Educational Data Mining / Learning Analytics

Agathe Merceron

Beuth University of Applied Sciences, Berlin
merceron@beuth-hochschule.de
Outline

- About EDM and LA
- Methods and Tasks:
  - Prediction
  - Clustering
  - Relationship Mining
  - Distillation of Data for Human Judgment
  - Discovery with Models
- Current Trends
- Conclusions
- List of References
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. http://www.educationaldatamining.org/
Learning analytics is the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimising learning and the environments in which it occurs. https://tekri.athabascau.ca/analytics/
About EDM and LA

Both fields organized with annual conference, open access journal and society.

EDM 2016
The 9th International Conference on Educational Data Mining
6/29 - 7/2/2016
Raleigh, NC - USA

LAK 16
The 6th International Learning Analytics & Knowledge Conference
University of Edinburgh, Edinburgh, UK, April 25-29, 2016
Methods (Baker & Yacef 2009) come mainly from data mining, machine learning, statistics, classical artificial intelligence, and increasingly from natural language processing.
Outline

- About EDM and LA
- Methods and Tasks:
  - Prediction
  - Clustering
  - Relationship Mining
  - Distillation of Data for Human Judgment
  - Discovery with Models
- Current Trends
- Conclusions
Important task: predict performance.

Different levels of granularity:

- Drop-off (Wolff & al. 2013)
- Pass/fail, mark in a degree (Zimmerman & al. 2015)
- Pass/fail, mark in a course (Lopez & al. 2012)
- Skill mastery in a tutoring system (Pardos & al. 2007).
Many works show that pass or fail, or even the interval of a mark in a degree or a course can be predicted with an accuracy of 70% or higher.

**No classifier that works best in all contexts** (Huang & Fang, 2013).

**No set of features that work best in all contexts**, though some works to predict the interval of the mark for a university degree suggest that including marks is essential (Golding & al. 2006, Zimmerman & al. 2015).
Example of classifier: Decision Tree (Asif & al. 2014).
Prediction: Mark in a degree

- Predict the interval of the degree mark: A, B, C, D or E (Asif & al. 2014).
- 4-years Bachelor Computing and Information Technology in a technical university of Pakistan.
- Competitive: selection on the marks in the High School Certificate (HSC) and entrance exam.
- **Conjecture**: academic records (no socio-economic feature) might be enough to predict the final mark with a reasonable accuracy: better than the baseline of predicting the majority interval C, 51.92%.
Prediction: Mark in a degree
Prediction: Mark in a degree

- **Which features?** HSC marks, marks of all modules from 1\textsuperscript{st} and 2\textsuperscript{nd} year and number of attempts.
- **Which classifiers?** Try all the well-known ones.
- **Validation:** one cohort as training set and the next cohort as test set (needs some stability in the curriculum) for generalization and pragmatic policy. *Different from other works which mostly use cross-validation.*
  - Cohort 1: 105 students graduated in 2012
  - Cohort 2: 104 students graduated in 2013
# Prediction: Mark in a degree

<table>
<thead>
<tr>
<th>Classifier</th>
<th>Accuracy / Kappa</th>
</tr>
</thead>
<tbody>
<tr>
<td>Decision Tree with Gini Index</td>
<td>68.27% / 0.493</td>
</tr>
<tr>
<td>Decision Tree with Information Gain</td>
<td>69.23% / 0.498</td>
</tr>
<tr>
<td>Decision Tree with Accuracy</td>
<td>60.58% / 0.325</td>
</tr>
<tr>
<td>Rule Induction with Information Gain</td>
<td>55.77% / 0.352</td>
</tr>
<tr>
<td>1- Nearest Neighbors</td>
<td>74.04% / 0.583</td>
</tr>
<tr>
<td>Naives Bayes</td>
<td>83.65% / 0.727</td>
</tr>
<tr>
<td>Neural Networks</td>
<td>62.50% / 0.447</td>
</tr>
<tr>
<td>Random Forest with Gini Index</td>
<td>71.15% / 0.543</td>
</tr>
<tr>
<td>Random Forest with Information Gain</td>
<td>69.23% / 0.426</td>
</tr>
<tr>
<td>Random Forest with Accuracy</td>
<td>62.50% / 0.269</td>
</tr>
</tbody>
</table>
Clustering

- Big variety of tasks.
- Variety of algorithms.
- Two works:
  - Clustering students to find out typical behaviours in a forum (Cobo & al. 2012)
  - Clustering utterances to find out speech acts or dialog acts (Ezen-Can & al. 2015).
Colors show an optimal clustering.

(Tan, Kumar & Steinbach, 2005)
Clustering: behaviours in forum

- **8 features**: 4 for writing, 4 for reading:
  - Number of initiated threads, number of reply posts, number of students replied, number of days with writing.

