Research Journal of Applied Sciences, Engineering and Technology 11(12): 1397-1403, 2015
DOI: 10.19026/rjaset.11.2246
ISSN: 2040-7459; e-ISSN: 2040-7467

Submitted: June 24, 2015 Accepted: August 2, 2015 Published: December 25, 2015

Research Article

Fuzzified MCDM Consistent Ranking Feature Selection with Hybrid Algorithm for Credit Risk Assessment

[1]Y. Beulah Jeba Jaya and [2]J. Jebamalar Tamilselvi
[1]Bharathiar University, Coimbatore,
[2]Department of MCA, Jaya Engineering College, Chennai, Tamil Nadu, India

Abstract: Feature selection algorithms that are based on different single evaluation criterions for determining the subset of features shows varying result sets which lead to inconsistency in ranks. In contrary, Multiple Criteria Decision Making (MCDM) with Fuzzified Feature Selection methodology brings consistency in feature selection ranking with optimal features and improving the classification performance of credit risks. By adopting multiple evaluation criteria inconsistent ranks to Fuzzy Analytic Hierarchy Process (FAHP) for feature selection along with hybrid algorithm (K-Means clustering-Logistic Regression classification) results in enabling Consistent Ranking Feature Selection (CRFS) and significant improvement over classification performance measures. When the proposed methodology is used with two different credit risk data set from the UCI repository, the experimental results show that the optimal features with hybrid algorithm, indicating improvements in the performance of classification in credit risk prediction over the current existing techniques.

Keywords: Credit risk, feature selection, fuzzy analytic hierarchy process, k-means clustering, logistic regression classification, multiple criteria decision making

INTRODUCTION

Existence of irrelevant features in a dataset, a data mining model performance tends to decrease (Jiliang *et al*., 2014). Feature selection algorithms aims to choose a small subset of the relevant features of the original dataset based on certain evaluation criterion, which leads to improved classification accuracy and reduce the computational complexity. To evaluate the quality of feature subsets, various evaluation measures are employed in feature selection algorithms. One of the feature selection algorithm ReliefF, distance evaluation measure is used to determine the features by means of distances between the instances (Yilmaz *et al*., 2012). Information evaluation measure is employed in many feature selection algorithms to find the information gain for the features (Koller and Sahami, 1996). The use of dependency measure in feature selection helps to determine the smaller subset sizes (Modrzejewski, 1993). But most of the feature selection algorithms in the literatures used single evaluation measures for selecting the subset of features which leads to inconsistency in ranking. To improve consistency in ranking, Multiple Criteria Decision Making Method (MCDM) provides a background to choose the best features on multiple criteria. Hence, a new Consistent Ranking Feature Selection (CRFS) is proposed which is based on multiple evaluation criteria (distance, dependency and information). Individual feature ranking is generated for each evaluation criteria. With the inconsistent ranked features on different single evaluation criteria, MCDM method (Fuzzy Analytic Hierarchy Process (FAHP)) (Van Laarhoven and Pedrcyz, 1983) are used to rank the features for consistency on multiple criteria. FAHP is an effective tool to handle the fuzziness of the data involved in deciding the alternatives from multiple criteria (Chan and Kumar, 2005).

The resulting optimal feature subset obtained is applied to hybrid algorithm for credit risk assessment. In this step also, an effective approach is used to combine K-Means clustering- Logistic Regression classification for credit risk assessment (Beulah Jeba Jaya and Tamilselvi, 2015).

The main objectives of this study are summarized as follows:

- Propose a Consistent Ranking Feature Selection (CRFS) for optimal features
- Use the effective hybrid algorithm with optimal features as input to increase the classification accuracy
- Compare the experimental results with the two UCI repository credit risk datasets and validate the proposed method with the current existing techniques.

Corresponding Author: Y. Beulah Jeba Jaya, Bharathiar University, Coimbatore, Tamil Nadu, India

LITERATURE REVIEW

Feature selection algorithms helps to better understand the insights of classification problem (Kohavi and Sommerfield, 1995). Previous researches in feature selection focus on independent criteria for selecting the feature subset. In filter or wrapper models of feature selection only independent criterion or single criteria is used (Liu and Yu, 2005). From the literatures, it is noticed that a feature subset determined using different evaluation criteria was different from the other (Dash and Liu, 1997).

