Towards fair, explainable and actionable clustering for learning analytics

Tai Le Quy
Leibniz University Hannover
Hannover, Germany
tai@l3s.de

Eirini Ntoutsi
Freie Universität Berlin
Berlin, Germany
eirini.ntoutsi@fu-berlin.de

ABSTRACT
Clustering is an important technique in learning analytics for partitioning students into groups of similar instances. Application examples include group assignments, students-class allocation, etc. However, traditional clustering does not ensure a fair-representation in terms of some protected attributes like gender or race, and as a result, the resulting clusters might be biased. Moreover, traditional clustering might result in clusters of varying cardinalities reducing their actionability for end user. In many applications, like group assignment, the capacity of the resulting clusters should be controllable to allow direct applicability of the resulting clusters. Furthermore, it is important to be able to explain why an instance/student is clustered into a specific cluster and/or which attributes play a crucial role in the clustering process. We believe that the aforementioned aspects of fairness, capacity and explainability are important for the successful application of clustering in the learning analytics domain.

Keywords
learning analytics, clustering, fairness, bias, explainability, capacity, actionability

1. INTRODUCTION
In education, machine learning (ML) has been used in a wide variety of decision-making tasks, for example, student dropout prediction [11], education admission decisions [25] or forecasting on-time graduation of students [16]. Recently, the incidents of discrimination in ML-based decision-making systems in education, such as grades prediction [4, 15], are an important reason for the increase of the attention to bias and fairness in ML of researchers [32]. Accordingly, the decisions made by the ML-based systems against groups or individuals on the basis of *protected attributes* like gender, race, etc. Bias in education has been studied in many aspects from different sources of bias in education [27], students' data analysis [3], racial bias [39] and gender bias [26]. However, ML-based decision-making systems have the potential to amplify prevalent biases or create new ones and therefore, fairness-aware ML approaches are required also for the learning environments.

In our research, we are focusing on the *fairness* of clustering methods in learning analytics since clustering is an effective method to analyze student data [8, 17, 28, 36]. Clustering algorithms are useful tools for partitioning students into groups of similar instances [3, 31]. Results from clustering methods are applicable in educational activities such as group assignments [10] and student team achievement divisions [37]. However, the traditional clustering algorithms do not take into account the fairness w.r.t. protected attributes like gender or race, as a consequence of focusing only on the similarity objective. Moreover, the cardinality of the resulting clusters is typically not part of the objective function and as a result clusters of very different cardinalities might be extracted reducing the usefulness of the results. Moreover, understanding the instances-to-clusters assignments, the important features for clustering and what characterizes each cluster (the so-called, cluster labels) is not always easy [33].

The aim of this research is to study the fairness, capacity and explainability requirements and challenges in the learning analytics domain and propose effective solutions that can be used by the domain experts. In this direction, we propose the concept of *fair-capacitated clustering* which extends traditional clustering focusing on clustering quality to also ensure fairness of representation in terms of some protected attribute(s) and the applicability of the resulting clusters by ensuring balanced cluster cardinalities. Such clusters can be exploited by different stakeholders in the learning environment: educators can better organize the learning activities, e.g., group assignments; students can learn better in a more inclusive and equitable environment.

In another direction, we plan to extend the *fair capacitated clustering with explainability* to give insights to the end users about how certain assignment decisions are made, what features are important for clustering and what the extracted clusters represent. Such information will allow educators to customize teaching activities to each group and improve the learning trajectory of each student, each group and the class in overall.

We believe that the results of our research are useful in other domains as well, for example business (clustering customers in marketing studies, salesmen areas distribution), traffic (vehicle routing) and communication (network design). Moreover, our research contributes to the further development of the domain of fairness and responsible AI with new methods (for the unsupervised learning problem) and application domain (learning analytics).

The rest of our paper is structured as follows: Section 2 overviews the related work. Research questions are presented in Section 3. Section 4 describes our ongoing work on fair-capacitated clustering and preliminary results. Finally, conclusions and outlook are presented in Section 5.

