

ON EXTRACTION OF NUTRITIONAL PATTERNS (NPS) USING FUZZY ASSOCIATION RULE MINING

M. Sulaiman Khan^{1,2}, Maybin Muyebeba¹, Frans Coenen²

¹ Liverpool Hope University, School of Computing, Liverpool, UK

² University of Liverpool, Department of Computer Science, UK

{khanm@hope.ac.uk, muyebebam@hope.ac.uk, frans@csc.liv.ac.uk}

Abstract. This paper proposes a framework for mining market basket data to generate Nutritional Patterns (NPs) and a method for analysing generated nutritional patterns using Fuzzy Association Rule Mining. Edible attributes are filtered from transactional input data by projections and are then converted to Required Daily Allowance (RDA) numeric values. The RDA database is then converted to a fuzzy database that contains expended normalized fuzzy attributes comprising of different fuzzy sets. Analysis of nutritional information is performed either from normal generated association rules or from a converted fuzzy transactional database. Our approach uses prototype support tool that extract Nutritional Patterns (NPs) and signifies the level of nutritional content in an association rule per item. The paper presents various performance tests and interestingness measures to demonstrate the effectiveness of the approach and concludes with experimental results and discussion on evaluating the proposed framework.

Keywords: Association rules, nutritional patterns, fuzzy association rules, nutrients.

1 INTRODUCTION

Association Rule Mining (ARM) [2] is a popular data mining technique that has been used to determine customer buying patterns from market basket data. General association rules are of the form $X \rightarrow Y$, which means customers who buy X also buy Y. The discovered rules indicate patterns of associating items. Such rules can be helpful in shelf arrangements, advertisement, sales promotion etc. However, nowadays health concerns are becoming increasingly

important to a large community of people including health practitioners, sporting organizations, governments and recently supermarkets. In data mining, association rules have been used to determine buying patterns (to the shop owner's benefit) but not nutritional pattern in general (to the customers health benefit).

People have recently become "healthy eating" conscious, but largely they are unaware of qualities, limitations and above all, constituents of food. For example, how often do people who buy baked beans bother with nutritional information other than looking at expiry dates, price and brand name? Unless the customer is diet conscious, there is no explicit way to determine nutritional requirements and consumption patterns. There are many dietary schemes and programmes that individuals follow that helps them determine how healthy they are but do not critically analyse nutritional elements that may affect their health. It is known that certain nutritional chemical elements when taken in large quantities do alter genetic material of a person, but also other elements are known to be more important for health than others.

Nowadays, nutritional information is usually labeled on supermarket products but is not used to determine actual nutritional patterns of every given customer transaction. This information would be useful for individual customers own health evaluation, supermarkets own reports on likely healthy buying patterns and many health related organizations including government health ministries. As modern society is concerned with health issues, association rules can be used to determine nutritional patterns by analysing product nutritional information, using market basket data. The approach signifies the level of nutritional content in an association rule per item. Most algorithms in the literature have concentrated on improving performance through efficient

implementations of the modified *Apriori* algorithm [3, 4, 13, 14]. Although improving performance and efficiency of various ARM algorithms is important, determining Nutritional Patterns (NPs) from customer transactions and association rules is also important. Extracting health related information using association rules from market basket data has mostly been overlooked.

In this paper we propose a fuzzy based approach for extracting nutritional patterns using fuzzy association rule mining, where a transactional database is converted into a database that contains the average RDA of nutrient values per item. This database is then converted into a fuzzy database with fuzzy attributes, according to the nutrients intake. The fuzzy database contains the actual contribution of nutrients per transaction or fuzzy membership degrees in fuzzy sets for each particular item (e.g. values 0.0, 0.3, 0.5, 0.2, 0.0 for fuzzy attributes very low, low, ideal, high and very high respectively) as shown in the figure 1.

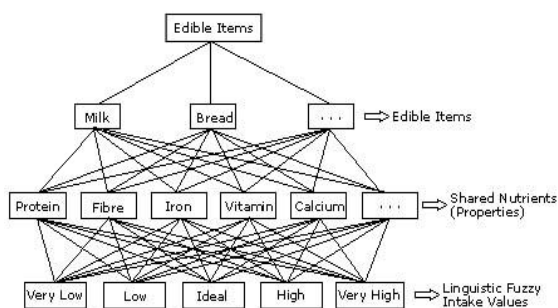


Figure 1: Edible items, Nutrients & Fuzzy Intervals

We show the effectiveness of this new method by applying it on different datasets. Our contributions are that edible attributes in market basket data are used with an RDA table, a fuzzy normalization process and correlation analysis produce effective rules and records good performance.