When multiple criteria are involved and compete with each other then Multiple Criteria Decision Making (MCDM) techniques can be implemented (Adomavicius and Kwon, 2007). Analytic Hierarchy Process is a powerful method for multiple criteria decision making problems and the method deals with a very uneven scale of decision (Sun, 2010). Implementing fuzzy values in uneven scale of decision produces more accurate result (Guner *et al*., 2005). Several Researchers employed Fuzzy theory with AHP to deal with the uncertainty (Buckley, 1985).

Important evaluation criteria for feature selection such as information, dependency and distance were identified from the literature (Ben-Bassat, 1982) and used in this study. Ranking the features shows the importance of the individual feature (Ramaswami and Bhaskaran, 2009). Feature selection algorithms such as ReliefF (Robnik-Sikonja and Kononenko, 2003) which shows good performance for classification problems, symmetrical uncertainty (Senthamarai Kannan and Ramaraj, 2010) has proved to be the successful feature selector to remove redundant features, information gain (Novakovic, 2009) attribute evaluation were used to select the most significant features.

Performance of classifiers was evaluated using important performance measures such as Overall accuracy, TP rate (Correctly classified good credits), TN rate (Correctly classified bad credits), F-Measure and Area under ROC for credit risk assessment and financial risk prediction (Yi *et al*., 2011; Beulah Jeba Jaya and Tamilselvi, 2014). To classify, Logistic Regression classifier was identified as a top classifier for financial risk prediction (Yi *et al*., 2011).

From the literatures, it was also observed that by combining clustering and classification methods gives high classification accuracy (Zeng *et al*., 2003) and gives improved performance when compared to single classification approach (Khanbabaei and Alborzi, 2013).

However, as such there is no research work based on multiple evaluation criteria (distance, dependency and information) ranks combined FAHP process (Van Laarhoven and Pedrcyz, 1983) along with hybrid algorithm for improved consistent ranking feature selection and classification in credit risk assessment.

PROPOSED WORK

Considering the existing feature selection algorithm drawbacks in producing consistent feature subsets for different evaluation criteria, a new methodology is proposed to improve the consistency in ranking and classification which is based on multiple evaluation criteria ranks combined FAHP process (Van Laarhoven and Pedrcyz, 1983) for determining the optimal features. Then the resulting optimal features are applied to hybrid algorithm which combines K-Means Clustering and Logistic Regression classification (Beulah Jeba Jaya and Tamilselvi, 2015) for assessing the credit risks.

The system architecture diagram for the proposed methodology is illustrated in Fig. 1 and the steps of the proposed methodology are as follows:

- Rank the features for each dataset based on distance, dependency and information criteria.
- Obtain pairwise comparison matrix based on the ranks for each criteria.
- Apply triangular fuzzy number to pairwise comparison matrix for determining fuzzy preference weights on each criteria using FAHP process.
- Compute overall fuzzy preference weights and greater overall fuzzy preference weights are ranked as high.
- Ranked Non-zero overall fuzzy preference weights are taken as optimal features.
- Using the optimal features as input, hybrid algorithm (K-Means-Logistic Regression) is applied to determine the improved performance measures for credit risk assessment.

Consistent ranking feature selection: Consistent Ranking Feature Selection (CRFS) is based on multiple evaluation criteria (distance, dependency and information) ranks with combined FAHP process (Van Laarhoven and Pedrcyz, 1983). The evaluation criterion 'distance' measures the distance between the instances that is closer to each other. The evaluation criterion 'dependency' measures the strong dependence between the two features. The evaluation criterion 'information' measures the amount of information in common between two features. Generalised ranking algorithm for feature selection based on single evaluation criteria is described below:

- Initialise with the starting instance of the dataset
- Evaluate the objective function by maximizing the evaluation measure
- Generate the random subset for evaluation from dataset
- Evaluate the current subset by using evaluation measure such as information, dependency and distance

