Proceedings, The Thirteenth AAAI Conference on
Artificial Intelligence and Interactive Digital Entertainment (AIIDE-17)

MTG: Context-Based Music Composition for Tabletop Role-Playing Games

Lucas N. Ferreira, Jim Whitehead
Department of Computational Media
University of California, Santa Cruz
Santa Cruz, CA, 95060

Abstract

This project aims to compose background music in real-time for tabletop role-playing games. To accomplish this goal, we propose a system called MTG that listens to players' speeches in order to recognize the context of the current scene and generate background music to match the scene. A speech recognition system is used to transcribe players' speeches to text and a supervised learning algorithm detects when scene transitions take place. In its current version, a scene transition occurs whenever the emotional state of the narrative changes. Moreover, the background music is not generated, but selected based on its emotion from a library of hand-authored pieces. As future work, we plan to generate the background music considering the current scene context and the probability of scene transition. We also consider to retrieve more information from the narrative to detect scene transitions, such as the scene's location and time of the day as well as actions taken by characters.

Introduction

This project proposes a system called MTG (Music for Tabletop Games) that automatically composes background music in real-time for tabletop role-playing games (RPG). Its main goal is to augment the experience of players in tabletop RPGs through algorithmic music (Nierhaus 2009). The object of our research is Dungeons and Dragons (D&D), a tabletop RPG in which players control characters, known as player characters (PC), in a story told by the dungeon master (DM). The motivation of this project is that RPG players often manually add background music to their sessions in order to enhance their experiences (Bergström and Björk 2014). Thus, unless one of the players constantly selects the background music according to the game context, the music might not match the current scene (e.g., the PCs could be battling a dragon while a calm music is being played). Having one of the players constantly selecting the background music is not ideal, as the player might get distracted from the game. By developing MTG we expect to answer the following research questions (RQs):

RQ 1: How to automatically detect scene transitions in spoken narratives?

RQ 2: To what extent do information retrieval methods support music composition in spoken narratives?

RQ 3: How can movie and video-game scoring techniques be applied to compose music for tabletop RPG?

RQ 4: To what extent can procedurally generated music augment the experience of tabletop RPG players?

MTG listens to players' speeches to recognize the context of the current scene and generate background music to match it. A speech recognition (SR) system is used to transcribe players' speeches to text and a supervised learning (SL) algorithm detects when scene transitions take place. In its current version, a scene transition occurs whenever the emotional state of the narrative changes. This is achieved by using the Naive Bayes (NB) approach to classify the narrative according to a model of emotion that includes the following emotions: Happy, Calm, Agitated, and Suspenseful. Moreover, in this current version, the background music is not generated, but selected based on its emotion from a library of hand-authored pieces. This initial system was evaluated via a user study in which we used it to select the background music for excerpts of videos of a D&D campaign.

The next step of this project consists of generating background music considering the current scene context and the probability of scene transition. We plan to evaluate this music generator via user studies with players in real D&D sessions. We also consider to retrieve more refined information from the narrative to detect scene transitions (such as characters' locations and actions) and use a reinforcement learning (RL) approach (Sutton and Barto 1998) to improve the classification model as the game progresses.

We believe that this project opens several directions for future research as the system is likely to be directly applicable to other tabletop games, especially storytelling-based games. Moreover, it could be applied beyond the realm of games. For example, problems such as background music generation for daily social events such as having dinner with friends or hosting a party.

Related Work

Algorithmic Music Composition (AMC) consists of using algorithms to combine musical parts in order to create a piece of music (Nierhaus 2009). AMC systems have been developed to compose music for different purposes, such as

jazz improvisation (Biles 1994), harmonization of chorales (Ames and Domino 1992) and musical therapy (Mattek 2011). However, there is a very scarce body of literature on AMC for narratives. Also, to the best of the authors' knowledge, MTG is the first AMC system for tabletop games. One of the few works directly related to MTG was proposed by Davis and Mohammad (2014). They used a lexicon-based approach to detect emotions in novels and an AMC technique to compose simple piano pieces that evoke these emotions. MTG differs from Davis and Mohammad's work because it is based on spoken narratives automatically transcribed to text through SR systems. Hence, MTG does not have access to the structure of the narrative (sentences, paragraph and chapter) and the text often is grammatically incorrect due to inaccuracies of speech recognition systems.

To compose meaninful music for narratives, we need to retrieve information from them, which can be achieved using Natural Language Processing (NLP) techniques such as Named Entity Recognition (NER) (Liu et al. 2011) and Text-based Emotion Recognition (TER) (Strapparava and Mihalcea 2008). TER is the process of computationally detecting emotions (e.g., anger, sadness and surprise) from text and its systems are usually lexicon-based or machine learned-based. For example, Strapparava and Mihalcea (2008) used a lexicon-based method similar to Davis and Mohammad (2014) to classify the emotions in newspaper headlines and a Naive Bayes classifier to detect Ekman's emotions in blog posts. NER labels sequences of words in a text which are the names of things, such as characters and location names (Liu et al. 2011). For example, Vala et al. (2015) presented an eight stage approach for character identification, which builds a graph where nodes are names and edges connect names belonging to the same character.