The paper is organised as follows: section 2 presents background and related work; section 3 gives a problem definition; section 4 discusses the proposed methodology; section 5 reviews experimental results, and section 6 concludes the paper with directions for future work.

2 BACKGROUND AND RELATED WORK

Many applications of association rule mining have been proposed in medical domain [9, 10, 15, 16] but

most of the researches in the literature have concentrated on improving performance through efficient implementation than producing effective rules [3, 4, 13, 14]. Again, in almost all ARM algorithms, thresholds (both confidence and support) are crisp values. This support specification may not suffice for queries and rule representations that require generating rules that have linguistic terms such as “low/Ideal/High for protein intake” etc. Fuzzy approaches [5, 6, 9, 11, 12] deal with quantitative attributes [7] by mapping numeric values to Boolean values. Detailed overviews for fuzzy association rules are given in [5, 6, 8].

Effective and efficient fuzzy algorithms supporting the mining process, i.e. the extraction of interesting associations from a database, have received less attention in the fuzzy community. This might be explained to some extent by the fact that, for fuzzy extensions of association analysis, standard algorithms can often be used or at least adapted in a relatively straightforward way. Still, some contributions have been made in this field. For instance, [17] describe the process of fuzzy association rules to obtain healthy buying patterns using binary Apriori algorithm.

Mining nutrient associations among itemsets is a new type of ARM technique which attempts to investigate Nutritional Patterns (NPs) by analysing nutrition consumption patterns. In [9], fuzzy associations are presented, where a reduced table is used to effectively minimise the complexity of mining such rules. The authors also present mining for nutrients in the antecedent part of the rule, but it is not clear how the fuzzy nutrient values are aggregated and largely, how membership functions are used. Our algorithm’s ultimate goal is to determine customers’ buying patterns for healthy foods, which can easily be evaluated using RDA standard tables. Other related work deals with building a classifier using fuzzy ARs in biomedical applications [18].

Fuzzy association rules have been used for medical data mining [9, 15, 16], but we propose a novel approach to determine whether customers are buying healthy food, which can easily be evaluated using required daily allowance (RDA) standard tables.

3 PROBLEM DEFINITION

In this section a sequence of formal definitions is presented to: (i) describe the concept of fuzzy association rule mining and (ii) the fuzzy approach

adopted by the authors. The normalization process for Fuzzy Transactions (T') and rules interestingness measures will also be discussed later in this section.

3.1 Fuzzy Association Rules

Mining fuzzy association rules is the discovery of association rules using fuzzy sets such that quantitative attributes can be handled [8]. A fuzzy quantitative rule represents each item as (item, value) pair. Fuzzy association rules are thus expressed in the following form:

$$\text{If } X \text{ is } A \text{ satisfies } Y \text{ is } B$$

For example *if (age is young) \rightarrow (salary is low)* where age and salary represents X and Y, and young and low are the discretised/linguistic values for attributes age and salary respectively representing A and B.

In the above rule, $X = \{x_1, x_2, \dots, x_n\}$ and $Y = \{y_1, y_2, \dots, y_n\}$ are itemsets, where $X \subset I, Y \subset I$, and $X \cap Y = \emptyset$. Sets $A = \{f_{x1}, f_{x2}, \dots, f_{xn}\}$ and $B = \{f_{y1}, f_{y2}, \dots, f_{yn}\}$ contain the fuzzy sets associated with the corresponding attributes in X and Y, for example (protein, low), (protein, ideal), (protein, high). The semantics of the rule is that when 'X is A' is satisfied, we can imply that 'Y is B' is also satisfied, which means there are sufficient records that contribute their votes to the attribute fuzzy set pairs and the sum of these votes is greater than the user specified threshold which could be crisp or fuzzy.

For a given database D with transactions $T = \{t_1, t_2, t_3, \dots, t_n\}$ with items $I = \{i_1, i_2, i_3, \dots, i_{|I|}\}$ and converted fuzzy transactions $T' = \{t'_1, t'_2, t'_3, \dots, t'_n\}$ with attributes $P = \{p_1, p_2, p_3, \dots, p_{|P|}\}$ and the fuzzy sets $F = \{fp_1, fp_2, \dots, fp_m\}$ associated with each attribute in P .

