Egyptian Informatics Journal (2015) **16**, 243–251

Cairo University

Egyptian Informatics Journal

www.elsevier.com/locate/eij
www.sciencedirect.com

FULL-LENGTH ARTICLE

Contextual motivation in physical activity by means of association rule mining

Sugam Sharma [a,*], Udoyara Sunday Tim [b], Marinelle Payton [c], Hari Cohly [d], Shashi Gadia [e], Johnny Wong [e], Sudharshanam Karakala [f]

[a] *Center for Survey Statistics & Methodology, Iowa State University, Ames, IA 50010, USA*
[b] *Department of Agricultural and Biosystems Engineering, Iowa State University, Ames, IA 50010, USA*
[c] *Center of Excellence in Minority Health and Health Disparities, Jackson State University, Jackson, MS, USA*
[d] *Department of Biology, Jackson State University, Jackson, MS 39217, USA*
[e] *Department of Computer Science, Iowa State University, Ames, IA 50010, USA*
[f] *Department of Epidemiology and Biostatistics, Jackson State University, Jackson, MS 39217, USA*

Received 31 August 2014; revised 27 June 2015; accepted 30 June 2015
Available online 21 August 2015

KEYWORDS

Association rule mining;
Physical activity;
Regularity

Abstract The primary thrust of this work is to demonstrate the applicability of association rule mining in public health domain, focusing on physical activity and exercising. In this paper, the concept of association rule mining is shown assisting to promote the physical exercise as regular human activity. Specifically, similar to the prototypical example of association rule mining, market basket analysis, our proposed novel approach considers two events – *exercise* (sporadic) and *sleep* (regular) as the two items of the frequent set; and associating the former, *exercise* event, with latter, the daily occurring activity *sleep* at night, helps strengthening the frequency of the *exercise* patterns. The regularity can further be enhanced, if the exercising instruments are kept in the vicinity of the bed and are within easy reach.

1. Introduction

The regularity in physical activity administers prodigious benefits to human health but the participation in any form of physical activity or exercise tends to decline over time and an average 50% dropout is reported within six months of initiation [11]. Embarking upon and establishing a regular pattern of a physical activity is the hardest part [23]. Our work addresses such issues and exploits an eminently popular and well researched computational technique, association rule mining (hereinafter referred simply as ARM) [24,4] for contextual

* Corresponding author.
E-mail address: sugamsha@iastate.edu (S. Sharma).
Peer review under responsibility of Faculty of Computers and Information, Cairo University.

Production and hosting by Elsevier

http://dx.doi.org/10.1016/j.eij.2015.06.003

motivation to promote regularity in physical activity and exercising patterns.

The concept of ARM was first introduced by Agrawal et al. [1] to discover correlations among items within commercial transactions. Since then, ARM techniques have been discussed quite extensively in the data mining literature and issues related to the efficient generation of such rules from large complex dataset have been addressed. Primarily, the objective of the ARM is to discover the intrigue relationships among the items in complex, and large structured or unstructured multidimensional datasets. Generally, association rules (hereinafter referred as ARs)[1] are the data mining strategies that uncover the hidden relationship between entities in large datasets that assists in better learning about that data, specifically in customer buying patterns in numerous business domains. ARM techniques have been applied to various disparate areas of research and applications; and market risk management, inventory control, and telecommunication are just to name a few. In this paper, we do not exploit the entire spectrum of ARM; rather focus on a widely appreciated theoretical potential modeling technique of ARM, called market basket analysis [2]. The technique fairly predicts the correlation in buying patterns of distinct groups of items and the algorithmic aspect of its performing is straightforward. To better understand the concept of market basket analysis, let us briefly illustrate it further with a hypothetical example. In a supermarket, in the entire day processing, there are several transactions committed, collectively forming a large dataset. Each transaction consists of the name of the items purchased. If bread, milk, and cheese, for example, together are the common items in most of the transactions, then this set {bread, milk, cheese} is termed as frequent set. So, a frequent set F can be defined as the set of items (zero or more) bought together in at least in T transactions, a user-defined threshold. Then, it is most likely that these three items should be kept close inside the business venue, presumably, resulting in product sale increase. As a prediction analytics technique, ARM has found applications in many areas of decision making such as cross-marketing, catalog-design, store layout and buying patterns. It represents knowledge embedded in large datasets as probabilistic implications and intimately associated unit computation of frequent itemsets or discovering frequent sets in data. In addition to the many applications of AR in mining categorical data, other studies in medical diagnostics and healthcare analytics have also been reported. For example, the AR techniques have been proposed for image analysis [16,25,20]. This concept, particularly in the form of the market basket analysis [12] has attained significant success in gaining the consumer insights in data warehouse [13], but its effectiveness is still restricted to the analysis of commercial items only, housed in business facilities to improve the product marketing and to increase the revenues. In this paper, we attempt to further expand the bandwidth of the usage of the AR concept beyond the commercial domains and focus on its potential application to the public health concerns and implications of physical activity; and this distinguished effort contributes to the novelty of this paper.

