VOL. 2, NO. 11, Dec 2012 ISSN 2225-7217

ARPN Journal of Science and Technology

http://www.ejournalofscience.org

Empirical Evaluation of Feature Selection Technique in Educational Data Mining

[1] A.S. Kavitha, [2] R. Kavitha, [3] J. Viji Gripsy
[1, 2, 3] Department Of Computer Science, PSGR Krishnammal College for Women, Coimbatore, India

ABSTRACT

In machine learning the classification task is commonly referred to as supervised learning. In supervised learning there is a specified set of classes and objects are labeled with the appropriate class. The goal is to generalize from the training objects that will enable novel objects to be identified as belonging to one of the classes. Evaluating the performance of learning algorithms is a fundamental aspect of machine learning. The primary objective of this thesis is to study the classification accuracy using feature selection with machine learning algorithms. Feature selection is considered successful if the dimensionality of the data is reduced and accuracy of a learning algorithm improves or remains the same. Hence our contribution in this research is to prepare an educational dataset with real time feedback from students and try to apply the same with weka tool to measure the classification accuracy. Some part of implementation is compiled with weka, which is written in java and experiment with weka explorer.

Keywords: *Feature Selection, One R, PART, K-means algorithm, RELIEF algorithm*

1. INTRODUCTION

Data mining is a term coined to describe the process of sifting through large databases for interesting patterns and relationships. With the declining cost of disk storage, the size of many corporate and industrial databases have grown to the point where analysis by anything but parallelized machine learning algorithms running on special parallel hardware is infeasible. Two approaches that enable standard machine learning algorithms to be applied to large databases are feature selection and sampling.

In 1991 it was alleged that the amount of stored information doubles every twenty months. Unfortunately, as the amount of machine readable information increases, the ability to understand and make use of it does not keep pace with its growth. Machine learning provides tools by which large quantities of data can be automatically analyzed. Fundamental to machine learning is feature selection. Feature selection for clustering or classification tasks can be accomplished on the basis of correlation between features, and that such a feature selection process can be beneficial to a variety of common machine learning algorithms. A technique for correlation-based feature selection, based on ideas from test theory, is developed and evaluated using common machine learning algorithms on a variety of natural and artificial problems. The feature selector is simple and fast to execute. It eliminates irrelevant and redundant data and, in some cases, improves the performance of learning algorithms.

Walters and Soyibo [36] conducted a study to determine Jamaican high school students' (population n=305) level of performance on five integrated science process skills with performance linked to gender, grade level, school location, school type, student type, and socio-economic background (SEB). The results revealed that there was a positive significant relationship between academic performance of the student and the nature of the school.

Khan [37] conducted a performance study on 400 students comprising 200 boys and 200 girls selected from the senior secondary school of Aligarh Muslim University, Aligarh, India with a main objective to establish the prognostic value of different measures of cognition, personality and demographic variables for success at higher secondary level in science stream.

Hijazi and Naqvi [38] conducted as study on the student performance by selecting a sample of 300 students (225 males, 75 females) from a group of colleges affiliated to Punjab university of Pakistan. The hypothesis that was stated as "Student's attitude towards attendance in class, hours spent in study on daily basis after college, students' family income, students' mother's age and mother's education are significantly related with student performance" was framed.

Kristjansson, Sigfusdottir and Allegrante [39] made a study to estimate the relationship between health behaviors, body mass index (BMI), self-esteem and the academic achievement of adolescents. The authors analyzed survey data related to 6,346 adolescents in Iceland and it was found that the factors like lower BMI, physical activity, and good dietary habits were well associated with higher academic achievement.

Moriana et al. [40] studied the possible influence of extra-curricular activities like study-related (tutoring or private classes, computers) and/or sports-related (indoor

ARPN Journal of Science and Technology

http://www.ejournalofscience.org

and outdoor games) on the academic performance of the secondary school students in Spain.

Cortez and Silva [46] attempted to predict failure in the two core classes (Mathematics and Portuguese) of two secondary school students from the Alentejo region of Portugal by utilizing 29 predictive variables. Four data mining algorithms such as Decision Tree (DT), Random Forest (RF), Neural Network (NN) and Support Vector Machine (SVM) were applied on a data set of 788 students, who appeared in 2006 examination. It was reported that DT and NN algorithms had the predictive accuracy of 93% and 91% for two-class dataset (pass/fail) respectively. It was also reported that both DT and NN algorithms had the predictive accuracy of 72% for a four-class dataset.

