Enhancing BMI-Based Student Clustering by Considering Fitness as Key Attribute

Erik Dovgan[1(✉)], Bojan Leskošek[2], Gregor Jurak[2], Gregor Starc[2], Maroje Sorić[2], and Mitja Luštrek[1]

[1] Department of Intelligent Systems, Jožef Stefan Institute, Jamova cesta 39, 1000 Ljubljana, Slovenia
{erik.dovgan,mitja.lustrek}@ijs.si

[2] Faculty of Sport, University of Ljubljana, Gortanova ulica 22, 1000 Ljubljana, Slovenia
{bojan.leskosek,gregor.jurak,gregor.starc,maroje.soric}@fsp.uni-lj.si

Abstract. The purpose of this study was to redefine health and fitness categories of students, which were defined based on body mass index (BMI). BMI enables identifying overweight and obese persons, however, it inappropriately classifies overweight-and-fit and normal-weight-and-non-fit persons. Such a classification is required when personalized advice on healthy life style and exercises is provided to students. To overcome this issue, we introduced a clustering-based approach that takes into account a fitness score of students. This approach identifies fit and not-fit students, and in combination with BMI, students that are overweight-and-fit and those that are normal-weight-and-non-fit. These results enable us to better target students with personalized advice based on their actual physical characteristics.

Keywords: Improving BMI-based classification · Fitness-based clustering · Multiobjective problem

1 Introduction

According to WHO, overweight and obesity have become urgent global health issues in recent decades [5]. Overweight and obese persons are classified according to the body mass index (BMI). This weight-to-height index enables defining categories of adolescents such as Overweight and Obese Adolescents (OOA) categories [1]. OOA defines four categories from the lowest to the highest BMI: underweight, normal weight, overweight, and obese. The BMI bounds for these categories are sex- and age-specific, and are typically given with sex-specific BMI-for-age charts.

This work is part of a project that has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no. 727560.

© Springer Nature Switzerland AG 2019
P. Kralj Novak et al. (Eds.): DS 2019, LNAI 11828, pp. 155–165, 2019.
https://doi.org/10.1007/978-3-030-33778-0_13

The main advantage of the BMI index and the resulting categories is its simplicity to measure. More precisely, it requires only two easy-to-obtain measurements: body weight and height. Its simplicity also represents its drawback: BMI fails to identify persons that, for example, have high muscle mass. Although they are overweight according to BMI, they are fit and should be treated differently than overweight persons without high muscle mass. This is a key issue when providing personalized advice on healthy life style and exercises, e.g., to students in high school. For example, the advice for students with high BMI and high muscle mass should be significantly different than for those with high BMI only.

BMI in combination with OOA has been widely used to study the correlation between obesity and health conditions in the last decades. For example, various risk factors were analyzed with respect to the OOA categories [1]. In some cases, however, BMI is not enough for accurate prediction. For example, it was shown that the prevalence of excess adiposity is overestimated by BMI in blacks within the pediatric population [10], which mirrors our own observation that BMI is not always appropriate for health-related clustering.

There were also studies on the relation between BMI and fitness. For example, cardiovascular risk profile was investigated in Caucasian males with at least 3 h of sports activity per week and the results showed that the threshold for an optimal BMI concerning cardiovascular risk factors might be far below $25\,\mathrm{kg/m^2}$ even if other lifestyle conditions are apparently optimal [7]. Heart failure mortality in men was studied in relation to cardiorespiratory fitness and BMI, and the results showed that the risk factor was significantly lower in fit compared with unfit men in normal and overweight body mass index but not in obese men [4].

The existing research shows that both BMI and fitness are important for assessing health status of persons and predicting health issues. In addition, it also shows that BMI and fitness score are two distinctive measurements: we cannot precisely predict one from the other, although some correlation exists. See, for example, Farrell et al. [4] who showed that there are unfit and normal weight persons, and those that are fit and obese. However, in contrast to BMI, there is no commonly used definition of fitness score. We propose to overcome this issue by considering a widely used test battery. This test battery is performed by students in Slovenian schools once a year and enables us to calculate an overall fitness score as well as access the main components of physical fitness (see Table 1). In contrast to related work, we do not predefine the clusters of fit and not fit persons, but we apply a multiobjective approach with three objectives to search for the best split into fit/non-fit clusters. In addition, the fitness score in combination with OOA categories enables the identification of persons that are overweight or obese but are fit, and those that have normal weight but are not fit. The resulting categories of students enable the teachers, parents and policy makers to create and provide personalized and better-targeted advice, recommendations and curricula.

