APPLICATION OF FUZZY THEORY FOR ADAPTIVE LEARNING DIAGNOSIS SYSTEM

Yu-Sheng Yang¹, Ching-Yi Huang², Yueh-Min Huang¹, Chiu-Sing Yang²

¹Department of Engineering Science, National Cheng Kung University, No.1, University Road, Tainan City 701, Taiwan
florencey1018@gmail.com, huang@mail.ncku.edu.tw

²Department of Electrical Engineering, National Cheng Kung University, No.1, University Road, Tainan City 701, Taiwan
q3696116@mail.ncku.edu.tw, csyang@ee.ncku.edu.tw

Abstract: This paper presents an application of Fuzzy Theory for Adaptive Learning Diagnosis System (FADS), which consists of five parts: calculating the concept similarity of different chapters; producing the personal adaptive reading mark automatically; calculating the learning degree of each chapter; constructing the fuzzy membership function; and evaluating the adaptive items selection mechanism. The system uses the similar concept method to compute the association between test items and teaching materials, and depends on a learner’s practice and test results when selecting and marking the important paragraphs from readings by bolding the text automatically. Each learner has personal adaptive marks on their teaching materials, used to identify the weak parts. Additionally, the system uses the associations and results from practices to compute the personal learning degree of each learner. FADS focus on the functions of a review and detailed explanation of answers, providing learners with diverse feedback and creating a personal self-study environment for each learner. The system transfers the difficulty of each chapter by using the fuzzy membership function to select appropriate items for the learner.

A total of 200 fourth-grade students in six social studies classes participated. The learners were divided into three groups, two classes per group; the groups were the experiment group, the comparison group, and the control group. The experimental method was quasi-experimental. The experiment group used FADS for learning diagnosis and assisted learning; the comparison group used traditional error rates for learning diagnosis and assisted learning; and the control group did not use a system for learning diagnosis and assisted learning. The results indicate that the use of FADS to diagnose and assist learning enabled the students in the experimental group to perform better than the two other groups. Since FADS makes online learning diagnosis analysis more effective and user-friendly, it can be an adaptive and flexible learning diagnosis system.

Keywords: adaptive, fuzzy theory, assessment system, self-study, on-line reading

1. Introduction
Assessment plays an important role in teaching and learning activities. It can document not only what students know and can do but it also influences learning. There are diverse methods for assessment in addition to written tests, including oral questioning, performance tasks, student self-assessment, and/or interaction and communication between teachers and students [5]. However, in traditional educational environments, most
teachers only use test scores to distinguish between students’ academic achievements; this has discouraged some students from learning.

Researchers suggest that adaptive teaching can create effective learning. Oppermann (1994) presents that systems able to adjust to the users automatically based on the system’s assumptions about user’s needs are called adaptive. In recent years, the application of adaptability has become a mainstream trend in computer-assisted learning, which can enhance testing efficiency, decrease testing time, lower testing errors, allow for diverse test questions, and the acquisition of test-related information.

In real life, most questions entail uncertainty, and such things that cannot be definitively expressed using numbers or models must be described using a language that expresses degrees. Recently, many scholars have applied the fuzzy theory to learning diagnosis evaluation, which attempts to improve upon traditional learning evaluations that are based only on an interpretation of scores. The aim of this study is to use the fuzzy theory as a basis to construct an adaptive online assisted learning system to improve upon the shortcomings of traditional evaluations.

2. Literature Review
In this paper, the related literature and research applied are explored as follows:

2.1 Assessment System
Assessment systems are significant sources of feedback information in the teaching process, the purpose of which is to assess whether teaching activities have achieved the teaching objectives, and to understand whether the learners have improved in terms of cognition, emotions, and skills after engaging in teaching activities.

2.2 Fuzzy Theory
Zadeh first introduced the Fuzzy set theory in 1965 as a way of handling imprecise data, and now this theory has many fields of application in information processing schemes. There are two methods of determining fuzzy membership functions. The first is to use human expert knowledge from the field, and the other is to use data collected from different detectors to determine membership functions. Wang (1997) stated that fuzzy systems are generally comprised of four parts: fuzzy rule database, fuzzy inference engine, fuzzification, and defuzzification.

