Accurate and Personalized Academic Advising

“M1 MLDM Internship defense presentation”

Mohammad Poul Doust

Supervisors: Luis Galárraga, Gonzalo Gabriel Méndez

IRISA, INRIA, Universite Rennes 1, 35042 Rennes, France

July 1st, 2019

Overview

Outline

Introduction:

- Electrical and Computer Engineering College of ESPOL University in Guayaquil Ecuador.
- Academic history for students from 1978 to 2012.

Goal: Use this data to build a proactive and interpretable Academic Advising System.

Introduction: Dataset

- A total of 8924 records of students grades from different departments.
- Grades are real values in the range [0 - 10].
- 6 is the passing grade.
- Part of the the data was scrapped (Course's Credit)

Introduction - Academic Advising

- There are selective and mandatory courses
- Some courses are difficult for some student, while easy for others.
- Students usually do not have idea about the nature of future courses
- Simulating student performance for future semesters will help to take proactive actions:
  - **Adjust semester load**
  - **Pay attention to specific courses**

Outline

Related Work - Educational Data Mining

- Educational Data Mining (EDM) is an emerging research field.
- Use Data Mining, Machine Learning and Statistics in educational settings.
- **Applications**:
  - Academic Advising
  - Curriculum Design
  - Identify academic problems
  - Dropout paths identification
  - etc..

Related Work

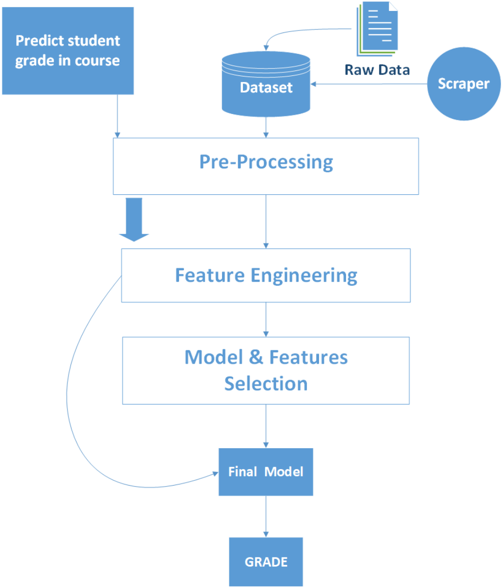
Various studies have been done in the area of Educational Data Mining:

- **Mendez et al.** [8]: Performs Data driven analysis to find Course Difficulty Estimators and Dropout paths **(2014)**
- **Burman et al.** [2]: Psychological features to predict student's performance using SVM **(2019)**
- **Nupur et al.** [5]: Predicting GPA using academic history using SVM, Decision Trees **(2019)**
- **Chan et al.** [4]: Predicting GPA using Collaborative Filtering **(2016)**

Our work focuses on predicting grade using academic history along with hand-crafted features

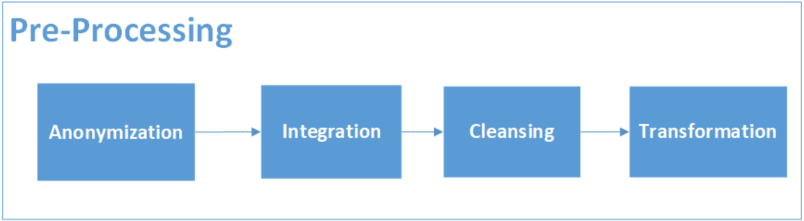
Outline

Approach - Overview



Approach: Data Pre-Processing

- Data Anonymization
- Data Integration
- Data Cleansing ("Garbage in, Garbage out")
  - Data Consistency
  - Data Integrity
  - Missing Values & Duplication Removal
- Data Transformation (Aggregating).



