Opportunities, Challenges and Methods for Higher Education Data Mining

Asmaa Elbadrawy[3] , Huzefa Rangwala[1],
Aditya Johri[2] and George Karypis[3]

[1]Computer Science & Engineering, George Mason University, VA

[2]Information Sciences &Technology, George Mason University, VA

[3]Computer Science &Engineering, University of Minnesota, MN

SIAM International Conference on Data Mining, April 2017

Tutorial Outline

- Part I: Introduction
  - Background
  - Big Problems in Higher EDM
- Part II: Problems and Methods
  - Knowledge Modeling
  - Performance Prediction
    - Next-Term Grade Prediction
    - In-Class Assessment Prediction
    - Drop-Out Prediction
  - Degree Planners
- Part III: Case Studies
  - SmartGPA
  - Academic Pathways
- Part IV: Pertinent Challenges

Part I: Introduction and Background

Educational Data Mining

"Educational Data Mining is an emerging discipline, concerned with developing methods for exploring the unique types of data that come from educational settings, and using those methods to better understand students, and the settings which they learn in."

www.educationaldatamining.org

Educational Data Mining (EDM)

- Emerged in the past two decades due to the large volume of educational data that was made available

- Concerned with developing and applying Data Mining (DM) methods to detect patterns in large amounts of educational data.

- Sources of Educational Data
  - K-12
  - Universities and Colleges
  - Open-Courseware and MOOCs
  - Informal Education and Learning
    - Museums and Online Communities

Tutorial Focus: Higher Education

TABLE 1: LEARNING AND ACADEMIC ANALYTICS

TYPE OF ANALYTICS	LEVEL OR OBJECT OF ANALYSIS	WHO BENEFITS?
Learning Analytics	**Course-level:** social networks, conceptual development, discourse analysis, "intelligent curriculum"	Learners, faculty
	Departmental: predictive modeling, patterns of success/failure	Learners, faculty
Academic Analytics	**Institutional:** learner profiles, performance of academics, knowledge flow	Administrators, funders, marketing
	Regional (state/provincial): comparisons between systems	Funders, administrators
	National and International	National governments, education authorities

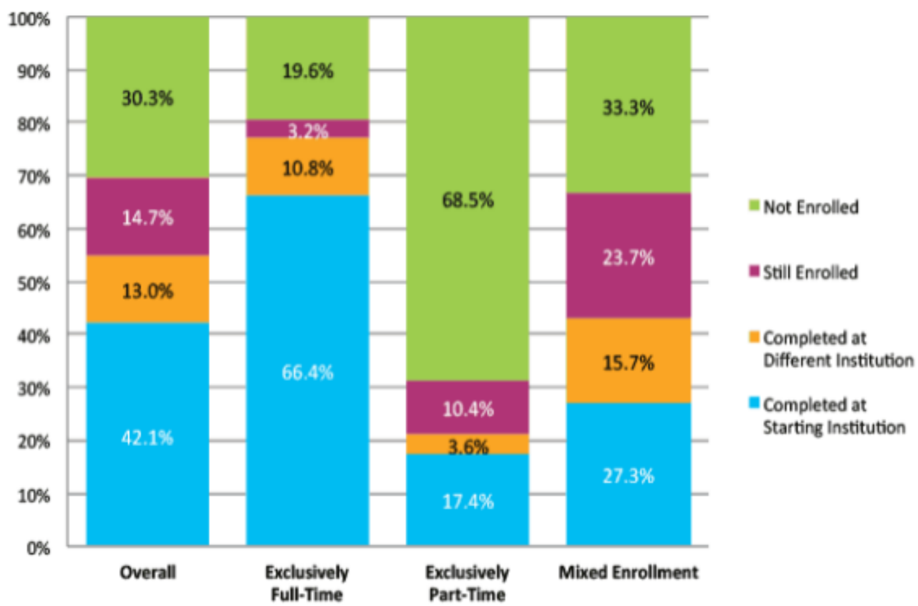
Long & Siemens (2011): http://er.educause.edu/articles/2011/9/penetrating-the-fog-analytics-in-learning-and-education

U.S. Higher Education Crisis

- High college drop-out rates
- Six years to finish a four-year college degree
- Inefficient college advising & lack of a clear enrollment plan

Figure 5. Six-Year Outcomes by Enrollment Intensity (N= 2,668,614)

(Click to Enlarge)



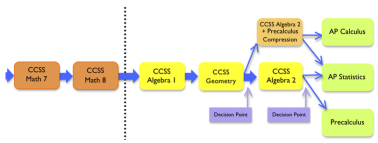
References:

- **Completing College: A National View of Student Attainment Rates. Report by [National Student Clearinghouse]**
- **Breaking the 4-year myth: Why students are taking longer to graduate. Article at [http://college.usatoday.com]**

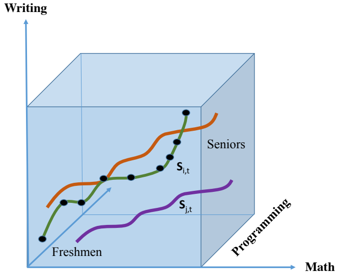
Career Planners

Dropout Prediction

Higher Educational Data Mining



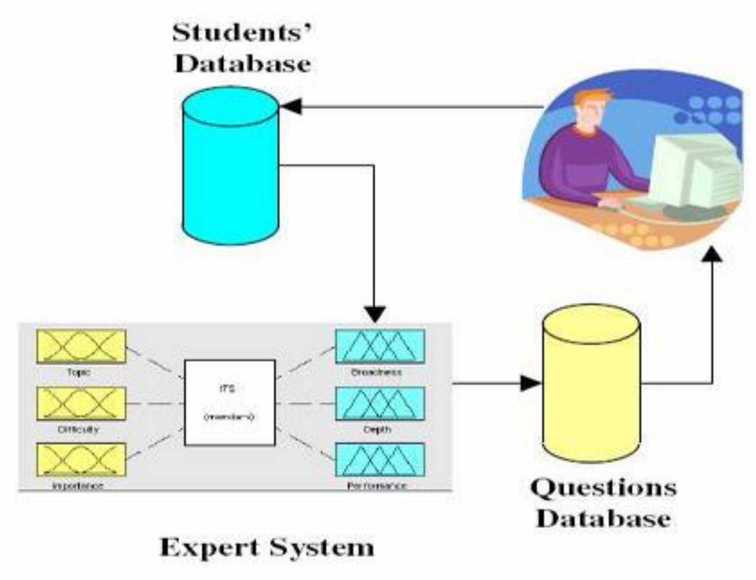
Degree Planners



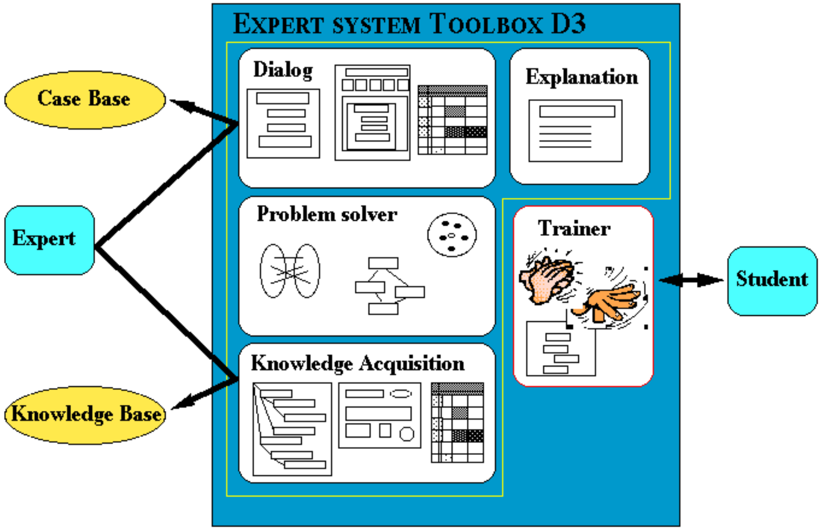
Modeling of Knowledge/Skill Acquisition

Course Recommendation (MOOCs)

AI-Based Intelligent Tutoring Systems



Wenger, E. 1987



Reinhardt, 1997

Intelligent Tutoring Systems

- Provide customized instruction and/or feedback without human intervention.
  - c.f [Anderson, J. et. al. 198]

- Adaptive Assessment Systems
  - Change Difficulty of Exam Questions based on Skill Level/Performance on Prior Questions.
  - Education Testing Services (ETS)

- Modeling of Knowledge/Skill Acquisition
  - Bayesian Knowledge Tracing (KT) [Corbett, T. et. al. 1994]
  - Various modified Knowledge Tracing (KT) models [Pardos, Z. et. al. 2011]

Motivating Higher Education Mining Project

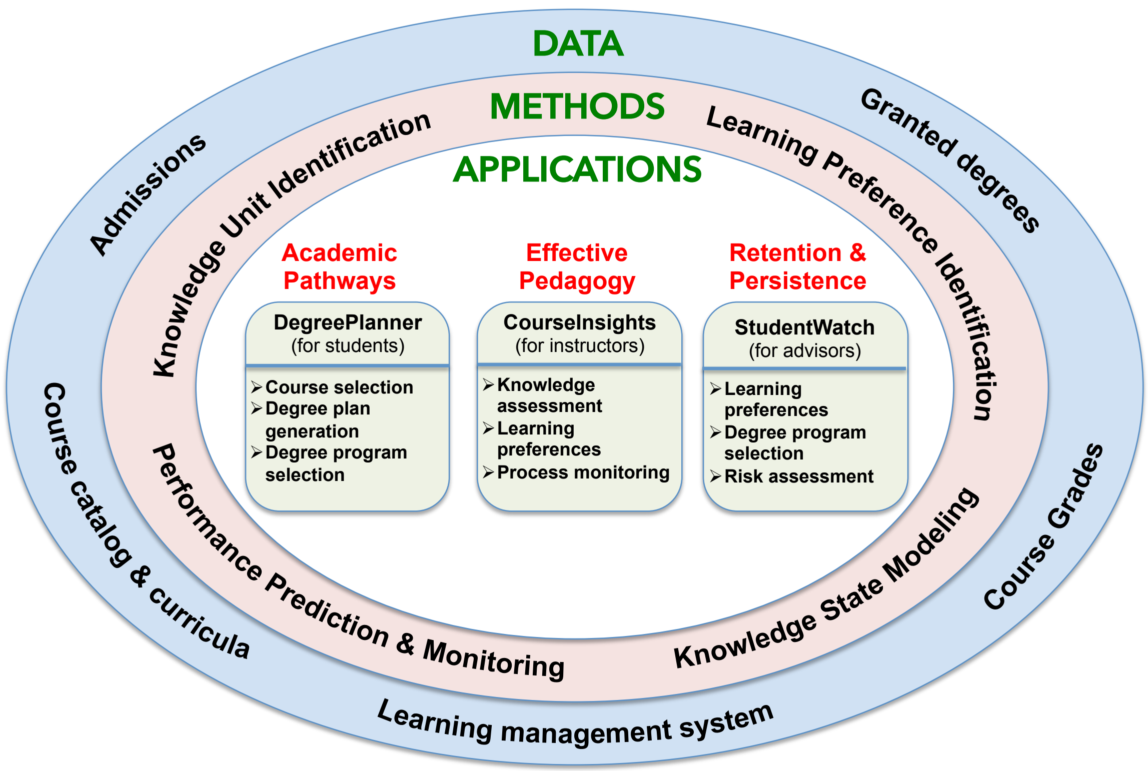
- Average National 6-year Graduation Rate is 59%
  - Equal Access Efforts Focus on Enrolment not Completion

- Higher Education Institutions need to develop innovative approaches to retain students, ensure their timely graduation, and are well-trained and workforce ready in their field of study.

- Need for better degree planners, early warning systems and intervention techniques that use student-related data.

Project Overview

NSF#1447489 & 14474488



Degree Planners &Degree Audit

Degree Planners

- Predict successful academic pathways and career trajectories.
- Predict Majors.
- Rank courses given what the students have taken.
- Make informed decisions about future enrollments.

Degree Audit

- E.g., Eulician Degree Works™ and uAchieve™
- Requirement Fulfillment
  - Find shortest path to fulfill degree requirements
  - Ensure that the degree requirements and pre-requisites are fulfilled within course recommendations
- What-If Analysis

Lifelong Learning

Economist, Jan. 2017

Lifelong Learning and Career Pathways

- Recommend educational material to a learner given:
  - Their current state of knowledge / current skills
  - The future job that the individual wants to pursue.

- LinkedIn + Lynda [Training Videos]
  - Learning paths (course sequences) toward a certain career
  - c.f. [https://www.lynda.com/learning-paths]

- Mooc Recommender System (moocrec.com)
  - Recommends MOOCs (pulled out from Coursera, EdX) to the users given their current knowledge status and dream job.

Automated Content Curation

Content Curation

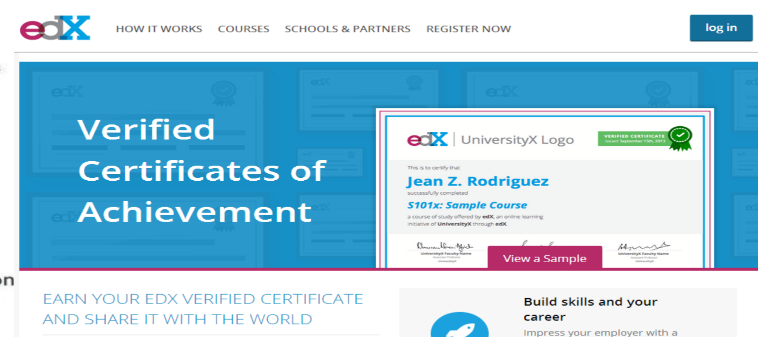
<u>Challenge</u>: Provide appropriate content based on learners' prior knowledge, learning preferences, and goals.