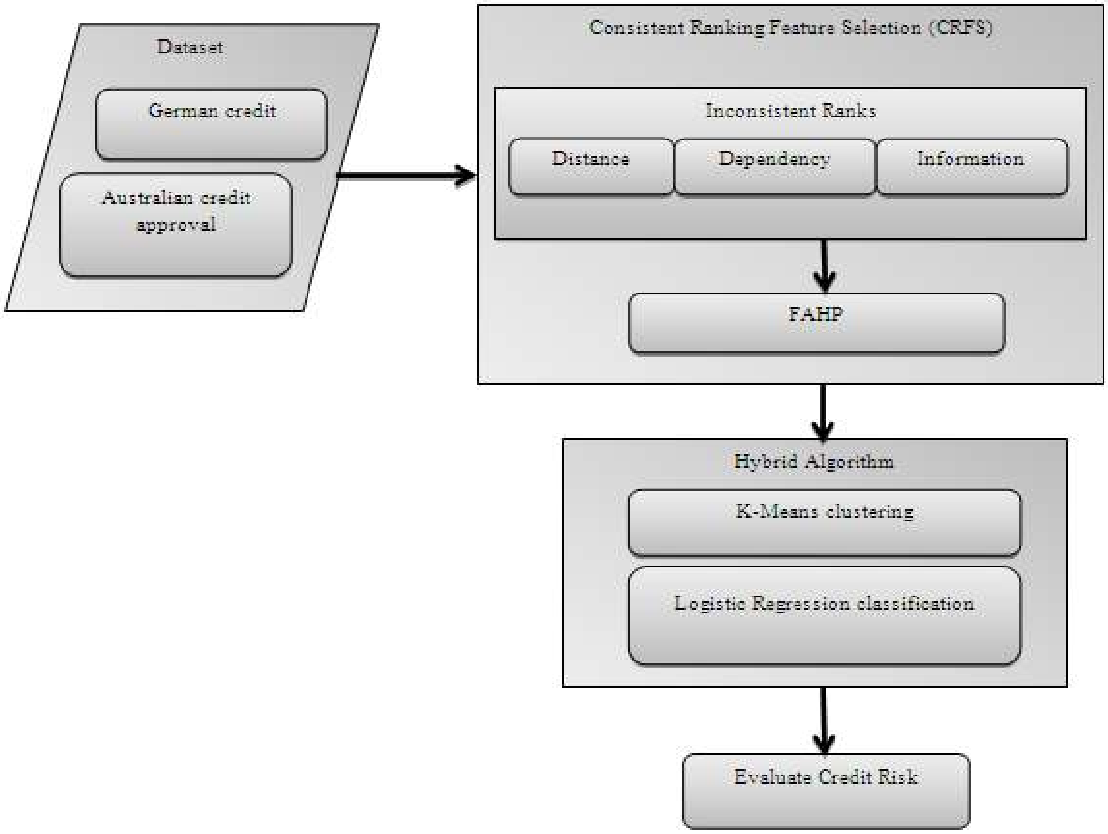

Fig. 1: System architecture diagram

- If the evaluation of current subset is maximized with the objective function then
  Feature subset = max (current subset)
- Repeat until it reaches the threshold

Based on the importance of generated ranks, pairwise comparison matrix is formed using Saaty scale of importance (Saaty, 1980).

The steps for the computation of pairwise comparison matrix are as follows:

- Saaty scale of importance = {a_{ij} =1-equal importance, 2, 4, 6, 8-intermediate, 5-strong importance, 7-very strong importance, 9-extreme importance}
- Pair Comparison Matrix [ ] = {a_{ij}} else if item i (row) is more important than item j (column) then the reciprocal of a_{ji}=1/a_{ij} is stored.