2.3 Application of Fuzzy Theory on Assessment Systems
Recently, many methods have been presented for applying the fuzzy set theory in assessment systems for enabling the efficiency. Bai and Chen (2008) presented a new method for dealing with students’ learning achievement evaluation using fuzzy membership functions and fuzzy rules. Lin (2008) used Item Response Theory (IRT) and S-P figures as the input of fuzzy rule models, and established Fuzzy structural modeling in order to individualize knowledge concept analysis for learners, and provide guidelines for remedial learning.

2.4 Similarity Computing
In a Chinese document, sentences are formed by various terms. Before analyzing sentences, it is first necessary to segment the phrases. The purpose of segmentation is to
turn Chinese sentences into several meaningful phrases and terms to expedite further analysis. The Academia Sinica Chinese Knowledge and Information Processing System contains a highly diverse corpus, which provides an XML format feedback of segmentation results; the accuracy of which can reach over 95 percent.

2.4.1 TF&IDF
Term Frequency (TF) is the frequency of phrases appearing in a text. It originated in experiments by Luhn (1958) on data searches, which found that without the high and low frequency words, the medium frequency words tend to be more meaningful. Inverse Document Frequency (IDF) refers to whether a phrase is prevalent in various documents. Therefore, the weight of keywords tends to be calculated as a product of TF and IDF (TF×IDF), and the weight of sentences is the sum of all the important keywords that appear in the sentence.

2.4.2 Intersection TF&IDF
From the perspective of set theory, if two sets have more overlaps, it means that the two sets are more similar, and vice versa. Thus, the correlation between two documents is determined by calculating the weights of the key phrases in each document as in the formula above, taking the overlapping phrases of the two documents, and the sum of weights for the overlaps represents the extent of the correlation.

3. System Design
FADS consists of five parts, including calculating concept similarity of different chapters, producing the personal adaptive reading mark automatically, calculating learning degree of each chapter, constructing the fuzzy membership function, and adaptive items selection mechanism. The related parameters were defined as follows:

\[
M = \{m_i \mid \text{for each } i, 1 \leq i \leq n\} : \text{The set of instructional materials all chapters}
\]

\[
Q = \{q_i \mid \text{for each } i, 1 \leq i \leq m\} : \text{The set of test items}
\]

\[
S = \{s_i \mid \text{for each } i, 1 \leq i \leq k\} : \text{The set of tested students}
\]

\[
TM = \{tm_i \mid \text{for each } i, 1 \leq i \leq n\} : \text{The set of instructional material keywords}
\]

\[
TQ = \{tq_i \mid \text{for each } i, 1 \leq i \leq m\} : \text{The set of test item keywords}
\]

\[
TS_i = \{ts_{ij} \mid \text{for each } i, j, 1 \leq i \leq n, 1 \leq j \leq s\} : \text{The set of instructional material paragraph keywords}
\]

\[
W\hat{Q}_i = \{wq_{ij} \mid \text{for each } i, j, 1 \leq i \leq m, 1 \leq j \leq n\} : \text{The weight vectors of test items}
\]

\[
\sum_{j=1}^{n} wq_{ij} = 1, \text{where } i = 1 \sim m
\]

\[
A = \{qs_{ij} \mid \text{for each } i, j, 1 \leq i \leq k, 1 \leq j \leq m\} : \text{The matrix of students' answers}
\]

\[
A = (qs_{ij})_{k \times m}, \text{where } qs_{ij} = \begin{cases} 0, & \text{i \_ answer } q_j \text{ incorrectly} \\ 1, & \text{i \_ answer } q_j \text{ correctly} \end{cases}
\]

\[
FQ = \{fq_i \mid \text{for each } i, 1 \leq i \leq m\} : \text{The paragraph to which test questions belong}
\]

\[
\text{Sim}(T_x, T_y) : \text{The similarity function for } T_x \text{ and } T_y
\]
Let $SD_i = \{D_j\}$ for each $i$, $1 \leq i \leq k$, $1 \leq j \leq n$: The student's learning degree of chapters. Of which, $SD_i = \{D_j\}_{k \times n}$, where $D_{ij} \in [0, 1]$

$SY_{ij} = \{Y_{ij}\}$ for each $i$, $1 \leq i \leq k$, $1 \leq j \leq n$, $t \in \{L, M, H\}$: The membership value of students for each chapter.