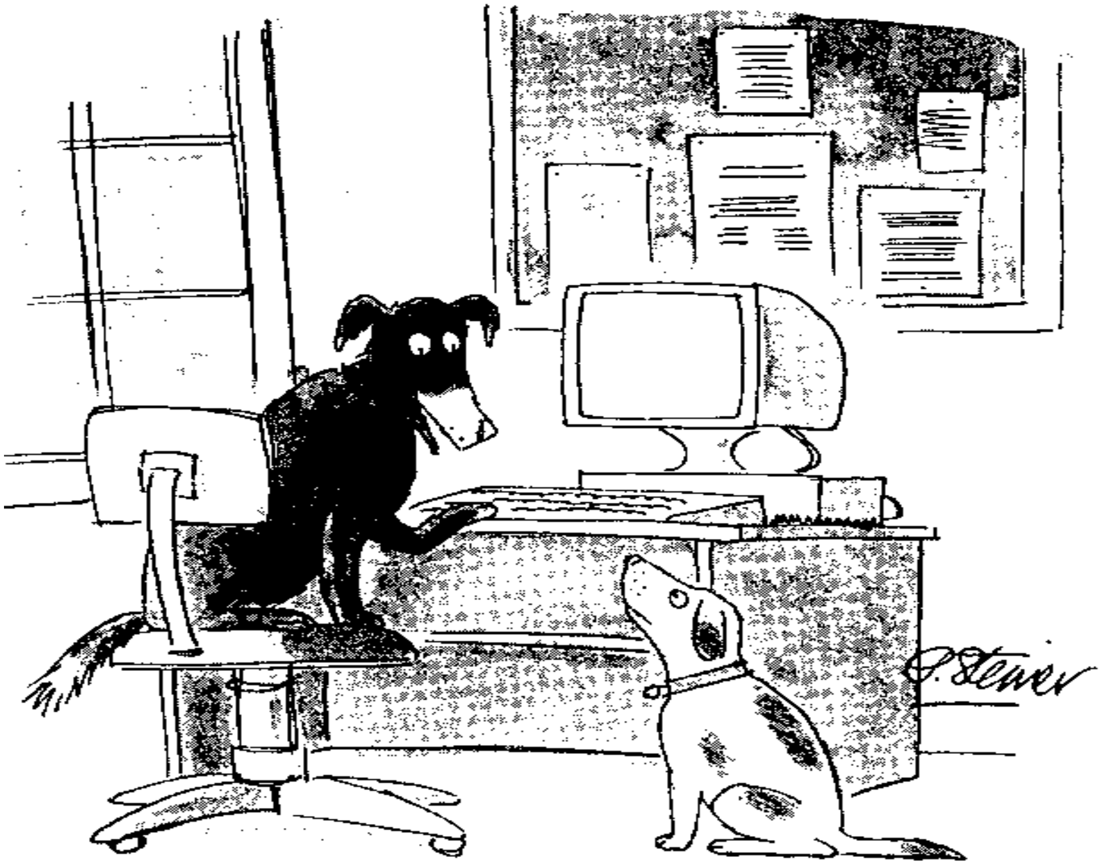
<u>Possible Solutions</u>: Recommendation, Aggregation, Crowdsourcing, Expert-sourcing through analytics and modeling of learner activity across mechanisms

Competency Certification

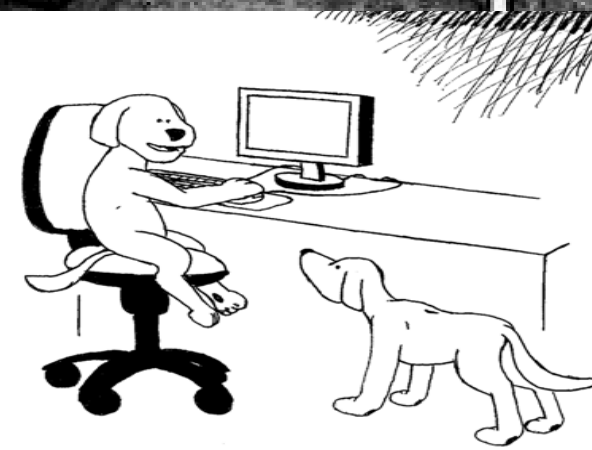
<u>Challenge</u>: Assess the competency of learners and certify it using an appropriate and useful metric.

<u>Possible Solutions</u>: Formative, dynamic, and summative assessment of learners activity within an environment. Assignment of visual or textual indicators of competency.





"On the Internet, nobody knows you're a dog."



"Today, on the Internet, no one no longer cares if you're a dog."

Learning is Formal or Informal

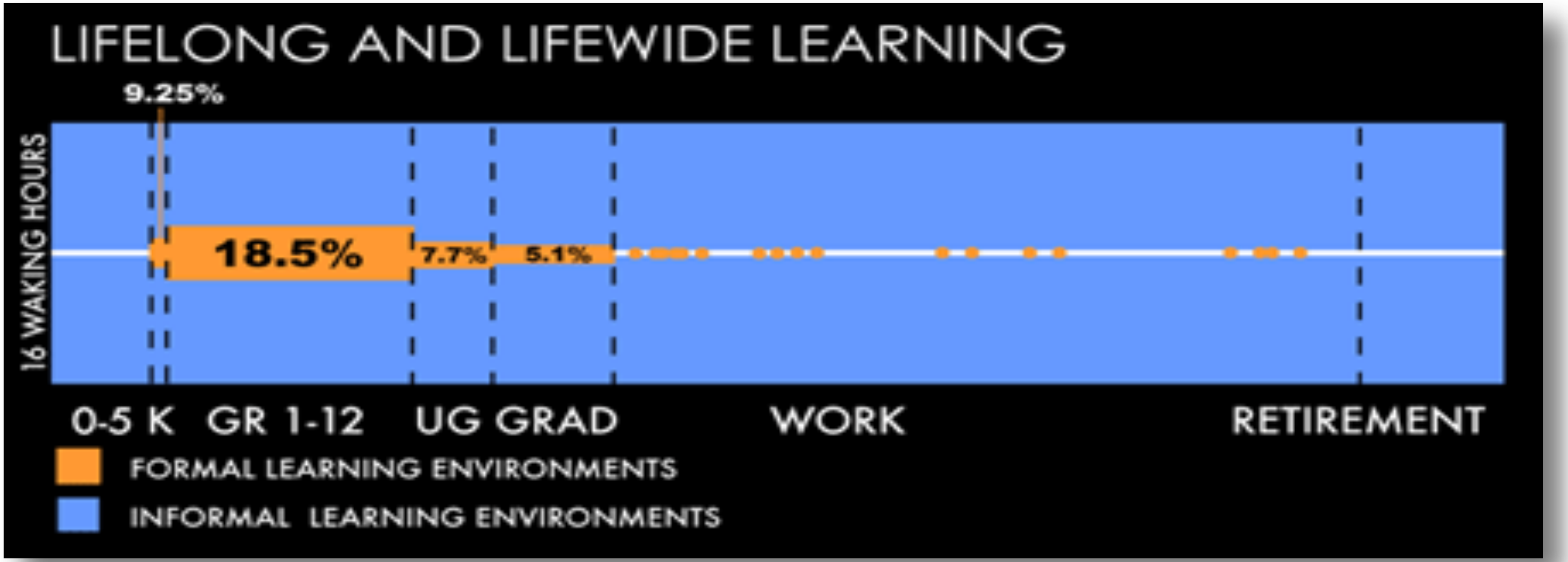


Image Credit: Life Center: www.life-slc.org

(Informal) Learning in Online Forums

- Online help forums and Q&A sites are gaining popularity across domains
- Supportive environments with high response rate, usually of high quality [Adamic et al. WWW 2008; Mamykina et al. CHI 2011]
- A core group of users supports the community [Yang et al. ICWSM 2009]

Research Study- New to Java [Hon, J. et.al. 2013]

Dataset: 10 years (2001-2009) of forum activity; 200K+ discussion messages, 37K+ discussion topics; Q/A ratio of 7.36

Post Count	Number of Users
0 to 10	14114
11 to 50	5076
51 to 500	1922
More than 500	326

User Profile by Status

Status	Required Duke Points	Members w/Status	Total Duke Points	Average Duke Points
Bronze	0	21179	21,326	1.01
Silver	100	180	36,055	200.31
Gold	500	36	24,026	667.39
Platinum	1000	43	81,904	1904.74

Quality of Help

Type of Help	Description
Framing	Help-givers gave framing help – to help them frame or reframe their questions – which might not be immediately useful to the help-seekers.
Provide Off-topic Opinion	*I don't think anybody here has a problem with someone asking for help with homework. The problem--or one of them--is when they don't provide a clear, precise question.*
Recommend Revisiting Original Source	*Maybe your instructor has made the (faulty) assumption that...You'd better ask him/her. Otherwise it appears that he/she is mistaken that you can simply derive the class name from the file name.*
Guiding	Help-givers offered guidance which can assist help-seekers with their task at hand but not entirely resolve their challenges.
Quote Directly from Existing Material	*Members are either declared in the type, or inherited because they are accessible members of a superclass or superinterface which are neither private nor hidden nor overridden.*
Provide Link to External Resources	*See if you can find record of this kind of GC [Garbage Collector]/finalizer bug at http://bugs.sun.com/.*
Advise to Use External Resource	*You have to declare your input variable before you use it, just like any other variable. Google for java while loop example. I'm sure there'll be plenty.*
Engaged	In discussion topics where expertise or context permitted, help-givers provided original codes, detailed explanation for a specific question and detailed procedures to assist with troubleshooting issues.
Write/Edit Code	*I hacked this up in about an hour as an example of how one should go about such stuff...*
Provide Detailed Explanations	*You are totally right Adams '==' operator compares the reference of 2 objects which cannot be the same at all cause ...[Followed by code]*
Provide Step-by-Step Instructions	*Here's the procedure for setting the classpath in Windows ...Create a new system variable called...*

Distribution of Quality of Help within a Thread

Type of Help	Avg. Count (Full Thread)	Avg. Count (1st Half)	Avg. Count (2nd Half)
Framing	**6.79**	**2.02**	**4.89**
Provide Off-topic Opinion	**5.40**	**1.43**	**4.04**
Recommend Revisiting Original Source	**1.38**	**0.43**	**0.85**
Guiding	**4.37**	**2.64**	**1.98**
Quote Directly from Existing Material	**1.45**	**0.81**	**0.64**
Provide Link to External Resources	**0.94**	**0.62**	**0.32**
Advise to Use External Resource	**1.98**	**0.83**	**1.15**
Engaged	**5.64**	**3.64**	**2.00**
Write/Edit Code	**1.51**	**0.96**	**0.55**
Provide Detailed Explanations	**1.96**	**1.15**	**0.81**
Provide Step-by-Step Instructions	**2.17**	**1.53**	**0.64**

Findings [Hon, J. et. al. 2013]

- A few expert members can support a large online help-giving forum.
- These experts are highly active and responsive – they provide help quickly, they provide help of high quality and do not duplicate their efforts.
- Help-giving does depend upon the quality of question and they guide help-seekers in framing of questions.

Where is the Data ?

- Traditional Data: Databases of local, state, or national level student and/or school demographics and performances.

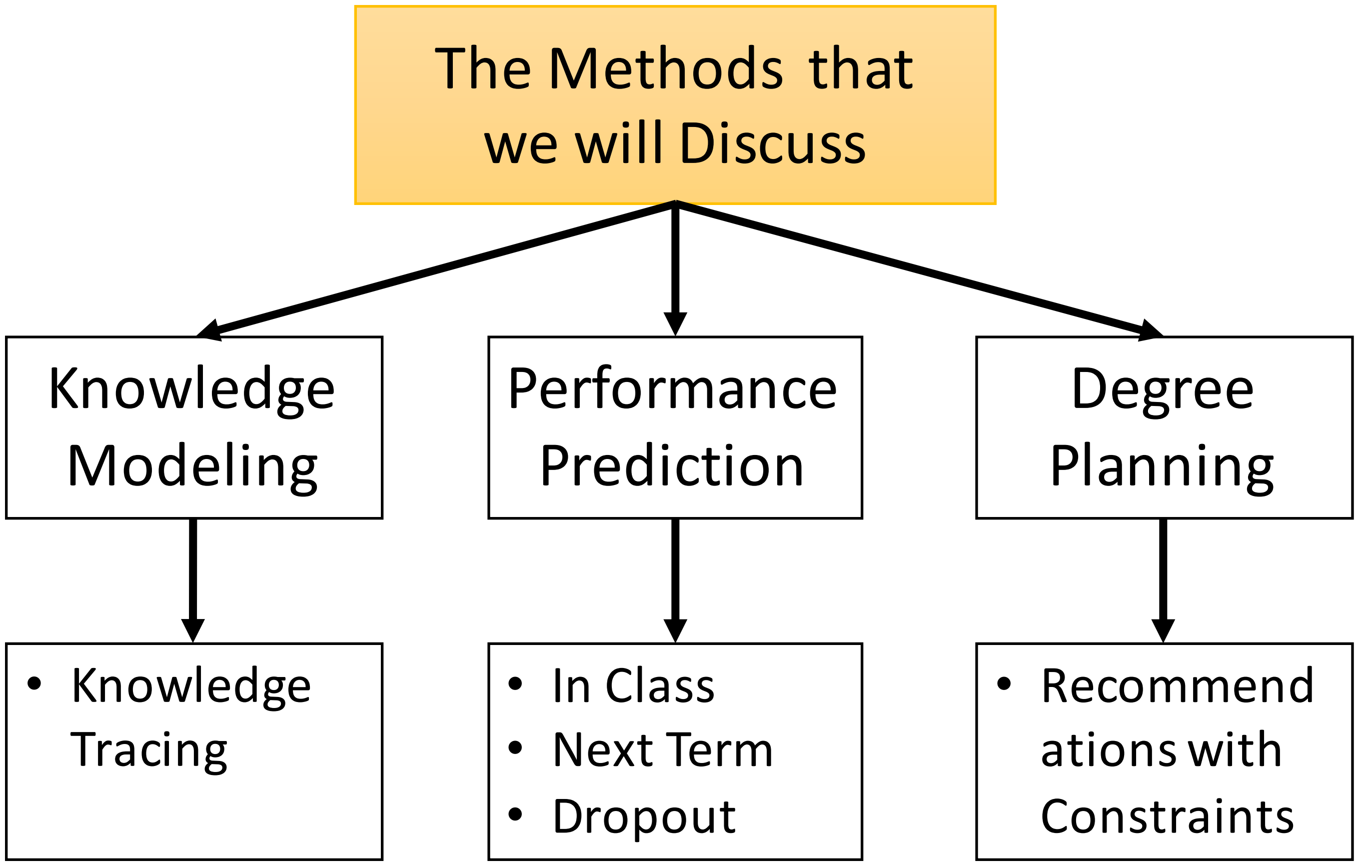
- Interactive Data: collected from learners interacting with systems like LMS, MOOCs, or Intelligent Tutoring Systems.

