Predicting Student Success Based on Prior Performance

Ahmad Slim, Gregory L. Heileman, Jarred Kozlick and Chaouki T. Abdallah

Department of Electrical and Computer Engineering
University of New Mexico
Albuquerque, NM 87131, USA
Email: {ahslim, heileman, jkozlick, chaouki}@unm.edu

***Abstract*—Colleges and universities are increasingly interested in tracking student progress as they monitor and work to improve their retention and graduation rates. Ideally, early indicators of student progress, or lack thereof, can be used to provide appropriate interventions that increase the likelihood of student success. In this paper we present a framework that uses machine learning, and in particular, a Bayesian Belief Network (BBN), to predict the performance of students early in their academic careers. The results obtained show that the proposed framework can predict student progress, specifically student grade point average (GPA) within the intended major, with minimal error after observing a single semester of performance. Furthermore, as additional performance is observed, the predicted GPA in subsequent semesters becomes increasingly accurate, providing the ability to advise students regarding likely success outcomes early in their academic careers.**

I. Introduction

Student success in higher education is under increasing scrutiny at both the state and federal levels [1]. This is driven by numerous factors, including the desire to improve institutional characteristics for rating purposes, the increasing trend of states tying institutional funding to student outcomes, as well as the fact that a bachelor's degree has become an increasingly necessary prerequisite for success in the work place—creating a moral imperative for colleges and universities to graduate the students they admit. Given these pressures, universities are collecting unprecedented amounts of information related to student performance and progress, and applying ever more sophisticated analytical techniques in efforts to determine the most important factors that contribute to attrition and persistence [2], [3]. Obviously, student progress within a degree program has a direct influence on graduation rates. Hence any effort aimed at enhancing or predicting the progress of a student in order to provide earlier advisement and/or interventions has the potential to positively impact graduation rates. Thus, in this paper we propose a probabilistic graphical model that allows us to reason about student performance and progress. In particular, we use a Bayesian Belief Network (BBN) model to represent the *curriculum graphs* of specific degree programs. Based upon the performance of a student in a given semester, we hypothesize that the BBN model can predict the "future" performance of the student in subsequent semesters. The model developed in this paper was applied to a number of different degree programs at the University of New Mexico (UNM), and was able to predict the GPA distribution of the students with minimal error.

The remainder of this paper is structured as follows. In Section II we provide background information and survey some related work in this research domain. In Section III we describe the theory behind the BBN machine learning technique employed in our framework. Section IV presents the details of our proposed framework. Implementation details and the decision-making algorithm are presented in Section V. Simulation methodology and results are presented in Section VI. Finally, Section VII presents some concluding remarks.

II. Background and Related Work

Many definitions of student success exist in the literature. While these vary from grades and persistence to self-improvement, most studies consider graduation the ultimate measure of student success [2]. For a college student, having a bachelor's degree has become a necessity with attainment rates topping 30% for adults over the age of 25 according to the latest census numbers [4]. From the university's perspective, and especially for public institutions, the definition of student success broadens from graduation into student retention rates and time-to-degree. These factors are important because many states have tied a percentage of the university's funding directly to such student success metrics [1]. This so-called "performance funding" has become a popular way to incentive universities to help students graduate in a timely fashion. Recent studies have addressed some new aspects associated with student progress analysis and curricular efficiency which have a direct influence on student success [5]–[7]. In [5], the authors developed various metrics related to curricular efficiency that correspond to the ease with which a student may satisfy the degree

978-1-4799-4518-4/14/$31.00 ©2014 IEEE

requirements associated with a given degree. These metrics were intended to measure the role that the structure of a curriculum plays in student academic success and accordingly suggest enhancement to the curriculum structure in an attempt to help students perform better in their respective academic programs. However this work did not take into account the performance of student in progress. In [6] and [7], the authors propose a *student progress ratio (SPR)* in order to study the progress of a student in a curriculum in each semester by investigating the structural properties of individual curricula, taking into account the degree to which individual courses in a curricula may impact student progress. This previous work takes into account the importance of particular courses in a curriculum, as well as the grades that students earn, in order to measure student progress. In this paper, we take into account this student progress in order to predict future performance.

