International Journal of Mathematics and Statistics Invention (IJMSI)
E-ISSN: 2321 – 4767 P-ISSN: 2321 - 4759
www.ijmsi.org Volume 4 Issue 1 || January. 2016 || PP-21-24

Mathematical Model of Affinity Predictive Model for Multi-Class Prediction

Hamdan O. Alanazi[1, 2], Abdul Hanan Abdullah [1], Moussa Larbani[3]
[1] *Department Computer Science, Faculty of Computing, Universiti Teknologi Malaysia, 81310 Johor,Malaysia*
[2] *Faculty of Applied Medical Science, Al-Majmaah University, Kingdom of Saudi Arabia*
[3] *Faculty of Economics, IIUM University, Jalan Gombak, Kuala Lumpur, 53100, Malaysia*

Abstract: *The notion of affinity which is one of the predictive models can bedefined as the distance or closeness between two objects.Unlike the fuzzy Set and Rough Set, the affinity can deal with third objects and deals with time dimension. In addition, it could deal with entities or abstract side by side with real objects. However, the existing model of affinity is developed for binary classification or prediction. In this paper, Affinity Predictive Modelhas been proposed in order to provide a multi-classprediction. This developed method can be used in many applications when multi-classpredictions are needed.*

Keywords: *Affinity Set, Predictive Model, Multi –Class Prediction*

I. INTRODUCTION

Nowadays, machine learning relates to computational models to improve the performance through automating the attainment of knowledge from past experience. Based on [1], the term predictive analytics might be used swapped with the term predictive model. In addition, the term pattern recognition, predictive model, machine learning and predictive analytics are used interchangeably [1-3]. It is important to predictive models to predict future outcomes [4]. Predictive models have many uses, including guiding healthcare policy; determining study eligibility of patients for new treatments; selecting appropriate tests and therapies in individual patient management including supporting decisions. Chen and Larbani initiated the theory of the affinity set and defined this set as the distance between two objects, where the distance measurement could be real or abstract [5]. This means the affinity set is a natural liking for or attraction to objects or abstracts. In order for this to happen, the affinity needs two elements namely the subjects between whom the affinity takes place and the affinity itself. Unlike the fuzzy Set and Rough Set, the affinity can deal with third objects and deals with time dimension. A notable feature is that it could deal with entities or abstract side by side with real objects.Huang et al. have used Affinity Set for measuring the performance of non-profit organization [6]. In the study of [7], they have provided a topology concept of Affinity Set as the data mining tool to classify and focus on the key attributes causing delayed diagnosis. Study of [8]provided a topology concept of Affinity Set as the data mining tool to classify and focus on the key attributes causing delayed diagnosis.Yuhet al.introduced the first classification model by using affinity Set [9]. The existing models of Affinity Predictive Models would not be able to provide a multi-class prediction. Thus, theAffinity Predictive Models should be improved to provide a multi-class prediction. The objective of this paper is to develop a mathematical model of affinity predictive model for multi-class prediction.

II. Mathematical Model Of AffinityPredictive Model For Multi- Class Prediction

Abstractly, Multi-class Affinity is a predictive model: given a problem instance to be classifiedusing affinity between entities, represented as following:

i. Definition of Affinity

Consider x and y are two entities. Then, the Affinity degree between entity (x) and entity (y) can be defined as follow:

AF_x^y *When* AF_x^y *= 1: Means the entity (x) has a very strong relationship with the entity (y).*
When AF_x^y *= 0: Means the entity (x) has not a relationship with the entity y.*
When $0 \leq AF_x^y \leq 1$*, Means that (x) has a relationship with the entity (y)*

ii. Definition of rule set

Rules in our rule set can be defined as follows:

$R_k : x_1\ is\ F_1^k\ \ldots\ldots x_n\ is\ F_n^k\ Then\ Class\ is\ O_1 with\ AF_1^k\ \ldots and\ O_c with\ AF_c^k$

Where R_k is the k-th rule$1 \leq k \leq L$ and$x_1, \ldots, x_n$ are the input variables *(Features)*, L is number of rules and $F_1^k, \ldots, F_n^k$ are the discrete values of the input variables, O_j is the class label, where j = {1,…,c} where c is number of classes. AF_j^k is the affinity degree of the rule R_k with the class O_j.