- Sensor Data: collected from instrumented learning environments such as video, audio, eye tracking, EEG, etc

- Exogenous Data: collected for other purposed (produced in other activities) that can be combined with data collected for education or learning, e.g. social media use

Modified from John Cherniavsky, NSF, AERA 2016 presentation

Part II:
Problems and Methods


The Methods that we will Discuss
Knowledge Modeling
Performance Prediction
Degree Planning
• Knowledge Tracing
• In Class
• Next Term
• Dropout
• Recommendations with Constraints

Knowledge Tracing

Knowledge Tracing (KT)

- Used in Intelligent Tutoring Systems for Modeling Knowledge/Skill Acquisition by students

- Model student knowledge over time to accurately predict how students will perform on future interactions

Bayesian Knowledge Tracing (BKT)

- Student knowledge is represented with binary variables
- One variable per skill
- The skill is either mastered by the student or not
- Observations are also binary: right or wrong answer to each problem

Bayesian Knowledge Tracing (BKT)

- A hidden Markov model
- Models probability of learning and correctly applying a skill
- Various modified Knowledge Tracing (KT) models

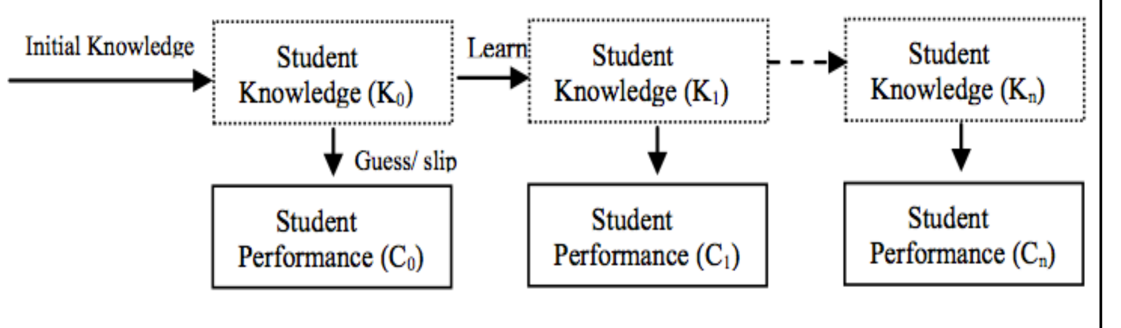
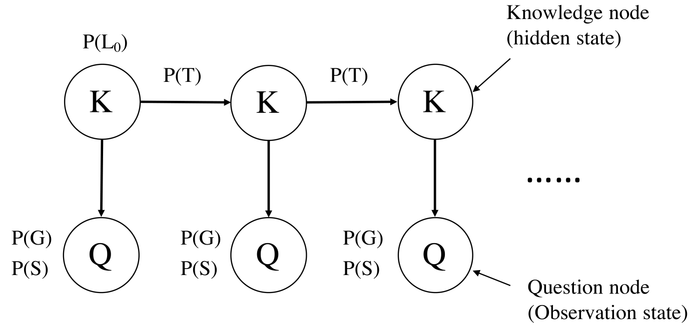


Fig. 1. Knowledge tracing model

Source: [30] Gong et at, ITS, 2010

KT with Item Difficulty (KT-IDEM)

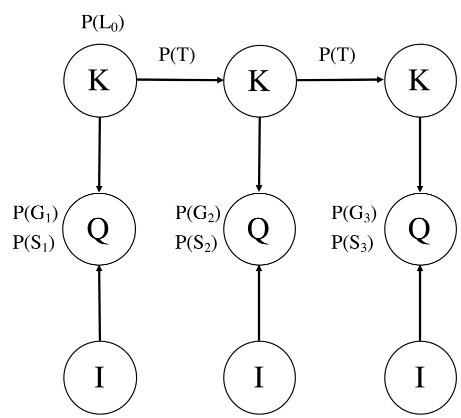
- Extends KT to capture item difficulty



Model parameters

$P(L_0)$ = Initial Knowledge
$P(T)$ = Probability of learning
$P(G)$ = Probability of guess
$P(S)$ = Probability of slip

- Improve prediction accuracy.



Model parameters

$P(L_0)$ = Initial Knowledge
$P(T)$ = Probability of learning
$P(G_{1\ldots n})$ = Probability of guess per question
$P(S_{1\ldots n})$ = Probability of slip per question

n denotes the number of all questions.

Source: [21] Z. A. Pardos et. al, 2011

KT with Item Difficulty (KT-IDEM)

- Evaluation using “The Cognitive Tutor: Mastery Learning datasets”

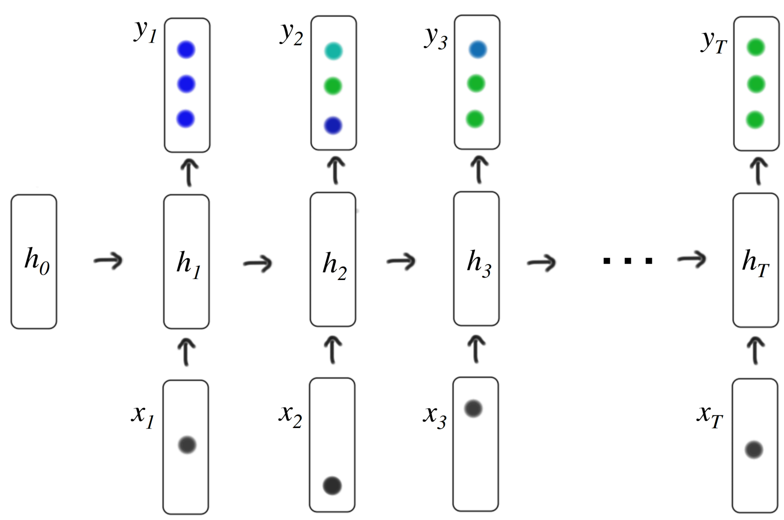
					AUC		
Skill	**#students**	**#prob**	**#data**	**#data/#prob**	**KT**	**KT-IDEM**	**Delta**
1	133	320	1274	3.98	0.722	0.687	- 0.035
2	149	102	1307	12.81	0.688	0.803	+0.115
3	116	345	1090	3.16	0.612	0.605	- 0.007
4	116	684	1062	1.55	0.694	0.653	- 0.041
5	159	177	1475	8.33	0.677	0.718	+0.041
6	116	396	1160	2.93	0.794	0.497	- 0.297
7	133	320	1267	3.96	0.612	0.574	- 0.038
8	116	743	968	1.30	0.679	0.597	- 0.082
9	149	172	1431	8.32	0.585	0.720	+0.135
10	148	177	1476	8.34	0.593	0.626	+0.033
11	149	172	1431	8.32	0.519	0.687	+0.168
12	123	128	708	5.53	0.574	0.562	- 0.012

Source: [21] Z. A. Pardos et. al, 2011

Deep KT

- Model student learning using Recurrent Neural Networks (RNNs) and a Long Short Term Memory (LSTM) model.

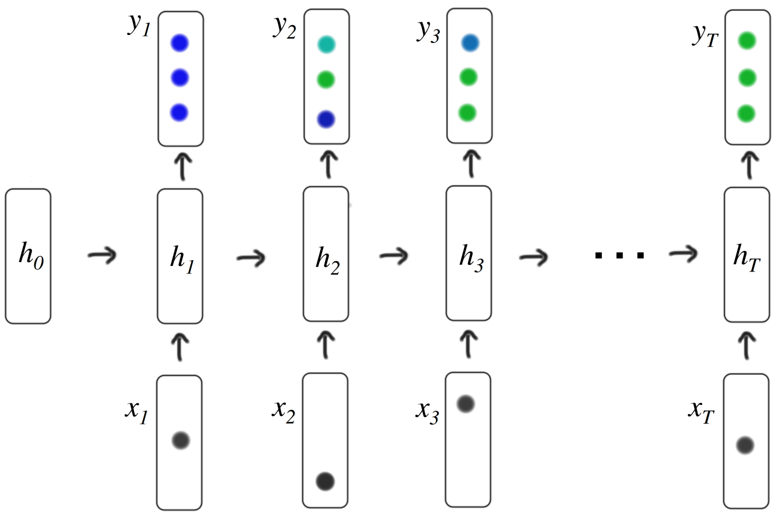
- Maps an input sequence of vectors $x_1 \ldots x_T$, to an output sequence of vectors $y_1 \ldots y_T$



Source: [22] C. Piech et. al., 2015

Deep KT

- Computing a sequence of 'hidden' states $h_1 \ldots h_T$
  - Encodes relevant information from past observations that are useful for future predictions



$$\mathbf{h}_t = \tanh\left(\mathbf{W}_{hx}\mathbf{x}_t + \mathbf{W}_{hh}\mathbf{h}_{t-1} + \mathbf{b}_h\right),$$
$$\mathbf{y}_t = \sigma\left(\mathbf{W}_{yh}\mathbf{h}_t + \mathbf{b}_y\right),$$

Source: [22] C. Piech et. al., 2015

Deep KT

- Input representation:
  - One-hot encoding for small number of exercises
  - Compressing the sparse exercises vector for large number of exercises

- Output representation:
  - A vector of length equal to the number of problems,
  - Each entry holds the predicted probability that the student would answer that particular problem correctly.
  - The prediction of answer a_{t+1} can be read from the entry in y_t corresponding to probability q_{t+1}

Source: [22] C. Piech et. al., 2015

Deep KT

- Evaluation
  - Improved AUC over KT

	Overview			*AUC*			
Dataset	Students	Exercise Tags	Answers	Marginal	BKT	BKT*	DKT
Simulated-5	4,000	50	200 K	0.64	0.54	-	0.75
Khan Math	47,495	69	1,435 K	0.63	0.68	-	0.85
Assistments	15,931	124	526 K	0.62	0.67	0.69	0.86

Table 1: AUC results for all datasets tested. BKT is the standard BKT. BKT is the best reported result from the literature for Assistments. DKT is the result of using LSTM Deep Knowledge Tracing.*

Source: [22] C. Piech et. al., 2015

Next-Term Grade Prediction

Data Utilized for Prediction

- Students Demographics
- High-school performance
- Grades in previous courses
- Student Academic Information: Majors, Academic Levels
- Course Information: Subjects, Content, Levels
- Information about instructors

Class of Methods

- Regression-based Methods
  - Logistic Regression
  - Personalized Multi-Regression
  - Course-Specific Regression

- Matrix Factorization
  - Typical MF
  - Factorization Machines
  - Course-Specific Matrix Factorization

- Domain Aware Methods
  - Popularity
  - User-based Collaborative Filtering
  - Matrix Factorization
  - Regression Models

Class of Methods

- **Regression-based Methods**
  - **Logistic Regression**
  - **Personalized Multi-Regression**
  - **Course-Specific Regression**

- Matrix Factorization
  - Typical MF
  - Factorization Machines
  - Course-Specific Matrix Factorization

- Domain Aware Methods
  - Popularity
  - User-based Collaborative Filtering
  - Matrix Factorization
  - Regression Models

Personalized Multi-Regression (PLMR)

- A linear combination of k regression models, weighted on a per-student basis.

$$\hat{g}_{ij} = w_0 + s_i + c_j + P_i W X_{ij}$$

Vector of Student/Course Feature Values

Grade of student i in course j

Global weight

Student bias

Cours e bias

Student Weight Vector

Regression Coefficients Matrix

- More personalized to each student
- Considers various student groups

Source: [9] A. Elbadrawy et al, 2016

Course-specific Regression (CSpR)

- Assumption: Previous courses provide necessary knowledge for future courses.

  ➔ Student's performance in a subset of the previous courses can predict her performance in a future course.

$$grade(s,c) = sparse\ linear\ combination(previous\ grades\ of\ s)$$

$$\boxed{\hat{y}^c = w_0^c + \mathbf{s}^T\mathbf{w}^c}$$

- A course-specific subset of the data is used to learn the model.

Source: [23] A. Polyzou et al, 2016

Class of Methods

- Regression-based Methods
  - Logistic Regression
  - Personalized Multi-Regression
  - Course-Specific Regression

- **Matrix Factorization**
  - **Typical MF**
  - **Factorization Machines**
  - **Course-Specific Matrix Factorization**

- Domain Aware Methods
  - Popularity
  - User-based Collaborative Filtering
  - Matrix Factorization
  - Regression Models

Matrix Factorization (MF)

- Student-Course-Grade data represented as a matrix
- Each student and course are represented via k dimensional latent feature vectors
    - Grade is estimated as vector inner product.

$$\hat{g}_{ij} = \sum_{f=1}^{k} v_{i,f} v_{j,f}$$

Courses

Students

					S				
	A								
							A		
				B-					
A-			B+						
						C			
		C							
								F	

Matrix Factorization (MF)

- Ignores the sequence in which a group of courses were taken.
  - A course latent representation can be influenced by courses taken afterward.

- Course-Specific Matrix Factorization (CSpMF)
  - Relies only on the subset of the data used by CSpR in order to estimate an MF model that is specific to each course.

- Factorization Machines (FM)
  - Utilize student/course features
  - Can predict grades for new students with no previous grades:
    - Replace student latent vector $\boldsymbol{v_s}$ by a linear transformation of his feature representation $\boldsymbol{P.f_s}$

Comparing the different methods

- Results: FM and PLMR outperformed Random Forests and Baselines.

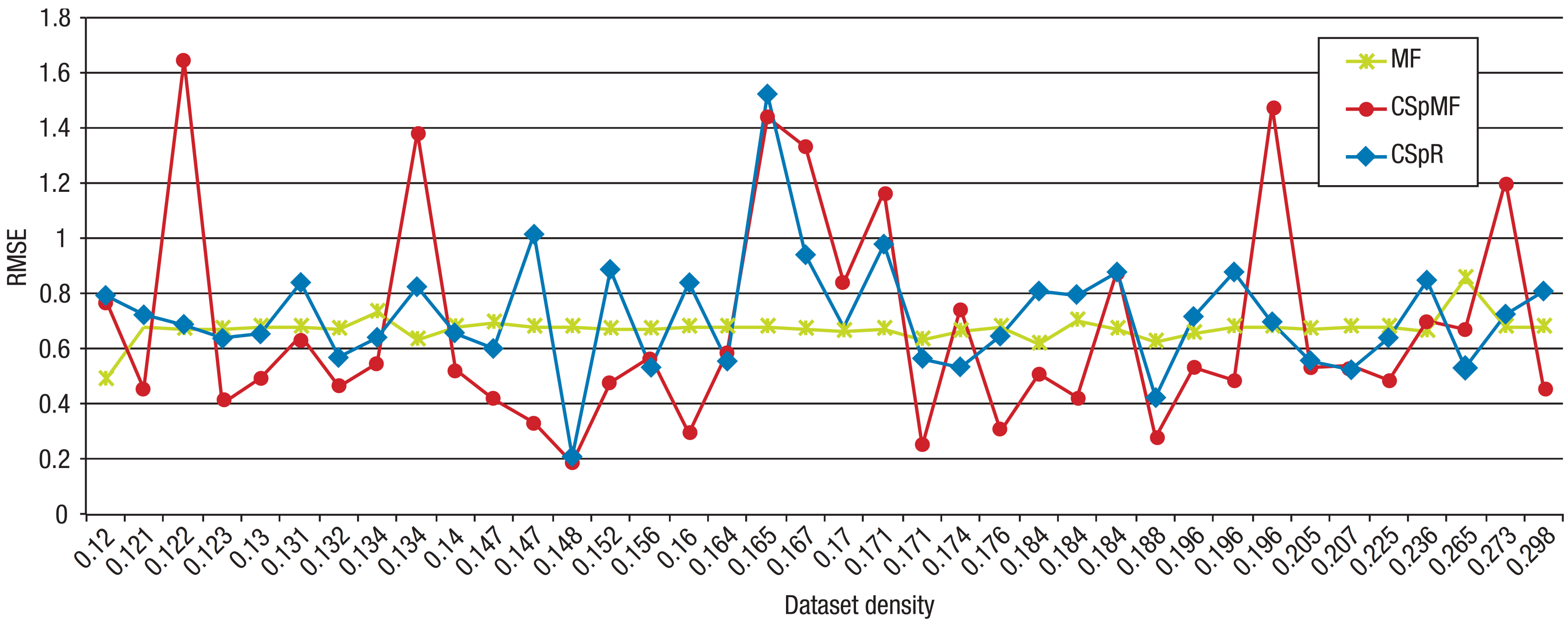
TABLE 1. Next-term grade prediction results on George Mason University transcript data.

