Semantic Hashing for Video Game Levels

Aaron Isaksen, New York University

Christoffer Holmgård, New York University

Julian Togelius, New York University

We use semantic hashing, an unsupervised machine learning technique based on a type of deep neural network called autoencoders, to categorise video game levels. We show how this technique can be used on dungeon room maps from The Legend of Zelda and segments of 2D-platformer levels from Super Mario Bros. We also discuss how this technique can be useful for game designers, AI-assisted game design, and procedural content generation.

1 Introduction

SEMANTIC hashing is an automated, non-linear method for finding similarities between instances of content using pretrained neural networks. The technique was originally developed for determining document similarity for searching through text documents to find related examples [1]. In this paper, we show how a variant of the semantic hashing technique [2] based on autoencoders [3, 4] can be used to find similarities in video game content.

Semantic hashing uses unsupervised learning to perform clustering of similar content. *Unsupervised* means the data does not need to be labelled by a game designer or player before processing. Instead, the system can take in level data – segmented into regions such as rooms in a dungeon or slices of a linear level – and determine which regions are similar given no additional knowledge of what the level data means.

Once a semantic hashing neural network is trained, it is simple to input new level data and have the network return the category which the new data is closest to. Thus, the network also allows us to classify unseen future data using the same structure.

We demonstrate the use of semantic hashing on two types of classic video game levels: top-down dungeon RPG maps from The Legend of Zelda [5] and side-scrolling platformer levels from Super Mario Bros. [6]. Level data is obtained from the Video Game Level Corpus [7], which takes level maps and preprocesses them to load into machine learning algorithms.

1.1 Application to Game Design

While this paper demonstrates a proof of concept to show that semantic hashing can effectively classify and cluster new game content automatically and quickly, we believe this method has a number of potentially useful applications in game design, procedural content generation, and user-generated content.

To begin, the statistical analysis of game content can help a designer understand which game content is common, unique, and under/over-represented in their design. For example, a designer may want to know how much variety and repetition there is in the game. Similarly, multiple designers working on the same game might wish to compare their content to see how much of it is similar. In a mixed-initiative system, an AI agent may offer suggestions to a human designer based on the type of work they are currently creating – to enable this, the machine must be able to identify what makes content similar or unique.

Procedural content generation allows a machine to generate new game assets [8], and a large category of such methods use machine learning to produce new content from existing game data [9]. This enables new methods in autonomous generation, co-creation, mixed-initiative design, and repair of content for games. With automated content classification, one can control the output of a procedural content generation system; e.g. if we wanted to generate levels with many bridges, we could use the classifier to check for this. Or, more generally, we could use an autoencoder to ensure that machine generated content has a certain typological distribution, ensuring variations or patterns in the output.

In games where players contribute content, semantic hashing could inform the game what the players are creating by assigning the new content to the most likely category already in the game. The autoencoder could learn these categories automatically from the existing game content, and can be retrained as more is created.

Isaksen, A., Holmgård, C. and Togelius, J., 'Semantic Hashing for Video Game Levels', *Game & Puzzle Design*, vol. 3, no. 1, 2017, pp. 10–16.