DEVELOPMENT OF TEACHING PERFORMANCE EVALUATION TOOL IN HIGHER EDUCATION USING ARTIFICIAL INTELLIGENT

WU JING

Thesis submitted in fulfilment of the requirements for the award of the degree of Master of Human Captical Resource

Faculty Of Manufacturing Engineering And Technology Management
Universiti Malaysia Pahang

July 2013

ABSTRACT

How to improve the teaching quality in higher education, has become the current focus of the work of education. However in universities , classroom teaching is the main channel for the implementation of education. Its quality at a large extent reflects and determine the quality of education in colleges and universities. The evaluation is key to improve teaching quality. So how to set up scientific justice evaluation of university classroom teaching quality system is very important problem.

Evaluation system includes three basic parts: evaluation indicators and weights, sources of information(the specific data of indicators) and the ways to deal with the information(models). We briefly discussed the evaluation indicators, weights and the impact of that to teacher evaluation. We mainly discussed and analyzed the way of data processing in the existent teacher evaluation system , and emphasized on utilizing a new way to solve the problems in traditional models.

Teacher evaluation is a highly non-linear relationship mixed with lots of qualitative and quantitative. But in the existing teacher evaluation system, its is linear models that are used mostly to deal with the information. First experts will set the specific indicators and the weights of each indicator, and then gain the final result of evaluation through the weighted average of data. Though the way is simple and easy to work, but the accuracy of the evaluation is not high, so generally it can just evaluate the quantitative indicators, and it's helpless to the qualitative or fuzzy indicators. In addition, the weights of the indicators, in the evaluation system artificially made, which will cause a big man-made effect on the evaluation process, that will result in some difference between the result of the evaluation and the actual situation. Though the fuzzy and analytic way that came out recently have solved problems on the qualitative and fuzzy indicators to a certain extent, but it has no much improvement in solving the excessively dependence of evaluation on subjective factors, and in reflecting the intrinsic relationship of indicators and the relationship between indicators and results, and on the accuracy of the results, which make the evaluation lose the objectivity and

scientific, and the reliability of the results is questionable.

This thesis attempts using the artificial neural networks method, appraises to the teaching performance evaluation. We analyzed the characters on structure, content and using means of kinds school's teacher evaluation through online search and surveys on the spot, considering lots of factors that can influence the results, putting forward improvement measures, establishing a BP net model to deal with information of teacher evaluation, and optimizing the model processing by utilizing strong functions of MATLAB toolbox. Finally 20 samples from a university in China, which are representative in indicators, had a emulate exercise and validated test. The result was analyzed. The data show that the model can objectively reflect the non-linear relationship between indicators and results, the results are accurate, the precision is high, and the result has a good agreement with the actual situation. According to the intrinsic relationship between indicator data and objectives based on network, all weights of indicator come out automatically, so it can better solve the problems of reliance of teacher evaluation on subjective factors, exclude the disturbance of personal factor, and advance the reliability of evaluation. All the results show that applying BP network to teacher evaluation is feasible and effective, and it will have a good prospect.

Keywords; teacher performance, teacher evaluation, BP neural network, MATLAB toolbox.

TABLE OF CONTENTS

CHAPTER 3 ARTIFICIAL NUORAL NETWORKS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AHP	Analytical Hierarchy Process
ANN	Artificial neural network
BP	Back Propagation
BPN	Back Propagation Network
SEEQ	The Students' Evaluation of Education Quality
LMBP	Levenberg-Marquardt back propagation
MLP	Multilayer Perceptron
MATLAB	matrix laboratory software

CHAPTER 1

INTRODUCTION

1.1 BACKGROUND

With the popularization of higher education and acceleration of college students, every year, colleges and universities are increasing scale. Expansion of the scale for colleges and universities, It broadens the scope for development, on the other hand, it also brings many problems, such as the issue of teaching quality, which is particularly prominent. And classroom teaching, is the key link throughout university teaching, but also is the core of teachers' work. Its teaching performance influences directly the quality of the whole university. Therefore, from the perspective of the educational value, using the scientific method to evaluate the quality of teaching comprehensively, reasonable and effectively, plays an important role in evaluating teaching level and teaching performance in university. However, because of their heavy workload, complicated statistical and other reason's evaluation of the teaching quality, often make a mere formality of the work, or simple enough system, or one-sided insufficient.