2. RELATED WORK

Chierichetti et al. [7] first introduced the fair clustering problem and presented a balance measure for computing fairness in the resulting clusters. They defined "fairlet" as a small cluster preserving fairness measure, and then they apply k-Center clustering algorithm on these fairlets to obtain the final clusters. In the later studies, Backurs et al. [1] described an algorithm for the fairlets computation in nearly linear time. The problem of fair clustering with multiple protected attributes is investigated in the researches of Rösner and Schmidt [34] and Bera et al. [2].

The capacitated clustering problem (CCP) was first introduced by Mulvey and Beck [30] with heuristic and subgradient algorithms. Later, researchers proposed approaches to solve the problem in the different clustering methods. For instance, Khuller and Sussmann [19] introduced an approximation algorithm for the capacitated k-Center problem. An improved version of k-Means algorithm for CCP was presented by Geetha et al. [12] with the use of a priority measure to assign points to their centroid. Lam and Mittenthal [20] proposed a heuristic hierarchical clustering method for CCP.

Quite a few researchers, recently, are interested in the usefulness of explainable and interpretable clustering models. Chen et al. [6] proposed a probabilistic discriminative model with the ability to learn rectangular decision rules for each cluster. Saisubramanian et al. [35] offered a voting method to consider which features are meaningful for the end user. Moshkovitz et al. [29] used an unsupervised decision tree to explain k-Means and k-Medians methods.

3. RESEARCH QUESTIONS

We organize the challenges into the research questions $Q1-Q3$ explained hereafter:

Q_1: **What is fairness in learning analytics and how to mitigate discrimination in clustering?** Fairness in education is an interesting topic researchers [5, 9, 13]. We investigate the fairness terminology in student analytics w.r.t protected attributes such as gender, race. Student performance can be considered as the protected attribute because in some cases no knowledge of the student's performance can help to prevent bias in the grading procedure [23, 24]. Related work in the fairness-aware ML area depicts a large variety of approaches that can be categorized into: i) pre-processing approaches that intervene at the input data [22]; ii) in-processing approaches that directly tweak the clustering algorithm to account for fairness [7] and iii) post-processing approaches that adjust the clustering results to ensure fairness [38]. We will mainly follow the in-processing approaches that directly incorporate fairness in the clustering process. However, such approaches depend on the clustering algorithm per se; our current work focuses on hierarchical and partitioning algorithms, in the future density-based clustering will be also investigated.

Q_2: **How to satisfy multiple objectives, namely capacity of clusters and fairness of representation on top of the (standard) cluster similarity objective?** As already mentioned, the actionability of the results is important. As a concrete example consider group assignments: groups should be comparable to allow for a fair allocation of work among students. In the capacitated clustering problem [30], they do not consider fairness, nor explainability. Likewise approaches for fair clustering also exist [7]. However, approaches that jointly consider the different objectives do not exist.

Q_3: **What is the explanation of a (fair-capacitated) clustering model and how to find it?** The importance of explainable clustering results for the end users has been already discussed. Explainability does not only allow for understanding how certain decisions are made but also allows for debugging of algorithmic decisions and corrections in case of decisions based on protected attributes like gender or race. There are different aspects to explainability in clustering: understanding how a certain assignment of an instance to a cluster was made, understanding what attributes contributed to clustering and explaining what each cluster is about (or cluster labeling). We will investigate the different aspects to allow educators to better understand the groups that are formed and to allow both educators and single users/students to understand how they fit into a particular cluster.

4. PRELIMINARY RESULTS ON FAIR CAPACITATED CLUSTERING

In this section, we present the preliminary results of our work namely *fair-capacitated clustering* [21] problem. The goal is to cluster students into fair-groups w.r.t. single protected attribute. *Gender*, typically, is chosen as the protected attribute. In other words, we would like to balance the number of males and females in the resulting clusters and our proposed methods should satisfy the size of group constraint in order to make the results more actionable.

We define the problem of (t, k, q)-fair-capacitated clustering as finding a clustering $\mathcal{C} = \{C_1, \cdots C_k\}$ that partitions the data X into k clusters such that the cardinality of each cluster $C_i \in \mathcal{C}$ does not exceed a threshold q, i.e., $|C_i| \leq q$ (*the capacity constraint*), the balance of each cluster is at least t, i.e., $balance(\mathcal{C}) \geq t$ (*the fairness constraint*), and minimizes *the objective function*. Parameters k, t, q are user-defined referring to the number of clusters, minimum balance threshold and maximum cluster capacity, respectively.

