



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.

© 2015 Production and hosting by Elsevier B.V. on behalf of Faculty of Computers and Information, Cairo University. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

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

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 multi-dimensional 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).

a technique that may better assist in diagnosing the critical medical conditions; and finally, a programming interface was presented for such diagnosis. The interface offers a novel diagnosis technique, based upon the relational ARs and the supervised learning algorithm and helps in software development to report the probable approximation of the illness. The interface is flexible enough and can accommodate new types of symptom of any disease with very minuscule efforts. Gupta et al. [10] seek the benefits of ARM concepts in understanding the behavior and pattern of amino acids in protein structure. They want to first understand the nature of the associations, the amino acids enjoy among themselves. Mapping to the popular concept of market basket analysis of AR, the amino acids may behave like the set of items and the protein sequences can be considered analogues to the basket that contains items. The authors claim to have important insights into the co-occurrence of some amino acids in certain proteins. Malerba et al. [14] sense the great potential of data mining techniques in public policy making and apply such techniques to census data to discover some associations to support good public policy and to underpin the efficient functioning of a democratic society [18]. The authors advice to split the original task of spatial ARM into sub-tasks: (a) finding large or frequent spatial patterns, (b) generating spatial ARs with high confidence; and claiming to have discovered some ARs that deliver new knowledge. Chaves et al. [6] use the concept of AR in the diagnosis of Alzheimer's disease, specifically to explore relationships among activated brain waves in single photon emission computed tomography. Pazhanikumar and Arumugaperumal [17] provide detail surveys of various applications of ARM.

As these case studies indicate, ARM has become an important mining technique in many cases of health informatics. Distinguishably, this paper attempts to fan out the focus of the ARM in the recent public health concerns and implications of physical activity to regulate exercise patterns as a strategy toward healthy living. The strong association between physical activity and health benefits is widely reported in the literature already and low levels of physical activity are attributed as a major contributing factor to the higher health risk.

3. Algorithmic overview of ARM

The ARs are considered as one of the central concepts in data mining that builds upon the capabilities of machine learning and artificial intelligence. As a popular data mining technique, AR helps reveal the hidden relationships between apparently unrelated data in a data repository [24]. These rules help to analyze and predict the behavior of the customer, which is important in market basket data analysis, clustering of the products, new layout of the stores. ARs are primarily dependent on two criteria – *support* and *confidence*. *Support* indicates the frequency of the appearance of the items in the database whereas the *Confidence* is an indication of how many times the statements are true – if one buys item x , she/he is $y\%$ more likely to buy item z . Both the parameters are primarily used to measure the usefulness of the rules to the user and in the data mining literature, such ways of measurement are termed

objective measures. In general, the ARs with high-support and high-confidence posit that a user who, for example, purchases items a , b , and c also purchases item d most of the time.

The objective measures analyze the data in the dataset statistically and *support* and *confidence* are used as two important analyzing parameters [1].

Let $I = \{i_1, i_2, \dots, i_n\}$ be a set of items
 Let us further assume that $R = \{T_1, T_2, T_3 \dots T_n\} \mid \forall T_i \subseteq I$
 Let us consider X and Y two call items $X, Y \subseteq I$
 The rule implies that $X \Rightarrow Y \mid \forall X, Y \subseteq I$ and $X \cap Y = \Phi$

Support: AR $X \Rightarrow Y$ is sustainable with a *support* s if there are $s\%$ of commercial transactions in R that are consisting of $X \cup Y$. Those ARs are considered to have minimum *support* if they hold for that value of s where $s \geq$ user-specified support.

In more general way, the *support* of an itemset (x_1, x_2, \dots, x_j) , where each x_i is an element, can be calculated as follows.

$\text{Support}(x_1, x_2, \dots) = \frac{\text{Number of transaction containing } x_1, x_2, \dots}{\text{Total number of transactions}}$

Confidence: AR $X \Rightarrow Y$ is sustainable with a *confidence*, c , out of the all transaction in R there exists $c\%$ of them, if consists of X , also consists of Y . If an AR consists of a value of c , surpassing the user-specified confidence is known to have minimum confidence.

The *confidence* of the fact that item (x_1, x_2, \dots) implies (y_1, y_2, \dots) an itemset $(x_1, x_2, \dots, y_1, y_2, \dots, y_k, \dots, x_j)$, where each x_i and y_k are elements, is calculated as below.