Table 1: Set of ordinary transactions

D	i_1	i_2	i_3
t_1	1	0	1
t_2	0	1	0
t_3	1	1	1
...

A fuzzy transaction is a special case of transformed ordinary transaction (table 1) and nonempty fuzzy subset of P where $T' \subseteq P$. In table 2 an item p_j and transaction t'_k contains a value v (membership degree) in $[0, 1]$. The membership degree of p_j in t'_k is $t_k(p_j(v_l))$. Without loss of generality, we also define edible set of items $E \subseteq I$ where any $i_j \in E$ consists of quantitative nutritional information $\bigcup_{p=1}^{|P|} i_j^P$, where each i_j^P is given as standard RDA numerical ranges and consists of $|P|$ nutrients.

Table 2: Set of edible fuzzy transactions

E	fp_1 (v_1)	fp_1 (v_2)	fp_1 (v_3)	fp_1 (v_4)	fp_1 (v_5)	fp_2 (v_1)	fp_2 (v_2)	fp_2 (v_3)	fp_2 (v_4)	fp_2 (v_5)
t'_1	0.0	0.7	0.3	0.0	0.0	0.0	0.0	0.8	0.2	0.0
t'_2	1.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0
t'_3	0.0	0.0	0.9	0.1	0.0	0.0	0.0	0.8	0.2	0.0
...

Each quantitative item p_j is divided into various fuzzy sets $f(p_j)$ and $m(l, v)$ denotes the membership degree of v in the fuzzy set l , $0 \leq m(l, v) \leq 1$ as shown in table 2.

3.1.1 Fuzzy Transactions Normalization Process

As mentioned above each quantitative item p_j in t'_k is divided into various fuzzy sets $f(p_j)$ and $m(l, v)$ denotes the membership degree of v in the fuzzy set l $0 \leq m(l, v) \leq 1$. For each fuzzy transaction $t' \in E$ (edible items), a normalization process to find significance of an items contribution to the degree of support of a transaction in order to guarantee a partition of unity is given by the equation (1):

$$m' = \frac{m(l, t'_k(p_j))}{\sum_{l=1}^{f(p_j)} m(l, t'_k(p_j))} \quad (1)$$

Without normalisation, support of an individual fuzzy item could increase in a transaction. The normalisation process ensures fuzzy membership values for each nutrient are consistent and are not affected by boundary values.

3.1.2 Fuzzy Support and Confidence

The problem of mining fuzzy association rules is given following a similar formulation in [15]. To generate Fuzzy Support (FS) value of an item set X with fuzzy set A , we use the equation (2):

$$FS(X, A) = \frac{\sum_{t_i \in T} \prod_{x_j \in X} m(t_i[x_j])}{|E|} \quad (2)$$

A quantitative rule represents each item as <item, value> pair. In the above equation we have used arithmetic mean averaging operator for fuzzy nutrients aggregation of candidate itemsets in a transactional database and used multiplication “mul” operator for fuzzy union of candidate items in a transaction. *min* or *max* operators can also be used but *mul* provides us the simplest and reasonable results as shown in table 3. In case when the fuzzy transactions are not normalized *mul* is more suitable because it takes the degrees of all items in a transaction into account.

Table 3: Effect of fuzzy mul operator

i_1	i_2	i_3	i_4		Max	Min	Mul
.2	.6	.7	.9	→	.9	.2	.075
.9	.8	.5	.6	→	.9	.5	.216
.7	0	.75	.8	→	.8	0	0
.3	.9	.7	.2	→	.9	.2	.037

For a rule $\langle X, A \rangle \rightarrow \langle Y, B \rangle$, the fuzzy confidence value (FC) where $X \cup Y = Z, A \cup B = C$ is given by equation (3):

$$FC(\langle X, A \rangle \rightarrow \langle Y, B \rangle) = \frac{\sum_{t_i \in T'} \prod_{z_j \in X} m(t_i[z_j])}{\sum_{t_i \in T'} \prod_{x_j \in X} m(t_i[x_j])} \quad (3)$$

where each $z \in \{X \cup Y\}$. For our approach, $X, Y \subset E$, where E is a projection of edible items from D . Depending on the query, each item i_j specified in the query and belonging to a particular transaction, is split or converted into $|P|$ nutrient parts $\bigcup_{p=1}^{|P|} i_j^p, 1 \leq j \leq |I|$. For each transaction t , the bought items contribute to an overall nutrient p by averaging the total values of