We present a two-step solution to the problem: i) we rely on fairlets [7] to generate minimal sets that satisfy the fair constraint and ii) we propose two approaches, namely hierarchical clustering (denoted by *hierarchical fair-capacitated*) and partitioning-based clustering (denoted by *k-Medoids fair-capacitated*, to obtain the fair-capacitated clustering. The hierarchical approach embeds the additional cardinality requirements during the merging step while the partitioning-based one alters the assignment step using a knapsack problem formulation to satisfy the additional requirements.

We experiment our proposed methods on four educational datasets: UCI Student performance[1], PISA test scores[2], OULAD[3], MOOC[4], containing the demographics, grades and school-related attributes of students. Table 1 in Appendix A summarizes the characteristics of datasets.

We report on clustering quality (measured as clustering cost, see Eq. 1), cluster fairness (expressed as cluster balance [7], see Eq. 2 and Eq. 3) and cluster capacity (expressed as cluster cardinality). The parameters are set as follows: the minimum threshold of balance $t = 0.5$, i.e., the proportion of the minority group is at least 50% in the resulting clusters; the maximum capacity of clusters $q = \left\lceil \frac{|X| * \epsilon}{k} \right\rceil$; ε is set to 1.01 and 1.2, for k-Medoids fair-capacitated and hierarchical fair-capacitated methods, respectively.

$$\mathcal{L}(X, \mathcal{C}) = \sum_{s_i \in S} \sum_{x \in C_i} d(x, s_i) \tag{1}$$

$$balance(C_i) = \min\left(\frac{|\{x \in C_i | \psi(x) = 0\}|}{|\{x \in C_i | \psi(x) = 1\}|}, \frac{|\{x \in C_i | \psi(x) = 1\}|}{|\{x \in C_i | \psi(x) = 0\}|} \right) \tag{2}$$

$$balance(\mathcal{C}) = \min_{C_i \in C} balance(C_i) \tag{3}$$

The baseline includes well-known clustering methods with fairness-aware approaches and a traditional algorithm. **1) k-Medoids**[18]. This is a traditional partitioning technique of clustering that uses the actual instances as centers (medoids) and divides the dataset into k clusters and minimizes the clustering cost. **2) Vanilla fairlet** [7]. A vanilla fairlet decomposition that ensures fair clusters is generated, then, a k-Center clustering algorithm [14] is applied to cluster those fairlets into k clusters. **3) MCF fairlet** [7]. It is an updated version of the *Vanilla fairlet* with The fairlet decomposition is transformed into a *minimum cost flow* (MCF) problem, by which an optimized version of fairlet decomposition in terms of cost value is computed.

The preliminary results show that our approaches deliver well-balanced clusters in terms of both fairness and cardinality while maintaining a good clustering quality. In terms of clustering cost (Figure 1-a) (Appendix B), our approaches outperform the vanilla fairlet and MCF fairlet methods although they are worse compared to the vanilla k-Medoids clustering. This is obvious due to the fact that our methods have to satisfy constraints on fairness or/and cardinality. *MCF fairlet hierarchical fair-capacitated* shows the best performance due to the optimization in the merging step. As illustrated in Figure 1-b regarding to fairness, our methods are comparative to the competitors. In which, the minimum threshold of balance t is visualized as a dashed line and the actual balance from the dataset is plotted as a dotted line. In Figure 1-c, the maximum capacity thresholds q are presented by the dashed and dotted lines. Our approaches are more preeminent with a lower dispersion, in terms of cardinality. The boxplots of our methods are drawn thicker because the variation of the capacity of resulting clusters is tiny in quite a few cases. MCF fairlet shows the worst performance, followed by Vanilla fairlet and vanilla k-Medoids algorithm.