In universities，teachers performance assessment is vital for the implementation of education. Its quality at a large extent reflects and determines the quality of education in colleges and universities. The performance evaluation is a key to improve teaching qualifty (Craft, 1998).

Teaching is the synthesis of a dynamic process of teaching and learning, there are many factors affecting it, and they affect it in different degree, the result of evaluation is hardly to use the equal math's analytic expression to show. It belongs to complicate the non-linear sort problem. It has brought the very major difficulty for the quality synthetic evaluation（Bryophyte et al.,1986). In the past, many systems are direct evaluated by setting up a mathematical model, such as weighted average method, hierarchical analysis process, fuzzy comprehensive evaluation method as mentioned before. The above methods in the assessment process require influencing factors (evaluation index) with a linear relationship. Thus, it is difficult to rule out a variety of stochastic and subjective evaluation results which are easily bring about many distortions and bias. Therefore, it is necessary to seek a new scientific assessment method.

The higher educational teaching performance evaluation is a complex problem. The teaching performance includes teaching condition, class difficulty, class teaching, learning effectiveness and other factors. These factors affect each other. At the same time, the relation between teacher and student also influence teaching performance(Chen, 1987). For now, there is no one teaching performance evaluation system can be consider perfect. From present research, it focuses on three aspects.

1.1.1 Research on Subject of Assessment

There are many methods and accesses evaluate teaching performance, such as a teacher evaluate by himself, confrere evaluation, leader evaluation, inspector group evaluation and student evaluation. Because of different roles have a different function, each evaluation method and result are only a part of teaching performance evaluation. It is not entire of teaching performance (Henkel, 2000). Due to teachers and evaluation, the evaluation at large not only spends much time and energy but also hard to practice because of the bias of people's relationship and unfamiliar with teaching process are a

difficulty to eliminate. So the method of the subject of student has to adapt widely by most of the higher educational organizations. From the end of 1990s, in China, a lot of universities developed the method which student evaluating teaching performance. The teaching performance has been promoted at that time.

1.1.2 Research on Teaching Performance Assessment System

The content is very widely of teaching performance assessment; there are two representative assessment indicator systems: one is class teaching quality indicator system by Steven M. Kimball, (2002). main viewpoint is distinguished teaching quality as two parts, the media indicator of a teaching process and the ultimate indicator of teaching effectiveness.

Another is the teaching evaluation indicator by an educationist named Babansky's from Soviet Russia; Babansky has built the indicator context from nine aspects as following:

1) Comprehend related subject
2) The skill of evaluating knowledge capability and skill
3) The skill of setting up work plan
4) The skill of complete the plan effectively
5) The skill of producing interesting in related subjects
6) Realize the relationship among subjects
7) Treat different student different ways
8) Producing student's study skill and capability
9) Understanding the principle of educational psychology for student

Babansky has present the basic outline of researching teaching, and also has established 4 grade evaluation criterion, it was maneuverability.

1.1. 3 The Methods of Teaching Quality

Through the hard effort by a lot of researchers, the educational evaluation has formed a mature theory system. There are some main evaluation methods, such as: weighted average method, Analytical Hierarchy Process (AHP), Fuzzy Comprehensive Evaluation and Artifical Neural Network (ANN) .

1) Weighted Average Method

Based on Weighted average method in educational evaluation, the teaching management department in university established each evaluation indicator and it's weight based on the importance of indicators, then students evaluate teaching performance by questionnaire or Internet, collect the data, and get all teachers' score with computer system calculation, finally gained the degree with the score.