The paper is further organized as follows. The fitness-based approach for clustering students is described in Sect. 2. Section 3 reports the experiments including the dataset and the results. Finally, Sect. 4 concludes the paper with ideas for future work.

Table 1. The physical fitness tests of the test battery. All the test measurements are in percentiles.

Fitness test	Measurement
PTSF	Thickness of triceps skinfold
PAPT	Reaction time during arm plate tapping
PSBJ	Distance jumped during standing broad jump
POCB	Time to pass a polygon backwards and on all fours
PSU	Number of sit-ups in 60 s
PSR	Distance between fingertips and toes when standing and bending forward
PBAH	Time in a bent arm position while hanging from a bar
P60m	Time to run 60 m
P600m	Time to run 600 m

2 A Fitness-Based Approach for Clustering Normal Weight, Overweight and Obese Students

There is no golden standard for deciding who is fit and who is not. The most straightforward approach to separate students who are fit from those who are not is to apply a threshold to the overall fitness score. However, it is not clear what this threshold should be, and since we have measurements of the main components of physical fitness available, we should consider whether they can be used to achieve a better separation. In our clustering, we explore these questions and finally propose an approach for separating the fit students from the non-fit.

2.1 Fitness Score

The fitness score is calculated by taking into account a set of physical fitness measurements. These measurements are obtained with the SLOfit test battery[1], i.e, a version of Eurofit Physical Fitness Test Battery [3], which is a set of physical fitness tests covering flexibility, speed, endurance, and strength. The selected set of measurements is shown in Table 1. For each measurement, a quantile (percentile) rank is calculated by taking into account sex and age. Utility functions then transform these ranks on a scale ranging from 0 to 100 points, where 0 is the worst possible score and 100 is the best. Finally, the points of all the measurements are summed up and the fitness score is determined as the quantile rank by taking into account reference population, sex and age.

2.2 Measuring Clustering Error

The fitness score enables the evaluation of clusters of students within a dataset: students within a cluster should have similar fitness score, while fitness score

[1] http://en.slofit.org/measurements/test-battery.

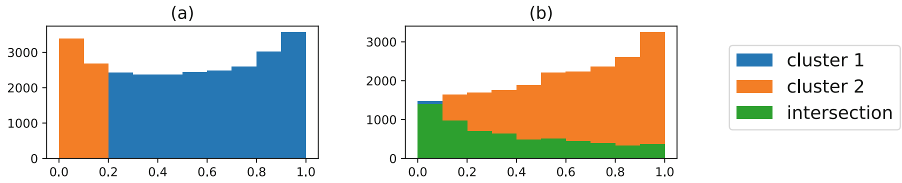

Fig. 1. Examples of clusters: (a) good clusters, i.e., there is no intersection between clusters; (b) bad clusters, i.e., intersection between the clusters is very high.

of various clusters should be different. To evaluate a pair of clusters, we firstly calculate the histograms of both clusters with respect to the fitness score. Next, we find the intersection between the histograms. The intersection represents the overlap between clusters, which ideally should be 0, since clusters should be disjunctive. Therefore, this intersection represents the error that is then normalized with respect to the size of both clusters. The resulting maximal error percentage between both clusters is then used as the amount of error with respect to the fitness score (e_f). Examples of histograms of clusters and intersections between them are presented in Fig. 1: Good clusters with no overlap are shown in Fig. 1a, while Fig. 1b depicts bad clusters with high percentage of overlap.

The same error function can be also applied to percentile ranks of fitness components (i.e., physical fitness measurements), which can be interpreted as follows: we want to find clusters in which students have similar percentile ranks of fitness components, while the percentile ranks between clusters should differ. As a consequence, the performance of fit and non-fit students with respect to individual components should be different. The error measure based on percentile ranks of fitness components (e_c) is thus calculated as the average of all the errors of individual fitness components.

