Structural Analysis of Literary Texts using Social Network Frameworks

Anon
Anon
Anon
Anon

Abstract—A work of literature shows different characteristics depending on its genre or author. Generally, features of literature can be revealed by linguistic analysis. However, the process of linguistic analysis is complicated and does not have a common standard. In this paper, we numerically calculate the relationship between agents that appear in literature and construct a relational network. The structure of the relational network is determined by the relationships between characters. A network that is composed of characters can be said to be a social network of a virtual world, so many existing social network analysis methods can be applied. We selected more than 20 novels including Lev N. Tolstoy’s “War and Peace”, J.K. Rowling’s “Harry Potter”, traditional novel “Three Kingdoms(三國志演義)” for an experiment. We introduce a visualization method for virtual social graphs and some useful analysis. The main contribution of this paper is that our model can be used to reveal the deep structure of a work of fiction using graph topology rather than traditional categories such as short or long novel.

Keywords—Social Network, Virtual characters, Graph theory, Minimum spanning tree, Literature analysis

I. MOTIVATION

In general, linguistic analysis methods, such as the study of the contextual meanings of words, are used for identifying features of literature. However, analysis of literature using semantically-oriented techniques make it difficult to automatically determine meaningful interpretations. In order to analyze the subtle meanings of words, it is necessary to read a work of literature manually from beginning to end. Therefore, literature analysis using the linguistic approach is complicated and time consuming. In our previous papers [8], and [9], we numerically calculated the relationships among agents appearing in a work of literature and constructed a Social Network from Fiction (SNF). The agents can be any types of words that are used in a work of literature. For example, we can construct SNF from relations between “Me” and “House,” “Movie” and “coke,” “eat” and “diner,” and the like. This method does not require any complicated linguistic analysis, so it can be applied in a relatively short time using computers. In this paper, we consider only characters as the agents. A network consisting of characters can be referred to as a social network of a virtual world. Therefore, many existing social network analysis methods can be applied to SNF. For example, we can extract major relations from a social network using a minimum distance spanning tree of the network. In addition, we can predict relations between two characters by examining their common neighbors.

Our approach involves three processes: (1) locating characters in the textual space of the fictional work, (2) computing the closeness of all character pairs based on textual distance, and (3) constructing the SNF by mapping characters to vertices and their strength of relations to edges between characters, with edge weights corresponding to the frequency and position of their appearance.

We perform this analysis on a corpus of more than 20 novels, including Tolstoy’s “War and Peace,” J.K. Rowling’s “Harry Potter,” and a well-known Chinese novel “Three Kingdoms(三國志演義)”.” Our results provide evidence that the features of a work of literature can be obtained from an SNF graph using a graph-theoretical approach. In the following sections, we survey related work in studies of literature and social networks. We then describe the methods we use to extract characters and construct an SNF graph, along with our graph-theoretical approach to analyzing their characteristics. Finally, we present and analyze experimental results.

II. RELATED WORK

A. Literature Analysis with Networks

Computer-based analysis of literature generally involves analysis at the word level. Such analysis focuses on discovering the authorial style and the lexical patterns of the words used [3]. In recent years, researchers have started to study the co-occurrence properties of words in the natural language space [7], [10].

The co-occurrence patterns of some word pairs may reveal important features hidden in the text, because all words in a text are closely related. One interesting application of this co-occurrence analysis provides clues for identifying spam or vicious reply messages [10]. Another application of word appearance patterns is biomedical text mining to find biologically and medically meaningful features by analyzing very large amount of textual data [2]. For example, hidden functions of genes can be estimated or predicted by biomedical text-mining tools, helping biologists produce a new drugs quickly and efficiently.