

Guiding Students towards Frequent High-Utility Paths in an Ill-Defined Domain

Igor Jugo, Božidar Kovačić, Vanja Slavuj

Department of Informatics, University of Rijeka, Croatia

„DITUS” SYSTEM ARCHITECTURE

New architecture of web-based tutoring system „DITUS” enables seamless communication with DM tools as a prerequisite for EDM process automatization. The goal of the new architecture is to provide hints on which units to learn before and/or after a currently selected unit, in order to streamline the learning process.

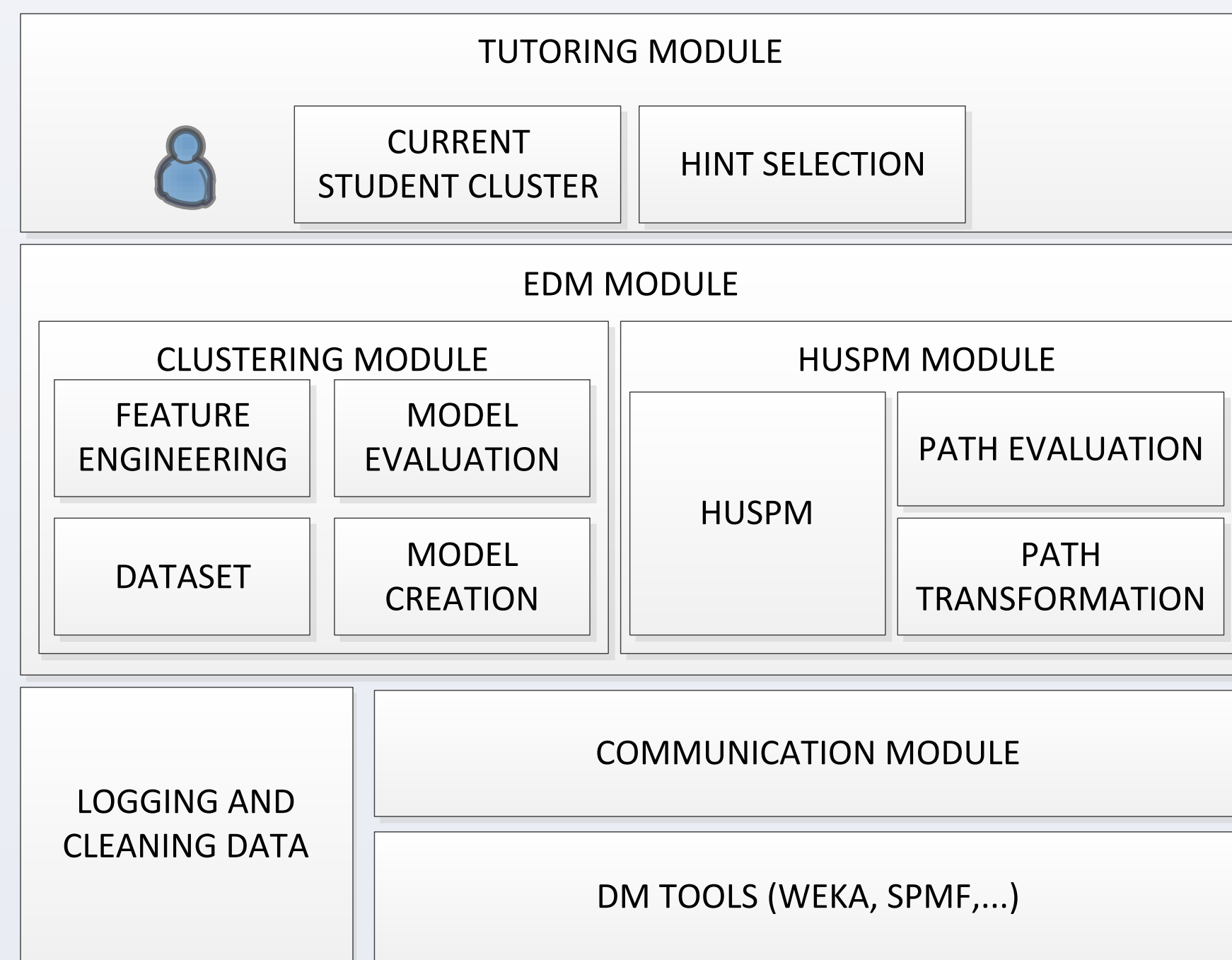


Figure 1. DITUS system architecture

PHASE 1 – CLUSTERING BASED ON ACTIVITY AND EFFICIENCY

In this phase, groups of students are discovered based on their interactions and learning efficiency. The clusters in the best model are then automatically graded based on the quality of their learning process. Student model is expanded by cluster affiliation.