III. Bayesian Belief Networks

A. Background

A BBN is a graphical structure that allows one to represent and reason about an uncertain domain. For a set of variables $X = (X_1, \ldots, X_n)$, a Bayesian network consists of a network structure S that encodes a set of conditional independence assertions about variables in X, and a set P of local probability distributions associated with each variable [8]. An example of a BBN which represents a subset of network behavior through variables namely, *Ahmad Oversleeps*, *Traffic* and *Ahmad Late* (as nodes) and two directed edges is shown in Fig. 1. An edge from one node to another implies a direct dependency between them, with a child and parent kind of relationship. To quantify the strength of relationships among the random variables, conditional probability functions are associated with each node, such that $P = \{p(X_1|\Pi_1), \ldots, p(X_n|\Pi_n)\}$ where Π_i is the parent set of X_i in X. If there is a link from X_i to X_j , then X_i is a parent of X_j and thus it belongs to Π_j . For discrete random variables the conditional probability functions are represented as tables, called Conditional Probability Tables (CPTs). For a typical node A , with parents $B_1, B_2, \ldots, B_n$, there is associated a CPT as given by $P(A|B_1, B_2, \ldots, B_n)$.

The main principle behind BBNs is Bayes rule:

$$P(H|e) = \frac{P(e|H)P(H)}{P(e)} \tag{1}$$

where $P(H)$ is the prior belief about a hypothesis H , $P(e|H)$ is the likelihood that evidence e results given H , and $P(H|e)$ is the posterior belief in the light of evidence e . This implies that belief concerning a given hypothesis is updated on observing some evidence.

B. Inference features

BBNs support three types of learning structural, parameter and sequential. The structure of the BBN can

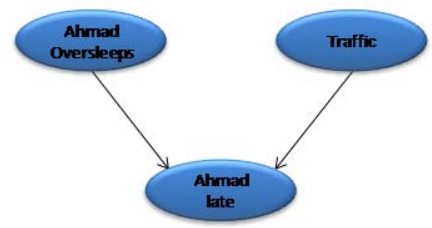

Fig. 1: An illustrative Bayesian Belief Network.

be constructed manually by a subject matter expert or through structure learning algorithms—PC and NPC algorithms [9], [10]. Parameter learning uses past data as the basis for learning the parameters through algorithms. One such algorithm, Expectation Maximization (EM), is particularly useful for parametric learning [11]. In order for the model to reflect behavior in the problem domain, the parameters of the model need to be updated based on observations. This process is termed sequential learning [12]. Evidence about a particular node is used to update the beliefs (posterior probabilities) of other nodes of the BBN. The BBN framework supports predictive and diagnostic reasoning and uses efficient algorithms for this purpose [13]. In this paper, predictive reasoning will be the main approach implemented in the BBN framework.

C. Application to our framework

The performance of a student in a given class can be used as a measure to predict competence or skills in later classes [14]. In other words, the history of a student's academic skills tells us something about future performance. For instance, an 'A' high school student is generally expected to do better in college than a 'C' student, other factors the same. Correspondingly, it makes sense that a college student who earns an 'A' in *Calculus II*, for instance, should be expected to earn a higher grade in *Calculus III* than those who earn a 'D'. In Fig. 2 , the application of BBN in the context of course network aimed at predicting the grades of the courses for a given student based on the evidence of previous grades, age, gender, educational level of parents, emotional factors, etc.

IV. BBN Characteristics for Curriculum Graphs

A. BBN Edges

Ultimately, degree attainment requires the satisfaction of all requirements associated with a degree program. The set of requirements associated with a particular degree program, along with the relationships between the individual requirements (e.g., course prerequisites) can be represented as directed acyclic graph, with a directed edge from node A to node B in the graph denoting that degree requirement A must be satisfied prior to the satisfaction of degree requirement B. Typically, a degree requirement is satisfied by passing a particular course,

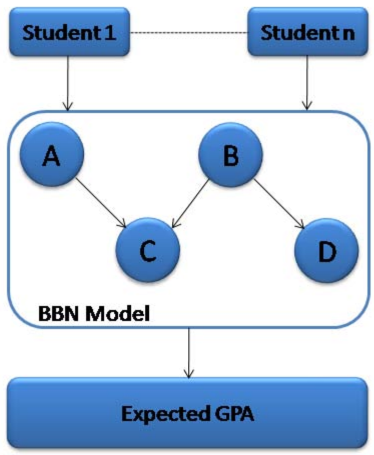

Fig. 2: BBN in the context of a course network.

