



UNIVERSITY
AT ALBANY

State University of New York

AIED 2019



Annotated Examples and Parameterized Exercises: Analyzing Students' Behavior Patterns

Mehrdad Mirzaei¹

Shaghayegh Sahebi¹

Peter Brusilovsky²

¹ Department of Computer Science, University at Albany - SUNY, Albany, USA

² School of Computing and Information, University of Pittsburgh, Pittsburgh, USA

Introduction

- Freedom to choose to work with any learning materials in modern online learning systems
- Various students' learning pace and repetitive activities
- Learning materials: parameterized problems and annotated examples

Research Questions

Questions:

- Do individual students exhibit stable behavioral patterns in their work with learning content
- To what extent student behavioral patterns are associated with their learning performance?

Definitions:

- Performance: Learning gain = normalized post score – normalized pre score
- Behavior: Student's interactions with problems and examples

Parametrized Java exercises in Mastery Grids

Introduction to object-oriented programming course

Dataset includes

- 83 Students
- 103 Parameterized exercises
- 42 Annotated examples
- 13796 Correct attempts
- 6233 Incorrect attempts
- 12713 Examples seen

```
public class Tester {  
    public static void main(String[] args) {  
  
        int i = 3;  
        int result = i * 1 + 2;  
  
    }  
}
```

What is the final value of result?

WRONG!

Your Answer:

4

Correct Answer:

5

Try Again

```
public class Tester {  
    public static void main(String[] args) {  
  
        int i = 3;  
        int result = i * 9 + 2;  
  
    }  
}
```

What is the final value of result?

Methodology

- Construct action sequences
 - Label students' actions such as exercises or studying course content
 - Construct sequences from labeled actions
- Extract patterns from sequences
 - Find frequent patterns using a sequential pattern mining algorithm
- Analyzing patterns
 - Compare extracted patterns to acquire meaningful patterns
 - Check the stability of the patterns
- Performance analysis
 - Find the correlation between patterns and student's performance

Building sequences

Attempt	Shorter than median	Longer than median
Successful attempt	s	S
Failed attempt	f	F
Reading example	e	E

Attempt labels
 s: Short success
 S: Long success
 f: Short failure
 F: Long failure
 e: Short exercise
 E: Long exercise

Micro patterns:
 (_Se_, _ee_, _sss_, _ffs_, _FS_, ...)

Student id	Sequence
1	_ee_ee_FFf_F_e_S_
2	_S_S_S_fF_S_se_S_e_Ss_
3	_SS_Sseee_eee_S_S_s_S_S_eee_...
4	_eee_e_eeeeeeee_e_Ssssss_ssssss_...
5	_SFsS_S_Se_Fs_S_Ffs_S_Ss_...

Pattern Vectors

Frequent sequential pattern mining

Find most frequent patterns from sequences by CM-SPAM algorithm

- Minimum support: 1%
- Minimum pattern length: 2

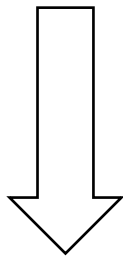
Rank	Pattern	Support	Rank	Pattern	Support
1	ss	1516	6	_Fs	901
2	Ss	1456	7	FS	828
3	ss_	1378	8	Fs	788
4	Fs	1153	9	sss	735
5	_Ss	974	10	Ss_	692

Top 10 extracted patterns ordered by support

Building pattern vectors

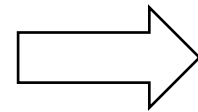
Build individual frequency vectors that show how frequently each pattern appears in student problem solving behavior vector.

P_1	P_2	P_3	P_4			P_{n-1}	P_n
f_1	f_2	f_3	f_4			f_{n-1}	f_n



L1 Normalization

P_1	P_2	P_3	P_4			P_{n-1}	P_n
f'_1	f'_2	f'_3	f'_4			f'_{n-1}	f'_n



$$\begin{bmatrix} f_0^{(1)} & f_1^{(1)} & \dots & f_n^{(1)} \\ f_0^{(2)} & f_1^{(2)} & \dots & f_n^{(2)} \\ \vdots & \vdots & \vdots & \vdots \\ f_0^{(m)} & f_1^{(m)} & \dots & f_n^{(m)} \end{bmatrix}$$

Students' vectors

Vectors Stability

- Are patterns representative of students' traits or depend on:
 - The time of the semester
 - Complexity of the problems
- Randomized Analysis
- Longitudinal Analysis
- Complexity Analysis