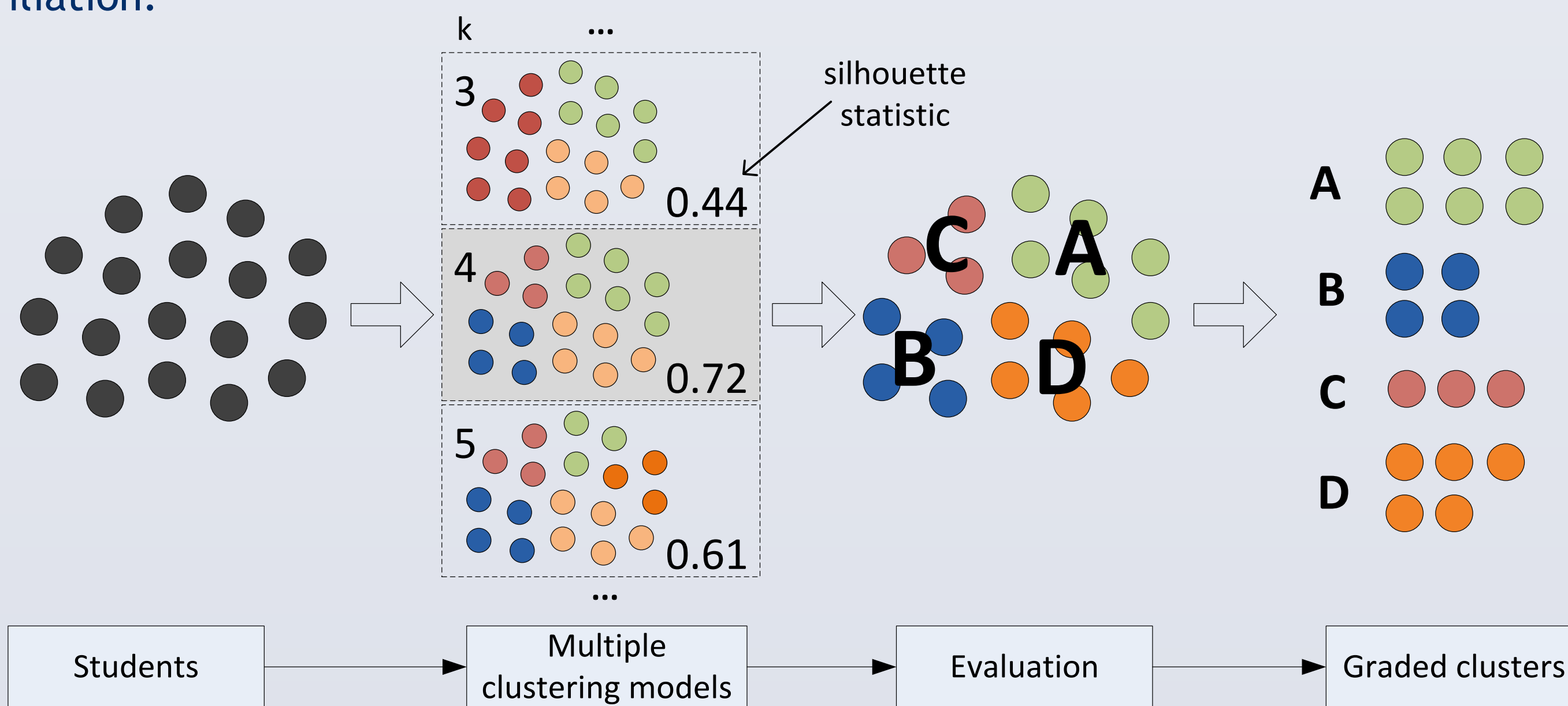


Figure 2. Automated clustering model creation, selection and interpretation

PHASE 2 – HIGH-UTILITY SPM

In this phase, high-utility sequential patterns are discovered from the dataset of student interactions. They provide a basis for hint selection. This process is performed for each cluster. Dataset preparation consists of the following steps: for each student in the cluster, evaluate all paths (*utility/profit*), transform paths into two sets (leading to/from each unit in knowledge domain), add to cluster dataset, perform HUSPM, store patterns ordered by *profit*.

