The Journal of Engineering

Research Article

eISSN 2051-3305
Received on 6th June 2019
Revised 29th June 2019
Accepted on 9th July 2019
E-First on 19th September 2019
doi: 10.1049/joe.2019.0938
www.ietdl.org

Educational data mining for students' performance based on fuzzy C-means clustering

Yu Li[1], Jin Gou[1] ✉, Zongwen Fan[2]
[1]College of Computer Science and Technology, Huaqiao University, Xiamen 361021, People's Republic of China
[2]School of Electrical Engineering and Computing, The University of Newcastle, Callaghan, NSW 2308, Australia
✉ E-mail: goujin@hqu.edu.cn

Abstract: Education greatly aids in the process of students' growth; therefore, education institutions try to provide high-quality education to their students. A possible remedy to provide high-quality education is by discovering knowledge from educational data. However, accurately evaluating students' performance is very challenging due to different sources and structures of educational data. In addition, different teaching strategies are required because students' learning ability are different. One way to discover the hidden knowledge from educational data is the use of clustering algorithms, which are capable of mining interesting patterns from educational data. Thus, this study presents a fuzzy C-means clustering algorithm using 2D and 3D clustering to evaluate students' performance based on their examination results (the examination grades from College of Computer Science and Technology, Huaqiao University for students enrolled in year 2014). Based on the experimental results from 2D and 3D clustering for evaluating students' performance, the educators can better understand the students' performance so as to build a pedagogical basis for decisions. Students can also receive some recommendations from the mining results about their performance.

1 Introduction

Education is a broad concept that consists of the process of facilitating learning, or the acquisition of knowledge, skills, values, beliefs, and habits [1, 2], which is not only very important for a person's life but also for a country's progress [3]. The level of education quality was found to have great influence for the growth rates of a country's income and mortality decline [4]. Mirowsky [5] discovered that education is also closely related with a person's social status and health, which to some extent helps people become better citizens [6]. Due to the importance of education, education institutions try to provide high-quality education to their students by discovering knowledge from educational data [7] to help decision makers to make the right adjustments of education [8].

With the large volumes of educational data obtained from various resources (files, documents, videos etc.), data mining was introduced in [9] to discover useful information for educators to make better decisions due to its successful use in many applications [10–12], including market basket analysis [13], and education [14]. However, mining useful knowledge from educational data (like the data from students' enrolment and attendance records, and their examination results [15]) is a challenging task. One way to address this problem is to apply educational data mining [16] because it is capable of discovering knowledge and patterns from educational data to better understand the students' performance [17]. To evaluate the students' performance, decision trees were applied in [18] by studying the key factors that could influence the students' performance in courses. Based on the students' records, Pandey and Pal [19] used Bayesian classification to evaluate the students' performance, which could be helpful for the institution to reduce the drop-out rate and improve the performance level of institutions.

With the extracted information like how many students will pass the final examination in a particular course, if a student can pass the particular examination, and which students will need assistance in order to graduate, the decision makers in colleges or universities can make use of educational resources to help students. In addition, the students would like to know what courses are best for them to achieve their goals based on the prediction of how well they will perform in the courses selected [8]. Thus, more and more researches have focused on knowledge discovery from educational data. For example, Jacob, Kotak, and Puthran [20] found out that with the use of data mining approaches, the stakeholders in the educational system can discover useful knowledge to help study, predict, and improve students' performance.

One of the most important steps for educational data mining is to visualise and analyse the data. A possible way for this is the use of clustering algorithms, which has been widely and successfully applied in data mining [21] by extracting useful inherent relationships in practical applications [22–24]. The main purpose of clustering algorithms is to discover interesting patterns from a given dataset based on sample similarity [14, 25]. Oyelade, Oladipupo, and Obagbuwa [26] applied K-means to predict students' academic performance; their results have shown K-means is useful to monitor students' performance in higher institution. K-means was also utilised as pre-processing in [27–30] to evaluate students' performance. Their results proved that K-means can provide helpful information for instructors to make better decisions. A variant of K-means, X-means, was used to analyse students' performance in e-learning training [31]. However, the above clustering algorithms are all hard-clustering algorithm [32, 33]. In some cases, each data sample may belong to two or more clusters [34] (e.g. a student's performance can be both 'good' and 'great'), but with different membership values [35] Unlike hard-clustering algorithms, fuzzy C-means clustering (FCM) extends hard clustering to soft condition [36, 37], where each data sample can belong to more than one cluster.