$Q_{Next}$: The total number of items in re-test test.

### 3.1 Calculating Concept Similarity

This experiment uses social studies as an example. For each semester, the course contents can be divided into several units, which include several chapters. In the existing question bank, each question usually belongs only to a certain paragraph; however, it is found that in the same unit, each paragraph is related; in the same unit, paragraphs may be conceptually related. Thus, if answers are incorrect for a certain paragraph, it not only means unfamiliarity with the paragraph, but may also mean incorrect concepts from other paragraphs.

In this paper, we introduce the method of calculating concept similarity, which first adjusts the concept similarity of the test questions and the chapters. Each question and item selection first undergoes segmentation through the Academia Sinica Chinese Knowledge and Information Processing System, and the redundant phrases are deleted. Each question is broken up into several terms, which are referred to as test keywords. The instructional materials in each chapter undergo the same process, and the results are referred to as instructional material keywords.

### 3.2 Adaptive Reading Mark

This study digitizes instructional materials as a supplementary tool to help each student review after testing. Adaptive reading marks are automatically generated in the online reading area of each learner. The calculation of adaptive reading marks is as follows:

**Input**: $TS_i, TQ$  
**Output**: $QG_i$

For $i = 1$ to $m$: \[ r = fq_i; \]

For $j = 1$ to $s$: \[ SG_j = Sim(tq_i, ts_{ij}); \]

$QG_i = MAX(SG)$;

After each question is calculated using adaptive reading marks, it is possible to find paragraph $QG_i$, which is closest to the test question.

### 3.3 Calculating the Learning of Paragraphs

After calculating the similarity between test questions and each paragraph or section, each question has its own weight ratio in the paragraph or section. The answers and the respective weight ratios are used to calculate the extent of learning of the chapters through the following formula:
Based on the number of chapters and sections in each unit, each section will calculate a value for the extent of learning for each learner; if there are three sections, then each learner would have three section extensions of learning values.

3.4 Construction of Fuzzy Membership Functions

Since students with different or similar grades do not necessarily have similar cognitive ability, and re-testing can have different difficulty levels, this study finds that a transfer formula is needed to help automate the question selection mechanism. The fuzzy membership function is used to construct the transfer function, as follows:

\[
Y_L = \begin{cases} 
1 & 0 \leq D_j \leq X - \sigma \\
\frac{X - D_j}{\sigma} & X - \sigma < D_j \leq X \\
0 & D_j > X 
\end{cases}
\]

\[
Y_M = \begin{cases} 
0 & 0 \leq D_j \leq X - \sigma \\
\frac{D_j - (X - \sigma)}{\sigma} & X - \sigma < D_j \leq X \\
\frac{(X + \sigma) - D_j}{\sigma} & X < D_j \leq X + \sigma \\
0 & D_j > X + \sigma 
\end{cases}
\]

\[
Y_H = \begin{cases} 
0 & D_j \leq X \\
\frac{D_j - X}{\sigma} & X < D_j \leq X + \sigma \\
1 & D_j > X + \sigma 
\end{cases}
\]

where, \( Y_L \) is the ratio of easy questions, \( Y_M \) is the ratio of medium-level questions, \( Y_H \) is the ratio of difficult questions, \( D_j \) is the extent of learning in paragraph \( j \), \( X \) is the percentage ratio of the total average, and \( \sigma \) is the percentage ratio of the standard deviation.