2. FEATURE SELECTION SYSTEM ARCHITECTURE

A simple three step feature selection approach is explained below. The goal of this architecture is to reduce a large set of features (on the order of thousands) to a small subset of features (on the order of tens), without significantly reducing the system's ability. The basic three steps of this system are:

- In first step the irrelevant features are removed.
- After that the redundant features are removed.
- And finally a feature selection algorithm is applied to the remaining features.

In this approach each step is working as a filter that reduces the number of candidate features, until finally only a small subset remains.

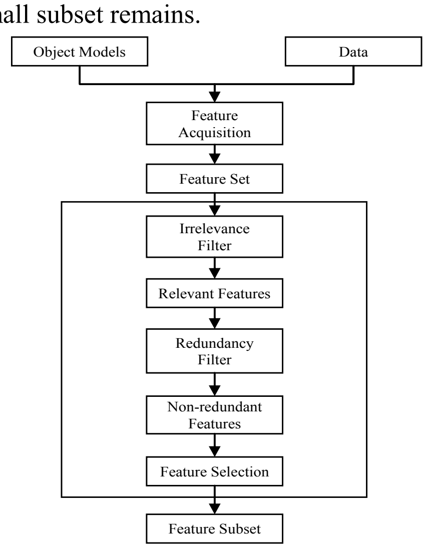

Fig 4.1: Feature Selection System Architecture

The first filter removes irrelevant features using a modified form of the Relief algorithm, which assigns relevance values to features by treating training samples as points in feature space. For each sample, it finds the nearest "hit" (another sample of the same class) and "miss" (a sample of a different class), and adjusts the relevance value of each feature according to the square of the feature difference between the sample and the hit and miss. There are several modifications to Relief to generalize it for continuous features and to make it more robust in the presence of noise. This system adopts Kononenko's modifications, and modifies Relief again to remove a bias against non-monotonic features, as described in [13].

Within this feature selection system, Relief is used as a relevance filter. Therefore it threshold the relevance values, to divide the feature set into relevant and irrelevant features. This can be done either by threshold the relevance value directly, or by selecting the highest n values and discarding the remaining features. In either case, relief does not detect redundancy, so the remaining feature set still contains redundant features.

The second step is a redundancy filter that uses the K-means algorithm [14] to cluster features according to how well they correlate to each other. When feature clusters are discovered, only the feature with the highest Relief score is kept; the other features in the cluster are removed from the feature set. This is an unusual application of K-means clustering, in that features are clustered (instead of samples), and correlation is used as the distance measure.

The third and final filter is a combinatorial feature selection algorithm. The following algorithms are used for to perform the above mentioned operations.

2.1 RELIEF algorithm

RELIEF is considered one of the most successful algorithms for assessing the quality of features due to its simplicity and effectiveness. It has been recently proved that RELIEF is an online algorithm that solves a convex optimization problem with a margin based objective function.

The simple intuition behind the RELIEF-F algorithm is that a good feature is a feature with little within class variance and generous amounts of between-class variance. A bad feature is characterized by within-class and between-class variances of magnitudes roughly equal. The computation of the weight vector is done in an iterative fashion, with all weights initially set to 0. All features in the training set are normalized (all values are set within the range [0 ... 1]) and thereafter used to update the weight vector as follows.

ARPN Journal of Science and Technology

http://www.ejournalofscience.org

On each iteration a random instance is chosen. The nearest hit is located, where hit is an instance of the same class as that of the randomly chosen instance. The nearest misses of all the classes but that of the randomly chosen instance is located, where a miss is an instance of a class different from that of the randomly chosen instance. If the difference between the attribute of the chosen instance and the nearest hit is lower than the corresponding value for the miss (es) then the weight value will be increased.

2.2 K-means Algorithm

The main idea is to define k centroids, one for each cluster. These centroids should be placed in a cunning way because of different location causes different result. So, the better choice is to place them as much as possible far away from each other. The next step is to take each point belonging to a given data set and associate it to the nearest centroid. When no point is pending, the first step is completed and an early group age is done. At this point it is in the need to re-calculate k new centroids as bar centers of the clusters resulting from the previous step. After take these k new centroids, a new binding has to be done between the same data set points and the nearest new centroid.

A loop has been generated. As a result of this loop, notice that the k centroids change their location step by step until no more changes are done. In other words centroids do not move any more. Finally, this algorithm aims at minimizing an objective function, in this case a squared error function. The objective function

$$J = \sum_{j=1}^{k}\sum_{i=1}^{n}\left\| x_i^{(j)} - c_j \right\|^2$$

where $\left\| x_i^{(j)} - c_j \right\|^2$ is a chosen distance measure between a data point $x_i^{(j)}$ and the cluster centre c_j, is an indicator of the distance of the n data points from their respective cluster centers.