Although the clusters of students with respect to the fitness score can significantly differ from the clusters based on OOA, it is reasonable to assume that the ratio between students with normal weight and those that are overweight or obese is similar to the ratio between fit and non-fit students. Note that the boundary between people with normal weight and those that are overweight or obese is to some degree arbitrary, and the same can be said for those who are fit or not. Therefore we assume the same ratio for the latter as for the former. As a consequence, the number of fat-and-fit students should be roughly the same as the number normal-weight-and-non-fit students. However, the exact numbers might differ, therefore we measure the error with respect to size difference (e_s) as the normalized difference between the size of normal weight students and the size of fit students.

The proposed approach enables us to evaluate and compare various clustering algorithms that aim at clustering students into fit and non-fit clusters. The comparison is done in three-objective space, where the errors (e_f, e_c, e_s) represent the dimensions, i.e., objectives, of this space: the error with respect to the total fitness (e_f), the average error with respect to individual fitness components (e_c), and the error with respect to the size (e_s). Note that all the errors should be minimized.

2.3 Clustering Based on Fitness Score

Besides applying existing clustering algorithms to solve the problem of finding clusters of fit and non-fit students, we also propose the following algorithm. First, the fitness score is discretized equidistantly. Second, each discretized value is used as a limit as follows: all the students with lower fitness score are added to the first cluster, while the students with higher score are added to the second cluster. Each such pair of clusters is evaluated with respect to the error functions (e_f, e_c, e_s). In comparison to other clustering algorithms, this approach has the advantage of being intuitive, easy to understand, and very effective. Its performance in comparison to other clustering algorithms is presented in the following section.

3 Experiments and Results

This section presents the dataset of students that were clustered, the clustering algorithms the were applied, and the obtained results with discussion.

3.1 Dataset of Physical Fitness Measurements

We evaluated our approach on a dataset of students from Slovenian schools, SLOfit[2]. More precisely, we only analyzed the data of high school students (ages 16–21). In addition, only the most recent year of measurements was used, i.e., 2018. The attributes for the clustering algorithms were percentile ranks of fitness components and are shown in Table 1. Moreover, only normal weight, overweight and obese students were selected. Note that the same approach can also be applied to underweight students, however, for the domain experts the most relevant division is between normal weight and overweight students. In total, 27,304 students were taken into account.

3.2 Clustering Algorithms

The clusters of fit and not-fit students were found with a set of clustering algorithms. Since the goal was to cluster in two clusters, only those algorithms that enabled defining the number of clusters were selected. However, several clustering algorithms have a high computational complexity, therefore, only a subset of data was clustered with those. In addition, some algorithms enabled creating a model on the subset and afterward cluster all the data with that model. The applied clustering algorithms and their characteristics are shown in Table 2. This table shows, for example, that spectral clustering has a high computational complexity, since only 5000 data could be clustered at once, and does not build

[2] http://www.slofit.org/.

Table 2. Evaluated clustering algorithms.

Clustering algorithm	Clustered data	Cluster all data with a model	Randomly set parameters
OOA (Default, based on [5])	All	Not needed	/
k-means [6]	All	Not needed	Random state
BIRCH [11]	5000	Yes	Threshold, Data sample
Spectral clustering [9]	5000	No	Random state, Data sample
Hierarchical clustering [8]	5000	No	Data sample
Fitness score (see Sect. 2.3)	All	Not needed	Fitness score bound

a model to cluster the entire dataset after clustering the subset of data. On the contrary, k-means has a lower computational complexity since it was able to cluster the entire dataset at once. Consequently, it was not required to use subset of data and build a model to cluster all the data. BIRCH is something in between: it has a high computational complexity, therefore it could cluster only subset of data. However, it enables building a model on this subset of data, which was then used to cluster the entire dataset.

In our experiment, all of these algorithms were run 1000 times with randomly set parameter values (and randomly selected subset of data, if all the data could not be clustered due to algorithm's high computational complexity).

3.3 Results of the Clustering Algorithms

All the algorithms had to cluster the students into two clusters, i.e., students that are fit and those that are not fit. As described in Sect. 3.2, 1000 runs of each algorithm were performed, therefore the results of all the runs are presented. Each algorithm run was evaluated and is presented in terms of three objective/error functions (e_f, e_c, e_s) as described in Sect. 2.3.

