of the domain which the system operates in and how valuable those products are considered to be. The criteria have been adopted by a number of researchers for reflective evaluation of the creativity demonstrated by their creative systems [3, 4, 5].

This work explores the theory that Ritchie's criteria can be adapted and exploited for the purposes of implementing a fitness function for creativity. The criteria are applied to a generation of jazz improvisors and used to select which of these improvisors should be carried forward to the next generation.

There has been some interesting work on using evolutionary techniques such as genetic algorithms (GAs) to generate music [6, 7, 8, 9, 10]. Ritchie's criteria require that a creative system should be able to generate artefacts in a specified style or domain. Jazz music improvisation has been chosen for this domain as it encompasses a wide variety of styles of music under the umbrella term of "jazz"; from "trad jazz", through the "bebop" style exemplified by Charlie Parker to free improvisation. This lays the foundation for much creative opportunities, to be exploited by evolutionary tangents taken by the system.

2 Evolving Creative Improvisation: Implementation

This system is written in Java, using the jGap[1] package to implement the genetic algorithm and the jMusic[2] package for music generation.

The system evolves a population of "Improvisors" in the form of a set of values for musical parameters. In each generation of the genetic algorithm, these parameters are used to generate MIDI music. The parameters control the maximum number of notes can sound at any one time, the total number of notes in the piece, the key of the music, the range of pitches used, note durations, tempo markings and proportions of notes to rests, as well as the amount of variability allowed in several of these areas. Notes used are restricted to those in the blues scale[3] for that key[4]. Within the constraints of the musical parameters, random choices are used for the generation of musical improvisations.

2.1 Using Ritchie's criteria as a fitness function

Ritchie's criteria rely on two ratings of the improvisations produced: how valuable these are as jazz improvisations and how typical they are of the genre. These ratings are made using information about the artefacts. Ritchie makes "no firm proposals on what this information should be" (p. 75) and leaves open the question of how the rating scheme should be implemented. In the present version of this improvisation system, these ratings are provided by human assessment, following the example set in [3] (although as discussed in Section 4, if these ratings can be automatically generated, this would speed up the evolution process).

[1] http://jgap.sourceforge.net
[2] http://jmusic.ci.qut.edu.au
[3] The blues scale is traditionally used for jazz music. In the key of C, the scale consists of the pitches: C, Eb, F, F#, G, Bb, C
[4] Future plans for the system are to allow some chromaticism in notes used, or to allow it to evolve the notes that should be used, as another parameter.

For each set of parameters, a number of improvisations are generated (the exact number is determined by one of the parameters). Two improvisations are selected and played to the human evaluator, who rates each improvisation on its typicality as an example of jazz and on how much they liked it.

Ratings are recorded for the two selected individual improvisations. If there are further improvisations by that Improvisor, the mean values for the two pairs of ratings are used as ratings for the remaining improvisations that had not been rated. In this way, the evaluator is presented only with a selection of improvisations to rate, making the process more time-efficient [10]. This is analogous to the evaluator being given a "demo" of the Improvisor rather than having to listen to all their productions.

At this stage, all improvisations have a value rating and typicality rating and the 18 criteria can be applied to the products of each Improvisor. Each criterion is specified formally in [1] such that a criterion is either true or false for a given Improvisor, depending on whether scores derived from the typicality and value ratings are greater than some threshold θ, by setting suitable parameters to represent high/low typicality and value ratings (α, β and γ). In [1] Ritchie chooses not to specify what values the threshold and parameters should take, but does highlight discussions on this [3, 5]. For simplicity, in the current implementation $\theta = \alpha = \beta = \gamma = 0.5$, but experimentation with these values may be profitable.

The fitness value for an individual Improvisor is a score between 0 (no creativity) and 1 (maximally creative). Again a simple approach[5] is taken:

$$fitness = \frac{number\ of\ criteria\ satisfied}{total\ number\ of\ criteria} \tag{1}$$

After all Improvisors have been evaluated for fitness, the highest scoring Improvisor parameters are used to generate a new set of Improvisors, to act as the new generation of this population of Improvisors. The whole process is then repeated once per generation, until the user wishes to halt evolution.

3 Preliminary Results

The current implementation of the evolutionary improvisation system was tested informally, with a jazz musician (the author) providing the ratings required by Ritchie's criteria. Over several runs, it was able to produce jazz improvisations which slowly evolved from what was essentially random noise, to become more pleasing and sound more like jazz to the human evaluator's ears.

The question of whether the system was able to evolve more creative behaviour is still unresolved and is the main focus of further work on this project.

Some interesting comments can be made on the implementation of Ritchie's criteria for creativity. Each criteria manipulates one or both of the ratings for typicality and value of the music produced during run-time. In [1], Ritchie left

[5] This approach assumes all 18 criteria contribute equally to the creativity of a system; again though this is open to experimentation

unresolved the issue of generating this rating information, (p. 75), concentrating on how the ratings should then be processed once obtained. This work suggests though that these ratings are crucial to the success of evaluation; without reliable, accurate rating schemes, the application of the criteria becomes less useful.

