

DOMAIN SPECIFICITY IN SUPERVISED MACHINE LEARNING ANALOGIES

Author: Mateo Nasse (M.A.Nasse@student.tudelft.nl)

Supervisors: Ilinca Rențea & Yuri Noviello

Responsible Professor: Gosia Migut

We need to learn how to teach Machine Learning



01INTRODUCTION

Despite artificial intelligence’s and machine learning’s widespread presence, the inner workings of these technologies are a mystery even to most users. Even to computer science students, learning in this area can be quite a challenge. The use of analogies to explain concepts proves to be useful [1] in general but can it be applied to machine learning? Literature on teaching machine learning through these methods has a significant gap [2]. Furthermore, is it possible to target a specific domain a person is familiar with to make analogies more understandable? The video game domain is an interest possibility given computer science students’ tendency to understand this area [3].

02RESEARCH QUESTION

How does domain specificity influence analogy-based explanations in supervised machine learning for first-year bachelor computer science students?

- How effective are analogy-based explanations using domain-specific concepts from **computer gameplay** in helping understand supervised machine learning for first-year bachelor computer science students?
- What insights do expert evaluations show about the strengths and weaknesses of domain-specificity in analogy-based explanations for teaching supervised machine learning?



03METHODODOLOGY

To answer the research question, analogies based on supervised ML concepts were created and experts helped pick two which were later used to test on the target group. Both chosen concepts had one analogy in the general domain and one in the gaming domain. A survey was created to test users’ knowledge gain and their motivation when subjected to both the definition of a concept and its analogy. Users were divided into two groups through A / B testing to compare results.

Analogy Design and Generation

Survey Implementation

Evaluation & Analysis

04RESULTS

A total of 15 experts of different ML knowledge levels participated in the evaluation. Not every expert reviewed every analogy. The responses were evenly distributed, with each analogy being reviewed by 3-4 experts.

Through A / B testing, two groups delivered a total of 24 responses. Group A, who focused on general domain analogies provided 14 answers and Group B, who focused on specific domain, answered 10 times.

Group A 58.3%

Group B 41.7%

In terms of knowledge gain:

- Group A’s accuracy when answering was **44.05%**.
- Group B’s accuracy when answering was **43.33%**.

Interestingly, the accuracy varied on different types of questions.

In terms of motivation, the groups were measured across four factors:

Motivation Factors by Group

Likert Scale (Not true - Very True)

Dimensions

Accuracy Distribution by Group

Accuracy (%)

Group

05CONCLUSIONS

Overall, it can be concluded that domain specificity has no significant effect on a learner’s understanding when subjected to supervised ML concepts.

Experts’s evaluations show that even though some analogies are considered well constructed, given the subjectivity of interpreting them, the overall agreement is not exceptionally high.

The user evaluation revealed no statistically significant difference between both group regarding knowledge gain or motivation.

These results imply that for analogies to be effective, there is no requirement for degree of specificity, as long as the domain is known to the reader. Even though, this study comes with limitations (small sample size, uneven group distribution), it opens the door to explore questions as the effect of specificity on different types of questions.

REFERENCES: 1. Vishnu S. Pendyala. Relating machine learning to the real-world: Analogies to enhance learning comprehension. In Kanubhai K. Patel, Gayatri Doctor, Atul Patel, and Pawan Lingras, editors, Soft Computing and its Engineering Applications, pages 127-139, Cham, 2022. Springer International Publishing. 2. Ignacio Evangelista, Germ'an Blesio, and Emanuel Benatti. Why are we not teaching machine learning at high school? a proposal. In 2018 World Engineering Education Forum - Global Engineering Deans Council (WEEF-GEDC), pages 1–6, 2018. 3. Zohal Shah, Chen Chen, Gerhard Sonnent, and Philip M. Sadler. The influences of computer gameplay and social media use on computer science identity and computer science career interests. Telematics and Informatics Reports, 9:100040, 2023.