contributing items i.e. if items i_3, i_4 and i_7 are in a transaction t_1 and all contain nutrient $p=5$ in any proportions, their contribution to nutrient 5 is

$$\sum \frac{|i_j^5|}{3}, j \in \{3,4,7\}. \text{ These values are then}$$

aggregated into an RDA table with a schema of nutrients (see table 5, section 4) and corresponding transactions. We use the same notation for an item

i_j with nutrient p, i_j^p as item or nutrient p_j in the RDA table. Given that items p_j are quantitative (fuzzy) and we need to find fuzzy support and fuzzy confidence as defined, we introduce membership functions for each nutrient or item since for a normal diet intake, ideal intakes for each nutrient vary. However, five (5) fuzzy sets for each item are defined as {very low, low, ideal, high, very high} based on expert analysis on nutrition.

Based on this analysis, examples of fuzzy membership functions for nutrient Protein is shown in figure 1. There are many different types of membership function and the type of representation of the membership function depends on the nature of the fuzzy set. In figure 2 the functions assume a trapezoidal shape since nutrient values in excess or in deficiency mean less than ideal intake according to expert knowledge. Ideal nutrients can assume value 1 naturally, but this value could be evaluated computationally to 0.8, 0.9 in practical terms.

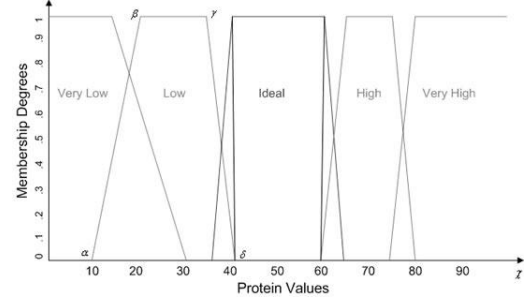


Figure 2: Fuzzy membership functions

$$\mu(x, \alpha, \beta, \gamma, \delta) = \begin{cases} 1, & x \in [\beta, \gamma] \\ \frac{(x - \alpha)}{(\beta - \alpha)}, & x \in [\alpha, \beta] \\ \frac{(\delta - x)}{(\delta - \gamma)}, & x \in [\gamma, \delta] \\ 0, & \text{otherwise} \end{cases} \quad (4)$$

Equation 4 [11] represents all nutrient membership degrees of a nutrient value “x”. The input database value x has “ideal” values of a nutrient between

β and γ , with lowest value α and highest value δ . The task is to determine a membership value of x in equation 4.

Note that equation 4 gives values equal to $m(l, v)$ in equations 1, 2 and 3. We can then handle any query after a series of data transformations and fuzzy function evaluations of associations between nutritional values. For missing nutrient values or so called “trace” elements, the fuzzy function evaluated zero degree membership.

3.2 Interestingness Measures

Measures of interestingness other than standard support and confidence are required in order to evaluate the quality of fuzzy association rules. The quality measure for a rule to be interesting is called certainty factor [15]. A rule can be considered interesting if the fuzzy set union of antecedent and the consequent has enough significance and the rule has adequate certainty. A measure of significance for a rule is similar to equation (3) and we have adopted it as the confidence of a rule. The certainty factor is determined by computing the fuzzy correlation of antecedent and the consequent of the rule. We have used Pearson’s product-moment correlation coefficient between attributes which is different from the general statistical usage of correlation because in association rule mining $X \Rightarrow Y \neq Y \Rightarrow X$.

The correlation $Corr(X, Y)$ between two variables X and Y with expected values $E(X)$ and $E(Y)$ and standard deviations σ_x and σ_y is defined as:

$$Corr(X, Y) = \frac{Cov(X, Y)}{\sigma_x \sigma_y}$$

$$Corr(X, Y) = \frac{E(X.Y) - E(X).E(Y)}{\sqrt{E(X^2) - E(X)^2} \sqrt{E(Y^2) - E(Y)^2}}$$

where E is the expected value of the variables and cov is covariance. We can transform the above correlation equation to find the certainty factor between two or more fuzzy attributes and can calculate fuzzy correlation as:

$$Corr_{fuzzy}(\langle X, A \rangle, \langle Y, B \rangle) = \frac{Cov(\langle X, A \rangle, \langle Y, B \rangle)}{\sqrt{Var(X, A)} \sqrt{Var(Y, B)}} \quad (5)$$