3.5 Adaptive Question Selection Mechanisms

FADS automatically calculates the learning degree of each chapter by learners; students with the same grades may have different learning degree of each chapter.

\[
CP_j = \frac{\sum_{j=1}^{n} D_j - D_j}{(n-1) \sum_{j=1}^{n} D_j}, \text{where } 1 \leq j \leq n
\]

\[
C_{jL} = Q_{\text{Need}} \times CP_j \times Y_L, \text{where } 1 \leq j \leq n
\]

\[
C_{jM} = Q_{\text{Need}} \times CP_j \times Y_M, \text{where } 1 \leq j \leq n
\]
\[ C_{jH} = Q_{Need} \times CP_j \times Y_j, \text{where} 1 \leq j \leq n \]

Through the formula above, the system can automatically generate adaptive question distribution ratios, the question distribution in each chapter, and the ratio of difficulty after learners undergo evaluation. Of which, \( CP_j \) is the question distribution ratio of chapter \( j \), \( C_{jL} \) is the distribution of easy questions in chapter \( j \), \( C_{jM} \) is the distribution of medium-level questions in chapter \( j \), and \( C_{jH} \) is the distribution of difficult questions in chapter \( j \).

4. System Implementation

This study uses the HGLS digital learning platform, which is applicable for various universities, as well as junior high and elementary schools. The HGLS system includes several sub-systems, such as a question creation system, a question dispensation system, a testing system, and a learning feedback mechanism. The FADS focuses on the functions of the learning feedback part, and provides learners with diverse learning feedback and assisted learning environments. These are explained as follows:

4.1 Answers Feedback

After learners complete the online tests, the system automatically calculates the evaluation results, thus, learners can view their own tests and detailed explanations, which allows them to understand their learning conditions.

4.2 Online Reading

The online instructional materials reading area is the interface in this system that provides learners the ability to reread instructional materials. After using the adaptive reading marks, the text portion would include adaptive marks for each student to note areas they are unfamiliar with, reminding them that when they read these texts, they should take notes, and think about the content more than one time.

4.3 Self-study

In self practice, the system offers learners an environment to learn on their own. After official evaluation, the learner can use the above adaptive question selection mechanism. The system automatically establishes a self-test practice sheet adapted to learners, allowing the learners to practice again.

5. Experiment Design

The experiment was conducted on 200 fourth grade elementary school students from six classes, and all students were taught by the same teacher. The experimental target was Unit 2 of the Social Studies for the second semester of the fourth grade.

The experimental method was quasi-experimental. The students were divided into three groups, two classes per group; the groups were the experiment group, the comparison group, and the control group. The experiment group used FADS for learning diagnosis and assisted learning, the comparison group used traditional error rates for learning diagnosis and assisted learning, and the control group did not use the system for learning diagnosis and assisted learning.

5.1 Reliability and Validity

This experimental study used the reliability and validity testing evaluation methods by Wang (2000) Education Testing and Evaluation, which is explained as follows:
5.1.1 Reliability
Reliability refers to the accuracy or precision of the measurement tool, or the extent of consistency of evaluation results. In reliability testing, a high \( \alpha \) value means that the evaluation is reliable. Statistically, the \( \alpha \) value must exceed 0.7 to be reliable; reliability is calculated as follows:

\[
\alpha = \left( \frac{m}{m-1} \right) \times \left( 1 - \frac{\sum_{i=1}^{m} \sigma_i^2}{\sigma^2} \right)
\]

where, \( m \) is the number of test questions, \( \sigma_i \) is the standard deviation of all students in the questions, and \( \sigma \) is the standard deviation of all students in the total test scores. The above equation shows that the pre-test evaluation \( \alpha = 0.81 \), and post-test evaluation \( \alpha = 0.82 \), which means both pre-test and post-test evaluations reach reliable levels.