Method	Root-mean-square error (RMSE)	Mean absolute error (MAE)
Factorization machine (FM)	0.7423	0.52 ± 0.53
Personalized linear multi-regression (PLMR)	0.7886	0.57 ± 0.55
Random forest (RF)	0.7936	0.58 ± 0.54
Mean of means	0.8643	0.64 ± 0.58
Uniform random guessing	1.8667	1.54 ± 1.06

Source: [9] A. Elbadrawy et al, 2016

Comparing the different methods

- Results: CSpMF outperformed MF and CSpR



Source: [9] A. Elbadrawy et al, 2016

Class of Methods

- Regression-based Methods
  - Logistic Regression
  - Personalized Multi-Regression
  - Course-Specific Regression

- Matrix Factorization
  - Typical MF
  - Factorization Machines
  - Course-Specific Matrix Factorization

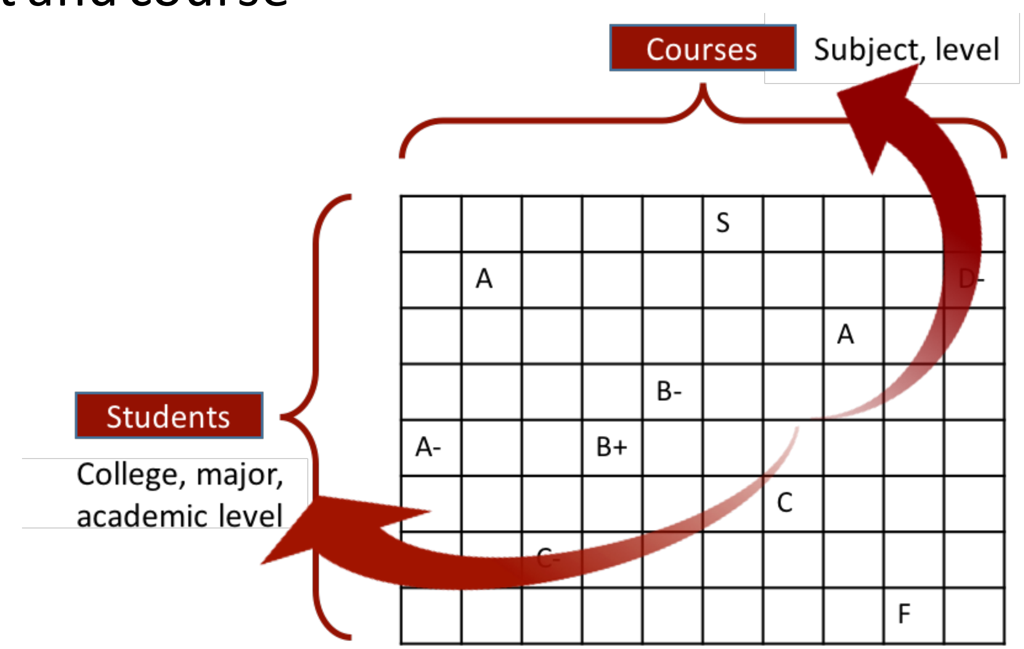
- **Domain Aware Methods**
  - **Popularity**
  - **User-based Collaborative Filtering**
  - **Matrix Factorization**
  - **Regression Models**

Domain Aware Methods

- Students of certain colleges, majors and academic levels tend to enroll in courses of certain subjects and levels.

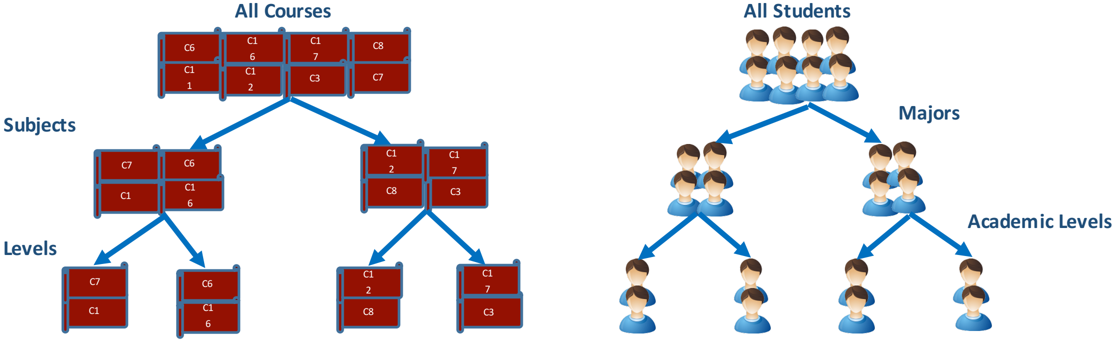
- **NMAR** data with **Grouping Structures**

- Patterns in the student-course matrix are determined by the student and course features.



Domain Aware Grade Prediction and Course Recommendation

- Define student and course groups at various levels of granularity.
- Finer groups ➔ more homogeneous, but less data points.
- Modify existing methods to incorporate these groups.



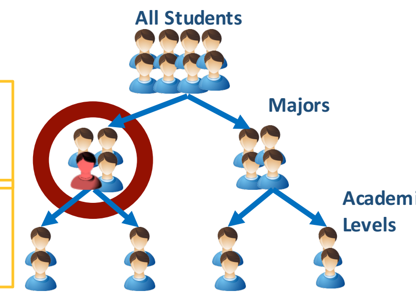
Domain Aware Grade Prediction and Course Recommendation

- Popularity-based Ranking:
  - Rank within group : $rank =$ $|\varphi_{s \rightarrow c}|$
- User-based Collaborative Filtering: Select neighbors from within group

n_n most similar students

same group as s

co-taken n_c courses



- Matrix Factorization: Define multiple row/column biases based on the groups.

$$\hat{r}_{s,c} = b_s^{\varphi_c} + b_c^{\varphi_s} + \boldsymbol{p}_s^t \boldsymbol{q}_c$$

Student bias of s for the context described by the features of course c

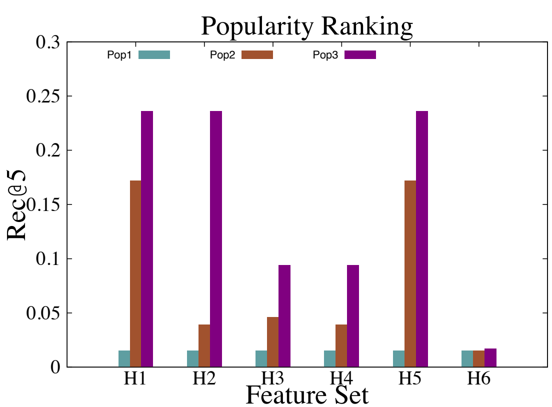
Course bias of c for the context described by the features of student s

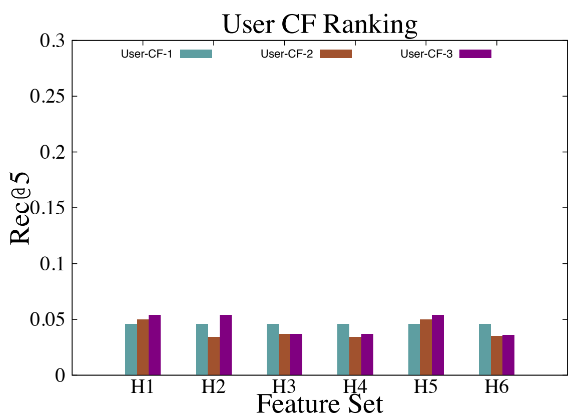
Student and course latent factors

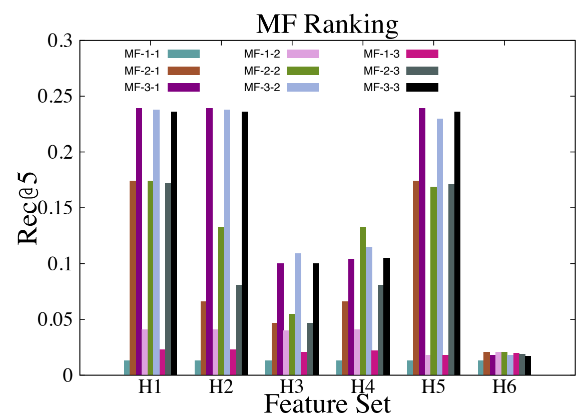
Source: [8] A. Elbadrawy el at, 2016

Domain Aware Grade Prediction and Course Recommendation

- Evaluation:
  - UMN dataset, 60K students, 10K courses, 1.7Million grades,
  - Six Feature Sets ➔ Six Student & Course Groups: H1,…H6
- Ranking Results (Recall@5):
  - Students majors & academic levels are the most important features

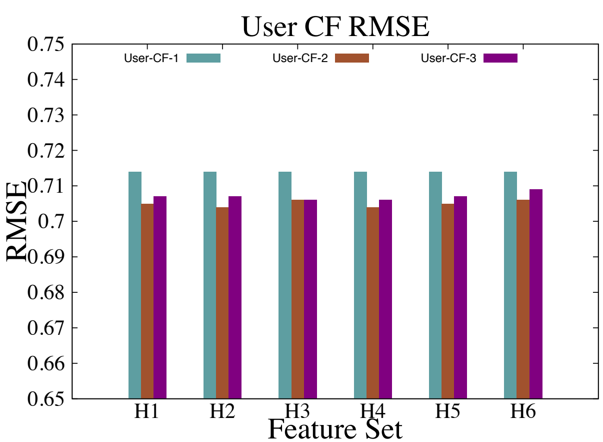


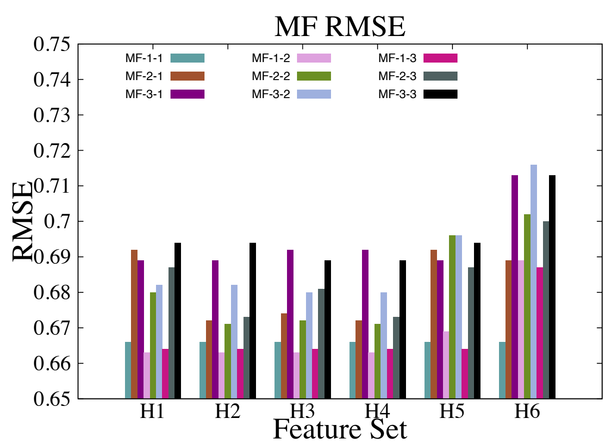


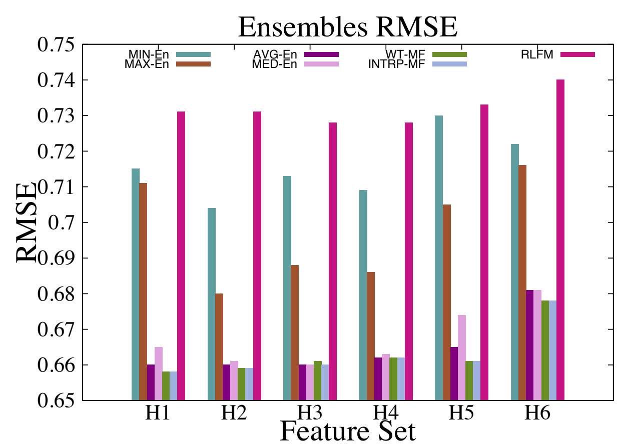


Domain Aware Grade Prediction and Course Recommendation

- Prediction Accuracy Results (RMSE):
  - Student majors are the most important features for User-CF
  - For MF, finer groups perform the worst due to small sample sizes with biases.
  - Ensemble with sample-size-based combination weight performs the best.







Cumulative Knowledge-based Regression(CKRM)

- Each course provides/requires a set of knowledge components.
- By taking a course, students acquire a set of knowledge components
- Assumes that
  - All courses can be represented in a space of knowledge components
  - A course provides a subset of components to students that taking it
  - Students can acquire the same (or similar) knowledge components by taking different subsets of courses.
- Models
  - Knowledge state of the student
  - Course's required and provided knowledge components

Cumulative Knowledge-based Regression(CKRM)

- Knowledge state of student s after taking j courses is computed as:

$$\mathbf{k}_{s,j} = \sum_{i=1}^{j} \left(\xi(s, c_j, c_i) \; g_{s,c_i} \; \mathbf{p}_{c_i} \right)$$

- ▸ g_{s,c_i} is the grade that student s obtained on course c_i
- ▸ $\xi(s, c_j, c_i)$ is a time-based exponential decaying function
- ▸ $\mathbf{p}_{c_i}$ is c_i's provided knowledge component vector

- Grade of student s in course c after taking j courses is estimated as:

$$\hat{g}_{s,c} = b_c + \mathbf{r}_c \; \mathbf{k}_{s,j}^T$$

- ▸ b_c is a course bias term
- ▸ $\mathbf{r}_c$ is c's required knowledge components vector
- ▸ $\mathbf{k}_{s,j}$ is the student's knowledge state vector (Eq. 1)

Source: [17] S. Morsy el at, 2016

Cumulative Knowledge-based Regression(CKRM)

Modeling the Knowledge Component Space

- Latent Knowledge Component Space
  - CKRMall: All courses share the set of knowledge components
  - CKRMdep: A different set of knowledge components for courses from the different departments

- Text-based Knowledge Component Space (CKRMtext)
  - Builds a course by knowledge component matrix using course descriptions in the University catalog
  - Knowledge components are the keywords

Cumulative Knowledge-based Regression(CKRM)

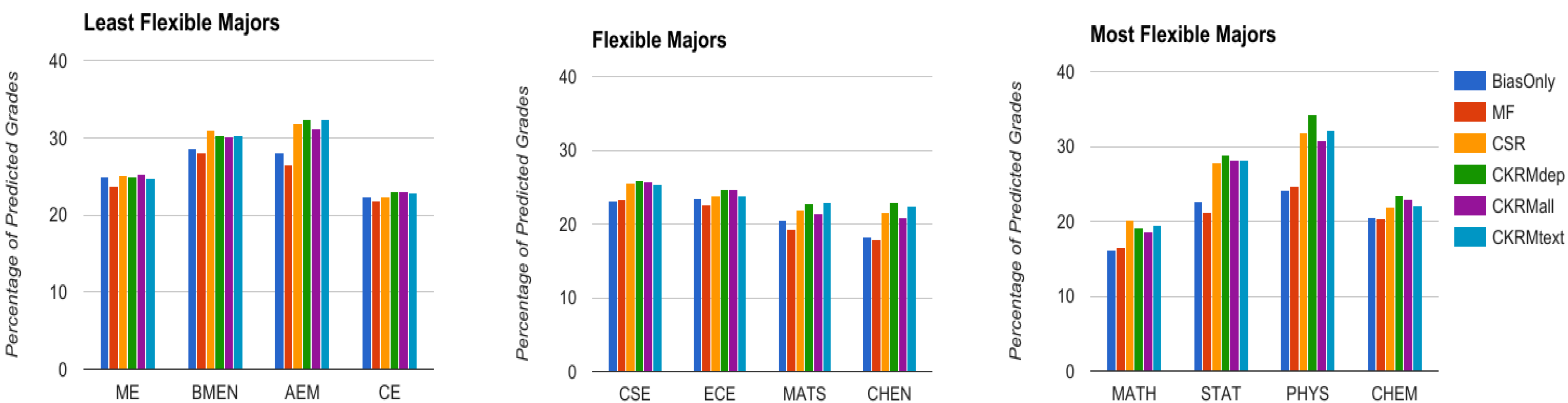
Evaluation:

- RMSE & Tick Error (the number of ticks the predicted grade is from the actual grade)
- Course (& Major) Flexibility:
  - ***A highly flexible course is a course does not share a lot of knowledge components with the courses taken before it.***
  - Flexibility of course offering c = (1 – JaccardCoef), with all courses that were taken prior to c
  - Major flexibility is the average over all course offerings within the major.