As the dimensionality of a domain expands, the number of features N increases. Finding an optimal feature subset is intractable and problems related feature selections have been proved to be NP-hard [44]. At this juncture, it is essential to describe traditional feature selection process, which consists of four basic steps, namely, subset generation, subset evaluation, stopping criterion, and validation [39]. Subset generation is a search process that produces candidate feature subsets for evaluation based on a certain search strategy

Ranking of features determines the importance of any individual feature, neglecting their possible interactions. Ranking methods are based on statistics, information theory, or on some functions of classifier's outputs [40]. Algorithms for feature selection fall into two broad categories namely wrappers that use the learning algorithm itself to evaluate the usefulness of features and filters that evaluate features according to heuristics based on general characteristics of the data [39].

2.3 Machine Learning Algorithms

The following are the classification algorithms used in the preprocessing data through three classifiers such as OneR, Naïve bayes and PART in WEKA.

2.3.1 One R

OneR, short for “One Rule”, is a simple classification algorithm that generates a one-level decision tree. OneR is able to infer typically simple, yet accurate, classification rules from a set of instances. Comprehensive studies of OneR’s performance have shown it produces rules only slightly less accurate than state-of-the-art learning schemes while producing rules that are simple for humans to interpret.

OneR is also able to handle missing values and numeric attributes showing adaptability despite simplicity. The OneR algorithm creates one rule for each attribute in the training data, then selects the rule with the smallest error rate as its ‘one rule’. To create a rule for an attribute, the most frequent class for each attribute value must be determined. A rule is simply a set of attribute values bound to their majority class; one such binding for each attribute value of the attribute the rule is based on.

2.3.2 Pseudo-code for the One R Algorithm

- For each attribute A,
- For each value VA of the attribute, make a rule as follows:
- count how often each class appears
- find the most frequent class Cf
- create a rule when A=VA; class attribute value = Cf
- End For-Each
- Calculate the error rate of all rules
- End For-Each
- Chose the rule with the smallest error rate

The error rate of a rule is the number of training data instances in which the class of an attribute value does not agree with the binding for that attribute value in the rule. OneR selects the rule with the lowest error rate. In the event that two or more rules have the same error rate, the rule is chosen at random.

VOL. 2, NO. 11, Dec 2012 ISSN 2225-7217

ARPN Journal of Science and Technology

http://www.ejournalofscience.org

OneR, short for "One Rule", is a simple, yet accurate, classification algorithm that generates one rule for each predictor in the data, then selects the rule with the smallest total error as its "one rule". To create a rule for a predictor, construct a frequency table for each predictor against the target. It has been shown that OneR produces rules only slightly less accurate than state-of-the-art classification algorithms while producing rules that are simple for humans to interpret

3. PART

Projective Adaptive Resonance Theory (PART) is a class of ANN that is capable of self learning, PART1 type accepts binary inputs used primarily in pattern classification applications like text clustering (Carpenter and Grosberg, 1987), where documents are presented as binary strings characterizing the occurrences of features, including: taxonomy generation, topic extraction and search engines hits grouping, which are quite useful in many modern applications like hierarchical web search. Although the supervised Text Categorization (TC) is the best for such applications in terms of quality, it lacks adaptability, requires expert's intervention and occasional retraining.

An interface layer neuron Xi (i = 1, …, n), where n is the number of input units, is connected to each neuron of Y layer Yj(j = 1,..., m), where m is the maximum allowed number of clusters, by two weighted pathways. Signals broadcast from neurons of layer X to neurons of layer Y over connections pathways with bottom-up weight matrix bij and from neurons of layer X to neurons of layer Y over connections pathways with top-down weight matrix tji (i = 1,.., n and j = 1,..,m).

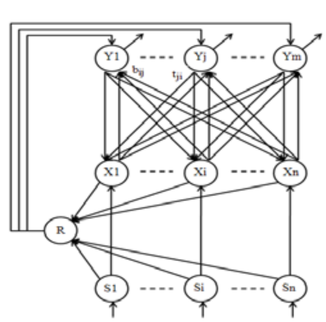

Fig 1: ART1 structure

4. RESULTS & DISCUSSION

The main source of data for this study is the responses obtained from students through a questionnaire with close-end questions. The responses give demographic details, family details, socio-economic details, previous academic performance at secondary level from different schools and other environmental details. A total of 150 higher secondary students from different schools in different districts of the state Tamil Nadu, India, supplied the details. The proposed work notice that the original feature vector of student performance data consisted of 35 features that were predictive variables. Besides, there was a two-case class variable result (pass / fail), which was considered as response variable. All these predictive and responsive variables shown in Table 1 belonged to the type of nominal data.

