PREDICTING FACULTY DEVELOPMENT TRAININGS AND PERFORMANCE USING RULE-BASED CLASSIFICATION ALGORITHM

Roxanne A. Ancheta*, Rosmina Joy M. Cabauatan, Bartolome T. Tanguilig Lorena, W. Rabago
Department of Information Technology Education, Technological Institute of the Philippines, Quezon City, Philippines

ARTICLE INFO

Corresponding Author:
Roxanne A. Ancheta
Technological Institute of the Philippines Quezon City, Philippines,
Department of Information Technology Education.

Key Words: Data Mining, Classification, Prediction

INTRODUCTION

The explosive growth of available data as a result of computerization [1] of almost every aspect of the day-to-day operations of organizations has instinctive contributions to the development of intelligent decision making technologies. A young yet promising of this kind is data mining. It stands out due to its wide-array of techniques from the different domains such as statistics, artificial intelligence, machine learning, algorithms, database systems and visualization[1][18]. These influences served as groundwork for its applications to business for which the academic institution is exceptionally classified. Generally, regardless of discipline, data mining has gained popularity due to its tools with potentials to identify trends within data [18] and turn them out into knowledge [1] mostly with predictive attributes that could significantly lead to better and strong bases for decision making.

Widely-held academic data mining researches are student-centric with much emphasis on predicting student learning outcomes like in [3, 19, 20] and [21]. Subsequently, this was extended to its applications to human resource management. In [24] for instance, data mining is used to measure the quality of a teacher. In other institutions as in [2],[4],[5] and [6], employee performance have been studied using data mining techniques.

From the educational context, once hired, a teacher takes part of the academic institution’s goal to increase productivity and to provide excellent customer service. While these roles are expected to be executed in the entire pedagogical process, human resource managers play a part in the enhancement of their performance. Training and productivity development programs are administered to help augment teacher’s inherent capabilities to perform well.

As highlighted in [5], these tasks involve a lot of managerial decision, which are closely dependent on various factors such as human experience, knowledge, preference and judgment. In most academic institutions, newly-hired teacher trainings are facilitated to let them acquire the needed skills to manage the teaching and learning process but there are no clear classifications and associations of these trainings to the measures used to evaluate their performances. Universal classifications of measures used to evaluate teaching performance include personal characteristics, educational and professional attributes [6]. To maintain consistency and disdain from partiality, this paper preserves these essential attributes used to gauge teacher performance. With data mining functionalities such as classification and association, this paper yearns for the extraction of knowledge significant for predicting training and development needs of teachers so as to lead to their expected performance. The knowledge to be discovered could support human resource managers in deciding apt vital enhancement trainings for teachers to efficiently respond to performance evaluations and expectation.

The rest of the paper is organized as follows: the second consolidates the related studies in educational data mining; third provides the research problem and objectives; fourth are detailed discussions of the methodology and process of knowledge discovery; section 5 exhibits the simulation, analysis and results; and finally, the paper ends with a conclusion and an outlook for future directions.

RELATED WORKS

There have been quite a lot of studies of data mining in the educational domain. These concerned about students and employees' performances.
Tair, et. al [3] used data mining to improve graduate student’s performance, and overcome the problem of low grades of graduate students using association, classification, clustering and outlier detection rules.

Related to this is the study of Bhardwaj and Pal [19] in which a data model was used to predict student’s performance with emphasis on identifying the difference of high learners and slow learners using bayes classification.

Decision tree as a classification algorithm has been utilized in [21] to predict the final grade of a student in a particular course. The same algorithm has been applied in [23] on past student performance data to generate a model to predict student performance with highlights on identifying dropouts and students who need special attention and allow teachers to provide appropriate advising or counselling.

Conversely, Pandey and Pal [24] have considered the qualities the teacher must possess in order to determine how to tackle the problems arising in teaching, key points to be remembered while teaching and the amount of knowledge of the teaching process. In the course of identifying significant recommendations, John Dewey’s principle of bipolar and Reybem's tri-polar education systems have been used to establish a model to evaluate the teachership on the basis of student feedback using data mining.

While [2] used classification technique to build models to predict new applicant’s performance, [5], [8], [9] and [11] used the same to forecast employee’s talents.

Another technique called fuzzy has been applied in [10] to build a practical model for improving the efficiency and effectiveness of human resource management while [15] has improved and employed it to evaluate the performance of employees of commercial banks.

Other noteworthy researches that have added significant contributions to this study may be referred from [16] to [24].

While cited studies substantiated the applications of data mining in the educational domain, there is none that has applied data mining to predict the training and development needs of employees based on their inherent characteristics which could be initially mined from entry credentials such as resume and other supporting documents. Also, classifications as to what training needs are required to individual and a group of employees were not included. Instead, predictions of performance and talents have been stressed out. However, in harmony with these applications, this paper strives to build a model for predicting training and development needs that is parallel to areas focusing on student data. For this reason, the Human Resource Department was considered as the paper’s area of interest.