However, none of the above research focus on both visualisation and clustering results for students' performance based on students' examination grades using FCM clustering algorithm. While with the visualisation of the clustering results, the students' performance can be evaluated and analysed [38]. In this case, the educators can make better decisions to provide quality education to students, which in return improves the students' performance [39]. Thus, in this paper, we applied FCM to evaluate students' performance with 2D and 3D clustering. The dataset was downloaded from the student achievement management system of Huaqiao University. The undergraduate students enrolled in year 2014 from College of Computer Science and Technology were used for students' performance evaluation. Pre-processing of the

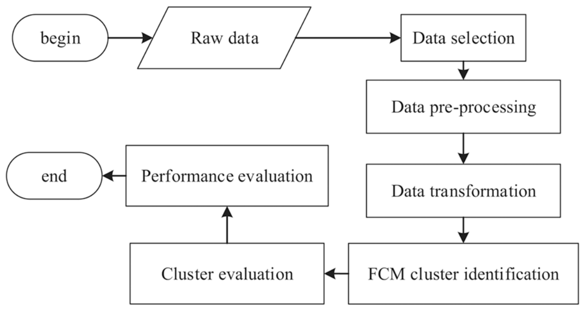

Fig. 1 *Flowchart for the evaluation of students' performance*

dataset was conducted before the use of FCM clustering algorithm. We analysed the students' performance for each class in the college based on 2D and 3D clustering. The experimental results have shown some interesting patterns, which give some useful indications for educators to evaluate student's performance.

The remainder of this paper is organised as follows. A brief introduction of FCM clustering algorithm is first given in Section 2. The steps for evaluating students' performance are presented in Section 3. In Section 4, the 2D and 3D clustering results for classes Digital Media, Network Engineering, Computer Science and Technology, Software Engineering class 1, Software Engineering class 2, and the whole college students are described. Then, the analyses for evaluating students' performance are discussed. Finally, conclusions are provided in Section 5.

2 Fuzzy C-means clustering (FCM) for the evaluation of students' performance

FCM, proposed by Dunn [40], is a soft-clustering approach that assigns a data sample to different clusters with different membership values [41]. Based on the similarity measures, the samples with high similarity are clustered into the same group while samples with low similarity are grouped into different clusters. In this case, FCM is capable of clustering data into C clusters, but with a certain membership belongs to the cluster centre [42], which is very helpful to find the inherent relationships of data samples in practical applications [43, 44]. The details of FCM are given as follows:

Step 1: Initialise the number of clusters C and the membership matrix $U^{(0)}$, respectively.
Step 2: For tth iteration, clustering centres are calculated by the weighted mean of all data samples.

$$c_i^t = \sum_{j=1}^{n} w_{\mathrm{ij}}^m x_j \tag{1}$$

$$w_{\mathrm{ij}}^m = \frac{\mu_{\mathrm{ij}}^m}{\sum_{k=1}^{n} \mu_{\mathrm{ik}}^m} \tag{2}$$

where n is the number of data samples, i is the index of cluster centre, j and k are indexes of data samples, $\mu_{\mathrm{ij}}^m \in [0, 1]$ is the membership value, and m is a weighted index.
Step 3: Update the membership matrix $U^{(k)}$. For each membership value μ_{ij}^m in $U^{(k)}$, it is calculated as (3)

$$\mu_{\mathrm{ij}}^m = \frac{1}{\sum_{k=1}^{C}\left(\frac{\| x_j - c_i \|}{\| x_j - c_k \|}\right)^{2/(m-1)}} \tag{3}$$

Step 4: Repeat from Step 2 to Step 4 until the terminal criteria is met. That is if the absolute difference between U^{t+1} and U^t is smaller than the pre-defined threshold ε, then stops the algorithm; else $t = t + 1$ and continues the iterations.

$$\| U^{t+1} - U^t \| < \varepsilon \tag{4}$$

Table 1 Ten samples of students' grade from class Network Engineering

Student number	College physics	CPMCS	CCLL
1425131001	81	77	83
1425131002	86	88	85
1425131003	82	79	77
1425131004	77	78	91
1425131005	84	91	71
1425131006	82	81	84
1425131007	80	76	77
1425131008	79	83	78
1425131009	88	86	84
1425131010	84	74	77

To better understand the procedure for the evaluation of students' performance based on FCM, the flowchart of evaluating students' performance using FCM is given in Fig. 1. As we can see from Fig. 1, the student dataset needs to be pre-processed (e.g. remove unnecessary features like student name and course name) by data selection, data pre-processing, and data transformation. When the data is prepared, FCM clustering algorithm is applied to obtain the clusters, where each cluster represents a performance level (including bad, average, good, and great). It is worth noting that the performance levels here are not exactly the same as the grade, but the performance among the students in the class or the college. Based on the discovered knowledge from clustering results, it is easier for the educators to build a pedagogical basis for decisions like which teaching approach is better for students. At the same time, students can get some recommendations from the mining results to further improve their performance.

3 Experimental results and analyses

In this section, the data description of the educational dataset is first given. Then, the experimental results for classes Digital Media, Network Engineering, Computer Science and Technology, Software Engineering class 1, Software Engineering class 2, and the whole college students are described. In each experiment, students' grades are clustered into four groups, namely bad, average, good, and great. Finally, the analyses of the clustering results are given.