Feature selection is normally done by searching the space of attribute subsets, evaluating each one. Greedy search method was performed to find out the best feature sets and they are listed below for reference:

a. Correlation-based Attribute evaluation (CB),
b. Chi-Square Attribute evaluation (CH),
c. Gain-Ratio Attribute evaluation (GR),
d. Information-Gain Attribute evaluation (IG),
e. Relief Attribute evaluation (RF) and
f. Symmetrical Uncertainty Attribute evaluation (SU)

These entire filter techniques mentioned above could assess the relevance of features [16] on the basis of the intrinsic properties of the data. Feature selection often increases classifier efficiency through the reduction of the size of the effective features. Therefore, there is a need to verify the relevance of all the features in the feature vector. In this connection, all the above six feature selection techniques based on different measures to choose the best subsets for a given cardinality are performed. The measures like ROC Values and Macro-Average F1-Measure values are used in the present investigation.

5. ASSESSING CLASIFICATION PREFORMANCE

It is known that there is no such thing as the best classification algorithm, when considering a broad scale of classification problems. It is the type of problem, prior knowledge, and other information that determines which type of classifier should provide the best performance. An important aspect thus, when confronted with a specific classification task is the evaluation and comparison of the performance of various classifiers and their parameters. To get a more or less reliable result, the trained classifier should be test on an independent set of patterns.

The performance of the classifier can be analyzed by considering the errors that have been made on the validation sets. For two-class classification problems, the

VOL. 2, NO. 11, Dec 2012 ISSN 2225-7217

ARPN Journal of Science and Technology

http://www.ejournalofscience.org

performance of a classifier can be summarized in four numbers.

- TP (True Positives) : actual positive patterns predicted as positives
- TN (True Negatives) : actual negative patterns predicted as negatives
- FP (False Positives) : actual negative patterns predicted as positives
- FN (False Negatives) : actual positive patterns predicted as negatives

From these numbers a series of performance measures have been derived, all aiming at summarizing the four numbers into one, to make it easier to compare classifiers:

- Error Rate
- Accuracy
- TP rate/recall sensitivity
- FP rate
- Precision / specificity
- Correlation coefficient
- Approximate correlation

These methods have been commonly used in machine learning research, yet have one main drawback: they assume a balanced class distribution. However in many real world datasets like biological datasets, the class distribution is highly imbalanced. An example negative patterns magnitudes higher than the number of actual sites. Imagine a dataset with 98% negative patterns and classifier always outputs the class negative. A measure that was recently introduced to incorporate content balancing for imbalanced dataset is the q9 measure.

Another way to evaluate the performance of a classifier is by analysis of the so-called Receiver Operator Curve (ROC), introduced in 1997 by Provost and Fawcett. On such a curve, the FP rate is plotted on the x-axis versus the TP rate (sensitivity or recall) on the y-axis. By varying the decision threshold, several values can be obtained for these rates.

Table1: Peak ROC Values

6. RESULTS

This section discuss about the results of feature selection techniques applied in educational data mining. The primary objective of this research is to evaluate the performance prediction of feature selection in EDM with respect to different dataset. It is an extension of existing research by M.Ramaswami work with 1500 records that was reduced to 150 in this work by using feature selection method. I try to confirm the quality dependents on dimensions or data. The dataset contains 150 records which were collected from college students.

The following table shows weighted mean ROC value of selected evaluation filters. the efficiency is not always extracted from the same sequence.

F1- Measure, which is another measure for evaluating the effectiveness of feature selection techniques, is the harmonic mean of the precision and recall and this metric is more relevant for evaluating the performance of the classifier.

VOL. 2, NO. 11, Dec 2012 ISSN 2225-7217

ARPN Journal of Science and Technology

http://www.ejournalofscience.org

Table1: Peak ROC Values

Feature Selection Evaluation Filters	Our Results		Previous Results	
	No. of Attributes	ROC Value	No. of Attributes	ROC Value
Chi-square Attribute	6	0.894	5	0.726
Gain Ratio Attribute	6	0.898	7	0.721
Information Gain Attribute	7	0.812	7	0.729
Relief Attribute	10	0.846	12	0.714
Symmetrical-Uncert. Attribute	8	0.822	6	0.721

It is observed from the above table that high ROC value (0.898) is identified in Gain Ratio Attribute evaluation, which is the optimal dimensionality in the dataset among selected filters. In the previous study, Information Gain attribute has

The generated macro-averaged F-measures, which are presented in Table 5.2, could show each of the feature selection attributes considered for this study and also shows the previous results.

The following table shows the peak values of f1-measure between newly extracted results and existing results.