Behavior Stability Analysis

- Split the sequences in two equal sets
- Build pattern vector for each pair
- Compare each half with other halves

	Self distance		Distance to others		t Stat	P-value
	Mean	SE	Mean	SE		
Random split	0.2082	0.0207	0.4639	0.0105	-16.0279	<0.0001
First half/second half	0.2995	0.0211	0.5207	0.0113	-12.3501	<0.0001
Random Split (Easy)	0.3644	0.0258	0.5769	0.0110	-9.9099	<0.0001
Random Split (Medium)	0.3266	0.0246	0.5465	0.0092	-11.1404	<0.0001
Random Split (Hard)	0.4219	0.0266	0.5703	0.0106	-6.4266	<0.0001

Comparing average of students' pattern vector distances with themselves vs. other students according to various splits using Jensen-Shannon divergence

Pattern Analysis

Apply clustering on student pattern vectors

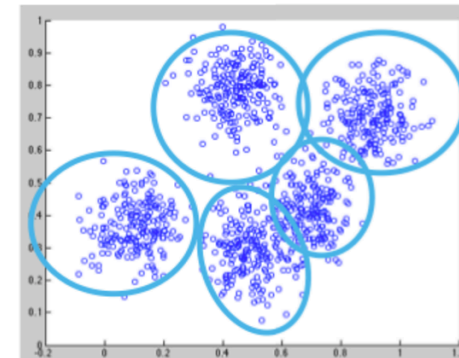
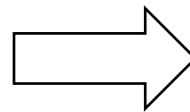
- Spectral clustering
- 3 clusters provide the best result

Compare their average pattern frequencies in the top 30 patterns.

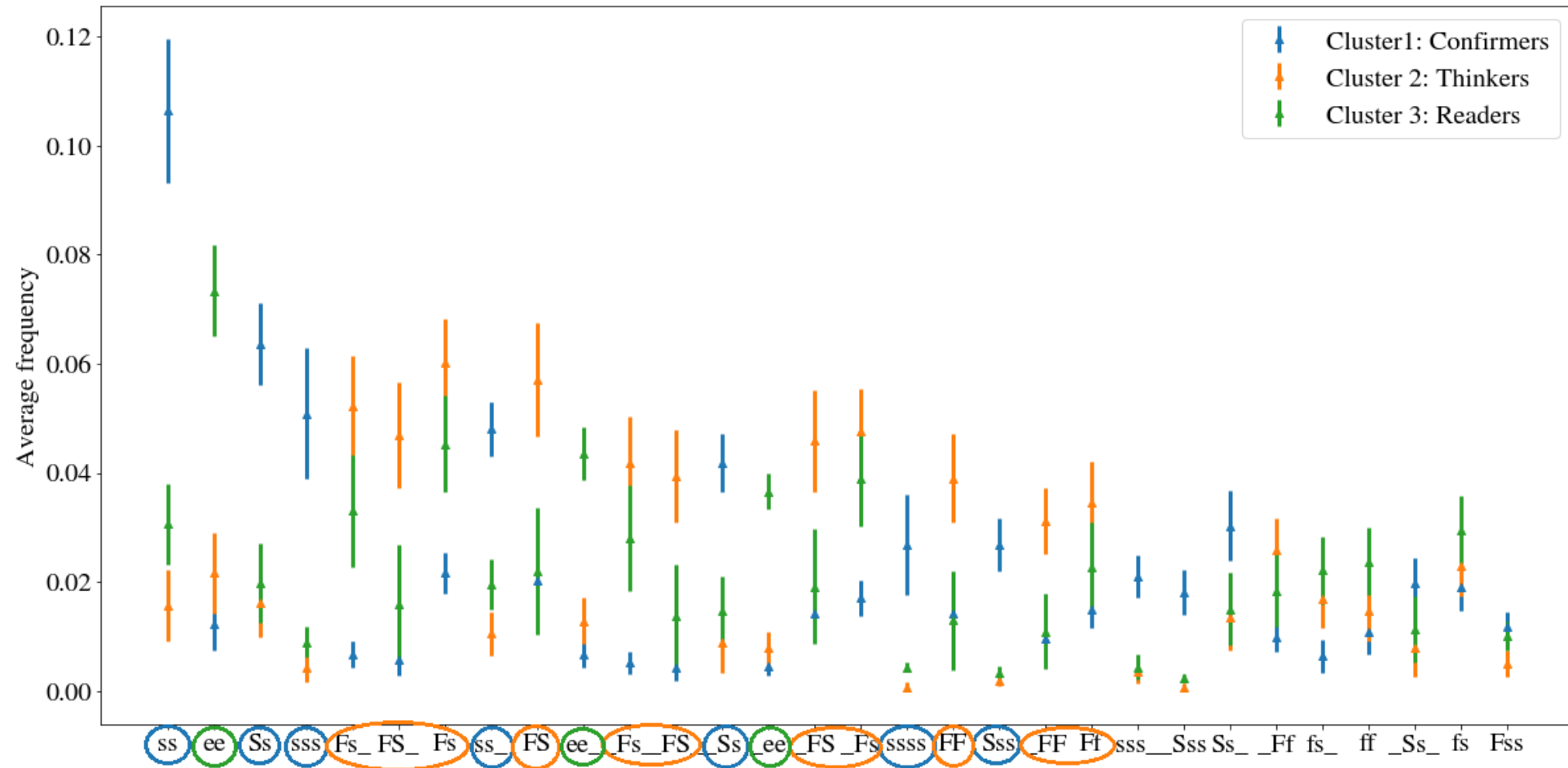
$$\begin{bmatrix} f_0^{(1)} & f_1^{(1)} & \dots & f_n^{(1)} \\ f_0^{(2)} & f_1^{(2)} & \dots & f_n^{(2)} \\ \vdots & \vdots & \vdots & \vdots \\ f_0^{(m)} & f_1^{(m)} & \dots & f_n^{(m)} \end{bmatrix}$$

