Analysis of Semestral Progress in Higher Technical Education with HMM Models

Ewa Lach*[1][0000−0001−5380−6508], Damian Grzechca[1][0000−0003−1391−8809], Andrzej Polański[1][0000−0002−1793−9546], Jerzy Rutkowski[1], and Michał Staniszewski[1][0000−0001−9659−7451]

Faculty of Automatic Control, Electronics and Computer Science, Silesian University of Technology, Akademicka 16, 44-100 Gliwice, Poland
{Ewa.Lach, Damian.Grzechca, Andrzej.Polanski, Jerzy.Rutkowski, Michal.Staniszewski}@polsl.pl

Abstract. Supporting educational processes with Hidden Markov Models (HMMs) has great potential. In this paper, we explore the possibility of identifying students' learning progress with HMMs. Students' grades are used to train the HMMs to find out if the analysis of obtained models lets us detect patterns emerging from student's results. We also try to predict the final students' results on the basis of their partial grades. A new, classification approach for this problem, using properties of HMMs is proposed: High and Low State Model (HLSM).

Keywords: Education · Hidden Markov Model · Students Classification

1 Introduction

Mathematical modeling of educational processes is a wide area, which can concern different aspects of learning/cognitive processes, such as mastering of skills, strategies, behaviors, studies on influence of many factors on effectiveness of learning/teaching, analyses and design of plans for educational scenarios. It can also concern different types of educational processes and different levels of education. Mathematical or statistical modeling of learning/cognitive processes has been a part of researches in psychometrics, psychology and cognitive psychology, over decades (e.g. [1],[2],[3],[4]). There is also a lot of on-going research activity in modeling educational processes. New studies in the area of mathematical and statistical modeling of educational processes head towards including many explanatory variables of educational processes in one model, e.g., expertise, motivation, organization, discovering distinct phases of the learning progress, as well as distinguishing levels of students competence (e.g.[5],[6],[7],[8],[9]). Recently published approaches also include a wide range of different mathematical and statistical tools, among which applications of artificial intelligence and machine learning algorithms for developing educational programs and scenarios, are especially interesting [10]. Studies in the area of applying formal models for educational processes highly benefit from wide availability of electronic records now routinely kept in educational institutions. Supporting educational processes

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77967-2_18

with mathematical modeling obviously has a great potential. Research towards aiding educational practice with mathematical modeling is considerably improving human education. Mathematical modeling allows for better understanding the learning process, factors behind it and contributes to better defining and measuring learning achievements and indexes of student progress. Formulating mathematical models can also help developing educational plans towards personalizing, profiling with respect to background, motivation level [11]. There are, nevertheless, many challenges in developing and applying models of educational processes. There are numerous factors that can influence learning processes. It might be disputable whether the factors are indeed important or which are stronger than others. Both factors influencing learning and indexes of the quality of the learning process, mastery, practice, experience, are not directly observed [12]. Researchers must design systems and models for their estimations. Students groups in studies are always heterogeneous, which poses difficulties for formulating conclusions of experiments and for repeatability of research results. Acquisition of skills is a dynamic process, whose dynamics is not trivial and again can be very heterogeneous. Learning processes are different for different levels, types and targets of education. One of the very promising mathematical approaches for modeling learning processes is by using hidden Markov models (HMM), which allow for capturing the key elements of the learning or skills acquisition processes. States of HMM can represent unobservable states of students learning progress while emission matrices can represent different testing procedures, which generate observations available to the tutor/researcher. HMM can be used for predicting results of education/learning and for estimating dynamics of the process. A special version of the HMM model of learning progress is the Bayesian Knowledge Tracing model (BKT)[13], with two unobservable states. There are numerous studies devoted to different aspects of applications of HMMs for modeling educational processes [5],[7],[14],[15].

On the basis of the collected data (grades of students from technical university for selected courses) we have pursued a research involving using HMM for modeling educational progress of students. We have fitted HMM models for the data on semestral courses for students groups. We performed statistical analyses of the obtained HMM models and we compared them with simpler descriptive models of the students learning progress. We used elaborated models for predicting exam results. In particular, we proposed new classifier (based on HMMs): High and Low State Model (HLSM). Comparisons between models and statistical analyses allowed us to study some questions regarding modeling of educational processes: "Can one observe (in semestral/yearly) horizon the dynamics of the learning process by using HMM?", "How does the scenario of the tutoring process influence the process of skills acquisition?'.

2 Educational Data

In our research, we study achievements of the students from technical university: The Silesian University of Technology, the Faculty of Automatic Control, Elec-

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77967-2_18

tronics and Computer Science, from different disciplines of studies: Informatics, Biotechnology and interdisciplinary studies: Control, Electronics, and Information Engineering (CEIE). The results were gathered during selected technical curses. A characteristic feature of courses at technical university is very strong feedback given to students throughout their duration. The feedback is given by partial grades, i.e., scores of assignments, short tests and scores of reports from laboratory exercises. We use two types of scores: partial grades, collected weekly or every two, or three weeks during laboratories or exercise classes and semestral/final scores collected as either result of examination sessions or as results of one or two semestral credit tests. Both types of scores are routinely kept as an element of administrative documentation of the education activity of the Faculty of Automatic Control, Electronics and Computer Science. Much of the available data, especially older recoreds, is still stored on paper. However, there is a clear trend of switching to electronic recording, which will allow for the expansion of data related to education and increase the possibilities of their analysis.

