早稲田大学審査学位論文
博士（人間科学）
概要書

個人化情報活用のためのパーソナルデータと挙動解析による統合モデリング手法
Unified Modeling and Analyzing of Personal Data and Behaviors for Individualized Information Utilization

２０１４年７月

早稲田大学大学院　人間科学研究科
周　暁康
ZHOU, Xiaokang

研究指導教員：　金　群　教授

With the rapid development of emerging computing paradigms, such as Ubiquitous Computing, Cloud Computing, Social Computing, and Mobile Computing, we have been continuously experiencing a fast change from all walks of our work, life, learning and entertainment. The high accessibility of SNS (Social Networking Service), coupled with the increasingly widespread adoption of wireless mobile computing devices, enables more and more populations to continuously generate larger amount of data from different environments, which represents more information in terms of an individual's behavioral habits and daily routines. It is said that a new era of big data has arrived. Among those big data, the so-called personal data, not only referring to the collection of data that is generated from an individual, but also any data that is related to an individual, has become a crucial source of innovation and value. It is indubitable that the valuable information hidden in personal big data can benefit individuals in various aspects.

Generally, information utilization is regarding various applications involving people and information together, including information seeking and recommendation for an individual, and information sharing and knowledge creation within a group. To take advantage of the considerable size of personal data for individualized information utilization, it is of crucial importance to build a well-structured user model, further make sense of individuals' intentions or needs, analyze various information behaviors and social activities, and better understand the user contexts. However, in many cases, the large scales of personal big data, dynamically generated from a variety of systems and different devices, is always with different data structures or no structure, and coupled with lots of useless noise data. It makes it difficult to find piece of relevant information that fits users' time-varying needs. Research works have tried to find a flexible and efficient solution, but it's still far from satisfaction due to the more complexity and heterogeneity of the personal big data.

In this study, to facilitate individualized information utilization and sharing not only for the individuals, but also for the groups or communities, we concentrate on the computational approaches to unified modeling and analyzing of the personal data and behaviors. The chaotic data will be systematically organized and managed to form the associative information, including time-varying individual intentions and additional information for the descriptions of data relations. The individuals' information behaviors and social activities will be analyzed to extract the behavioral features and calculate the similarities among them. Furthermore, the related individuals will be connected in a dynamically socialized networking according to the calculation of their dynamical and potential correlations. Their multi-dimensional profiling will be built, and the social communities will be discovered based on the outcomes from the analysis of personal data. Finally, both the behavior patterns and user correlations will be considered together to develop an integrated recommendation mechanism to provide the users with individualized support.

Firstly, a unified framework of data integration and organization called *Organic Streams*, is proposed to analyze and organize the personal big data. The new concept of organic stream, which is designed as a flexibly extensible data carrier, is introduced and defined to provide a simple but efficient means to formulate, organize, and represent the personal big data with inherent and potential relations. It can also be regarded as a logic metaphor to meaningfully analyze and process the raw stream data into an associatively organized form based on the individual needs. A heuristic mechanism is developed and applied to capture users' time-varying interests or needs, and aggregate and integrate the relevant data together to obtain the associative information.

A behavioral analysis approach is proposed to detect and calculate the social influence hidden in the individual behaviors, and model and analyze the sequential behaviors in the task-oriented processes with formal descriptions. The action patterns are extracted from an individual user's sequential behaviors toward a certain purpose, and the behavioral similarities among a group of users are then calculated and described based on the action patterns. The perceived social influence can be utilized to analyze and describe users' social relationships, and the extracted action patterns as well as their similarities can help improve the quality of user contexts and assist the recommendations in the task-oriented processes.

Based on these basic models and methods, the *DSUN* (Dynamically Socialized User Networking) model, as a viable alternative way to obtain larger information sources and connect more and more people together, is constructed to describe and represent users' implicit and explicit social relationships, namely the characteristics - based relationships and influence-based relationships, using the valuable outcomes of analyzing personal data and individual behaviors. A set of measures are introduced and defined to measure and describe the detail of user correlations, which can dynamically calculate and build a connection between two related people in a specific time period, according to their static and dynamical feature based similarity and interactional behavior based social influence. A series of attributes are defined and analyzed to build the multi-dimensional user profiling, which can facilitate the search of information sources by finding favorable users in both global (e.g., *hub user* and *promotion user*) and personalized (e.g., *contribution user* and *reference user*) way. Three algorithms are developed to discover the multi-types of social communities considering both the dynamical user correlations (e.g., *strong correlation-based tie* and *weak correlation-based tie*) and profiling (e.g., *user profiling-based tie*), which can recommend users to join different communities in accordance with their different intentions, so as to promote the information sharing and collaborative work.

Moreover, as an application of the proposed approaches, an integrated recommendation method is proposed to provide users with the individualized learning guidance and support. A hierarchical model is presented to

describe the relations among learning actions, activities, sub-tasks and tasks within a user community for the task-oriented learning process. The *LA-Pattern* (Learning Action Pattern) is defined to discover and represent an individual user's learning behavior patterns extracted from sequences of learning actions, and the *Goal-driven Learning Group* is proposed to analyze and describe the similarities of learning behaviors among a group of users. Based on these, an integrated mechanism is developed for the goal-driven learning recommendation in accordance with the analysis of behavior patterns and user correlations, which can provide the target user with the most suitable learning action as the appropriate next learning step to complete a specific learning goal.

To demonstrate the feasibility and effectiveness of our methods, two experimental studies are conducted respectively. The experimental results with the analyses conducted using the Twitter data demonstrate the high usability and practicability of our proposed *DSUN* model which can assist personalized information utilization and sharing in both favorable user finding and social community discovering. The empirical analysis results conducted in a community-based learning system illustrated that the calculated LA-Patterns and Goal-driven Learning Groups can correctly describe the users' learning behaviors and their similarities as well, which can be applied to frequency-based learning pattern recognition and categorization according to different learning goals. And the evaluation results showed the usefulness of our proposed recommendation method that can effectively guide users to pursue their learning purposes and facilitate the task-oriented learning process.

This study is expected to benefit both individuals and communities, not only in the systematical processing of personal big data which can capture the time-varying individual needs and generate associative information from the chaotic data to facilitate the personalized information seeking and social knowledge creation, but also in the dynamical constructing of social relationships which can help build a well-structured user model and involve increasing people into a well-connected social networking to promote the information sharing and recommending. The unified modeling and analyzing approach presented in this study can facilitate the individualized information utilization from chaotic data to associative information, and further to connected people.