Students' vectors

Clustering

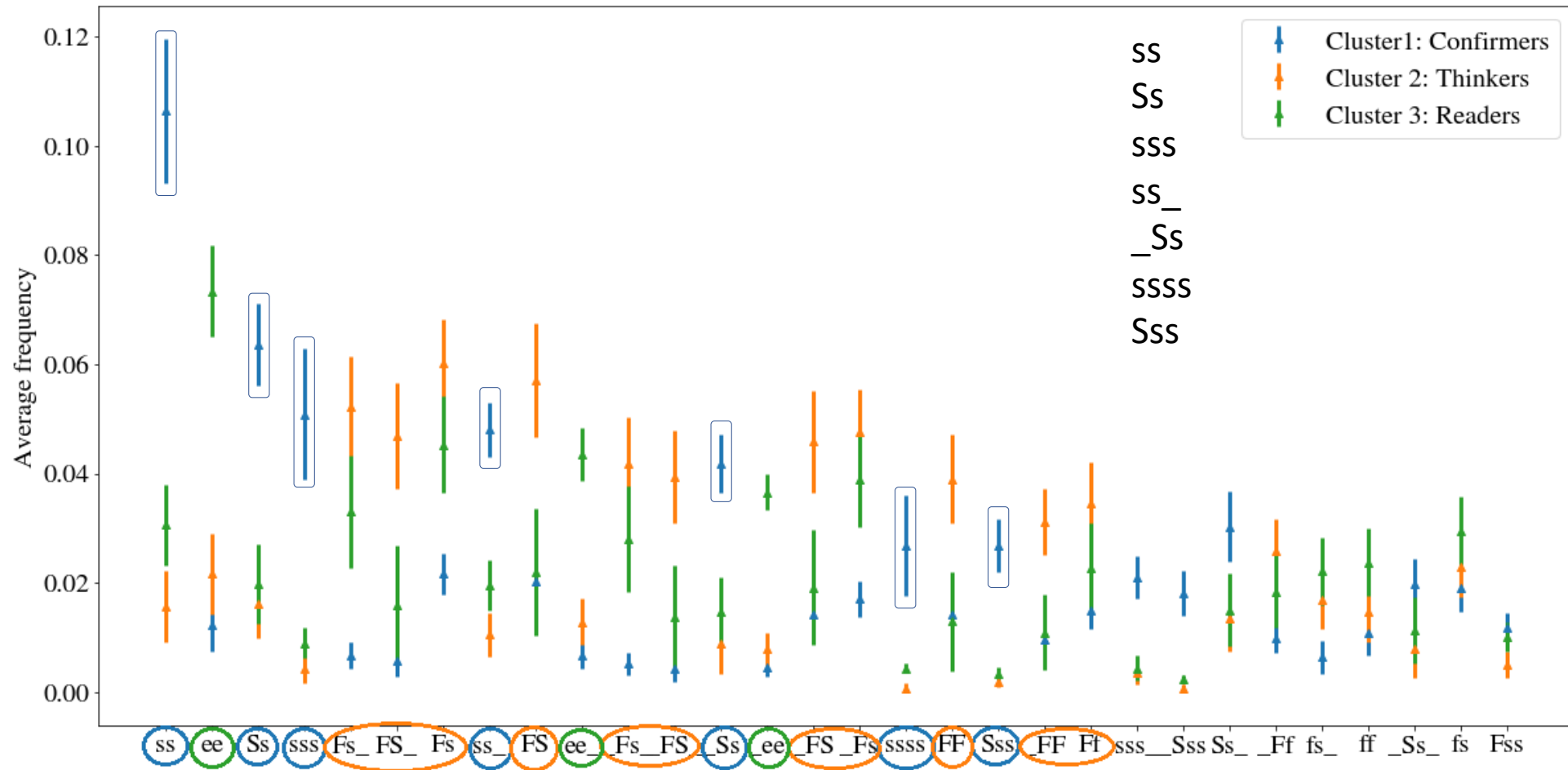


Pattern analysis



