Tools for Videogame Discovery Built Using Latent Semantic Analysis

James Owen Ryan[1,2], Eric Kaltman[1], Michael Mateas[1], and Noah Wardrip-Fruin[1]
[1] Expressive Intelligence Studio
[2] Natural Language and Dialogue Systems Lab
University of California, Santa Cruz
{jor, ekaltman, michaelm, nwf}@soe.ucsc.edu

ABSTRACT

We introduce two tools for videogame discovery built using latent semantic analysis, a foundational technique in statistical natural language processing. The first, GameNet, is an explorable network in which 11,829 games are linked according to how related they are. The second tool, GameSage, takes a user's free-text description of an idea for a game and integrates with GameNet to present an interactive listing of the existing games that are most related to her idea. Both tools are hosted online as web apps.

1. INTRODUCTION

There is a scarcity of tools online by which game scholars, game designers, and general game players can seek out new games of interest. These three groups have different (but complementary) needs for such prospective tools: a scholar may want to find games related to a game (or broader topic) of study; a designer may want to find existing games that are similar to her own idea for a game, or to a game she is developing; a player may seek a certain type of game to play. For all of these examples, the most promising existing method is likely using Google and some combination of disconnected websites. In this paper, we introduce two tools for game discovery that together address all of these use cases.

2. LATENT SEMANTIC ANALYSIS MODEL

Both of the tools we present here are underpinned by a *latent semantic analysis* (LSA) model trained on Wikipedia articles describing videogames[1]. LSA is a natural language processing (NLP) technique by which words are attributed vectorial semantic representations according to their contextual distributions across a large collection of text [1]. From a corpus of text, a *co-occurrence matrix* of its words and documents is built; this matrix specifies which words occurred in which documents (and thereby which documents words occurred in). The columns and rows in this matrix can be thought of as vectors that represent the meanings, in an approximate sense, of the words and documents that they correspond to—this is called a *vector space model* of semantics. LSA is an example of such a model, but its hallmark is that it reduces the dimensionality of these vectors by a matrix factorization algorithm. Remarkably, doing this allows the model to infer semantic associations that are not encoded in the full co-occurrence matrix. This ability to learn global associations from local co-occurrences is the achievement of LSA and what led to it becoming one of the major NLP techniques of the last twenty years. Having an LSA model, one can easily calculate how semantically related any of its documents are by taking the cosine between their LSA vectors. In corpora in which each document pertains to a specific individual concept, these relatedness scores can reasonably be utilized as a measure of the relatedness of the concepts themselves. Relying on this notion, we trained an LSA model on a corpus comprising Wikipedia articles for 11,829 videogames. By this model, we can quantify how related any two of these games are by taking the cosine between their LSA vectors.

[1]We invite the interested reader to consult [3], in which we discuss LSA and our model at much greater depth.

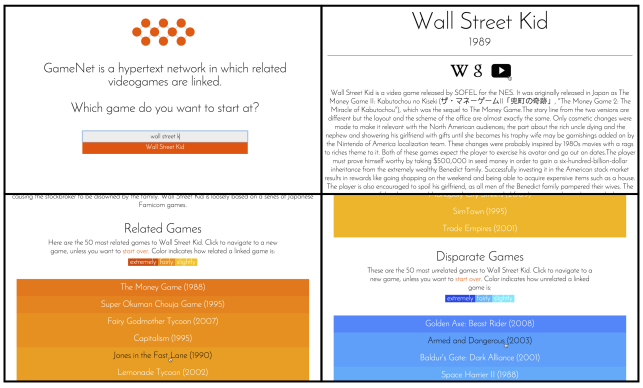

Figure 1: GameNet entry for *Wall Street Kid*.

3. GAMENET

GameNet is a tool for game discovery in the form of a network in which related games are linked. The tool is composed of entries for the 11,829 games that are included in our LSA model. Each game's entry includes links to GameNet entries for other games that are related to that game, as well as to gameplay videos and other informative sources found elsewhere on the web. Figure 1 shows excerpts from the GameNet homepage and its entry for the Nintendo Entertainment System game *Wall Street Kid* (1990).

At the GameNet home page, the user indicates which game she wishes to start at and is brought to that game's

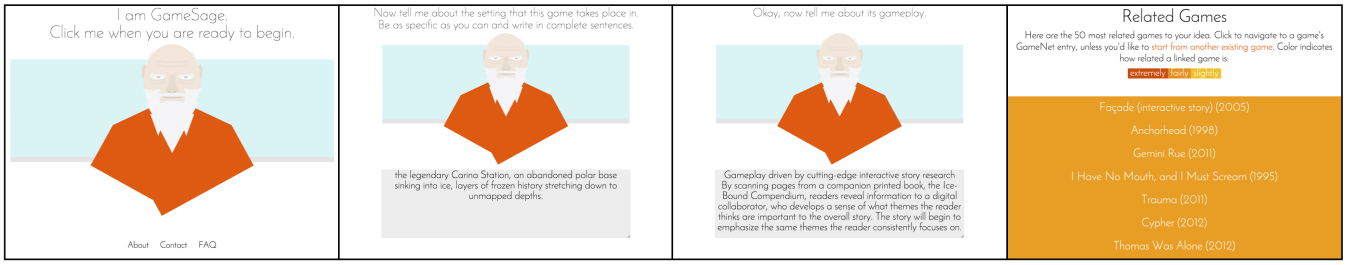

Figure 2: Excerpts from a GameSage session; the user describes *Ice-Bound* and obtains a list of related games.