Management plays the role of ensuring this by closely adhering to the standards set by the higher management or by some heuristic needs of applicants with distinctive qualifications and potentials.

A. Business Understanding

With proper endorsement and approval of some school administrators, questions as to how data mining functionalities are best applied in Technological Institute of the Philippines has been identified. Recent studies had focused on student data. For this reason, the Human Resource Department was considered as the paper’s area of interest.

Along with the responsibilities of an academic institution to provide quality education to students is they examine on
how its human resources particularly the teachers perform in the realization of its goals as clearly defined by the so-called graduate attributes. Commissioned to carry out these duties is the Human Resource Department. With respect to the needs of each department, qualified applicants are selected. Once hired, employees potential are re-examined to see how well they can execute their duties. To evaluate their performance, standard evaluation tool of the institution is used. Apart from the development trainings and seminars administered for newly hired teachers such as start-of-semester orientation, integrity of the workplace, on-the-job training program, and ICT Skills Enhancement, there is no such tool used to link these trainings to entry profile of teachers as well as a device to make these parallel to the criteria of performance evaluation.

To digest then, the current state of administering these HR roles and policies, processes and evaluation tool were identified and reviewed. To substantiate these are actual data such as faculty profile, and performance evaluation criteria. These were used to predict the training needs of teachers as well as to forecast their performance to the significant areas of evaluation. To figure out the interrelationships among the attributes affecting teacher performance, Table 1 and 2 are exhibited.

Table 1. Newly-hired Faculty Profile and Performance Evaluation Criteria Matrix

<table>
<thead>
<tr>
<th>CLASS</th>
<th>Performance Evaluation Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Professional</td>
<td>Personal</td>
</tr>
<tr>
<td>Teaching</td>
<td></td>
</tr>
<tr>
<td>Research</td>
<td></td>
</tr>
<tr>
<td>Service</td>
<td></td>
</tr>
<tr>
<td>Personality</td>
<td></td>
</tr>
</tbody>
</table>

Table 2 shows the matrix of TIP's faculty performance evaluation criteria and its relationship to trainings and seminars. Both Table 1 and 2 provide the bases to distinguish the trainings and development needs appropriate to each kind of data set instance or tuple.

B. Data Understanding

To have an in-depth comprehension of the essential data needed for the realization of the objectives, the profile of newly-hired teachers, performance evaluation attributes, and training plans were gathered and cleaned so as to remove inconsistencies and noise unnecessary to generate reliable processing results. Since these data are stored in different directories of the HR database, these were collated to easily examine their intrinsic relationships to one another and identify their relevance to the objective. The result was used as the foundation of properly structuring the final data set. Table 3 exhibits how the data are organized into tuples representing each instance of the whole data set.

Table 3. Newly-hired Faculty Profile and Development Training Matrix

<table>
<thead>
<tr>
<th>CLASS</th>
<th>Performance Evaluation Criteria</th>
<th>Training and Seminars</th>
</tr>
</thead>
<tbody>
<tr>
<td>Professional</td>
<td>Personal</td>
<td></td>
</tr>
<tr>
<td>Teaching</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Research</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Service</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Personality</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

C. Data Preparation

Tuples structured using the template presented in Table 3 were discretized to fit in with the requirements of data classification. Table 4 shows the possible values of each attribute of the data set. For instance, original values for graduate and undergraduate courses were changed to vertically-aligned, aligned, and not-aligned. The process of discretization was performed to most attributes in order to avoid numerical values.
Table 4. Discretized Data Set

<table>
<thead>
<tr>
<th>Personal/</th>
<th>Organizational</th>
<th>Decisional</th>
<th>Class Label</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>7</td>
<td>9</td>
<td>2</td>
</tr>
</tbody>
</table>

With the priorities highlighted in Table 1 and the possible values of discretized data set in Table 4, class tuples of the data set were labeled based on their heuristic importance to the objectives. Class labels include professional training and personal training.

The Waikato Environment for Knowledge Analysis (WEKA) 3.60 serves as an intelligent tool for data analysis and predictive modeling. It is in this regard that aggregated data was transformed into Attribute-Relation File Format (ARFF), a WEKA readable format of data set. Figure 2 shows the WEKA Explorer with the loaded arff-formatted data set ready for analysis.

With WEKA’s wide collection of free analytical tools and data mining algorithms, it was chosen among others to be the primary tool for classifying and associating teacher attributes with training and performance. Generally, data mining classification technique is a two-step process consisting of learning step where classification model is constructed and classification step where the model is used to predict class labels for a given data [1].