- **Hierarchical agglomerative clustering**:
  - All features calculated as ratio.
  - 2 clusterings: writing features and reading features.
  - Normalized Euclidean distance, complete link.
  - Adaptation of inconsistency criterion to isolate the best clusters.
  - Clusters from the 2 clusterings are combined.
Clustering: behaviours in forum

- Find known results: less students write than read.
- The smaller the reading, the higher the drop-off rate and fail rate.
- Results (Cobo & al. 2012)
Clustering: dialog acts

- When we talk, we do something (http://en.tintin.com/).
Dialog acts:

- Question: “What is an anonymous class?”.  
- Answer: “An anonymous class is a class without name.”.  
- Issue, problem: “this program does not compile”.  
- Statement: “this assignment is long”.  
- Reference: “An interesting video about Bubblesort.”.  
- Positive, negative acknowledgment: “Thanks, I got it”, “I am still confused”.  

Problem: classify automatically sentences in forums or tutorial dialogs in dialog acts.
Clustering: dialog acts

- Classical approach is supervised:
  - Annotate manually a large corpus (bottle neck).
  - Identify cues or features: punctuation, unigram, bigram, position of unigram in the sentence, preceding dialog act, etc. (Kim & al. 2010).
  - Train a classifier. Support Vector Machine (Kim & al. 2010):
    - Positive_ack: F-Score 0.54 (9.20% of the sentences).
    - Questions: F-Score 0.95 (55.31% of the sentences).
Clustering: dialog acts

- **Unsupervised approach** ([Ezen-Can & al. 2015](#)). Dialogues come from a computer mediated environment to tutor students on programming. Students recorded by Kinect cameras.

- **Features to describe sentences:**
  - **Lexical features**: unigram, word ordering, punctuation.
  - **Dialog-context features**: position in the dialog, length, author of previous message (tutor, student), etc..
  - **Task features**: task before the utterance (writing, compiling), status of most recent coding action, etc..
  - **Posture features**: head distance, torso distance.
  - **Gesture features**: one hand and two hands to head.
Clustering: dialog acts

- K-Medoids algorithm with Bayesian Information Criterion (BIC) to infer the optimal number of clusters.
- Distance between utterances: cosine + longest common subsequence for lexical features.
- 7 clusterings according to the previous dialog act of tutors.
- The majority vote in each cluster gives the dialog act.
- A new utterance is predicted according to the cluster with the nearest center.
- Leave-on-Student-out validation: 67% average accuracy, 61% without posture and gesture features.
Relationship Mining

- **Association rules mining:** if students make mistake A, they also make mistake B (Merceron & Yacef 2005).
- **Correlation mining:** negative correlation between gaming a tutoring system and post-test (Baker & al. 2004).
- Preliminary statistics.
- Visualizations. Here too data preparation is crucial.
  - LeMo project (Fortenbacher & al. 2013)
Heatmap: marks of all students in all courses of a 4 years Bachelor degree, technical university:

- First year courses on the left, then 2nd year courses, 3rd year courses and on the right 4th year courses.
X-means clustering year wise (Asif et al. 2015):
- Euclidean distance, Tool: Rapid Miner.
- Gives 4 clusterings.

Clusterings are combined.

Heatmap shows now the groups of students with low marks, average marks and high marks, and give hints about courses that could act as detectors.
Building on (Baker & al. 2004), (Baker & al. 2006) proposes a model for gaming the system. Features include:

- Number of times a specific problem is wrong across all problems.
- Probability that a student knows a skill.
- Various times: time taken for the last 3 actions, 5 actions
- Etc…

Generalize to new lessons and new students.

This detector is used with new data to discover more patterns such as in (SanPedro & al. 2015): What happens to students who game the system?
Outline

- About EDM and LA
- Methods and Tasks:
  - Prediction
  - Clustering
  - Relationship Mining
  - Distillation of Data for Human Judgment
  - Discovery with Models
- Current Trends
- Conclusions
Current Trends

- Natural Language Processing: tutorial dialogues, essays, forums.
- Multimodal Analysis: data from the educational system + data from camera, from EEG etc.
- Multilevel Analysis: different levels of analysis with the data recorded by the system.
Relating level forum and performance level (Merceron 2014) in a programming course of a LMS over 4 years:

- Posts manually labelled with dialog acts: questions, issues, answers, references, positive and negative acknowledgments.
- Hypothesis: questions and issues come preliminary from low achieving students.
After removing an outlier: high achieving students had more questions (and much more answers) than low achieving students.
Outline

- About EDM and LA
- Methods and Tasks:
  - Prediction
  - Clustering
  - Relationship Mining
  - Distillation of Data for Human Judgment
  - Discovery with Models
- Current Trends
- Conclusions
Conclusions

- Numerous approaches.
- Numerous tasks.
- Numerous findings.

What is not a reality yet is the analysis of educational data on a routine basis to understand learning and teaching better and to improve them.
Conclusions

Challenges:

- Privacy: Opt-in. Limit the available data, hence the findings and validity of the results.
- Generalizability: is a classifier to predict performance still valid 2 years later, or in another degree? Not sure. Most probably Data Scientists needed.
Comments? Ideas? Questions?
Thank you for your attention!

Data Science Group, Beuth University of Applied Science
https://projekt.beuth-hochschule.de/data-science/


References


References


References


References


