Stage Predicting Student Stay Time Length on Webpages of Online Courses based on Grey Models

Qingsheng Zhang, Kinshuk, Sabine Graf, and Ting-Wen Chang
Athabasca University, Canada
National Chung Cheng University, Taiwan
sabineg@athabascau.ca

Motivation

Student Modelling:

- try to get various information about a student

- More and more research is done on automatic student modelling

- Automatic student modelling means to infer students' characteristic (e.g., learning styles, cognitive abilities, etc.) from their behaviour in a online course

Motivation

- One of the most often used variables for automatic student modelling is the time that students spent on certain learning objects (e.g., content).

- However, time is a problematic variable since it can include a lot of noise (e.g., student is doing something else, last learning object of the learning session, etc.)

- In this paper, we look into the prediction of stay time length of students using stage prediction, power law, and grey models

Research Question and Contributions

- **How to predict the stay time length of students on content?**

- **An approach for such prediction can help in**
  - Filtering noise from real data and therefore, provide more accurate stay time length data
    → improves automatic student modelling
    → improves learner analytics
  - Compare actual length with predicted length and response to significant differences (→ content object might be too difficult or too trivial)
    → improves course design

Looking into Power Law

- Power Law is a specific relationship between two quantities:

$$P(x) = c * x^{-k}$$

- Many relationships are based on this formula

- In the educational domain, this ranges from short term perceptional tasks to team-based long term tasks (Ritter and Schooler, 2001), where the power law describes the relationship between practices and performance

- For more complex skills, decomposition of the skills in each underlying skill again shows power law relationships (Kenneth and Santosh, 2004)

Data

- Data from an online course

- We looked only into data about content

- 91,084 learning events from 459 students

- Threshold for low noise: 2 sec.

- Threshold for high noise: 300, 600, 900, 1200, 1800 sec.

→ More than 50,000 data are used for testing after filtering

Experiment

- **Predicting data using power law and two grey models:**
  - GM (1, 1) for exponential type sequences
  - Verhulst for sequences with saturated trend
- **Prediction is based on the 3 most recent history data**
  - → Subsequence of 3 data is used to predict the next one
  - → Shift to the next subsequence of 3 data and predict the next one
  - → Etc.

Experiment

- **Compare predicted data with actual data**

- **Considered new knowledge concepts by observing the ratio between actual data and predicted data**

- **If this ration exceeds a certain threshold → assume that the student starts learning a new knowledge concept → using next 3 data for constructing a new predicting model**

Results

Ratio	NLV(s)	NHV(s)	AMMRE (%)	Number of predicted points
1	2	600	89.51	104
2	2	600	38.67	4,552
3	2	600	59.00	5,693
4	2	600	76.96	5,773
5	2	600	92.43	5,629
6	2	600	104.84	5,417
7	2	600	118.40	5,188

Ratio	NLV(s)	NHV(s)	AMMRE (%)	Predicted points numbers
1	2	900	90.82	107
2	2	900	38.36	4,548
3	2	900	58.99	5,720
4	2	900	77.75	5,911
5	2	900	93.06	5,802
6	2	900	105.40	5,612
7	2	900	119.21	5,399

Conclusions

- Relative error of 38% is not too bad (e.g., actual value is 1 minute, predicted value is 1:20)

- Results show that using power law and grey models can to a certain extend predict stay time of learners on content pages

- Future research will deal with refining our approach (e.g., by looking into other predictive models, considering complexity of content pages, etc.)