2.1 Description of courses

In our work, we used data from the following five courses:

Introduction to System Dynamics (SD) is a one-semester course for third semester of CEIE, at the engineering level of studies. The course includes 15 lectures (30 hours) and 7 laboratories (15 hours). The course is taught in English. The final score is the result of the written exam. The aim of the course is making students familiar with problems and methods related to modeling physical dynamical systems and dynamical systems as engineering constructions. During the classes part students are obligated to solve three different tasks which are marked by a tutor: modeling physical and engineering systems by using the method of balances (2 classes), Lagrange equations I and II (2 classes) and electromechanical analogies I and II (1 or 2 classes depending on schedule).

Fundamentals of Computer Programming (FoCP) is a one–semester course taught in Polish at the engineering level, in the first year of Informatics studies. The course includes 15 lectures (30 hours) and 15 laboratories (30 hours). The final score is the result of the written exam. FoCP course provides knowledge required to understand, design and write computer programs in the C++ language.

Statistical Inference (SI) . It is a two-semester course given to the students of of the discipline Bio-engineering, specializing in bioinformatics, at the MSc. level of studies. The course includes 15 lectures (30 hours) and 15 laboratories (30 hours). The course is taught in Polish. Topics covered by the course include statistical learning theory, construction and validation of classifiers, as well as numerous examples of using statistical learning and classification in biotechnology and bioinformatics.

Introduction to Electric and Electronic Circuits (CT), formerly Circuit Theory Course. It is a two-semester course provided for students of CEiE,

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77967-2_18

at the engineering level of studies. The course is taught in English. The aim of the course is to provide students a trade-off between theory (mainly lectures), practical problems (computational problems solved on the classroom tutorials), and practice (physical observations and measurements in the laboratory). The laboratory exercises give an understanding of electric and electronic circuit connections – students translate circuit diagrams into real circuit connections. First semester of CT covers topics related to DC (Direct Current Circuit analysis), the second semester is mainly focused on AC (Alternating Current Analysis).

Biostatistics and Biometry (Bib). It is a one-semester course for students of the discipline Bio-engineering, at the engineering level of studies. The course includes 15 lectures (30 hours), 7 laboratories (15 hours) and 7 classsroom exercises (15 hours). The final score is the result (average) of two semestral credit tests. The course is taught in Polish. Topics covered by the course include distributions of random variables, data pre-processing, parameter estimation, parametric and non-parametric statistical tests.

2.2 Description of the collected data

SD. Data contain students' classroom exercises test marks recorded bi-weekly as well as final/semestral scores. Grades come from 206 different students attending the course over the years 2015–2018 with pass rate 78%. Students received partial grades from the eight-element set.

FoCP. Data contain results of students laboratory tests (10 short tests) recorded weekly as well as final/semestral scores. Grades come from 872 different students attending the course over the years 2016 - 2019 with pass rate 36%. Students received partial grades from the three-element set.

SI. Data contain results of students six laboratory tests as well as final/semestral scores. Grades come from 74 different students attending the course over one year (2015) with pass rate 49%. Students received partial grades from the five-element set.

CT. Data contain students' classroom exercises test marks recorded weekly (15 tests) as well as final/semestral scores. Grades come from 62 different students attending the course over one year (2016/17) with pass rate 35%. Students received partial grades from the twelve-element set.

Bib. Data contain students' classroom exercises tests marks recorded weekly as well as final/semestral scores. Grades come from 96 different students attending the course over one year (2015) with pass rate 45%. Students received partial grades from the eleven-element set.

3 Methodology of the analysis of the educational progress of students with the use of HMMs

HMMs are particularly suitable for analyzing the process of learning. They can model the hypothesis that students' levels of understanding of the material can

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77967-2_18

be represented by a number of hidden states. Finding the unobservable states of HMMs in which students can be during the learning process and estimating possible transitions between states, can shed light to the behavioral patterns emerging in education. The choice of the HMMs for modeling learning process is also concordant to the assumption that during the educational process, not only what grades are given to students is important, but also in which order they are given. Accordingly, the fact that a student gets grades: $\{3, 4, 5\}$, and not $\{5, 4, 3\}$ provides important (additional) information about the student's learning process and is often taken into account by the teacher.

The language of HMMs is as follows. N is a number of states, M–number of observations, $S = \{S_1, S_2, \dots, S_N\}$–set of states, $V = \{1, 2, \dots, M-1\}$–set of observations, T–sequence length, $Q = \{q_1; q_2; \dots; q_T\}$–sequence of states, $O = \{o_1; o_2; \dots; o_T\}$–sequence of observations. Hidden Markov models (HMMs) are stochastic models identifying processes unfolding over time: movement through a sequence of states Q, that are not directly observed. Each state is described by probabilities of observed variables, that for all states generate Emission probabilities matrix B:

$$b_{j(k)} = P(o_t = k | q_t = S_j) \tag{1}$$

,where $1 \leq t \leq T$, $1 \leq j \leq N$, $1 \leq k \leq M$, $\sum_{k=1}^{M} b_{j(k)} = 1$. State Transition Probability matrix A describes the likelihood of moving from one state to each other state in the next time period:

$$a_{ij} = P(q_t = S_j | q_{t-1} = S_i) \tag{2}$$

,where $1 \leq i, j \leq N$, $\sum_{j=1}^{N} a_{ij} = 1$. Π is a vector of initial distribution transitions probabilities. The figure 2 shows the matrices A and B of an illustrative six-states model. HMM can be estimated with the Baum-Welch algorithm [16] using the sequences of observable variables O. The Viterbi algorithm [17] can be used to estimate Q for selected HMM and O .