$\text{Confidence}((x_1, x_2, \dots) \Rightarrow (y_1, y_2, \dots))$
 $= \frac{\text{Support}((x_1, x_2, \dots) \Rightarrow (y_1, y_2, \dots))}{\text{Support}(x_1, x_2, \dots)}$

The concept of *support* and *confidence* can be further illustrated using the following sample case study.

3.1. Sample case study

A fictional retail store sells hardware items such as tiles, cement, and paints. The store is interested in improving sales and wants to understand the consumers' shopping patterns and preferences. To undertake such analysis, it decides to adopt the concept behind the market basket analysis and analyzes all the transactions every day. For simplicity, here in this paper, only four random such transactions are depicted in Table 1 as an instance on a random day that can be further represented as a simplified array in Table 2.

Using the data in Table 1, various associations of the purchased items are analyzed in terms of *support* and *confidence*.

Table 1 Data for random five transactions on a random day.

Transaction ID	Purchases
1	Tiling Cement, Tiles
2321	Paint, White Spirit
31	Paint, Wallpaper, Plaster
204	Paint, Plaster, Tiling Cement, Tiles

Support (Paint, White Spirit) = $(1/4) = 25\%$
 Support (Paint, Plaster) = $(2/4) = 50\%$
 Support (Tiling Cement, Tiles) = $2/4 = 50\%$
 Confidence (Paint \Rightarrow White Spirit) = $1/3 = 33\%$
 Confidence (Paint, Plaster \Rightarrow Wallpaper) = $1/2 = 50\%$
 Confidence (Tiling Cement \Rightarrow Tiles) = $2/2 = 100\%$

Let us discuss one interesting case of association between the items – *Tiling Cement* and *Tiles*. As the above calculations indicate, there is a 50% support for the fact that a customer purchases *Tiling Cement* and *Tiles* together, whereas there is 100% confidence in the fact, if someone purchases *Tiling Cement*, he will surely buy *Tiles* too.

In this paper, we applied the concept of ARM to help explain and address the issue of physical activity and exercise patterns. More specifically, we associate mild exercise with regular sleeping pattern; there is good *support* and *confidence* that people will perform the exercise along with sleeping event. We propose that one should engage herself/himself in a mild exercise just before sleep. We further propose that the type of exercise is an open choice, but to avoid any disturbance in sleeping order, we recommend a light exercise. The next section describes the experimental evaluation of this study, utilizing the AR concept.

4. Methodology

Aside from the systematic approach of identifying the relationship between characteristic attributes, this study hypothesizes that the spatial and temporal aspects play some role in regulating physical activity and the exercise patterns. The dependencies of the activities are experienced in the form of AR. The experiment involves the informal participation. The protocol considers three locations, spatially varied from subject's bed, considered as the origin, and are as follows: (1) fitness facility, equipped with a large number of modern excising equipment, (2) living room, housing exercising equipment, and (3) bed

side, where light exercising equipment is in easy reach. The distance of the fitness center is assumed to be approximately half-mile from the residence of the participant. Any monetary burdens, such as the fitness center membership fee (monthly or annually) or costs of additional fitness equipment, that may possibly influence the study outcomes, are not considered. For each spatial aspect, the study is conducted for four weeks, and participation is documented on daily basis. Similar to the other studies that involve human participation, this study can involve a large cohort of adults of any race, ethnicity, or gender. However, this study is preliminary and seen as a proof of concept. So, we do not present extensive experiments to assess the quality of results and performance. We believe that an avatar can satisfy the demonstration requirements here, but to have the feel of real testing environment for the demonstration purpose, a real subject is requested for the participation instead of an avatar, which satisfactorily suffices the requirements. So, within the scope of this paper, the choice of the large dataset is optional. Therefore, in this paper, a cohort of size one (one participant) is used. The characteristics of the subject are – a healthy adult male, age 35 years. The participation is purely informal and completely voluntary and participant can leave anytime during the entire course of the study and there is no penalty for discontinuing the participation. There is no compensation for the participation. The risks of harm to a subject by participating in this experiment are minimal and no different from those ordinarily encountered in daily life. The identity of the participant is kept confidential and is not used anywhere throughout the entire study. When the study is completed and the data are analyzed, the identity is destroyed. The name of the participant is not used in any report.