The value of correlation ranges from -1 to +1. Value -1 means no correlation and +1 means maximum

correlation. In our problem, only positive values can be considered as the degree of relation. As the certainty value increases from 0 to 1, the more related the attributes are and consequently the more interesting they are. Therefore if the rule “IF *Protein* is *low* THEN *Vitamin A* is *high*” holds, then the certainty value should be at least greater than zero. This could mean customers prefer to buy more vitamin related items to protein ones and the HBP value is simply the certainty value obtained (see section 6.1

4 PROPOSED METHODOLOGY

The proposed methodology consists of various phases, each of which is evaluated using fuzzy sets for quantitative attributes (Nutrients) as mentioned earlier. We have developed an algorithm called Fuzzy Healthy Association Rule Mining algorithm (FHARM). FHARM can deal with other kinds of transactional and relational databases to generate fuzzy association rules using quantitative attributes. We have discovered two techniques to obtain Nutritional Patterns as described in the next sections.

4.1 Nutritional Fuzzy ARM Mining

To mine from the transactional file (table 6), input data is projected into edible database on-the-fly thereby reducing the number of items in the transactions and possibly transactions too. The latter occurs because some transactions may contain non-edible items which are not needed for nutrition evaluation. This new input data is converted into an RDA transaction file (table 7) using RDA table (definition 6) with each edible item expressed as a quantitative attribute and then aggregating all such items per transaction (see definition 2, equation 1).

Table 6: Market Basket Data **Table 7:** Converted RDA transactions

TID	Items	TID	Pr	Fe	Ca	Cu
1	X, Z	1	45	150	86	28
2	Y	2	9	0	47	1.5
3	X, Y, Z	3	54	150	133	29.5
4	..	4

At this point, two solutions may exist for the next mining step. One is to discretised nutrients into intervals and converts RDA Transactions into discretised transactions with boolean values (Table 8) for each nutrient and corresponding value for

each interval per nutrient. Each transaction then (table 8), will have repeated fuzzy values {very low, low, ideal, high, very high} for each nutrient present in every item of that transaction. Table 8 actually shows only two nutrients.

Table 8: Discretised (Boolean) transaction file

TID	Protein (Pr)					Iron (Fe)					...	
	VL	L	Ideal	H	VH	VL	L	Ideal	H	VH		
1	0	1	0	0	0	0	0	1	0	0	0	...
2	1	0	0	0	0	1	0	0	0	0	0	...
3	0	0	1	0	0	0	0	1	0	0	0	...
4

Thus nutrients can have only values [0, 1] and only 1 intake value out of five in table 8, which represents its complete membership in that interval. The discretised boolean data can be mined by any binary type association rule algorithm to find frequent item sets and hence association rules. This approach only gives us, for instance, the total support of various fuzzy sets per nutrient and not the degree of support as expressed in equations 2 and 4.

The other approach (which we have adopted) is to convert RDA transactions (table 7) to linguistic values for each nutrient and corresponding degrees of membership for the fuzzy sets they represent above or equal to a fuzzy support threshold. Each transaction then (table 9), will have repeated fuzzy values {very low, low, ideal, high, very high} for each nutrient present in every item of that transaction. Table 9 actually shows only two nutrients.

Table 9: Linguistic (Fuzzy) transaction file

TID	Protein (Pr)					Iron (Fe)					...	
	VL	L	Ideal	H	VH	VL	L	Ideal	H	VH		
1	0.0	0.7	0.3	0.0	0.0	0.0	0.0	0.8	0.2	0.0	0.0	...
2	1.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	...
3	0.0	0.0	0.9	0.1	0.0	0.0	0.0	0.8	0.2	0.0	0.0	...
4

A data structure is then used to store these values (linguistic value and degree of membership) and large itemsets are found based on the fuzzy support threshold. To obtain the degree of fuzzy support, we use equations 2 and 4 on each fuzzy set for each nutrient and then obtain ARs (Nutritional Patterns) in the normal way e.g.

IF Protein intake is High AND Vitamin A intake is Low THEN Fat intake is High.

4.2 Rule Query on Nutrient Associations

To mine a specific rule, $X \rightarrow Y$, for nutritional content, the rule base (table 10) is scanned first for this rule and if found, converted into an RDA table (table 11) otherwise, the transactional database is mined for this specific rule. The latter involves projecting the database with attributes in the query, thus reducing the number of attributes in the transactions, and mining as described in 4.1.

