International Journal of Engineering & Technology, 7 (4) (2018) 5851-5855

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET
doi: 10.14419/ijet.v7i4.14874
Research paper

Proposing a new methodology on vague association rule mining for the diagnosis of heart disease hesitation patterns

P. Umasankar [1] *, V. Thiagarasu [2]

[1] Research Scholar, Department of Computer Science, Manonmaniam Sundaranar University, Tirunelveli, Tamil Nadu
[2] Associate Professor, Department of Computer Science, Gobi Arts and Science College, Erode
**Corresponding author E-mail: pus.msu2013@gmail.com*

Abstract

In the realistic situation, the health care has which contain imprecisely specified data. This imprecise data indicates the presence of vagueness, incompleteness and uncertainty which causes the problem during important decision-making task in the prediction of heart disease. Traditional Association Rule Mining has limitations as it only deals with the features that are actually present in the prediction of heart disease and ignores the features that are almost not considered for the heart disease prediction. Furthermore, these features may be placed with predictive feature by imposing its attractiveness measure; disease prediction pattern mining in the scenario of imprecise and vague environment is very difficult which is frequent in recent years. For the effectiveness of retrieved hesitated patterns and rules, the concept of vague set theory is used. For the same consideration, heart disease dataset features as weighting factor is used for generation of disease prediction patterns.

Keywords: *Heart Disease, Vague Set Theory; Disease Prediction Pattern; Weighting Factor; Attractiveness Measure; Hesitant Pattern; Association Rule Mining.*

1. Introduction

The problem of decision making, especially in health care is a crucial task in every hospital [9-12]. In past years, data mining technology follows the conventional approach that offers only statistical analysis and discovers rules [1]. The main technique uses support and confidence measures for generating rules. But since the data have become complex today it is crucial to find the solution to deal with such problems [13-15].

Association rules can effectively use to uncover unknown relationship producing the result that can provide a basis for forecasting and decision making [2]. Traditional association rule mining model considers that items have the same significance without taking consideration of item weight within the transaction and also ignores vague status of item which is not always case. Considering an example [TV →vcd,1%,80%] may be more important than [bread →milk,3%,80%] although former holds a lower support because those items in the first rule usually come with more disease prediction per hospital, but conventional association rule is disregarded this variance.

2. Preliminaries

2.1. Vague set

Definition 1: *A vague set A [19] in the universe of discourse U is characterized by two membership functions given by:*

1) A truth membership function

$t_A : U \rightarrow [0,1]$

and

2) A False membership function

$f_A : U \rightarrow [0,1]$

where $t_A(u)$ is a lower bound of the grade of membership of u derived from the 'evidence for u', and $f_A(u)$ is a lower bound of the negation of u derived from the 'evidence against u and

$$t_A(u) + f_A(u) \leq 1$$

Thus the grade of membership of u in the vague set A is bounded by a sub interval $[t_A(u), 1 - f_A(u)]$ of [0, 1]. This indicates that if the actual grade of membership is μ(u), then

$$t_A(u) \leq t_A(u)$$

The vague set is written as

$$A = \{(u, [t_A(u), f_A(u)]) \mid u \in U\}$$

where the interval $[t_A(u), 1 - f_A(u)]$ is called the 'vague value' of u in A and is denoted by $V_A(u)$.

It is worth to mention here that interval-valued fuzzy sets (i-v fuzzy sets) [3] are not vague sets. In i-v fuzzy sets, an interval valued membership value is assigned to each element of the universe considering the 'evidence for u' only, without considering 'evidence against u' [6]. In vague sets both are independently proposed by the decision maker. This makes a major difference in the judgment about the grade of membership [3].

Definition 2: *A vague set A of a set U is called*

Copyright © 2018 P. Umasankar, V. Thiagarasu. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

1) the zero vague set of U if $t_A(u) = 0$ and $f_A(u) = 1$ for all $u \in U$
2) the unit vague set of U if $t_A(u) = 1$ and $f_A(u) = 0$ for all $u \in U$
3) the α -vague set of U if tA(u) = α and fA(u) = 1− α for all u 2 U, where $\alpha \in (0, 1)$.

