A Preliminary Study on the Behavioral Perception of Social Clusters in Uyghur Network

Liu Shuxian 1*, Cui Qing 1

1School of Information Science and Engineering, Xinjiang University, Urumqi, Xinjiang Province, 830046, China

*Corresponding author’s e-mail: lsx623@163.com

Abstract. The virtual society of the Internet is becoming a new front for the breeding, spreading, planning, and implementation of criminal organizations. Implementing network information monitoring is one of the important means to ensure network information security and maintain social stability. At present, there are many researches on Chinese network content monitoring. The monitoring research on Uyghur network is in the exploration stage, and it mainly focuses on content security research. This paper conducts a research on behavioral perception of network social clusters based on social computing. Based on the theory of social computing, it studies the basic theories of perception, analysis, prediction and management of Uyghur network group behavior, and reveals the behavioral characteristics and dynamic evolution of social clusters, and provides technical support for combating criminal activities on the Internet.

1. Introduction

At present, the monitoring of the network is mainly based on content monitoring, and its main processing method is based on keyword filtering. This technology is simple and is one of the basic means of network monitoring. However, this technology cannot involve semantic understanding, and the context is not clear. In recent years, domestic research methods for Chinese network monitoring have gradually transitioned to semantic analysis. The content of Chinese is very rich, including Chinese and Chinese minority languages. As a region where ethnic minorities live in China, Xinjiang uses a large number of ethnic languages such as Uighur and Kazakh in addition to Chinese. However, the research on network monitoring of minority languages started late.

Uyghur as a part of Chinese, is a type of pinyin text, horizontally typeset from right to left. The word is the smallest language unit in Uighur that can be used independently. Words are separated from words by spaces, and there is no problem of word segmentation like Chinese. Therefore, the collected information can be directly filtered by sensitive words. Related technologies are mature in the study of natural language information processing. The Uyghur processing method can draw on these results.

The words in Uyghur are composed of stems and additional components. The same word will show different ways of writing in different contexts, and the same letter can also be written in multiple ways. Therefore, in order to ensure the correct implementation of the research, it is necessary to complete the semantic expansion of the basic dictionary. In addition to defining semantic synonyms and contextual relationships, it is necessary to supplement the different expressions of the words. It is feasible in computer information processing.

There is a big difference in the syntactic structure between Uyghur and Chinese. Chinese belongs to the SVO(Subject + Verb + Object) language, while Uyghur is the SOV(Subject + Object + Verb)
language. Its grammatical structure is relatively standardized, so it is feasible to conduct semantic orientation analysis on Uyghur.

With the popularization of network applications, the network monitoring efforts are intensified in time. We acquire the basic theory of behavior perception, analysis, prediction and management based on social computing network social clusters, and the behavioral characteristics of social clusters and their dynamic evolution laws are obtained. It not only provides technical support and assistance for network security supervision, but also can be applied to the classification of user groups and the investigation of real intentions in e-commerce websites, the development and design of Uyghur intelligent search engines, and the development and construction of intelligent expert systems.

2. Research status
The related research progress on the public opinion application system mainly includes: Websense Enterprise developed by Websense Company[1], based on the website classification database Websense Master Database for network content filtering. But it does not support content restore and backup functions, and can’t support after-the-fact forensics. SurfControl[2] has three product lines: web filtering system, mail security gateway and anti-spyware software. However, it only supports web applications, and does not support content restoration and backup and after-the-fact forensics. Niksun NetDeteetor of Niksun [3] supports recombination functions of HTTP, Email (SMP/POP3/IMAP4), Telnet, FTP, MIME attachments and IM (ICQ, Yahoo, MSN, AOL). The University of Wisconsin's NLAGARA project[4-5] studies the collection and monitoring of information from the Internet and then wraps it into XML data streams for retrieval and filtering.

In recent years, due to the out-of-control of network public opinion, mass incidents have frequently erupted, which has caused people's attention to network public opinion monitoring, which has formed a huge market demand. Aiming at the huge potential market demand, many research institutes and enterprises have developed a number of network public opinion monitoring application systems.

