

2013

Using Lexicons Obtained from Online Reviews to Classify Computer Games

Miaoqi Zhu

DePaul University, mzhu8@mail.depaul.edu

Xiaowen Feng

DePaul University, xfang@cdm.depaul.edu

Follow this and additional works at: <http://aisel.aisnet.org/sighci2013>

Recommended Citation

Zhu, Miaoqi and Feng, Xiaowen, "Using Lexicons Obtained from Online Reviews to Classify Computer Games" (2013). *SIGHCI 2013 Proceedings*. 24.

<http://aisel.aisnet.org/sighci2013/24>

This material is brought to you by the Special Interest Group on Human-Computer Interaction at AIS Electronic Library (AISeL). It has been accepted for inclusion in SIGHCI 2013 Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.

Using Lexicons Obtained from Online Reviews to Classify Computer Games

Miaoqi Zhu

School of Computing
DePaul University
mzhu8@mail.depaul.edu

Xiaowen Fang

School of Computing
DePaul University
xfang@cdm.depaul.edu

ABSTRACT

This paper presents a new method for characterizing computer games based on lexicons obtained from online game reviews. Inspired by the lexical approach to define personality traits (Ashton, 2007), we hypothesize that game players would have used natural language in describing computer games and play experience over the time and the fundamental traits of computer games would be encapsulated in player's languages. Therefore, the traits of computer games could be explored by investigating descriptive terms within game reviews.

The lexical hypothesis states that people will want to talk about personality traits that they view as having important consequences in their lives (Ashton, 2007). As a result, people will inevitably invent some words to describe those who exhibit high or low levels of these essential traits. Over long periods of time, words that describe important traits should become established in every language (Ashton, 2007). Likewise, computer game traits can be characterized by a set of adjective terms. They are consistently perceived by different stakeholders (e.g., game players, developers) at various points of time. Therefore, it is argued that the lexical approach is applicable to study.

A revised lexical approach was adopted and five phases were introduced in our approach: 1) Collecting game review 30,252,536 records of contents (e.g., user reviews, product specifications, etc.) were downloaded from three major game websites (i.e., Gamespot.com, Gamestop.com, IGN.com); 2) Building a dictionary of game descriptive adjectives. Natural Language Processing applications were developed. WordNet was employed as the English dictionary to determine the Parts of Speech. Approximately 11,000 unique adjectives were found. 3) Extracting game player's ratings of adjectives. The online reviews were converted to another dataset as follows: first, each adjective term in the list from the 2nd phase was treated as an individual item, which was then saved as the field name of a table; second, all online reviews were retrieved one at a time. For each review, if an adjective was present in this review, the value for this adjective (field) was set to 1. Otherwise, 0 was registered. The end product of this phase

was a table (696,801 by 11,000) of values "0" and "1". 4) Factor analysis. An exploratory factor analysis with Varimax rotation was conducted. Challenges were confronted and solved due to the size of table to be loaded. 5) Exploring computer game traits. The factor analysis discovered 70 distinct factors. However, some initial observations motivated additional efforts to re-organize them. For instance, some factors seemed to contain adjectives with similar meanings. Hence, three more steps were taken such as pre-processing the factor, labeling the factors, and grouping the factors into traits.

14 traits were identified. Taking a few for illustrations, Trait T3 reflects the power felt during game play. The power was indicated by adjectives like "murderous", "merciless", "almighty", and "dominant." Trait T4 is no doubt pointing to a category of strategy games. Some of the examples are "strategic", "tactical", "historical", "cooperative", and "economic." Trait T5 refers to the type or nature of actions (e.g., simulating, racing). Adjectives vary for different type of actions. For example, "driving", "high-speed", and "fast" were used to describe racing actions while "fighting", "martial", and "mortal" were used to describe fighting actions. Trait T7 embodies the theme of narratives. The adjectives in this group provided different themes of game stories such as "mysterious", "magical", and "sci-fi". Trait T13 is mobility of computer games, of which most popular adjective was "mobile." Trait T9 describes the degree of realism or fantasy of computer games. Some example adjectives are "artificial", "animate", and "surreal."

Instead of assigning a game to one or more genres that are overlapping and ambiguous, a researcher could refer to these 14 traits to characterize it more precisely. For instance, "Sid Meier's Civilization" is viewed as a turn-based strategy game, one of which features resource macro-management. The key traits of it would be T3, T4 and T5.

The 14 traits provide a more objective and reliable way to investigate computer game development by better characterizing games. Moreover, this revised lexical approach provides behavioral researchers a revolutionary tool for analyzing large dataset contributed by the user.