Cumulative Knowledge-based Regression(CKRM)

- Percentage of grades with no error (0 Tick Error)

- CKRM outperforms CSpR, especially on the most flexible majors.



Source: [17] S. Morsy el at, 2016

Cumulative Knowledge-based Regression(CKRM)

- Text Analysis: Qualitative Analysis on CKRMtext
  - For the students who took each course c, extract the top words with the highest weights in their knowledge states prior to taking c

CSCI 3081W -- Program Design and Development
Top keywords: data:122.58, analysi:76.56, advanc:69.97, fundament:59.17, structur:49.26, program:45.31, algebra:35.43, comput:34.43, set:27.91, system:27.01, languag:25.3, tree:24.71, softwar:23.55, topic:23.09, permit:21.66
CSCI 5523 -- Introduction to Data Mining
Top keywords: analysi:17.09, develop:4.98, advanc:4.73, program:4.15, model:2.66, algorithm:2.41, data:2.4, fundament:2.05, structur:1.86, topic:1.74, system:1.73, languag:1.6, calculu:1.45, logic:1.26, softwar:1.05

Words in red denote those that appear in the listed course's pre-reqsuites descriptions, whereas words in blue denote those that appear in the course's description

Source: [17] S. Morsy el at, 2016

In-Class Assessment Prediction

In-Class Assessment Prediction

- Predicting a student's performance on ***in-class assessments*** like quizzes and homework assignments.

- Can potentially provide the needed early intervention for students that are at risk of failing a course or dropping out.

Data utilized for Prediction

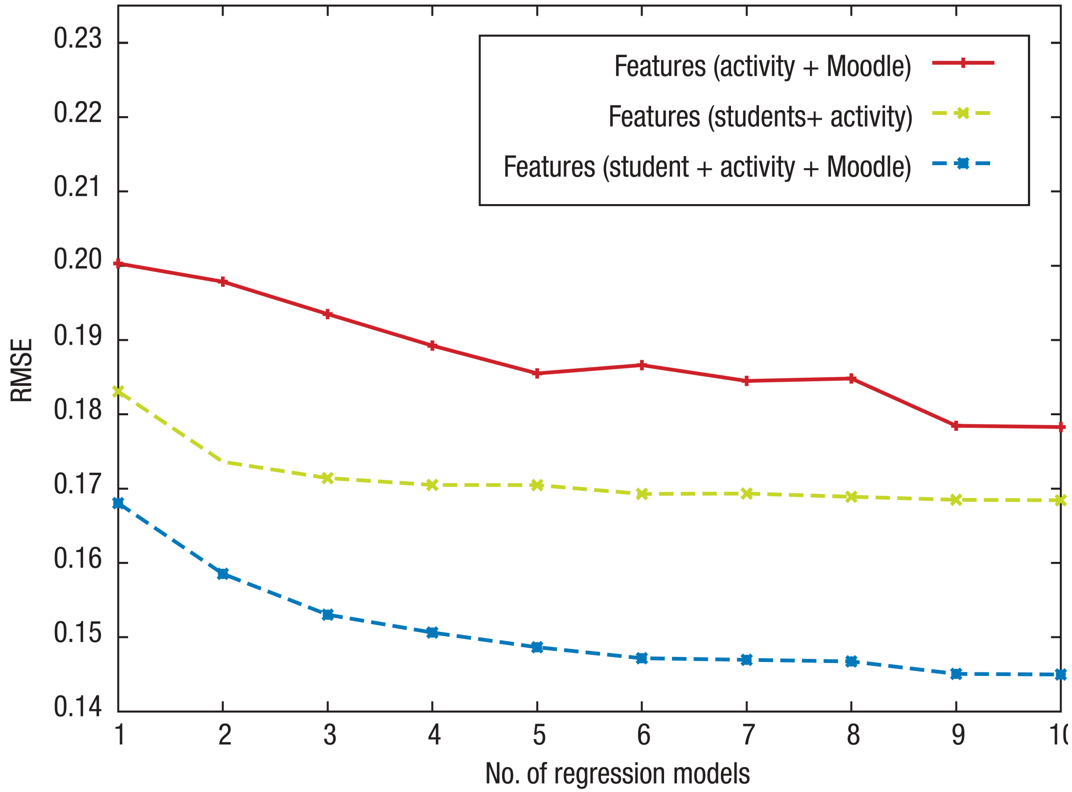
- Learning Management System (LMS) data:
  - Information about students access course material, posting on forums, reading other students postings, … etc.
- MOOC data:
  - Click-stream server logs indicates watching of class videos, access to other materials, … etc.
- Student data: previous performance, major

Methods

- Logistic Regression
  - One global model, not personalized.
- Matrix Factorization
  - Ignores students' interactions with the LMS, which can provide more granular forecasts
  - Only applicable in the case of fixed recurring assessments within each course
- PLMR
  - Analyze click-stream server logs to extract features
  - Personalized prediction of the student performance in the next graded assessment
  - Analyze relative contribution of the different features to the predictions
    - Enforce non-negativity constrain on parameters to ensure additive contribution to predicted grades

PLMR

- RMSE using different features with PLRM



UMN Moodle Dataset

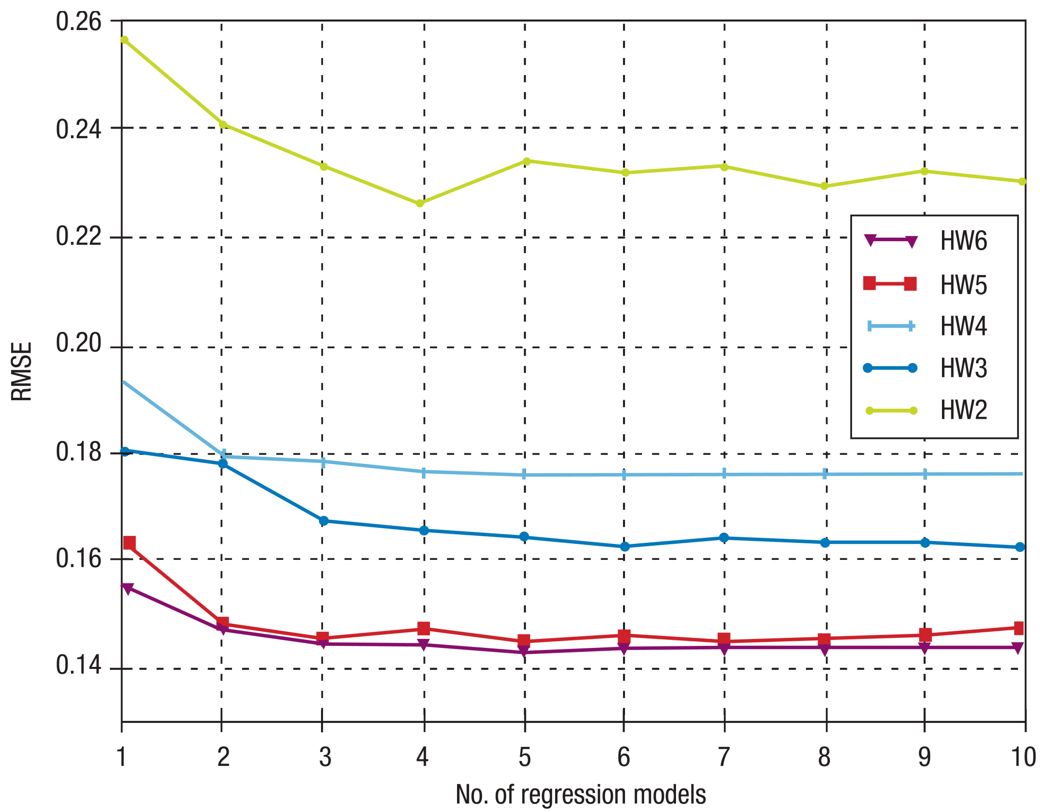
Improved accuracy using student-LMS interaction features

Improved accuracy with increasing number of regression models

Source: [9] A. Elbadrawy et al, 2016

PLMR

- RMSE on different assessments (Homework) with PLRM



Stanford MOOC Dataset

Improved accuracy with increasing number of regression models

Source: [9] A. Elbadrawy et al, 2016

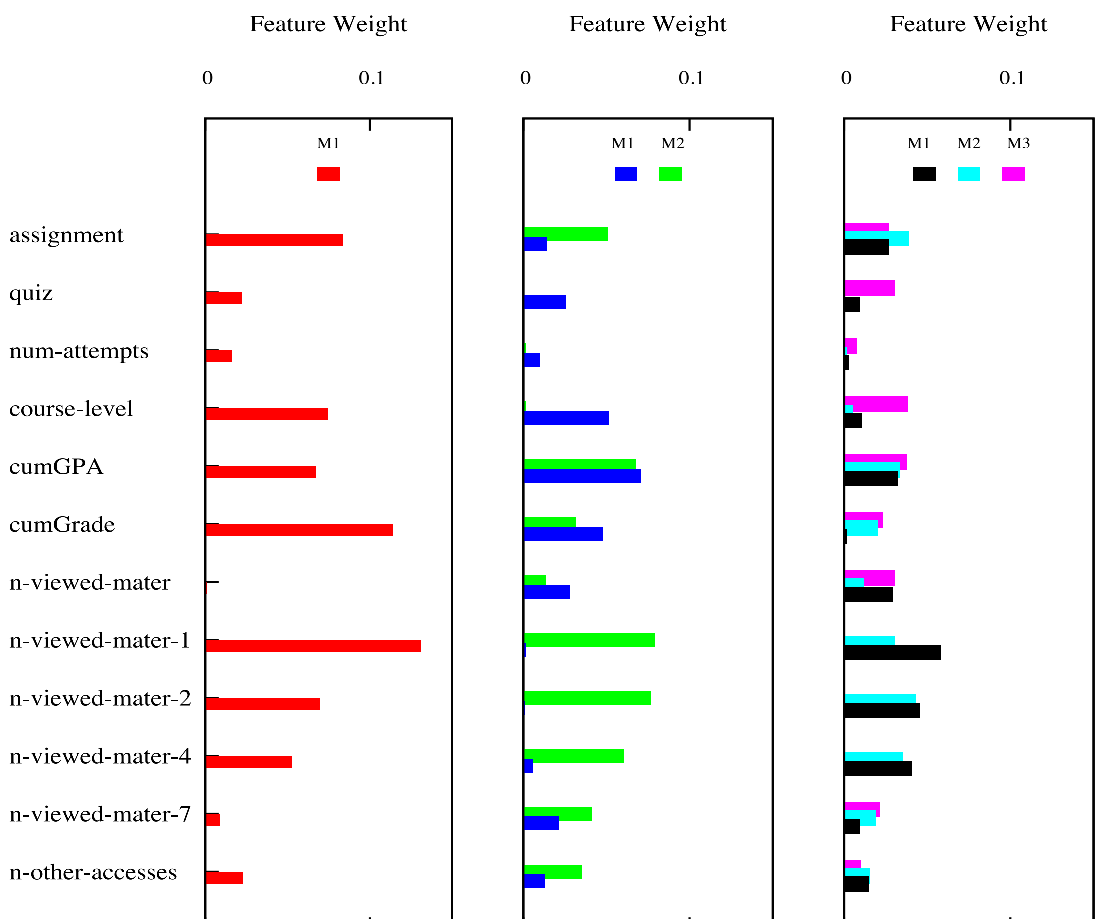
PLMR

- Feature Contribution to Predicted Grades

- Highest contributing features:

Previous Performance

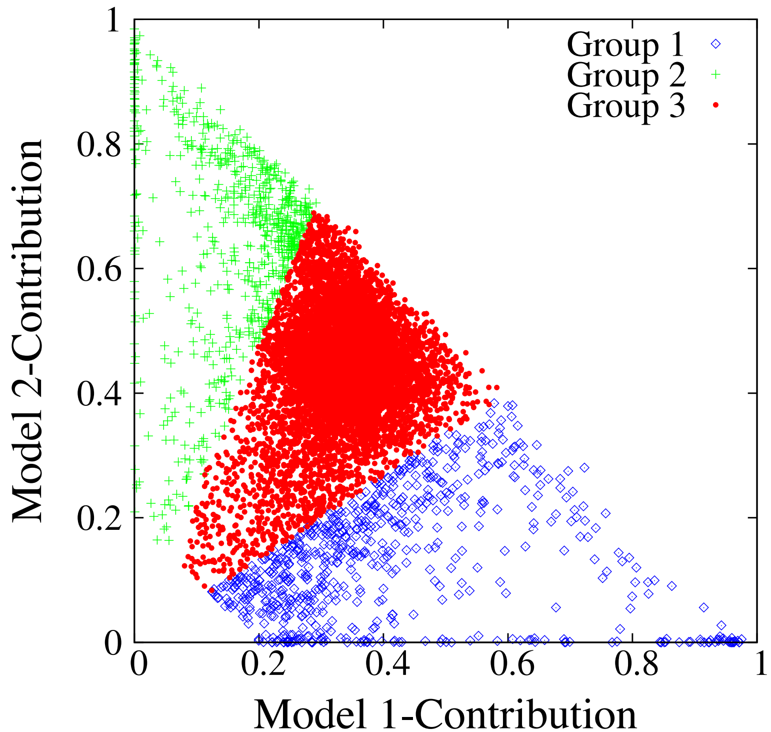
View Course Material



Source: [10] A. Elbadrawy, et al, 2015

PLMR

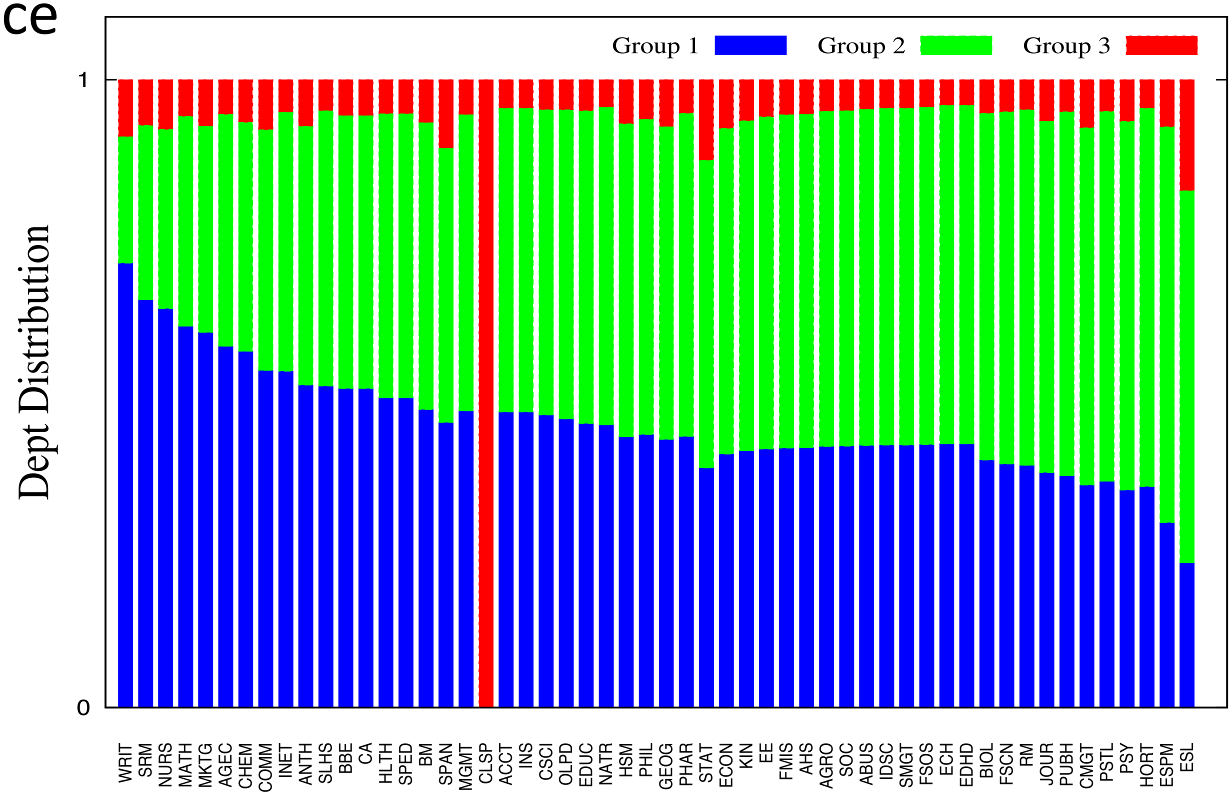
- Analysis of Student Membership Weights
- For some students, their LMS interactions are more predictive of their performance than other students