Let D[0, 1] denotes the family of all closed sub-intervals of [0, 1]. Now we define refined minimum (briefly, rmin) and order" $\leq$ " on elements $D_1 = [a_1, b_1]$ and $D_2 = [a_2, b_2]$ of D [0, 1] as:

rmin $(D_1, D_2) = [\min\{a_1, a_2\}, \min\{a_1, a_2\}\ ]$

$D_1 \leq D_2 \leftrightarrow\ a_1\ \leq a_2 \wedge b_1\ \leq b_2$

Similarly, we can define $\geq$, = and rmax. Then concept of rmin and rmax could be extended to define rinf and rsup of infinite number of elements of D [0, 1].

It is that L = {D [0, 1], rinf, rsup, $\leq$} is a lattice with universal bounds [0, 0] and [1, 1].

2.2. Median and imprecision membership

In order to compare vague values we used two memberships: median membership and imprecision membership. It is noteworthy that for a given vague value [α(x),1− β(x)] we have unique median membership M_m and imprecision membership M_i and vice versa. Median membership is defined as

$$M_m = \frac{1}{2}(\alpha + (1 - \beta))$$

which represents the overall evidence contained in a vague value. It can be checked that $0 \leq M_m \leq 1$. Obviously, the vague value [1,1] has the highest M_m which means the corresponding object definitely belongs to the vague set (i.e., a crisp value). While the vague value [0, 0] has the lowest M_m, this means that the corresponding object definitely does not belong to the vague set.

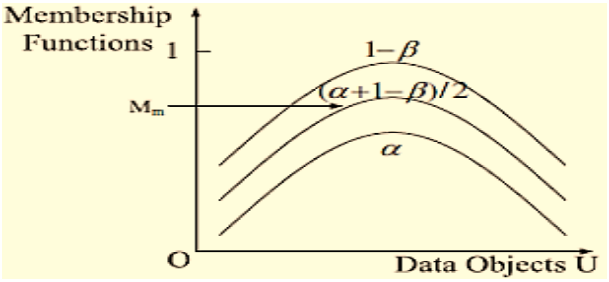


Fig. 1: Median Membership of Vague Set.

Imprecision membership is defined as

$$M_i = ((1 - \beta) - \alpha)$$

which satisfies the overall imprecision of a vague value. It can be Veteran that $0 \leq M_i \leq 1$. The vague value of [P, P] ($p \in [0,1]$) has the lowest M_i which means that the membership of the corresponding object is exact (i.e a fuzzy value). While the vague value [0,1] has the highest M_i this means that do not have an y information about the membership of the corresponding object.

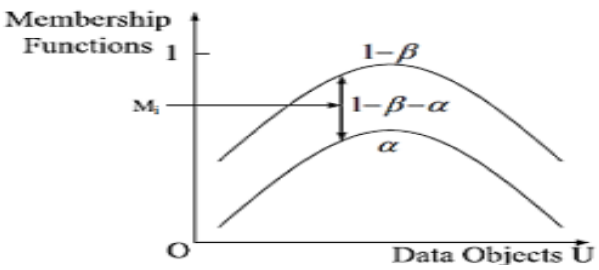


Fig. 2: Imprecision Membership of a Vague Set.

The median membership and imprecision membership are employed to measure the attractiveness and hesitation of the feature respect to a patient disease.

2.3. Hesitation and overall hesitation

Given an item $x \in I$ and a set of HSs $S = \{S_1, S_2, \ldots, S_n\}$ with a partial order $\leq$. The hesitation of x with respect to a hesitation status HSs $s_i \in S$ is a function $h_i(x)$: I → [0,1] such that

$$\alpha(x) + \beta(x) + \sum_{i=1}^{n} h_i(x) = 1$$

where $h_i(x)$ represents the evidence for the HS s_i of x. The overall hesitation of x with respect to S is given by

$$H(x) = \sum_{i=1}^{n} h_i(x)$$

This can be easily find from the above definition that [4]

$$H(x) = 1 - \alpha(x) - \beta(x)$$

2.4. Intent and overall intent

Given a set of HSs (S, ≤), the intent of an item x with respect to an HSs $s_i \in S$, denoted as int(x, s_i) is a vague value $[\alpha_i(x), 1 - \beta_i(x)]$ which is sub interval of $[\alpha(x), 1 - \beta(x)]$. The overall intent of x denoted as INT (x) is the interval $[\alpha(x), 1 - \beta(x)]$