Approach: Data Pre-Processing - Integration & Cleansing

- Data Integration

StudentID	HighschoolType	GradDate	CourseID	Semester	GRADE
1	Particular	14/12/2006	FIEC04341	1S	8,30
1	Particular	14/12/2006	FIEC04358	1S	7
41	Fiscal	14/08/2003	FIEC04341	1T	9
41	Fiscal	14/08/2003	ICF00471	1S	8,60
120	Nacional	Not Graduate	ICF00471	2T	8,10

- Data Cleansing

StudentID	HighschoolType	GradDate	CourseID	Semester	GRADE
1	Particular	14/12/2006	FIEC04341	1S	8.30
1	Particular	14/12/2006	FIEC04358	1S	7
41	Fiscal	14/08/2003	FIEC04341	1S	9
41	Fiscal	14/08/2003	ICF00471	1S	8.60
120	Fiscal	NA	ICF00471	2S	8.10

Approach: Data Pre-Processing - Data Transformation

- Data Transformation

StudentID	HighschoolType	GradDate	FIEC04341	FIEC04358	ICF00471
1	Particular	14/12/2006	8.30	7	NA
41	Fiscal	14/08/2003	9	NA	8.60
120	Fiscal	NA	NA	NA	8.10

Table 1: Data Transformation Sample

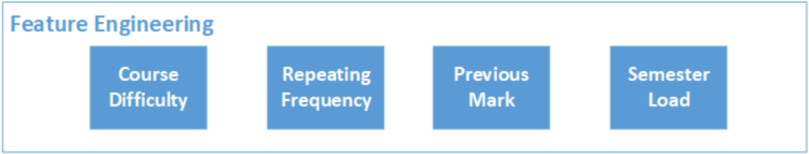
Approach: Features Engineering

Features Engineering involves employing domain knowledge from human expertise to extract meaningful indicators from the raw data [7]:

- Is one of the most fundamental phases
- Affects all the following steps

In this study, several features have been used:

1. Repeating Frequency
2. Last achieved mark in a course (accounting for non repeaters)
3. Semester Load
4. Course Difficulty Estimators [3]



Approach: Features Engineering - Difficulty Estimators

- Course Difficulty Estimators:

$$\alpha_j = \frac{\sum_i GPA_i^2}{\sum_i (r_{ij} * GPA_i)}$$

$$\beta_j = \frac{\sum_i (GPA_i - r_{ij})}{N_s^j}$$

Where:

- GPA_i is the total GPA for student i
- r_{ij} is student's i grade on course j
- N_s^j is the overall number of students enrolled in course j

Approach: Problem and Machine Learning

The main problem is to predict a student grade given previous academic history

- **Input:** Feature vector illustrates student's previous academic history
- **Output:** Numerical value represents a prediction for student's grade in future course.

For **Regression** problem:

- Regularized Linear Regression (**L2 Norm**) has been used for Interpretability and Simplicity.
- Features selection using M5' Algorithm
- Parameter Tuning

Approach: Linear Regression

- **Linear Regression**:

$$J(\theta) = \frac{1}{2m}\left[\sum_{i=1}^{m}\left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right)^2 + \lambda\sum_{j=1}^{n}\theta_j^2\right]$$

$$\begin{aligned} y^{(i)} = &\text{ the real value to predict for instance i} \\ h_\theta(x^{(i)}) = &\text{ the predicted value for instance i} \\ \theta_j = &\text{ the weight for feature j} \\ \lambda = &\text{ ridge parameter} \\ m = &\text{ size of dataset} \\ n = &\text{ size of features set} \end{aligned} \tag{1}$$

- **Example**:

Statistics = (0.09 * CalculusI) + (0.2 * LinearAlgebra) + (0.15 * CalculusII) - (0.15 * CountCalclusII) + (0.07 * PrevCalculusI) - (0.15 * CountLinearAlgebra) + (0.07 * PrevStatistics) + (0.4 * IS_REPEATER) - (0.01 * LOAD) - (0.07 * Beta) + 3.2

Approach: Challenges

Is it better to build one compact unified model, or to have different models for each important course ?

- We have around **2097** different courses in the data
- **4800** different students
- Many courses have low enrolment rate

Building one model (data matrix 4800 x 2097) would be challenging !.

1. Large # Features (Curse of Dimensionality)
2. Average courses Load is 5 (Sparsity)
3. Sparsity might cause Bias.
4. Redundant features (2097 grades to predict one grade !)
5. Heterogeneous Data (Different departments, students)
6. Need Complex model (not interpretable)

Approach: Case Studies

We split the data into different self-contained case studies.