Now the FAHP process is applied to pairwise comparison matrix. The procedure for FAHP process to produce the consistent ranking on multiple evaluation criteria are described in steps below:

1. First, the Triangular Fuzzy Number (T_F_N) is formed by three parameters $\overline{a_{ij}}$ = (lower, modal, upper) where lower is the lower bound, upper is the upper bound and modal is the modal value of the fuzzy number. Based on the modal value, the T_F_N is defined as (modal-δ, modal, modal+δ) and the inverse T_F_N is defined as ($\frac{1}{\text{modal}+\delta}$, modal, $\frac{1}{\text{modal}-\delta}$) where the degree of fuzziness δ is considered as 1. The degree of fuzziness is more appropriate between 0.5 and 1 (Tang and Beynon, 2005).
2. Assign triangular fuzzy number to the pairwise comparison matrix to form a fuzzy comparison matrix (Chang, 1996)
   if Pair Comparison Matrix (i,j)> = 1 and i<>j
   $\overline{\{b_{ij}\}}$ = T_F_N
   Else
   $\overline{\{b_{ij}\}}$ = reciprocal of T_F_N
3. Calculate Fuzzy Synthetic Vector (FSN) by fuzzy addition operations (Chang, 1996)
   FSV1[ ]=$\sum_{i=1}^{m}$ ($\sum_{j=1}^{m} lower_j$,$\sum_{j=1}^{m} modal_j$, $\sum_{j=1}^{m} upper_j$)
   FSV2[ ]= $\sum_{i=1}^{m}$ ($\frac{1}{\sum_{j=1}^{modal} lower_j}$,$\frac{1}{\sum_{j=1}^{modal} modal_j}$, $\frac{1}{\sum_{j=1}^{modal} upper_j}$)
4. Calculate degree of possibility by the rule as

M2($lower_2$, $modal_2$, $upper_2$) ≥ M1($lower_1$, $modal_1$, $upper_1$) and defined as

- Equal to '1' if modal2≥modal1
- Equal to '0' if lower1≥upper2
- otherwise (1ower1-upper2) / (modal2-upper2)-(modal1-lower1)

5. Calculate Fuzzy preference weights by choosing the minimum between the degree of possibility of fuzzy number and 'm' degree of possibility of fuzzy numbers (where i = 1, 2……,.m)
6. Normalize the weights obtained in step 5.
7. Repeat step 2 to 6 for each evaluation criteria
8. Compute overall fuzzy preference weight
9. [ ]=$\sum_{i=1}^{m}$ W_{ii} * W_m
10. Obtain non-zero ranked overall fuzzy preference weights as optimal feature set

Hybrid algorithm: Using the optimal feature set, the performance measures are evaluated using effective hybrid algorithm (Beulah Jeba Jaya and Tamilselvi, 2015) (K-Means clustering-Logistic Regression classification). K-Means is one of the popular algorithm for clustering (Jain, 2010). In this study, it is used to divide the dataset into two homogeneous clusters based on class labels using the ranked optimal features as input and the implementation details are shown below:

1. Initialize K = Number of class labels in the dataset.
2. Determine the centroid coordinate and update the cluster center.
3. Calculate the Euclidean distance for all objects based on cluster center.
4. Group the objects based on minimum Euclidean distance from step 3.
5. Repeat step 3 and 4 until no changes in the cluster center.

From the clustered dataset, Logistic Regression classifier is applied to determine the performance measures. Logistic Regression is a popular classification method used in many data mining applications (Liu *et al*., 2009) and identified as one of the top classifier in detecting the unidentified fraud cases (Yi *et al*., 2011). The implementation steps of the Logistic Regression classifier is given below:

- Estimate the class probability (Y) which lies in the range between 0 and 1:

$$Y = \frac{1}{e^{(-w_0 - w_1 a_1 - w_2 a_2 \ldots\ldots - w_k a_k)}}$$

where, Wi is the weights and ai is the variable on dimensions i = 1, 2……k (no. of dimensions)

- Choose the weights to maximize the log-likelihood function

Maximize:

$$\sum_{i=1}^{n} (1 - x^i) \log(1 - prob[1|a_1^1, a_2^2 \ldots \ldots, a_k^k]) + x^i \log(1 - prob[1|a_1^1, a_2^2 \ldots \ldots, a_k^k])$$

- If Y greater than 0.5 then choose class 1 else class 0.