The participant is provided with a questionnaire before the study commences and instructions are given about the procedure of the study. The daily response from the participant, whether that individual has executed the physical exercise, is collected. The questionnaire is very simple and contains places for entries for the entire month. The response requires putting a YES or NO, whether the exercise is performed for that particular day or not, in the designated location on a daily basis. The cohort is allowed to start the excise process at the fitness facility for the first four weeks. And, the data are collected for those days, when the participant visited the facility to perform the exercise. The type of the exercise and its duration are not the part of data collection, and hence ignored in this study. For the next four weeks, the cohort is asked to carry out the exercise at the home only and the exercising instruments are placed in the living room of the participant's home. Unlike the fitness facility, in the living room the exercising instruments are only dumbbells, though, the choice of the instruments is not constraint-specific. For this spatial location also, the data are collected in the similar manner as done for fitness facility.

Table 2 Synthetic data for confidence measure in ARM.

Transaction ID	Tiling Cement	Tiles	Paint	White Spirit	Wallpaper	Plaster
1	Y	Y	N	N	N	N
2321	N	N	Y	Y	N	N
31	N	N	Y	N	Y	Y
204	Y	Y	Y	N	N	Y

In the third phase of this study, the exercising instruments are moved to the vicinity of the bed, where the participant usually sleeps. In this case, the subject uses a hard surface bed of low-altitude and the participant has easy access to the exercising instruments from it. In this phase, the participant is asked to perform the exercise just before sleep. It is worth mentioning again, the count of only those particular days, when the exercise is performed, is the data of interest. The type of the exercise and the duration of the exercise execution are not important from the perspective of this study. Thus, the entire study is finished in 3×4 weeks = 12 weeks i.e. 3 months.

4.1. Alternative daily physical activities

- Yoga –This is another mode of physical activity that had been practiced for centuries, especially in ancient India [5]. However it has earned a wide spread attention and reputation around the globe for healthy living in a very recent time. Last year, the United Nations also recognized the importance of yoga in human life as daily physical activity and declared June 21 as the international yoga day [15]. On June 21st, 2015, around the world, the people were observed practicing yoga to celebrate that day. There are some asanas such as vajrasana [3] in yoga that are recommend for healthy life and can be practiced before sleep. A resident can perform yoga before going to sleep.
- Stationary bicycling – This is another daily activity that can be considered as a viable option in the context of this research work. Stationary bicycle can be placed in the living or bedroom and inhabitant can work out before going to sleep. However installation of this exercise instrument at the later location is likely to observe higher occurrence of physical activity.
- Treadmill running – Running on a treadmill can be considered as another alternative exercise or physical activity. Similarly, this exercise instrument can be placed in living or bedroom and resident can perform exercise daily.

4.2. Limitations

The core intent of the paper is to focus on demonstrating the utility and the applicability of the ARM in public health, rather making any conclusive statement regarding the sample dataset. For this proof of concept, we believe that even an avatar can satisfy the demonstration requirements of this work, but we prefer to choose a real subject to test the demonstration in real environment and that satisfactorily suffices the requirements of this work; so, the choice of large dataset is optional.

Although the large sample size is not a primary requirement of this work, we still reckon that statistically, a sizable cohort always ensures a finer rationalization of any relationship. Although the use of large dataset is optional in this study, this miniscule sample size, from some critics' point of view, can be considered as one of the limitations of this paper.

5. Results and discussion

In this section, the results, especially in terms of the occurrence of the physical activity are analyzed, discussed and compared. As shown in the Fig. 1(a)–(d), the study is carried out for four

weeks for each spatial location. For each individual week, the number of times the activity is performed is counted. Results are collected separately for individual spatial location, and considered for exercise execution. It can be observed, in Fig. 1(a) that in the beginning of the activity the frequency is higher i.e. the participant visits the fitness facility more often, but as the time passes by, the frequency degrades slowly. It can be noticed that in the last week of the month, the subject does not visit the fitness center at all and the obvious human behavior can be considered the cause behind it. For the spatial location, living room, the frequency of the activity is relatively better, but over the period of time, the occurrence of the activity decreases in this case also. The results in Fig. 1(b) and (c) apparently establish some kind of relationship between the occurrence of the exercise activity and the spatial aspect.