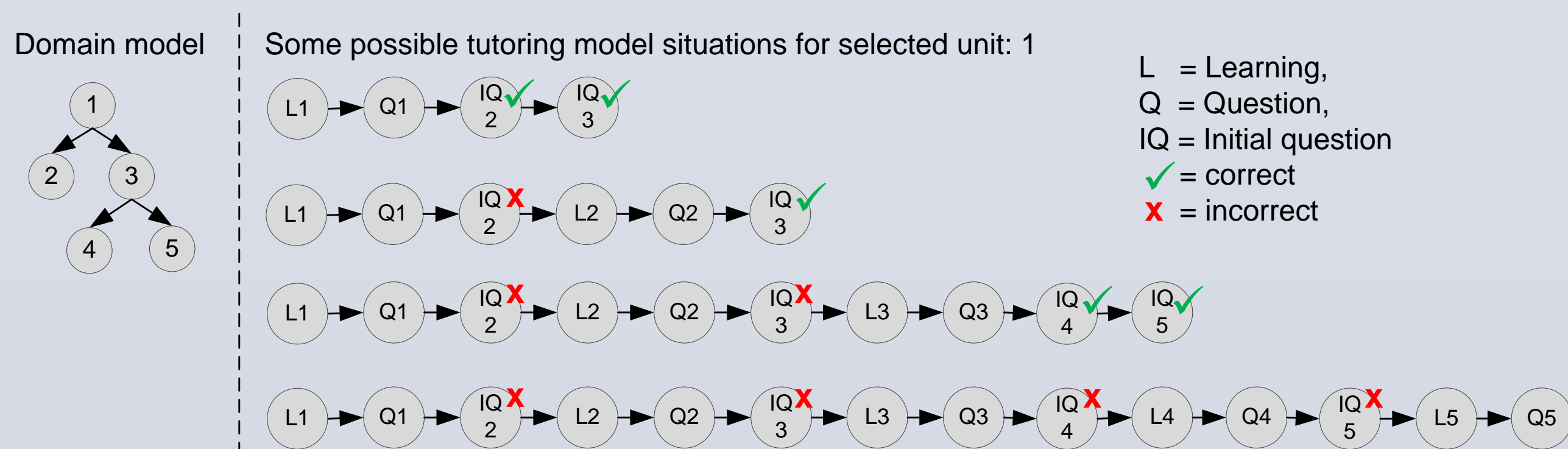


Figure 3. Differences in path efficiency

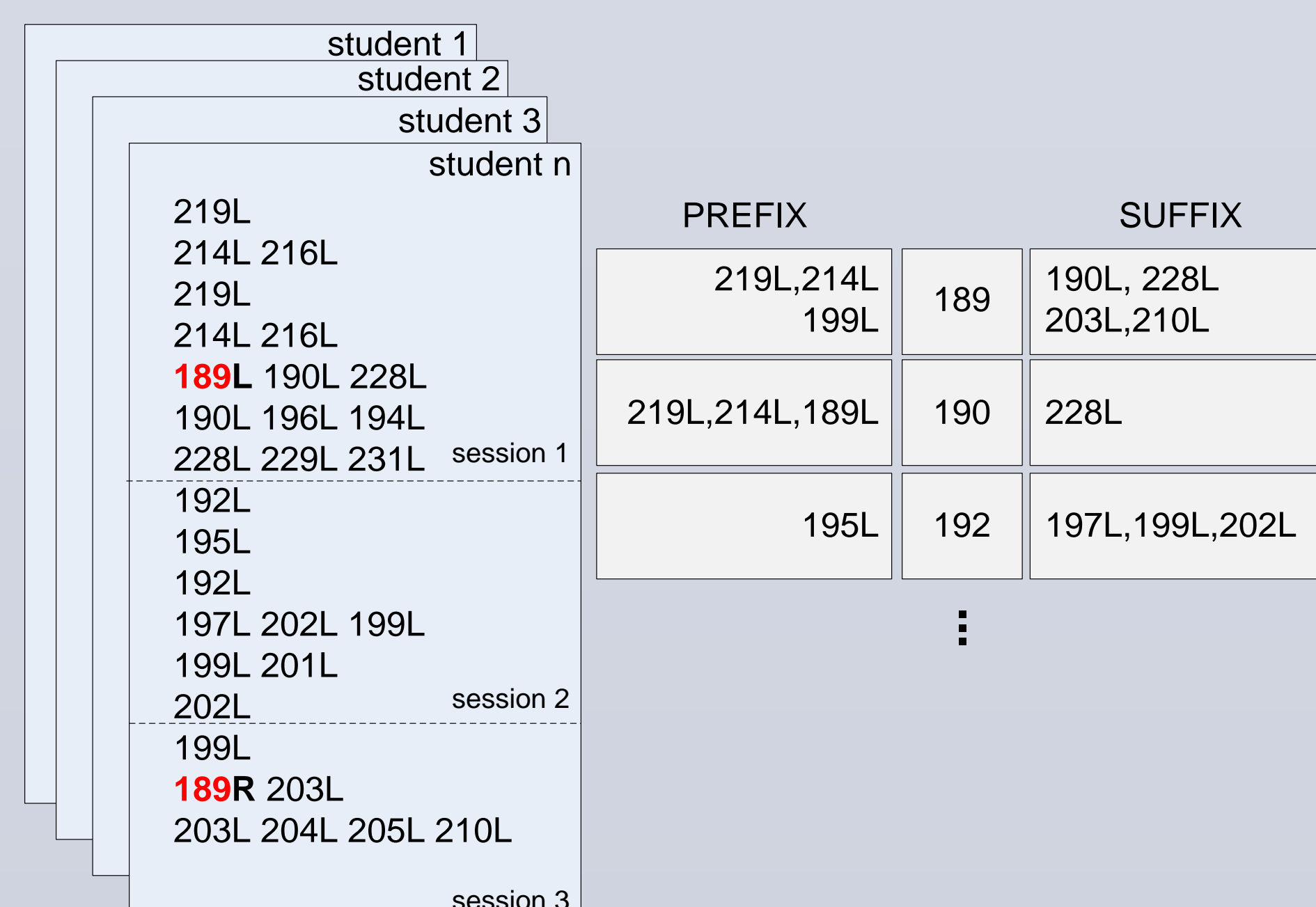


Figure 4. Transforming students' paths into prefix and suffix datasets

PHASE 3 – PRESENTING HINTS

In this phase, the new tutoring model performs hint selection based on the students' current cluster affiliation, currently selected unit and previously presented hints. Prefix hints are presented at the start of the learning activity (for the selected unit), while suffix hints are presented immediately after the learning activity is completed.

EXPERIMENT

Two knowledge domains: D1 („Web technologies”) = 31 units, D2 („Design patterns”) = 36 units. Control and test group: n = 35 3rd year undergraduate students (in each group). Duration : D1 = 4 days, D2 = 7 days. Phase 1 and 2 performed on a growing dataset (every 400 new interactions): D1 = 9, D2 = 13 runs.