At present, there are many researches on Chinese network content monitoring in China [6-7]. The monitoring research on Uyghur network is in its infancy, and it mainly focuses on content security research. From the perspective of technological development, combined with the characteristics of Uyghur, the research results of technologies such as social computing cannot be directly copied into the realization of Uyghur network monitoring, but the related algorithms have been improved can provide theoretical support and technical support for Uyghur network supervision and analysis.

3. The problem to be solved
In order to break through the key technologies of the two stages of perception and analysis involved in the orderly management of social cluster behavior under the Uyghur network, two key scientific issues need to be resolved.

3.1. Scientific Issue 1
The Semantic Convergence of Uyghur Network Cluster Behavior. The research of this scientific problem aims to reveal the intrinsic semantic association and hierarchy between the massive data of Uyghur network cluster behavior information, and provide support for subsequent analysis. The Uyghur network contains information such as text, image, video, motion and space that exhibit multi-source, heterogeneous and massive characteristics. The subject behavior of these information is widely distributed, with long duration and redundant information. In order to effectively realize the social cluster information perception under the cover of massive heterogeneous noise information in the Uyghur network, it is necessary to combine the results of databases, statistical learning, information theory and other fields to study the clustering of behavioral awareness and the aggregation and management of massive heterogeneous information, and to realize cluster behavior description, multi-level time-varying association between entities, information consistency and integrity processing, integrated management and retrieval of distributed massive heterogeneous data, and to establish a social security information intelligence acquisition and perception framework.
3.2. Scientific Question 2
The Evolution of Complex Network Dynamics of Cluster Behavior. The research of this scientific problem aims to reveal the human behavior characteristics of social clusters in Uyghur network environment, the mapping law of Uyghur network and social cluster model structure, and the evolution mechanism of social cluster behavior dynamics. The life cycle of social cluster behaviors that are closely related to events is essentially the process of the formation and evolution of social clusters. In order to grasp the formation and evolution of social cluster behavior, it is necessary to use information technology to dig out the social structure of cluster behavior hidden under data and representation, and use the system model to describe the operational state and evolution process of cluster behavior social structure. Cluster behavior not only has obvious human behavioral characteristics, but also has complex network structure characteristics. Using human behavioral characteristics, we can detect and identify cluster behaviors. Using its complex network structure characteristics, we can grasp the mechanism of abnormal cluster behavior formation and evolution. In complex networks, abnormally active individual behaviors follow the principle of maximizing their value to find associated goals. High-value individuals are more likely to become common association targets, and the structural features with high agglomeration coefficients emerge to form cluster behavior eigensystems. Some individuals of the intrinsic social structure corresponding to the cluster behavior are driven by specific task objectives, after dynamic evolution to form a complex network. Under the guidance of accidental sudden factors, specific cluster behavior emerges. The intrinsic social structure and dynamic evolution of cluster behavior are keys to understanding the internal mechanism of the formation, evolution, subsidence and rebound of the social cluster behavior of Uyghur network. The formation of complex network of cluster behavior is closely related to the individual behavior. Agent's individual behavior choice is the dynamic mechanism of cluster behavior evolution. Therefore, studying the human behavioral representation method and discovering the social structure of cluster behavior and the evolution law of cluster dynamics are the key issues in the study of social cluster behavior in Uyghur network.

(1) The issue of vocabulary diversity. As a sticky language, Uyghur has a rich system of inflections. The relationship between Uyghur words and words, the grammatical function of words in sentences is mainly reflected by morphological changes. In order to achieve network monitoring, it is necessary to solve the problems of vocabulary diversity, synonyms and polysemy.

(2) A regular analysis model for the evolution of network group activities. In order to realize the monitoring of the network user group and avoid the occurrence of realistic mass incidents, establishing a good analysis model of the evolution law of group activities is a key issue that must be solved in this project.