entry. Here, in a header, the game's title and year of release are prominent, as well as links to the game's Wikipedia article and Google Images and YouTube search results found using autogenerated queries; included as well is a summary of the game that was extracted from Wikipedia. Below these elements is the core of the entry, which is a color-coded listing of the fifty most related games to the game at hand. As alluded to in the previous section, GameNet judges how related any two games are by taking the cosine between their documents' LSA vectors. To promote exploration, the related games are stylized as hyperlinks to their own GameNet entries. Finally, below the listing of related games is a listing of the most *unrelated* games to the game at hand. These links can potentially serve as portals to corners of the medium that were previously unknown to the user.

GameNet is intended primarily as a tool for game scholars to find games related to a research topic. In [3], we conducted an expert evaluation of GameNet in which six published game scholars indicated that the tool is especially useful for the scholar who wishes to explore a relatively unfamiliar area of games, but that it may also be used to discover unforeseen cases related to topics that have already been thoroughly researched. Additionally, we envision the tool being useful more generally to game enthusiasts who may seek out new games to play or may use the tool to learn more about the medium, for instance, by exploring the network to find and learn about games of historical interest.

4. GAMESAGE

GameSage is a tool that takes free-text input describing an idea for a videogame and lists the existing games that are most related to that idea. This tool utilizes the notion in LSA of *folding in*, whereby a new document that was not used during model training is fitted with a representation in the semantic space derived by the model. By treating the user's input text (which specifies her game idea) as a corpus document (on par with the videogame Wikipedia articles we used to train our LSA model) and folding it in, we are able to derive an LSA vector for the idea. From here, we determine which existing games (from among GameNet's 11,829 games) are most related to the game idea by using cosine similarity, just as we did in constructing GameNet.

Figure 2 shows excerpts from a session with the tool in which a user describes the in-development indie game *Ice-Bound* [2] and gets back a list of related games (including other indie games, story-focused games, and games used in narrative-technology research). At the GameSage home page, the system is personified as an animated sage character who proceeds to ask the user seven questions about her game idea, each pertaining to a particular facet of the (prospective) game. After the final question is answered, the system concatenates the responses and preprocesses this text using the same procedure we enacted to preprocess our corpus (see [3]). From here, the preprocessed text is attributed an LSA vector by folding it into our model, and GameNet's 11,829 games are then ranked according to how related they are to the user's idea. Finally, GameSage makes a request to GameNet to generate an entry for the game idea, which the user is then taken to.

The obvious usage of this tool is finding games that are related to an idea for a game, which could prove helpful as a way of gathering insight during the early stages of game development. To evaluate GameSage's utility in this regard, we have just conducted a user study in which more than 200 novice game designers used it to find games related to their own works in progress; this data is currently being analyzed.

5. CONCLUSION

In this paper, we have introduced two tools for game discovery and outlined their major intended usages: GameNet is intended for scholars to find games related to their research topics; GameSage may be used by game designers to find games related to their works in progress. We also envision several other use cases for these tools. They could be used to rediscover a game whose title one has forgotten: assuming the game is included in GameNet, a user can describe it to GameSage and then search for it among the games most related to her description. Further, a game player could describe the type of game she wishes to play. Extending this, library patrons could choose games in physical collections by describing the type of thing they are seeking—we are currently exploring this very notion. While Google and other search engines perform poorly on long search queries, GameSage performs better as game ideas get longer and more specific. As such, we see these tools as instantiating a fundamentally new way of searching for games.

6. LINKS

Try the tools at `http://gamecip.soe.ucsc.edu/projects`.

7. ACKNOWLEDGMENTS

This project was made possible in part by Institute of Museum and Library Services grant LG-06-13-0205-13.

8. REFERENCES

[1] S. T. Dumais. Latent semantic analysis. *Annual Review of Information Science and Technology*, 2004.

[2] A. A. Reed, J. Garbe, N. Wardrip-Fruin, and M. Mateas. Ice-bound: Combining richly-realized story with expressive gameplay. *Proc. FDG*, 2014.

[3] J. O. Ryan, E. Kaltman, M. Mateas, and N. Wardrip-Fruin. What we talk about when we talk about games: Bottom-up game studies using natural language processing. In *Proc. FDG*, 2015.