EXPERIMENTAL RESULTS

The proposed methodology is evaluated on two credit risk dataset from the UCI machine learning repository (Asuncion and Newman, 2007). The German credit dataset contains 1000 instances with 20 predictor attributes and 1 class attribute. 700 instances are good cases and 300 instances are bad cases. The Australian credit approval dataset contains 690 instances with 14 predictor attributes and 1 class attribute. The experiment is evaluated according to the proposed methodology and implemented using MATLAB (Version 7.9) for FAHP process based on multiple evaluation criteria ranks and WEKA 3.7 (Witten *et al*., 1999) to rank the features on different evaluation criteria and classification of credit risks.

Using ReliefFAttributeEval (Kira and Rendell, 1992), Symmetrical Uncert Attributet Eval (Yu and Liu, 2003) and Info Gain Attribute Eval feature selection algorithms within WEKA, multiple evaluation criteria ranks are generated. The result in Table 1 shows that there are inconsistencies in the ranks used by different algorithms and measures for German credit dataset. Similar inconsistencies in the ranks are observed with the Australian credit approval dataset (data not shown). The ranks in Table 1 are based on the ordering of the attributes in the German credit dataset. The attributes of German credit dataset are as follows: checking_status, duration, credit_history, purpose, credit_amount, savings_status, employment, installment_commitment, personal_status, other_parties, residence_since, property_magnitude, age, other_payment_plans, housing, existing_credits, job, num_dependents, own_telephone, foreign_worker.

Inconsistent multiple evaluation criteria ranks are applied to FAHP process. The pairwise comparison matrix (Saaty, 1980) is formed based on the ranks for each criterion and given as input to FAHP process to determine the fuzzy comparison matrix for different evaluation criteria and the overall fuzzy preference weights. The overall fuzzy preference weights and the ranked features in non-increasing order for the German credit dataset are shown in Table 2. Non-zero fuzzy preference weights for the features obtained are taken as optimal features.

Using these optimal features, hybrid algorithm (Beulah Jeba Jaya and Tamilselvi, 2015) is applied to evaluate the important identified performance measures (Beulah Jeba Jaya and Tamilselvi, 2014) for credit risk assessment.

Table 1: Inconsistent ranking on different evaluation criteria

Criteria	Ranks																			
Distance	1	3	4	6	7	9	12	8	19	2	14	10	13	18	17	11	5	16	15	20
Dependency	1	3	2	5	6	13	4	15	12	20	14	7	10	9	17	19	18	8	11	16
Information	1	3	2	6	4	5	12	7	15	13	14	9	20	10	17	19	18	8	11	16

Table 2: Overall fuzzy preference weights and the ranks

Criteria/German credit attributes	Distance	Dependency	Information	Overall weights	Overall rank
	0.076912	0.356325	0.566762		
Checking_status	0.1674	0.1698	0.1698	0.1696	1
Duration	0	0.1698	0.1698	0.1567	3
Credit_history	0.1674	0.1698	0.1698	0.1696	2
Purpose	0.1674	0.0500	0.1135	0.0950	6
Credit_amount	0	0.1135	0.1135	0.1048	5
Savings_status	0.1125	0.1135	0.1135	0.1134	4
Employment	0.1125	0	0.0500	0.0370	10
Installment_commitment	0.0534	0	0	0.0041	12
Personal_status	0.1125	0	0	0.0087	11
Other_parties	0	0	0	0	14
Residence_since	0	0	0	0	15
Property_magnitude	0.0534	0.0500	0.0500	0.0503	7
Age	0	0.1135	0	0.0405	9
Other_payment_plans	0	0	0	0	16
Housing	0	0.0500	0.0500	0.0462	8
Existing_credits	0	0	0	0	17
Job	0	0	0	0	18
Num_dependents	0	0	0	0	19
Own_telephone	0.0534	0	0	0.0041	13
Foreign_worker	0	0	0	0	20