In our research, we use partial grades of students as available observations to estimate/train the HMMs to further find out if the analysis of obtained models lets us detect patterns emerging from student's results. As neither the true number of unobservable states nor values of model parameters (the matrices A and B) are known they are estimated on the basis of available observations using the Baum-Welch algorithm, which estimation quality depends on the initial values of parameters. For the first try, we use the same values for each parameter:

$$a_{ij} = 1/N, \quad b_{j(k)} = 1/M \tag{3}$$

If a Baum-Welch algorithm cannot differentiate between states, we set initial differentiating values. First, we change transition probabilities, and if it does not help we change emission probabilities. The second initial values for transition probabilities use the assumption that it is easier to stay in the same state than to move to another. We use:

$$a_{ij} = \begin{cases} 0.8 & \text{for } i = j \\ \frac{0.2}{N-1} & \text{for } i \neq j \end{cases} \tag{4}$$

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77967-2_18

To define the second proposition of initial values for emission probabilities, we assume, since observables are grades, that the probability that similar grades will occur together is greater than that there will be more distant grades in one state. Additionally, to solve the problem with the division of scores into states, if the obtained value are not an integer, greater variety among higher grades is assumed. We use:

$$b_{j(k)} = \begin{cases} \frac{2}{M+fMdN} & \text{for } \begin{array}{l} k > (fMdN*(j-1)+(M-(fMdN*N))) \text{ and} \\ k \leq (fMdN*j+(M-(fMdN*N))) \end{array} \\ \frac{1}{M+fMdN} & \text{for } \begin{array}{l} k \leq (fMdN*(j-1)+(M-(fMdN*N))) \text{ and} \\ k > (fMdN*j+(M-(fMdN*N))) \end{array} \end{cases} \tag{5}$$

, where $fMdN = floor(M/N)$ and function $floor(x)$ rounds the real value x to the next lower integer, in the $-\infty$ direction.

As part of the analysis of the educational process, we look at the state transition probability matrix A and emission probabilities matrix B of obtained HMM models, looking for patterns and rules. We compare HMMs for all students (HMMall), students who passed (HMMpass), and failed (HMMfail) courses.

We also compare the goodness-of-fit, which describes how well a model fits a set of observations, as assessed by the log-likelihood function (l). The function (l) is a logarithmic transformation of the likelihood function, that measure the probability that the model describe this particular data (sequences of observations). When fitting HMM models, it is possible to increase the likelihood by adding states, but doing so may result in overfitting. We compare log-likelihood for training and testing dataset.

The number of hidden states is estimated by multiple runs of the Baum-Welch algorithm with different values of N ($2 \leq N \leq M$) and using Akaike information criterion (AIC) [18] and Bayesian information criterion (BIC) [19]. BIC and AIC are used to compare models for a given set of data, trading-off between the goodness-of-fit of the model and the simplicity of the model. by introducing a penalty term for the number of parameters in the model. The penalty term is larger in BIC than in AIC.

$$AIC = (-2)L + 2k \quad BIC = (-2)L + ln(n)k \tag{6}$$

,where: L–maximum log-likelihood, k–number of independently adjusted parameters within the model, n–the sample size (T*(number of sequences)). The model with the lowest BIC or AIC criterion is preferred.

To test if adding states makes the HMM models overfitted in practice we use the cross-validation technique to randomly divide the data in half 200 times, calculating the mean L for the training and testing dataset for HMMs with different values of N ($2 \leq N \leq M$).

3.1 Results of fitting HMMs to students partial scores data

In this part of the work, we have fitted HMMs to all data on students' partial scores to find patterns and conclusions regarding students passing the courses.

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77967-2_18

First, initial values of transition probabilities needed to be set according to the second type of initial transition probabilities (eq. 4), because the training algorithm wasn't able to vary states with all probabilities set equally. For models with three and more states for courses SD and CT initial values for emission probabilities needed to be varied also (eq. 5). Without that emission probabilities for the second and subsequent states were identical.

Table 1 presents AIC and BIC calculated for HMMs of five analyzed courses. It shows that for our data adding states to models increases the value of AIC and BIC. Considering this selection criterion, we should use two-state models to analyze our data.