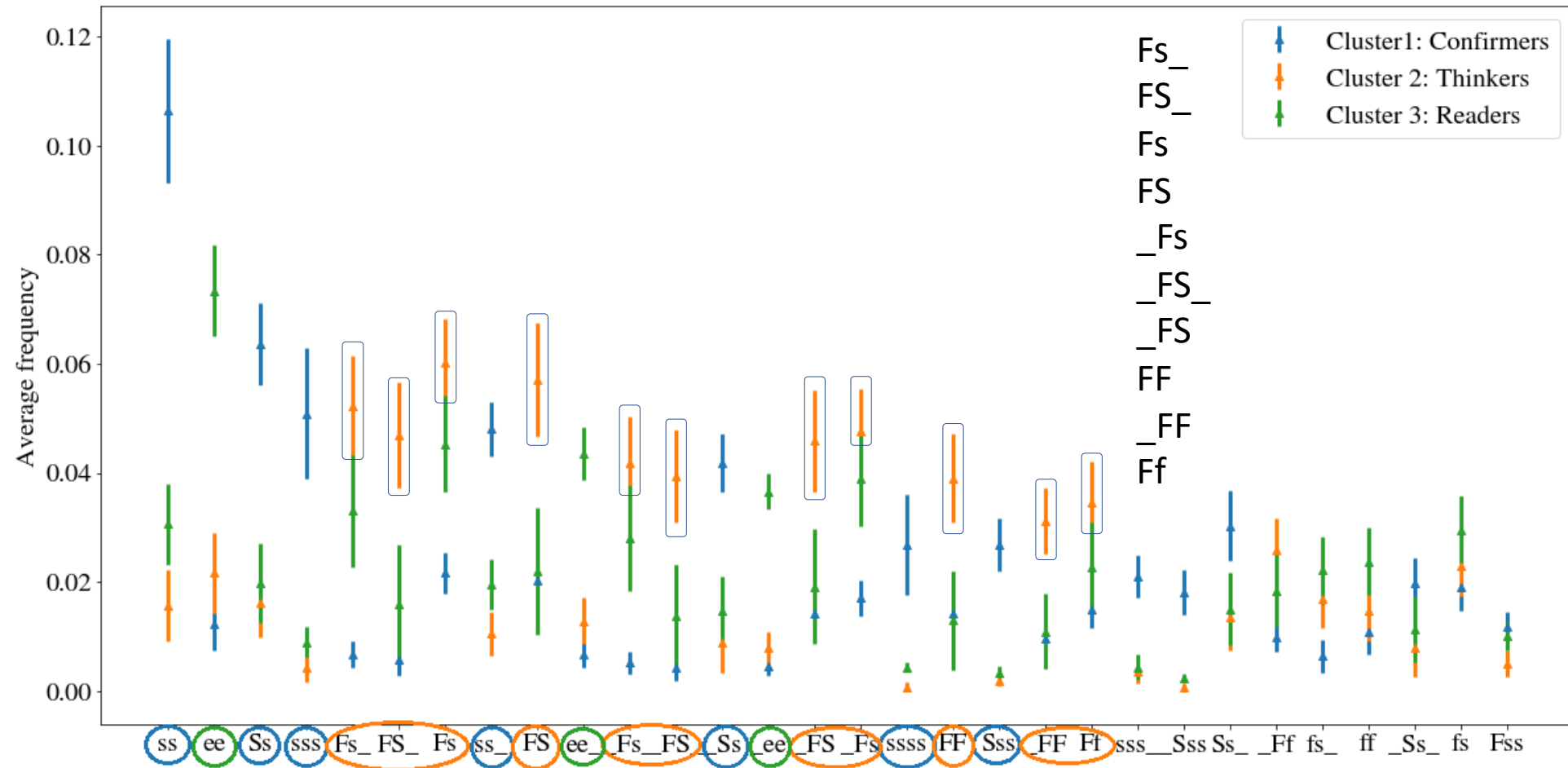
Top 30 patterns and their frequencies in 3 clusters