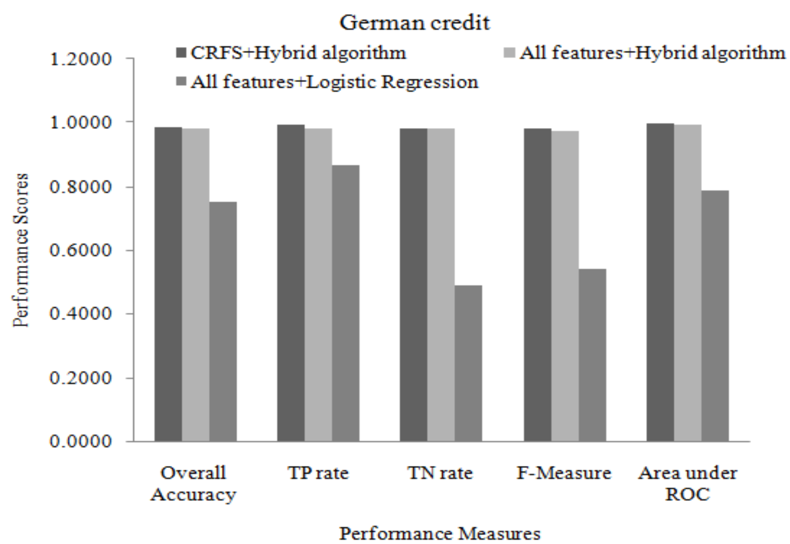

Fig. 2: Comparison-existing vs proposed methodology (Performance measures of German credit dataset)

Performance analysis: In this study, various comparative analysis has been done to show the improved results of proposed methodology for credit risk assessment. Fig. 2 shows the graphical comparison of performance measures between existing techniques and proposed methodology for German credit dataset. From Fig. 2 it is observed that the performance scores have shown significant improvement when proposed methodology is implemented. The proposed methodology (CRFS and hybrid algorithm) gives 98.6% overall accuracy, 99.2% TP rate, 98.0% TN rate, 98.1% F-Measure, 99.8% area under ROC which is 0.5% higher in overall accuracy, 1.1% higher in TP rate, equal with TN rate, 0.7% higher in F-Measure and 0.4% higher in area under ROC when compared with all features and hybrid algorithm (Beulah Jeba Jaya and Tamilselvi, 2015) and 23.4% higher in overall accuracy, 12.8% higher in TP rate, 49% with TN rate, 43.9% higher in F-Measure, 21.3% higher in Area under ROC when compared with all features and Logistic Regression classification.

Similarly, the comparative analysis of performance measures between existing techniques and proposed methodology for Australian credit approval dataset is shown in Fig. 3. It is observed that the performance scores are also greatly improved when proposed

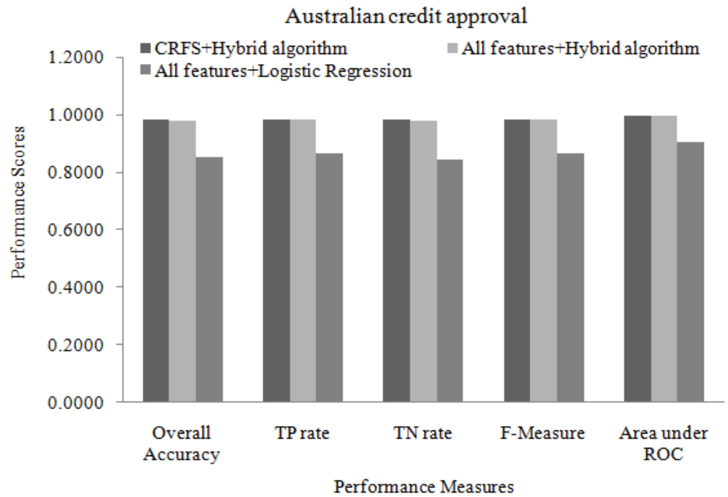

Fig. 3: Comparison-existing vs proposed methodology (Performance measures of Australian credit approval dataset)

Table 3: Performance measures improvement (CRFS + Hybrid algorithm Vs All features + Hybrid algorithm) with respect to dataset