Table 1. BIC and AIC for HMM models

	FoCP		Bib		CT		SD		SI	
N	AIC	BIC	AIC	BIC	AIC	BIC	AIC	BIC	AIC	BIC
2	34.46	98.26	104.97	215.97	104.33	218.06	53.89	148.17	44.61	89.70
3	52.02	166.87	137.08	316.39	133.90	316.25	79.83	235.39	64.23	141.51
4	70.44	249.09	172.46	428.60	167.82	426.63	109.78	336.06	88.01	203.95
5			211.99	553.52	204.86	547.97	143.74	450.16	115.85	276.86
6			255.48	690.93	249.08	684.32	181.68	577.67	147.62	360.16
7			303.24	841.15	295.89	831.10	223.64	718.63		
8			354.93	1003.83	347.31	990.33	269.57	872.98		
9			410.79	1179.23	402.79	1161.46				
10			470.34	1366.85	462.41	1344.57				
11			534.10	1567.22	526.41	1539.89				
12			601.54	1779.81	593.33	1745.97				

The same result was obtained for experiments comparing the mean L for the training and testing set for successive N values. We can see in Figure 1(b-e) an increasing trend for L, calculated for the training set, and a decreasing trend for the testing set for all courses. The same behavior can be observed for the HMMs trained with data from earlier years and tested with data from subsequent years (Fig. 1a). So we can observe that with more states we get models that are better fitted for the training datasets, but less universal.

Finally, we looked at the matrices for HMMs with successive numbers of states. For two states, there is a clear division of emission probabilities for individual states. As the number of states increases, this division becomes blurred. With two states, one state has a high probability of having higher grades, the other state has a high probability of having lower grades. Adding additional states does not result in the emergence of intermediate grade distributions. Instead, there are states for which we cannot make conclusive deductions about students' abilities based on their grading probability distribution. Within one state there is, for example, the probability at the level of 0.1 - 0.3 for 1/3 - 2/3 grades (e.g. fig 2). There may also be states with one or two grades with high probability, but the likelihood of leaving this state is greater than for the other states (e.g. fig 2). We observe, also, a decreasing trend in the average probability of staying in the current state (a_{ii}) with a simultaneous increase in the

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77967-2_18

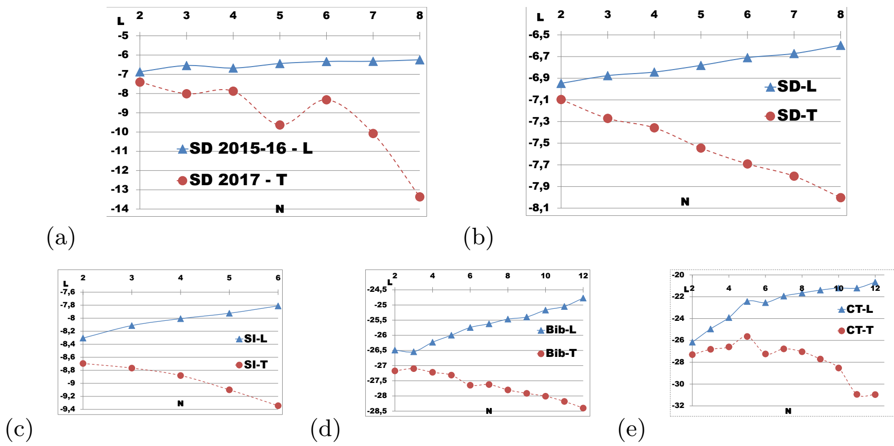


Fig. 1. Maximum log-likelihood values calculated for the training (suffix 'L') and testing (suffix 'T') datasets for a subsequent number of states of HMM models.

A:

state	1	2	3	4	5	6
1	0.153	0.555	0.000	0.097	0.195	0.000
2	0.000	0.751	0.000	0.000	0.000	0.249
3	0.000	0.000	1.000	0.000	0.000	0.000
4	0.000	0.289	0.000	0.711	0.000	0.000
5	0.000	0.000	0.068	0.000	0.932	0.000
6	0.000	0.000	0.000	0.000	0.000	1.000

B:

grade \state	**0**	**2.0**	**2.5**	**3.0**	**3.5**	**4.0**	**4.5**	**5.0**
1	0.000	0.000	0.194	0.000	0.000	0.793	0.000	0.012
2	0.000	0.000	0.029	0.197	0.276	0.215	0.190	0.093
3	0.006	0.362	0.076	0.317	0.000	0.000	0.238	0.000
4	0.000	0.125	0.023	0.126	0.725	0.000	0.000	0.000
5	0.045	0.000	0.027	0.295	0.212	0.418	0.003	0.000
6	0.016	0.000	0.052	0.010	0.101	0.198	0.300	0.323

Fig. 2. State Transition Probability matrix A and Emission probabilities matrix B for six-sates HMM model for course SD.

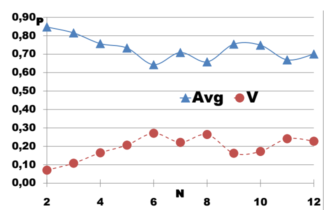


Fig. 3. Average probability for all courses (Avg) of staying in the current state and its variance (V) for a subsequent number of HMM states.

variance of a_{ii} (Fig. 3). This means frequent movement between states with a large variety of probabilities describing the transition between states. There are states in which the probability of staying is close to 0 and states from which it is impossible to get out ($a_{ii} = 1$).This means that, from an analytical point of view, it is difficult to establish any patterns in student behavior for HMMs with more states.

The above results prompted us to focus on the analysis of two-state HMMs. For each course, the two-state model divides the student skills into the high state HS (with a high probability of obtaining higher grades) and the low state LS (with a high probability of obtaining poor grades). Regardless of the course type, number, and type of intermediate grades, we can observe the presence of the high state and low one .

