# Over-Fitting and Error Detection for Online Role Mining

*Victor W. Chu, University of New South Wales, Sydney, NSW, Australia*

*Raymond K. Wong, University of New South Wales, Sydney, NSW, Australia*

*Chi-Hung Chi, Intelligent Sensing and Systems Laboratory, CSIRO, Hobart, TAS, Australia*

## ABSTRACT:

*Recent research has attempted to use role-based approaches to recommend mobile services to other members among the same group in a context dependent manner. However, the traditional role mining approaches originated from the domain of security control tend to be rigid and may not be able to capture human behaviors adequately. In particular, during the course of role mining process, these approaches easily result in over-fitting, i.e., too many roles with slightly different service consumption patterns are found. As a result, they fail to reveal the true common preferences within the user community. This paper proposes an online role mining algorithm with a residual term and an error term, that automatically group users according to their interests and habits without losing sight of their individual preferences and random errors. Moreover, to resolve the over-fitting problem, the authors relax the role definition in role mining mechanism by introducing quasi-roles based on the concept of quasi-bicliques. Most importantly, the new concept allows us to propose a monitoring framework to detect and correct over-fitting in online role mining such that recommendations can be made based on the latest and genuine common preferences. To the best of the authors' knowledge, this is a new area in service recommendation that is yet to be fully explored.*

*Keywords:* *Incremental, Online Role Mining, Over-Fitting, Pseudo-Roles, Quasi-Bicliques, Service Recommendation*

## INTRODUCTION

The popularity of smartphones equipped with context aware capability, e.g., location detection, offers many opportunities for service providers to provide personalized services to their customers, e.g., to make these devices more intelligent and adaptive to user behaviors by automatically recommending suitable services (Adomavicius & Tuzhilin, 2011; Wörndl, Schulze, & Schlichter, n.d.). This functionality is especially useful when a time-poor individual is on the move but is in need of some services shortly after.

Most of the service recommendation techniques are based on individual intelligence or the local knowledge of users (Su & Khoshgoftaar, 2009) rather than the common knowledge among the user community. In this paper, we use the notion of "roles" to represent

DOI: 10.4018/jwsr.2012100101

user groups. Users who play the same role are likely to share the same preferences and behavior patterns. Once a user's role in a certain context is recognized, the services closely related to the role can be recommended to the user. Furthermore, as pointed out by previous work on role modeling, e.g., (Wong, Chau, & Lochovsky, 1997), roles can have different granularity or levels and are better represented in a hierarchy, e.g., in form of ontology.

Although the concept of roles has been used in task/service recommendation, e.g., (Fukazawa, Naganuma, Fujii, & Kurakake, 2006), most of the existing approaches typically implement the role-task ontology manually. This is not practical for mobile environments where both roles and spatiotemporal contexts are changing dynamically while their owners move. This paper adapts the concept of roles from the domain of Role-Based Access Control (RBAC) (Sandhu, Coyne, Feinstein, & Youman, 1996) and organizes user interests and habits according to roles.

Existing role mining approaches in the field of RBAC (Frank, Buhmann, & Basin, 2010) only consider two dimension parameters, namely <*user, permission*>, and neglect "context", which is important to mobile service recommendation. In this paper, similar to Wang et al. (2012) and Wong, Chu, Hao, and Wang (2012), we extend the behavior pattern recognition method (Cao, Bao, Yang, Chen, & Tian, 2010) to identify context-aware roles from multi-user behavior patterns. Different from Wang et al. (2012), Wong et al. (2012) and all other related work to this paper, our proposed framework and algorithms support incremental role mining.

Most RBAC research efforts are on static data and disregard the dynamic nature of user preferences and behavior patterns, i.e., changes over time. This problem is more critical to mobile service recommendation than traditional access control and security, since mobile users are usually on the move. They need to respond to the changing environments and hence contexts. As a result, the proposed role mining algorithm needs to be incremental (aka online) and support updates efficiently. To the best of our knowledge, this is the first proposal in the field of context-aware role mining. Moreover, the traditional solutions (Wang et al., 2012; Wong et al., 2012) around RBAC tend to be rigid. The user preferences and behaviors based on role mining are likely to be over-fitted without filtering out user-specific preferences or errors. It is similar to the over-fitting problem discussed in process mining (Van der Aalst et al., 2010).

We propose a framework based on the concept of quasi-biclique (Abello, Resende, & Sudarsky, 2002; Mishra, Ron, & Swaminathan, 2004; Sim, Li, Gopalkrishnan, & Liu, 2006; Sun & Nobel, 2008) to allow for an improved generalization of role mining; but yet, at the same time capturing the user-specific preferences by a residual term, to allow for this crucial information about each individual to be used in service recommendation. In addition, we develop a method to evaluate if there is any over-fitting and also establish an overall framework for parameter adjustment to make the online role mining adaptive to a changing environment. An error term is added on top of the residual term to allow for the capturing of random errors, which are unavoidable in a dynamic environment involving human behaviors. We apply a statistical test of significant to verify user-specific preferences allowing the detection of errors by assuming each online role mining cycle as a trinomial trial.

In order to illustrate the problem of over-fitting in role mining, we provide a synthetic dataset in Figure 1 to demonstrate the problem. By using traditional role mining method, e.g., (Wong et al., 2012) and the basic algorithm $(\mathcal{A}_{B1})$ in this paper, eight roles are mined from the dataset as shown in Figure 2. However, one would expect only half of the roles (four) by accepting minor variation of human preference. The expected role set is shown in Figure 3. This example shows that a minor variation of preferences among members could significantly increase the size of the role set resulting in merely another representation of the data without any inference capability.

The organization of this paper is as follows. Section *Related Work* summarizes the related work. Section *Problem Definition* defines the