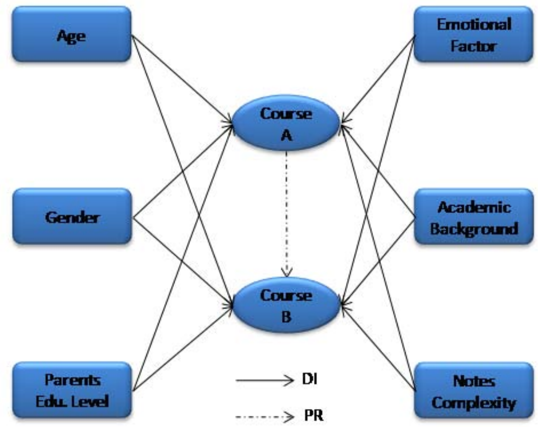

Fig. 3: BBN model of the curriculum graph.

and the precedence relationships in the graph correspond to course prerequisites. Thus, we will refer to graph as a *curriculum graph*. A student satisfies all degree requirements, and therefore receives the associated degree, once they have traversed this graph, visiting every node according to the precedence relationships in the graph. In our proposed framework,however, the edges of the BBN for the *curriculum graph* are not only restricted to prerequisite relationships. Basically a directed edge from node A to node B in the BBN of the *curriculum graph* denotes that the student performance in degree requirement A has a *direct influence* on predicting that of degree requirement B. Thus the presence of a *direct influence* edge between two requirements in the BBN does not imply the presence of a prerequisite relationship, however we hypothesize that the opposite is true. Accordingly—denoting a *direct influence* relationship as *DI* and a prerequisite relationship as *PR*— *PR* edges are a subset of *DI* edges that is $PR \subseteq DI$. In other words, we considered in this work that the presence of prerequisite relationships among the courses in the BBN indicate a *direct influence* on predicting the level of performance in the direction of the edge.

B. BBN Nodes

Basically the variables influencing the predication of the performance of a student in a given course are not only restricted to the performance of previous courses. Many factors, other than performance of previous courses, have direct impact on predicting "future" performance. Studies have shown that age and gender [15], [16], academic background [17], educational level of parents [18], emotional and social factors [19], and even the complexity measure of teacher's lecture notes [20] have direct influence on student progress. Typically, the BBN of a *curriculum graph* would be something similar to that illustrated in Fig. 3.

V. Implementation Aspects

A. Assumptions

For the purpose of a proof of concept, in this paper we present a basic network topology. In particular, the only variable (i.e., node) that will be illustrated in the BBN is the performance of a student in a "course" (i.e., no variables related to age, gender, educational level of parents, emotional factors, etc.), which will be a discrete variable. The states of the "course" variable are the letter grades associated with the "course" that is $A+, A, A-, B+, B, B-, C+, C, C-, D+, D, D-$ and F. Also it is assumed that the only edge type present in the BBN is the prerequisite relationship *PR*. Hence, in this work, we model the BBN of a curriculum C consisting of n degree requirements as a directed graph $G_C = (V, E)$, where each vertex $v_1,\ldots,v_n \in V$ represents a "course" in C, and there is a directed edge $(v_i, v_j) \in E$ from course v_i to v_j if v_i must be satisfied prior to the satisfaction of v_j . The final structure of the BBN for a *curriculum graph* will be something similar to that shown in Fig. 4.

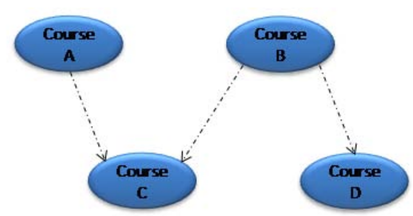

Fig. 4: BBN model of the curriculum graph implemented in our framework. Note that the "course" variable is the only node presented in this BBN model and *PR* edges are the only links relating these type of nodes.