Source: [10] A. Elbadrawy, et al, 2015

PLMR

- Student Groups within Departments
- Some departments tend to have more students whose interactions with the LMS are not predictive of their performance



Source: [10] A. Elbadrawy, et al, 2015

Drop-Out Prediction

Predicting Students who Drop Out

- Methodology:
  - Binary Classification: Students who graduated vs. students who dropped out
- Features:
  - # previously taken courses in Science, Math, … etc
  - # Attempts & Performance in previous courses
- Classifiers:
  - Decision Trees (CART, J48), Bayesian Classifier (BayesNet), Rule-based Classifier (JRip), Simple Logistic Regression (Logit), Random Forests (RF)
- **Most Important Predictors (CART, EE Department):**
  - **Grades in Linear Algebra, Calculus, Networks, Avg. Grade in High School Science Courses**

Classifiers	*OneR*	*CART*	*J48 -M 2*	*J48 -M 10*	*BayesNet*	*Logit*	*JRip*	*RF*
Accuracy	0.76	0.81 o	0.79	0.79	0.75	0.79	0.78	0.80

Source: [7] Dekker et al, 2009

Predicting Dropout in E-Learning Courses

- Binary Classification
- Features
    - Student Information System (SIS) features **(constant)**
        - ACT score
        - Financial support
        - Credit hours completed, previous drops, degree seeking status
        - Previous online course completion
        - Demographics: gender and age
    - Course Management System (CMS) features **(changing)**
        - Total # logins to CMS
        - # days since last login
        - Total time spent in CMS
        - Course features: credit hours, level, prefix

Predicting Dropout in E-Learning Courses

- Generate baseline and dynamic dropout risk scores
- Create/send alerts to students & instructors
- recommend customized interventions
  - Example: A student with high risk score due to financial aid status is directed to work with the financial aid staff.

Source: [4] R. Bukralia et al, 2015

Predicting Dropout in E-Learning Courses

- Highest accuracy achieved by Decision Trees

		Dropped (0)	Completed (1)	Sum	Per class accuracy (%)	Overall accuracy (%)	ROC area
DT (J.48)	**Dropped (0)**	177	53	230	76.95	73.52	0.739
	Completed (1)	86	209	295	70.84		
Naïve Bayes	**Dropped (0)**	165	71	236	69.10	67.80	0.759
	Completed (1)	98	191	289	66.08		
LR	**Dropped (0)**	183	71	254	72.04	71.23	0.786
	Completed (1)	80	191	271	70.47		
ANN (MLP)	**Dropped (0)**	184	64	248	74.19	72.76	0.778
	Completed (1)	79	198	277	71.48		
SVM	**Dropped (0)**	173	60	233	74.24	71.42	0.714
	Completed (1)	90	202	292	69.17		

Source: [4] R. Bukralia et al, 2015

Predicting Dropout in E-Learning Courses

- Boosted Decision Trees provided even higher accuracy.
- Credit hours <= 68, GPA <=3.563, Age <= 22 , then the student would drop out.

		Dropped (0)	Completed (1)	Sum	Per class accuracy (%)	Overall accuracy (%)	ROC area
DT (J.48)	**Dropped (0)**	177	53	230	76.95	73.52	0.739
	Completed (1)	86	209	295	70.84		
DT (C5.0 without Boosting)	**Dropped (0)**	218	27	245	88.97	86.29	0.886
	Completed (1)	45	235	280	83.92		
DT (C5.0 Boosted)	**Dropped (0)**	229	15	244	93.85	90.67	0.965
	Completed (1)	34	247	281	87.90		

Source: [4] R. Bukralia et al, 2015

Degree Planners

Degree Planners

- Given:
  - A student, his major and the courses that he took

- Find a set of courses that
  - Satisfy degree requirements
  - Satisfy some timing constraints
  - Satisfy prerequisites
  - Students are expected to perform well at

Recommendations with Prerequisites

- take course prerequisites into consideration in order to generate valid course recommendations

- find a short path to fulfill degree requirements and reduce time to graduation

- Problem Formulation
  - Given pre-computed course recommendation scores
  - Recommend a set A of k courses whose pre-requisites were satisfied or are part of the recommended set s.t. the aggregate over the course scores in A is maximized.
  - NP-hard
  - Use approximate algorithms

[Source: 20] A. G. Parameswaran et al, 2009

Recommendations with Prerequisites

- Algorithm 1
  - Set A: select k courses whose prerequisites were satisfied and have highest scores
  - Greedily replace a course inside A with a course from outside A until total score cannot be further grown or there are no more eligible courses.
- Algorithm 2
  - Each set = a course and its prerequisites
  - Sets are sorted by average course score and inserted into a priority queue
  - Insert one set at a time in A until it has k items, and update sets in queue
- Algorithm 3
  - Set A: k course with highest scores regardless prerequisite satisfaction.
  - Prerequisite courses are incrementally added to replace items that can be removed without interrupting any prerequisites.

[Source: 20] A. G. Parameswaran et al, 2009

Part III:
Case Studies

Student Life Dataset

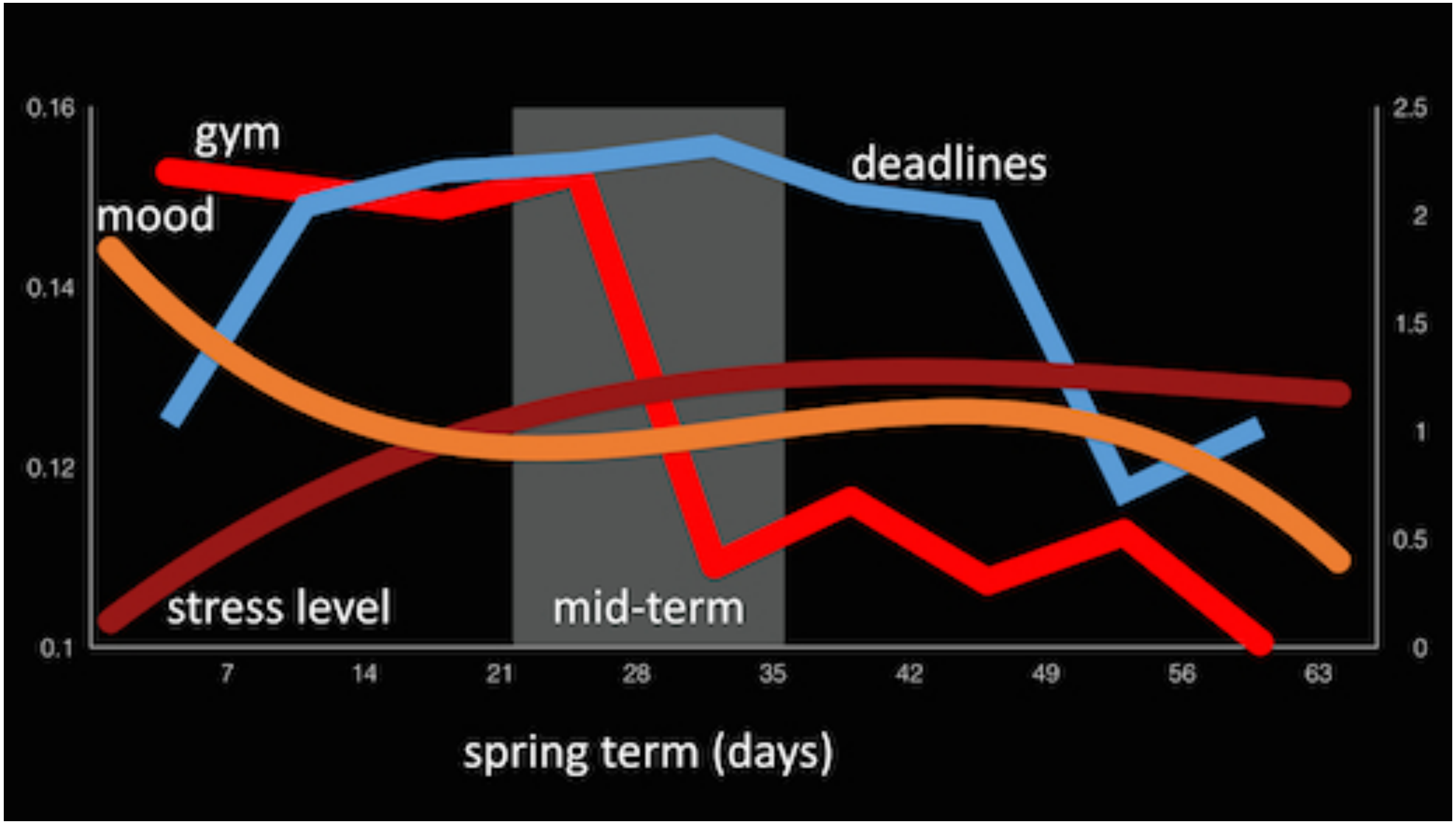
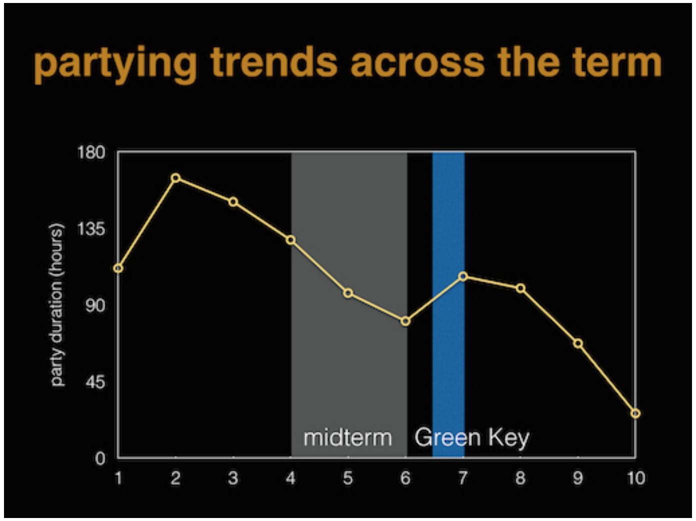


Image Credit: http://studentlife.cs.dartmouth.edu/
[28] R. Wang et. al. 2014]

SmartGPA [Wang et. al. 2015]

- Use passive sensing data (from phones) and self-report to understand different student behaviors between high and low-performing students.

- Predict GPA using ubiquitous smart phones.



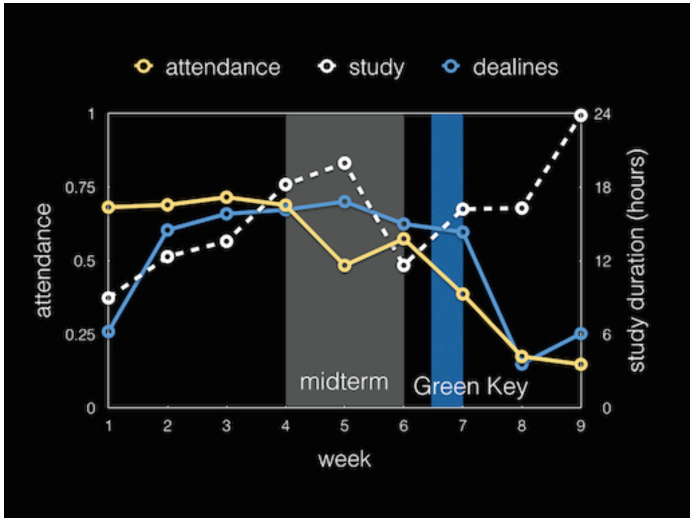


Image Credit: http://studentlife.cs.dartmouth.edu/

SmartGPA

- GPA Correlation Analysis

Table 3: Spring Term GPA Correlations.

	features	r	p-value
automatic sensing	activity term-slope	-0.551	0.002
	activity post-slope	-0.576	0.001
	activity night term-slope	-0.431	0.017
	activity night post-slope	-0.654	< 0.001
	activity day term-slope	-0.411	0.024
	activity day post-slope	-0.442	0.016
	activity evening term-slope	-0.485	0.007
	conversation freq night breakpoint	0.379	0.039
	indoor mobility term-slope	-0.606	< 0.001
	indoor mobility pre-slope	0.423	0.020
	indoor mobility post-slope	-0.515	0.004
	indoor mobility night term-slope	-0.529	0.003
	indoor mobility night pre-slope	0.365	0.047
	indoor mobility night post-slope	-0.543	0.002
	indoor mobility day term-slope	-0.568	0.001
	indoor mobility day post-slope	-0.371	0.048
	indoor mobility evening term-slope	-0.552	0.002
	dorm duration term-slope	0.437	0.016
	social duration dorm pre-slope	0.363	0.049
	party duration mean	-0.398	0.029
	study duration mean	0.381	0.038
	study duration pre-slope	0.397	0.030
survey	Perceived Stress Scale (post)	-0.405	0.050

Source: [29] R. Wang et al, 2015

Example: Academic Pathways

- Goal: Increase awareness of academic advisors and prospective students about the course choices that can lead to improved academic success.
- Approach: Uncover academic path (sequence of courses) that high- and low- achieving students follow & identify contrasting patterns between the two cohorts.
- Data

Major	Group	# students	# courses
CS	High achieving (GPA>=3.0)	208	330
	Low achieving (GPA<3.0)	199	295
INFT	High achieving (GPA>=3.0)	73	224
	Low achieving (GPA<3.0)	53	183

- Method: **Apriori-based frequent pattern mining**
  - A set of courses are considered frequent if the % of students taking them at the same semester is more than a pre-specified threshold (support set to 25%).

