Exploring and Following Students' Strategies When Completing Their Weekly Tasks

Jessica McBroom, Bryn Jeffries, Irena Koprinska and Kalina Yacef   School of Information Technologies, University of Sydney

Introduction

We wish to understand how students manage their learning when presented with tasks of varying difficulty by analysing the order in which students attempt such tasks. We investigate the following questions

1. What strategies do students take in attempting the different tasks each week?
2. Are there differences between the strategies of the regular and advanced students?

We show how to represent and mine data from student tasks with different difficulty levels.

The Data

The data comes from a third year database course with a regular and advanced stream. Students completed weekly homework tasks with the following properties:

1. Different difficulty levels: easy, medium and hard
2. Marks awarded based on most difficult task completed
3. Automated feedback on student submissions given
4. Unlimited attempts allowed before deadline

The data extracted consisted of the marks for every student's attempt on each task.

Strategies

Figures 1 and 2 show the relative frequency of the different strategies taken by all students overall and each week respectively. Strategies are labelled based on the tasks completed, e.g. EH indicates a student completes that week's Easy task first, followed by the Hard task. There are 16 possible strategies. The most common strategies observed are None (30%), E (31%), EM (11%) and EMH (15%).

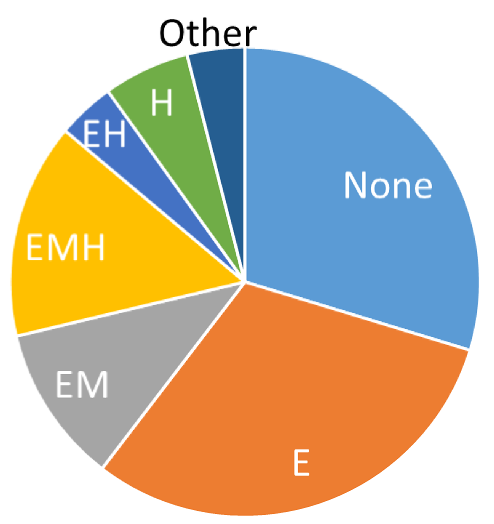

Figure 1: Overall student strategy distribution

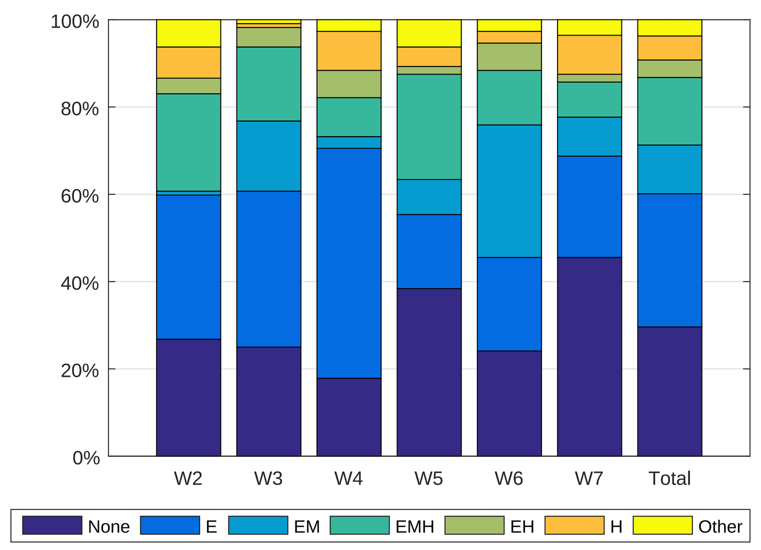

Figure 2: Weekly student strategy distributions

Clustering with missing information

Students were able to test their code at home, which meant we didn't always have complete information about their strategies. For example, a student might have completed all three tasks but only submitted the hardest one!

The Solution

We clustered students based only on the highest difficulty task completed each week (using k-means, k=5, centroids: Fig 2). We then took the mode weekly strategy for each cluster (Table 1), which allowed us to compare student strategies despite the missing information.