In the first step, a classifier (commonly referred to as model) is built describing a predetermined set of data classes. It is in this stage where a classification algorithm builds the classifier by analyzing or learning from a training set made up of data set tuples and their associated class labels. Because the class label of each training tuple is provided, this step is known as supervised learning [1].

In the second step, test data are used to estimate the accuracy of the classification rules. If the accuracy is considered acceptable, the rules can be applied to the classification of new data tuples [1].

Various evaluation metrics for predictive accuracy of a classifier include hold-out and random sub-sampling, cross-validation and bootstrap methods. These methods share similar characteristics since they are all based on randomly sampled partitions of a given data but differ in terms of processes and techniques [1].

In hold-out method, the given data are randomly partitioned into two independent sets, a training set and a test set. Typically, two-thirds of the data are allocated to the training set, and the remaining one-third is allocated to the test set [1].

In cross-validation, the initial data are randomly partitioned into mutually exclusive subsets or folds, each of approximately equal size. Training and testing is performed several times depending on set folds. In the first iteration, a partition is reserved as test set and the remaining partitions are collectively used to train a model [1].

The bootstrap method samples the given training tuples uniformly with replacement. That is, each time a tuple is selected, it is equally likely to be selected again and re-added to the training set [1].

For simplicity, clarity and dependability, the hold-out method was chosen to be the method for selecting a training set to derive a model and a test set to assess the accuracy of the generated model. In Weka, this is performed using percentage split [28]. The process is simplified for illustrative purposes in Figure 3.

Classification and association have been chosen as the most appropriate data mining functionalities for training and performance predictions. The former is a form of data analysis that extracts models describing important data classes. Such models, called classifiers, predict categorically discrete, unordered class labels [1].

The basic classification techniques include decision tree, bayesian, and rule-based. In contrast to the first two techniques, rule-based classification learned model is represented as a set of IF-THEN rules. These rules are generated either from a decision tree or directly from the training data using sequential covering algorithm (SCA). An if-then rule is an expression of the form IF condition THEN conclusion in which the “IF” part (left side) is the rule antecedent or precondition and “THEN” part (right side) is the rule consequent [1].

Rule-based classification is also attributed to accuracy and coverage. These measures whether or not a rule antecedent are satisfied and the rule covers the data set tuples. From a class-labeled data set, coverage refers to the number of tuples covered by the rule while accuracy determines the tuples correctly classified by the rule. Significant in rule-quality measures that considers both
accuracy and coverage is the First-Order Inductive Learner (FOIL), a sequential covering algorithm that learns the first-order logic rules [1]. Definitions of these measures are provided in Figure 4.

At this point, to aggregate the classification tool's specific supports, this study used a rule-based classification technique using sequential covering algorithm (SCA) and hold-out method of partitioning sets of data. For rules quality, FOIL was chosen as it considers both coverage and accuracy. In WEKA, a cloned Ripper algorithm called Jrip [29] is designed to execute classification of data sets while simulating the process of sequential covering algorithm. Selected due to its capability to directly classify data set without having to base rules on a decision tree, Figure 5 provides evidence of its simplicity.

![Figure 4. Coverage and Accuracy Classification Measures](image)

![Figure 5. Sequential Covering Algorithm](image)

Strategically, rules in sequential covering algorithm are learned one at a time. Each time a rule is learned, the tuples covered by the rule are removed, and the process repeats on the remaining tuples. The process continues until the terminating condition is met, such as there are no more training tuples or the quality of a rule retuned is below a user-specified threshold. The Learn-One-Rule procedure finds the best rule for the current class, given the current set of training tuples. Rules are grown from general to specific manner. This begins from an empty rule then gradually keeps appending attribute test to it. Attribute test is added as a logical conjunct to the existing condition of the rule antecedent. For measures of rule quality, every time it considers an attribute test, it must check to see if appending such a test to the current rule's condition will result in an improved rule. These characteristics made this study to choose SCA as an algorithm to simulate the process of predicting training and development needs of newly-hired faculty members of TIP. Presented in Table 5 are the details of the result of the classification learning stage.

Based on the results of rule generation, the accuracy rate obtained is 87%. This is supported by the positive tuples (TP Rate) correctly labelled by the classifier of 87% and negative tuples (FP Rate) that incorrectly labelled as positive of 13%. Also, the rule classifies that 75 out of 90 newly-hired teachers need professional training while three (3) need personal or organizationally-related training. Heuristically, this proves that the classifier is reliable enough to be used for evaluating new set of data. This is then saved and be utilized for the next data set evaluation cycle, the actual classification stage.

![Table 5. Result of Model Generation (Learning Stage)](image)

To evaluate the classifier performance, accuracy of prediction is very much important. In order to comprehend whether or not the rule generated is accurate to evaluate a test set, basic terminologies and notations are hereby presented in Table 6.

