

## A NOVEL ALGORITHM FOR MINING FUZZY HIGH UTILITY ITEMSETS

CHENG-PING LAI<sup>1,2</sup>, PAU-CHOO CHUNG<sup>1</sup> AND VINCENT S. TSENG<sup>3</sup>

<sup>1</sup>Institute of Computer and Communication Engineering, Department of Electrical Engineering

<sup>3</sup>Department of Computer Science and Informatics Engineering

National Cheng Kung University

No. 1, University Road, Tainan City 701, Taiwan

lcp@mail.hosp.ncku.edu.tw; pcchung@ee.ncku.edu.tw; tsengsm@mail.ncku.edu.tw

<sup>2</sup>Department of Medical Informatics

National Cheng Kung University Hospital

No. 138, Sheng Li Road, Tainan City 704, Taiwan

Received May 2009; revised December 2009

**ABSTRACT.** *Utility mining is to find the itemsets in a transaction database with high utility values like profits. Although a number of algorithms on high utility mining have been proposed, they did not reflect the fuzzy degree of quantity and profit level for mined high utility itemsets, which are essential for decision making in various applications like stock control and sales analysis. In this paper, we explore to apply fuzzy sets theory to the utility mining problem and propose a novel method, namely FHUI (Fuzzy High Utility Itemsets)-Mine, for mining fuzzy high utility itemsets. In addition to reflecting the fuzzy degree for quantity and profit regions of high utility itemsets, FHUI-Mine also provides a fuzzy threshold range that may include itemsets with profits slightly less than the designated threshold value. To prove the feasibility of FHUI-Mine, it was compared with the well-known Two-Phase algorithm through experimental evaluation. The results show that FHUI-Mine delivers higher mining capability since it can not only mine all high utility itemsets found by Two-Phase algorithm but also discover additional itemsets that are potentially high utility ones.*

**Keywords:** Utility mining, High utility itemset, Fuzzy sets theory, Fuzzy data mining

1. **Introduction.** Utility mining problem was defined by Yao *et al.* [22], where the utility relationships among itemsets were analyzed and the mathematical model of utility mining was defined. The goal of utility mining is to find all itemsets in a transaction database with utility values greater than the user specified threshold. Intuitively, utility is a measure of how profitable an itemset is.

A number of algorithms on high utility itemsets mining have been proposed [1,4,6,11,16]. Two-Phase algorithm was proposed by Liu *et al.* [11], which decomposes the mining process into two phases. The *transaction-weighted Downward Closure Property* is utilized in transaction-weighted utilization mining in Phase I and those overestimated itemsets are pruned off in Phase II. The *CTU-Mine* algorithm [4] that extends the pattern growth approach, mines the high utility itemsets from relatively dense or long pattern datasets. *THUI-Mine* algorithm [16] can discover temporal high utility itemsets from data stream efficiently. Many of the proposed algorithms [16,20,23] were based on the Two-Phase algorithm due to the advantages of less candidates, higher accuracy and lower arithmetic complexity [11].

However, there exist some major weaknesses in the existing methods for high utility itemsets mining: 1) In the past studies, once the minimum utility threshold is decided,