Peloponnese OptLearn Algorithm

Spiros Sirmakessis

1 Introduction

Active learning, emphasized in education, engages students in their learning process. Research shows it's more effective than traditional methods, fostering deeper understanding and participation. Students analyze, debate, research, and produce, individually or in groups, transforming from passive listeners to active contributors. This chance encourages educators to adopt innovative teaching tools and methodologies for interactive and appealing learning environments [1, 5, 7].

In this context, the MathE platform emerges. MathE (mathe.pixel-online.org) is an online educational platform designed to support students struggling with college-level mathematics and those seeking to enhance their understanding of various mathematical concepts. It offers a range of resources such as videos, exercises, practice tests, and instructional materials covering diverse areas of higher education mathematics, accessible globally for free. Considering the preliminary results of the MathE platform [3], efforts to improve it have aligned with its mission to provide a transformative learning environment adaptable to the user's needs.

This technical report describes the OptLearn algorithm proposed by the University of Peloponnese. The algorithm aims to recommend the questions for the students, considering their preliminary results on the platform.

2 Methodology

In this study, we employed the k-means clustering technique to categorize students' performance based on the responses they submitted through the platform system. Clustering serves as an unsupervised technique for segmenting data, with its primary objective being the organization of dataset elements into distinct groups (clusters) by assessing their similarities and differences focusing on discovering underlying patterns in an unsupervised manner [9].

Partitioning clustering methods aim to obtain partitions on the set to be classified. In general, the partition method applied to a set of X elements starts from a subset of k points, considered the centers of the aggregation classes. The method is iterated, recalculating the centers at each stage, and the elements of the set to be classified are reallocated according to their dissimilarity to the centers. Usually, the stopping criterion is defined as having no modifications after two successive iterations. In practice, the algorithm is typically run multiple times with different initial states, and the best configuration obtained from all stages is used as the final cluster. These types of methods are usually based on a central point, denoted as cluster centroid [10, 6].

The k-means algorithm is a popular method used for clustering data into groups. Its simplicity, efficiency, and effectiveness make it one of the most widely used clustering algorithms in several fields such as machine learning, data mining, and pattern recognition [4]. It consists of separating samples into groups of equal variance, minimizing a criterion known as the inertia or within-cluster sum-of-squares (WSS) [2]. As k-means is not an automatic clustering algorithm, it requires the definition of the initial parameter k, which represents the number of clusters division.

Thereby, the k-means algorithm aims to choose centroids that minimize the inertia, or within-cluster sum-of-squares criterion, presented in Equation 1 [2].

$$WSS = \min \sum_{i=1}^{k} \sum_{x \in C_i} ||X - \mu_i||^2, \tag{1}$$

From these centers, clustering is defined, grouping data points according to the center to which each point is assigned.

3 Proposed Algorithm

When a lecturer submits a question in the MathE platform, it is required that they inform the difficulty level of this question. As default, the platform comprises 5 difficult level, denoted as $score_q$, considering the Likert scale [8]. The questions with the lowest difficulty level are assigned a score of 1, while those with higher difficulty levels receive a score of 5. The remaining questions are distributed among the intermediate levels according to the difficulty level assigned by the lecturer.

Once the lecturer's opinion is obtained through the platform system, it is necessary to obtain the student's performance level, which is based on their student's historical performance $H_{student}$. For this, the k-means algorithm categorizes the students in 5 clusters cq according to their historical answers in the platform, considering the answers in the topic. This categorization defines the level of the initial question for each test for each student. Thus, if a given student is categorized in cluster 3 regarding his performance, the initial question of its test will belong to difficult level 3. The next question selection will be guided by the answer to the previous question, considering the following logic: If the answer is correct, the next question will be selected from cluster $cq + 1$. But if the answer is incorrect, the next question will be chosen from cluster $cq - 1$. Finally, if the student marks the option "I don't know," the level is kept, and another question about the current level is selected.

Algorithm 1 presents the proposed Algorithm developed by the University of Peloponnese, named the Peloponnese OptLearn Algorithm. Thus, consider a given question q, that belongs to the set $Q = \{1, ..., nq\}$, which is associated with a score that defines the difficult level $score_q \in \{1, ..., Level_{max}\}$.

Algorithm 1 Peloponnese OptLearn Algorithm

Consider the input parameters: $Q = \{1, ..., nq\}$, $score_q \in \{1, ..., Level_{max}\}$.
$H_{student} \longleftarrow$ Student success rate in the platform.
$k \longleftarrow$ number of cluster partitioning;
Consider $score = \{score_q, ..., score_{nq}\}$ as input parameters;
$k \longleftarrow$ number of cluster partitioning;
Obtain the clustering $k - means(H_{student}, k)$ associating each question to a given cluster cs.
For a given student, do:
while Number of questions selected for the current test is not reach **do**
    Verify the student cluster, cs.
    Select randomly an initial question belonging to cluster cq.
    if q has a corrected answer **then**
        the next question is randomly selected from $cq + 1$ cluster.
        if q has a incorrect answer **then**
            the next question is randomly selected from $cq - 1$ cluster.
        elseThe answer is "I don't know"
            select another question of cluster cq
        end if
    end if
end while

References

[1] Alhawiti, N.M.: The influence of active learning on the development of learner capabilities in the college of applied medical sciences: Mixed-methods study. Advances in medical education and practice **14**, 87–99 (2023). https://doi.org/doi.org/10.2147/AMEP.S392875

[2] Arthur, D., Vassilvitskii, S.: K-means++: The advantages of careful seeding. In: Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms. p. 1027–1035. SODA '07, Society for Industrial and Applied Mathematics, USA (2007)

[3] Azevedo, B.F., Pereira, A.I., Fernandes, F.P., Pacheco, M.F.: Mathematics learning and assessment using mathe platform: A case study. Education and Information Technologies (2021). https://doi.org/10.1007/s10639-021-10669-y

[4] Azevedo, B.F., Rocha, A.M.A.C., Pereira, A.I.: Hybrid approaches to optimization and machine learning methods: a systematic literature review. Journal of Machine Learning (2024)

[5] Deslauriers, L., McCarty, L.S., Miller, K., Callaghan, K., Kestin, G.: Measuring actual learning versus feeling of learning in response to being actively engaged in the classroom. Proceedings of the National Academy of Sciences **116**(39), 19251–19257 (2019). https://doi.org/10.1073/pnas.1821936116

[6] Jain, A.: Data clustering: 50 years beyond k-means. Pattern Recognition Letters **31**(8), 651–666 (2010)

[7] Keiler, L.S.: Teachers' roles and identities in student-centered classrooms. International Journal of STEM Education **5**(34) (2018)

[8] Norman, G.: Likert Scales, Levels of Measurement and the "Laws" of Statistics. Advances in Health Sciences Education **15**(5), 625–632 (2010). https://doi.org/10.1007/s10459-010-9222-y

[9] Rehman, A.U., Belhaouari, S.B.: Divide well to merge better: A novel clustering algorithm. Pattern Recognition **122**, 108305 (2022)

[10] Steinley, D.: K-means clustering: a half-century synthesis. British Journal of Mathematical and Statistical Psychology **59**(1), 1–34 (2006)