Dataset	Overall accuracy	TP rate	TN rate	F-Measure	Area under ROC
German credit	0.5%	1.1%	0%	0.7%	0.4%
Australian credit approval	0.3%	0.3%	0.3%	0.2%	0%

methodology is implemented with Australian credit approval dataset. The combination of CRFS and hybrid algorithm have shown 98.3% overall accuracy, 98.4% TP rate, 98.2% TN rate, 98.4% F-Measure, 99.4% area under ROC which is higher by 0.3% in overall accuracy, TP rate, TN rate, 0.2% higher in F-Measure and equal with area under ROC when compared with all features and hybrid algorithm (Beulah Jeba Jaya and Tamilselvi, 2015). The proposed methodology is also compared with all features and Logistic Regression classification and observed that it is 12.9% higher in overall accuracy, 11.8% higher in TP rate, 13.9% with TN rate and 11.9% higher in F-Measure and 8.8% higher in area under ROC.

The improvement of performance measures by the proposed methodology Vs all features and hybrid algorithm (Beulah Jeba Jaya and Tamilselvi, 2015) with respect to two different datasets are discussed in Table 3.

Based on the above results, the proposed methodology for credit risk assessment provides better performance scores than existing technique. Thus, consistent ranking feature selection along with hybrid algorithm shows a significant role in improving the credit risk assessment.

CONCLUSION

Existing feature selection algorithms gives different reduced dataset for different single evaluation criteria. To improve consistency, consistent ranking with FAHP process applied to determine the optimal feature set. In addition, hybrid algorithm is also used in this study (Beulah Jeba Jaya and Tamilselvi, 2015) to improve the classification performance measures. The proposed methodology is implemented and compared with the existing techniques using two different datasets from UCI repository and obtained consistent ranking using multi-criteria decision making with significant improvement in performance measures such as overall accuracy by 0.5%, TP rate by 1.1%, equal with TN rate, F-Measure by 0.7% and Area under ROC by 0.4% using German credit dataset. For Australian credit approval dataset, the performance measures are improved with the proposed methodology by 0.3% in overall accuracy, TP rate, TN rate, 0.2 % in F-Measure and equal with area under ROC.As a future work, a fuzzy concept will be extended for clustering and classification to provide best feasible solution for credit risk assessment.

REFERENCES

Adomavicius, G. and Y.O. Kwon, 2007. New recommendation techniques for multi-criteria rating systems. IEEE Intell. Syst., 22(3): 48-55.

Asuncion, A. and D. Newman, 2007. UCI machine learning repository. School of Information and Computer Science. University of California, Irvine, CA. Retrieved from: http://mlearn.ics.uci.edu/MLRepository.html.

Ben-Bassat, M., 1982. Pattern Recognition and Reduction of Dimensionality. In: Krishnaiah, P.R. and Kanal, L.N. (Eds.), Handbook of Statistics-II. North-Holland, Amsterdam, pp: 773-791.

Beulah Jeba Jaya, Y. and J.J. Tamilselvi, 2014. Simplified MCDM analytical weighted model for ranking classifiers in financial risk datasets. Proceeding of the International Conference on Intelligent Computing Applications (ICICA, 2014). Coimbatore. India, pp: 158-161.

Beulah Jeba Jaya, Y. and J.J. Tamilselvi, 2015. Multiple criteria decision making based credit risk prediction using optimal cut-off point approach. Int. J. Appl. Eng. Res., 10(8): 20041-20054.

Buckley, J.J., 1985. Fuzzy hierarchical analysis. Fuzzy Set. Syst., 17(1): 233-247.

Chan, F.T.S. and N. Kumar, 2005. Global supplier development considering risk factors using fuzzy extended AHP-based approach. Omega, 35(4): 417-431.

Chang, D.Y., 1996. Applications of the extent analysis method on fuzzy AHP. Eur. J. Oper. Res., 95: 649-655.

Dash, M. and H. Liu, 1997. Feature selection for classification. Int. J. Intell. Data Anal., 1(3): 131-156.

Guner, H., O. Mutlu and O. Kulak, 2005. Supplier selection in fuzzy environment. In: Durmusoglu, B. and C. Kahraman (Eds.), Proceeding of the 35th International Conference on Computers and Industrial Engineering. Turkey, Istanbul, pp: 839-844.

Jain, A.K., 2010. Data clustering: 50 years beyond k-means. Pattern Recogn. Lett., 31(8): 651-666.