2.5. Attractiveness and overall attractiveness

The attractiveness of x with respect to an HSs$_i$, denoted as att(x, s_i) is defined as the median membership of x with respect to s_i that is $\frac{1}{2}(\alpha_i(x) + (1 - \beta_i(x)))$. The overall attractiveness of x is a function ATT(x): I → [0,1] such that [5]

$$ATT(x) = \frac{1}{2}(\alpha(x) + (1 - \beta(x)))$$

2.6. AH-pair transaction database

An AH-pair database is sequence of AH-pair transactions. An AH-pair transaction T is a tuple $\langle v_1, v_2, \ldots, v_m \rangle$ on an itemset $I_T = \{X_1, X_2, \ldots, X_m\}$ where $I_T \in I$ and $< v_j = M_A(x_j), M_H(x_j) >$ is an AH-pair of the item x_j with respect to a given HS or the overall hesitation for $1 \leq j \leq m$

3. Vague association rule mining

A Vague Association Rule (VAR), r = (X => Y), is an association rule obtained from an AH-pair database. Based on the attractiveness and hesitation of an item, we define four different types of support and confidence of a VAR depending on what kind of knowledge we want to acquire. For clarity, we use A to denote Attractiveness and H to denote Hesitation.

3.1. Support

For a given AH-pair database, D, four types of support for an item set Z or a VARX $\Longrightarrow$Y where $X \cup Y = Z$ as follows:

1) The attractiveness support (A-Support) of Z is defined as $\frac{\sum_{T \in D} \prod_{z \in Z} M_A(Z)}{|D|}$
2) The hesitation support (H-Support) of Z is defined as $\frac{\sum_{T \in D} \prod_{z \in Z} M_H(Z)}{|D|}$
3) The attractiveness-hesitation (AH-support) of Z is defined as $\frac{\sum_{T \in D} \prod_{x \in X, y \in Y} M_A(x) M_H(y)}{|D|}$
4) The hesitation-attractiveness (HA-support) of Z is defined as $\frac{\sum_{T \in D} \prod_{x \in X, y \in Y} M_H(x) M_A(y)}{|D|}$

Z is an A (or H or AH or HA) frequent item set FI if the A-support (or H-support or AH-Support or HA-support) support or Z is no less than the (respective A or H or AH or HA) minimum support threshold σ.

3.2. Confidence

For a given AH-pair database four types of confidence for an item set Z or a VAR, $r = (X \implies Y)$ where $X \cup Y = Z$ as follows:

1) If both X and Y is A FIs, then the confidence of rule called the A-confidence of rule is defined as

$$\frac{Asupp\ (Z)}{Asupp\ (x)}$$

2) If both X and Y is H FIs, then the confidence of rule, called the H-confidence of rule is defined as

$$\frac{Hsupp\ (Z)}{Hsupp\ (x)}$$

3) If X is an A FI and Y is an H FI, then the confidence of rule, called the AH-Confidence is defined as

$$\frac{AHsupp\ (Z)}{AHsupp\ (x)}$$

4) If X is an H FI and Y is an A FI, then the confidence of rule, called the HA-confidence is defined as

$$\frac{HAsupp\ (Z)}{HAsupp\ (x)}$$

4. Proposed methodology for diagnosis of hesitation pattern of the heart disease

On critical analyzed it is examining that hesitation for the feature decrease the attractiveness of the feature and hence reduce the probability of analyzing the feature that results in decreasing the finding the severity of the disease. The characteristic of the feature can be a major factor that increases the hesitation or attractiveness toward the heart disease and hence it is an important factor in mining the hesitation information [16-18] [20]. The proposed an algorithm mine the vague association rules that is used for the heart disease patterns generation which is used to increase the finding severity of the disease. The following step by step procedure represents the mining of vague information pattern of heart disease diagnosis dataset.

