

Chapter 6

Students' Performance Prediction Using Multi-Channel Decision Fusion

H. Moradi, S. Abbas Moradi and L. Kashani

Abstract A teacher or an artificial instructor, embedded in an intelligent tutoring system, is interested in predicting the performance of his/her students to better adjust the educational materials and strategies throughout the learning process. In this chapter, a multi-channel decision fusion approach, based on using the performance in “assignment categories”, such as homework assignments, is introduced to determine the overall performance of a student. In the proposed approach, the data gathered are used to determine four classes of “expert”, “good”, “average”, and “weak” performance levels. This classification is conducted on both overall performance and the performance in assignment categories. Then, a mapping from the performances in “assignment categories” is learned, and is used to predict the overall performance. The main advantage of the proposed approach is in its capability to estimate students' performance after a few assignments. Consequently, it can help the instructors better manage their class and adjust educational materials to prevent underachievement.

Keywords Learning analytics · Student performance prediction · Educational data mining · Decision fusion · Assignment categories

H. Moradi (✉) · S. A. Moradi
School of ECE, University of Tehran, North Kargar St,
P.O. Box 14395-515 Tehran, Iran
e-mail: moradih@ut.ac.ir

S. A. Moradi
e-mail: s.a.moradi@ut.ac.ir

L. Kashani
Department of Psychology and Special Education, University of Tehran,
Jalal Al-e-Ahmad Avenue, P.O. Box 11455/6456 Tehran, Iran
e-mail: lkashani@ut.ac.ir

Abbreviations

BKT	Bayesian knowledge tracing
CART	Classification and regression tree
CHAID	Chi-square automatic interaction detection
EDM	Educational data mining
ITS	Intelligent tutoring systems
LMS	Learning management system

6.1 Introduction

Performance prediction is very important in conducting in-time interventions for the students, whose performances are not at the expected level, especially the students at-risk¹ for poor academic performance [1]. The detection of the at-risk college students becomes highly important in the critical first year [2]. The first year failing students are more likely to fail throughout the rest of that academic program [1]. It should also be noted that the typical performance evaluation at the middle of a learning process, i.e. the midterm exams, is too late for detection of the at-risk students [3].

According to Tinto's Student Integration Model [2], student's academic performance is one of the factors influencing student's integration in academic life, which impacts the student's decision to stay in the program of study (retention) or to leave the program of the study (attrition). Since retention is a formidable issue in any educational institute, academic performance and its prediction is of significant importance.

From another point of view, performance prediction is also necessary for an instructor to be able to plan the instructions personalized according to the learners' zone of proximal development [4]. According to Vygotsky's zone of proximal development, any given learner has a level of learning without any guidance; the learner also has another level of performance, in which the student cannot complete tasks unaided, but can perform well with guidance. The difference between these two levels indicates any learner's zone of proximal development, which is very crucial for the instructor to find and plan the instructional materials and strategies accordingly. In other words, the performance prediction helps an instructor to determine whether or not a student falls in his/her zone of proximal development. Then the teacher can adjust the teaching materials and strategies to the learners' potentials. Furthermore, the instructions can be planned to be more challenging than the current level of knowledge, but still within the possible range of potential capabilities and competencies.

¹ At-risk students are the ones who do not experience success at school, more likely to fail academically, and may drop out. It is important to detect them and help them as early as possible to avoid future failures and eventual drop out.

The professional instructors have a natural talent or have built an expertise over the time to predict the performance of their students as soon as possible. Unfortunately, this expertise does not come easily and it is not acquired in a short time; it cannot be easily transferred to pre service or in-service instructors, and is not available for e-Learning systems. Consequently, it is important to develop methods and algorithms that can help teachers or artificial instructors to better predict the performance of their students as early as possible. That has been the drive behind many research studies in the area of performance prediction and skill estimation [5–11].

The development of such methods, to predict the performance of students, can help a teacher in the form of a program, such as a module for a spread sheet or a standalone program, to process the assignments' grades. The prediction made by the program can be used by the instructor to monitor his/her students' progress, and decide on his/her next educational activities accordingly. On the other hand, these methods can be implemented as a module for a Learning Management System (LMS) to automatically retrieve the assignments' grades from the database, and predict the performance level. The results would be accessible to the instructor for planning his/her activities. Similarly, an artificial tutor in an Intelligent Tutoring Systems (ITS) automatically uses these methods to predict the performance level of its learner(s) and adjust the educational materials accordingly.

From another point of view, such a method can be used in traditional, semi-online,² and online learning environments to predict the learners' performance, and adjust the learning materials accordingly. The experiments reported in this chapter are conducted in a semi-online setup, in which the students receive the lectures in a face-to-face format while they receive the materials and complete quizzes through an LMS.

In the subsequent sections, the importance of performance prediction and the related research is further examined. Furthermore, the data mining approaches in performance prediction has been reviewed. In Sect. 6.3, the multi-channel decision fusion performance prediction is introduced. Section 6.4 is dedicated to experimental setups and the performance of the multi-channel decision fusion approach. Discussion on the results and future works are explained in Sects. 6.5 and 6.6.

6.2 Student Modeling

Student modeling module is one of the 4 major modules in an ITS, including the student module, the expert module, the tutoring module, and the user interface module. These intelligent educational systems gain much of their power from having the student model that describes the learner's proficiencies at various

² Also referred as the mixed methods.

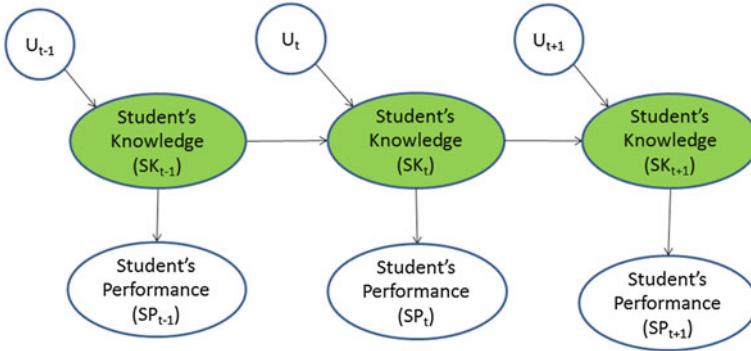


Fig. 6.1 The dynamic Bayesian Network that characterizes the evolution of a student's knowledge based on the previous knowledge (SK_{t-1}) and the learning activities (U_t). The student's performance is the measurable outcome of the knowledge. The unwanted effects during the learning process can be considered as noise for the learning activity

aspects of the domain to be learned. A student's model [12] represents the student's long term characteristics, such as personality, and short term characteristics, such as motivation and emotion. These characteristics can be used to diagnose errors and mistakes in a learner's knowledge, to determine learners' misconceptions, to predict a possible learner's reaction to the learning materials, and to evaluate learner's performance.