The results in three-dimensional objective space are shown in Fig. 2a. In addition, Figs. 2b–c show two additional perspective of the objective space: the first one focuses on the fitness score error, while the second one focuses on fitness components' and delta size errors. Since all three objectives are minimized, the optimal solution would be in (0, 0, 0), which is at the bottom left of all three figures.

These results show that the OOA clustering is not good with respect to the fitness score and fitness components' errors, since all the other algorithms are better in these two objectives. On the other hand, it is optimal with respect to the delta size error, which is true by definition, since delta size error measures the difference of sizes of the obtained clusters compared to the OOA clusters. In addition, k-means, spectral clustering and Fitness score clustering find the best splits with respect to the fitness components' error, while the Fitness score

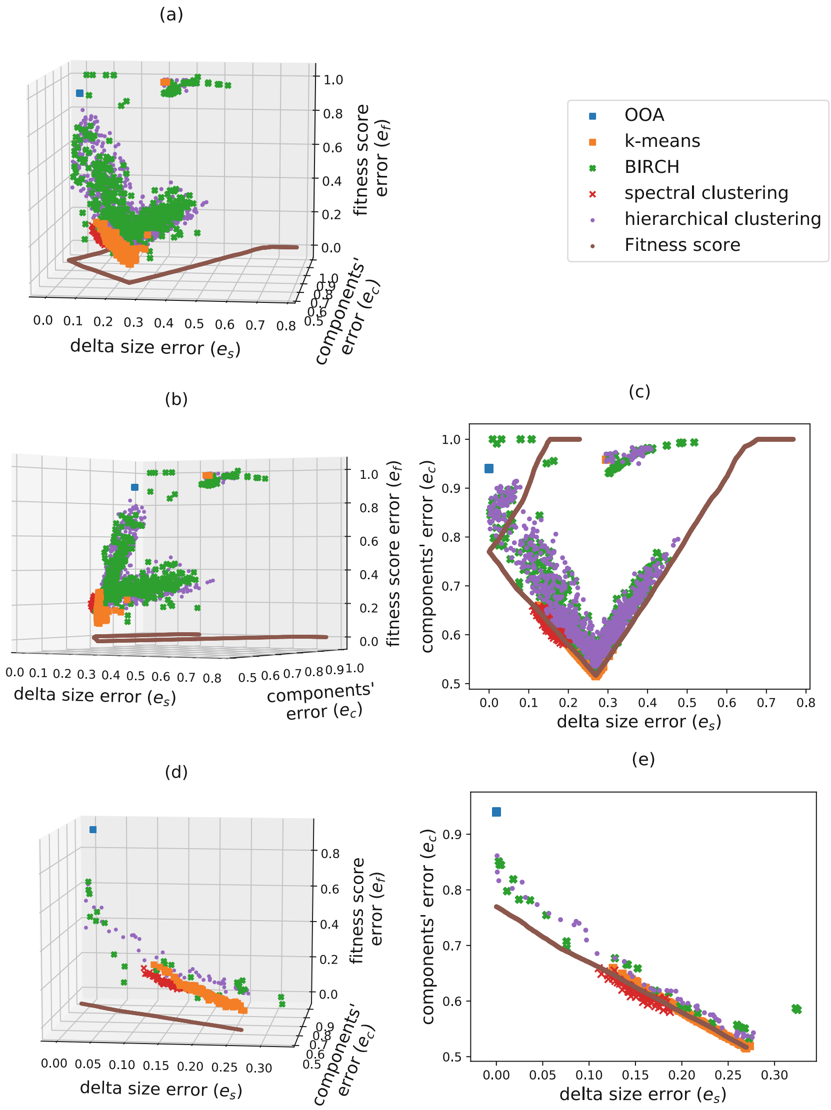

Fig. 2. Results of the clustering algorithms: (a) three-dimensional objective space; (b) focus on fitness score error; (c) projection on two objectives: fitness components' and delta size errors; (d) nondominated solutions in the three-dimensional objective space; (e) nondominated solutions projected on two objectives: fitness components' and delta size errors.

Table 3. Hypervolume of the clustering algorithms.