Table 5.2: Peak Values of F1-Measure

Feature Selection Evaluation Filters	Our Results		Previous Results	
	No. of Attributes	F1 Value	No. of Attributes	F1 Value
Chi-square Attribute	14	0.635	12	0.584
Gain Ratio Attribute	14	0.680	12	0.592
Information Gain Attribute	14	0.678	12	0.588
Relief Attribute	14	0.681	12	0.582
Symmetrical-Uncert. Attribute	14	0.680	12	0.592

It is evident from the above table that among the selected evaluation filters, newly extracted result shows relief attribute has high value 0.681.

The following Figure 1 shows the peak values of f1-measure of existing results.

ARPN Journal of Science and Technology

http://www.ejournalofscience.org

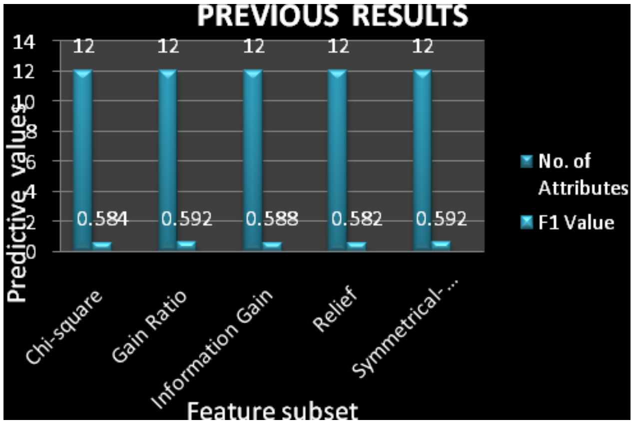

Fig 1: F1 Value of existing results

Further, the new performance results were compared using paired t-test about the overall classification accuracy between Naivebayes, OneR and PART. Naivebayes extracted 78.93%, OneR extracted 96% and PART extracted 95.27%. The naivebayes method found significant difference with OneR and PART.

Independent t-test is applied to evaluate the significant performance difference among newly extracted result and existing result. The t-test result
confirms there is no significant difference among the EDM dataset performance.

The following Figure 5.6 shows the peak values of f1-measure of newly extracted results.

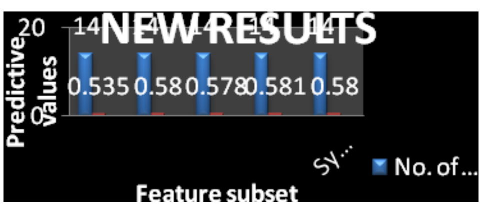

Fig 2: F1 Value of existing results

Compare to the above figure it clearly shows the difference in no of attribute and increase value in F1 measure between existing and newly extracted results.

7. PROPOSED FEATURE OPTIMIZATION

Feature extraction and selection are two fundamental problems in machine learning research. Nosst only can their proper design reduce system complexity and processing time, but they can also enhance system performance in many cases. A commonly used method for feature extraction and selection is to pre-multiply a projection matrix to pattern vectors with the aim to optimize a certain criterion function.

Feature weighting considered as well-established optimization techniques and allows for efficient algorithm implementation. Among the existing feature weighting algorithms, RELIEF is considered one of the most successful ones due to its simplicity and effectiveness. It has been recently shown that RELIEF is an online algorithm that solves a convex optimization problem aimed at maximizing a margin-based objective function.

We first present a brief review of RELIEF. Let $D = f(x_n, y_n)^N n{=}1 \subset R^i \times \{\pm 1\}$ be a training dataset, where x_n is the n-th data sample and y_n is its corresponding class label. The basic idea of RELIEF is to iteratively estimate feature weights according to their ability to discriminate between neighboring patterns. In each iteration, a pattern x is randomly selected and then two nearest neighbors of x are found, one from the same class (termed the nearest hit or NH) and the other from the opposite class (termed the nearest miss or NM). The weight of the i-th feature is then updated as: $wi = wi + |x^{(i)} - NM^{(i)}(x)| - |x^{(i)} - NH^{(i)}(x)|$, for $1 \leq i \leq I$. We provide a mathematical interpretation for the seemingly heuristic RELIEF algorithm. We define the margin for a pattern x_n as $P_n = d(x_n - NM(x_n)) - d(x_n - NH(x_n))$, where $NM(x_n)$ and $NH(x_n)$ are the nearest miss and hit of x_n, respectively, and $d(\cdot)$ is a distance function. For the purpose of our proposal, we define $d(x) = \Sigma_i |x_i|$, which is consistent with the distance function used in the original RELIEF algorithm.