B. Decision-Making Policy

To meet the objective of predicting the grades of the courses to be taken by a student attending a given degree program, we need to design a policy to assign a grade for the courses to be taken in the future once we determine their respective marginal probabilities based on the evidence of the grades of previously taken courses. Denote by $L = \{A+, A, A-, B+, B, B-, C+, C, C-, D+, D, D-, F\}$ the set of grade letters assigned to a course and $G = \{4.3, 4, 3.7, 3.4, 3, 2.7, 2.3, 2, 1.7, 1.4, 1, 0.7, 0\}$ the set of grades mapping L. For a course i, upon retrieving an evidence e, a decision is made using two methods:

1) Maximum a Posteriori Probability (MAP) estimate:

$$g = \underset{g \in G}{\operatorname{argmax}}\, p(g|e), \tag{2}$$

where $p(g|e)$ is the marginal probability of course i states based on evidence e which is the set of grades of previous courses.

2) Expected Grade (EG) estimate:

$$\hat{g} = \mathbb{E}(G) = \sum_{g} gp(g|e). \tag{3}$$

Note that one of the predicted letter grades for a given course might be F. In this case no data is available to fill in the CPTs due to the fact that students cannot move to another class if they fail its respective pre-requisite(s). For instance, we cannot fill a CPT row querying about the probability of a student earning a C on $CalculusIII$ conditioned on getting F on $CalculusII$. Simply, the student must get D or higher on $CalculusII$ to go for $CalculusIII$. In other words, the student has to repeat the course and pass it. To overcome this problem, we apply Markov chain model. The transition probabilities are graphically represented by the transition diagram shown in Fig. 5 with a 13 state Markov chain model representing the letter grades. Thus in case the BBN model predicts a F grade for a given course, it will use the Markov chain model to choose the letter grade (other than F) with the maximum transitional probability, that is

$$g = \underset{l \in L}{\operatorname{argmax}}\, p(l) \tag{4}$$

where $p(l)$ is the transition probability.

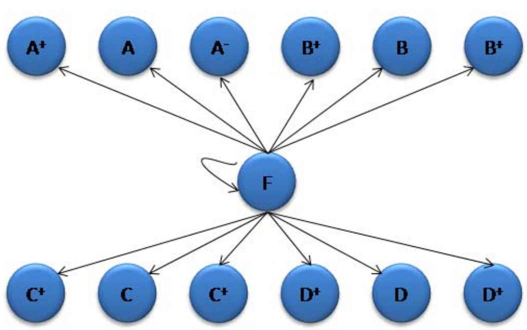

Fig. 5: A 13-state Markov chain model.

VI. Simulation Results

A. Simulation Setup

In an attempt to empirically validate our proposed BBN framework, we analyzed actual university data from the University of New Mexico (UNM)[1]. For this we used the data of 115,746 students to generate the CPTs for all the courses in the BBN. Then we chose 400 students, who already earned their degree programs, randomly from different departments (i.e., mechanical engineering department, chemical engineering department, electrical engineering department and nuclear engineering department) to test the framework. The performance of the framework was measured using mean squared error (MSE):

$$err^t = \frac{1}{n}\sum_{i=1}^{n}(\hat{Y}_i - Y_i)^2, \tag{5}$$

where err^t is the MSE measured based on the evidence of the grades of the courses taken between semester 1 and semester t; n is the number of students;$\hat{Y}$ is a vector of n GPA predictions; $Y$is the vector of the true GPA (the actual GPA values of the students).

B. Data Pre-processing

It is well known that building relationships between courses based on pre-requisite links is not trivial. For example a course i may be a co- or pre-requisite to another course j and/or vise-versa. In order to deal with such relationships, some assumptions were made:

1) If course i is a co- or pre-requisite to course j, we assume that i is a pre-requisite to j. In other words, we assume the worst case scenario where course i and j cannot be taken in the same academic term.
2) If course i is a co- or pre-requisite to j and vice-versa or in other words if courses i and j are co-requisites, we consider the worst case scenario in which one of the courses is considered to be the pre-requisite of the other. In this case we eliminate cycles from our graph.