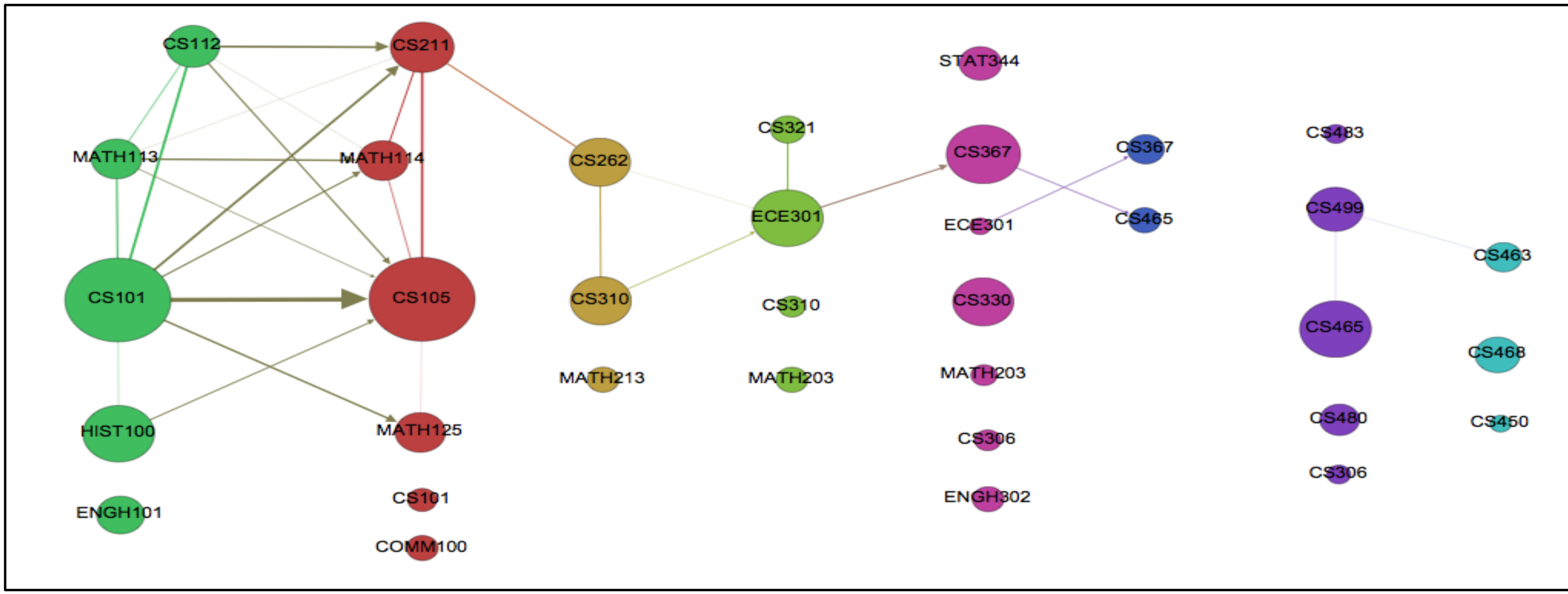


Figure 1: Trajectory of frequent courses for high achieving students (CS)

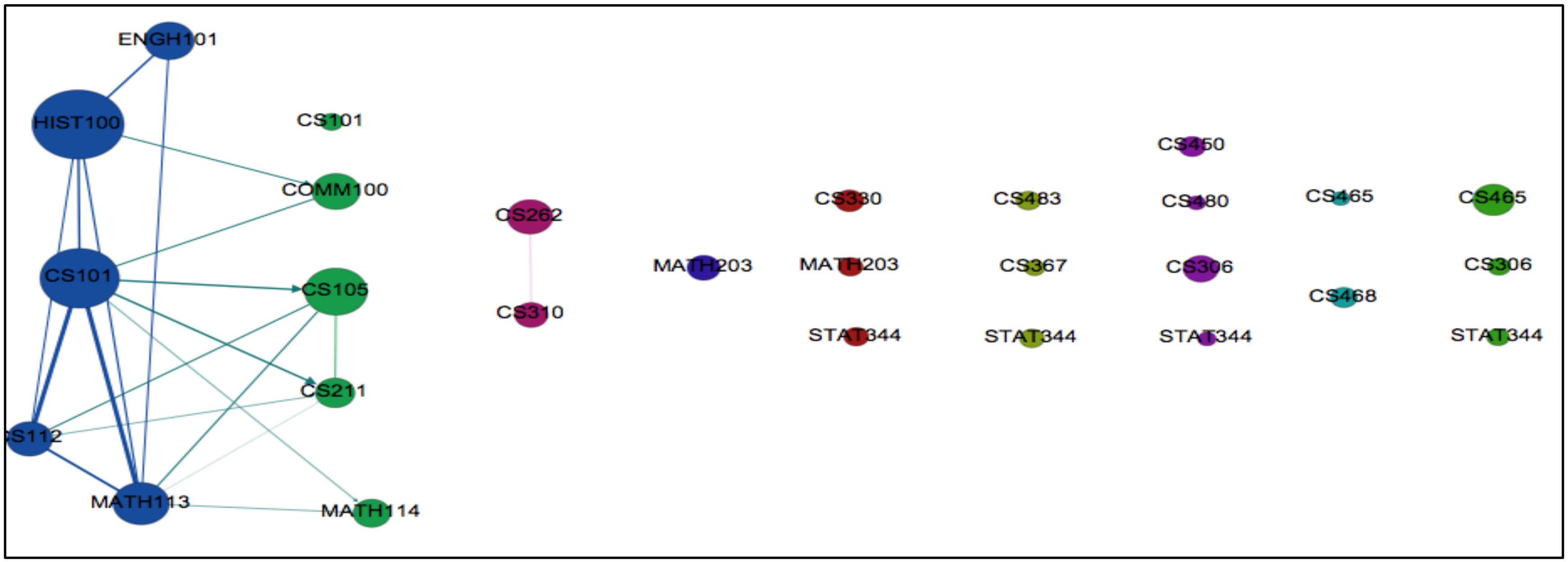


Figure 2: Trajectory of frequent courses for low achieving students (CS)

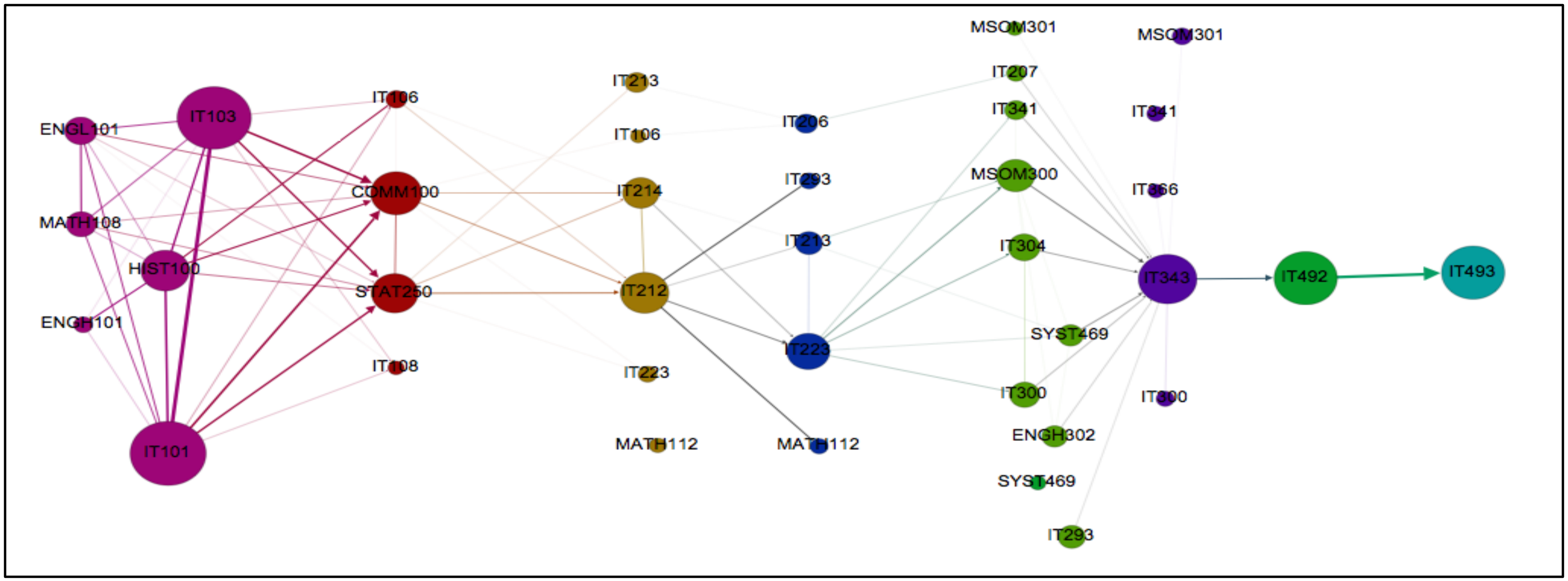


Figure 3: Trajectory of frequent courses for high achieving students (INFT)

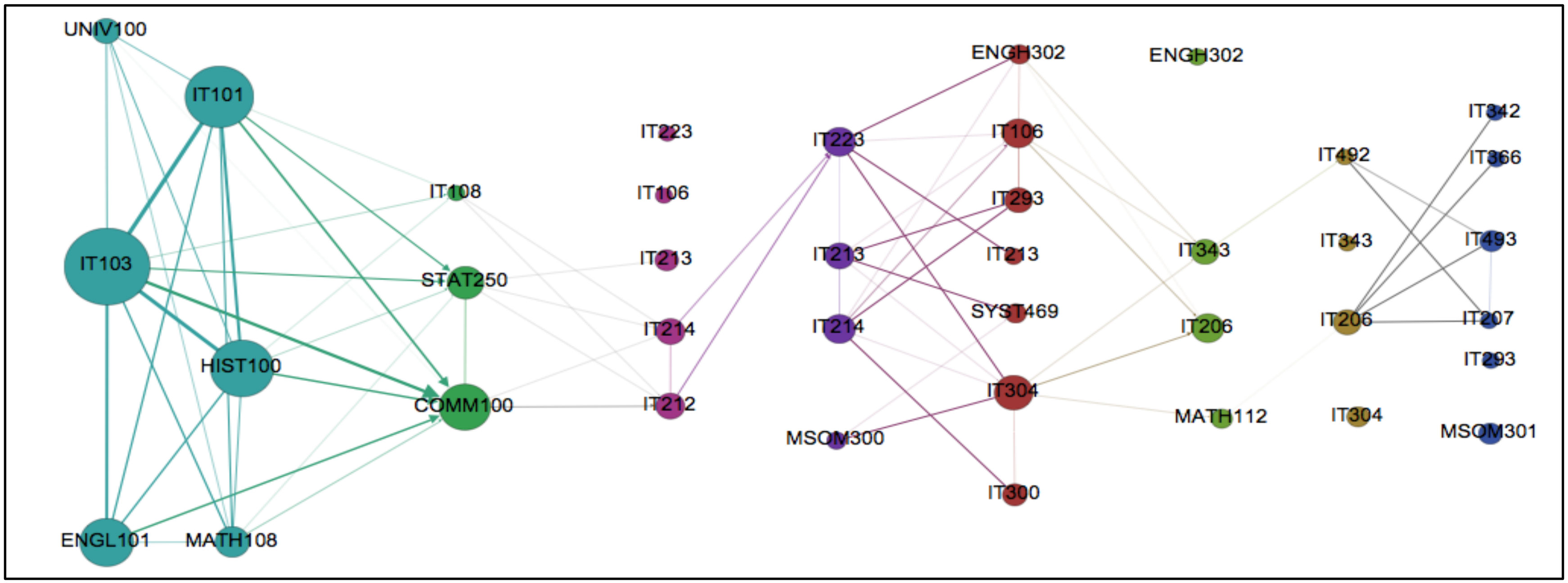


Figure 4: Trajectory of frequent courses for low achieving students (INFT)

Findings

1. Low performers enrolled in multiple courses together at the same semester that their high performers do not usually take concurrently

2. Low performers postponed some courses until the end of the program

3. Low and high performers take certain courses in different semester than the other group

Part IV:
Pertinent Challenges

Analytics for Educational Attainment [FTC Report]

- “Identify students for advanced classes who would otherwise not have been eligible for such classes based on teacher recommendations alone.”

- “Identify students who are at risk of dropping out and in need of early intervention strategies”

- “Big data analytics to demonstrate how certain disciplinary practices, such as school suspensions, affect African-American students far more than Caucasian students, thereby partly explaining the large discrepancy between the graduation rates of these two groups”.

Fairness, Ethics and Privacy

- Big Data: A Tool for Inclusion or Exclusion? Understanding the Issues [FTC Report]

- Risk Scenarios:
  - Results in some individuals mistakenly denied opportunities based on the action of others (Algorithm Bias/Data Bias)

  - “Schools might be tempted to push students into programs of study based on predictive analytics, instead of personal passions and interest. Flagging students as at-risk might discourage them further or negatively alter their professor’s opinion of them.” – See Higher Educatioan Marketing Blog

  - Privacy Challenges Surrounding Sharing of Datasets [FERPA Laws]

Responsible Education Data Research

See: http://gsd.su.domains/sample-policies/ (Mitchell Stevens)

- Shared Understanding – Different stakeholders produce data and should have a shared understanding regarding the limits of use.
- Transparency – Clarifying the process and evaluating each component in a complex environment.
- Informed Improvement – Improve educational processes & contribute to general learning.
- Open Futures – Always enable opportunity and never foreclose it.

Purdue Signals Project – Actionable Insight?

- "Detects early warning signs and provides intervention to learners who may not be performing to the best of their abilities before they reach a critical point." – Course Signals Homepage.