In the former case, NPS are generated and the rule is stored in the new rule base with appropriate support, for example [proteins, ideal] \rightarrow [carbohydrates, low], 35%. A rule of the form "Diet Coke \rightarrow Horlicks, 24%" could be evaluated to many rules including for example, [Proteins, ideal] \rightarrow [Carbohydrates, low], 45%; where, according to rule representations shown in section 3, X is "Proteins", A is "ideal" and Y is "Carbohydrates", B is "low" etc. The same transformation to an RDA table occurs and the average value per nutrient is calculated before conversion to membership degrees or linguistic values. Using equations 2, 3, 4 and 5, we evaluate final rules expressed as linguistic values. The following example shows a typical query as described in 4.1 where TID is transaction ID, X, Y, Z are items and P (protein), Fe (Iron), Ca (calcium), Cu (Copper) are nutritional elements and support (Supp) and confidence (Conf) is given:

Table 10: Rule base

Table 11: RDA table and HBP rule

Rules	Support		Pr	Fe	Ca	Cu	..	
$X \rightarrow Y$	24%	\rightarrow	$X \rightarrow Y$	20	10	30	60	..
$Y \rightarrow Z$	47%	
$X, Y \rightarrow Z$	33%	

Nutritional Pattern

$X \rightarrow Y$ [Proteins, Very Low] \rightarrow [Carbohydrates, Low], Supp=45%, Conf=20%;

5 EXPERIMENTAL RESULTS

In order to show the defined frameworks effectiveness, we performed experiments using the prototype tool (figure 3) using T10I4D100K dataset containing simulated market basket data [generated by the IBM Almaden Quest research group][1]. The data contains 100K transactions and 1000 items. We considered 600 edible items out of the 1000 and used a real nutritional standard RDA table to derive fuzzy sets. Nutritional Patterns are then generated

from T10I4D100K dataset using methodology as described in section 4.

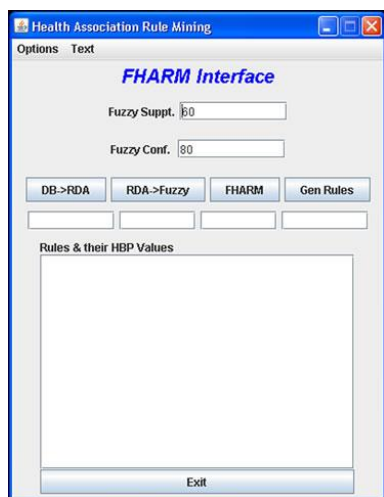


Figure 3: Prototype Tool

5.1 Experiment One

Experiment for method mentioned in 4.1 shows how our approach produces NPs in terms of interesting rules. We use all the 27 nutrients with T10I4D100K dataset. Figure 4 shows difference between number of frequent items produced by fuzzy method (with and without normalisation) and discrete method (discretised boolean data).

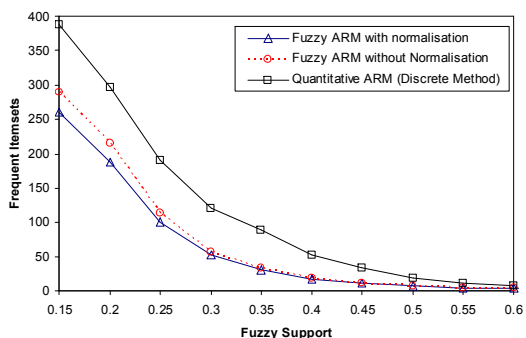


Figure 4: Number of Interesting Rules using fuzzy correlation

From the results, it is clear that the approach with boolean data produces more and irrelevant rules than the fuzzy approach. This is because using discrete method we cannot get the actual membership degree of nutrients intake for different intervals in each transaction thus considering full membership in any of the interval.

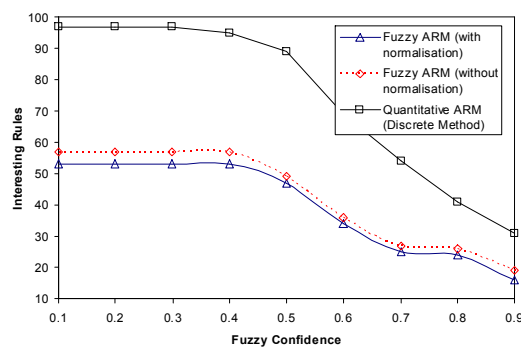


Figure 5: Number of Interesting Rules using fuzzy confidence

Figure 5 and Figure 6 shows the number of interesting rules using user specified fuzzy confidence and fuzzy correlation values respectively. Correlation has not been applied to quantitative ARM algorithm due to the boolean data and so only Fuzzy approaches (with and without normalisation) have been shown in figure 6.