- Different in size
- Different in nature
- Each one will be evaluated separately

Based on the results, we will explore the possibility of building Unified model.

Approach: Case Studies

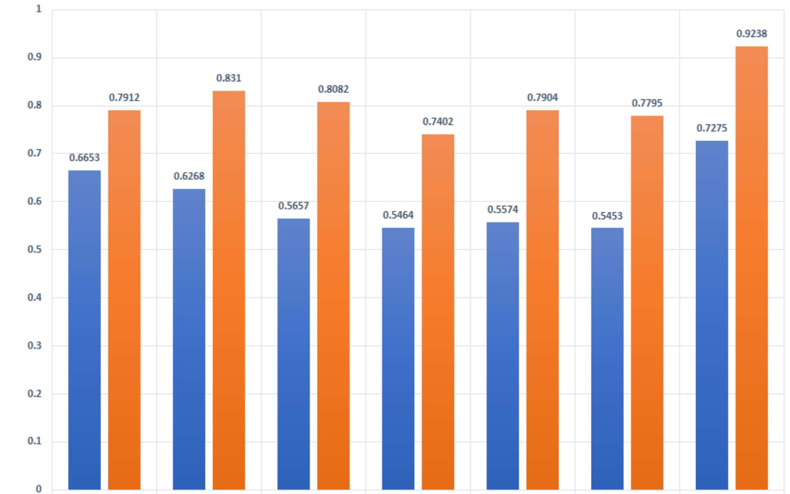
Case Study	Courses	Dataset Size
A - Algorithms Analysis	Programming Fundamentals, Data Structure and Discrete Mathematics	87
B - Statistics I	Multivariable Calculus, Linear Algebra, Differential Calculus and Integral Calculus	654
C - Statistics II (Non-Repeaters[1])	Calculus I, Calculus II and Linear Algebra	938
D - Statistics II (Repeaters[2])	Calculus I, Calculus II and Linear Algebra	566
E - Statistics II (All)	Calculus I, Calculus II and Linear Algebra	1503
F - Analysis of Electrical Networks	Calculus I, Calculus II, Calculus III, Linear Algebra, Physics I and Physics II	974

[1] Students who never took target course before

[2] Students who took the course before

Outline

Results



	Case Study A	Case Study B	Case Study C	Case Study D	Case Study E	Case Study F	Case Study G
MAE	0.6653	0.6268	0.5657	0.5464	0.5574	0.5453	0.7275
RMSE	0.7912	0.831	0.8082	0.7402	0.7904	0.7795	0.9238

MAE RMSE

Results

- Using 10-fold Cross Validation for evaluating

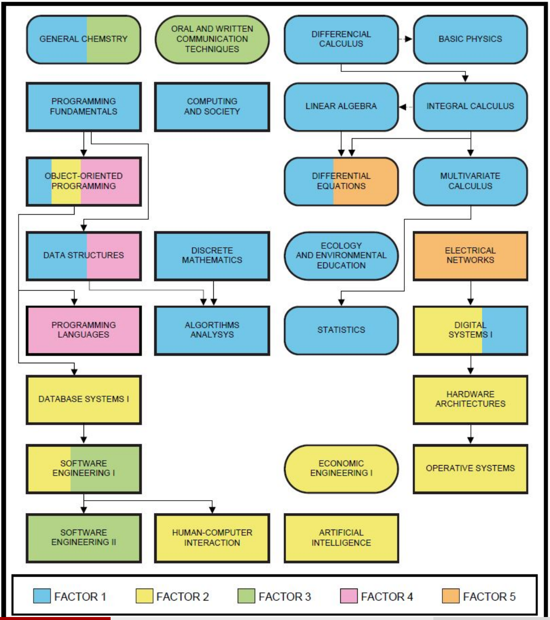
Case Study	MAE	RMSE
A - Algorithms Analysis	0.6653	0.7912
B - Statistics I	0.6268	0.831
C - Statistics II (Non-Repeaters)	0.5657	0.8082
D - Statistics II (Repeaters)	0.5464	0.7402
E - Statistics II (All)	0.5453	0.7795
F - Analysis of Electrical Networks	0.7275	0.9238