Clustering algorithm	Hypervolume
OOA	0.006
k-means	0.368
BIRCH	0.372
Spectral clustering	0.309
Hierarchical clustering	0.343
Fitness score	0.450

clustering also finds the best splits with respect to the delta size error. Moreover, the Fitness score clustering outperforms all the other algorithms with respect to the fitness score error, which is expected since the Fitness score clustering defines clusters that do not overlap with respect to fitness score (thus fitness score error is 0).

Figures 2a–c show all the solutions found by the clustering algorithms. However, when comparing the algorithms, it is simpler to only show nondominated solutions of each algorithm. A solution is nondominated if none of the objective functions can be improved in value without degrading some of the other objective values [2]. Therefore, a dominated solution can be discarded since there exists at least one (nondominated) solution that is equal or better in all the objectives. The nondominated solutions of the clustering algorithms are shown in Figs. 2d–e. These solutions confirm the above described comparison between the clustering algorithms.

Objective space enables us to compare results of clustering algorithms visually. However, a more appropriate approach for algorithm comparison consists of applying a unary operator suitable for multiobjective problems. A commonly used operator is the hypervolume [12]. Hypervolume measures the volume of the portion of the objective space that is dominated by the (nondominated) solutions. As a consequence, a higher hypervolume is preferable. The hypervolumes covered by the solutions of the clustering algorithms are listed in Table 3. This table shows that the Fitness score clustering found solutions that better cover the objective space in comparison to the other algorithms. Another argument in favor of the Fitness score clustering is that other algorithms only rarely outperform it in terms of fitness components and delta size error (as best seen in Figs. 2c and e), while the Fitness score clustering significantly outperforms the other algorithms in terms of the fitness score error (as best seen in Fig. 2b).

A solution found with the Fitness score clustering is presented in Fig. 3 in terms of distributions of BMI, OOA, fitness score and fitness components between the two clusters. This figure shows the clusters divided by fitness score 0.5, i.e., the division with the lowest fitness components' error.

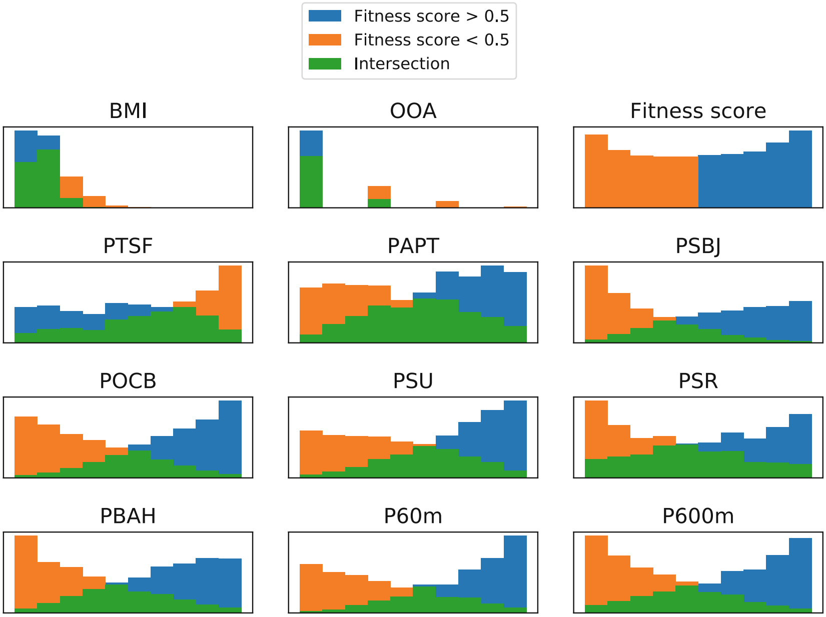

Fig. 3. Distribution of data with respect to BMI, OOA, fitness score and fitness components, which are additionally clustered with the Fitness score clustering in students with fitness score <0.5 and students with fitness score >0.5.

3.4 Discussion

The presented experiment has shown that the best clusters with respect to the three objectives are found by the Fitness score clustering. This can be seen in visual representation of the solutions in the objective space, and it is also confirmed by the hypervolumes obtained by the clustering algorithms. In addition, the Fitness score clustering enables finding clusters with the lowest (zero) delta size error, and with the lowest fitness components' error (the same fitness components' error is also achieved by the k-means algorithm). Even more, all the clusters of the Fitness score clustering have zero fitness score error, while none of the other clustering algorithms found clusters with zero fitness score error. Therefore, the Fitness score clustering performed the best among the tested algorithms.