To illustrate our approach, let us consider that *exercise* – sporadic event – and *sleep* – regular event – are the two items in the frequent set. Analogous to the prototypical example of ARM in business domain – market basket analysis – where the AR suggests keeping the associated items together to increase sales, the mild *exercise* event can be associated with *sleep* event. As already mentioned, the *sleep* event is a regularly occurring event, the association of the *exercise* event with it, will propel the latter for a drift behavioral inclination toward the regular event classification. In other words, being the items in the frequent itemset, if a *sleep* event occurs, there is a very high priority that *exercise* event occurs as well. The focus of this research is to enhance the regularity in exercise occurrences using the concept of ARM and the spatial aspect: the close proximity of the exercising equipment to the bed, and the temporal aspect: the time to execute the exercise with respect to the sleeping time, have important roles in it. However, the type of the exercise is an open choice. Once the exercise is over, the weariness caused, may further help the subject to fall asleep quickly. Our proposed approach does not intent to disrupt the sleeping patterns rather to encourage the regularity in performing the mild exercise, just before sleep that may help deepening the sleeping patterns.

The remainder of the paper is organized as follows. Section 2 represents a brief literature review on applications of ARM in numerous areas. Section 3 elaborates on the concept and algorithm associated with ARM. Section 4 details the methodological approach used in this study. Section 5 is the results and discussion section and the paper is concluded in section 6.

2. Literature review

A number of studies have attempted to extend the concept of ARM beyond the traditional business domain. Stilou et al. [21] exploit the importance of ARM in healthcare system for intelligent diagnosis and extraction of invaluable information from medical databases, especially diabetic data repository that eventually helps in developing the knowledge base automatically and quickly. Semenova et al. [22] develop an AR algorithm for large health databases, focusing on itemsets that offer knowledge and useful insights, unlike other algorithms that mainly focus on all frequent itemsets. Brossette et al. [4] utilize the concept of AR in developing a data analysis process, called Data Mining Surveillance System (DMSS), which integrates the hospital infection control and public health surveillance data to identify new, unexpected, and interesting patterns from the dataset. Their study explains also the importance of the surveillance systems in detecting the new and re-emerging threats of infectious agents in public health and hospital settings. Concaro et al. [7,8] develop general methodologies for mining of the temporal AR on sequences of hybrid events – the events that possess heterogeneous temporal elements such as *time interval* and *time point*. Algorithms are developed to extract the temporal AR in the sequences of hybrid events. Gamberger et al. [9] use the concept of ARs to generate the confirmation rules with high quality of predictions in medical diagnosis that may assists physicians for better cure of the patients. They apply the confirmation rules in coronary artery disease diagnosis and claim that the AR is provenly very useful in developing the reliable confirmation rules. Serban et al. [19] first redefine the ordinal ARs as relational AR, supported by mathematical formulation. The relational AR helps proposing

[1] ARs (association rules) are the plural representation of AR (association rule).