How to Teach Unsupervised Machine Learning with Analogies

A Study on the Effectiveness of Analogies in Teaching Unsupervised Machine Learning

Vincent Ruijgrok   *CSE3000 Research Project*   *Delft University of Technology*

Supervisors:
G. Migut, I. Rentea, Y. Noviello

TUDelft

① Introduction
- The importance and use of ML is growing across fields [1]
- There is a growing need for ML teaching methods [2]
- Unsupervised learning is a relevant field, and some concepts struggle with interpretability [3]
- Analogies often provide effective teaching, since they are good for making mental model [4]

Research Question
How does the use of analogies in teaching unsupervised learning affect the knowledge gain measured by the answering of theoretical questions?

② Background
- An analogy is "a comparison between things that have similar features" [5]
- They have a mapping between the target concept (what is explained) and the source concept (what the target is compared to. A mapping consists of a set of relations between features of the target and source [6]
- There is quite some research on analogies in education, but not on ML education. Only one research exists, which lists analogies without evaluating them [7]

③ Analogy Generation
- Lecture notes from the ML course in the Computer Science bachelor at Delft University of Technology were used for the concept list
- LLMs were used to generate analogies. They are good for that purpose, if their output is verified by experts.
- Results**:** 10 analogies for 10 concepts in unsupervised ML with a full mapping.

Concept list
1. Unsupervised learning
2. Clustering
3. Intra-cluster cohesion
4. Inter-cluster cohesion
5. K-means clustering
6. Single linkage
7. Complete
8. Average linkage
9. Agglomerative dendrogram
10. Divisive dendrogram

Example Analogy
Complete Linkage: If two cities want to ensure their farthest apart homes are still within reach, complete linkage measures the distance between the two farthest houses before connecting the cities.
- Clusters = Cities
- Data points = Houses
- Linkage distance = Max distance between two houses across cities
- Result = clusters only merge when elements are close

④ Expert Evaluation
- 15 experts (students who passed ML course, ML course TA's, ML lecturer) evaluated the analogies on target-concept coverage, mapping strength and metaphoricity
- Results: See Figure 1 and 2. Analogies 1-3 were rated best with high means and some agreement among experts. These continue to the student evaluation
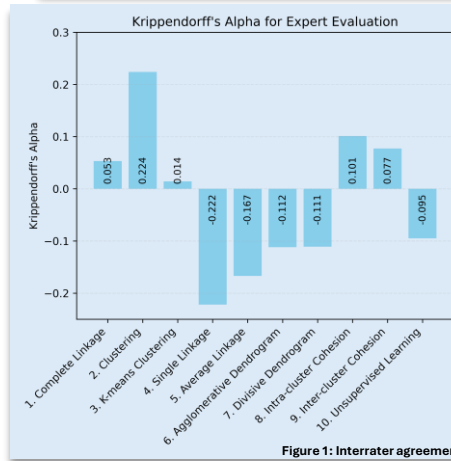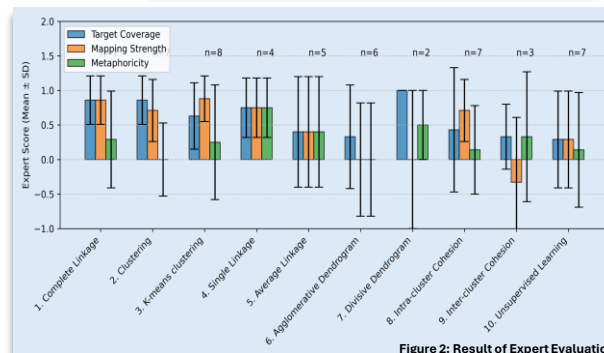

Figure 1: Interrater agreement

Figure 2: Result of Expert Evaluation

⑤ Student Survey
- A/B test to evaluate effect of analogies on knowledge gain of first-year CSE bachelor students. Group A got explanation with analogies. Group B got generic explanations.
- Both groups did a pre-test and post-test with questions on the level of 'Understanding' on Bloom's taxonomy. Students also did non-cognitive evaluation: the Reduced Instructional Materials Motivation Survey.
- Results: See Figure 3 and 4. There was a higher knowledge gain for 2 of 3 analogies and post-test scores for Group A were higher/similar than B. Students with analogies were more motivated and confident. No results were statistically significant.
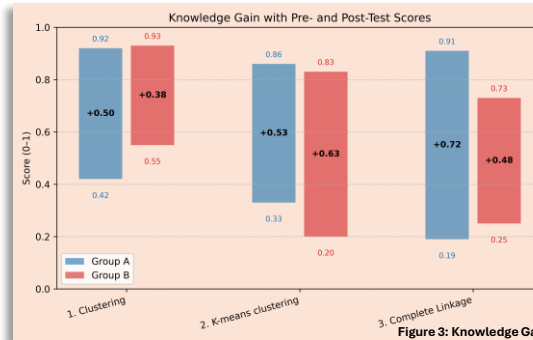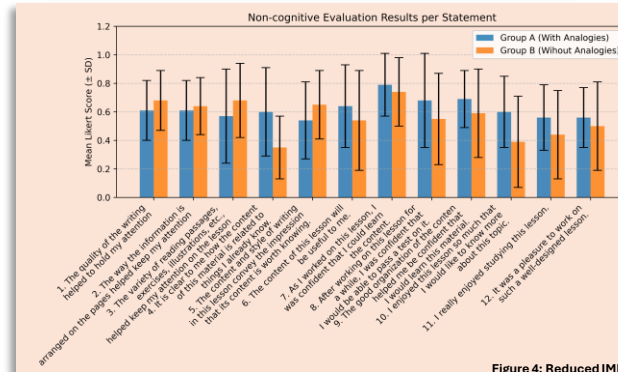

Figure 3: Knowledge Gain

Figure 4: Reduced IMMS

⑥ Takeaways
- 3 generated analogies had a positive effect on students' learning
- There is a great potential for more analogy-use in unsupervised ML education.
- All analogies and evaluation results can be found on ml-teaching-analogies.github.io

⑦ Future work
- Develop more analogies for these concepts or for different concepts
- Evaluate analogies further to conclusively prove their effectiveness
- Investigate the long-term effect or effect on different levels on Bloom's taxonomy

References
[1] Shapiro & Fiebrink (2019). Launching Agenda for Learning ML, ACM TOCE 19(4).
[2] Hazzan & Mike (2023). Pedagogical Challenge of ML Education, Springer.
[3] Almuqati et al. (2024). Challenges in Super- & Unsupervised Learning, IJASEIT 14(4).
[4] Dagher (1995). Instructional Analogies in Science Ed, Sci Educ 79(3).
[5] Analogy – Def.: "a comparison" (Cambridge Dictionary).
[6] Bhavya et al. (2024). Long-Form Analogy Evaluation, INLG Gen Challenges.
[7] Pendyala (2022). ML Real-World Analogies, Springer.