Context-based Music Composition for Tabletop RPGs

Up to now, our system uses the NB classifier to detect scene transitions based on the variations of emotion in the D&D narrative. We model this solution as a finite state machine (FSM) where the NB decides when to change states. A state transition occurs whenever the emotion with the highest probability of transition exceeds a threshold d_t. Whenever the FSM changes to a state s, a song that evokes the emotion of s is picked randomly. This system does not account for the structure of the scenes, so it plays the same song in the beginning, in the middle and in the end of the scene. Moreover, the transitions between two scenes are sharp, since one song is simply played after another and the end of the previous can be different from the beginning of the next one.

In order to solve the sharp transition problem, the next step of this project consists of generating background music instead of just selecting hand-authored pieces. We envision to generate music in a two steps approach: defining a set of musical instruments to use and then scoring a musical piece for that set. Music will be generated in a constructive way, i.e., based on rule-sets where musical notes are fundamental blocks. Musical theory will be used to encode these rules, which will determine various elements of the generated music (such as tempo, tonality, rhythm, etc) according to the context of the scene. These rules will allow MTG to generate musical structures like intro blocks, loops, and transition blocks. MTG will use the probability of transition from one state to another in order to decide what structure to use.

We need to retrieve information from the narrative in order to discover the context of the scene as well as when a transition is about to happen. Currently, we only retrieve the emotional state of the scene. We plan to retrieve more refined information such as what characters are in the scene as well as their location, time, and actions. This will support the generation of more dramatic and meaningful background music. For example, if the PCs are in a dense forest at night, the system might compose a suspenseful music. Whereas if they are in an open forest during the day, it might compose a more calm song. Retrieving such information will also allow us to create music themes for specific characters and locations.

The current method for detecting scene transition uses a SL approach. We plan to improve it by using a hybrid learning approach: an initial model is created through SL which can be improved continuously with player feedback via RL (Sutton and Barto 1998). Thus, players interact with MTG through voice commands to give it positive feedback whenever they like the generated music. With this hybrid approach, we expect MTG to compose sufficiently meaningful music for any starting player and to improve its compositions over time. This is an important improvement since different players have different play and speaking styles.

Preliminary and Expected Results

We created a dataset of sentences from a D&D campaign labeled according to our model of emotion. It was used to evaluate the accuracy of NB on detecting scene transitions (RQ1). NB had an accuracy of 64% on average when detecting scene transitions. The quality of NB on creating music selection was evaluated with a user study: human subjects watched and evaluated short videos of a D&D game session with background music selected by the system. This evaluation is performed in comparison to selections made by the video authors. The results of this study showed a clear preference by the participants for the system's selections.

Since this study was performed with short videos, it did not provide insights on how the system performs in a long game session. The videos were on average 1 minute long and they had at most 4 scene transitions. Thus, we don't know how bored the users would be if they were exposed to longer videos. Moreover, with this video-based evaluation, we can't evaluate to what extent the generated background music can augment the experience of players.

The advantage of this video-based evaluation is that it is considerably fast, so we plan to use it to evaluate the quality of the generated songs (RQ2 and RQ3). Once the output of MTG is good enough, we evaluate how it can improve players experience in a set of real D&D sessions (RQ4). We expect that MTG will be able to augment the experience of D&D players by making them more immersed in the game. Specially, we expect to see players using more and better role-playing (acting within the narrative).

Acknowledgments

We would like to thank Levi Lelis and Rafael Padovani for their collaboration and support on this project, including all the interesting ideas and the many rounds of reviews. Lucas N. Ferreira also acknowledges financial support from CNPq Scholarship (200367/2015-3).

References

Ames, C., and Domino, M. 1992. Cybernetic composer: an overview. In *Understanding music with AI*, 186–205. MIT Press.

Bergström, K., and Björk, S. 2014. The case for computer-augmented games. *Transactions of the Digital Games Research Association* 1(3).

Biles, J. A. 1994. Genjam: A genetic algorithm for generating jazz solos. In *ICMC*, volume 94, 131–137.

Davis, H., and Mohammad, S. M. 2014. Generating music from literature. *arXiv preprint arXiv:1403.2124*.

Liu, X.; Zhang, S.; Wei, F.; and Zhou, M. 2011. Recognizing named entities in tweets. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1*, 359–367. Association for Computational Linguistics.

Mattek, A. 2011. Emotional communication in computer generated music: Experimenting with affective algorithms. In *Proceedings of the 26th Annual Conference of the Society for Electro-Acoustic Music in the United States*.

Nierhaus, G. 2009. *Algorithmic composition: paradigms of automated music generation*. Springer Science & Business Media.

Strapparava, C., and Mihalcea, R. 2008. Learning to identify emotions in text. In *Proceedings of the 2008 ACM symposium on Applied computing*, 1556–1560. ACM.

Sutton, R. S., and Barto, A. G. 1998. *Introduction to reinforcement learning*, volume 135. MIT Press Cambridge.

Vala, H.; Jurgens, D.; Piper, A.; and Ruths, D. 2015. Mr. bennet, his coachman, and the archbishop walk into a bar but only one of them gets recognized: On the difficulty of detecting characters in literary texts. In *EMNLP*, 769–774.