Using a human evaluator as a "rating scheme" is easy for the system implementor but causes problems at run time. Even with restrictions placed on how many products the human must evaluate, any reliance on human intervention introduces a *fitness bottleneck* [6] into the system, such that the progress of evolution is significantly slowed down by having to wait for the evaluator to listen to and rate the music samples. Levels of expertise, fatigue during system runtime, individual bias in preferences and varying levels of concentration also affect the reliability of using human interaction in this way. Issues with using a human as part of a fitness function are discussed in greater depth in [10].

The system generates random improvisations using evolved parameters, without making use of any examples to guide the production of improvisations. A side effect of this is that there is no "inspiring set" of examples: as used by many of the criteria. Therefore those criteria currently do not add any new information to the creativity evaluation and do not contribute to the fitness function.[6]

Although nothing extraordinary has been generated thus far by the system, the need for human intervention has restricted any longer-term evolution of parameters from being attempted. It is proposed that at least some of the information currently supplied by user interaction can be derived or estimated automatically; then the system (and Ritchie's criteria) can be tested over more generations of evolution. This is discussed further in section 4.

4 Plans for Future Work

To attempt to escape the problems caused by the fitness bottleneck, automated methods of rating the improvisations for typicality and value are being explored.

- Genre classification methods are being investigated to help judge how typical an improvisation is of a specified genre.
- In this work, the value of an improvisation is interpreted as how pleasing an improvisation is to listen to. Currently, a value rating function is being implemented based on the perceptual principles described in [11].

Once the system has been extended to a degree where evolution can take place over a reasonable time frame, it will be tested over several runs. The results of evolution will be compared to similar tests carried out with human participants who will be asked to rate creativity of several of the improvisations. This will allow a fairer investigation of the appropriateness and accuracy of Ritchie's criteria for evaluating creativity.

On a longer term basis, this approach could also be used to test other theories of how best to evaluate the creativity of the products of a creative system (by

[6] If machine learning methods are used to automate typicality ratings, though, the inspiring set will then consist of the examples used during the learning process.

implementing them as a fitness function for creativity, in various domains). The various theories can be compared and contrasted to each other and to human judgements. We can also consider the efficacy of evaluating the creativity of a system based solely on the artefacts it produces, in comparison to evaluative frameworks that also take into account the creative process, or details about the system itself or the environment it operates in (e.g. as discussed briefly in [12]). Hence this proves a useful tool to enable us to move closer towards the goal of discovering how best to replicate human evaluation of creativity.

5 Acknowledgements

This work has benefitted from discussions with Nick Collins and Chris Thornton. Chris Kiefer has provided useful advice during the implementation of this system.

References

[1] Ritchie, G.: Some empirical criteria for attributing creativity to a computer program. Minds and Machines **17** (2007) 67–99

[2] Ritchie, G.: Assessing creativity. In: Proceedings of AISB Symposium on AI and Creativity in Arts and Science. (2001)

[3] Gervas, P.: Exploring quantitative evaluations of the creativity of automatic poets. In: Proceedings of the 2nd. Workshop on Creative Systems, Approaches to Creativity in Artificial Intelligence and Cognitive Science (ECAI 2002). (2002)

[4] Haenen, J., Rauchas, S.: Investigating artificial creativity by generating melodies, using connectionist knowledge representation. Proc. of ECAI Workshop on Computational Creativity (2006)

[5] Pereira, F., Mendes, M., Gervas, P., Cardoso, A.: Experiments with assessment of creative systems: an application of Ritchie's criteria. In: Proceedings of the Workshop on Computational Creativity (IJCAI 05). (2005)

[6] Biles, J.A.: Genjam: A genetic algorithm for generating jazz solos. In: Proceedings of the International Computer Music Conference, Denmark (1994)

[7] Burton, A.R., Vladimirova, T.: Generation of musical sequences with genetic techniques. Computer Music Journal **23**(4) (1999) 59–73

[8] Dahlstedt, P.: Thoughts on creative evolution: A meta-generative approach to composition. Contemporary Music Review **28**(1) (2009) 43–55

[9] Gartland-Jones, A., Copley, P.: The suitability of genetic algorithms for musical composition. Contemporary Music Review **22**(3) (2003) 43–55

[10] Todd, P.M., Werner, G.M.: Frankensteinian methods for evolutionary music composition. In: Musical Networks: Parallel Distributed Perception and Performance. MIT Press, Cambridge, MA (1999) 313–340

[11] Huron, D.: Tone and Voice: A Derivation of the Rules of Voice-Leading from Perceptual Principles. Music Perception **19**(1) (2001) 1–64

[12] Jordanous, A.: Defining creativity: Finding keywords for creativity using corpus linguistics techniques. Proceedings of the ICCCX Concerence (2010)