One of a learner's characteristics is his/her current knowledge, and the change in his/her knowledge, which is the most significant issue in any learning process. Several widely used models, such as overlay model, differential model, perturbation model, and constraint-based model, are used to model learners' knowledge [13]. However, since knowledge is not directly measurable, performance is used as a measurable outcome of the knowledge at any given time. That is why evaluation tools are used throughout the human history to indirectly measure the knowledge level of a person. To model the relationship between the knowledge and performance evaluation, Bayesian Networks can be used (Fig. 6.1).

The learning process, i.e. the change in the knowledge, can be modeled using a Bayesian Network based on Markov assumption, in which the current knowledge of a learner is shaped based on his/her prior knowledge and the learning activities (Fig. 6.1). In this model, the performance evaluation is the measurable outcome of the knowledge directly affected by knowledge. The knowledge at time $t-1$ can be acquired through previous studies or through interaction with friends or the environment. The change in the knowledge represents the probability of acquiring new knowledge, i.e. changing the level of the knowledge, after facing new learning materials (Eq. 6.1).

$$p(SK_t | SK_{t-1}, U_t) \quad (6.1)$$

In which SK_{t-1} and SK_t are learner's knowledge at times $t-1$ and t accordingly. U_t represents the learning materials in the given learning period. The unmodeled or undesired learning can be represented as noise for U_t . In other words, a student's knowledge can be estimated by his/her previous knowledge and the learning activity. To improve this knowledge estimation, the Bayesian filtering approach can be used (Eqs. 6.2 and 6.3), in which the estimated belief in a knowledge level is the summation of all possible previous knowledge levels that may reach the estimated knowledge level (Eq. 6.2), i.e.:

$$\bar{bel}(SK_t) = \sum_{SK_{t-1}} p(SK_t|SK_{t-1}, U_t)bel(SK_{t-1}) \quad (6.2)$$

In which $\bar{bel}(SK_t)$ is the estimated student's knowledge at time t and $bel(SK_{t-1})$ is the belief in the previous learner's knowledge. An effective, useful, and yet simple partitioning of the performance level is to classify it into four categories of weak, average, good, and excellent knowledge levels. In other words, the stereotype model is used to represent the knowledge level of a learner into four stereotypes, i.e. stereotype one as weak learners, stereotype two as average learners, stereotype three as good learners, and stereotype four as excellent learners. Obviously, it is possible to change the stereotype models based on the domain knowledge if needed, such as the case that has been conducted in knowledge analysis studied in PeRSIVA [14] in which 8 stereotypes are defined. The estimated belief in the knowledge can be corrected (modified) by observing the knowledge through performance evaluation (Eq. 6.3).

$$bel(SK_t) = \eta p(SK_t|SK_{t-1}, U_t) \bar{bel}(SK_t) \quad (6.3)$$

In other words, by having the learning materials and activities, it is possible to guess the extent to which the students have met the learning objectives. By measuring the performance, we can correct the guess, and determine the real learning acquired. It should be noted that determining how much a learning material or activity can affect learners' knowledge is an important issue. This is what talented and experienced educators have empirically learned.

On the other hand, the probability of a performance assuming a given knowledge (Eq. 6.4) depends on many factors such as the probability of slip or guess [11] and the performance evaluation materials.

$$p(SP_t|SK_t) \quad (6.4)$$

Since the extent to which a student has met the learning goals in the process of learning is the measure of successfulness of an educational system, it becomes important to accurately determine the knowledge level of a learner.

Although the above model can help to better evaluate the current knowledge of a learner, the early prediction of his/her knowledge remains an important issue in order to improve the learning process and the educational system. The following sections address this issue in more detail.

6.3 Performance Prediction

One of a learner's characteristics is his/her current knowledge, which is evaluated using his/her performance in an assignment or examination. From the learner's point of view, the performance evaluation can be formative [15], such that it would give appropriate feedback to the learner on his/her current state, or summative [15] that summarizes the knowledge of the learner at the end of the learning period. From the system's point of view, or an instructor's point of view, this evaluation can be used in diagnostic, predictive and evaluative forms to improve the learning process.

Due to the importance of the performance prediction, which allows on the fly adjustment to the learning materials in improving the learning process outcomes, this chapter is devoted to performance prediction using assignment categories. In other words, based on the current performance, i.e. SP_t , in different assignments the most accurate predicted performances for the midterm, i.e. $SP_{t=midterm}$, and for the final exam, i.e. $SP_{t=final}$, are needed.

In the following, the position of the performance prediction in a general student modeling module is explained. Later, its position in an ITS and the data mining approaches used for performance prediction are explained. A comprehensive list of research studies using Educational Data Mining (EDM) has been prepared by Romero and Ventura [16].

6.3.1 Performance Prediction in ITS

Although performance prediction is useful for all learning environments, from traditional face-to-face learning environments, to e-Learning and ITS, nonetheless, it becomes more valuable for e-Learning and ITS, since it tries to compensate the shortcomings caused by the lack of face-to-face interaction of a human expert.

It should be noted that assessing a learner's knowledge, especially through performance evaluation, is difficult because (1) part of the learner's proficiency evaluation comes from visual observation which is not available in online systems, in ITS, and even in large traditional classrooms since the instructor is incapable of having direct interaction with most of the learners, (2) due to low reliability and validity of most available testing instruments, the learner's performance in an exam or quiz may not be a perfect reflection of the learner's knowledge and proficiency in a field, and (3) the fact that the state of the learner's knowledge changes over time.

Additionally, the information gathered through human-computer interaction might not clearly and/or uniquely represent real world situation. Consequently, it is important to design a data mining system capable of correctly and efficiently process the data to predict the performance.

It should also be noted that the predicted performance can be represented in the typical scale format, i.e. 0–4 or 0–20 depending on the educational setup, or it can be represented in a few major classes such as weak, normal, good, and excellent. The later one is more appropriate since all the predictions are not 100 % accurate, and providing the predicted performance in the scale might create unnecessary expectations.

6.3.2 Data Mining Approaches for Prediction

Based on the fact that the performance, in any measurable setup such as exams or homework assignments, is a representation of the knowledge of a learner, a wide range of research studies have been performed to model and estimate performance. Decision Trees, classification methods and Bayesian Networks [13, 16, 17] are from the most widely used approaches. For instance, TELEOS is a learning environment in which Bayesian network is used to diagnose the student's knowledge state [18].

A set of research studies using Data Mining approaches have focused on considering the possibility of answering questions correctly/incorrectly, i.e. slip and guess, by chance [11]. The effects of slip and guess have also been considered in the Bayesian Knowledge-Tracing (BKT) approach [14].

Classification and regression trees [19, 20] are typical machine-learning methods for constructing prediction models from data. The models are obtained by recursively partitioning the data space and fitting a simple prediction model within each partition.

C4.5 and CART [20] (Classification and Regression Tree) are two classification tree algorithms that follow the general recursive tree building approach. C4.5 uses entropy for its impurity function, whereas CART uses a generalization of the binomial variance called the Gini index. These approaches first grow an overly large tree, and then prune it to a smaller size to minimize an estimate of the misclassification error. By default, CART employs tenfold cross validation, whereas C4.5 uses a heuristic formula to estimate error rates. CART estimates the dependent variable, while C4.5 estimates the class to which dependent variable belongs. In the proposed multi-channel decision fusion approach, in which the nearest neighbor approach is employed for classification, CART is used.