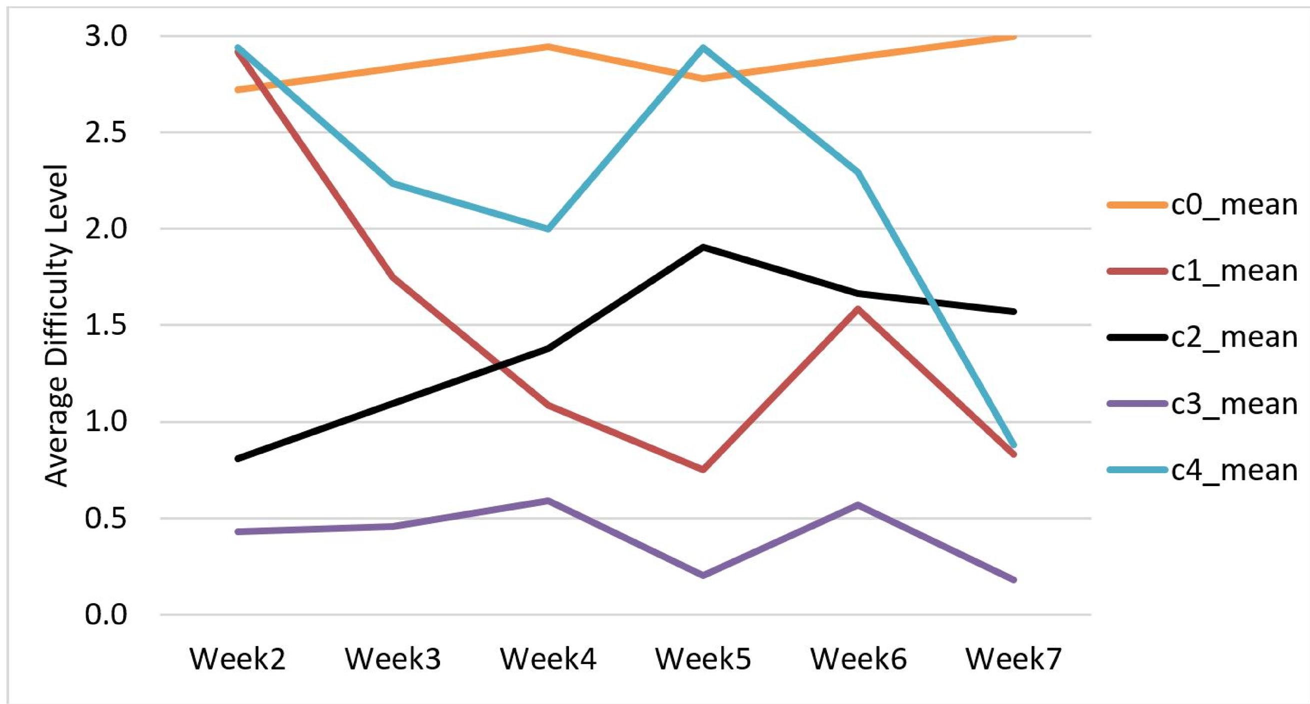

Figure 3: Average highest difficulty task completed by students in each cluster, each week

Cluster	W2	W3	W4	W5	W6	W7	%(adv)
0	EMH	EMH	EMH	EMH	EMH	EMH	9(50)
1	EMH	EM	E	None	EM	None	13(0)
2	E	E	E	E	E	None	18(20)
3	None	None	E	None	None	None	45(15)
4	EMH	EM	E	EMH	EM	E	15(15)

Table 1: Mode weekly strategy per cluster. Last column shows proportion of regular and advanced (in parentheses) students.

Findings

1. In general, both the regular and advanced students began with the easy task and worked up to the hard one
2. Advanced students generally performed better than regular students
3. Many regular students and some advanced students consistently made few attempts, possibly due to the low task weighting

Sliding Window Rule Mining

To find trends in changes of strategy we looked for association rules $X \rightarrow Y$ in which X occurred before Y in time and restricted our analysis to periods of three weeks. We used a sliding 3-week window over each student's strategy vector, similar to rule mining in time-series subsequences [1], but with the time encoded into each item. From these item sets (n = 448) we searched for rules 1a,2b $\rightarrow$ 3c where a, b and c were the strategies used in consecutive weeks. The 5 highest confidence rules are shown in Table 2.

Rule	Support	Confidence	Lift
1None,2None → 3None	14%	85%	2.7
1EMH,2EMH → 3EMH	5%	62%	4.63
1EMH,2EM → 3E	3%	57%	2.00
1None,2E → 3E	3%	45%	1.58
1None,2E → 3None	3%	45%	1.43

Table 2: Highest-confidence rules found using length-3 sliding window rule mining technique

We can analyse these rules as follows:

1. High likelihood of not attempting the task if the student had not submitted two previous tasks

2,3. Student is likely to work through all three tasks progressively if they did so in the previous two tasks

4,5. Many students' strategies were on the borderline between completing the task only or none at all

Conclusion

1. Clustering and rule mining can be applied to data from tasks in which students have choices between several activities
2. It is possible to handle missing information

References and Acknowledgements

[1] Chen Y. Campana B. Hu B. Zakaria J. Keogh E Shokoohi-Yekta, M. Discovery of meaningful rules in time series. *Proc. KDD 2015*, 2015.

[2] A Radu and J. Stretton. Pasta. `http://www.it.usyd.edu.au/~bjef8061/pasta/`. Accessed: 2016-05-20.

Work was funded by the Human-Centred Technology Cluster of the University of Sydney. Poster template adapted from "baposter Landscape Poster", originally created by Brian Amberg, accessed at `http://www.latextemplates.com/template/baposter-landscape-poster`, 25 June 2016.