Apparently, the proximity of the exercising instruments motivates the participant to engage in exercise more and this can be cited by Fig. 1(c), where the exercise instruments are kept just near the bed side that helps to have a consistency in the occurrences of the event.

Finally, the results are combined in Fig. 1(d) that depicts clearly, how the number of the occurrences of the exercising event increases as the spatial location of exercising instruments advances to the sleeping venue.

In order to evaluate the usefulness of the rule, exploited for three different spatial locations – fitness center, living room at home, and bed side, we calculate the *support* and *confidence*. Although, unlike the general AR, there is no commercial transaction involved, there are no sale items. Rather *sleep*, and *exercise* events are considered two items of frequent set that will be used to calculate the *support* and *confidence*.

So, $X = \text{sleep (S)}$, and $Y = \text{exercise (E)}$.

In this paper, we attempt to calculate the *support* and *confidence* on weekly basis for all the spatial locations of the exercise equipment and are compiled in the tables below.

Tables 3.1–3.4 summarize the *support* and *confidence* calculated using the concept of association data mining rule, wherein the individual person needs to walk, drive or travel to a fitness facility located away from the place of residence. More specifically, Table 3.1 demonstrates how to calculate the daily basis *support* focusing on sleep (S) and exercise (E) as the primary physical activities. Tables 3.3 and 3.4 are similar to Tables 3.1 and 3.2, except the focus on data related to *confidence*. Table 3.4 is similar to Table 3.2 except that it contains the calculated *confidence*, obtained using the association mining rule.

Tables 4.1–4.4 contain the data for the *support* and *confidence*, calculated for AR, when the exercising instruments are located inside the living room of the home and a person does not need to walk or travel to fitness facility. The descriptions of Tables 4.1–4.4 are same to that of Tables 3.1–3.4.

Tables 5.1–5.4 show the *support* and *confidence*, calculated for the rule, when the exercising instruments are spatially located near the bed only and are in the easy access of the inhabitant.

Figs. 2 and 3 are the graphical representations of *support* and *confidence* of the ARs for all three spatial locations, respectively. In both the graphs, the *support* and *confidence*

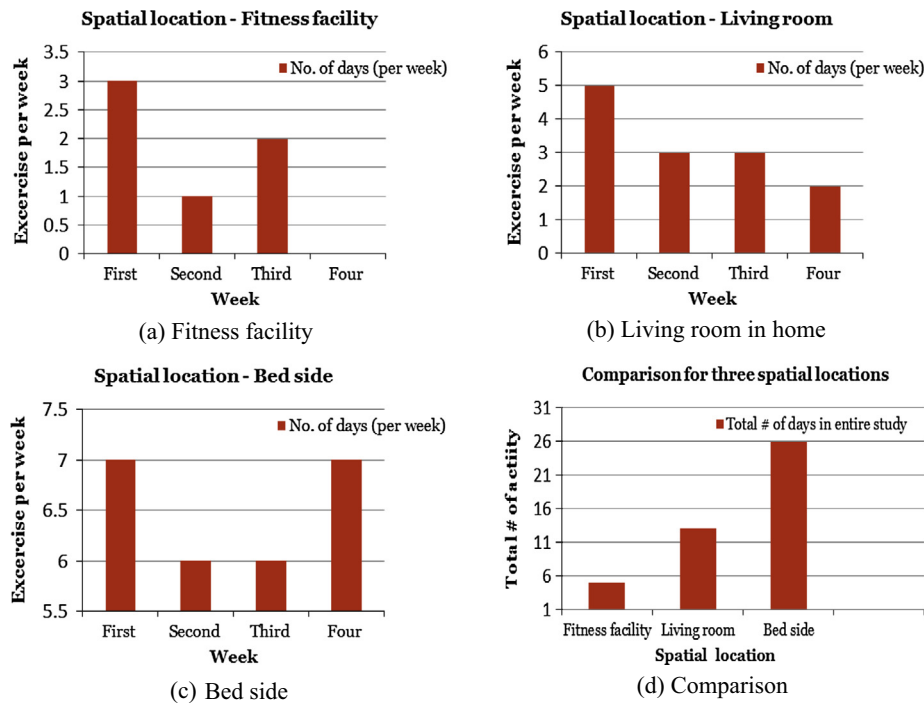


Figure 1 Spatial position vs. exercise occurrences.

