Quantitatively Evaluating the Influence of Online Social Interactions in the Community-Assisted Digital Library

YongHong Tian
Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100080, P.R. China
+86 010 82649529
yhtian@ict.ac.cn

TieJun Huang
Graduate School of Chinese Academy of Sciences
Beijing 100039, P.R. China
+86 010 82649514
tjhuang@ict.ac.cn

Wen Gao
Graduate School of Chinese Academy of Sciences
Beijing 100039, P.R. China
+86 010 82649208
wgao@ict.ac.cn

ABSTRACT

Online social interactions are useful in information seeking from digital libraries, but how to measure their influence on the user's information access actions has not yet been revealed. Studies on this problem give us interesting insights into the workings of human dynamics in the context of information access from digital libraries. On the basis, we wish to improve the technological supports to provide more intelligent services in the ongoing China-America Million Books Digital Library so that it can reach its potential in serving human needs.

Our research aims at developing a common framework to model online social interaction process in community-assisted digital libraries. The underlying philosophy of our work is that the online social interaction can be viewed as a dynamic process, and the next state of each participant in this process (e.g., personal information access competency) depends on the value of the previous states of all participants involving interactions in the period. Hence, considering the dynamics of interaction process, we model each participant with a Hidden Markov Model (HMM) chain and then employ *the Influence Model*, which was developed by C. Asavathiratham as a Dynamic Bayes Net (DBN) of representing the influences a number of Markov chains have on each other, to analyze the effects of participants influencing each other. Therefore, one can think of the entire interaction process as a DBN framework having two levels of structure: the local level and the network level. Each participant ***i*** has a local HMM chain $\Gamma(A_i)$ which characterizes the transition of his internal states in the interaction process with state-transition probability $\sum_j d_{ij} P(S_t^i \mid S_{t-1}^j)$ (Here *states* are his personal information access competence in different periods, while *observations* are his information access actions). Meanwhile, the network level, which is described by a *network graph* $\Gamma(D^T)$ where $D = \{d_{ij}\}$ is the *influence factor matrix*, represents the interacting relations between participants. The strength of each connection, d_{ij}, describes the *influence factor* of the participant ***j*** at its begin on the one ***i*** at its end. Hence, this model describes the dynamic inter-influence process of the internal states of all participants involving online interactions.

To automatically build the model, we need firstly to extract observed features from the data of online social interactions and information access actions. Obviously, the effects of interactions are stronger if messages are exchanged more frequently, or the participants access more information in the online digital libraries during the period of time. Based on this consideration, we select the *interaction measure* $IM_t^{i,j}$ and the *amount of information* IA_t^i as the estimation features of x_t^i. The interaction measure $IM_t^{i,j}$ and the amount of information IA_t^i parameterize the features calculated automatically from the data of online social interactions between the participants ***i*** and ***j***, and the features calculated from the data of information access actions respectively. Secondly, we need to develop a mechanism for learning the parameters d_{ij} and $P(S_t^i \mid S_{t-1}^j)$. Given sequences of observations $\{x_t^i\}$ for each chain ***i***, we may easily utilize the Expectation-Maximization algorithm or the gradient-based learning algorithm to get their estimation equations.

We ran our experiments in the online digital library of W3C Consortium (www.w3c.org), which contains a mass of news, electronic papers or other materials related to web technologies. Users may access and download any information and materials in this digital library, and also may free discuss on any related technological problems by means of its mailing lists. Six users were selected in our experiments to collaboratively perform paper -gathering tasks related to four given topics. Any user might call for help from the others through the mailing lists when had difficulties in this process. All participants were required to record subjective evaluations of the effects that the others influenced his tasks. Each experiment was scheduled by ten phases. And in each phase, we sampled $IM_t^{i,j}$ and IA_t^i for each participant and then fed them into the learning algorithms to automatically build the influence model. By comparing with the subjective influence graphs, the experimental results show that the influence model can estimate approximately the influences of online social interactions.

Categories and Subject Descriptors: H.1.2 [Models and Principles]: User/Machine Systems -Human factors, Human information processing.

General Terms: Algorithms, Experimentation, Human Factors

Keywords: Community-Assisted Digital Library, Online Social Interactions, the Influence Model, Statistical Feature Extraction

JCDL'02, July 13-17, 2002, Portland, Oregon, USA.
ACM 1-58113-513-0/02/0007.