[1]https://archive.ics.uci.edu/ml/datasets/Student+Performance
[2]https://www.kaggle.com/econdata/pisa-test-scores
[3]https://analyse.kmi.open.ac.uk/open_dataset
[4]https://github.com/kanika-narang/MOOC_Data_Analysis

5. CONCLUSION AND OUTLOOK

The investigations of the fairness, capacity and explainability requirements in the learning analytics domain are the main goals of our research. In this paper, we present the challenges of our work with 3 research questions. The preliminary results on the fair-capacitated clustering problem show that our approaches can satisfy multiple objectives namely fairness, capacity and clustering cost. In the next step, we want to deploy the implementation of an explainable fair clustering algorithm to achieve the clarification of the assignment in a fair clustering method.

Acknowledgements

The work of the first author is supported by the Ministry of Science and Education of Lower Saxony, Germany, within the PhD program "LernMINT: Data-assisted teaching in the MINT subjects", for which the second author is a principal investigator.

6. REFERENCES

[1] A. Backurs, P. Indyk, K. Onak, B. Schieber, A. Vakilian, and T. Wagner. Scalable fair clustering. In *International Conference on Machine Learning*, pages 405–413. PMLR, 2019.

[2] S. Bera, D. Chakrabarty, N. Flores, and M. Negahbani. Fair algorithms for clustering. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, editors, *NIPS*, volume 32. Curran Associates, Inc., 2019.

[3] S. Bharara, S. Sabitha, and A. Bansal. Application of learning analytics using clustering data mining for students' disposition analysis. *Education and Information Technologies*, 23(2):957–984, 2018.

[4] K. Bhopal and M. Myers. The impact of covid-19 on a level students in england. *SocArXiv*, 2020.

[5] S. Bøyum. Fairness in education–a normative analysis of oecd policy documents. *Journal of Education Policy*, 29(6):856–870, 2014.

[6] J. Chen, Y. Chang, B. Hobbs, P. Castaldi, M. Cho, E. Silverman, and J. Dy. Interpretable clustering via discriminative rectangle mixture model. In *2016 IEEE 16th International Conference on Data Mining (ICDM)*, pages 823–828. IEEE, 2016.

[7] F. Chierichetti, R. Kumar, S. Lattanzi, and S. Vassilvitskii. Fair clustering through fairlets. In *NIPS*, pages 5036–5044, 2017.

[8] A. M. De Morais, J. M. Araujo, and E. B. Costa. Monitoring student performance using data clustering and predictive modelling. In *2014 IEEE frontiers in education conference (FIE) proceedings*, pages 1–8. IEEE, 2014.

[9] N. J. Dorans and L. L. Cook. *Fairness in educational assessment and measurement*. Routledge, 2016.

[10] M. Ford and J. Morice. How fair are group assignments? a survey of students and faculty and a modest proposal. *Journal of Information Technology Education: Research*, 2(1):367–378, 2003.

[11] J. Gardner, C. Brooks, and R. Baker. Evaluating the fairness of predictive student models through slicing analysis. In *LAK'19*, pages 225–234, 2019.

[12] S. Geetha, G. Poonthalir, and P. Vanathi. Improved k-means algorithm for capacitated clustering problem. *INFOCOMP*, 8(4):52–59, 2009.

[13] C. Gipps and G. Stobart. Fairness in assessment. In *Educational assessment in the 21st century*, pages 105–118. Springer, 2009.

[14] T. F. Gonzalez. Clustering to minimize the maximum intercluster distance. *Theoretical computer science*, 38:293–306, 1985.

[15] S. Hubble and P. Bolton. A level results in england and the impact on university admissions in 2020-21. *House of Commons Library*, 2020.

[16] S. Hutt, M. Gardner, A. L. Duckworth, and S. K. D'Mello. Evaluating fairness and generalizability in models predicting on-time graduation from college applications. *Proceedings of The 12th International Conference on Educational Data Mining (EDM 2019)*, 2019.

[17] Y. Jayabal and C. Ramanathan. Clustering students based on student's performance-a partial least squares path modeling (pls-pm) study. In *International Workshop on Machine Learning and Data Mining in Pattern Recognition*, pages 393–407. Springer, 2014.

[18] L. Kaufman and P. J. Rousseeuw. Partitioning around medoids (program pam). *Finding groups in data: an introduction to cluster analysis*, 344:68–125, 1990.