Chi-squared Automatic Interaction Detection (CHAID) [21] employs yet another strategy. If the input is ordered, its data values in the node are split into 10 intervals and one child node is assigned to each interval. If the input is unordered, one child node is assigned to each value of the input. Then, CHAID uses significance tests and Bonferroni corrections to try to iteratively merge pairs of child nodes. This approach has two consequences. First, a few nodes may be split into more than two child nodes. Second, considering the sequential nature of the tests and the inaccuracy in the grading, the method is biased toward selecting variables with fewer distinct values.

Due to the fuzzy nature of the human performance, fuzzy approaches have been widely used in the field [22, 23]. At the end, it should be mentioned that Pardos et al. [24] showed that an ensemble of different approaches could result in better predicting a student's knowledge level.

6.4 Multi-Channel Decision Fusion Performance Prediction

To predict performance in a course or in a learning setup, available grades from different assignments such as homework, lab assignments, projects, and online quizzes, which we will refer to them as assignment categories, can be used. As mentioned earlier, one approach to performance evaluation and/or prediction is to use all available grades from different assignments to determine the performance level, and to predict the performance in the midterm or final exam (Fig. 6.2a). However, this approach does not directly consider the importance and impact of each assignment category in the prediction and evaluation of the performance.

This becomes important when assignments are from different categories and each has its own characteristics and importance. For instance, the performance level that is represented by the homework assignment category, in which the students have adequate time to think, consult friends, or the teaching assistants, is completely different from the performance level that in class quizzes represent, in which the students have limited time and should analyze and answer the questions individually. Figure 6.2b represents a multi-channel decision fusion approach, in which the performance in each assignment category is determined, and those performance levels are used to determine the overall performance.

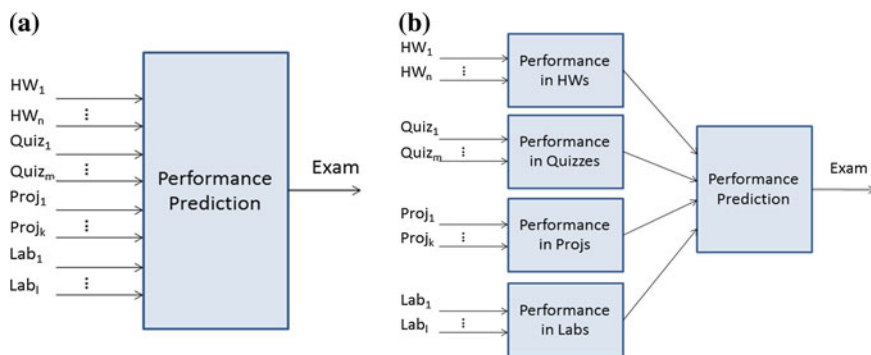


Fig. 6.2 Different approaches in performance prediction. **a** One shot performance prediction, in which the performances in assignments are used to predict the performance in the exam. **b** The multi-channel decision fusion for performance prediction in which each assignment category has a performance level, based on the related assignments, and the overall performance level is determined using only the performance in assignment categories

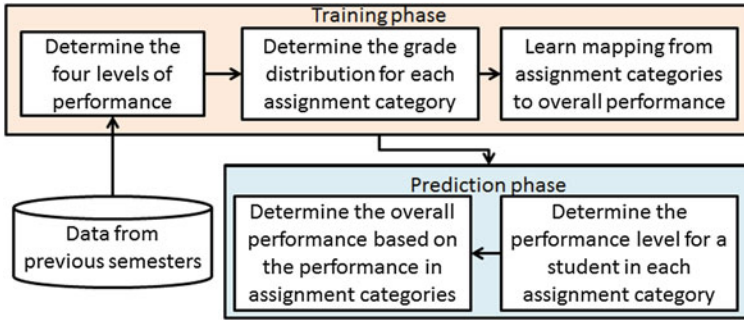


Fig. 6.3 The proposed multi-channel decision fusion approach consists of two phases, the training phase and the estimation phase. In the training phase, the data from previous semesters is used to create the mapping, from assignment categories to overall performance, which can be used in the current or future semesters

In order to conduct the performance prediction, a training phase is needed to learn the relationship between the performance in assignment categories and the overall performance, which is normally conducted based on the data provided from the previous runs of a course. In the case of the multi-channel decision fusion performance prediction, this training phase should be performed in order to determine the performance in assignment categories, the overall performance level, and the mapping between the two, i.e. from the performance in assignment categories to the overall performance. The overall training and estimation phases are shown in Fig. 6.3. In the following subsections, these three are explained.

6.4.1 Determining the Performance Level in Assignment Categories

As shown in Fig. 6.3, the first step in the training phase is to classify students into a few groups, which normally consists of four groups of expert, good, average and weak. Each group is represented by a normal distribution, i.e. $N(\mu, \sigma^2)$, with μ representing the mean, and σ^2 representing the variance. These groups are determined based on the data collected from previous runs of the course, which includes all the grades and final performance of the students in the course.

This classification can be made manually by an instructor or an expert, or by using intelligent methods such as K-means. The number of groups can also be determined manually or detected automatically. It does worth mentioning that the automatic approach normally results into more accurate classification than the fixed human classification. In the human classification, typically the mean for the expert performance level is set for 18.5 and its minimum is set for 17 in a scale of 20. However, in an automatic classification, 18.2 were determined to be the

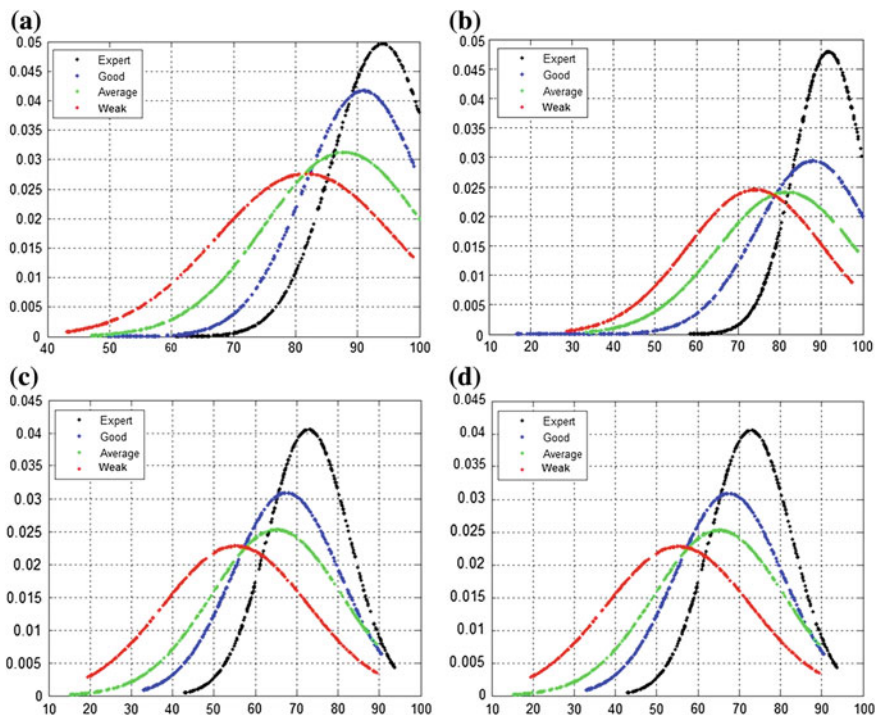


Fig. 6.4 The grade distribution for each assignment category in the four performance levels: **a** lab assignments, **b** homework, **c** quizzes, and **d** projects. The performance distributions are represented using Gaussian distribution (Table 6.3), in which the horizontal axes are the grades out of 100