4 Conclusion

This paper presented a new approach for identifying students that are overweight and fit, and those that have normal weight, but are not fit. This classification enhances the widely used BMI index that is suitable to classify students only as normal weight or overweight/obese. The presented approach introduces the fitness score calculated based on a set of physical fitness measurements performed in schools by all the students once a year. In addition, it also defines three objectives, i.e., (fitness score error, fitness components' error and delta size error), which are used to assess the quality of the clustering algorithms that find clusters of students. Furthermore, the Fitness score clustering is developed, which clusters students with respect to their fitness score. The results show that the Fitness score clustering finds better clusters of students in comparison to widely-used general-purpose clustering algorithms. The obtained clusters enable the identification of students that are overweight or obese but are fit, and those that have normal weight but are not fit, which makes it possible to define personalized and better targeted advice, recommendations and curricula for the students.

In our future work we will evaluate the proposed approach on additional datasets of students from Slovenia and abroad. This approach will be also combined with algorithms that predict students' future performance in order to assess whether the discovered clusters can improve this prediction. A particular challenge also represents the definition/generation of personalized and better-targeted advice, recommendations and curricula.

References

1. Bacha, F., Saad, R., Gungor, N., Janosky, J., Arslanian, S.A.: Obesity, regional fat distribution, and syndrome X in obese black versus white adolescents: race differential in diabetogenic and atherogenic risk factors. J. Clin. Endocrinol. Metab. **88**, 2534–2540 (2003)
2. Deb, K., Pratap, A., Agrawal, S., Meyarivan, T.: A fast and elitist multiobjective genetic algorithm: NSGA-II. IEEE Trans. Evol. Comput. **6**(2), 182–197 (2002)
3. Eurofit: Eurofit Tests of Physical Fitness. Council of Europe, Strasbourg, 2 edn. (1993)
4. Farrell, S.W., Finley, C.E., Radford, N.B., Haskell, W.L.: Cardiorespiratory fitness, body mass index, and heart failure mortality in men. Circ. Hear. Fail. **6**(5), 898–905 (2013)
5. Kallioinen, M., Granheim, S.I.: Overweight and obesity in the western pacific region. Technical report, World Health Organization (2017)
6. Kanungo, T., Mount, D.M., Netanyahu, N.S., Piatko, C.D., Silverman, R., Wu, A.Y.: An efficient k-means clustering algorithm: analysis and implementation. IEEE Trans. Pattern Anal. Mach. Intell. **24**, 881–892 (2002)
7. Ortlepp, J.R., Metrikat, J., Albrecht, M., Maya-Pelzer, P., Pongratz, H., Hoffmann, R.: Relation of body mass index, physical fitness, and the cardiovascular risk profile in 3127 young normal weight men with an apparently optimal lifestyle. Int. J. Obes. **27**, 979–982 (2003)
8. Rokach, L., Maimon, O.: Clustering methods. In: Maimon, O., Rokach, L. (eds.) Data Mining and Knowledge Discovery Handbook, pp. 321–352. Springer, Boston (2005). https://doi.org/10.1007/0-387-25465-X_15
9. Shi, J., Malik, J.: Normalized cuts and image segmentation. IEEE Trans. Pattern Anal. Mach. Intell. **22**, 888–905 (2000)

10. Weber, D.R., Moore, R.H., Leonard, M.B., Zemel, B.S.: Fat and lean BMI reference curves in children and adolescents and their utility in identifying excess adiposity compared with BMI and percentage body fat. Am. J. Clin. Nutr. **98**(1), 49–56 (2013)
11. Zhang, T., Ramakrishnan, R., Livny, M.: Birch: an efficient data clustering method for very large databases. In: Proceedings of the 1996 ACM SIGMOD International Conference on Management of Data, pp. 103–114 (1996)
12. Zitzler, E., Thiele, L.: Multiobjective evolutionary algorithms: a comparative case study and the strength Pareto approach. IEEE Trans. Evol. Comput. **3**(4), 257–271 (1999)