2) Analytical Hierarchy Process (AHP)

The AHP is a structured technique for dealing with complex decisions. Rather than prescribing a "correct" decision, the AHP helps decision makers find one that best suits, their goal and their understanding of the problem—it is a process of organizing decisions that people are already dealing with, but trying to do in their heads.

Based on mathematics and psychology, Thomas (1970) developed the AHP and has been extensively studied and refined since then. It provides a comprehensive and rational framework for structuring a decision problem, for representing and quantifying its elements, for relating those elements to overall goals, and for evaluating alternative solutions. It is used around the world in a wide variety of decision situations, in fields such as government, business, industry, healthcare, and education.

Several firms supply computer software to assist in using the process. Users of

the AHP first decompose their decision problem into a hierarchy of more easily comprehended sub-problems, each of which can be analyzed independently. The elements of the hierarchy can relate to any aspect of the decision problem—tangible or intangible, carefully measured or roughly estimated, well- or poorly-understood—anything at all that applies to the decision at hand.

Once the hierarchy is built, the decision makers systematically evaluate its various elements by comparing them to one another two at a time, with respect to their impact on an element above them in the hierarchy. In making the comparisons, the decision makers can use concrete data about the elements, or they can use their judgments about the elements' relative meaning and importance (Simenon et al., 1950). It is the essence of the AHP that human judgments, and not just the underlying information, can be used in performing the evaluations.

The AHP converts these evaluations to numerical values that can be processed and compared over the entire range of the problem. A numerical weight or priority is derived for each element of the hierarchy, allowing diverse and often incommensurable elements to be compared to one another in a rational and consistent way. This capability distinguishes the AHP from other decision making techniques.

In the final step of the process, numerical priorities are calculated for each of the decision alternatives. These numbers represent the alternatives' relative ability to achieve the decision goal, so they allow a straightforward consideration of the various courses of action.

3) Fuzzy Comprehensive Evaluation

Fuzzy comprehensive evaluation based on fuzzy mathematics, applied the principle of synthesis of fuzzy relations with some of the ill-defined and difficult to

quantify the quantitative factors, from a number of factors being evaluated under the hierarchy of things to undertake a comprehensive assessment of the situation.

4) Artificial neural network

An Artificial Neural Network (ANN), usually called Neural Network (NN), is a mathematical model or computational model that is inspired by the structure and/or functional aspects of biological neural networks. A neural network consists of an interconnected group of artificial neurons, and it processes information using a connectionist approach to computation. In most cases an ANN is an adaptive system that changes its structure based on external or internal information that flows through the network during the learning phase. Modern neural networks are non-linear statistical data modeling tools. They are usually used to model complex relationships between inputs and outputs or to find patterns in data.

1.2 PROBLEM STATEMENT

From above research, understanding of teaching quality evaluation situation at present, it included who evaluate, evaluation indicator and evaluation method. Who evaluate? There is a teacher evaluate by himself. Leader evaluates experts group evaluates and student evaluated teacher (Daniel et al., 2000). The evaluation indicator selection, usually using the general high educational evaluation indicator, face to various universities, also different character, but the same evaluation system has been used. It is irrationality. The analysis following is about the common evaluation methods have mentioned above:

1.2.1 The Indicator Weighted Average

Traditional indicators weighted average method, is teaching by the school administration to develop the evaluation of various indicators, and the importance of

individual indicators according to the weight of each indicator set, and organize the students through the online questionnaire or evaluation, to obtain data then calculated by the computer system to evaluate the object of all the scores, then scores to determine the grade.

The method even though the data transaction process is simpler, but this method has defined the simple linear relation among the indicator by people, and the weight of each evaluation factor according to experience, apparently it cannot prove it has the possibility of linear increasing among the evaluation indicator, also cannot prove the irrationality of the weight.