mean and 17.40 as the minimum for the expert class. This happens to be more realistic based on the specific grades achieved in a specific course and was verified by an expert. Also the standard deviation is determined more accurately and can be updated over time.

After classifying the students into four performance levels, the grade distribution for each assignment category is determined (Fig. 6.4). By analyzing Fig. 6.4 several interesting points which are consistent with the reality can be observed:

1. The impact of plagiarism: There is no significant difference between groups of “average” and “good” learners in online quizzes. Perhaps it could be due to the fact that the learners collaborated together in answering the questions, which was confirmed by our indirect observations.³ Plagiarism also shows its impact in homework assignments. Although it was possible to detect plagiarism in

³ A survey was performed to validate this observation which has been explained at the end of Section 6.5.

homework assignments, it is not as easy as detecting plagiarism in programming projects, for which MOSS⁴ is used to detect similar codes and the students had to orally deliver the projects. In the case of lab assignments, the fixed structure of the lab assignments does not allow adequate discrepancy between performance levels.

2. The impact of interaction between close performance level groups: It is fairly visible that in case of collaboration and/or plagiarism, the learners tend to interact with peers in their own performance group or in a performance group close to them. For instance, the learners in the “weak” group tend to interact and get help from the learners in the “average” group. That is why their scores are closer to each other. Similarly, the learners in the “average” group tend to get help from the learners in the “good” performance level group. This phenomenon is more visible between the top three groups rather than the weak group. The reason could be that the weak learners, who mostly constitute at-risk students, tend to lose their motivation for better achievement, while the learners in average and good performance level groups still have hope to gain better grades through cheating and collaboration with their stronger peers.

Although the above two facts reduce the effectiveness of assignment categories in predicting the overall performance level, especially for lab and quiz assignments, these still provide clear distinction between “weak” and other groups, especially the “excellent” group.

That is actually the main reason that multi-channel decision fusion provides better prediction since these features/characteristics of the assignment categories can be clearly considered in the prediction.

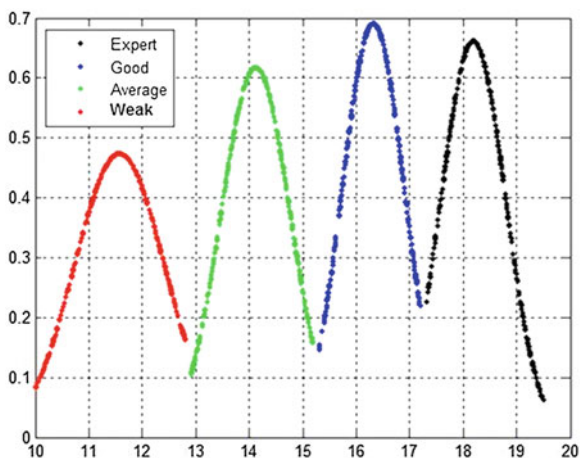
6.4.2 Determining Overall Performance Levels

The overall performance is typically determined based on the result in the final assessments, i.e. final exams in an educational system. On the other hand, midterm exams are used to provide feedback to both learners and instructors such that they can adjust their activities for better results. These methods of performance assessment are chosen as the ground truth of the performance level since:

1. These assessment methods suffer less from noise since these are conducted in a controlled setup, in which there is a lower possibility of plagiarism.
2. These assessments show the sole understanding and knowledge of a learner, since learners should answer questions on their own. In other words, in case of homework, they cannot get help from others or hide behind the performance of others, in case of team projects.

⁴ <http://moss.stanford.edu>.

Fig. 6.5 The clustering of the students into four distinctive performance levels



3. They represent the intermediate and final performance levels, one as a formative assessment and the other as a summative assessment.
4. These are accepted assessments approaches for performance evaluation.

Consequently, to determine the overall performance, the midterm and final grades from previous semesters are classified into the four groups. Similar to the approaches considered to determine the performance level in assignment categories, the four groups of performance level in midterm and final exams can be determined using manual or automatic classification.

Figure 6.5 shows the clustering performed on the final exam for a course using K-means. In this specific course, the four performance levels have clear distinction from each other. A comparison between Figs. 6.4 and 6.5 shows how midterm and final exams have stronger differentiating capability between the performance level groups, compared to assignment categories. This could be due to the noise involved in the assignment categories so it cannot be a clear representation of the learner's knowledge.

6.4.3 Mapping from the Performance in Assignment Categories to Overall Performance

Based on the discussion in the previous subsection, the mapping from the performance in assignment categories to overall performance is based on the midterm and final exam grades. To develop the mapping between the performance level in assignment categories and exam grades, different data mining approaches can be used. The simplest approach can be linear regression. Other possible approaches can be CHAID and CART. Even an ensemble of different methods can be used to take advantage of the strength of each method. A study on a specific course shows

that linear regression has the best performance among these three, which would be discussed in the experimental results section.

After selecting the right method for mapping between the assignment categories and the overall performance, the training data is used to train the map. In the prediction phase, the mapping is used to find the overall performance of a learner based on the available grades in the assignment categories.

The nearest neighbor, i.e. the performance level with the closest average performance to the performances determined in the training phase, is considered as the performance level of the learner. The nearest neighbor is determined by Euclidean distance between the means of two groups. In the case that the distances between the means are too close, the distance between variances is also considered.

In the last step, the accuracy of the mapping between the performance levels in the assignment categories and the final performance level is measured using the current midterm and final exams. Then, the mapping is updated to reduce the mapping error.

6.4.4 The Characteristics of Assignment Categories

To better understand the importance of multi-channel decision fusion performance prediction, it is necessary to compare the assignment categories based on the features that have impact on evaluating the knowledge level.