3.1 Dataset description

The dataset used in this paper was derived from the student achievement management system of Huaqiao University. The undergraduate students enrolled in year 2014 from College of Computer Science and Technology were used for performance evaluation. There are five classes (Digital Media, Network Engineering, Computer Science and Technology, Software Engineering class 1, Software Engineering class 2 with 45, 34, 55, 48, and 64 students, respectively) for four majors (Digital Media, Network Engineering, Computer Science and Technology, Software Engineering). Thus, the total number of students for the college is 246. Three courses, including College Physics, Career Planning and Management for College Students (CPMCS), College Chinese Language and Literature (CCLL), are used to evaluate the students' performance. To better understand the dataset, some samples of students’ grades are given in Table 1 from class network engineering after pre-processing (e.g. removing missing values, courses' names, and students' names). While other classes have the same data format.

3.2 Experimental results and analyses

The clustering results for Digital Media, Network Engineering, Computer Science and Technology, Software Engineering class 1, Software Engineering class 2, all the college students are given in Figs. 2–7, respectively, for both 2D and 3D clustering. In addition, clustering centres for classes Digital Media, Network Engineering, Computer Science and Technology, Software Engineering class 1,

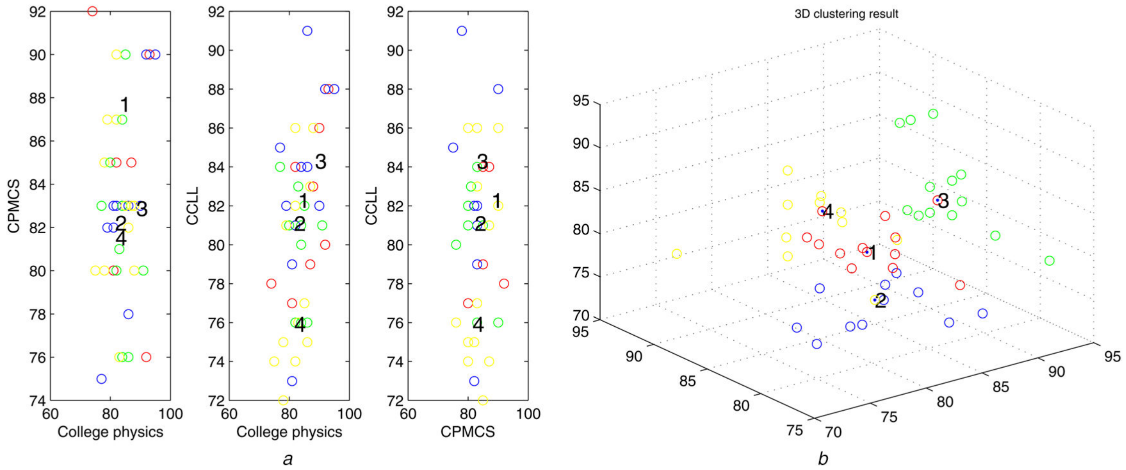

Fig. 2 *2D and 3D clustering for class Digital Media*
(a) 2D clustering, ***(b)*** 3D clustering

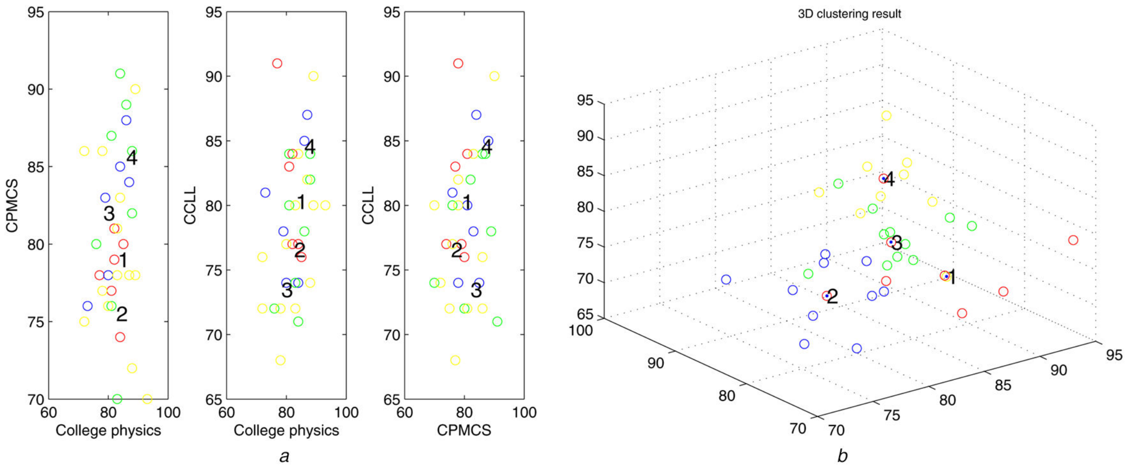

Fig. 3 *2D and 3D clustering for class Network Engineering*
(a) 2D clustering, ***(b)*** 3D clustering