4. Feasibility analysis
A large amount of information on the Internet, such as blogs, forums, chats, consumption records, and e-mails, is a mapping of people and organizational behavior in real life. Network data can be used to analyze the behavior patterns of individuals and groups, thus deepening understanding of people's lives, organization and social. The study of computational sociology involves three interrelated problems: the way people interact, the form of social group networks, and their evolution. The study of these questions can help people answer many social questions. The μsik (micro simulation kernel) parallel simulation engine is an open source parallel and distributed simulation micro-kernel simulation engine developed by University of Georgia Perumalla. Its design draws on the micro-kernel idea in the operating system and divides the functions in the parallel simulation engine into different levels of services, the most basic service is placed at the core, other services are placed on the periphery, and the outer layer can call the inner service. National University of Defense Technology based on μsik realized the design of a social computing experimental dynamic simulation engine, which provided a simulation implementation idea for the application of social computing in Uyghur network information detection.
From the key technologies of the typical social clusters in the Uyghur network, which can be collected and controlled, the common scientific problems are extracted. Based on the massive data in the Uyghur network, combined with social psychology and social structure theory, we will research on the perception, mechanism and management of social cluster behavior, establish a regular model of social cluster behavior, and establish a computational framework for social cluster behavior analysis and management under the Uyghur network.

4.1. Research on scientific issues
It starts from three levels: cluster behavior perception, pattern and law mining, and management guidance mechanism. Cluster behavioral awareness aims to focus attention on the cluster behavior of specific topics from the Uyghur network data representation. Cluster behavior patterns and law mining aim to explore the inherent scientific laws of cluster behavior, and master the rules and patterns of cluster behavior self-evolution and stimulated evolution. The combination of cluster behavior perception and law is the decision-making basis for optimizing the management decision of cluster behavior and realizing the clusterable behavior.

4.2. Cluster behavioral awareness
Uyghur network emotional vocabulary is used to form reinforcement learning mechanism to improve the credibility of information acquisition. By using the Internet mining system to collect the initial clues of the virtual world social cluster, the spatial and temporal directionality of the corresponding real world social behavior is preliminarily judged, and then the selective attention mechanism is used to verify whether the Uyghur network community activity semantics are consistent with the online clues. Such cross-enhancement, mutual confirmation, to obtain timely, accurate and comprehensive cluster behavior attention focus and semantic understanding.

4.3. Pattern and law mining aspects
In the aspect of pattern and law mining, using the idea of collaborative training technology in the field of machine learning, cross-reference interactive verification and co-evolution based on the data mining-based computational model and the computational model based on complex network analysis, explore the simplicity, invariance and changing laws of the social cluster model under the complex representation of the Uyghur network.

4.4. Computing framework design
In the design of computing framework, qualitative analysis and quantitative calculation are combined, through the mutual transformation of qualitative knowledge and quantitative data, integration of data, model, knowledge, experience, and psychological, ‘human-machine combination, people-oriented’ collaborative work and other means, we can establish a social cluster behavioral informatics computing framework, build a verification and evaluation system, design configurable, reusable, scalable computing modules and data and knowledge interfaces to make it suitable for social cluster behavior management services under the Uyghur network.
The purpose of this paper is to use social computing technology to study the group behavior perception in the network, analyze the evolution law of group activities, and provide technical ideas for purifying the network and ensuring network security.

**Project funding:** National Natural Science Foundation of China (61762085)

**References**

[1] Websense, Inc. WebSecurity, Internet Filtering and Internet Security Software-Websense, Inc. http://www.websense.com/global/zh-cn/index.php

[2] Surfcontrol. Content security defense expert. http://www.surfcontrol.cn

[3] Niksun.NIKSUN – Forensics&Compliance. http://www.niksun.com

[4] Jianjun Chen, David J. Dewitt, Feng Tian, Yuan Wang. NiagaraCQ: A scalable Continuous Query System for Internet Databases. SIGMOD, 2000:379-390

[5] Jeffrey Naughton, David Dewitt, David Maier. The Niagara Internet Query System Electronics Letters, 1999, vol.35, no.1:14-16

[6] Zhang Rui. (2012) Design and implementation of SMS network public opinion information monitoring system. Modern intelligence, 32(3): 68-71

[7] Wang Feiyue. (2011) Research on Dynamic Netizen Group Based on Social Computing and Parallel System. Journal of University of Shanghai For Science and Technology, 33(1): 8-17