Repeat practicing even after success



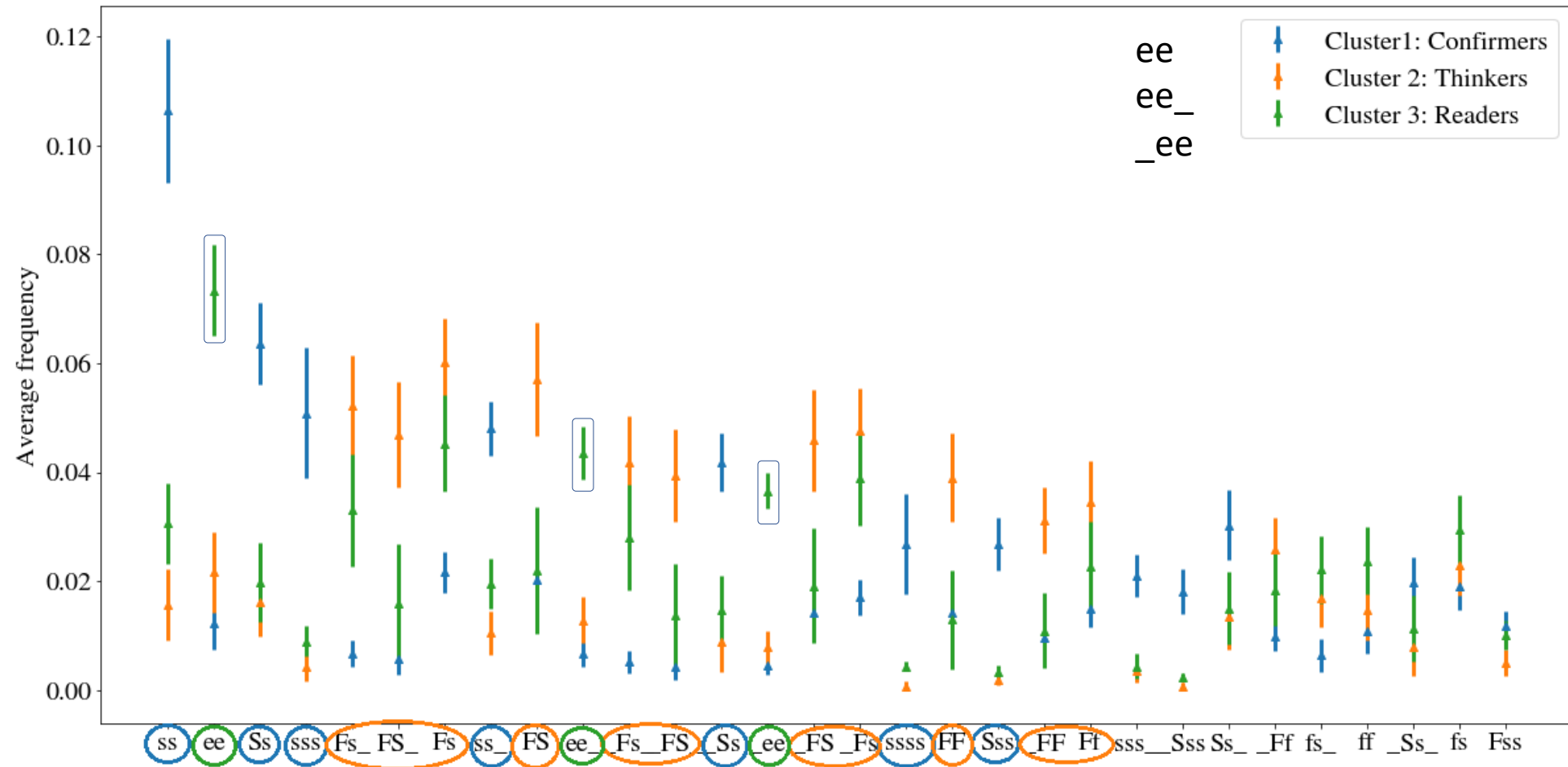
Top 30 patterns and their frequencies in 3 clusters