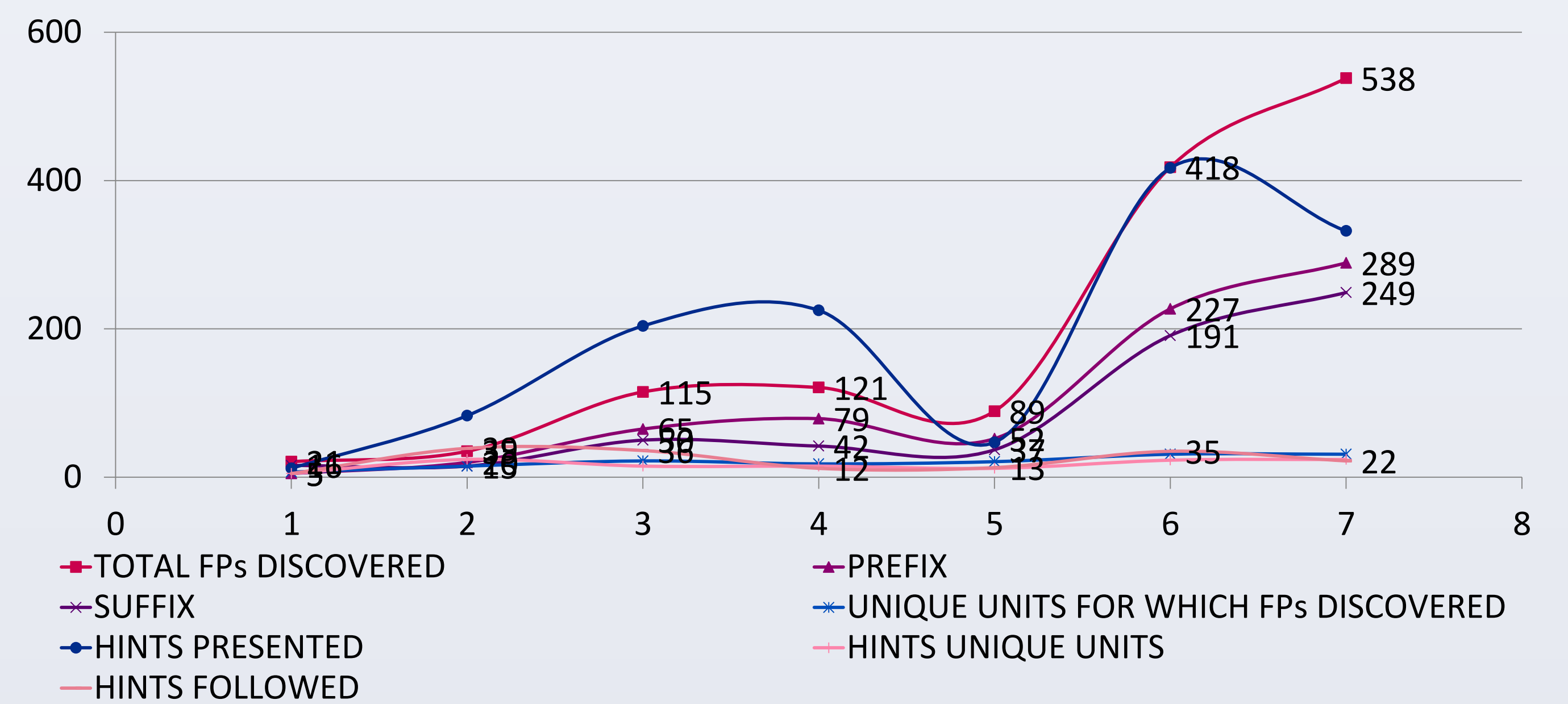


Figure 5. Results of HUSPM and hint selection during D1 experiment

No	Dataset size	Total FPs discovered	Prefix	Suffix	Unique units	Hints presented	Hints unique units	Hints followed
1	1206	21	5	16	7	12	5	5
2	1616	35	20	15	15	83	24	39
3	2068	115	65	50	22	204	15	36
4	2504	121	79	42	18	225	15	12
5	2912	89	52	37	21	47	12	13
6	3314	418	227	191	31	417	23	35
7	3604	538	289	249	31	332	24	22

Table 1. Data for Figure 5

RESULTS

Test group results show improved learning process efficiency - lower number of interactions and less time required to complete the domain (D1 ~ 15%, D2 ~ 5%). Progress visualisation for D1 below: Control group = 12, Test group = 17 students in fastest group.

