Abstract—Skill level estimation is very important since it allows an instructor, a human or an artificial instructor through an intelligent tutoring system, to predict the level of a student and adjust the learning materials accordingly. In this paper, a new approach based on 1-NN (First Nearest Neighbor) is introduced to determine the skill level of a student based on the pattern of skill levels learned over time in the same course. The data over several years are used to determine four clusters of expert, good, average and bad skill level. The advantage of the proposed approach is in its capability to adjust the levels over time based on the new data received each year. Furthermore, it can estimate the skill level after a few homework or project assignments. Consequently it can help an instructor to better conduct its class. The proposed approach has been implemented and tested on an introductory computer programming course and the results prove the validity of the approach.

Index Terms—Intelligent Tutoring Systems, Skill level estimation, Student Modeling.

I. INTRODUCTION

Skill level prediction is a mastery that good instructors and teachers have that helps them in their teaching by helping students who may not perform as expected. Unfortunately, this expertise does not come easily and in a short time and is not available for eLearning system. Consequently, it is important to develop methods and algorithms that can help human instructors, the combined systems of human instructors and Learning Management Systems (LMS), and fully automated systems such as Intelligent Tutoring Systems (ITS), to predict the skill level of the students as early as possible in order to help the students in their learning process.

In the latter case, i.e. the intelligent tutoring systems, there are four different subsystems or modules: the interface module, the expert module, the student module, and the tutor module. These intelligent educational systems derive much of their power from having a student model [3] that describes the learner’s proficiencies at various aspects of the domain to be learned. In such a module, the skill level prediction becomes very important.

However, assessing a student’s knowledge is difficult because 1) part of the student’s proficiency evaluation comes from visual observation which is not available in online systems, and even in large traditional classrooms since the instructor does not have direct interaction with most of the students, 2) the student’s performance in an exam or quiz may not be a perfect reflection of the student’s knowledge and proficiency in a field, and 3) the state of the student’s knowledge changes over time [4]. It also should be noted that the information gathered through human-computer interaction is imprecise, error prone, and its interpretation is vague and uncertain. Consequently, it is important to design a system to be able to correctly and efficiently process the data.

Therefore, skill level modeling module, as part of the student module, should be able to estimate the level of a student’s performance as soon as possible, observe changes in his/her skill level and classify students in groups according to their skill level. Such embedded module in an ITS, provides customization and personalization of the environment according to skill level and training required to improve educational quality. Such module would be very helpful for traditional classrooms too if the instructor uses an LMS such as Moodle or blackboard in which all the students’ grades and activities are available. The skill level estimator can provide early warning to the instructor about his/her students who are falling behind the class and need extra attention and help to catch up with the class.

In this paper, a skill level estimator has been proposed that uses the performance of students in different learning activities and estimates their skill level and their performance in the future. An advantage of the proposed approach is in its early rough estimation that is helpful for the instructors and the intelligent system.

The rest of the paper has been organized as follows. In section II, the other research related to this topic has been evaluated. The proposed method has been explained section III. In section IV the results and comparison with similar methods have been presented.

II. BACKGROUND AND RELATED WORK

Knowledge tracking is the new term that has been used for skill level modeling and estimation. A wide range of researches have been performed from Boolean logic to fuzzy logic-based approaches due to fuzziness involved in the skill level estimation and representation. Decision Trees, classification methods and Bayesian Networks [1][2][7][17][18][19][20][21][22][23] are from the most widely used approaches.

The Bayesian Networks use a hidden Markov model to model the skill level (Fig. 1). The hidden parameters are the initial knowledge, i.e. p(L0) for the first time and p(Lt) as the current knowledge before facing a new learning object, and the learned knowledge level at each step. This initial knowledge can be acquired through previous studies or through interaction with friends or environment. The change in the knowledge represents the probability of acquiring new knowledge, i.e. changing the level of the knowledge, after facing a new learning object.
In this model, the student is observed through questions and assignments in which \( P(g) \), i.e. the probability of answering a question from a higher knowledge level than the current level and \( P(s) \) is the probability of inability to answer a question at a lower level of the current knowledge level, are estimated. \( P(g) \) and \( P(s) \) are representations of a student level. These parameters are easier to estimate when only one skill level is evaluated compared to cases in which a number of skill levels are evaluated at the same time.