The modified chi-square evaluator tested through java implementation. According to the result, we try to compute the difference among original chi-squared evaluator and modified one, which shows an effective performance. Since, the modified algorithm using RELIEF weighted correlation. Hence, irrelevant / classification error can be reduced. The following figure exhibits the performance of modified chi-squared attributor evaluator with classification error sequence.

VOL. 2, NO. 11, Dec 2012 ISSN 2225-7217

ARPN Journal of Science and Technology

http://www.ejournalofscience.org

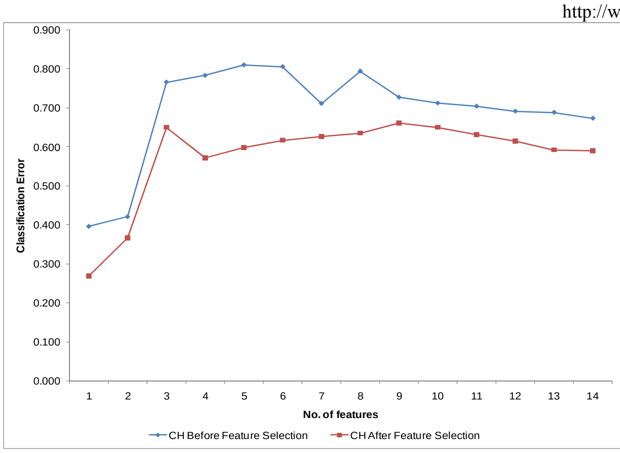

Fig 3: Modified Chi-squared Attribute Evaluation

The modified algorithm is playing a significant role in reducing classification error. It is still need to be evaluated with other attribute evaluators to ensure the consistent quality of the RELIEF over other algorithms.

8. CONCLUSION

Feature selection has been an active and fruitful field of research area in pattern recognition, machine learning, statistics and data mining communities. The main objective of feature selection is to choose a subset of input variables by eliminating features, which are irrelevant or of no predictive information. This study contains educational data mining dataset with 150 records and total of 34 attributes finally considered and out of which HSC grade is a response variable and other variables are predictive variables. This study result reveals that consideration of feature selection model is effective, but the efficiency is not extracted in the same sequence always. Hence it is suggested to reconstruction of this model is required for consistent performance in educational data mining dataset. The F1-measure also confirms its variation in result. In order to evaluate the performance difference, independent t-test is applied. It shows that there is a significant difference among the EDM performance among previous work.

Reconstruction of existing model is not quite easy job. Hence, we focused on classification errors for enhancement. It is observed from the result that all algorithms are having minimum level of classification errors. The RELIEF is an optimization algorithm, which is used to improve the quality of feature selection. This concept has been tested with chi-squared attribute evaluator and the result shows significant improvement in reducing the classification errors. It is suggested to test the same sequence to other feature selection techniques to ensure the quality.

REFERENCES

[1] Introduction to Data Mining and Knowledge Discovery, Third Edition ISBN: 1-892095-02-5, Two Crows Corporation, 10500 Falls Road, Potomac, MD 20854 (U.S.A.), 1999.

[2] Larose, D. T., "Discovering Knowledge in Data: An Introduction to Data Mining", ISBN 0-471-66657-2, John Wiley & Sons, Inc, 2005.

[3] Dunham, M. H., Sridhar S., "Data Mining: Introductory and Advanced Topics", Pearson Education, New Delhi, ISBN: 81-7758-785-4, 1st Edition, 2006.

[4] Chapman, P., Clinton, J., Kerber, R., Khabaza, T.,Reinartz, T., Shearer, C. and Wirth, R.. "CRISP-DM 1.0 : Step-by-step data mining guide, NCR Systems Engineering Copenhagen (USA and Denmark), DaimlerChrysler AG (Germany), SPSS Inc. (USA) and OHRA Verzekeringenen Bank Group B.V (The Netherlands), 2000".

[5] Fayyad, U., Piatetsky-Shapiro, G., and Smyth P., "From Data Mining to Knowledge Discovery in Databases," AI Magazine, American Association for Artificial Intelligence, 1996.

[6] Bernstein, A. and Provost, F., "An Intelligent Assistant for the Knowledge Discovery Process", Working Paper of the Center for Digital Economy Research, New York University and also presented at the IJCAI 2001 Workshop on Wrappers for Performance Enhancement in Knowledge Discovery in Databases.

[7] Baazaoui, Z., H., Faiz, S., and Ben Ghezala, H., "A Framework for Data Mining Based Multi-Agent: An Application to Spatial Data, volume 5, ISSN 1307-6884," Proceedings of World Academy of Science, Engineering and Technology, April 2005.