Spend more time on solving a problem



Top 30 patterns and their frequencies in 3 clusters

Spend more time on reading examples

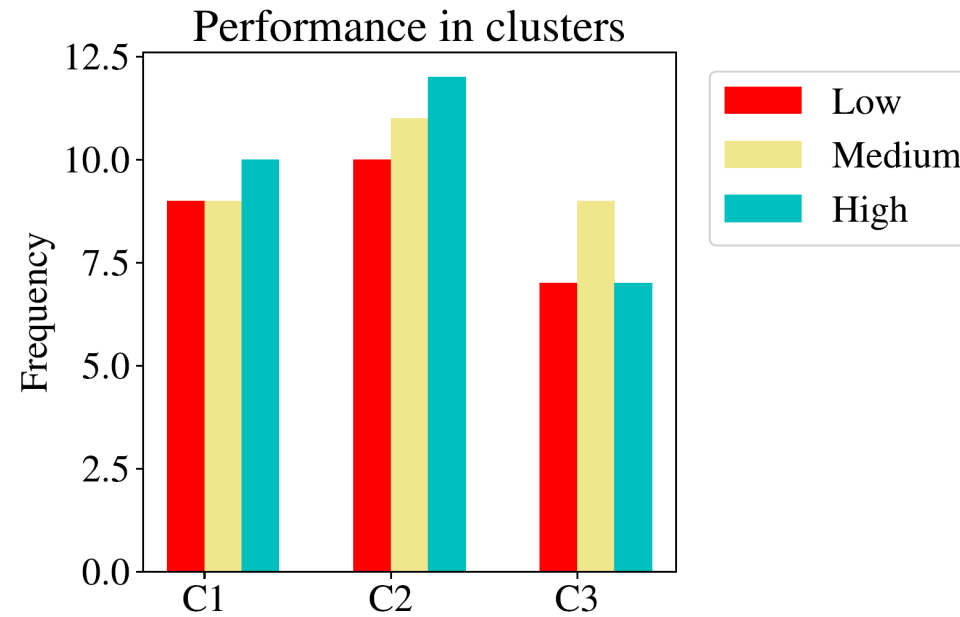


Top 30 patterns and their frequencies in 3 clusters

Conclusion

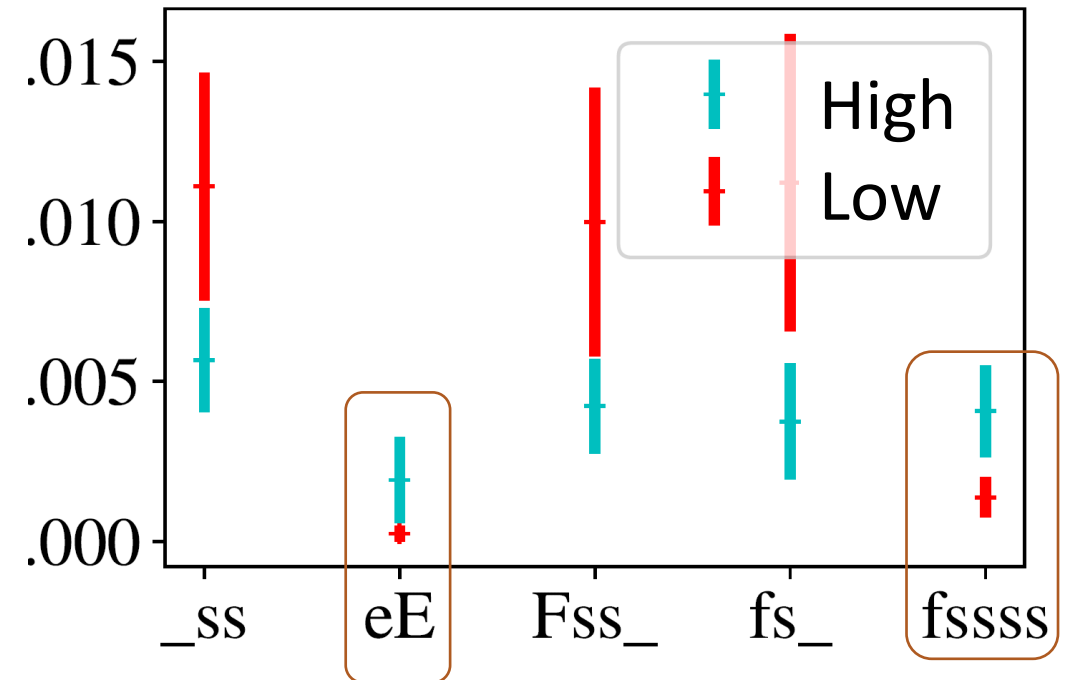
- Most frequent patterns in cluster 1 : ss, Ss, sss
 - Students tend to repeat practicing even if they succeed.
 - Confirmers
- Most frequent patterns in cluster 2: Fs_,Fs, FS, FS_
 - Students tend to spend more time on solving a problem
 - Thinkers
- Most frequent patterns in cluster 3: ee, _ee
 - Students tend to spend more time on reading annotated examples
 - Readers