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77967-2_18

Table 2. Values for courses' HMMs.

course	FoCP			SD			Bib			SI			CT			FoCP				SD			
	All	Pass	Fail	All	Pass	Fail	All	Pass	Fail	All	Pass	Fail	All	Pass	Fail	2016	2017	2018	2019	2015	2016	2017	2018
GHPr	0.70	0.80	0.59	0.73	0.72	0.52	0.96	0.98	0.63	0.80	0.95	0.57	0.86	0.96	0.71	0.72	0.72	0.66	0.73	0.93	0.81	0.77	0.74
GLPr	0.92	0.75	0.97	0.90	0.93	0.87	0.61	0.58	0.94	0.95	0.84	0.92	0.94	0.82	0.98	0.89	0.95	0.92	0.93	0.91	0.88	0.79	0.94
HS->HS	0.92	0.94	0.91	1.00	1.00	0.70	0.64	0.67	0.95	0.84	0.80	0.95	0.86	0.90	0.78	0.94	0.90	0.94	0.89	1.00	1.00	1.00	1.00
LS->LS	0.95	0.92	0.96	0.83	0.77	0.85	0.83	0.81	0.76	0.76	0.78	1.00	0.98	0.93	0.96	0.95	0.94	1.00	0.94	0.60	0.57	0.66	0.84
L	-7.89	-6.68	-8.31	-6.97	-6.83	-7.10	-26.65	-25.97	-26.82	-8.39	-7.85	-8.48	-24.74	-26.72	-22.66	-7.79	-8.05	-7.80	-7.82	-8.08	-9.55	-6.78	-8.26
L-var	7.39	7.58	5.57	1.61	1.50	1.57	9.56	5.37	9.10	2.82	5.45	1.64	69.07	58.98	82.31	7.23	7.80	7.30	7.38	3.33	1.86	0.79	3.27

In order to compare two-state HMMs, additional values were introduced: $GHPr$ (the probability of good grades in HS) and $GLPr$ (the probability of poor grades in LS). These values are calculated by dividing the emission matrix B into two parts and summing the probabilities of the columns from the different parts for each state: for HS for good grades, for LS for poor grades.

We can observe similarities and differences in student behavior for different courses (table 2). For example, the $GHPr$ and $GLPr$ values are similar for 4 out of 5 courses. Except for Bib, the $GLPr$ value is greater than the $GHPr$ and is around 0.93, which tells us that in LS, a student has a high chance of getting a poor grade (around 0.93). This chance is greater than the chance of getting a good grade in HS (for two courses it is around 0.71, for the other two it is around 0.83). As a participant of the Bib course, student has a good chance of a good grade in HS (0.96) and as much as 0.39 in LS. This shows that students get many more good grades during Bib course and average Bib students have a greater chance of being assigned to LS, than for the rest courses. The likelihood of a transition between these states varies with course. For FoCP, students have a high probability of remaining in the current state (about 0.93); for SD there is a 0.17 probability of a transition to HS from LS and no possibility to leave HS; for CT, there is a 0.14 probability of going to LS from HS and only 0.02 of leaving the LS. BiB and SI students have more options for transitions between states. This may be influenced by the existence of dependencies between subsequent graded tasks (without understanding the previous graded task, it is impossible to understand the next topic) or the lack of it (each mark grades a separate topic) and the ease of catching up. Looking at the variance of L for all courses, we can assume that the behavior of the students of CT is more varied then for rest courses. The comparison of the values of $GHPr$, $GLPr$, state transition probabilities and L for HMMs from individual years for two courses: FoCP and SD shows that students behave similarly within the same courses (tab. 2).

Interesting conclusions can be drawn from the comparison of HMMs for all students, students who passed, and failed courses (tab. 2). It can be expected that: the $GHPr$ will be higher for HMMpass and lower for HMMfail, and the opposite for the $GLPr$, the probability of staying in HS will be higher for HMMpass and lower for HMMfail, and the opposite for the probability of staying in LS. As for L, greater fitness for HMMpass and HMMfail is expected. The results, however, are not so obvious. The above-described situation occurs only for the FCoP, for the others, there are greater or lesser deviations from the expected results. This would indicate that, within a given course, passing it or not more or less can be explained by partial grades. On this basis, we can

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77967-2_18

assume that other factors determine to a greater extent the passing or failing of some courses. As for the HMMs log-likelihood - L, some models achieved a slight improvement, but also some models received a lower L. On this basis, we can conclude that there is a variation in behavior among students who pass or fail a course (this is especially visible for students who fail: 4 out of 5 courses recorded lower L for HMMfail).

4 Predicting students final results

Analyzing student partial grades the problem of predicting with them student's exam results was explored. In order to prepare a comprehensive comparison of classification results, five different classical approaches plus a new High and Low State Model (HLSM), based on results from section 3, were applied.

The obtained course data is randomly split in the proportion 67:33 200 times [24]. 67 percent of the data is used as a training dataset, the rest as a testing dataset. In addition, we check whether models trained on data from previous years for a specific course can be used to predict exam results in subsequent years for that course.

4.1 Models

In order to prepare comprehensive comparison of classification results, five different classical approaches were applied.