[19] S. Khuller and Y. J. Sussmann. The capacitated k-center problem. *SIAM Journal on Discrete Mathematics*, 13(3):403–418, 2000.

[20] M. Lam and J. Mittenthal. Capacitated hierarchical clustering heuristic for multi depot location-routing problems. *Int. J. Logist. Res. Appl.*, 16(5):433–444, 2013.

[21] T. Le Quy, A. Roy, G. Friege, and E. Ntoutsi. Fair-capacitated clustering. *The 14th International Conference on Educational Data Mining (EDM 2021)*, 2021.

[22] B. T. Luong, S. Ruggieri, and F. Turini. k-nn as an implementation of situation testing for discrimination discovery and prevention. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 502–510, 2011.

[23] J. M. Malouff, A. J. Emmerton, and N. S. Schutte. The risk of a halo bias as a reason to keep students anonymous during grading. *Teaching of Psychology*, 40(3):233–237, 2013.

[24] J. M. Malouff, S. J. Stein, L. N. Bothma, K. Coulter, and A. J. Emmerton. Preventing halo bias in grading the work of university students. *Cogent Psychology*, 1(1):988937, 2014.

[25] F. Marcinkowski, K. Kieslich, C. Starke, and M. Lünich. Implications of ai (un-) fairness in higher education admissions: the effects of perceived ai (un-) fairness on exit, voice and organizational reputation. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, pages 122–130, 2020.

[26] T. Masterson. An empirical analysis of gender bias in education spending in paraguay. *World Development*, 40(3):583–593, 2012.

[27] M. Meaney and T. Fikes. Early-adopter iteration bias and research-praxis bias in the learning analytics ecosystem. In *Companion Proceeding of the 9th International Conference on Learning Analytics & Knowledge (LAK'19), Fairness and Equity in Learning Analytics Systems Workshop*, pages 14–20, 2019.

[28] A. Merceron and K. Yacef. Clustering students to help evaluate learning. In *IFIP World Computer Congress, TC 3*, pages 31–42. Springer, 2004.

[29] M. Moshkovitz, S. Dasgupta, C. Rashtchian, and N. Frost. Explainable k-means and k-medians clustering. In *International Conference on Machine Learning*, pages 7055–7065. PMLR, 2020.

[30] J. M. Mulvey and M. P. Beck. Solving capacitated clustering problems. *European Journal of Operational Research*, 18(3):339–348, 1984.

[31] Á. A. M. Navarro and P. M. Ger. Comparison of clustering algorithms for learning analytics with educational datasets. *IJIMAI*, 5(2):9–16, 2018.

[32] E. Ntoutsi, P. Fafalios, U. Gadiraju, V. Iosifidis, W. Nejdl, M.-E. Vidal, S. Ruggieri, F. Turini, S. Papadopoulos, E. Krasanakis, et al. Bias in data-driven artificial intelligence systems-an introductory survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 10(3):e1356, 2020.

[33] E. Ntoutsi, M. Spiliopoulou, and Y. Theodoridis. Tracing cluster transitions for different cluster types. *Control & Cybernetics*, 38(1), 2009.

[34] C. Rösner and M. Schmidt. Privacy preserving clustering with constraints. *arXiv preprint arXiv:1802.02497*, 2018.

[35] S. Saisubramanian, S. Galhotra, and S. Zilberstein. Balancing the tradeoff between clustering value and interpretability. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, pages 351–357, 2020.

[36] M. Tanai, J. Kim, and J. H. Chang. Model-based clustering analysis of student data. In *International Conference on Hybrid Information Technology*, pages 669–676. Springer, 2011.

[37] M. Tiantong and S. Teemuangsai. Student team achievement divisions (stad) technique through the moodle to enhance learning achievement. *International Education Studies*, 6(4):85–92, 2013.

[38] C. Vrain, I. Davidson, et al. Constrained clustering via post-processing. In *International Conference on Discovery Science*, pages 53–67. Springer, 2020.

[39] N. Warikoo, S. Sinclair, J. Fei, and D. Jacoby-Senghor. Examining racial bias in education: A new approach. *Educational Researcher*, 45(9):508–514, 2016.