![Table 6. Terminologies on Evaluating Classifier Performance](image)

By testing the classifier with a new set of data, obtained accuracy rate is 95%. Proofs to verify the accuracy rate include a TP rate of 96% for professional training and 89% for personal training. The confusion matrix also proves that most newly-hired teachers need professional training than personal training.

**E. Evaluation**

In contrast to the result of predicting training and development needs of newly-hired faculty members of TIP in which the generated rule predicts that for faculty...
members with the profile attributes such as aligned undergraduate and graduate courses even with both teaching and industry experiences need professional training would likely obtain low ratings in the faculty performance evaluation. Based from Table 1, the first two attributes have bearings to mastery of the subject matter, teaching competence, methods and strategies, and command of the language of instruction and with less impact on personal or institutional attributes.

Table 7. Result of Model Testing (Classification Stage)

<table>
<thead>
<tr>
<th>Scheme</th>
<th>1er classifier_rules_Reln</th>
<th>1er classifier_rules_Reln</th>
<th>1er classifier_rules_Reln</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.952</td>
<td>0.111</td>
<td>0.983</td>
<td>0.902</td>
</tr>
<tr>
<td>Number of Rules Generated</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>

F. Deployment

The result of this study intends not to replace the HR roles to design and plan the appropriate training needs of newly-hired faculty members but to provide scientific and sound bases for selecting more trainings associated to the identified weaknesses. Specifically, the rules generated as predictors for human resource development needs will be recommended as a supplementary tool for devising strategic plans for faculty empowerment programs.

CONCLUSION AND FUTURE WORK

Discovering knowledge based from previous learning makes data mining an efficient and effective tool to predict future occurrences which may affect decisions of current situations. In this study, predictions were made on the performance of newly-hired faculty members based on their current training needs. This concludes that sensitive issues like these would bring out implications as to how and what programs are needed to enhance the inherent potentials of faculty members. Moreover, this study introduces potential applications of data mining in the educational domain not just for human resource management and student-related concerns but for other academic and non-academic areas. This study further outlooks for applications of results to analyze enhancement programs for senior faculty members and to identify patterns affecting both teacher and student performance using other data mining techniques such as association rules.

REFERENCES

[1] Jiawei Han, Michelleline Kamber, & Jian Pei (2008). Data mining: concepts and techniques. Morgan Kaufmann publishers
[17] Yuan Qi and Rosalind W. Picard (2002). Context-sensitive Bayesian Classifiers and Application to Mouse Pressure Pattern Classification
[22] Surjeet Kumar Yadav, Brijesh Bharadwaj (2012). Data Mining Applications: A comparative Study for Predicting Student’s performance
[23] Umesh Kumar Pandey (2012). Mining Data to Find Adept Teachers in Dealing with Students. IJ. Intelligent Systems and Applications
[27] Data Mining - Classification II. http://www-staff.lboro.ac.uk/~comds2/coc131/COC131_tutorial_w7.pdf

Authors

Roxanne A. Ancheta graduated with a BS Information Technology degree from Technological Institute of the Philippines-Manila (TIP-Manila) in 2009. She obtained her Master in Information Technology from (TIP-QC). She became a full-time faculty member of TIP-QC College of Information Technology Education in November 2009.

Rosmina Joy M. Cabauatan is Assistant Professor of both Graduate School and College of Information Technology Education of Technological Institute of the Philippines. With her research interests in educational games, and applications of data mining in the computing and educational disciplines, she has been commended for his contributions to institutional researches. Pursuing her Doctoral Studies in Information Technology, she is currently undertaking her Dissertation on application of data mining in mobile computing.

Lorena W. Rabago graduated with a BS Computer Science degree from the Philippine Christian University in 1990. She obtained her Master in Information Technology from Technological University of the Philippines. She is also a graduate of Doctor of Technology at the same university. Currently, the Department Head of the Information Technology in Technological Institute of the Philippines, a member of the Technical Committee on Information System of the Commission on Higher Education. She is also the Vice-President of the Philippine Society of IT Educators – National Capital Region Chapter and an Accreditor for Information Technology Education Faculty, Instruction and Laboratories of the Philippines Association of the College and Universities of the Commission on Accreditation.

Bartolome T. Tanguilig III graduated with BS in Computer Engineering degree from Pamantasan ng Lungsod ng Maynila in 1991. He obtained his Master in Computer Science from De La Salle University in 1999. He is also a graduate of Doctor of Philosophy in Technology Management from Technological University of the Philippines in 2003. He is a founder of the Junior Philippine ITE Researchers. Currently, the Dean of the College of the Information Technology Education in Technological Institute of the Philippines, a member of the Technical Panel in Information Technology Education and the Chair of the Technical Committee on Information Technology of the Commission on Higher Education, national board member of the Philippine Society of IT Educators. He became an accreditor of Graduate Programs of the Philippine Association of Colleges and Universities of the Commission on Accreditation in 2010.