Table 3.1 Calculating *support* for first week.

Transaction ID (Day)	Items	Support{SE}
1	SE	$4/7 = 57.14\%$
2	S	
3	SE	
4	S	
5	SE	
6	S	
7	SE	

Table 3.3 Calculating *confidence* for first week.

Transaction ID (Day)	Items	Confidence {S \Rightarrow E}
1	SE	$4/7 = 57.14\%$
2	S	
3	SE	
4	S	
5	SE	
6	S	
7	SE	

are highest for the bed side spatial location. The higher percentage of *support* and *confidence* cites for higher usefulness of the rule.

In contrary to the fact that the regularity of an exercising activity diminishes over a short period time, this research by associating the *exercise* with *sleep*, the regular event motivates for more participation in exercising activities and encourages the regularity in the exercise patterns. The association of an infrequent event *exercise* with a regular event *sleep* enhances the likelihood of happening *exercise* event, along with the *sleep*

Table 3.4 Calculating *confidence* for the entire month (4 weeks).

Week	Total transactions	Items		Confidence {S \Rightarrow E} (%)
		S	E	
1	7	7	4	57.14
2	7	7	2	28.57
3	7	7	2	28.57
4	7	7	0	0
Average confidence				28.57

Table 3.2 Calculating *support* for the entire month (4 weeks).

Week	Total transactions	Items		Support{SE} (%)
		S	E	
1	7	7	4	57.14
2	7	7	2	28.57
3	7	7	2	28.57
4	7	7	0	0
Average support				28.57

event. The occurrence of an *exercise* event is also influenced by the spatial aspect of the exercising site and as per the Figs. 2 and 3, bed side spatial location (higher *support* and *confidence* values) of the exercising instruments further helps in regulating the exercise patterns.

It can be noticed, in these Figs. 2 and 3, the *support* and *confidence* are exhibiting the same values, but generally, they possess disparate values and have their own significance in assessing the effectiveness of the ARs.

Table 4.1 Calculating *support* for first week.

Transaction ID (Day)	Items	Support{SE}
1	SE	$6/7 = 85.71\%$
2	S	
3	SE	
4	SE	
5	SE	
6	SE	
7	SE	

Table 4.2 Calculating *support* for the entire month (4 weeks).

Week	Total transactions	Items		Support{SE} (%)
		S	E	
1	7	7	6	85.71
2	7	7	4	57.14
3	7	7	4	57.14
4	7	7	3	42.85
Average support				60.71

Table 4.3 Calculating *confidence* for first week.

Transaction ID (Days)	Items	Confidence {S \Rightarrow E}
1	SE	$6/7 = 85.71\%$
2	S	
3	SE	
4	SE	
5	SE	
6	SE	
7	SE	

Table 4.4 Calculating *confidence* for the entire month (4 weeks).

Week	Total transactions	Items		Confidence {S \Rightarrow E} (%)
		S	E	
1	7	7	6	85.71
2	7	7	4	57.14
3	7	7	4	57.14
4	7	7	3	42.85
Average confidence				60.71

Table 5.1 Calculating *support* for first week.

Transaction ID (Days)	Items	Support{SE}
1	SE	$7/7 = 100\%$
2	SE	
3	SE	
4	SE	
5	SE	
6	SE	
7	SE	

Table 5.2 Calculating *support* for the entire month (4 weeks).

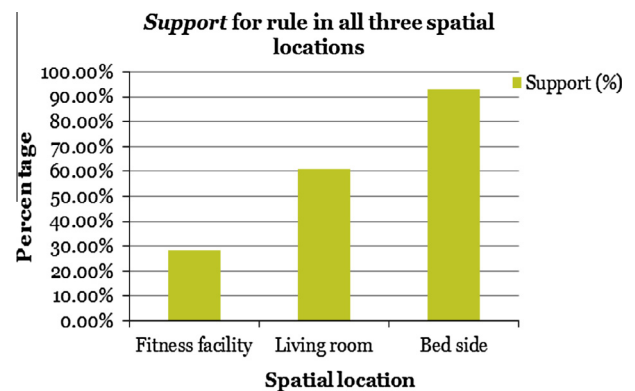
Week	Total transactions	Items		Support{SE} (%)
		S	E	
1	7	7	7	100
2	7	7	6	85.71
3	7	7	6	85.71
4	7	7	7	100
Average support				92.85

Table 5.3 Calculating *confidence* for first week.