The following list (Table 6.1) is a set of important features proposed in this chapter. Other features may be added in the future. The importance of a feature in

Table 6.1 Typical assignment categories and their features

	Plagiarism	Plagiarism detection	Time limit	Discriminatory	Legitimate help	Slip and guess	Team work
e-Homework	+	+	-	+	0	-	-
Homework	+	-	-	+	+	-	-
Coding Projects	+	+	-	+	+	-	-
Hardware projects	0	0	-	+	+	-	-
Online quiz	+	-	-	0	-	+	-
In-class quiz	-	-	+	0	-	+	-
Laboratory	0	0	0	-	+	-	-
e-Essay	+	+	-	+	+	-	-
Essay	+	-	-	+	+	-	-
Take home exam	+	-	0	+	-	-	-
Exams	-	-	+	+	-	0	-
Team projects	0	+	0	0	+	-	+

any of the assignment categories is shown by “-”, when it is not important, “0” when it is neutral, and “+” when it is important.

- **Plagiarism:** This feature shows the possibility of cheating in a given assignment. For example, it is harder to plagiarize in essay exams, compared to coding projects.
- **Detecting Plagiarism:** This feature shows the capability of detecting plagiarized materials. For instance, hard copy assignments are difficult to be checked against others while checking electronics assignments is easier.
- **Time limit:** This feature shows if a learner can be under pressure due to the limited time for completing the assignment.
- **Discriminatory:** Which shows that the assignment can clearly discriminate between different knowledge levels or not. For instance, laboratory assignments are too structured and less discriminatory since the answers are normally fixed.
- **Legitimate help:** Which shows if asking for help and guidance is allowed for a given assignment or not. It shows if the instructed materials have been completely absorbed, or the learner still needs guidance to make use of his/her knowledge.
- **Slip or guess:** That shows if slip and guess can easily happen in a given assignment. For instance, in coding projects, guessing is hardly possible, while in quizzes, assuming multiple choice or true/false quizzes, it is easy.
- **Team work:** This refers to the fact that in the assignments completed in groups, it is difficult to evaluate the contribution shares, i.e. the knowledge level and the effort of each team member.

Table 6.1 shows the importance of each of the above features in a set of typical assignment categories. If a new assignment is designed, its features can be compared to the listed features in this table to decide whether it should be considered as a new category or it can be included into an existing category.

6.5 Experimental Results and Discussion

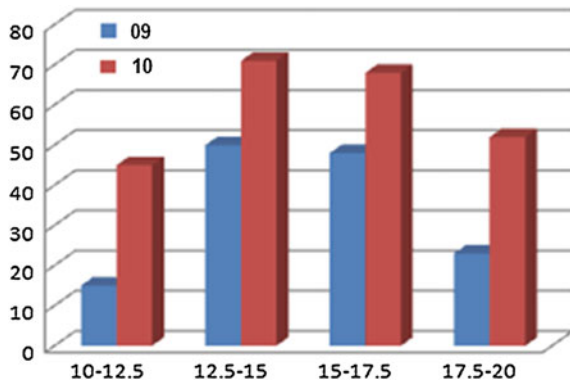
In order to evaluate the proposed method, 387 students who took the “Introduction to Computers and Programming” course at the school of ECE, University of Tehran, in the fall semesters of 2009–2010, have been selected. The majority of these students are freshmen, taking this course at the university level for the first time and in the first semester after entering college. Since these students have to pass the national entrance exam to enter the school, most of the students are among the top 1,000 students in the country.

The course includes four assignment categories, i.e. online quizzes, laboratory assignments, homework, and projects with 5, 15, 5 and 15 % of the total grade respectively. The course is conducted in the combined traditional face-to-face and online format, in which Moodle is used as LMS to deliver quizzes, slides

Table 6.2 Extracted skill levels from K-means classification

	Min	Max	Mean	SD	Total
Expert	17.40	19.5	18.20	0.55	92
Good	15.30	17.3	16.35	0.56	81
Average	12.90	15.2	14.12	0.63	80
Weak	10.00	12.8	11.53	0.81	51

Fig. 6.6 The distribution of the grades in the four categories of expert, good, average and weak. The percentages of the grades in each range are shown for two different semesters, i.e. fall 2009 and fall 2010. The grades are out of 20



and readings, assignments, grades, and to provide online collaborative features, such as discussion forums and news. Consequently, the LMS contains all the grades and data about the course. 80 % of the data is used for training and classification and 20 % is used for testing. The data processing and normalizations are performed using Weka.⁵

Table 6.2 shows automatic data classification that has distinguished four different performance levels, i.e. expert, good, average, and weak, which matches human intuition. 30.16 % of the students have been categorized with expert performance level, 26.6 as Good, 26.2 as average and 16.7 % as weak (Fig. 6.6). Total represents the total number of students in each performance level. It should be noted that the number of students below 10 were very small, and did not constitute a group with adequate data.

Figure 6.5 shows four clear levels of performance existing in the course. It should be noted that k-means is used to perform clustering in this step. We have observed that k-means clusters in four or five groups based on the given data in the class. Thus, k-means is setup to get four groups to have a fixed set of groups for all the data sets, i.e. grades in different assignment categories. Also, this automatic approach has more accurate clustering than the fixed human clustering since the learning setup would be slightly/greatly different from semester to semester. This difference could be based on the individual differences between students, instructors teaching the course, changes in the course materials and assignments,

⁵ Waikato Environment for Knowledge Analysis.

and the variations in the teaching assistants helping with the course. In the human clustering, normally the mean for an expert is set for 18.5 and its minimum is set for 17. However, in automatic clustering, 18.2 is determined to be the mean and 17.40 as the minimum for the expert class. Also the standard deviation is determined more accurately, and can be updated over time.

After determining the four overall performance levels in the course, the grades' normal distribution for each assignment category in each class of performance level is determined. Figure 6.4 shows the results of analyzing the assignment category distribution into four performance levels.

After determining the four overall performance levels in the course, the grade distribution for each assignment category in each class of performance level is determined. It is interesting to see that in the labs and online quizzes, the grade distribution between different performances levels do not differ significantly. In contrast, the grade distribution for the homework and projects differ between the four performance levels.

This could be due to the fact that the lab assignments are so systematic that the results are fairly close to each other, and does not allow differentiation between different performance levels. On the other hand, the possible reason that the online quizzes are not reliable measures to differentiate between different performance levels is that the students might cheat and work together to answer the online quizzes. It should be noted that although the labs and online quizzes are not good measures for differentiation among all the performance levels, however, they can be used to differentiate between expert and weak performance levels.

This can be justified based on the fact that the students who work together to answer quizzes are within the group or groups who feel closer together to collaborate with each other. For instance, weak and average students may work together to answer an online quiz, while students in good and expert groups tend to work together. Table 6.3 shows the distribution of the grades in each assignment category in the four performance levels.

In this step, the distributions of students' grades are calculated for these four assignment categories. Then each performance level that has the closest mean and standard deviation from the distribution of a student's grades is considered as the performance level of that specific student.

