

REVERSE ENGINEERING A HIDDEN MARKOV MODEL FOR COMPLEX SOCIAL SYSTEMS

By

Hung-Ching Chen

An Abstract of a Thesis Submitted to the Graduate

Faculty of Rensselaer Polytechnic Institute

in Partial Fulfillment of the

Requirements for the Degree of

DOCTOR OF PHILOSOPHY

Major Subject: COMPUTER SCIENCE

The original of the complete thesis is on file
in the Rensselaer Polytechnic Institute Library

Examining Committee:

Dr. Malik Magdon-Ismail, Thesis Adviser

Dr. Mark Goldberg, Member

Dr. William (Al) Wallace, Member

Dr. Mohammed Zaki, Member

Rensselaer Polytechnic Institute
Troy, New York

May 2008
(For Graduation August 2008)

ABSTRACT

We present a machine learning methodology (models, algorithms, and experimental data) to discover the agent dynamics that drive the evolution of the social groups in a community. Agent dynamics are governed by *micro-laws*, which determines an agent’s behaviors. There are a number of challenges, the first and foremost being the complex nature of the *micro-laws* needs to represent even a very simple society. Agents may have discrete attributes together with continuous parameters, which inevitably lead to a mixed optimization problem, and each agent has its own attributes, which also interact with other agents’ attributes, suffering from combinatorial and dimensionality curses. Another challenge is that the data upon which to answer the question is not available—typically social groups (especially online groups) do not announce their memberships, and a researcher has to infer groups from observable macro-quantities such as communication statistics. We set up the problem by introducing an agent-based hidden Markov model for the agent dynamics: an agent’s actions are determined by *micro-laws*. Nonetheless, we learn the agent dynamics from the observed communications without knowing state transitions. Our approach is to identify the appropriate *micro-laws* corresponding to an identification of the appropriate parameters in the model. The model identification problem is then formulated as a mixed optimization problem. To solve the problem, we develop a multistage learning process for determining the group structure, the group evolution, and the *micro-laws* of a community based on the observed set of communications among actors, without knowing the semantic contents. Finally, to test the quality of our approximations and the feasibility of the approach, we present the results of extensive experiments on synthetic data as well as the results on real communities, such as Enron email and Movie newsgroups. Insight into agent dynamics helps us to understand the driving forces behind social evolution.