[8] Botia, J. A., Garijo, M. y Velasco, J. R., Skarmeta, A. F., "A Generic Data mining System basic design and implementation guidelines", A Technical Project Report of CYCYT project of Spanish Government. 1998. Web Site: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.53.1935

[9] Sirgo, J., Lopez, A., Janez, R., Blanco, R., Abajo, N., Tarrio, M., Perez, R., "A Data Mining Engine based

ARPN Journal of Science and Technology

http://www.ejournalofscience.org

on Internet, Emerging Technologies and Factory Automation," Proceedings ETFA '03, IEEE Conference, 16-19 Sept. 2003.WebSite:www.citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.11.8955

[10] Vajirkar, P., Singh, S., and Lee, Y., "Context-Aware Data Mining Framework for Wireless Medical Application". Lecture Notes in Computer Science (LNCS), Volume 2736, Springer-Verlag. ISBN 3-540- 40806-1, pp. 381 – 391.

[11] Romero, C., Ventura, S. and De-Bra, P. "Knowledge Discovery with Genetic Programming for Providing Feedback to Courseware Authors, Kluwer Academic Publishers, Printed in the Netherlands, 30/08/2004".

[12] Kusiak, A., Kernstine, K.H., Kern, J.A., McLaughlin, K.A., and Tseng, T.L., "Data Mining: Medical and Engineering Case Studies". Proceedings of the Industrial Engineering Research 2000 Conference, Cleveland, Ohio, pp. 1-7,May 21-23, 2000.

[13] Jadhav, S. R., and Kumbargoudar, P., "Multimedia Data Mining in Digital Libraries: Standards and Features".Proceedings of conference Recent advances in Information Science and Technology READIT – 2007, pp 54-59, Organized by Madras Library Association - Kalpakkam Chapter & Scintific information Resource Division, Indira Gandhi Center for Atomic research, Department of Atomic Energy, Kalpakkam, Tamilnadu,India. 12-13 July 2007

[14] Witten, I.H., Frank, E. (2005) Data Mining: Practical Machine Learning Tools and Techniques: 2nd Edition. San Francisco: Morgan Kaufmann.

[15] Barnes, T., Bitzer, D., Vouk, M. (2005) Experimental Analysis of the Q-Matrix Method in Knowledge Discovery. Lecture Notes in Computer Science 3488: Foundations of Intelligent Systems, 603-611.

[16] Pavlik, P., Cen, H., Wu, L., Koedinger, K. (2008) Using Item-Type Performance Covariance to Improve the Skill Model of an Existing Tutor. Proceedings of the First International Conference on Educational Data Mining, 77-86.

[17] Romero, C., Ventura, S. (2007) Educational Data Mining: A Survey from 1995 to 2005. Expert Systems with Applications, 33(1), 135-146.

[18] Moore, A. (2005) Statistical Data Mining Tutorials. Available online at http://www.autonlab.org/tutorials/ . Retrieved June 27, 2008

[19] Kay, J., N. Maisonneuve, K. Yacef and P. Reimann (2006). The Big Five and Visualisations of Team Work Activity. Proceedings of Intelligent Tutoring Systems (ITS06), M. Ikeda, K. D. Ashley & T-W. Chan (eds). Taiwan. Lecture Notes in Computer Science 4053, Springer-Verlag, 197-206.

[20] Beck, J.E., Mostow, J. (2008). How who should practice: Using learning decomposition to evaluate the efficacy of different types of practice for different types of students. Proceedings of the 9th International Conference on Intelligent Tutoring Systems, 353-362.

[21] Desmarais, M.C., Maluf, A., Liu, J. (1996) User-expertise modeling with empirically derived probabilistic implication networks. User Modeling and User-Adapted Interaction, 5, 3-4, 283-315.

[22] Romero, C., Ventura, S., Espejo, P.G., Hervas, C. (2008) Data Mining Algorithms to Classify Students. Proceedings of the First International Conference on Educational Data Mining, 8-17.

[23] Baker, R.S.J.d., Barnes, T., Beck, J.E. (Eds.) (2008) Proceedings of the 1st International Conference on Educational Data Mining.

[24] Strube, M.J. (1985) Combining and comparing significance levels from non-independent hypothesis tests. Psychological Bulletin, 97, 334-341.

[25] Amershi, S., Conati, C. (2006) Automatic Recognition of Learner Groups in Exploratory Learning Environments. Proceedings of ITS 2006, 8th International Conference on Intelligent Tutoring Systems.

[26] Scheines, R., Sprites, P., Glymour, C., Meek, C. (1994) Tetrad II: Tools for Discovery. Lawrence Erlbaum Associates: Hillsdale, NJ.