Target weight is a number of indicators; it should be objective and reflect the importance and role of indicators of size. Important indicator that the weight of greater value, rather than less important indicator of the weight, therefore, the index weight also has oriented, in practice. People are always the main focus on important indicators of the important indicators of the requirements and standards to give more attention to, and for secondary indicators are often paid less attention, sometimes even ignored. If the index for the same level of equal treatment, does not distinguish between the importance and role of the size, it runs counter to objectivity, it cannot reflect the reality, the evaluation results will be distorted.

1.2.2 The Analytic Hierarchy Process (AHP)

Analytic hierarchy process (AHP) basic principle: The AHP law the first question hierarchical, decomposes according to the question nature and the general objective this question the order the level, invites the expert again carries on a more objective judgment after each level's various factors, correspondingly gives the relative important quota expression; Then the establishment mathematical model, calculates each level complete factor the relative important weight, and sorts; Finally carries on the

plan decision-making and the choice according to the sorting result solves the question measure.

The method can gain the evaluation indicator weight, but it also require the linear relation among the indicator , otherwise it would influence the effectiveness of the AHP model .Then which judgment matrix is real better frame when diverse matrix have been given by experts , AHP model cannot answer this question.

1.2.3 Fuzzy Comprehensive Evaluation

The fuzzy mathematics is the research and processing fuzziness phenomenon mathematics. It is by American cybernetics expert establishment, the so-called fuzzy mathematics multi-level synthesis judgment's principles. This is needed to judge the identical thing first many kinds of factors, divides into certain factors according to some attribute, then carries on at the beginning of the level synthesis judgment to some big factor, based on this carries on the top-level synthesis judgment again to at the beginning of a level synthesis judgment's results.

The fuzzy comprehensive judgment is the use mathematics method, draws the teacher classroom quality rating conclusion through the mathematics synthesis judgment model one method. Its success application, the key lays in the correct stipulation evaluation the sets of sub factors and the reasonable structure fuzzy evaluation matrix. The method can obtain the value level or priority of the evaluation object, but this method requires establishing a judgment matrix of a suitable object. It would result in the diverse matrix by different expert, at last getting the inconsistent evaluation result.

These three methods all require the linear relation among the factors (indicator). However teaching synthesis of the dynamic process of teaching and learning, there are

many factors affecting it, and they affect it in different degree, the result of evaluation is hardly to use the equal math's analytic expression to show. It belongs to complicate a non-linear sort problem. It has brought the very major difficulty for the quality synthetic evaluation. So, three methods above all are irrationality for teaching performance evaluation.

1.3 Artificial Neural Network (ANN)

ANN as a new technique can approach adequately to the random complicated nonlinear relation, and it can establish a model without knowing the reason of data production. Back Propagation (BP) neural network through the training according to the exit's knowledge (learning sample data) to obtain the value model of an object, it can solve the nonlinear comprehensive problem and decrease the infection of the decision results by people. However the different neural network method can influence the evaluation result highly, so how to choose? Which method is suitable for teaching performance evaluation? Because of standard BP method has some limitation such as: the speed of BP training very slowly and hard to master the training. It is easy to get into the partial infinitesimal points and hard to go away from; exist network paralysis. So if using BP to implement net training, it is slow down the speed of convergence, even not convergence result. So this thesis will be using improved BP too suitable to teaching performance evaluation.

Be aimed at the characteristics of teacher's performance evaluation and the existent problems in the universities，the research puts forward an Artificial Neural Network(ANN) technology for system performance evaluation. Artificial Neural Network technology has been used in the many field's performance evaluations. The basic characteristic of ANN is non-linear mapping, learning classifier and real time optimization. Neural networks have been the broad application in industrial control, classification, forecasting. Data mining area's etc. Artificial neural network also has

better fault tolerance, filter noise and characteristics of online application.