5.1.2 Validity
Validity is the extent to which the results of the evaluation fit the results, or the determination of whether an evaluation is meaningful and can cover the research topic. The results show that pre-test questions comprise 29%, 15%, 23%, and 33% of the total test bank ability indicator ratio, and post-test questions comprise 25%, 30%, 23%, and 33% of the total test bank ability indicator ratio. The test questions are approximately between 20%~30% of the total test bank distribution ratio, meaning that this test is valid for testing students’ cognition and learning.

5.2 Data Analysis
The data are analyzed as follows:

5.2.1 Calculation of Test Difficulty
The post-test in this study is designed to be more difficult than the pre-test test; however, scores are incomparable due to the different levels of difficulty, thus, the formula for calculating test difficulty is as follows:

\[
PD_k = \frac{1}{m} \sum_{i=1}^{m} \left( \lambda + (\epsilon_i - \bar{\epsilon}) \times 10 \right), \text{where } k = 2
\]

where, the number of questions on tests is \( m \), \( \lambda \) is the difficulty value of test questions, \( \lambda \in N \), \( \epsilon_i \) is the incorrect rate of all students for question \( i \), \( \bar{\epsilon} \) is the average incorrect rate of students. The above equation can be used to obtain the difficulty index \( PD_k \) after the test is modified.

5.2.2 Normalization of Raw Scores
Raw average scores in the pre-test and post-test evaluations must be normalized before they can be compared for analysis; the formula for calculating the normalization of raw scores is as follows:

\[
E_k' = E_k \times \frac{PD_k}{PD}, \text{where } k = 2
\]

where, \( E_k \) is the average score of a group’s pre- or post-test evaluation, and \( E_k' \) is the average score after modification and normalization.
5.2.3 Statistical Tests

This study uses mean difference to test whether there are significant differences between the two populations; however, since the experiment includes three groups, two groups are used for the statistical test each time, and must be conducted three times. Statistical testing analysis is as follows: null hypothesis \( H_0 : \mu_1 - \mu_2 = 0 \): represents the hypothesis that there are no differences between the two populations, alternative hypothesis \( H_1 : \mu_1 - \mu_2 \neq 0 \): represents the hypothesis that there are differences between the two populations. Since each group of students exceeds 30 people, they can be referred to as large samples, thus, the two-tailed Z-test is used, and the level of significance for \( \alpha \) value is set to 0.1. The formula for calculating the threshold value is as follows:

\[
C = (\mu_1 - \mu_2)H_0 \pm Z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}
\]

where, \( \mu_i \) is the mean value of the population \( i \), \( \sigma_i \) is the standard deviation of the population, and \( n_i \) is the number of populations. If the difference between the means of two populations is greater than the threshold value \( C \), it means \( H_0 \) is rejected due to significant differences between the two populations; and is accepted \( H_0 \) if vice versa, meaning there are significant differences between the two populations.

Analysis of the statistical test results for the experiment group and control group showed no significance between pre-test scores of the two groups; however, in the post-test scores, the experimental group scored significantly higher than the control group. For the comparison group and control group, there are no significant differences in their pre-test scores, and post-test scores of the comparison group are not significantly higher than that of the control group. For the experiment group and comparison group, there is no significance between pre-test scores of the two groups; however, in the post-test scores, the experiment group is higher than the comparison group, though insignificantly.

6. Conclusions

This paper proposes an adaptive learning diagnostic system that applies the fuzzy theory, FADS. The results indicate that FADS can make online learning diagnosis analysis more adaptive. Moreover, the adaptive questions can be used to make tests more flexible, and the online reading adaptive marks can help each learner reviewing the emphasized points of the instructional materials. According to the results of the experiment, the post-test evaluation of the experimental group, which used FADS, has the greatest improvement in terms of scores, with post-test evaluation scores being significantly higher than that of the control group and the comparison group.

FADS still has room for improvement, such as how to adjust the calculation methods of test concept similarities, and the transfer function of the fuzzy membership functions can be adjusted. Future study will continue to develop online real-time learning diagnosis systems that are more adaptive and flexible, so that digital learning platforms can have more diverse development and applications.

REFERENCES