Performance analysis



Performance analysis in cluster 1 (Confirmers)

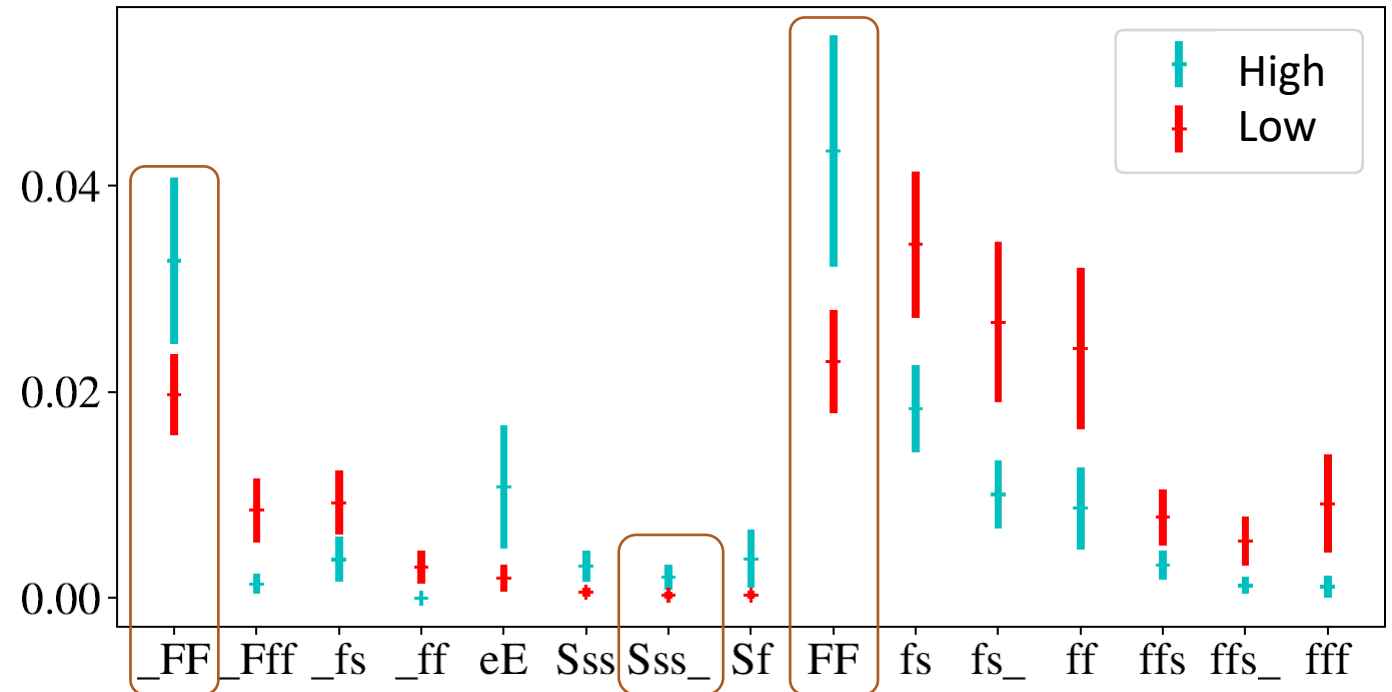
- Weak students:
 - Patterns: 'Fss_' and '_ss'
 - Repeat after an initial success
 - Short repetitions and quit after failure
- Strong students:
 - Patterns: 'fssss' and 'eE'
 - More repetition after an initial failure
 - Repeat reading examples