The estimation of the final performance through the assignment categories can be done using different methods. As mentioned in the previous section, three methods have been used and compared to each other for this estimation. The

Table 6.3 Calculated grade range for the learning objects

	Lab		Quiz		Homework		Project	
	μ	σ	μ	σ	μ	σ	μ	σ
Expert	94.2	7.7	73.7	9.7	92.0	8.4	97.2	5.3
Good	91.0	8.5	68.3	13.0	88.4	12.2	87.8	14.3
Average	88.1	12.2	64.8	15.9	81.6	16.2	75.9	20.5
Weak	82.1	14.6	54.2	18.4	74.7	17.1	61.7	23.2

Table 6.4 The comparison of different methods for performance level classification. It is clear that, over all, regression has better results than the other two

	CHAID	CART	Regression	CHAID	CART	Regression
<i>Estimated</i>	<i>Was Expert</i>			<i>Was Good</i>		
Expert	64.3	82.1	96.6	19.2	42.3	26.9
Good	35.7	14.3	3.4	46.2	42.3	73.1
Average	00.0	3.6	00.0	34.6	15.4	00.0
Weak	00.0	00.0	00.0	00.0	00.0	00.0
<i>Estimated</i>	<i>Was average</i>			<i>Was weak</i>		
Expert	25.0	15.4	0.0	0.0	0.0	00.0
Good	50.0	15.4	15.4	6.3	6.3	00.0
Average	33.3	7.7	38.5	6.3	00.0	12.5
Weak	00.0	61.5	61.5	87.5	93.7	87.5

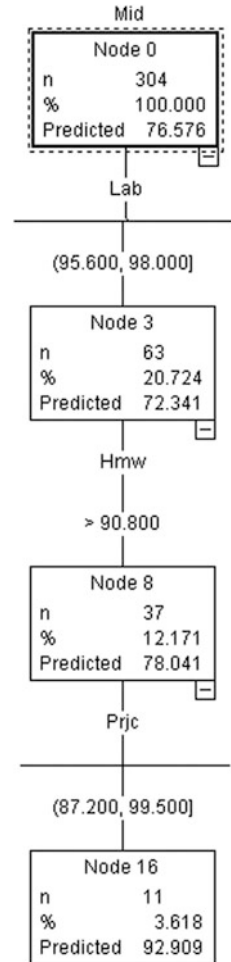
results are shown in Table 6.4. An advantage of CART and CHAID is that the results are shown in a hierarchical tree, and the results can be analyzed easier. Figure 6.7 shows a branch of the tree generated by CHAID and Figs. 6.8 and 6.9 show the result of CART and CHAID trees, respectively.

In this branch, the root node consists of all samples. At this point, it is possible to predict the midterm grade with 76.5 % accuracy. At the second level, the algorithms try to come up with a range, i.e. 95.6–98, in the LAB grades to improve the prediction. The accuracy has dropped to 72.3 %. Including the homework grades increased the accuracy to 78 % at the 3rd level and the projects could increase the accuracy to 92.9 %. It should be noted that at the root level, no assignment category is considered and the classification is done based on the raw assignment grades. Thus, it clearly shows the advantage of multi-channel decision fusion approach to the basic approach with 16 % increase in the prediction rate.

As shown in Table 6.4, CHAID classified the students with 64.3, 46.2, 33.3 and 87.5 % accuracy in expert, good, average, and weak classes, respectively. As it can be seen, CHAID performs better in good class than CART, while CART performs better in expert and weak. Both approaches were not very successful in classifying students in the good group. As it can be seen in the table, regression approach outperforms the other two approaches. Consequently, linear regression is more suitable for performance level estimation than the other two approaches.

Prediction of a student's final performance level, as soon as possible, is very important in adjusting the course materials. Also, as mentioned earlier, this is crucial for helping at-risk students, especially in the first year of college or university [1]. Consequently, the lower the number of assignment categories needed to effectively predict the performance of a student, more suitable the item to be used for performance level prediction. That is why the results in the "Introduction to Computers and Programming" course has been analyzed to determine the best learning objects for performance level prediction. The result shows that the performance of a student can be predicted by using two to three homework or project grades. Using three to four grades can determine the performance level with high confidence.

Fig. 6.7 The figure shows a branch of the tree generated by CHAID which is shown in Fig. 6.8. In this example, it is shown how the midterm prediction for a skill level is conducted based on lab, then HW, followed by the project. “*n*” represents the number of students predicted, “%” the percentage of the students predicted in this group and “predicted” represents the accuracy of prediction



As mentioned earlier, the grades for quizzes and lab assignments do not have clear differentiation power between all four performance levels and cannot be used for this purpose. Furthermore, four to six grades are needed to be able to conduct performance level classification. Consequently, those would not be used for the purpose of early classification.

To further evaluate the effectiveness of the proposed approach, the approach was conducted in the Artificial Intelligence course, at the school of electrical and computer engineering, University of Tehran, with 60 registered students. Two assignment categories have been used in this course. Since the study was only performed in one semester and the data from other semesters was not available, both training and testing were performed on the same data from one semester. Consequently, to train the system, beside the assignment categories, the midterm results are used as the performance level ground truth. Also, 80 % of the data is