1. Support vector machine (SVM) is mainly used for two-class classification. SVM finds classes by applying the best hyperplane to distinguish one class from another (by means of largest margin). In this context, margin refers to the maximum width of a plate parallel to the hyperplane without internal data points. SVM was applied with the Gaussian kernel function [20].
2. K nearest neighbour (KNN) can give the k closest points in the feature points. Classification results are based on a neighbour's plural voting, where the object is assigned to the class most frequently found among its k closest neighbors. The number of neighbours in the algorithm was set to one with a single standardization [21].
3. Linear regression (LR) models can be treated as a linear relationship of response with respect to one or more predictive factors. LR was applied along with binomial distribution [22].
4. Naive Bayes classifier (NB) in statistics is known as a probabilistic classifier which is based on Bayes theorem assuming strong (or rather naive) independence between given features [23].
5. HMM classic classification method (HMMC) consists of training separate HMMs for each class. The tested object is assigned to the class (HMM model) for which the log-likelihood is the highest [7] .
6. High and Low State Model (HLSM) is a new classification model proposed by the authors for the selected problem. It is based on the analysis of two-state HMM models trained on the partial grades received by students during

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77967-2_18

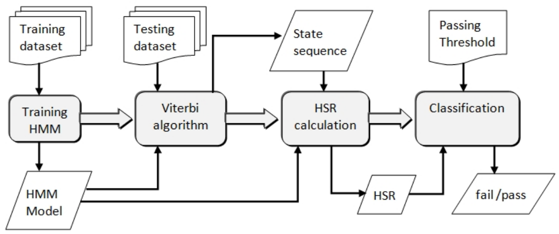


Fig. 4. HLSM classification process.

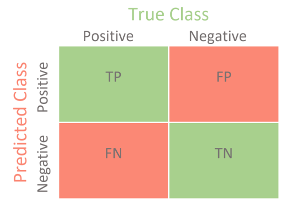


Fig. 5. The Confusion Matrix.

specified technical courses. For each of the analyzed courses, the two-state HMM models were characterized by one state with an assigned high probability of receiving higher partial grades and a low probability of receiving lower partial grades (HS), and a second state with opposite probabilities (LS). The proposed HLSM model assumes that by finding out the degree of membership to a HS of a student's behavior, we can determine the chance of a student to pass an exam. The classification process starts with the training of the two-state HMM model. Then, as a result of comparing the probabilities in the emission matrix for the extreme one or two grades, the high and low states are identified. Next, we use the Viterbi algorithm to calculate for a given sequence of students' grades, the most likely path (sequence of states) through the trained HMM model. For a generated sequence of states, we calculate how many times the student is in a high state in relation to all states in which he is staying during the entire course: High State Rate (HSR). Finally, we use HSR to predict whether the student will pass or not the final exam comparing this value to the passing threshold (PT). The diagram in the figure 4 shows the whole classification process.

For all classifying models, we can define four parameters that assess their performance: True Positive (TP), False Positive (FP), True Negative (TN), False Negative (FN) as presented in the confusion matrix (fig. 5). Unfortunately, none of these values can be used alone to judge the quality of a classifier. This results in the definition of various measures that are using TP, FP, TN, and FN values in different configurations. In the paper, we will use the area under the ROC curve (AUC) that can give an idea about the usefulness of the classifier in general. Greater AUC means a better model for a specific classification. The ROC (Receiver Operating Characteristic) curve (fig. 6(b)) shows in a graphical way the trade-off between sensitivity and specificity for every possible cut-off, where:

$$Sensitivity = TP/(TP + FN), \quad Specificity = TN/(TN + FP) \tag{7}$$

In addition to comparing models, the ROC curve was used, also, to determine the cut-off point for HLSM: passing threshold (PT). We use Youden's index J.

$$J = Sensitivity + Specificity - 1 \tag{8}$$

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77967-2_18

4.2 Results

During the conducted experiments the highest AUC values were obtained by the LR and HLSM models (fig. 6). For three courses (SD, FoCP and SI) LR model gets the highest results, for two (CT and Bib) - HLSM. For the course SD, the lead of a model LR is very small (0.003) and when we look at the classification for the following semesters of SD, we can see that HLSM has the best results for the three years, and only in one semester LR is doing better. If we average the results for subsequent years of SD, HLSM has a lead of 0.049. For FoCP and SI, LR has a lead (about 0.023), higher than for SD. The analysis of AUC for the following semesters of FoCP shows LR leading for 3 out of 4 semesters (0.028 on average). Only in 2019, HLSM has an AUC higher by 0.002. However, for CT and Bib, HLSM has a greater difference in AUC value to LR (Bib-0.178, CT-0.099). Because of that, when averaging the results for all courses, HLSM has the highest value. The situation is similar when predicting behaviors of students based on models generated for previous years. The LR and HLSM models achieve the best results alternately. When averaging the results LR is better for the course SD, HLSM for FoCP.

When comparing the predictions based on the models for previous years for course SD, the HMMC model presented interesting behaviors (out of 9 classifications, it achieves the best results twice, four times the lowest). Ultimately, if we compare the averaged values of the classification for models for previous years for the SD course with the result of the classification for SD, HMMC recorded the largest decrease in AUC value. In the remaining experiments, the HMMC obtained positive results, but never exceeded the HLSM.

The courses are also characterized by different values of AUC. For SD, Bib, and SI, the AUC is quite low (0.61-0.64) for FoCP it is higher (0.747) and CT has the best chance of a good classification (0.831). This is in line with the correlation calculated between the partial grades and the passing of the course (accordingly: Bib-0.19, SD-0.27, SI-0.28, FoCP-0.42, CT-0.62).