The Bayesian Network approach [1][7] has been used with and without considering the time spent answering questions. Liue et. al [7] used Bayesian belief network to assess the skill level and the slip probability for incorrectly answering each easy question and the lucky guess probability for correctly answering each hard question using the following information:

- The number of questions on a topic at each level answered correctly and incorrectly,
- The time spent on answering each question.

Classification and regression trees [13][14][15][16] 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. Fig. 2 shows a pseudo code of such methods.

C5 and C&RT [24] (Classification & Regression Tree) are two classification tree algorithms that follow the general recursive tree building method. C4.5 uses entropy for its impurity function, whereas C&RT 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. C&RT employs 10-fold (default) cross validation, whereas C4.5 uses a heuristic formula to estimate error rates. C&RT estimates the dependent variable while C5 estimates the class to which dependent variable belongs to. In this study C&RT is used since the nearest neighbor approach is used.

Chi-squared Automatic Interaction Detection (CHAID) employs yet another strategy. If the input is ordered variable, 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 to the sequential nature of the tests and the inaccuracy in the grading, the method is biased toward selecting variables with few distinct values.

Due to the fuzzy nature of skill level estimation soft computing techniques such as Fuzzy set theory have been used in applications of educational assessment [8-12]. Nolan in [8-9] has shown that an Expert Fuzzy classification scoring system can help teachers in making an assessment in less time and with a level of accuracy comparable to the best teacher. Ma and Zhou [10] proposed a fuzzy grade scale approach to assess and evaluate the outcome of the student-centered learning. Echauz and Vchtevanos [11] proposed a fuzzy grading method that utilizes a student’s performance measures to produce a fair mark distribution. Ranti [12] described a fuzzy set to evaluate the student’s answer script.

In this paper we present a method which estimates the skill level considering the learning objects, such as projects, quizzes, labs and homework. In contrast to other proposed methods which consider all learning objects together, the proposed method separates the skill level in each learning object resulting in more accurate estimation.

III. THE PROPOSED METHOD

As mentioned earlier, in this method the skill level in each learning object, which are called the sub-skill levels, has been considered separately. This is due to the fact that in a learning object such as lab assignments, the students follow exact predetermined steps and they perform closer to each other. In other words, their overall skill level would not be very dominant in the results. However, in learning objects such as homework or projects, they more rely on their own skill levels and these learning objects are better indication of their skill level.

As mentioned earlier, there are researches focused on using the overall skill levels to determine the final skill level. However, we used the sub-skill levels to determine the intermediate and final skill levels. The proposed approach consists of two major phases of training and estimation that are broken down into the following four steps:

1. Determining the number of skill levels, typically four
2. Determining the grade range in each skill level
3. Determining the reference/major final skill level
4. Determining the mapping between the reference skill level and the learning objects.

Fig. 3 shows the proposed framework for estimating the final skill level. It should be mentioned that the learning objects discussed here are the assessment-based learning objects such as quizzes, exams, homework and labs. The interaction of the user with other learning objects can be used for better estimation in the future.

In the training phase, first, the students are classified into a few groups, which normally happen to be in four...
groups of expert, good, average and bad. These groups are determined based on the data collected from previous runs of a course which includes all the grades and final performance of the students in the course. This classification can be done manually by the instructor or an expert, or by using intelligent methods such as K-means. The number of the groups can also be determined manually or detected automatically.

After classifying the students into four skill levels, the grade distribution for each learning object, for the four skill levels, are determined (Fig. 4). These two steps, i.e. determining the number of skill levels and the distribution of each skill level, constitute the training phase to determine the skill level in each learning object. In other words, the pattern from previous data is extracted to be used for estimation in the next phase.

To determine the overall skill level of a student and measure the accuracy of this estimation, the midterm and final grades are used as the reference learning object. The reason for selecting midterm and final exams as the ground truth, i.e. the referenced skill level, is that these show the intermediate and final skill levels. Furthermore, these assessments suffer less from noise since these are conducted in a controlled setup and there is a lower possibility of plagiarism. Finally, there is a mapping between the sub-skill levels measured in the learning objects and the skill levels measured in the midterm and final.

Thus, in the training phase (Fig 3), linear regression and classification and regression tree method, such as CART and CHAID, is used to determine the impact of each learning object in the midterm and the final grade of each class of skill levels. In other words, as the midterm and final grades are representations of students’ skill level, the output of step 4 is a mapping from sub-skill levels to overall skill levels.