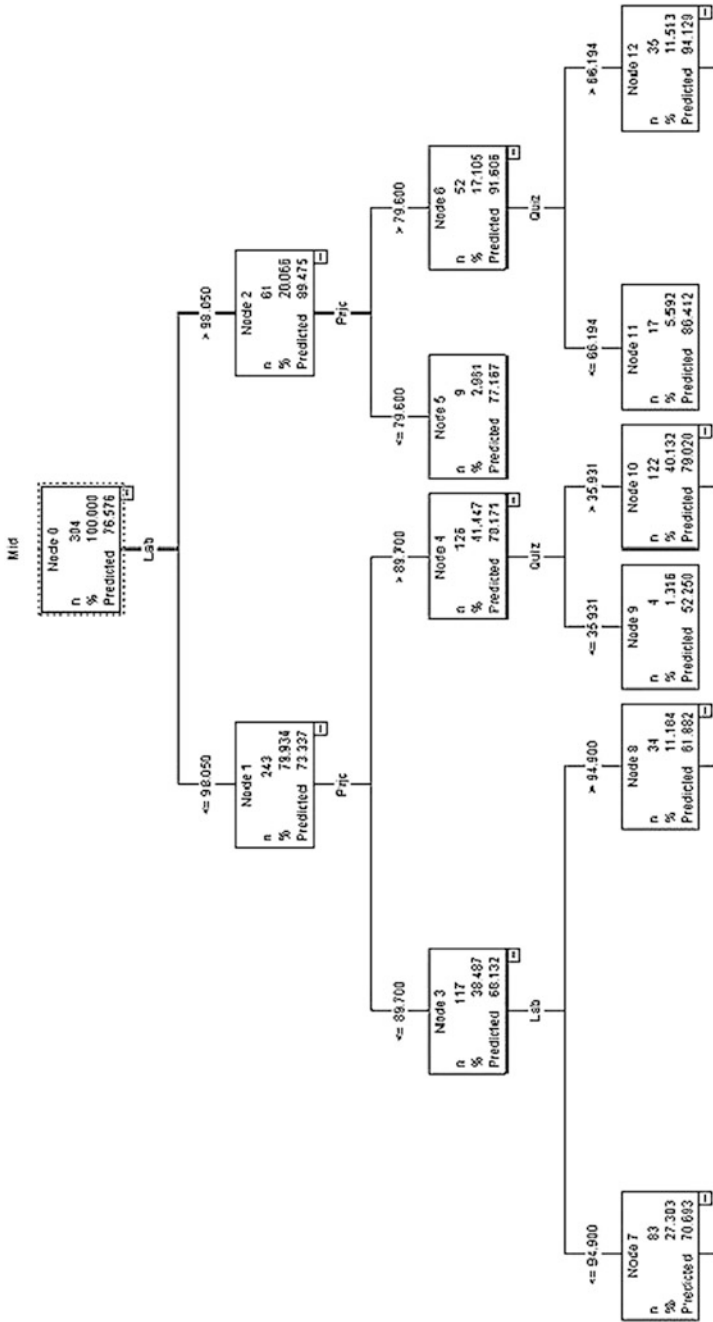


Fig. 6.8 Four levels of CART decision tree

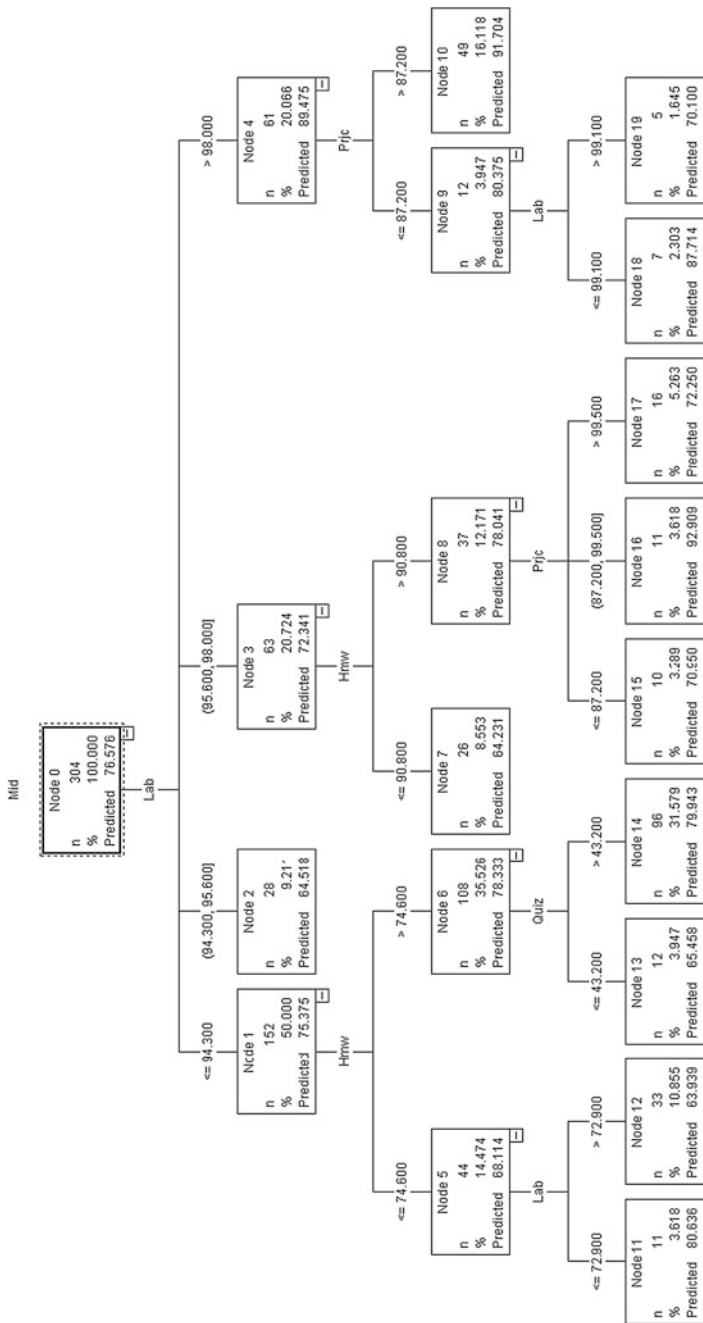


Fig. 6.9 CHAID decision tree

Table 6.5 The distribution of the grades in the Artificial Intelligence course in three groups

	Homework		CA		Midterm		Final		Total	
	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ
Expert	84.67	6.62	85.91	3.73	78.40	6.64	73.00	7.87	15.93	0.80
Good	72.88	7.79	78.31	10.17	66.53	5.22	63.41	7.17	13.94	0.75
Average	44.31	9.22	53.33	20.18	57.22	14.06	50.00	10.63	11.17	1.00

used as the training set and 20 % as the test set. Since the students normally perform well in this course, k-means clustered the students in three groups based on their performances, i.e. excellent, good, and average. Interestingly, the importance of the homework assignment category in performance prediction was lower than the importance of the projects.

This could be due to the fact that the students could easily cheat in homework, while it was harder to plagiarize in projects. Furthermore, the result shows 75 % accuracy in predicting the midterm exam grades only by having two homework grades and one project grade. This shows that this approach is effective even in a new course with limited assignment categories, in order to predict the performance level of a learner. Table 6.5 shows the distribution of the grades in the Artificial Intelligence course, in which the grades are classified into three groups.

It should be mentioned that in this course, the same conclusion, i.e. regression outperforms CART and CHAID, has been made. Regression correctly predicted the performance level of 12 students out of the 16 test cases. Meanwhile, CART only predicted five cases correctly, and CHAID could not make the decision tree in two performance levels.

The reason behind the better performance of linear regression over CART and CHAID can be in the fact that linear regression does not ignore the possible correlation between the variables, i.e. the assignment categories in this problem. However, CART and CHAID consider each variable separately to make the classification. The results show that the correlation between these variables is not negligible, and should be considered.

As mentioned earlier, it was observed that the interaction between students in groups may reduce the differentiation between the groups, specially the average and good groups which are at the middle of the group spectrum. To validate this observation, the students were asked to give the name of three classmates with whom they have interacted the most in completing their assignments. Eighty-three students responded to this survey. The average grade of each set of students, whom interacted with each other were, used to classify them to their nearest performance level group. The standard deviation of these set of students is 1.96, confirming the fact that the students have high tendency to interact within their performance level group or close groups.

6.6 Conclusion and Future Work

Performance level prediction is important, because it can be used to adjust the learning materials to improve the learning experience of students. It becomes very crucial for helping at-risk students, especially in the first year of college or university. The result of performance level prediction can be used by a human instructor or by an intelligent agent in an intelligent tutoring system. In this chapter, a multi-channel decision fusion approach is proposed to determine the performance level as early as possible based on determining the performance levels in assignment categories such as homework and projects.