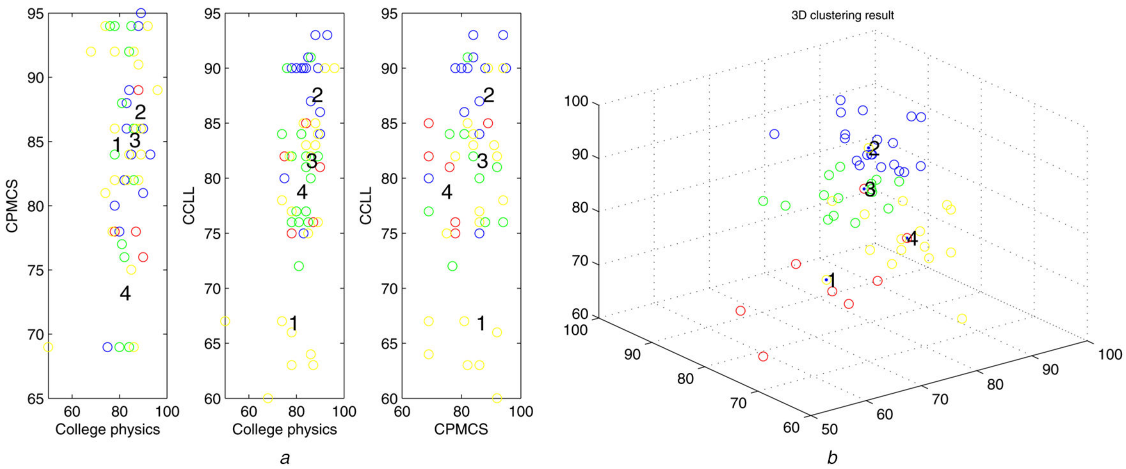

Fig. 4 *2D and 3D clustering for class Computer Science and Technology*
(a) 2D clustering, ***(b)*** 3D clustering

Software Engineering class 2, and all the college students are given in Table 2. The format of each clustering centre is represented as (College Physics, Career Planning and Management for College Students, College Chinese Language and Literature), and the best results based on average grades for each performance level are highlighted in bold.

As we can see from Fig. 2, students with performance level 'great' in class Digital Media are mainly in cluster 3 for 3D clustering even though some data samples have worse results than cluster 1 in (College physics, CPMCS) for 2D clustering. This is because the clustering is based on the overall performance by considering the grades of all three courses. Thus, a student with one course grade lower than the other student does not mean his/her overall performance is also lower. Fig. 3 shows that the students with great performance in class Network Engineering are all in cluster 4 for both 2D and 3D clustering. Thus, the students in cluster 4 have the best performance in the class. According to Figs. 4 and 5, similar to Fig. 3, the students with performance level 'great' in class Computer Science and Technology and Software Engineering class 1 are all focus on cluster 2 and cluster 4,

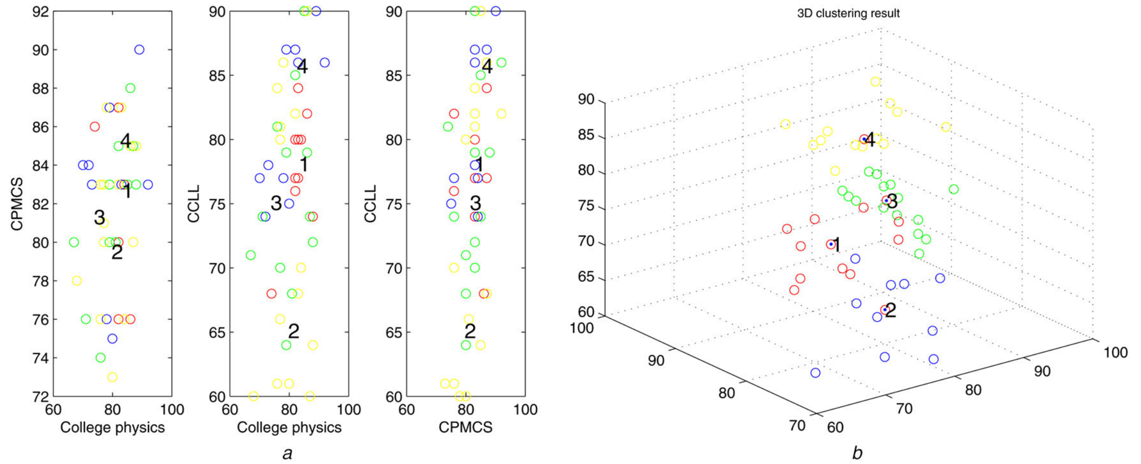

Fig. 5 *2D and 3D clustering for Software Engineering class 1*
(a) 2D clustering, ***(b)*** 3D clustering

