Difficulty and Complexity of Introductory Programming Problems

Tomáš Effenberger **Jaroslav Čechák** Radek Pelánek

Adaptive Learning
Research group, Masaryk university Brno

CSEDM 2019

Difficulty vs. Complexity

Difficulty relates to the observed performance.

Complexity is an intrinsic characteristic of problems.

Why bother?

- problem sequencing
- cold start problem of difficulty measures
- fine tuning problem difficulty
- detecting anomalies among problems

Our Goal

Explore relationships between:

- complexity measures
- difficulty measures
- complexity and difficulty measures

Our Data

Exercise	Interface	Problems	Students	Attempts
RoboMission	blocks	85	3,800	62,500
Turtle Blockly	blocks	77	11,000	63,600
Turtle Python	text	51	2,400	11,900
Python	text	73	2,000	10,700

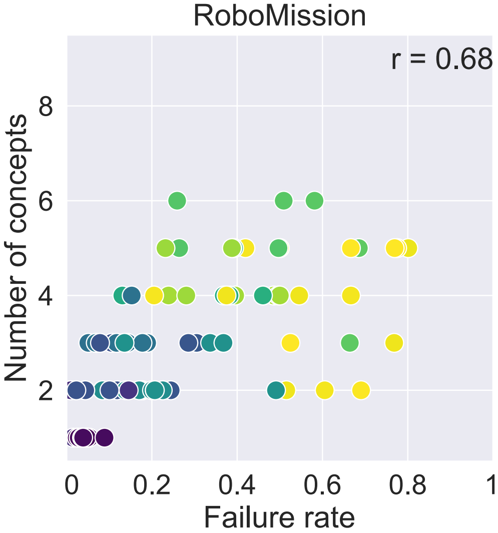
Used Measures

- complexity
    - instruction length
    - code length
    - number of unique concepts
- difficulty
    - failure rate
    - median solving time
    - median number of attempts

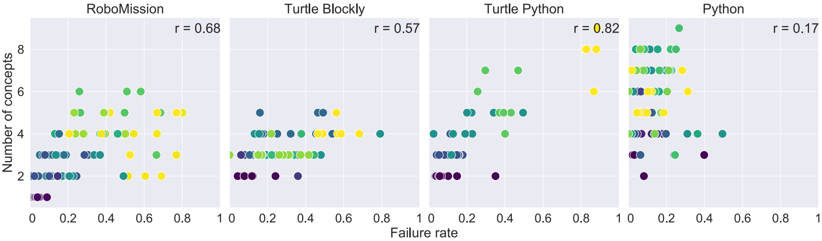
Methods

- Spearman's correlation
  - comparing pairs of measures
  - more indepth
- Principal Component Analysis (PCA)
  - broader overview
  - multiple measures at the same time

Spearman’s Correlation Example



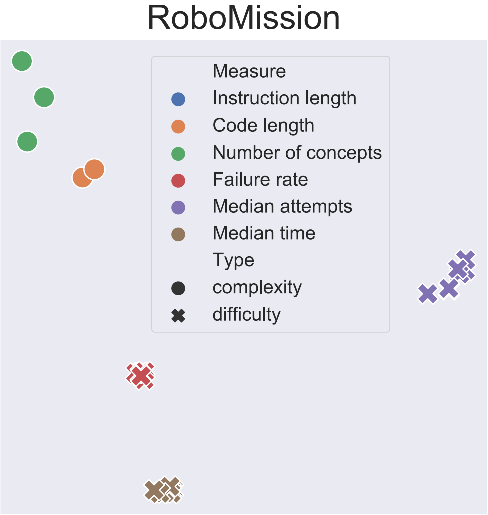
Spearman's Correlation Examples



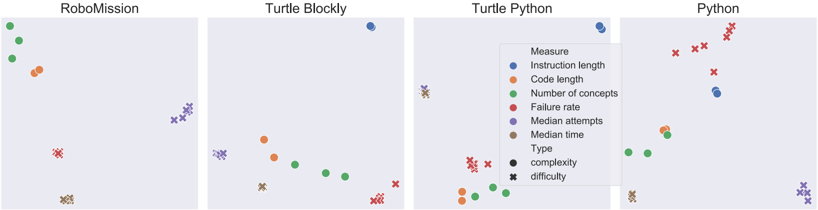
Spearman's Correlation Conclusion

- conclusions drawn from a single exercise may not generalize
- variants of the same measure correlate well ($r \geq 0.7$)

PCA Example


RoboMission
Measure
Instruction length
Code length
Number of concepts
Failure rate
Median attempts
Median time
Type
complexity
difficulty

PCA Results


RoboMission
Turtle Blockly
Turtle Python
Python
Measure
Instruction length
Code length
Number of concepts
Failure rate
Median attempts
Median time
Type
complexity
difficulty

PCA Conclusion

- variants of the same measure form tight clusters
- the same type of measures tend to be closer together

Complexity for Difficulty Estimation

- using lines of code and number of concepts
- estimate solve time, number of attempts, and failure rate

Conclusion

- conclusions drawn from a single exercise may not generalize
- filtering resulted in highly similar measures
- predicting difficulty using complexity is difficult