To determine the skill level of a student, i.e. the midterm skill level or the final exam skill level, the learning objects’ grades for a given student are mapped into the skill levels using the mapping determined in the training phase. The results of this mapping are compared to the midterm or final grade of each level that determined in the training phase. The first nearest neighbor, i.e. a skill level with the closest average grade to the grades determined in the training phase, is considered as the skill level of the student.

In the last step, the accuracy of the mapping between the sub-skill levels and the final skill level is measured using the current midterm and final exams. Then the mapping is updated to reduce the mapping error.

IV. IMPLEMENTATION AND RESULTS

To evaluate the proposed method 387 students from “Introduction to Computers and Programming” course in 2009 and 2010 have been selected. The course includes four learning objects, i.e. online quizzes, labs, homework, and projects. 80% of the data is used for training and classification and 20% is used for testing. Table 1 shows automatic data clustering that has distinguished four different skill levels, i.e. expert, good, average, and bad, which matches human intuition. 30.16% of students have been categorized with expert skill level, 26.6 as Good, 26.2 as average and 16.7% as bad. Total represents the total number of students in each skill level.

Figure 4 shows four clear levels of skill existing in the course. It should be noted that k-means is used to perform clustering in this step. Furthermore, it should be noted this automatic approach has more accurate clustering than the fixed human clustering. 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 skill levels in the course, the grade distribution for each learning object in each class of skill level is determined. Figure 5 shows the results of analyzing the learning objects distribution in each four skill levels.

### Table 1. Extracted Skill Levels from K-means Clustering

|       | Min | Max  | Mean | STDDEV | 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    |
| Bad   | 10.00| 12.8 | 11.53| 0.81   | 51    |

http://www.i-jep.org
Figure 5. The grade distribution for each learning object in the four skill levels, a) labs, b) homework, c) quizzes, and d) projects. The skill distributions are represented using Gaussian distribution (Table II) in which the horizontal axes are the grades out of 100.

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

This could be due to the fact that the lab assignments are so systematic and the results are fairly close to each other that does not allow differentiation between different skill levels. On the other hand, the possible reason that the online quiz is not a good feature to differentiate between different skill levels is that the students tend to cheat and work together to answer the quizzes. It should be noted that although the labs and online quizzes are not good features between all the skill levels, however, they can be used to differentiate between expert and bad skill levels. This can be justified as the students who work together to answer quizzes are from similar group or groups which are closer together. For instance, bad and average students may work together to answer an online quiz while good and experts work together. Table II shows the distribution of the grades in each learning object in the 4 skill levels.

In this step, the distributions of students’ grades are calculated for these four learning objects. Then each skill that has the closest mean and standard deviation from the distribution of a student’s grade is considered as that student’s skill level.

The estimation of the final skill through these sub-skills 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 results are shown in table III. An advantage of C&RT and CHAID is that the results are shown in a hierarchical tree and the results can be analyzed easier. Fig. 6 shows a branch of the tree generated by CHAID and Figures 7 and 8 show the result of C&RT 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 LAB grades, with skill levels ranging from 95.6 to 98%, are included. 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%.

As shown in Table III, CHAID classified the students with 64.3, 46.2, 33.3 and 87.5 percent accuracy in expert, good, average, and bad classes respectively. As it can be seen, CHAID performs better in good class than C&RT while C&RT performed better in expert and bad. Both approaches were not very successful in classifying good students. As it can be seen in the table, regression approach outperforms the other two approaches. Consequently, linear regression is more suitable for skill level estimation than the other two approaches.

Prediction of a student’s final skill level, as soon as possible, is very important for adjusting the course and the course materials. Consequently, the lower the number of a learning objects needed to effectively predict the perform
figure of a student, the better the item to be used for skill 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 skill level prediction. The result show that by using two to three homework or project grades the skill of a student can be predicted. Using three to four grades can determine the skill level with high confidence.

As mentioned earlier, the grades for quizzes and labs do not have clear differentiation between all four skill levels and cannot be used for this purpose. Furthermore, four to six grades are needed to be able perform skill level classification. Consequently, those would not be used for this purpose.

V. DISCUSSION AND CONCLUSION

Skill level prediction is important because it would be used to adjust the learning materials to improve the learning experience of students. The result of skill level prediction can be used by a human instructor or by an intelligent agent in an intelligent tutoring system. In this paper a new method is proposed to determine the skill level as early as possible based on determining the sub-skill levels in learning objects such as homework and projects.