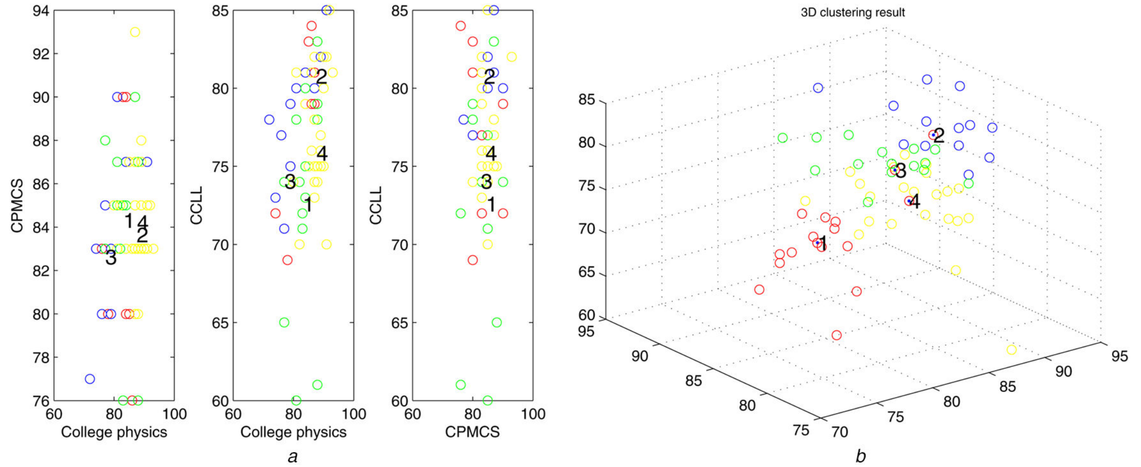

Fig. 6 *2D and 3D clustering for Software Engineering class 2*
(a) 2D clustering, ***(b)*** 3D clustering

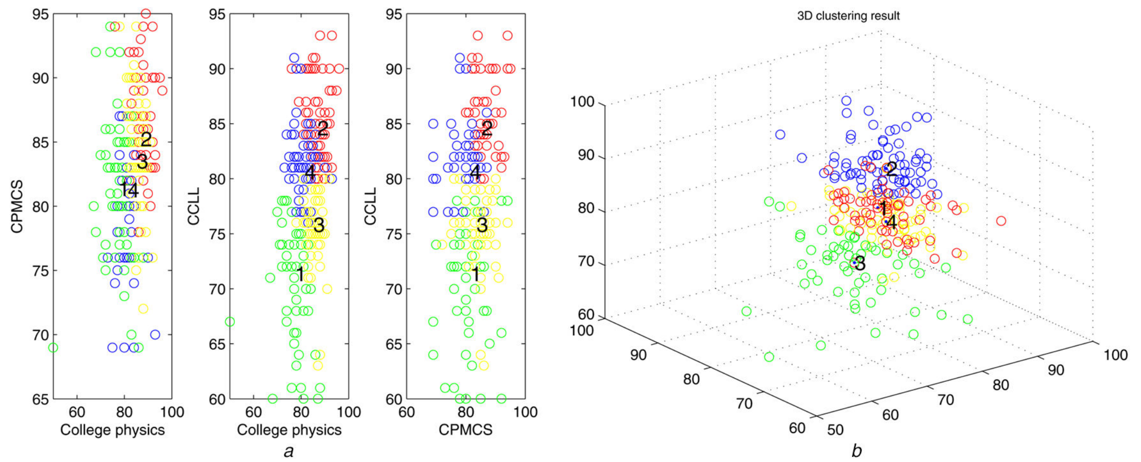

Fig. 7 *2D and 3D clustering for all students in the college*
(a) 2D clustering, ***(b)*** 3D clustering

respectively, for both 2D and 3D clustering. While in Fig. 6, although the students in cluster 2 have the best performance, some students are not as good as those in clusters 1 and 4 in (College physics, CPMCS) for 2D clustering. This is because there will be some overlaps between two clusters. For example, the worst performance for level 'great' may or may not better than the best performance for level 'good'. Finally, the grades for all students in college are shown in Fig. 7, and the students with the best performance are fall into cluster 2.

While for the performance level 'bad', students in cluster 4 in class Digital Media for 2D clustering have the worst performance, but for 3D clustering, some samples in cluster 2 have worse performance. The main reason is that the best performance in level 'bad' has some probabilities outperform the worst performance in level 'average' according to the overall evaluation. The same reason for classes Network Engineering, Computer Science and Technology, Software Engineering class 1, and Software Engineering class 2 in clusters 3, 1, 2, and 1, respectively. Educators can give some help for students who fall into these clusters to improve their overall performance.