APPENDIX
A. DATASET

Table 1: An overview of the datasets

Dataset	#instances	#attributes	Protected attribute	Balance score
UCI student performance-Mathematics	395	33	Gender (F: 208, M: 187)	0.899
UCI student performance-Portuguese	649	33	Gender (F: 383; M: 266)	0.695
PISA test scores	3,404	24	Male (1: 1,697; 0: 1,707)	0.994
OULAD	4,000	12	Gender (F: 2,000; M: 2,000)	1
MOOC	4,000	21	Gender (F: 2,000; M: 2,000)	1

B. UCI STUDENT PERFORMANCE DATASET

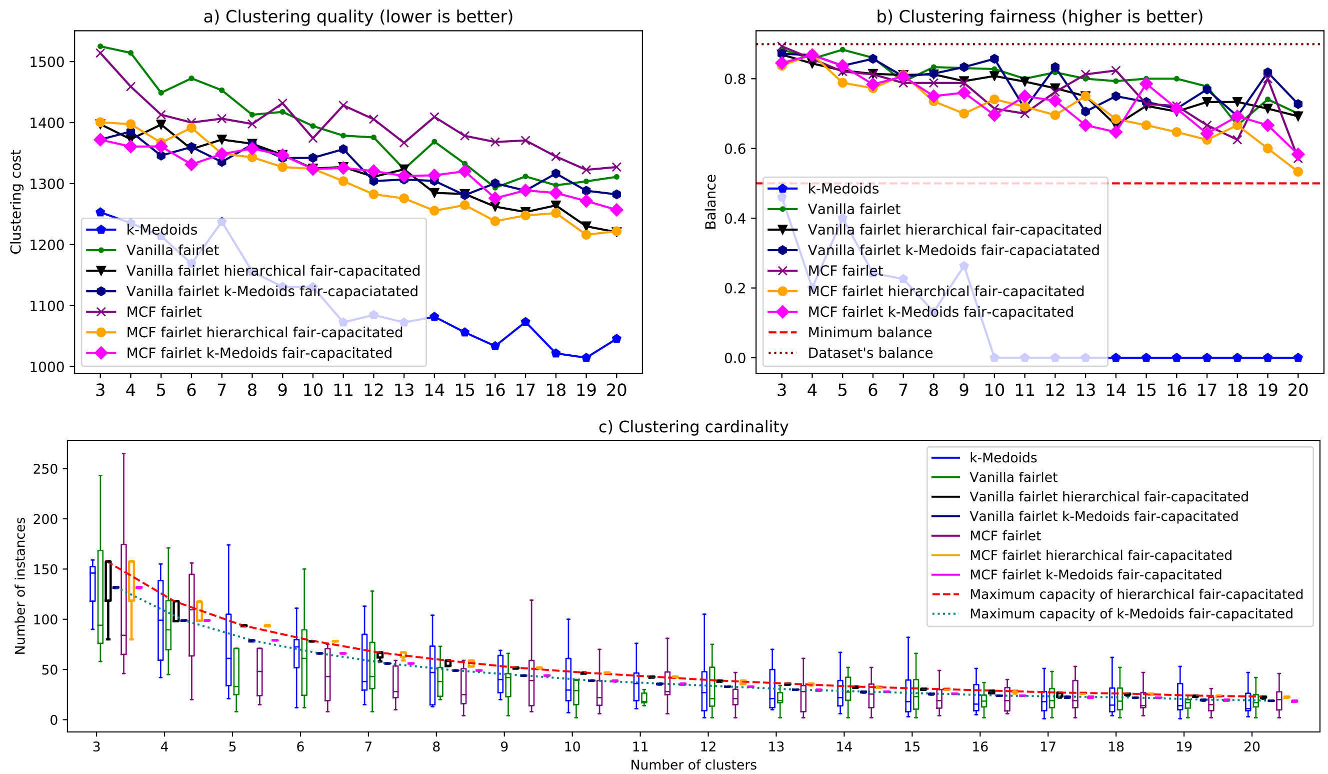

Figure 1: Performance of different methods on UCI student performance dataset - Mathematics subject