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

1. The student's skill level can be determined at early stages of a course based on a few important learning objects. In the case of our “Introduction to Computers and Programming”, 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 skill level.

2. The system can determine which learning object helps 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, is in using sub-skill levels rather than the overall skill levels. It should be mentioned that since normal distribution is used to model the skill level, at least 30 samples are needed for each level to correctly model the skill 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 learning objects and skill level. Furthermore, we will study the use of fuzzy logic to better represent the fuzziness in the data. Finally, the possibility of using the interaction of the user with system through non-assessment learning objects would be investigated.
PAPER

USING SUB SKILLS TO MODEL AND ESTIMATE FINAL SKILL LEVEL

Figure 7. Four levels of C&RT decision tree

Figure 8. CHAID decision tree

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

|               | was Expert | was Good | was Average | was Bad |
|---------------|------------|----------|-------------|---------|
|               | CHAID | C&RT | Regression | CHAID | C&RT | Regression | CHAID | C&RT | Regression | CHAID | C&RT | Regression |
| **Estimated** |      |      |            |      |      |            |      |      |            |      |      |            |
| **Expert**    | 64.29 | 82.14 | **96.60**  | 19.23 | 42.31 | 26.92      | 25.00 | 15.38 | 0.00       | 0.00  | 0.00  | 0.00       |
| **Good**      | 35.71 | 14.29 | 3.40       | 46.15 | 42.31 | **73.08**  | 50.00 | 15.38 | 15.38      | 6.25  | 6.25  | 0.00       |
| **Average**   | 0.00  | 0.00  | 0.00       | 34.62 | 15.38 | 0.00       | 33.33 | 7.69  | **38.46**  | 6.25  | 0.00  | 12.59      |
| **Bad**       | 0.00  | 0.00  | 0.00       | 0.00  | 0.00  | 0.00       | 0.00  | 61.54 | 61.54      | 87.50 | 93.75 | **87.50**  |
REFERENCES

[1] H. Liu, H. Shi, Y.S. hang, and S. Chen, “SAM: A student assessment and modeling system,” in Proceedings of international Conference on Artificial intelligence (IC-M‘03), June 2003.

[2] W.R. Murray, “A practical approach to Bayesian student modeling,” in Proceedings of the international Conference on intelligent Tutoring System, August 1998. http://dx.doi.org/10.1007/3-540-48716-5_48

[3] Wooff, B. “Artificial Intelligence in Education. Encyclopedia of Artificial Intelligence”, John Wiley & Sons: New York (1992) 434-444

[4] Conati, C., Gertner, A., VanLehn, K., “Using Bayesian networks to manage uncertainty in student modeling”, User Modeling and User-Adapted Interaction 12 (2002)371-417 http://dx.doi.org/10.1023/A:1021258506583

[5] P. Brusilovsky, “The construction and application of student models in intelligent tutoring systems”, Journal of Computer and System Sciences International, 32(1), 70-89 (1994).

[6] Nykänen, “Inducing Fuzzy Models for Student Classification”, Educational Technology & Society, . (2006).

[7] Hairong Liu; Hongchi Shi; Yi Shang; Su-Shing Chen,”Student modeling with timed assessment information” Information Technology: Research and Education, 2004. ITRE 2004. 2nd International Conference on , vol., no., pp. 121- 125, 28 June-1 July 2004

[8] James R. Nolan, “A prototype Application of Fuzzy Logic and Expert Systems in Education Assessment”, Proceedings Fifteenth National Conference on Artificial Intelligence (AAAI-98). Tenth Conference on Innovative Applications of Artificial Intelligence. AAAI Press/MIT Press, Menlo Park, CA, USA. pp. 1134-9; 1998.

[9] James R. Nolan, “An expert fuzzy classification system for supporting the grading pf student writing samples” Expert Systems with applications. Vol. 15, Iss. I , pp. 59-68; July 1998. http://dx.doi.org/10.1016/S0957-4174(98)00011-6

[10] Jian Ma and Dunning Zhou, “Fuzzy Set Approach to the assessment of Student-Centered Learning”, IEEE transaction on Education, Vol, 43, No. 2, May 2000. http://dx.doi.org/10.1109/13.848079

[11] J.R. Echauz and G. J. Vachtesvano; “Fuzzy grading system”, IEEE trans. Educ., vol. 38, no.2, pp. 158-164, 1995. http://dx.doi.org/10.1109/13.387218

[12] B. Ranjit, “An application of fuzzy set in students evaluation”, Fuzzy sets and systems, vol. 74, pp. 187-194, 1995.