C. Numerical Results

As mentioned previously, 400 student were chosen randomly from four different departments as our test collection. The courses taken by these students are spread over 22 semesters (i.e., 7 years). The MSE is measured at each semester where the grades of the courses taken by the students up until that semester are entered as evidence to the BBN framework. The MSE is measured using two different methods illustrated in Eq. (2) and Eq. (3) to calculate the predicted GPA vector $\hat{Y}$. To demonstrate the practicality of our approach, we compared our framework to another one where no

[1]All the UNM data used in this work are found at s3.amazonaws.com/employing-bayesian-belief-networks-for-course-networks/bbn-data.zip

edges are present. In other words, we generated another graph where we assumed that the performance of a student in one class does not have any influence on the performance of other classes (i.e. no edges are present). As in the BBN framework, we also measured the MSE using Eq. (2) and Eq. (3) to calculate the predicted GPA vector $\hat{Y}$. Note that no evidence is existing anymore in Eq. (2) and Eq. (3) regarding the second framework. Figure 6 illustrates the performance of both frameworks. This figure shows that the MSE values are decreasing gradually throughout the 22 semesters upon receiving new evidence e. The red curves show the MSE values for the BBN framework whereas the blue ones show those of the second framework. Besides, the dashed curves presents the MSE values using the MAP estimate method illustrated by Eq. (2) whereas the solid ones presents those using the EG estimate method illustrated by Eq. (3). Apparently, from the figure, it is seen that the MAP estimate method outperforms that of the EG estimate in both frameworks. Besides, the curves show that the BBN framework outperforms the other framework in both methods (i.e., MAP and EG). These results clearly illustrate the influence of a student's present performance on predicting his "future" performance. Using the BBN framework, upon receiving the grades of the first semester (i.e., evidence e), for instance, the MSE value (using MAP estimate) is measured to be 0.16. However, using the second framework, the MSE value is measured to be 0.76. On the other hand, comparing the MSE values for both frameworks, using the EG estimate method, upon receiving the grades of the first semester, shows a big gap. For semester one, the MSE value, using the BBN framework, is 0.37 whereas that, using the second framework, it is 1.94. On a scale of 4.3 (i.e., the maximum GPA value that can be achieved), it is obvious that the MSE value, using the second framework, is significantly high. This result illustrates the significance of the BBN framework in providing a better probability distribution, compared to the second framework, of the letter grades for a given course upon receiving an evidence e (i.e., marginal probability). Basically, the results show that the BBN framework gives students a more accurate prediction about the probability distribution of the letter grades for their "future" courses.

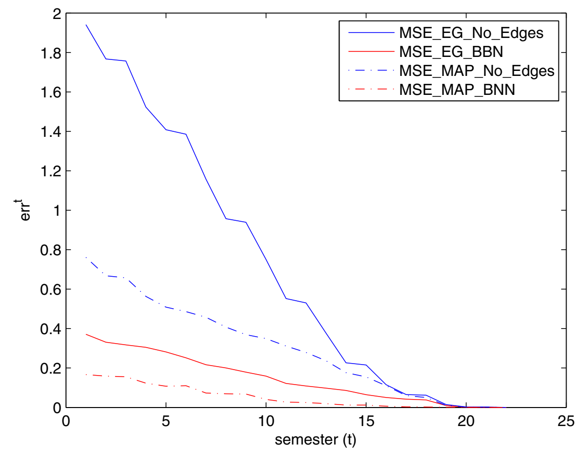

Fig. 6: MSE values of the two frameworks for 22 semesters with 3 semesters per year. The red curves show the MSE values for the BBN framework whereas the blue ones show those of the second framework (i.e. no edges). Besides, the dashed curves presents the MSE values using the MAP estimate method illustrated by Eq. (2) whereas the solid ones presents those using the EG estimate method illustrated by Eq. (3).