[27] Mostow, J. (2008). Experience from a Reading Tutor that listens: Evaluation purposes, excuses, and methods. In C. K. Kinzer & L. Verhoeven (Eds.), Interactive Literacy Education: Facilitating Literacy Environments through Technology, pp. 117-148. New York: Lawrence Erlbaum Associates, Taylor & Francis Group.

[28] Merceron, A., Yacef, K. (2008) Interestingness Measures for Association Rules in Educational Data.

ARPN Journal of Science and Technology

http://www.ejournalofscience.org

Proceedings of the First International Conference on Educational Data Mining, 57-66.

[29] Walonoski, J., Heffernan, N.T. (2006). Detection and Analysis of Off-Task Gaming Behavior in Intelligent Tutoring Systems. Proceedings of the 8th International Conference on Intelligent Tutoring Systems, 382-391.

[30] Baker, R.S.J.d., Beck, J.E., Berendt, B., Kroner, A., Menasalvas, E., Weibelzahl, S. (2007) Complete On-Line Proceedings of the Workshop on Data Mining for User Modeling, at the 11th International Conference on User Modeling (UM 2007).

[31] Hershkovitz, A., Nachmias, R. (2008) Developing a Log-Based Motivation Measuring Tool. Proceedings of the First International Conference on Educational Data Mining, 226-233.

[32] Baker, R.S.J.d., de Carvalho, A.M.J.A. (2008). Labeling Student Behavior Faster and More Precisely with Text Replays. Proceedings of the First International Conference on Educational Data Mining, 38-47.

[33] Shih, B., Koedinger, K.R., Scheines, R. (2008) A Response-Time Model for Bottom-Out Hints as Worked Examples. Proceedings of the First International Conference on Educational Data Mining, 117-126.

[34] Baker, R.S.J.d., Corbett, A.T., Roll, I., Koedinger, K.R. (2008) Developing a Generalizable Detector of When Students Game the System. User Modeling and User-Adapted Interaction,18, 3, 287-314.

[35] Cen, H., Koedinger, K., Junker, B. (2006) Learning Factors Analysis - A General Method for Cognitive Model Evaluation and Improvement. Proceedings of the 8th International Conference on Intelligent Tutoring Systems, 12-19.

[36] Y. B. Walters, and K. Soyibo, "An Analysis of High School Students' Performance on Five Integrated Science Process Skills", Research in Science & Technical Education, Vol. 19, No. 2, 2001, pp.133 – 145.

[37] Z. N. Khan, "Scholastic Achievement of Higher Secondary Students in Science Stream", Journal of Social Sciences, Vol. 1, No. 2, 2005, pp. 84-87.

[38] S. T. Hijazi, and R. S. M. M. Naqvi, "Factors Affecting Student's Performance: A Case of Private Colleges", Bangladesh e-Journal of Sociology, Vol. 3, No. 1, 2006.

[39] A. L. Kristjansson, I. G. Sigfusdottir, and J. P. Allegrante, "Health Behavior and Academic Achievement Among Adolescents: The Relative Contribution of Dietary Habits, Physical Activity, Body Mass Index, and Self-Esteem", Health Education & Behavior, (In Press).

[40] J. A. Moriana, F. Alos, R. Alcala, M. J. Pino, J. Herruzo, and R. Ruiz, "Extra Curricular Activities and Academic Performance in Secondary Students", Electronic Journal of Research in Educational Psychology,Vol. 4, No. 1, 2006, pp. 35-46.

[41] H. A. Camdeviren, A. C. Yazici, Z. Akkus, R. Bugdayci, and M. A. Sungur, "Comparison of Logistic Regression Model and Classification Tree: An Application to Postpartum Depression Data", Expert Systems with Applications, Vol. 32, No. 4, 2007, pp. 987-994.

[42] G. V. Kass, "An Exploratory Technique for Investigating Large Quantities of Categorical Data", Applied Statistic, Vol. 29, 1980, and pp. 119-127.

[43] I. H. Witten, and E. Frank, Data Mining – Practical Machine Learning Tools and Techniques (2nd ed.), San Francisco, CA: Morgan Kaufmann Publisher, 2005.

[44] H. Liu, and R. Setiono, "Chi-square: Feature Selection and Discretization of Numeric Attributes", Proceedings of IEEE 7th International Conference on Tools with Artificial Intelligence, Vol. 338, No. 391, 1995.

[45] S. Ganesh, "Data Mining: Should it be Included in the Statistics Curriculum?" The 6th international conference on teaching statistics (ICOTS-6), Cape Town, South Africa, 2002.

[46] K. Ramasamy, "Mother Tongue and Medium of Instruction – A Continuing Battle", Language in India, Vol. 1, No. 6