This approach consists of two phase of training and estimation. The advantages of the proposed method are:

- The student's performance level can be determined at early stages of a course, based on a few important assignment categories. In the case of our "Introduction to Computers and Programming" course, it can be determined up to 5 weeks after the beginning of the semester based on the homework and projects. In other words, evaluation via the assignments and the programming projects will accelerate the recognition of performance level.
- The system can determine which assignment category helps improve the quality of a student's performance. Consequently, a human instructor or an intelligent tutoring system can use this information to tailor the course for the best performance.

The importance of the proposed approach, compared to the other proposed approaches such as fuzzy skill level estimation, Bayesian networks, and Factorization methods, is in using assignment categories levels, rather than the overall performance levels to estimate the future performance levels.

It should be mentioned that since normal distribution is used to model the performance levels, at least 30 samples are needed for each level to correctly model the performance level. If lower number of samples is available, then Z or T distributions may be used.

The future work would focus on using neural networks for better learning the mapping between the performance level in assignment categories and the final performance level. Furthermore, we will study the use of fuzzy logic to better represent the fuzziness in the data. Also, as it was discussed in [Sect. 6.5](#), the correlation between the assignment categories could be very important and it should be further investigated.

Although we planned to consider the learning style effects in the performance prediction, the learning style of the students at the engineering school might be limited to certain classes. Consequently, a wider study needed to be done to analyze the impact of the learning styles in performance prediction. Finally, the possibility of using the interaction of the user with system through non-assessment learning objects would be investigated.

Acknowledgements The authors like to thank Babak Araabi and Maryam Mirian for their feedback on this chapter. Furthermore, the authors like to thank the ITS group at the Advanced Robotics and Intelligent Systems Laboratory for their help throughout this research.

References

1. Singell, L.D., Waddell, G.R.: Modeling retention at a large public university: can at-risk students be identified early enough to treat? *Res. High. Educ.* **51**(6), 546–572 (2010)
2. Tinto, V.: From theory to action: exploring the institutional conditions for student retention. In: Smart, J.C. (ed.) *Higher Education: Handbook of Theory and Research*, vol. 25, pp. 51–89. Springer, Heidelberg (2010)
3. Donnelly, J.: Use of web-based academic alert system for identification of underachieving students at an urban research institution. *Coll. Univ.* **85**(4), 39–42 (2010)
4. Wass, R., Harland, T., Mercer, A.: Scaffolding critical thinking in the zone of proximal development. *High. Educ. Res. Dev.* **30**(1), 317–328 (2011)
5. Torabi, R., Moradi, P., Khantaimoori, A.R.: Predict student scores using bayesian networks. *Procedia Soc. Behav. Sci.* **46**, 4476–4480 (2012)
6. Ramaswami, M., Bhaskaran, R.: A CHAID based performance prediction model in educational data mining. *Int. J. Comput. Sci. Issues* **7**(1), 10–18 (2010)
7. Kotsiantis, S.B.: Use of machine learning techniques for educational purposes: a decision support system for forecasting students' grades. *Artif. Intell. Rev.* **37**(4), 331–344 (2012)
8. Baker, R., Pardos, Z., Gowda, S., Nooraei, B., Heffernan, N.: Ensembling Predictions of student knowledge within intelligent tutoring systems. In: Konstan, J., Conejo, R., Marzo, J., Oliver, N. (eds.) *User Modeling, Adaption and Personalization. LNCS*, vol. 6787, pp. 13–24. Springer, Heidelberg (2011)
9. Thai-Nghe, N., Drumond, L., Krohn-Grimberghe, A., Schmidt-Thieme, L.: Recommender system for predicting student performance. *Procedia Comput. Sci.* **1**(2), 2811–2819 (2010)
10. Ghazarian, A., Noorhosseini, S.M.: Automatic detection of users' skill levels using high-frequency user interface events. *J. User Model. User-Adap. Inter.* **20**(2), 109–146 (2010)
11. Thai-Nghe, N., Drumond, L., Horvath, T., Krohn-Grimberghe, A., Nanopoulos, A., Schmidt-Thieme, L.: Factorization techniques for predicting student performance. In: Santos O. C., Boticario J. G. (eds.) *Educational Recommender Systems and Technologies: Practices and Challenges*, pp. 129–153. IGI Global, Hershey (2012)
12. Chrysafiadi, K., Virvou, M.: Student modeling approaches: a literature review for the last decade. *Expert Syst. Appl.* **40**(11), 4715–4729 (2013)
13. Millán, E., Loboda, T., Pérez-de-la-Cruz, J.L.: Bayesian networks for student model engineering. *Comput. Educ.* **55**(4), 1663–1683 (2010)
14. Desmarais, M.C., d Baker, R.S.J.: A review of recent advances in learner and skill modeling in intelligent learning environments. *User Model. User Adapt. Inter.* **22**(1–2), 9–38 (2012)
15. Black, P.: Pedagogy in theory and in practice: formative and summative assessments in classrooms and in systems. In: Corrigan, D., Gunstone, R., Jones, A. (eds.) *Valuing Assessment in Science Education: Pedagogy, Curriculum, Policy*, pp. 207–229. Springer, Heidelberg (2013)
16. Romero, C., Ventura, S.: Educational data mining: a review of the state of the art. *IEEE Trans. Syst. Man Cybern. Part C: Appl. Rev.* **40**(6), 601–618 (2010)
17. Baker, R. S. J. d.: Data mining for education. In: McGaw, B., Peterson, P., Baker, E. (eds.) *International Encyclopedia of Education (3rd edn.)*, vol. 7, pp. 112–118. Elsevier, Oxford (2010)
18. Chieu, V.M., Luengo, V., Vadcard, L., Tonetti, J.: Student modeling in orthopedic surgery training: exploiting symbiosis between temporal bayesian networks and fine-grained didactic analysis. *J. Artif. Intell. Educ.* **20**(3), 269–301 (2010)

19. Theodoridis, S., Koutroumbas, K.: *Pattern Recognition*, 4th edn. Academic Press, Oxford (2008)
20. Han, J., Kamber, M., Pei, J.: *Data Mining: Concepts and Techniques*, 3rd edn. Elsevier, Oxford (2011)
21. Soman, K.P., Diwakar, S., Ajay, V.: *Insight into Data Mining: Theory and Practice*. Prentice Hall, India (2006)
22. Jia, B., Zhong, S., Zheng, T., Liu, Z.: The study and design of adaptive learning system based on fuzzy set theory. In: Cheok, Z.A.D., Müller, W., Zhang, X., Wong, K., (eds.) *Transactions on Edutainment IV*. LNCS, vol. 6250, pp. 1–11(2010)
23. Chrysafiadi, K., Virvou, M.: PeRSIVA: an empirical evaluation method of a student model of an intelligent e-learning environment for computer programming. *Comput. Educ.* **68**, 322–333 (2013)
24. Pardos, Z.A., Baker, R.S.J.d., Gowda, S.M., Heffernan, N.T.: The sum is greater than the parts: ensembling models of student knowledge in educational software. *SIGKDD Explor.* **13**(2), 37–44 (2011)