To determine if our data wasn't skewed by chance, we calculated statistical significance for the AUC values (t-test and Wilcoxon rank-sum test) and obtained results indicating that our experiments are statistically significant at the 0.05 significance level.

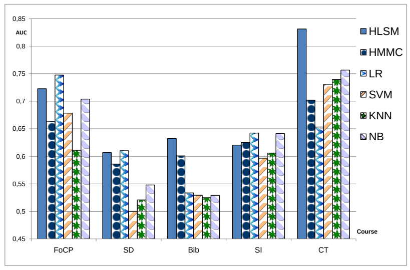
5 Conclusion

When looking at the trained HMM models for different courses, it can be noticed that, while one can always distinguish high and low states for each course, the models differ significantly. As shown in section 3, it is possible to distinguish patterns for different courses based on generated HMMs. These observations can be interpreted that by using HMM models, we can draw conclusions about the educational process in individual courses.

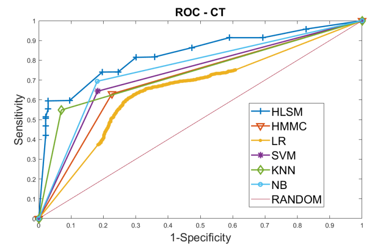
The performed experiments also showed that it is possible, to some extent, to predict the students' final results based on their partial grades. In particular, the proposed new classifier, based on HMMs, HLSM achieved good results.

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77967-2_18

(a)



(b)



	FoCP	SD	Bib	SI	CT	AVG (courses)	FoCP16	FoCP17	FoCP18	FoCP19	AVG (FoCP)	SD15	SD16	SD17	SD18	AVG (SD)
HLSM	0.723	0.607	0.632	0.620	0.831	0.683	0.731	0.666	0.682	0.799	0.719	0.616	0.727	0.636	0.661	0.660
HMMC	0.664	0.586	0.601	0.625	0.702	0.636	0.721	0.611	0.682	0.718	0.683	0.596	0.695	0.592	0.660	0.636
LR	0.747	0.610	0.533	0.642	0.653	0.637	0.740	0.685	0.739	0.797	0.740	0.519	0.645	0.593	0.689	0.612
SVM	0.678	0.500	0.529	0.597	0.731	0.607	0.661	0.559	0.611	0.725	0.639	0.506	0.589	0.505	0.587	0.547
KNN	0.611	0.521	0.525	0.606	0.739	0.600	0.603	0.524	0.598	0.714	0.610	0.507	0.500	0.500	0.556	0.516
NB	0.704	0.548	0.529	0.641	0.756	0.636	0.699	0.661	0.702	0.734	0.699	0.525	0.688	0.533	0.662	0.602

	FoCP									
	16->17	16->18	16->19	17->18	17->19	18->19	16/17->18	16/17->19	16/17/18->19	AVG
HLSM	0.667	0.714	0.783	0.719	0.793	0.757	0.718	0.800	0.791	0.749
HMMC	0.609	0.692	0.768	0.666	0.667	0.703	0.695	0.681	0.672	0.684
LR	0.689	0.689	0.827	0.717	0.681	0.765	0.708	0.791	0.802	0.741
SVM	0.636	0.592	0.721	0.544	0.534	0.638	0.691	0.703	0.643	0.634
KNN	0.587	0.586	0.665	0.508	0.517	0.613	0.617	0.621	0.651	0.596
NB	0.638	0.721	0.716	0.710	0.729	0.681	0.724	0.742	0.725	0.710

	SD									
	15n->16	15->17	15->18	16->17	16->18	17->18	15/16->17	15/16->18	15/16/17->18	AVG
HLSM	0.411	0.569	0.621	0.579	0.698	0.629	0.639	0.537	0.647	0.592
HMMC	0.387	0.481	0.667	0.526	0.437	0.507	0.569	0.719	0.389	0.520
LR	0.597	0.495	0.438	0.702	0.649	0.730	0.543	0.604	0.744	0.611
SVM	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500
KNN	0.500	0.500	0.500	0.500	0.500	0.413	0.486	0.515	0.533	0.494
NB	0.500	0.486	0.515	0.486	0.659	0.574	0.500	0.548	0.524	0.532

Fig. 6. AUC values calculated for different classifiers for five courses and for following semesters of SD and FoCP and an illustrative ROC curve for course CT (b).

Acknowledgment

This publication was supported by the Department of Graphics, Computer Vision and Digital Systems, under statue research project (Rau6, 2021), Silesian University of Technology (Gliwice, Poland). The authors would like to thank for involvement in data collection especially to prof. J. Polańska, prof. S. Deorowicz, prof. K. Simiński, dr F. Binczyk, dr A. Papież and dr J. Żyła.