iii. Generating all possible rules

Generate possible rules where the number of rules $N_p = n_1 * \ldots * n_n$, where $n_1, \ldots, n_n$ represent the number of discrete values of the input variables $x_1, \ldots, x_n$, respectively.

iv. Affinity between rules and classes through training set

Finding the Affinity between rules and classes through training set can be illustrated as following:

1. Calculate the frequency FR_j^k which is the number of R_k occurrences in the training data set forclass O_j.
2. Finding the Affinity Degree $AFkc_j^k$between each rule R_k and the class pattern O_j using the following formula:

$$AFkc_j^k = \frac{Wk_j^k * FR_j^k}{\sum_{j=1}^{c} Wk_j^k \times FR_j^k}, j = 1, \ldots, c$$

This can be formulated as:

$$AFkc_j^k = \frac{Wk_j^k * FR_j^k}{FR_1^k \times W_1^k + \cdots + FR_c^k \times W_c^k}, j = 1, \ldots, c$$

Where the weights of rule R_k with the class pattern O_j, where Wk_j^k where$\mathrm{Wk}_j^k \in [0,1]$, $\sum_{j=1}^{c} Wk_j = 1$

v. Affinity between rules and classes through other rules within $core - r$

Finding the Affinity between rules and classes through other rules AFklc within $core - r$ can be illustrated as following:

1. Finding the Affinity between Rules

Affinity between two rules k and l can be calculated with this equation

$$AFkl\ \frac{\sum_{m=1}^{n} F_m}{n}$$

Where $F_m = 1\ when\ F_i^k = F_i^l$otherwise $F_m = 0$, Where F_i^k , F_i^l are the discrete values of the *i-th* input variable for the rules k and l, respectively where $i \in [1, \mathrm{n}]$

2. Finding the rule list within assigned core

Finding the core-set of the k-th rule CS_{R_k} among the total number of rules NR where the affinity between R_k and these rules are greater or equal than $core - r$where $core - r \in [0,1]$. These rules are selected from the all rulesTR_list as follows:

$CS_{R_k} = \{0\}$;
$index\_r1 = 1$;

For $index\_r$ =1: NR

$$If\ AFkl \geq core - r$$
$$CS_{R_k}\{index\_r1\} = TR\_list\{index\_r\} + CS_{R_k}$$
$$index\_r1 = index\_r1 + 1;$$

End
End

3. Finding the Affinity between rules and classes through other rules within $core - r$

4. Calculate the $AFklc$ between R_k and class O_j through the other CS_{R_k} as follows:

$AFklc = 0$;

For $index\_r$=1: N_list % N_list is the number of rules in CS_{R_k}

$$AFklc_k^j = \frac{Wk_j^{index\_r} * FR_j^{index\_r}}{\sum_{j=1}^{c} Wk_j^{index\_r} \times FR_j^{index\_r}} + AFklc_k^j$$

End

vi. <u>Affinity between rules and classes through Super rules within core– s</u>

Finding the Affinity between rules and classes through Super rules within core –s ($AFksc$) can

1. Finding super rules

To find the super rules these following steps need to be followed:

- Find the super rules SR_list through rules which occur at least one time in the training set.
- Calculate the frequency FS_j^k which is the number of SR_listoccurrences in the training data set forclass O_j.

vii. <u>Finding the Affinity between rules and super rules</u>

Finding the Affinity between rules k and super rules s can be calculated with this equation

$$AFks \frac{\sum_{m=1}^{n} F_m}{n}$$

Where $F_m = 1$ $when$ $F_t^k = F_i^s$otherwise $F_m = 0$, Where F_t^k , F_t^s are the discrete values of the *t-th* input variable for the rules k and s, respectively, where $1 \leq t \leq n$

viii. <u>Finding the rule list of super rules within assigned core</u>

Finding the Core-Super-Set of the k-th rule CSS_{R_s} among the super rules where the affinity between R_k and these rules are greater or equal than $core - s$ where $core - s \in [0,1]$. These rules are selected from the all super rules SR_list as follows:

$CS_{R_k} = \{0\}$;
$CSS_{R_s} = \{0\}$;
$index\_s1 = 1$;
For $index\_s = 1$: SR

$$If\ AFks \geq core - s$$
$$CSS_{R_s}\{index\_s1\} = SR\_list\{index\_s\} + CSS_{R_s}$$
$$index\_s1 = index\_s1 + 1;$$

End
End

ix. <u>Finding the Affinity between rules and classes through super rules within $core - s$</u>

Calculating the $AFksc_j^k$ between R_k and class O_j through the super rules CSS_{R_s}as follows:

$AFksc = 0$;

For $index\_s$ =1: S_list

$$AFksc_j^k = \frac{Ws_j^{index\_s} * FS_j^{index\_s}}{\sum_{j=1}^{c} Ws_j^{index\_s} \times FS_j^{index\_s}} + AFksc_j^k$$

End

Where the weights of rule R_k with the class pattern O_j, where S_list is the number of rules in CSS_{R_s} and Ws_j^s where$\mathrm{Wk}_j^s \in [0,1]$, $\sum_{j=1}^{c} Ws_j = 1$

x. Affinity between rules and classes through Frequentist probability

Finding the Affinity between rules and classes through Frequentist probability can be found with this equation:

$$AFkfc_k^j = \frac{N_j}{N}$$

Where N_j is the number of training patterns of class O_j and $N$is the total number of training patterns

xi. Final Affinity between rules and classes through all affinity relationships

Finding Final Affinity AFF between each rule R_k and class O_j through all affinity relationships can be found with this equation:

$$AFF_j^k = \frac{AFk_j^k + AFklc_j^k + AFksc_j^k + AFkfc_j^k}{N\_Aff}$$

Where AFk: Affinity between rules and classes through training set, AFklc: Affinity between rules and classes through other rules within$core - \mathrm{r}$, $AFksc$: Affinity between rules and classes through Super rules within $\mathrm{core} - \mathrm{s}$, AFkfc: Affinity between rules and classes through Frequentist probability, N_Aff : Number of Affinity relationships which equals to 4 and $1 \leq \mathrm{k} \leq L$ and $1 \leq \mathrm{j} \leq c$

III. CONCLUSION

Predictive models are using widely for providing aclassification and prediction. The affinity predictive model can be used for classification and prediction problems. A review of related literature reveals that existing models of Affinity Predictive provide a binary classification. In multi-class classification, the input is to be classified into one, and only one, of non-overlapping outcomes. Indeed, the multi- class prediction is needed in many applications. Therefore, a mathematical model of Affinity have been presented in this paper.

REFERENCES

[1] Burke, Jason (2013). Health analytics: Gaining the insights to transform health care. Vol. 71. John Wiley & Sons,

[2] Omer A., Dominique L.(2015). Predictive Marketing: Easy Ways Every Marketer Can Use Customer Analytics and Big Data, book. DOI: 10.1002/9781119175803

[3] Hildebrandt M. and Katja D.(2013). Privacy, Due Process and the Computational Turn: The Philosophy of Law Meets the Philosophy of Technology. Routledge.

[4] Vogenberg, F. Randy(2009). "Predictive and prognostic models: implications for healthcare decision-making in a modern recession." American health & drug benefits 2.6: 218.

[5] Chen Y, Larbani M (2006) Developing the affinity set and its applications. In: Proceeding of the Distinguished Scholar Workshop by National Science Council, July, pp. 14-18.

[6] Huang, Wen-Tsung, and Yuh-Wen Chen(2012). "Qualitative data envelopment analysis by affinity set: a survey of subjective opinions for NPOs." Quality & Quantity: 1-15.

[7] Chen, Y.W., Larbani, M., Shen, C.M. and Chen, C.W.(2009) "Using Affinity Set on Finding the Key Attributes of Delayed Diagnosis." Applied Mathematical Sciences 3.7: 297-316.

[8] Larbani, Moussa, and Yuh-WehChen (2009). "A Fuzzy Set Based Framework for Concept of Affinity." Applied Mathematical Sciences 3.7: 317-332.

[9] Chen YW, Larbani M, Hsieh CY, Chen CW (2009) Introduction of affinity set and its application in data-mining example of delayed diagnosis. Expert SystAppl 36:10883-10889.