1.4 RESEARCH OBJECTIVES

The existing assessment method is mostly quite unitary, has not formed the complete system; The weight target and the appraisal goal are linear relationships, is not genuine representative the actual relations; The quality of teaching is the complex quality synthetic evaluation question, is unable to carry on the expression with the simple mathematical expression. How to choose the weight target, how to determine that the appraisal relations mathematical expression, how does carry on the proof technique, whether is suitable is our final goal. So more specific objectives are following:

i. To identify the flaws of traditional performance assessment;
ii. To establish a new mathematical evaluation model by utilizing artificial intelligent;
iii. To prove the proposed model is effective in evaluating teaching performance.

This study plans to use the BP network theory establishment teaching performance evaluation system. It first formulates the science reasonable evaluating indicator system, then the determination model, the choice algorithm carries on the training to the sampled data, obtains the measure results, and carries on the confirmation again to the confirmation data. Based on the following questions, we want to confirm the above method the feasibility, whether to establish:

i. What are the problems of the current performance assessment?
ii. What mathematical evaluation model can be proposed by utilizing artificial intelligent?
iii. Does the proposed model is effective in evaluating teaching performance?

1.5 THESIS OUTLINE

The chapter one discussed several kinds of common teaching performance emphatically the evaluation method, the comparative analysis each method well and bad points. It proposed the present existence's question and the goal, how and brief elaboration did achieve the goal the simple step.

The chapter two reviews teacher evaluation history so as to find research target and methodology.

The chapter three narrated the ANN historical development and several kind of different network type briefly, introduced in detail based on the BP network concrete algorithm and the flow chat.

The chapter four further introduced the ANN algorithm and has explained with examples the ANN algorithm concrete application through a sub-system, uses the LMBP algorithm the optimized method, has proven its feasibility with the mathematical method.

The chapter five establishes teaching performance evaluation the system procedure. Though it compares three kinds of different function, the use collected data, carries on the training to the network, then carries on the confirmation, through the above result truncation chart, had proven carries on the quality of teaching appraisal using ANN is completely feasible, although has the place which some needs to improve.

The chapter six make conclusion and recommdations in the future.

CHAPTER 2

LITERATURE REVIEW

2.1 INTRODUCTION

The development of teacher evaluation along with the education evaluation development. Take the US as the representative, generally thought that the education evaluation has experienced the seed time, the formative year, the period of expansion and the mature period four stages roughly. The teacher evaluation is educates the evaluation an important component. Its development also experienced the germination period, to form the time and the development time approximately, each time's teacher appraised has the bright theory background and the practice characteristic.

American education psychologist Thorndike (1904) has published "Mental test Law", introduced systematically the statistical method and the establishment survey's basic principle, and proposed “every existence's thing has a quantity thing measurable quantity” the famous judgment, worked the survey objectification for the education survey and the teacher, the standardization to lay the rational.

Donald et al.,(1984) describe succinctly the modern history of formal teacher evaluation–that period from the turn of the twentieth century to about 1980. This history might be divided into three overlapping periods: (1) The Search for Great Teachers; (2)

Inferring Teacher Quality from Student Learning; and (3) Examining Teaching Performance. At the beginning of the twenty-first century, teacher evaluation appears to be entering a new phase of disequilibrium; that is, a transition to a period of Evaluating Teaching as Professional Behavior.

The Search for Great Teachers began, in early, in 1896 with the report of a study conducted by Robert et al. (2002) asked 2411 students from the second through the eighth grades in Sioux City, Iowa, to describe the characteristics of their best teachers. He thought that by making desirable characteristics explicit he could establish a benchmark against which all teachers might be judged. Some 87 percent of those young Iowans mentioned, "helpfulness" as the most important teacher characteristic. However a stunning 58 percent mentioned, "personal appearance" as the next most influential factor.

