A Study on Online Social Networks Theme Semantic Computing Model

Chen Fu, Department of Computer Science, Beijing Foreign Studies University, Beijing, China
Xu Yuemei, Department of Computer Science, Beijing Foreign Studies University, Beijing, China
Ni Yihan, Department of Computer Science, Beijing Foreign Studies University, Beijing, China

ABSTRACT

The widespread use of Mobile Intelligent Terminals and ubiquitous access to networks has enabled online information sources including Weibo and Wechat to bring huge impact to the society. Only a few words of network information can expand rapidly and catalyze the generation of a huge amount of information. The highly real-time content, fission-like spreading rate and enormous public opinion guiding forces created in this process will cast great influence on the society. Thus, semantic computing on online social networks and research on topics about emergencies have great significance. In this article, a numerical model of text semantic analysis based on artificial neural network is proposed, and a semantic computational algorithm for social network texts as well as a discovery algorithm for emergencies is provided with reference to the information provided by the social nodes itself and the semantic of the text. Through the numerization of text, the calculation and comparison of semantic distance, the classification of nodes and the discovery of community can be realized. In this article, semantic vector of micro-information for nodes and closure extension of semantic extensions are defined in order to build up an equivalence of short sentences, and in turn realize the discovery of emergencies. Then, huge quantities of Sina Weibo contents are collected to verify the model and algorithm put forward in this article. In the end, outlooks for future jobs are provided.

KEYWORDS:
Artificial Neural Nets, Burst Topics Discovering, Online Social Networks, Semantic Distance Theme Mining, Theme Semantic Computing

1. INTRODUCTION

With the rapid development of information technology, the Internet network has formed a huge cyberspace. Mass network content, flat transmission ability of network and various disruptive innovation network applications have a deep influence on the social development and human progress. The widespread use of Mobile Intelligent Terminals and ubiquitous access to networks, in particular, has enabled online information sources including Weibo and Wechat to bring huge impact to the society. The huge impact to the society brought by the trace network information of very few words enables the comprehensive analysis of the microblogging social characteristics and the research of utility maximization positive energy transmission to have great significance. The highly real-time content, fission-like spreading rate, smaller transmission load and amazing public opinion guiding forces of the online networks give it social influence that should not be underestimated.

The analyses of the influence of online social networks have got the attention of the world’s first-rate research institution a long time ago (Wolfe, 1997). Online social networks such as Twitter,
Facebook, LinedIn and Sina Weibo have great differences from traditional social networks on their behavior characteristics, transmission means and influence. The most obvious external characteristic is the huge flow of instantaneous and short text such as the 140 characters of Twitter and Weibo, SMS and search query. Even if there is no limit to the number of characters in the post massage, we also have reason to believe that people still prefer to using short text to express the information needed spreading widely and rapidly. Therefore, strengthening the processing of short text information is of great significance.

The semantic of short text plays a decisive role in the spreading of it. Semantic computing of short text is of great significance in community discovery of online social network, network topology structure analysis, recommendation of nodes, targeted advertising, management of organization structure, identification of terrorist organizations and so on. For instance, when someone raises a question on the quiz module of Sina Weibo, deciding to whom to send the question to get an accurate answer is a typical problem about organization structure management. If the semantic vector coordinates of the node and the question are known, we can get the semantic distance as well as the set of nodes which forward that question by computing the semantic norm. Also, the traditional online community discovery often recognizes the community network topology through the connection between nodes instead of the calculation of semantic distance (Pan-Pan, & Jian, 2013; McPherson, Smith-Lovin, & Cook, 2001). The measurement of information without the support of semantic information lacks sufficient foundations. Using vectors to describe the information of short text online has become a problem of great significance. Considering the characters that a Weibo post is short but rich in semantics, has strong semantic impact and is high real-time, and combining it with the semantic information and historical information of the node itself, in this paper, we put forward a semantic vector measurement model to describe Weibo content. We calculate the semantic distance by describing semantics through multidimensional vectors. We can detect emergent topics, identify opinion leaders, conduct rumor analysis and recommend Weibo content and nodes based on this model.

Combining with the features above and the current work, the contributions of this thesis include:

- In this article, we build the corpus according to plenty of Weibo short texts and do training about numeralization and vectorization of text in combination with general news corpus.
- Combining with the characteristic that texts are short, we “amplify” the semantic meanings of keywords and “narrow” the semantic meanings of words that are not important in order to build a clearer profile of the semantic meanings of short texts.
- In order to classify and organize Weibo short texts depending on numerical and vectorial results, besides amplifying and narrowing semantic meanings, we also build up a model of the equivalence classes of short texts, and improve its ability of classification through closure extension of semantic meanings.

The amplification of structural semantics and textual semantics. The model of deep semantics of short texts modeling is built up based on artificial neural network in order to acquire statistical semantic vectors of short texts in Weibo. And then we amplify semantic meanings based on the vector distance and the semantically correlative lexicon.

2. RELATED WORK

Research work related to this paper includes the following three parts and we will discuss them respectively.
An Ontology-Based Content Model for Intelligent Web Content Access Services
www.igi-global.com/article/ontology-based-content-model-intelligent/3079?camid=4v1a

bpCMon: A Rule-Based Monitoring Framework for Business Processes Compliance
www.igi-global.com/article/bpcmon/181301?camid=4v1a