In addition, as we can see from Figs. 2–7 in both 2D and 3D clustering, the students with great performance are more likely to

Table 2 Clustering centres for classes Digital Media, Network Engineering, Computer Science and Technology, Software Engineering class 1, Software Engineering class 2, and all the college students

Class name	Performance level			
	Bad	Average	Good	Great
Digital Media	(**81.672, 81.524, 75.915**)	(**81.608, 82.205, 81.102**)	(**82.884, 87.696, 82.279**)	(88.568, 82.878, 84.301)
Network Engineering	(79.175, 77.424, 74.536)	(83.391, 77.445, 79.839)	(81.410, 83.954, 74.588)	(85.856, 85.110, 84.695)
Computer Science and Technology	(76.521, 84.800, 66.866)	(80.025, 73.205, 78.786)	(84.072, 85.090, 81.646)	(**86.437, 87.356, 87.593**)
Software Engineering class 1	(79.645, 79.521, 65.114)	(73.661, 81.318, 75.093)	(83.093, 82.695, 78.114)	(82.540, 85.297, 85.740)
Software Engineering class 2	(77.009, 82.626, 74.050)	(83.027, 84.316, 72.597)	(87.614, 84.207, 75.842)	(87.475, 83.649, 80.777)
All students	(77.352, 81.429, 71.299)	(85.164, 83.603, 75.946)	(81.357, 81.092, 80.504)	(86.778, 85.263, 84.639)

The format of each clustering centre is represented as (College Physics, Career Planning and Management for College Students, College Chinese Language and Literature), and the best results based on average grades for each performance level are highlighted in bold.

have higher density on the high-grade part while the students with bad performance are more likely to be sparse or lower density in the low-grade part. This makes sense because if a student has great performance in each course, then this student has more potential to have the overall high performance. While for a student with low performance in all or most part of the courses, this student has high probability to have low overall performance.

Finally, according to Table 2, we can see that the students in class Digital Media have very stable performance because for four performance levels (bad, average, good, and great), they all have very similar grades in clusters; and the grades in performance levels bad, good, and great have similar grades, and they are all higher than the other classes. While for class Computer Science and Technology, students have higher great performance than class Digital Media but also have worse performance than the other classes in the performance levels bad, good, and great. In this case, there are both very high performance and low performance students in this class.

In summary, clustering technique is a useful tool to discover interesting knowledge and patterns from educational data by analysing the clusters and the clustering centres. In addition, the visualisation of data is also very important to explore the deeper relationship between data samples based on 2D or 3D clustering. The results have shown that the evaluation of students' performance is possible with proper strategies, algorithms, and analyses. It is worth noting that the performance levels in the experiments are based on the students' overall performance in the class or the college. Thus, it is not exactly the same as the grade (e.g. the grade is marked as 'bad' if the grade is <60 points (total score is 100), or 'great' if the grade is equal or >90). Based on the discovered knowledge from clustering results, the educators have better knowledge about students' performance to build a pedagogical basis for decisions like which teaching approach is better for students. Students can also receive some recommendations from the mining results about their performance.

4 Conclusion

To discover knowledge from educational data (students' enrolment and attendance records, examination results, study attitudes, and behaviours), data mining is a useful tool to extract some interesting patterns [45]. In this paper, we applied soft-margin clustering algorithm, FCM, to visualise and cluster students' examination grades to evaluate students' performance based on their examination records with 2D and 3D clustering. Three courses (College Physics, Career Planning and Management for College Students, and College Chinese Language and Literature) for students in four classes (Digital Media, Network Engineering, Computer Science and Technology, Software Engineering class 1, and Software Engineering class 2) and the whole college students are evaluated. The results from 2D and 3D clustering have shown that the students with great performance have more potential to have higher density on the high grade part while the students with bad performance have lower potential to the high-grade part (e.g. Bad (76.521, 84.800, 66.866) and Great (86.437, 87.356, 87.593) in Computer Science and Technology, respectively). By visualising the 2D and 3D clustering, even though some students may have low performance in one or two courses, the overall performance can still be good (with low probability). Thus, with the visualisation of the clustering results, it is easier for educator to evaluate and analyse students' performance, which makes it useful for educators to make better decisions so as to provide quality education for students.

However, we only consider students' grades in our clustering results while there are other types of educational data (e.g. students' enrolment and attendance records) we can use. Hence, for future work, more sources and structures of educational data can be utilised for clustering. In addition, regression and classification algorithms like support vector machines [46] and artificial neural networks [47] can be applied to predict students' grades in the next examination or predict if a student can successfully graduate. In this case, educators are able to achieve useful knowledge that can help them make better decisions like the course design or other pedagogical strategies.

5 Acknowledgments

This work was supported by the Major Education Teaching Reform Research Project of Fujian Universities (FBJG20180016), the Innovation and Entrepreneurship Education Reform Project of Huaqiao University (612-50319003, 600012-50000226), Huaqiao University Fund Project (2019HJY01), and Key Laboratory of Cloud Computing and Internet-of-Thing Technology (Quanzhou University of Information Engineering), Fujian Province University (YJSW1804).