Transaction ID (Days)	Items	Confidence {S \Rightarrow E}
1	SE	$7/7 = 100\%$
2	SE	
3	SE	
4	SE	
5	SE	
6	SE	
7	SE	

Table 5.4 Calculating *confidence* for the entire month (4 weeks).

Week	Total transactions	Items		Confidence {S \Rightarrow E} (%)
		S	E	
1	7	7	7	100
2	7	7	6	85.71
3	7	7	6	85.71
4	7	7	7	100
Average confidence				92.85

**Figure 2** Support in all three spatial locations.

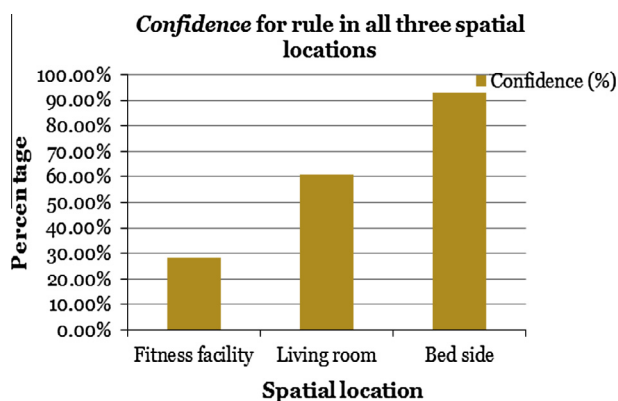


Figure 3 Confidence in all three spatial locations.

6. Conclusions

As per the general human behavior, the exercises start off with great enthusiasm, energy and regularity, but are often faded down over a short period time and become the sporadic events. Regulating even a mild exercise is not an easy task for everybody. In this research, we adopted the concept of ARM that helped in regulating the exercise patterns. The *exercise*, the sporadic event was placed in the frequency set with *sleep*, the regular event. This enhanced the likelihood of happening *exercise* event, along with the *sleep* event. In this paper, the problem was addressed from the spatiotemporal point of view. The spatial component represented the location of the exercise instruments, and the temporal component implies the time, when the exercise should be performed with respect to *sleep*, the regular pattern in human daily activity. In this study, we considered three spatial locations for exercise execution: (1) fitness facility, (2) living room, and (3) bed side. We proposed and validated that the proximity of the exercise instruments to the bed, the easy reach for the user, further contributed toward the exercising regularity. The *support* and *confidence* were used as the evaluating parameters to measure the usefulness of the created AR for different spatial locations. We advocated the mild exercise rather than the intensive ones, but encouraged the continuity or increase in the frequency of exercising. Thus, this study primarily encourages the regularity in the exercise patterns. Supported by the very fact that the regular exercise helps in obesity prevention and control, this work indirectly contributes to the ongoing efforts in curbing the obesity pandemic.

It is already stated that the core focus of the paper is to demonstrate the applicability of the ARM in health informatics instead of drawing conclusive statements about the sample dataset. This study is warranted on discovering the full potential of this important area of data mining research. The extension of ARM to health informatics presents several challenging issues, for the development of effective and efficient methods and the sample of the dataset. Additional work is also needed to improve the efficiency of applying the concept to large dataset to deduce more proven concrete conclusions. We do, however, consider our approach as the first step toward more exhaustive applications of the concept that has, for the most part, focused on business applications. But, we hope this proof of concept study will invigorate additional

investigation of applications in health care, especially given its potential in advancing decision making.