Step 1: *Initialize array to store intent, AH-pair and frequent item set;*
Step 2: *Initialize favor α and against (β) variable with value zero;*
Step 3: *Increment favor (α) by one when value in database indicates that the feature is taken;*
Step 4: *Increment against (β) by one when value in database indicates that the feature is not considered;*
Step 5: *Generate intent $[\alpha, 1-\beta]$ using favor and against and store in intent array;*
Step 6: *Store the value as attractiveness and hesitation in AH-pair array.*
Step 7: *Mine all features whose attractiveness and hesitation is greater than minimum support and minimum weighted support.*
Step 8: *Generate the candidate set of size two as C_{A_2} from A_1, $C_{A_1H_1}$ from A_1, and H_1 and C_{H_2} from H_1.*
Step 9: *Generate all frequent itemset of size two whose support and weighted support is greater than minimum threshold;*
Step 10: *Similarly generate frequent itemset for all size and store in array of frequent itemset;*
Step 11: *Generate subsets of frequent items such that subset is contained in superset of frequent items;*
Step 12: *If subset is weighted vague frequent itemset, then find rule otherwise obliterate the subset the from the itemset list;*
Step 13: *Return weighted vague rule;*

5. Experimental result and discussion

The description of the Heart Disease dataset [7] has given below in the table 1.

Table 1: Description about the Heart Disease Dataset

No	Name	Description
1	Age	Age in Years
2	Sex	1=male, 0=female
3	cp	chest pain type(1 = typical angina, 2 =atypical angina, 3 = non-anginal pain, 4 = asymptomatic)
4	trestbps	Resting blood sugar(in mm Hg on admission to hospital)
5	chol	Serum cholesterol in mg/dl
6	fbs	Fasting blood sugar>120 mg/dl(1= true, 0=false)
7	restecg	Resting electrocardiographic results (0 = normal, 1 = having ST-T wave abnormality, 2 = left ventricular hypertrophy)
8	thalach	Maximum heart rate
9	exang	Exercise induced angina
10	oldpeak	ST depression induced by exercise relative to rest
11	slope	Slope of the peak exercise ST segment (1=upsloping, 2=flat, 3= downsloping)
12	Major Vessels Colored (ca)	Number of major vessels colored by fluoroscopy
13	thal	3= normal, 6=fixed defect, 7= reversible defect
14	num	Class (0=healthy, 1=have heart disease)

In the pre-processing stage, Genetic Algorithm is used to reduce the size of the heart disease dataset. Genetic algorithm [8] has selected only 4 features from heart disease dataset. Chest Pain, ExerInd, Major Vessels Colored and Thal.

5.1. Intent calculation of the heart disease features

Following table 2 gives the Intent calculation for mining the hesitation pattern of the heart disease feature.

Table 2: Intent Calculation of the Heart Disease Feature

Feature Name	Support	Confidence
Chest Pain	0.800000	0.866667
ExerInd	0.733333	0.833333
Major Vessels Colored (ca)	0.333333	0.666667
Thal	0.233333	0.566667

5.2. Intent calculation of the heart disease features

Following table 2 gives the Intent calculation for mining the hesitation pattern of the heart disease feature.

Table 2: Intent Calculation of the Heart Disease Feature

Feature Name	Support	Confidence
Chest Pain	0.800000	0.866667
ExerInd	0.733333	0.833333
Major Vessels Colored (ca)	0.333333	0.666667
Thal	0.233333	0.566667

5.3. Attractiveness Hesitation (AH) Pair of the selected features

Table 3: AH Pair of the Selected Features

Feature Name	Attractiveness	Hesitation
Chest Pain	0.833333	0.066667
ExerInd	0.716667	0.033333
Major Vessels Colored (ca)	0.500000	0.333333
Thal	0.400000	0.333333

Above table 3 gives the AH Pair of the selected features for the prediction of the heart disease.

5.4. Rule generation using association rule mining

Following table 4 depicts the rule generated by using Association Rule mining using Support and Confident level

Table 4: Pattern and Its Support for the 3 Types of Pattern Size

Pattern Size	Pattern Number	Pattern and its support
2	1	chestpain → Exerind (Support - 21)
	2	chestpain → ca (Support - 20)
	3	chestpain → thal (Support - 12)
	4	Exerind → ca (Support - 11)
	5	Exerind → thal (Support - 8)
	6	ca → thal (Support - 7)
3	1	chestpain, Exerind → ca (Support - 20)
	2	Exerind, ca → thal (Support - 11)
	3	chestpain, ca → thal (Support - 14)
	4	chestpain, Exerind → thal (Support - 11)
4	1	chestpain, Exerind,ca → thal (Support - 7)

5.5. E. Attractiveness (A), hesitation (h), attractiveness-hesitation (ah) and hesitation-attractiveness (ha) for 3 types of pattern

Table 5 and Table 6 represents the A, H, AH and HA support and confidence for the pattern size 2.