Jiliang, T., A. Salem and L. Huan, 2014. Feature Selection for Classification: A Review. In: Charu, A. (Ed.), Data Classification: Algorithms and Applications. CRC Press, 2014.

Kira, K. and L.A. Rendell, 1992. A practical approach to feature selection. Proceeding of 9th International Workshop on Machine Learning, pp: 249-256.

Khanbabaei, M. and M. Alborzi, 2013. The use of genetic algorithm, clustering and feature selection techniques in construction of decision tree models for credit scoring. Int. J. Manag. Inform. Technol. (IJMIT), 5(4): 13-31.

Kohavi, R. and D. Sommerfield, 1995. Feature subset selection using the wrapper method: Over fitting and dynamic search space topology. Proceeding of 1st International Conference on Knowledge Discovery and Data Mining, pp: 192-197.

Koller, D. and M. Sahami, 1996. Toward optimal feature selection. Proceeding of International Conference on Machine Learning. Bari, Italy, pp: 284-292.

Liu, H. and L. Yu, 2005. Toward integrating feature selection algorithms for classification and clustering. IEEE T. Knowl. Data En., 17(4): 491-502.

Liu, J., J. Chen and J. Ye, 2009. Large-scale sparse logistic regression. Proceeding of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Paris, France, pp: 547-556.

Modrzejewski, M., 1993. Feature selection using rough sets theory. Proceeding of the European Conference on Machine Learning. Vienna, Austria, pp: 213-226.

Novakovic, J., 2009. Using information gain attribute evaluation to classify sonar targets. Proceeding of 17 Telecommunications Forum. Telfor, Belgrade, pp: 1351-1354.

Ramaswami, M. and R. Bhaskaran, 2009. Study on feature selection techniques in educational data mining. J. Comput., 1(1): 7-11.

Robnik-Sikonja, M. and I. Kononenko, 2003. Theoretical and empirical analysis of ReliefF and RreliefF. Mach. Learn., 53: 23-69.

Saaty, T.L., 1980. The Analytic Hierarchy Process. McGraw Hill International, New York.

Senthamarai Kannan, S. and N. Ramaraj, 2010. A novel hybrid feature selection via symmetrical uncertainty ranking based local memetic search algorithm. Knowl-Based Syst., 23(6): 580-585.

Sun, C.C., 2010. A performance evaluation model by integrating fuzzy AHP and fuzzy TOPSIS methods. Expert Syst. Appl., 37(12): 7745-7754.

Tang, Y.C. and M.J. Beynon, 2005. Application and development of a fuzzy analytic hierarchy process within a capital investment study. J. Econ. Manage., 1(2): 207-230.

Van Laarhoven, P.J.M. and W. Pedrcyz, 1983. A fuzzy extension of Saaty's priority theory. Fuzzy Set. Syst., 11: 229-241.

Witten, I.H., E. Frank, L. Trigg, M. Hall, G. Holmes and S.J. Cunningham, 1999. Weka: Practical machine learning tools and techniques with java implementations. Proceeding of the ICONIP/ANZIIS/ANNES'99 Workshop on Emerging Knowledge Engineering and Connectionist-based Information Systems, pp: 192-196.

Yi, P., W. Guoxun, K. Gang and S. Yong, 2011. An empirical study of classification algorithm evaluation for financial risk prediction. Appl. Soft Comput., 11(2): 2906-2915.

Yilmaz, T., E. Gulen, A. Yazici and M. Kitsuregawa, 2012. A relief-based modality weighting approach for multimodal information retrieval. Proceeding of ACM International Conference on Multimedia Retrieval (ACM ICMR), Hong Kong.

Yu, L. and H. Liu, 2003. Feature selection for high-dimensional data: A fast correlation-based filter solution. Proceeding of the 20th International Conference on Machine Learning, pp: 856-863.

Zeng, H.J., X.H. Wang, Z. Chen, H. Lu and W.Y. Ma, 2003. CBC: Clustering based text classification requiring minimal labelled data. Proceeding of the IEEE International Conference on Data Mining, pp: 443-450.