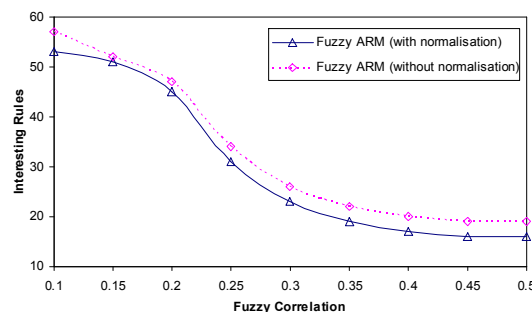


Figure 6: Number of Interesting Rules using fuzzy correlation

Figure 6 presents less and more interesting rules than figure 5 because it uses the correlation value for evaluation of interestingness between the antecedent and the consequent of a rule. The experiments show that normalization before applying correlation yields significantly less rules. In addition, the novelty of the approach is in being able to analyse nutritional content of itemsets or rules. Some interesting rules produced by our approach are as follows:

- IF *Protein* intake is *Ideal* THEN *Carbohydrate* intake is *low*.
- IF *Protein* intake is *Low* THEN *Vitamin A* intake is *High*.
- IF *Protein* intake is *High* AND *Vitamin A* intake is *Low* THEN *Fat* intake is *High*.

Depending on expert analysis, these rules are useful in analysing customer buying behaviour concerning their nutrition.

5.2 Experiment Two

We also implemented the algorithms for analysing rule queries and calculating fuzzy support and fuzzy confidence for methodology describes in section 4.2. For missing nutrient values or so called “trace” elements, the fuzzy function evaluated zero degree membership. We run classical ARM AprioriTFP algorithm on the data to produce a rule base. Some of the rule queries are as follows:

Rule 1: Milk → Honey, Support=29%

The rule is evaluated accordingly (see 4.2) as

44% - Very Low in [Calcium Cholesterol Fats Iodine Magnesium Manganese Phosphorus Sodium VitaminA VitaminC VitaminD VitaminK]
3% - Low in [VitaminB12]
14% - Ideal in [Fiber Protein VitaminB6 Zinc]
7% - High in [Niacin VitaminE]
29% - Very High in [Biotin Carbohydrate Copper Folacin Iron Riboflavin Selenium Thiamin]

Rule 2: Cheese, Eggs → Honey, Support=19%

37% - Very Low in [Calcium Fats Iodine Magnesium Phosphorus VitaminA VitaminB12 VitaminC VitaminD VitaminK]
3% - Low in [Carbohydrate]
22% - Ideal in [Manganese Protein Sodium VitaminB6 VitaminE Zinc]
3% - High in [Cholesterol]
33% - Very High in [Biotin Copper Fiber Folacin Iron Niacin Riboflavin Selenium Thiamin]

Rule 3: Jam → Milk, Support=31%

48% - Very Low in [Calcium Cholesterol Fats Iodine Iron Magnesium Phosphorus] etc.

It is surprising to see that for most rules (at least these shown here), calcium purchases from calcium rich products like milk and cheese are very low. Contrary, Biotin (Vitamin H, rules 1 and 2) deficiency that causes cholesterol, loss of appetite, hair loss etc is very high possibly because it is found in egg yolks and milk (dry skimmed). These inferences could be useful in real data applications.

6 CONCLUSION AND FUTURE WORK

In this paper, we presented a novel framework for extracting nutritional patterns (NPs) from customer transactions. Projections are made on input data into

edible attributes to find fuzzy association rules using nutrients actual membership degrees. Standard health information for each nutrient is provided as fuzzy RDA data. Fuzzy support and confidence as well as correlation are used as interestingness measures. A user can extract interesting HBP (Healthy Buying Patterns) rules from the transactions or from a given rule in the rule base.

In future, we intend to evaluate our approach on real and larger customer data. The determination of our presented approach for nutrient analysis is also important and viable future work, which should involve expert knowledge in evaluating the real value of nutritional patterns in health terms. Defuzzification of rules will bring new insights to the real healthy trends described in the paper. Similarly, non-coded food items (those without nutritional information) like fruits, vegetable in general etc need to be fuzzified in some other way and evaluated as shown.