Patterns with significant difference for low and high performance (learning gain) students in Cluster 1

Performance analysis in cluster 2 (Thinkers)

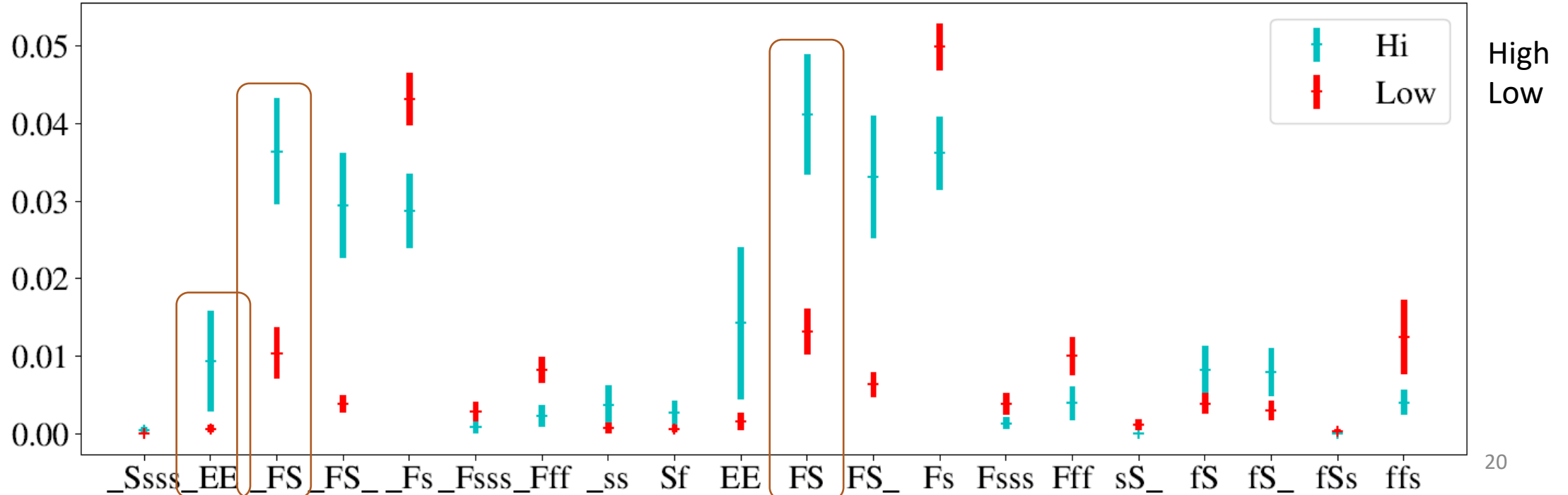
- Weak students:
 - Patterns: 'fff', 'ff', 'ffs_', 'fs'
 - Frequently try to guess and fail in solving problems
- Strong students:
 - Patterns: '_FF', 'FF', 'Sss'
 - Try a problem, until it is sufficiently understood



Patterns with significant difference for low and high performance (learning gain) students in Cluster 2

Performance analysis in cluster 3 (Readers)

- Weak Students:
 - Patterns: 'ffs', '_Fs' and 'Fs' - Do not spend enough time on their attempts
- Strong students:
 - Patterns: 'EE', '_FS', and 'FS' - Work with examples more carefully – No rush after failure



Sample Recommendations

Confirmers

Repetitions after failure in problem solving

Thinkers

Continue to think deeper for each problem

Readers

Working more carefully with examples and spending more time to think



CM-SPAM

- **CM-SPAM** is a sequential pattern mining algorithm based on the SPAM algorithm.
- CM-SPAM utilizes a new technique named co-occurrence pruning to prune the search space
- The **support of a sequential pattern** is the number of sequences where the pattern occurs divided by the total number of sequences in the database.
- A **frequent sequential pattern** is a sequential pattern having a support no less than the *minsup* parameter provided by the user.