Table 5: Four Types of Support for Pattern Size 2

Pattern Number	Attractiveness (A)	Hesitation (H)	Attractiveness-Hesitation (AH)	Hesitation-Attractive (HA)
1	0.298611	0.001111	0.023889	0.013889
2	0.217593	0.002222	0.017407	0.027778
3	0.128472	0.005000	0.010278	0.062500
4	0.083333	0.004444	0.006667	0.055556
5	0.055556	0.003704	0.004444	0.046296
6	0.280694	0.001667	0.013056	0.035833

Table 6: Four Types of Confidence for Pattern Size 2

Pattern Number	Attractiveness (A)	Hesitation (H)	Attractiveness-Hesitation (AH)	Hesitation-Attractive (HA)
1	0.358333	0.016667	0.028667	0.208333
2	0.261111	0.033333	0.020889	0.416667
3	0.154167	0.075000	0.012333	0.937500
4	0.100000	0.066667	0.008000	0.833333
5	0.066667	0.055556	0.005333	0.694444
6	0.391667	0.050000	0.018217	1.075000

Table 7 and Table 8 depicts the A, H, AH and HA support and confidence for the pattern size 3. Table 9 gives the support and Confidence of A, H, AH and HA for the pattern size 4.

Table 7: Four Types of Support for Pattern Size 3

Pattern Number	Attractiveness (A)	Hesitation (H)	Attractiveness-Hesitation (AH)	Hesitation-Attractive (HA)
1	0.077971	0.000037	0.000769	0.003627
2	0.046036	0.000083	0.001792	0.002141
3	0.029861	0.000074	0.001593	0.001389
4	0.019907	0.000062	0.001327	0.000926

Table 8: Four Types of Confidence for Pattern Size 3

Pattern Number	Attractiveness (A)	Hesitation (H)	Attractiveness-Hesitation (AH)	Hesitation-Attractive (HA)
1	0.093565	0.000556	0.000956	0.054398
2	0.055243	0.001250	0.002150	0.032118
3	0.035833	0.001111	0.001911	0.020833
4	0.023889	0.000926	0.001593	0.013889

Table 9: Support and Confidence Value of A, H, AH and HA for 4th Pattern Size

Type	Pattern number	A	H	AH	HA
Support	1	0.012020	0.000003	0.000060	0.000559
Confidence	1	0.014425	0.000042	0.000072	0.008386

6. Conclusion

To identify the heart disease, the hospitals need to provide the most appropriate attributes to their patients for the diagnosis. In this paper, in the pre-processing stage, Genetic algorithm has utilized to reduce the size of the heart disease dataset. Using vague set theory, a limitation in the traditional Association Rule has addressed by ignoring the hesitation information of the features in the dataset using proposed methodology for the calculation of Attractiveness (A), Hesitation (H), Attractiveness-Hesitation (AH), and Hesitation-Attractiveness (HA) patterns. In the second stage, using vague set, a new methodology has proposed to derive the Attractiveness (A), Hesitation (H), Attractiveness-Hesitation (AH), and Hesitation-Attractiveness (HA) patterns for diagnosis the heart disease with pre-processed features. The support and confidence value of each pattern has calculated for each pattern size.

References

[1] Patel, Vimla L., et al. "The coming of age of artificial intelligence in medicine." *Artificial intelligence in medicine* 46.1 (2009): 5-17. https://doi.org/10.1016/j.artmed.2008.07.017.

[2] Wu, Xianguo, et al. "A hybrid information fusion approach to safety risk perception using sensor data under uncertainty." *Stochastic Environmental Research and Risk Assessment*32.1 (2018): 105-122. https://doi.org/10.1007/s00477-017-1389-9.

[3] Thiagarasu, P. Umasankar V. "Mining Correlation Rules for Interval-Vague Sets." *International Journal Of Engineering And Computer Science* 6.2 (2017).

[4] Thiagarasu, V., and P. Umasankar. "Mining Correlation Rules for Multiple Attribute Group Decision Making Models with Vague Sets." *International Journal of Applied Engineering Research*11.16 (2016): 8848-8857.

[5] V. Thiagarasu, P. Umasankar. "Mining MAGDM Problems with Triangular Vague Sets" *International Journal of Scientific and Research Publications*, 11.16 (2016): 297-310.