Some research compendium of research on teaching competence noted that supervisors' ratings of teachers were the metric of choice. A few researchers, however, examined average gains in student achievement for Inferring Teacher Quality from Student Learning. They assumed, for a good reason, that supervisors' opinions of teachers revealed little or nothing about student learning. Indeed, according to Medley and his colleagues, these early findings were "most discouraging." The average correlation between teacher characteristics and student learning, as measured often by achievement tests, was zero. Some characteristics related positively to student achievement gains in one study and negatively in another study. Simeon, et al., (1950) reviewed more than 1,000 studies of teacher characteristics, defined in nearly every way imaginable, and found no clear direction for evaluators. Jacob and Philip, (1963) called once and for all for an end to research, and evaluation aimed at linking teacher characteristics to student learning, arguing it was an idea without merit Li et al (2009).

Medley et al., (1984) note several reasons for the failure of early efforts to judge

teachers by student outcomes. First, student achievement varied, and relying on average measures of achievement masked differences. Secondary, researchers failed to control for the regression effect in student achievement–extreme high and low scores automatically regress toward the mean in second administrations of tests. Third, achievement tests were, for a variety of reasons, poor measures of student success. Perhaps most important, as the researchers who ushered in the period of Examining Teaching Performance were to suggest, these early approaches were conceptually inadequate, and even misleading (Barr, 1984). Student learning as measured by standardized achievement tests simply did not depend on a teacher's education, intelligence, gender, age, personality, attitudes, or any other attribute. What mattered was how teachers behaved when they were in classrooms.

The period of Examining Teaching Performance abandoned efforts to identify desirable teacher characteristics and concentrated instead on identifying effective teaching behaviors; that is, those behaviors that were linked to student learning. The tack was to describe clearly and precisely teaching behaviors and relate them to student learning–as measured most often by standardized achievement test scores. In rare instances, researchers conducted experiments for arguing that certain teaching behaviors actually caused student learning. Like Kratz(1985) a century earlier, these investigators assumed that "principles of effective teaching" would serve as new and improved benchmarks for guiding both the evaluation and education of teachers. Brophy and Thomas (1986) produced the most conceptually elaborate, while Marjorie and Joseph, (1984) extensive bibliography of research done from 1965 to 1980 is a useful reference.

2.2 GOALS OF TEACHER EVALUATION

Although there are multiple goals of teacher evaluation, they are perhaps most often described as either formative or summative in nature. Formative evaluation

consists of evaluation practices meant to shape, compose, or improve teachers' performances. Clinical supervisors observe teachers, collect data on teaching behavior, organize these data, and share the results in conferences with the teachers observed. The supervisors' intent is to help teachers improve their practice. In contrast, summative evaluation, as the term implies, has as its aim the development and use of data to inform summary judgments of teachers. A principal observes teachers in action, works with them on committees, and examines their students' work, talks with parents, and the like. These actions, aimed at least in part at obtaining evaluative information about teachers' work, inform the principal's decision to recommend teachers either for continuing a teacher's contract or for termination of employment. Decisions about initial licensure, hiring, promoting, rewarding, and terminating are examples of the class of summative evaluation decisions.

The goals of summative and formative evaluation may not be as different as they appear at first glance. If an evaluator is examining teachers collectively in a school system, some summary judgments of individuals might be considered formative in terms of improving the teaching staff as a whole. For instance, the summative decision to add a single strong teacher to a group of other strong teachers results in improving the capacity and value of the whole staff.

In a slightly different way, individual performance and group, performance affects discussions of merit and worth. *Merit* deals with the notion of how a single teacher measures up on some scale of desirable characteristics. Does the person exhibit motivating behavior in the classroom? Does she take advantage of opportunities to continue professional development? Do her students do well on standardized achievement tests? If the answers to these types of questions are "yes,", then the teacher might be said to be "meritorious." Assume, for a moment that the same teacher is one of six members of a high school social studies team in a rural school district. Assume also that one of the two physics teachers just quit, the special-education population is