References

- [1] Agrawal R, Imielinski T, Swami A. Mining association rules between set of items in large databases. In: Proceedings of ACM SIGMOD; 1993. p. 207–16.
- [2] Berry MJ, Linoff GS. Data mining techniques: for marketing, sales, and customer relationship management. John Wiley & Sons; 2004.
- [3] Broad WJ. The science of yoga: the risks and the rewards. Simon and Schuster; 2012.
- [4] Brossette SE, Sprague AP, Hardin JM, Waites KB, Jones WT, Moser SA. Association rules and data mining in hospital infection control and public health surveillance. J Am Med Inform Assoc 1998;5(5):373–81.
- [5] Bryant EF. The yoga sutras of Patanjali: a new edition, translation, and commentary. North Point Press; 2015.
- [6] Chaves R, Ramírez J, Górriz JM, Puntónet CG. Association rule-based feature selection method for Alzheimer's disease diagnosis. Expert Syst. Appl. 2012;39(14):11766–74.
- [7] Concaro S, Sacchi L, Cerra C, Fratino P, Bellazzi P. Mining healthcare data with temporal association rules: improvements and assessment for a practical use. In: Proceedings of artificial intelligence in medicine, Verona, Italy, 18–22 July, 2009. p. 16–25
- [8] Concaro S, Sacchi L, Cerra C, Fratino P, Bellazzi P. Mining healthcare administrative data with temporal association rules on hybrid events. Method Inform Med 2011;50(2):166–79.
- [9] Gamberger D, Lavrac N, Jovanoski V. High confidence association rules for medical diagnosis. In: Proceedings of intelligent data analysis in medicine and pharmacology, DC, 6–10 November, 1999. p. 42–51.
- [10] Gupta N, Mangal N, Tiwari K, Mitra P. Mining quantitative association rules in protein sequences. In: Proceedings of Australasian conference on knowledge discovery and data mining, vol. 3755; 2006. p. 273–81.
- [11] Hausenblas HA, Carron AV, Mack DE. Application of the theories of reasoned action and planned behavior to exercise behavior: a meta-analysis. J Sport Exercise Psychol 1997;36–51.
- [12] Jukic N, Nestorov S. Comprehensive data warehouse exploration with qualified association-rule mining. Decis Support Syst 2006;42:859–78.
- [13] Kimball R. The data warehouse toolkit: the complete guide to dimensional modeling. John Wiley & Sons; 2006.
- [14] Malerba D, Esposito F, Lisi FA, Appice A. Mining spatial association rules in census data. Res Offic Stat 2002;5(1):19–44.
- [15] Mane AB. International yoga day: positive step toward global health. Int J Yoga 2015;8(2):163.
- [16] Ordonez C, Santana C, Braal L. Discovering interesting association rules in medical data. In: Proceedings of the ACM SIGMOD workshop on research issues in data mining and knowledge discovery, Dallas, Texas, USA, May 14, 2000. p. 78–85.
- [17] Pazhanikumar K, Arumugaperumal S. Association rule mining and medical application: a detailed survey. Int J Comput Appl 2013;80(17):10–9.
- [18] Saporta G. Data mining and official statistics. In: Proceedings of Quinta Conferenza Nazionale di Statistica, ISTAT, Roma, 15 November, 2000. p. 15–7.
- [19] Serban G, Czibula IG, Campan A. A programming interface for medical diagnosis prediction. Studia Universitatis, Babes-Bolyai, Informatica, LI(1); 2006. p. 21–30.
- [20] Sheela LJ, Shanthi V. An image analysis and classification protocol for characterization of normal and abnormal memory loss in aging from structural MRI. In: International conference on signal processing systems, Singapore, 15–17 May, 2009. p. 639–43.

- [21] Stilou S, Bamidis PD, Maglaveras N, Pappas C. Mining association rules from clinical databases: an intelligent diagnostic process in the healthcare. *Stud Health Technol Inform* 2001;84(Pt 2):1399–403.
- [22] Semenova T, Hegland M, Graco W, Williams G. Effectiveness of mining association rules for identifying trends in large health databases, vol. 1. In: Workshop on integrating data mining and knowledge management – the 2001 IEEE international conference on data mining, CA, USA, 29 November–2 December, 2001.
- [23] The Human Heart. Healthy Heart. In The Franklin Institute; 2014 [retrieved 18.04.14] <<http://www.fi.edu/learn/heart/healthy/exercise.html>> .
- [24] Zhang S, Wu X. Fundamentals of association rules in data mining and knowledge discovery. *Wiley Interdiscipl Rev Data Min Knowl Discov* 2011;1(2):97–116.
- [25] Zubi ZS, Saad RA. Using some data mining techniques for early diagnosis of Lung cancer. Recent researches in artificial intelligence, knowledge engineering and data bases. In: 10th WSEAS international conference on artificial intelligence, knowledge engineering and data bases, Cambridge, UK, 20–22 February, 2011. p 32–7.