6 References

[1] Lamichhane, C.D.: 'Understanding the education philosophy and its implications', *NCC J.*, 2018, **3**, (1), pp. 24–29

[2] 'Education', https://en.wikipedia.org/wiki/Education, accessed 6 June 2019

[3] Hoot, J.L., Szente, J., Mebratu, B.: 'Early education in Ethiopia: progress and prospects', *Early. Child. Educ. J.*, 2004, **32**, (1), pp. 3–8

[4] Jamison, E.A., Jamison, D.T., Hanushek, E.A.: 'The effects of education quality on income growth and mortality decline', *Econ. Educ. Rev.*, 2007, **26**, (6), pp. 771–788

[5] Mirowsky, J.: '*Education, social Status, and health*' (Routledge, Philadelphia, 2017)

[6] Power, C.: 'The power of education', in Maclean, R., Watanabe, R., Symaco, L.P. (Eds.): '*The power of education*' (Springer, Melbourne, 2015), pp. 1–14

[7] Baradwaj, B.K., Pal, S.: 'Mining educational data to analyze students' performance', *Int. J. Adv. Comput. Sci. Appl.*, 2011, **2**, (6), pp. 63–69

[8] Romero, C., Ventura, S.: 'Educational data mining: a survey from 1995 to 2005', *Expert Syst. Appl.*, 2007, **33**, (1), pp. 135–146

[9] Romero, C., Ventura, S.: 'Data mining in education', *Wiley Interdiscip. Rev.: Data Mining Know. Discov.*, 2013, **3**, (1), pp. 12–27

[10] Shmueli, G., Bruce, P.C., Yahav, I., *et al.*: '*Data mining for business analytics: concepts, techniques, and applications in R*' (John Wiley & Sons, Hoboken, 2017)

[11] Soni, J., Ansari, U., Sharma, D., *et al.*: 'Predictive data mining for medical diagnosis: an overview of heart disease prediction', *Int. J. Comput. Appl.*, 2011, **17**, (8), pp. 43–48

[12] Agrawal, S., Agrawal, J.: 'Survey on anomaly detection using data mining techniques', *Procedia Comput. Sci.*, 2015, **60**, pp. 708–713

[13] Kaur, M., Kang, S.: 'Market basket analysis: identify the changing trends of market data using association rule mining', *Procedia Comput. Sci.*, 2016, **85**, pp. 78–85

[14] Koedinger, K.R., D'Mello, S., McLaughlin, E.A., *et al.*: 'Data mining and education', *Wiley Interdiscip. Rev.: Cogn. Sci.*, 2015, **6**, (4), pp. 333–353

[15] Dutt, A., Ismail, M.A., Herawan, T.: 'A systematic review on educational data mining', *IEEE. Access.*, 2017, **5**, pp. 15991–16005

[16] Dutt, A., Aghabozrgi, S.M., Ismail, A.B., *et al.*: 'Clustering algorithms applied in educational data mining', *Int. J. Inf. Electron. Eng.*, 2015, **5**, (2), pp. 112–116

[17] Romero, C., Ventura, S.: 'Educational data mining: a review of the state of the art', *IEEE Trans. Syst., Man, Cybern., Part C (Applications and Reviews)*, 2010, **40**, (6), pp. 601–618

[18] Al-Radaideh, Q.A., Al-Shawakfa, E.M, Al-Najjar, M.I.: 'Mining student data using decision trees'. Int. Arab Conf. on Information Technology (ACIT'2006), Yarmouk University, Jordan, November 2006, pp. 1–5

[19] Pandey, U.K., Pal, S.: 'Data mining: a prediction of performer or underperformer using classification', *Int. J. Comput. Sci. Inf. Technol.*, 2011, **2**, (2), pp. 686–690

[20] Jacob, J., Jha, K., Kotak, P., *et al.*: 'Educational data mining techniques and their applications'. 2015 IEEE Int. Conf. on Green Computing and Internet of Things (ICGCIoT), Noida, India, 2015, pp. 1344–1348

[21] Xu, D., Tian, Y.: 'A comprehensive survey of clustering algorithms', *Ann. Data Sci.*, 2015, **2**, (2), pp. 165–193

[22] Boada, R., Borkowski, R., Monroy, I.T.: 'Clustering algorithms for Stokes space modulation format recognition', *Opt. Express*, 2015, **23**, (12), pp. 15521–15531

[23] Spakov, O., Miniotas, D.: 'Application of clustering algorithms in eye gaze visualizations', *Inf. Technol. Control*, 2015, **36**, (2), pp. 214–216

[24] Darabkh, K.A., Albtoush, W.Y., Jafar, I.F.: 'Improved clustering algorithms for target tracking in wireless sensor networks', *J. Supercomput.*, 2017, **73**, (5), pp. 1952–1977

[25] Berkhin, P.: 'A survey of clustering data mining techniques', in Kogan, J., Teboulle, M., Nicholas, C. (Eds.): '*Grouping multidimensional data*' (Springer, Berlin, 2006), pp. 25–71