Outline

Future Works

- **New Features**:
  - **Fine-Grained**: Difficulty by Professor, Years and Department
  - **Other Difficulty Indicators:** Skewness
- **Unified Model**: **Dimensionality Reduction** to represent courses in term of latent features.
- Student's general performance as features by clustering courses (Mathematics, Humanity,.) or department. (Domain Knowledge)

Future Work - Dimensionality Reduction [8]



Outline

Conclusion

- We were able to predict student grade with low error rate in all experiments (encouraging enough to be deployed).
- The method is robust (**low deviation** in all results between different case studies)
- Provides interpretable information.
- Features:
  - Non-academic features had low impact on the performance
  - Academic history is efficient indicators for the performance
  - Introduced features reduced RMSE by 23%, and MAE by 16%

Questions?

Approach - Features Selection (M5' Algorithm)

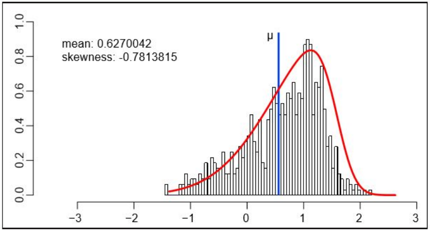
- M5' is tree-based features selection algorithm [1]
- Removes the attribute with the least impurity factor (Standard Deviation Reduction) and evaluates the model each time.
- Quality increased then the attribute is removed.
- This process is repeated until all remaining attributes are important for model quality

Evaluation - 10-fold Cross Validation

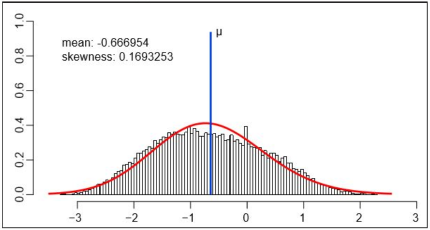
- The evaluations is done using repeated 10-fold cross validation method [6].
- Dataset is divided into 10 equal-sized groups.
- Each time, one group is considered as a testing set while the remaining are used for training a model.
- This process is repeated 10 times and a separate model is trained and evaluated each time.
- The final evaluation would be the average of evaluation for the 10 models.

Future Work - Skewness [8]

- Evaluate Skewness for a course a difficulty measure (The more Negative, The more difficult)



(a) Algorithms Analysis



(b) Oral and Written Communication Techniques

Francisco Azuaje. “Witten IH, Frank E: Data Mining: Practical Machine Learning Tools and Techniques 2nd edition”. In: *BioMedical Engineering OnLine* 5.1 (Sept. 2006), p. 51. ISSN: 1475-925X. DOI: 10.1186/1475-925X-5-51. URL: https://doi.org/10.1186/1475-925X-5-51.

Iti Burman and Subhranil Som. “Predicting Students Academic Performance Using Support Vector Machine”. In: *2019 Amity International Conference on Artificial Intelligence (AICAI)*. IEEE. 2019, pp. 756–759.

Jonathan P Caulkins, Patrick D Larkey and Jifa Wei. “Adjusting GPA to reflect course difficulty”. In: (1996).

Pei-Chann Chang, Cheng-Hui Lin and Meng-Hui Chen. “A hybrid course recommendation system by integrating collaborative filtering and artificial immune systems”. In: *Algorithms* 9.3 (2016), p. 47.

Nupur Chauhan et al. "Prediction of Student's Performance Using Machine Learning". In: *Available at SSRN 3370802* (2019).

Jerome Friedman, Trevor Hastie and Robert Tibshirani. *The elements of statistical learning.* Vol. 1. 10. Springer series in statistics New York, 2001.

Isabelle Guyon et al. *Feature extraction: foundations and applications.* Vol. 207. Springer, 2008.

Gonzalo Mendez et al. "Curricular design analysis: a data-driven perspective". In: *Journal of Learning Analytics* 1.3 (2014), pp. 84–119.