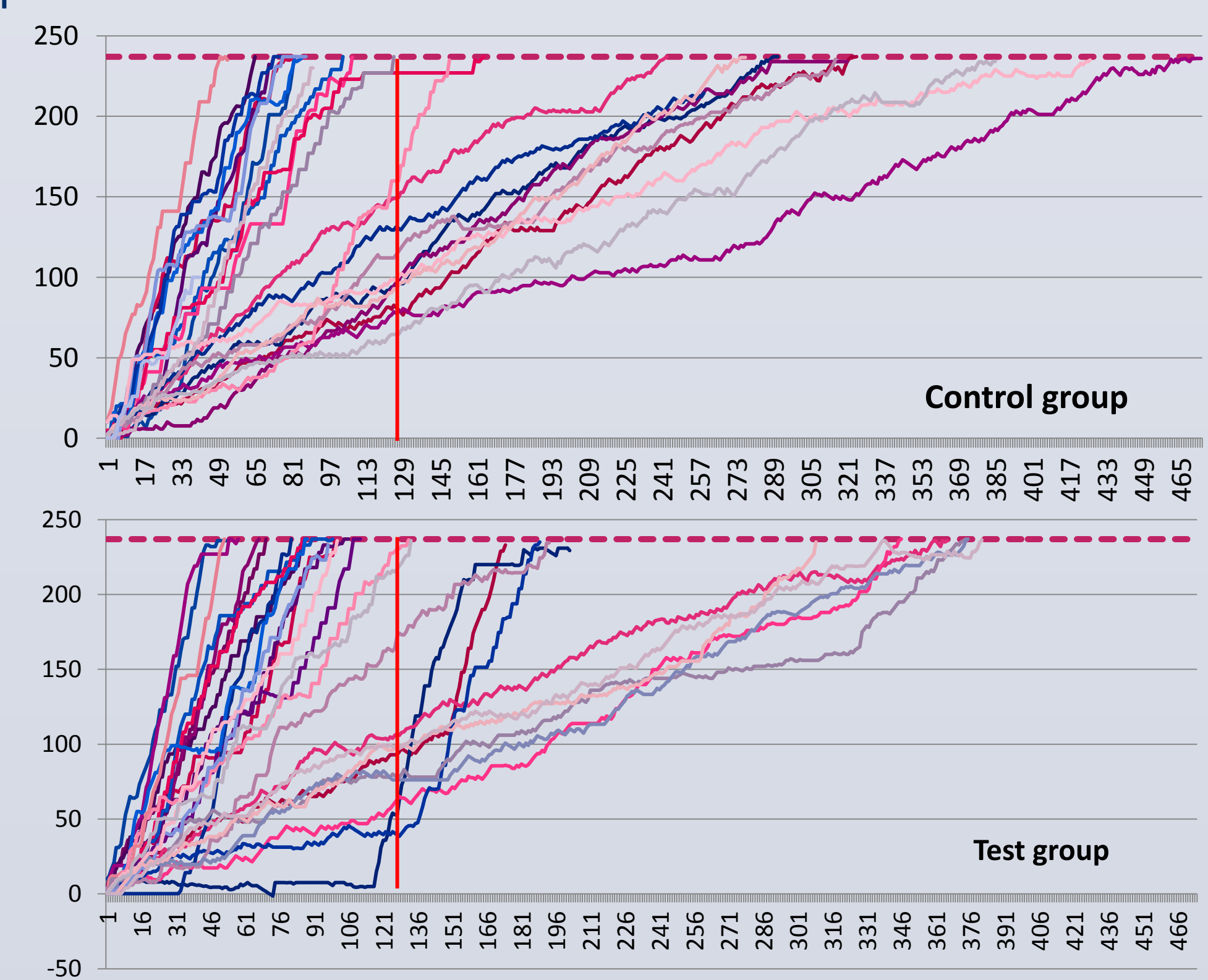


Figure 6. Student progress visualisation for D1 (dashed line = domain completed)

FUTURE WORK

- improvement of clustering interpretation algorithm (exclude inactive students before clustering, add them to a lowest graded cluster, add outlier detection),
- improvement of hint selection algorithm (first select paths regarding selected unit from both current and one grade higher cluster, then order by „profit”, and finally choose unit to present.

ACKNOWLEDGEMENTS

This research is a part of the Project "Enhancing the efficiency of an e-learning system based on data mining", code: 13.13.1.2.02., funded by the University of Rijeka, Croatia.