[26] Oyelade, O.J., Oladipupo, O.O., Obagbuwa, I.C.: 'Application of K-means clustering algorithm for prediction of students' academic performance', *Int. J. Comput. Sci. Inf. Sec.*, 2010, **7**, (1), pp. 292–295

[27] Chen, S.Y., Liu, X.: 'An integrated approach for modeling learning patterns of students in web-based instruction: a cognitive style perspective', *ACM Trans. Comput.-Hum. Interact. (TOCHI)*, 2008, **15**, (1), pp. 1–28

[28] Wook, M., Yahaya, Y.H., Wahab, N., *et al.*: 'Predicting NDUM student's academic performance using data mining techniques'. 2009 Second Int. Conf. on Computer and Electrical Engineering, Dubai, United Arab Emirates, December 2009, Vol. 2, pp. 357–361

[29] Shovon, M.H.I., Haque, M.: 'An approach of improving student's academic performance by using K-means clustering algorithm and decision tree', *Int. J. Adv. Comput. Sci. Appl.*, 2012, **3**, (8), pp. 146–149

[30] Durairaj, M., Vijitha, C.: 'Educational data mining for prediction of student performance using clustering algorithms', *Int. J. Comput. Sci. Inf. Technol.*, 2014, **5**, (4), pp. 5987–5991

[31] Bovo, A., Sanchez, S., Héguy, O., *et al.*: 'Clustering moodle data as a tool for profiling students'. 2013 Second Int. Conf. on E-Learning and E-Technologies in Education (ICEEE), Lodz, Poland, September 2013, pp. 121–126

[32] Celebi, M.E., Kingravi, H.A., Vela, P.A.: 'A comparative study of efficient initialization methods for the K-means clustering algorithm', *Expert Syst. Appl.*, 2013, **40**, (1), pp. 200–210

[33] Heil, J., Häring, V., Marschner, B., *et al.*: 'Advantages of fuzzy K-means over K-means clustering in the classification of diffuse reflectance soil spectra: a case study with west african soils', *Geoderma*, 2019, **337**, pp. 11–21

[34] Rai, P., Singh, S.: 'A survey of clustering techniques', *Int. J. Comput. Appl.*, 2010, **7**, (12), pp. 1–5

[35] Fan, Z., Chiong, R., Hu, Z., *et al.*: 'A two-layer wang-mendel fuzzy approach for predicting the residuary resistance of sailing yachts', *J. Intell. Fuzzy Syst.*, 2019, **36**, (6), pp. 1–11

[36] Hung, M.C., , Yang, D.L.: 'An efficient fuzzy C-means clustering algorithm'. Proc. IEEE Int. Conf. on Data Mining (ICDM), San Jose, USA, 2001, pp. 225–232

[37] Suganya, R., Shanthi, R.: 'Fuzzy C-means algorithm-a review', *Int. J. Sci. Res. Publications*, 2012, **2**, (11), p. 1

[38] Vandamme, J.P., Meskens, N., Superby, J.F.: 'Predicting academic performance by data mining methods', *Educ. Econ.*, 2007, **15**, (4), pp. 405–419

[39] Siemens, G., Long, P.: 'Penetrating the fog: analytics in learning and education', *EDUCAUSE Rev.*, 2011, **46**, (5), pp. 30–40

[40] Dunn, J.C.: 'A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters', *J. Cybern.*, 1973, **3**, (3), pp. 32–57

[41] Fan, Z., Chiong, R., Hu, Z., *et al.*: 'Investigating the effects of varying cluster numbers on anomalies detected in mining machines'. 2017 Int. Conf. on Computer and Drone Applications (IConDA), Kuching, Malaysia, November 2017, pp. 82–86

[42] Krinidis, S., Chatzis, V.: 'A robust fuzzy local information c-means clustering algorithm', *IEEE Trans. Image Process.*, 2010, **19**, (5), pp. 1328–1337

[43] Gou, J., Hou, F., Chen, W., *et al.*: 'Improving wang-mendel method performance in fuzzy rules generation using the fuzzy C-means clustering algorithm', *Neurocomputing*, 2015, **151**, pp. 1293–1304

[44] Wu, Z., Wu, Z., Zhang, J.: 'An improved FCM algorithm with adaptive weights based on SA-PSO', *Neural Comput. Appl.*, 2017, **28**, (10), pp. 3113–3118

[45] Witten, I.H., Frank, E., Hall, M.A., *et al.*: '*Data mining: practical machine learning tools and techniques*' (Morgan Kaufmann, Burlington, 2016)

[46] Xu, J., Xu, C., Zou, B., *et al.*: 'New incremental learning algorithm with support vector machines', *IEEE Trans. Syst., Man, Cybern.: Syst.*, 2018, (99), pp. 1–12, doi: 10.1109/TSMC.2018.2791511

[47] Carleo, G., Troyer, M.: 'Solving the quantum many-body problem with artificial neural networks', *Science*, 2017, **355**, (6325), pp. 602–606