[6] Durairaj M, Poornappriya T S, "Survey on Vague Set Theory for Decision Making in Various Applications", International Journal of Emerging Technology and Advanced Engineering, Volume 8, Special Issue 2, pp. 104-107, 2018.

[7] http://archive.ics.uci.edu/ml/datasets/heart+disease

[8] Bhatt, Anurag, Sanjay Kumar Dubey, and Ashutosh Kumar Bhatt. "Analytical Study on Cardiovascular Health Issues Prediction Using Decision Model-Based Predictive Analytic Techniques." *Soft Computing: Theories and Applications*. Springer, Singapore, 2018. 289-299. https://doi.org/10.1007/978-981-10-5699-4_28.

[9] LakshmanaprabuS.K, Sachi Nandan Mohanty, K. Shankar, Arunkumar N, GustavoRamireze. "Optimal deep learning model for classification of lung cancer on CT images", Future Generation Computer Systems. 2018. https://doi.org/10.1016/j.future.2018.10.009.

[10] Lakshmanaprabu SK, K. Shankar, Deepak Gupta, Ashish Khanna, Joel J. P. C. Rodrigues, Plácido R. Pinheiro, Victor Hugo C. de Albuquerque. "Ranking Analysis for Online Customer Reviews of Products Using Opinion Mining with Clustering". Complexity, 2018: 1-9. https://doi.org/10.1155/2018/3569351.

[11] K. Shankar, Lakshmanaprabu S.K, Deepak Gupta, Andino Maseleno, Victor Hugo C. de Albuquerque. Optimal Features Based Multi Kernel SVM Approach for Thyroid Disease Classification. The Journal of Supercomputing, 2018. https://doi.org/10.1007/s11227-018-2469-4.

[12] Lakshmanaprabu SK, K. Shankar, Ashish Khanna, Deepak Gupta, Joel J. P. C. Rodrigues, Plácido R. Pinheiro, Victor Hugo C. de Albuquerque. Effective Features to Classify Big Data using Social Internet of Things. IEEE Access, 6 (2018): 24196-24204. https://doi.org/10.1109/ACCESS.2018.2830651.

[13] Andino Maseleno, Alicia Y.C. Tang, Moamin A. Mahmoud, Marini Othman, Suntiaji Yudo Negoro, Soukaina Boukri, K. Shankar, Satria Abadi, Muhamad Muslihudin. "The Application of Decision

Support System by Using Fuzzy Saw Method in Determining the Feasibility of Electrical Installations in Customer's House". International Journal of Pure and Applied Mathematics, 119. 16 (2018): 4277-4286.

[14] Muhammad Muslihudin, Risma Wanti, Hardono, Nurfaizal, K. Shankar, Ilayaraja M, Andino Maseleno, Fauzi, Dwi Rohmadi Mustofa, Muhammad Masrur, Siti Mukodimah, "Prediction of Layer Chicken Disease using Fuzzy Analytical Hierarcy Process", International Journal of Engineering & Technology, 7. 2.26 (2018): 90-94.

[15] Eka Sugiyarti, Kamarul Azmi Jasmi, Bushrah Basiron, Miftachul Huda, K. Shankar, Andino Maseleno, "Decision Support System of Scholarship Grantee Selection using Data Mining", International Journal of Pure and Applied Mathematics, 119.15 (2018): 2239-2249.

[16] Tri Susilowati, Kamarul Azmi Jasmi, Bushrah Basiron, Miftachul Huda, K. Shankar, Andino Maseleno, Anis Julia, Sucipto, "Determination of Scholarship Recipients using Simple Additive Weighting Method", International Journal of Pure and Applied Mathematics, 119. 15 (2018): 2231-2238.

[17] K. Shankar. "Prediction of Most Risk Factors in Hepatitis Disease using Apriori Algorithm", Research Journal of Pharmaceutical, Biological and Chemical Sciences, 8. 5 (2017): 477-484.

[18] Nur Aminudin, Eni Sundari, K. Shankar, P. Deepalakshmi, Fauzi, Rita Irviani, Andino Maseleno. "Weighted Product and Its Application to Measure Employee Performance", International Journal of Engineering & Technology, 7. 2.26 (2018): 102-108. https://doi.org/10.14419/ijet.v7i2.27.11574.