Overall, the approach presented here is effective, efficient and could be very useful for both the customer and health organizations.

References

1. Agrawal R. and Srikant R., Quest Synthetic Data Generator. IBM Almaden Research Center, http://www.almaden.ibm.com/cs/projects/iis/hdb/Projects/data_mining/datasets/data/assoc.gen.tar.Z
2. Agrawal, R., Imielinski, T. and Swami, A. N.: Mining Association Rules Between Sets of Items in Large Databases. In Proceedings of the 1993 ACM SIGMOD Conference on Management of Data, (1993), 207-216.
3. Bodon, F.: A Fast Apriori Implementation. Proceedings of the IEEE ICDM Workshop on Frequent Itemset Mining Implementations, Vol. 90, (2003).
4. Lee, C.-H., Chen, M.-S., Lin, C.-R. : Progressive Partition Miner: An Efficient Algorithm for Mining General Temporal Association Rules. IEEE Transactions on Knowledge and Data Engineering, Vol. 15, No. 4, (2003), 1004 - 1017.
5. Chen, G. and Wei, Q.: Fuzzy Association Rules and the Extended Mining Algorithms, Information Sciences-Informatics and Computer Science: An International Journal archive, Vol. 147, No. (1-4), (2002), 201 - 228.
6. Au, W-H. and Chan, K. : Farm: A Data Mining System for Discovering Fuzzy Association Rules. In

- Proceedings of the 18th IEEE Conference on Fuzzy Systems, (1999), 1217-1222.
7. Srikant, R. and Agrawal, R.: Mining Quantitative Association Rules in Large Relational Tables. In Proceedings of ACM SIGMOD Conference on Management of Data. ACM Press, (1996), 1-12.
 8. Dubois, D., Hüllermeier, E. and Prade, H.: A Systematic Approach to the Assessment of Fuzzy Association Rules. To appear in Data Mining and Knowledge Discovery Journal, (2006).
 9. Xie, D. W.: Fuzzy Association Rules discovered on Effective Reduced Database Algorithm, In Proceedings of IEEE Conference on Fuzzy Systems, 2005.
 10. He, Y., Tang, Y., Zhang, Y-Q. and Synderraman, R.: Adaptive Fuzzy Association Rule Mining for Effective Decision Support in Biomedical Applications, International Journal Data Mining and Bioinformatics, Vol. 1, No. 1, (2006), 3-18.
 11. Gyenesei, A.: A Fuzzy Approach for Mining Quantitative Association Rules, Acta Cybernetical, Vol. 15, No. 2, (2001), 305-320.
 12. Klir, G. J. and Yuan, B.: Fuzzy Sets and Fuzzy Logic, Theory and Applications, Prentice-hall, (1995).
 13. F. Coenen, Leng, P. , Goulbourne, G. : Tree Structures for Mining Association Rules. Journal of Data Mining and Knowledge Discovery, Vol. 15, No. 7, (2004), 391-398.
 14. Wang, C. and Tjortjis, C.: PRICES: An Efficient Algorithm for Mining Association Rules. In Proceedings of the 5th Conference on Intelligent Data Engineering Automated Learning, Lecture Notes in Computer Science Series, Vol. 3177, Springer-Verlag, (2004), 352-358.
 15. LavraĀc N, Keravnou ET, Zupan B, Intelligent data analysis in medicine and pharmacology. In: First International Workshop on Intelligent Data Analysis in Medicine and Pharmacology (IDAMAP-96). Dordrecht: Kluwer Academic Publishers, 1996.
 16. Delgado M, SaĀnchez D, Vila MA. Acquisition of fuzzy association rules from medical data. In: Barro, MarĀn R, editors. Fuzzy logic in medicine. Wurzburg (Wien): Physical-Verlag, 2000, in press.
 17. Muyeba M, M. Sulaiman Khan, M. Malik, Z. and Tjortjis, C. "Towards Healthy Association Rule Mining (HARM), A Fuzzy Quantitative Approach", *In Proc. 7th Conf. Intelligent Data Engineering Automated Learning*, LNCS Series, Vol. 4224, Springer-Verlag, (2006), 1014-1022.
 18. He, Y., Tang, Y., Zhang, Y-Q. and Synderraman, R.: Adaptive Fuzzy Association Rule Mining for Effective Decision Support in Biomedical Applications, International Journal Data Mining and Bioinformatics, Vol. 1, No. 1, (2006), 3-18.