VII. Conclusion

The results shown in this paper demonstrate that BBNs can easily model a curriculum graph and can be used to predict the future progress of a student. It has been shown that a marginal error of 0.16 can be achieved upon receiving the grades of the first semester (i.e., evidence). Furthermore, the results show that the MSE decreases gradually upon receiving additional evidence, demonstrating the viability of the proposed BBN model. This initial work will be extended in the future to model multiple variables in the BBN model in addition to the "course" variable (e.g., student initial condition, age, gender, educational level of parents, emotional factors, instructor difficulty, etc.). We anticipate that this additional information will improve the performance of our framework. This application may prove useful in tracking the progress of students in order to provide an early intervene alert aimed out improving student outcomes.

References

[1] "Performance funding for higher education," National Conference of State Legislatures, Feb. 2012. [Online]. Available: http://www.ncsl.org/research/education/performance-funding.aspx

[2] A. Venezia, P. M. Callan, J. E. Finney, M. W. Kirst, and M. D. Usdan, "The governance divide: A report on a four-state study on improving college readiness and success," The National Center for Public Policy and Higher Education, San Jose, CA, Tech. Rep., Sep. 2005.

[3] L. Zhang, "Does state funding affect graduation rates at public four-year colleges and universities?" *Educational Policy*, vol. 23, no. 5, pp. 714–731, Sep. 2009.

[4] "Educational attainment in the united states," United States Census Bureau, 2012. [Online]. Available: http://www.census.gov/hhes/socdemo/education/data/cps/2012/tables.html

[5] J. Wigdahl, G. L. Heileman, A. Slim, and C. T. Abdallah, "Curricular efficiency: What role does it play in student success?" in *Proceedings of the the 121st ASEE Annual Conference and Exposition*. Indianapolis, Indiana, USA: IEEE, 2014.

[6] A. Slim, J. Kozlick, G. L. Heileman, J. Wigdahl, and C. T. Abdallah, "Network analysis of university courses," in *Proceedings of the 6th Annual Workshop on Simplifying Complex Networks for Practitioners*. Seoul, Korea: ACM, 2014.

[7] ——, "The complexity of university curricula according to course cruciality," in *Proceedings of the 8th International Conference on Complex, Intelligent, and Software Intensive Systems*. Birmingham City University, Birmingham, UK: IEEE, 2014.

[8] D. Heckerman, "A tutorial on learning with bayesian networks," Learning in Graphical Models, Tech. Rep., 1996.

[9] C. N. G. Peter Spirtes and R. Scheines, *Causation, Prediction, and Search*, second edition ed. MIT Press.
[10] H. Steck, "Constraint-based structural learning in bayesian networks using finite data sets."
[11] F. V. Jensen and T. D. Nielsen, *Bayesian Networks and Decision Graphs*, second edition ed. Springer Co.
[12] K. P. Murphy, "Dynamic bayesian networks: Representation, inference and learning," 2002.
[13] K. B. Korb and A. E. Nicholson, *Bayesian Artificial Intelligence*, second edition ed. CRC Press Co.
[14] A. T. Chamillard, "Using student performance predictions in a computer science curriculum," in *In ITICSE '06: Proceedings of the 11th Annual Conference on Innovation and Technology in Computer Science Education*, 2006, pp. 260–264.
[15] R. F. Deckro and H. W. Woundenberg, "Mba admission criteria and academic success," vol. 8, no. 4, pp. 765–769.
[16] ——, "Identifying factors that influence performance of non-computing majors in the business computer information systems course," vol. 21, no. 4, pp. 431–446.
[17] D. F. Butcher and W. A. Muth, "Predicting performance in an introductory computer science course," *Commun. ACM*, vol. 28, no. 3, pp. 263–268, Mar. 1985. [Online]. Available: http://doi.acm.org/10.1145/3166.3167
[18] S.-M. R. Ting and T. L. Robinson, "First-year academic success: A prediction combining cognitive and psychosocial variables for caucasian and african american students."
[19] J. Bennedsen and M. E. Caspersen, "Optimists have more fun, but do they learn better? : On the influence of emotional and social factors on learning cs and math," vol. 18, no. 1, pp. 1–16.
[20] K. J. Keen and L. Etzkorn, "Predicting students' grades in computer science courses based on complexity measures of teacher's lecture notes," *J. Comput. Sci. Coll.*, vol. 24, no. 5, pp. 44–48, May 2009. [Online]. Available: http://dl.acm.org/citation.cfm?id=1516595.1516605