[13] Delavari, N., Beizkadeh, M.R. & Shirazi, M.R.A. (2004). “A new model for using data mining in higher educational system”. 5th International Conference on Information Technology Based Higher Education and Training.

[14] Romero, C., Ventura, S., Espejo, P.G. & Hervs, C. (2008). “Data mining algorithms to classify students”. In The 1st International Conference on Educational Data Mining (EDM 2008), 8-17, Montréal, Canada.

[15] Tabanedd, Y. & Sami, A. (2010). “Classification of tutor system logs with high categorical Features”. In Proceedings of the KDD Cup 2010 Workshop on Improving Cognitive Models with Educational Data Mining.

[16] Yan, R., Liu, Y., Jin, R. & Hauptmann, A. (2003). “On predicting rare classes with svm ensembles in scene classification”. In The IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP 2003), III-21-4, IEEE Xplore.

[17] Gong, Y., Beck, J. & Heffernan, N. (2010a). Comparing knowledge tracing and performance factor analysis by using multiple model tting procedures. In V. Aleven, J. Kay & J. Mostow, eds., Intelligent Tutoring Systems, vol. 6094 of Lecture Notes in Computer Science, 35(44, Springer Berlin / Heidelberg.

[18] Baker, R., Pardos, Z., Gowda, S., Nooraei, B. & Heffernan, N. (2011). Ensembling predictions of student knowledge within intelligent tutoring systems. In J. Konstan, R. Conejo, J. Marzo & N. Oliver, eds., User Modeling, Adaptation and Personalization, vol. 6787 of lecture Notes in Computer Science, Springer Berlin / Heidelberg.

[19] Baker, R.S., Corbett, A.T. & Aleven, V. (2008a). Improving contextual models of guessing and slipping with a truncated training set. In The first International Conference on Educational Data Mining (EDM 2008).

[20] Baker, R.S., Corbett, A.T. & Aleven, V. (2008b). More accurate student modeling through contextual estimation of slip and guess probabilities in bayesian knowledge tracing. In Proceedings of the 9th International Conference on Intelligent Tutoring Systems, ITS ’08, 406-415, Springer-Verlag, Berlin, Heidelberg.

[21] Baker, R.S., Corbett, A.T., Gowda, S.M., Wagner, A.Z., MacLaren, B.A., Kaufman, L.R., Mitchell, A.P. & Giguere, S. (2010). Contextual slip and prediction of student performance after use of an intelligent tutor. In the 18th International Conference on User Modeling, Adaptation and Personalization (UMAP 2010), Lecture Notes in Computer Science, 52-63, Springer.

[22] Beck, J.E. & Chang, K.M. (2007). Identifiability: A fundamental problem of student modeling. In Proceedings of the 11th international conference on User Modeling, UM07, 137-146, Springer-Verlag, Berlin, Heidelberg.

[23] Bekele, R. & Menzel, W. (2005). A bayesian approach to predict performance of a student (bapps): A case with ethiopian students. In Proceedings of the International Conference on Artificial Intelligence and Applications, 189-194, Vienna, Austria.

[24] Loh, W.-Y. (2011). Classification and regression trees, Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, vol.1, 14-23 http://dx.doi.org/10.1002/widm.8

AUTHORS

Seyed Abbas Moradi is with the Advanced Robotics and Intelligent Systems Lab, School of ECE, College of Eng., University of Tehran, Tehran, Iran (e-mail: s.a.moradi@ut.ac.ir).

Hadi Moradi and Masoud Asadpour are with ARIS lab and Control and Intelligent Processing Center of Excellence, School of ECE, College of Engineering, University of Tehran, Tehran, Iran (e-mail: moradhi@ut.ac.ir, asadpour@ut.ac.ir).

This article is an extended and modified version of a paper presented at the the Interactive Mobile and Computer Aided Learning (ICML2012), held in November 2012, at Princess Sumaya University for Technology, Amman, Jordan. Received 26 January 2013. Published as resubmitted by the authors 18 March 2013.