References

1. Brainerd,C.J.: Markovian interpretations of conservation learning. Psychological Review, vol. 86, no.3, pp. 181–213 (1979) https://doi.org/10.1037/0033-295X.86.3.181
2. Brainerd,C.J.:Developmental invariance in a mathematical model of associative learning. Child Development,51(2),349–363(1980) https://doi.org/10.2307/1129267
3. Brainerd,C.J.: Three-state models of memory development: A review of advances in statistical methodology. Journal of Experimental Child Psychology, vol. 40, no.3, pp. 375–394 (1985) https://doi.org/10.1016/0022-0965(85)90072-4
4. Cleeremans A., McClelland, J. L.: Learning the structure of event sequences. Journal of experimental psychology. General vol. 120, no. 3, pp. 235–253(1991) https://doi.org/10.1037/0096-3445.120.3.235
5. Geigle, C., Zhai, C.: Modeling Student Behavior With Two-Layer Hidden Markov Models. JEDM, vol.9, no.1, pp.1–24 (2017) https://doi.org/10.5281/zenodo.3554623
6. Kassarnig, V. et al.: Academic performance and behavioral patterns. EPJ Data Science, vol. 7, no. 10, (2018) https://doi.org/10.1140/epjds/s13688-018-0138-8
7. Witteveen, D., Attewell, P.: The College Completion Puzzle: A Hidden Markov Model Approach. Res High Educ, vol. 58, pp. 449–467 (2017) https://doi.org/0.1007/s11162-016-9430-2

8. Tenison, C., Anderson, J.R.: Modeling the Distinct Phases of Skill Acquisition. Journal of Experimental Psychology: Learning, Memory, and Cognition, vol. 42, no. 5, pp. 749–767 (2016) https://doi.org/10.1037/xlm0000204
9. Logue A.W., Douglas, D., Watanabe-Rose, M.: Corequisite Mathematics Remediation: Results Over Time and in Different Contexts. Educational Evaluation and Policy Analysis, 41(3), 294–315 (2019) https://doi.org/10.3102/0162373719848777
10. Chi, M.T.H. et al.: Translating the ICAP Theory of Cognitive Engagement Into Practice. Cognitive Science vol. 42, pp. 1777–1832 (2018) https://doi.org/10.1111/cogs.12626
11. D'Mello, S., Olney, A., Williams, C., Hays, P.: Gaze tutor: A gaze-reactive intelligent tutoring system. International Journal of Human - Computer Studies, vol. 70, no. 5, pp. 377–398 (2012) https://doi.org/10.1016/j.ijhcs.2012.01.004
12. Beck, J.E., Chang, K.: Identifiability: A Fundamental Problem of Student Modeling. In: Conati, C., McCoy, K., Paliouras, G. (eds) User Modeling 2007, LNCS, vol. 4511, pp 137–146 (2007) https://doi.org/10.1007/978-3-540-73078-1_17
13. Corbett, A.T., Anderson, J.R.: Knowledge tracing: Modeling the acquisition of procedural knowledge. User Model. User-Adap Inter, vol. 4, pp. 253-–278 (1994) https://doi.org/10.1007/BF01099821
14. Wang G., Tang Y., Li J., Hu X.: Modeling Student Learning Behaviors in ALEKS: A Two-Layer Hidden Markov Modeling Approach. In: Penstein Rosé C. et al. (eds) Artificial Intelligence in Education. AIED 2018, LNCS, vol 10948 (2018) https://doi.org/10.1007/978-3-319-93846-270
15. Hu Q., Rangwala H.: Course-Specific Markovian Models for Grade Prediction. In: Phung D., Tseng V., Webb G., Ho B., Ganji M., Rashidi L. (eds) Advances in Knowledge Discovery and Data Mining. PAKDD 2018, LNCS, vol 10938 (2018) https://doi.org/10.1007/978-3-319-93037-43
16. Baum, L.E., Petrie, T.: Statistical Inference for Probabilistic Functions of Finite State Markov Chains. The Annals of Mathematical Statistics, vol. 37, no. 6, pp. 1554–1563 (1966) https://doi.org/10.1214/aoms/1177699147
17. Viterbi, A.: Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. IEEE Transactions on Information Theory, vol. 13, no. 2, pp. 260-269 (1967) https://doi.org/10.1109/TIT.1967.1054010
18. Akaike,H.: A new look at the statistical model identification. IEEE T Automat Contr, vol.19, no.6, pp.716–723(1974) https://doi.org/10.1109/ TAC.1974.1100705
19. Schwarz, G.: Estimating the dimension of a model. Annals of Statistics, vol. 6, no.2, pp. 461–464 (1978) https://doi.org/10.1214/aos/1176344136
20. Christianini, N., Shawe-Taylor, J.: An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods. Cambridge University Press, Cambridge, UK (2000) https://doi.org/10.1017/CBO9780511801389
21. Altman, N.S.: An introduction to kernel and nearest-neighbor nonparametric regression. The American Statistician, vol. 46, no. 3, pp. 175–185 (1992) https://doi.org/10.1080/00031305.1992.10475879
22. Collett, D.: Modeling Binary Data. New York: Chapman and Hall, 2002.
23. Trevor, H., Tibshirani, R., Friedman, F: The Elements of Statistical Learning: Data Mining, Inference, and Prediction. 2nd ed. Springer Series in Statistics (2009) https://doi.org/10.1007/978-0-387-84858-7
24. Xu, Y., Goodacre, R.: Splitting Training and Validation Set: A Comparative Study of Cross-Validation, Bootstrap and Systematic Sampling for Estimating the Generalization Performance of Supervised Learning. Journal of Analysis and Testing, vol. 2, pp. 249–262 (2018) https://doi.org/10.1007/s41664-018-0068-2

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77967-2_18