- Features:
  - Early Interventions starting at the second week of class
  - Real time, Frequent and ongoing feedback

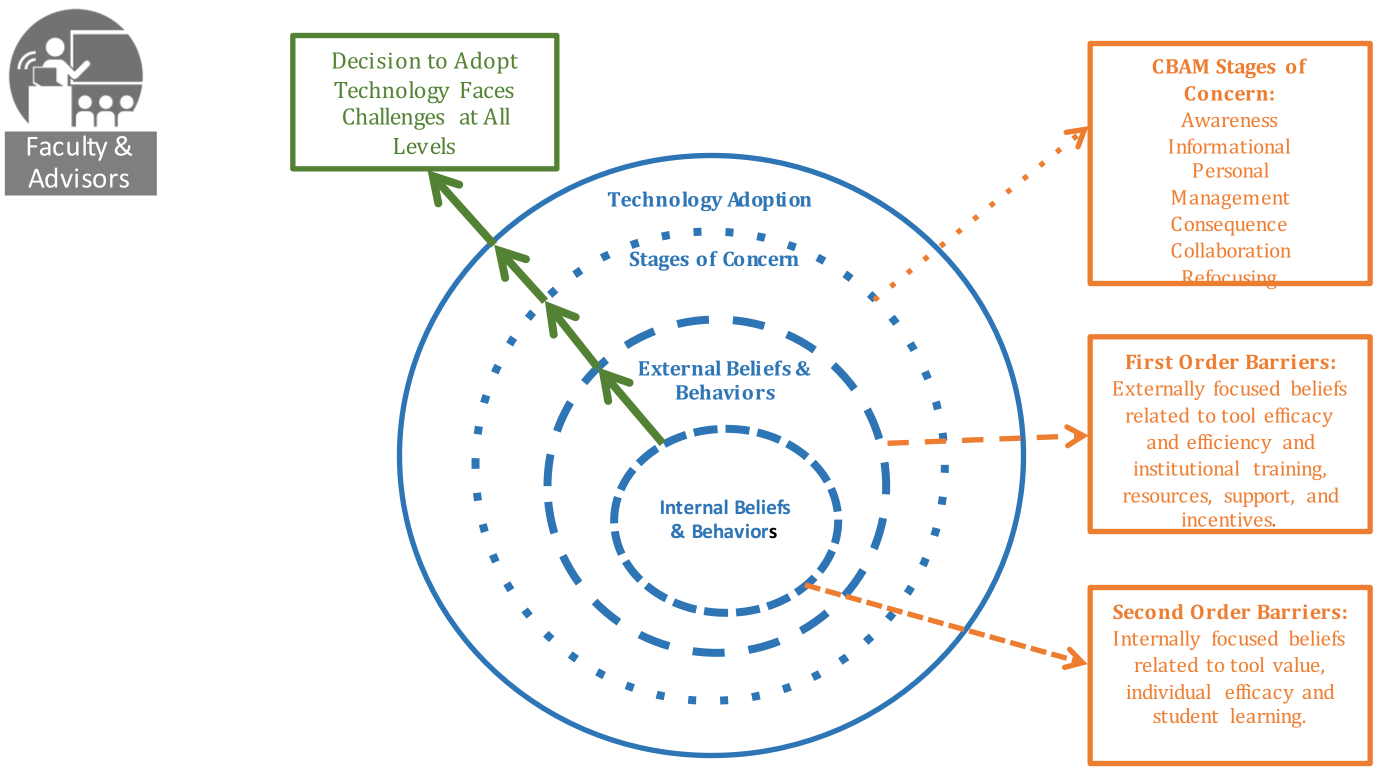
Evaluation of Learning & Advising Dashboards: Incentives & Barriers to Faculty, Advisor, & Student Use

Carrie Klein, Jaime Lester, Huzefa Rangwala & Aditya Johri
George Mason University

	& Infrastructur	Institution & Resources	Individual Decision-Making
Faculty & Advisors	• Component use is role dependent • Used to communicate, build community • Tool alignment necessary • Clear, accurate, real-time data visualizations • Desire tailored training and education • Concerns about bias of predictive data	• Desire for more guidance purpose, policies (FERPA), and process • Want inclusion in choice and implementation • Want planning, training, recognition and rewards for use	• Dependent on trust of LD accuracy, and efficacy of tool • Dependent on user confidence, prior experience, and biases related • Impacted by time, institutional support, leadership mandates, and rewards systems • Impacted by assessment of tool
Students	• Use phones and computers to access • Want more dynamic mobile options • Want anonymity and personalization • Want contextual data visualizations • Want control over data and communication • Desire for a one-stop shop for academic record • Want academic and job data alignment	• Want in-person support and translation of data • Feel there is no choice but to use tool • Frustrated by lack of alignment between tool and university structures • Wish all faculty would use tool	• Trust academic data from perceived experts (faculty and advisors) • Distrust of predictive data, especially for majors suggestions • Visualizations without context are disregarded • Will not use data that prescribes outcomes • Will use data that suggests possibilities

Faculty & Advisor ET Tool Nested Adoption Model:

Carrie Klein, Jaime Lester, Huzefa Rangwala & Aditya Johri
George Mason University



Learning Dashboard Data: Student Decision-Making Tension Points

Carrie Klein, Jaime Lester, Huzefa Rangwala & Aditya Johri
George Mason University

Students

Desire for personalized and tailored communication and assessment

Desire to remain anonymous and not be "singled out"

Desire to receive information and alerts based on performance

Often ignore automated alerts or want to decide when and how to receive information

Desire to know more about themselves and their potential

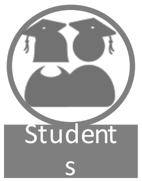
Mistrust in predictive data: "No one can tell me what I can or can't do"

Trust in visualized data when accurate, contextual, and legitimized (via expertise of faculty or advisor)

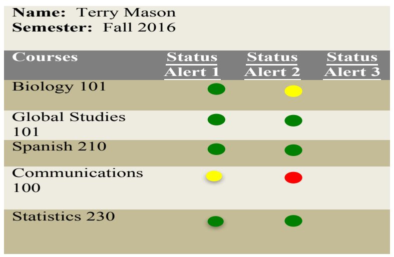
Despite trust, may question data or act in opposition of intervention (believing own assessment over algorithm)

Examples of Student-Evaluated Dashboard Mockups

Carrie Klein, Jaime Lester, Huzefa Rangwala & Aditya Johri
George Mason University



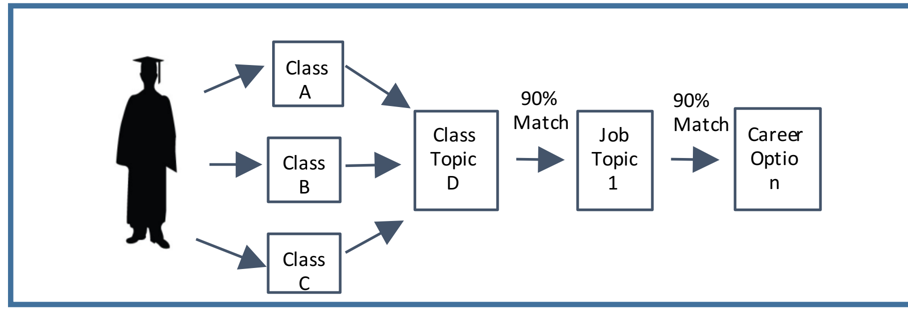
Signals-Based Alert Model



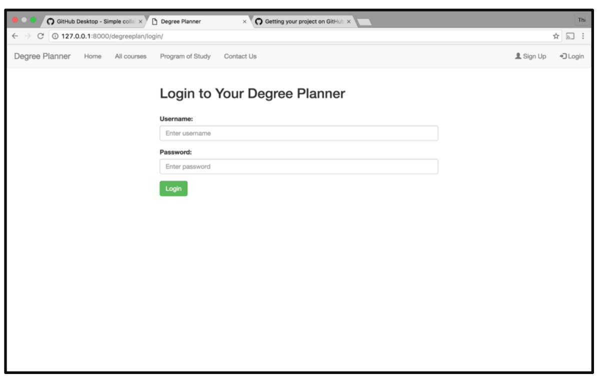
Name: Terry Mason
Semester: Fall 2016

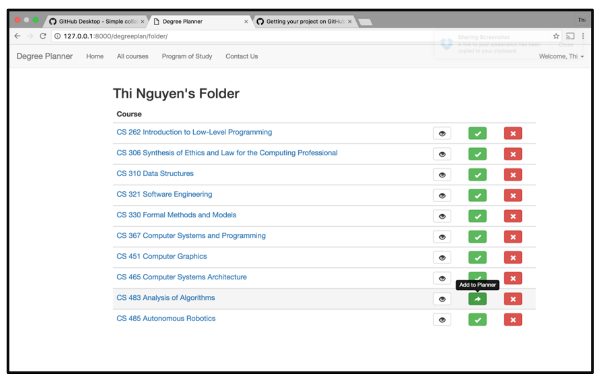
Courses	Status Alert 1	Status Alert 2	Status Alert 3
Biology 101			
Global Studies 101			
Spanish 210			
Communications 100			
Statistics 230			

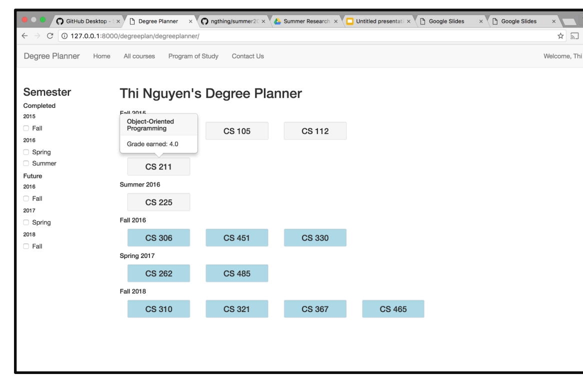
Suggested Career Paths Dashboard Designed by Abigail Justen



Suggest Comprehensive Degree Planner Designed by Thi Nguyen







Take Home Messages

- Opportunities Galore for Data Mining Application and Innovation
  - Predictive Methods – Matrix Factorization
  - Grouping and Clustering
  - Rule Mining
- Several Key Challenges Remain
  - Data is sensitive.
  - Disconnect between Algorithm Output and Decisions.
  - Actions/Decisions have Consequences.
- Impact Unbounded
  - Assisting all stakeholders: Students, Instructors, Administrators and Public.
  - Learning is Life-Long.

Acknowledgements

- BIGDATA: Learning Data Analytics: Providing Actionable Insights to Increase College Student Success
- Grant Numbers: 1447489 and 144778

Students:

- Mack Sweeney, Michel Rouly, Carrie Klein, Zhiyun Ren, Qian Hu, Abigail Jungsten, Thi Nguyen, Ameer Takaddien, GMU
- Asma Elbadrawy, Agoritsa Polyzou and Sara Morsy, UMN

Co-Investigators:

- Profs. Huzefa Rangwala, Aditya Johri and Jaime Lester, GMU
- Profs. George Karypis, Thomas Brothen and Nikos Sidiropoulus, UMN

References

[1] J. R. Anderson, C. F. Boyle, and B. J. Reiser. Intelligent tutoring systems. Science (Washington), 228(4698):456-462, 1985.

[2] R. Barber and M. Sharkey. Course correction: Using analytics to predict course success. In Conference on Learning Analytics and Knowledge, LAK, 2012.

[3] D. Britz. Recurrent neural networks tutorial, part 1, introduction to rnns.

[4] R. Bukralia, A. V. Deokar, and S. Sarnikar. Using Academic Analytics to Predict Dropout Risk in E-Learning Courses, pages 67-93. Springer International Publishing, 2015.

[5] F. T. Commission et al. Big data: A tool for inclusion or exclusion? understanding the issues. FTC Report, 2016.

[6] A. T. Corbett and J. R. Anderson. Knowledge tracing: Modeling the acquisition of procedural knowledge. User modeling and user-adapted interaction, 4(4):253-278, 1994.

[7] G. Dekker, M. Pechenizkiy, and J. Vleeshouwers. Predicting students drop out: A case study. In EDM, 2009.

References

[8] A. Elbadrawy and G. Karypis. Domain-aware grade prediction and top-n course recommendation. In RecSys, pages 183-190, 2016.

[9] A. Elbadrawy, A. Polyzou, Z. Ren, M. Sweeney, G. Karypis, and H. Rangwala. Predicting student performance using personalized analytics. IEEE Computer, 49(4):61-69, 2016.

[10] A. Elbadrawy, R. S. Studham, and G. Karypis. Collaborative multi-regression models for predicting students' performance in course activities. Learning Analytics And Knowledge, 2015.

[11] B. R. Hawkins W.J., Heernan N.T. Learning Bayesian knowledge tracing parameters with a knowledge heuristic and empirical probabilities. 2014.

[12] J. L. Herlocker, J. A. Konstan, A. Borchers, and J. Riedl. An algorithmic framework for performing collaborative filtering. In SIGIR, 1999.

[13] T. Hon Jie and A. Johri. Experts learn more (than newcomers): An exploratory study of argumentation in an online help forum. In Computer Supported Collaborative Learning, 2013.

[14] M. Khajah, R. V. Lindsey, and M. C. Mozer. How deep is knowledge tracing? CoRR, 2016.

References

[15] Y. Koren, R. Bell, and C. Volinsky. Matrix factorization techniques for recommender systems. Computer, 42(8), 2009.

[16] P. Leitner, M. Khalil, and M. Ebner. Learning Analytics in Higher Education – A Literature Review, pages 1-23. Springer International Publishing, 2017.

[17] S. Morsy and G. Karypis. Cumulative knowledge-based regression models for next-term grade prediction.

[18] C. Olah. Understanding LSTM networks. August 27, 2015.

[19] A. Parameswaran, P. Venetis, and H. Garcia-Molina. Recommendation systems with complex constraints: A course recommendation perspective. ACM Trans. Inf. Syst., 2011.

[20] A. G. Parameswaran and H. Garcia-Molina. Recommendations with prerequisites. RecSys '09, pages 353-356, 2009.

[21] Z. A. Pardos and N. T. Heernan. Kt-idem: introducing item difficulty to the knowledge tracing model. In International Conference on User Modeling, Adaptation, and Personalization, pages 243-254. Springer, 2011.

References

[22] C. Piech, J. Spencer, J. Huang, S. Ganguli, M. Sahami, L. J. Guibas, and J. Sohl-Dickstein. Deep knowledge tracing. CoRR, 2015.

[23] K. G. Polyzou A. Grade prediction with course and student specific model. 2016.

[24] B. Reinhardt. Generating case oriented intelligent tutoring systems.

[25] S. Rendle. Factorization machines with libFM. ACM Trans. Intell. Syst. Technol., 3(3):57:1-57:22, May 2012.

[26] M. A. Santana, E. B. Costa, B. F. dos S. Neto, I. C. L. Silva, and J. B. A. Rego. A predictive model for identifying students with dropout proles in online courses. In EDM (Workshops), volume 1446 of CEUR Workshop Proceedings, 2015.

[27] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl. Item-based collaborative ltering recommendation algorithms. WWW '01, New York, NY, USA, 2001.

[28] R. Wang, F. Chen, Z. Chen, T. Li, G. Harari, S. Tignor, X. Zhou, D. Ben-Zeev, and A. T. Campbell. StudentLife: assessing mental health, academic performance and behavioral trends of college students using smartphones. In Proceedings of the Joint Conference on Pervasive and Ubiquitous Computing, pages 3-14. 2014.

References

[29] R. Wang, P. Hao, X. Zhou, and A. T. Campbell. SmartGPA: How smartphones can assess and predict academic performance of college students. GetMobile, 19(4):13-17, 2015.

[30] Gong Y., Beck J.E., Heffernan N.T. Comparing Knowledge Tracing and Performance Factor Analysis by Using Multiple Model Fitting Procedures. In Intelligent Tutoring Systems. ITS 2010.

[31] E. Wenger. Articial Intelligence and Tutoring Systems: Computational and Cognitive Approaches to the Communication of Knowledge. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1987.

[32] G. G. Yudelson M.V., Koedinger K.R. Individualized Bayesian knowledge tracing models.