Social Network Analysis for Virtual Characters Appeared in Fiction Literatures

ABSTRACT
The literature shows different characteristics depending on its genre and 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 calculated the relationship between agents that appear in literature, and constructed the relation network. Structure of relation network is determined by relationship between characters. A network that made up of the characters can be said the social networks of the virtual world, so many existing social network analysis methods can be applied. We selected more than 20 novels for experiment, which include Lev N. Tolstoy's "War and Peace", J.K. Rowling's "Harry Potter", traditional novel "Three Kingdoms(三國志演義)". We introduce the visualization method for virtual social graph and come useful analysis.

Categories and Subject Descriptors
H.5.1 [INFORMATION INTERFACES AND PRESENTATION]: Multimedia Information Systems—Artificial, augmented, and virtual realities

General Terms
Theory, Measurement

Keywords
Social Network, virtual characters, entropy, minimum spanning tree, graph theory

1. MOTIVATION
In generally, linguistic analysis such as studying contextual meaning of the word are used for identifying feature of literatures. However, the process of linguistic analysis is complicated and does not have a common standard, so features of literature vary over analysts. In order to analyze the subtle meaning of words, it is necessary to read literature from beginning to the end. Therefore, analyzing by human is complicated and time consuming. In our previous paper [7], we numerically calculated the relationship between agents that appear in literature, and construct the relation network. The agent can be any words that were used in literature. For example, we can construct the network from relations of "Me" and "House", "Movie" and "coke", "eat" and "diner", etc. This construction method does not require any complicate linguistic analysis, so it is possible to construct in a short time by rule. In this paper, we only considered characters as agent. A network that made up of the characters can be said the social networks of the virtual world, so many existing social network analysis methods can be applied. For example, we can extract major relations from networks by Minimum distance spanning tree of agent network. And, we can measure the importance level of characters by analysing depth of the spanning tree. Experimentally, we show features of literature in both numerically and visually by analysing fiction network. Furthermore, we prove other useful information of literature can be obtained through a graph theoretical approach.

2. RELATED WORK
Recently, researchers started to study properties of co-occurrence words in natural language space [6, 9]. Since all words in a text are closely related, the co-occurrence pattern of some pair of words may reveal the important features hidden in the text. One interesting application of this co-occurrence analysis gave the clue to identify spam or vicious replying messages [8]. Another application of word appearance pattern is biomedical text mining to find the biologically and medically meaningful features by analyzing the huge textual data [2].

Typical social network studies based on textual data focus on the structured data, including email and Tweet messages and phone SMS messaging texts for social network construction [5]. The main goal of this work is to find the most influential person. However, until now, little work has been done on extracting social relations between individuals from text. One notable work has published on constructing conversational networks from literature [10]. Elsas et al proposed a method for extracting social networks from nineteenth-century British novels [4]. Their conclusion is twofold. First there is an inverse correlation between the amount of dialogues in a novel and the number of characters in that novel. Second a significant difference of social interaction in the 19th century English novel is basically geographical.