

RESEARCH ARTICLE

Analysis of college students' canteen consumption by broad learning clustering: A case study in Guangdong Province, China

Chun Yang¹, Hongwei Wen¹, Darui Jiang², Lijuan Xu², Shaoyong Hong^{1,2*}

1 School of Accounting, Guangzhou Huashang College, Guangzhou, China, **2** School of Data Science, Guangzhou Huashang College, Guangzhou, China

* shy2002021@163.com

**OPEN ACCESS**

Citation: Yang C, Wen H, Jiang D, Xu L, Hong S (2022) Analysis of college students' canteen consumption by broad learning clustering: A case study in Guangdong Province, China. PLoS ONE 17(10): e0276006. <https://doi.org/10.1371/journal.pone.0276006>

Editor: Brij Bhooshan Gupta, Asia University, TAIWAN

Received: July 19, 2022

Accepted: September 27, 2022

Published: October 13, 2022

Copyright: © 2022 Yang et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the manuscript and its [Supporting information](#) files.

Funding: This work is funded by Innovation Team Project of Humanities and Social Sciences in Colleges and Universities of Guangdong Province (2020WCXTD008), The Project of Philosophy and Social Sciences in Guangdong Province (GD17XGL19), The Scientific Project of Guangzhou Huashang College (Grant: 2020HSDS03, 2021HSKT01). The funders had no role in study

Abstract

Investigation on college students' consumption ability help classify them as from rich or relative poor family, thus to distinguish the students who are in urgent need for government's economic support. As canteen consumption is the main part of the expenses of the college students, we proposed the adjusted K-means clustering methods for discrimination of the college students at different economic levels. To improve the discrimination accuracy, a broad learning network architecture was built up for extracting informative features from the students' canteen consumption records. A fuzzy transformed technique was combined in the network architecture to extend the candidate range for identifying implicit informative variables from the single type of consumption data. Then, the broad learning network model is fully trained. We specially designed to train the network parameters in an iterative tuning mode, in order to find the precise properties that reflect the consumption characteristics. The selected feature variables are further delivered to establish the adjusted K-means clustering model. For the case study, the framework of combining the broad learning network with the adjusted K-means method was applied for the discrimination of the canteen consumption data of the college students in Guangdong province, China. Results show that the most optimal broad learning architecture is structured with 14 hidden nodes, the model training and testing results are appreciating. The results indicated that the framework was feasible to classify the students into different economic levels by analyzing their canteen consumption data, so that we are able to distinguish the students who are in need for financial aid.

1. Introduction

Consumption is a special behavior that reflects the people's possession of objects. The possession of objective ability is a necessary condition for the implementation of behavior [1]. Consumption ability is a concept accumulated by the rapid development of the social economic system. For the consumption behaviors, in addition to the knowledge and skills of consumer action strategies, the identification, comparison and selection of products will take extra time

design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

and energy, while the price the products account for most of the causal ratio [2]. Awareness and attention about the well-being of people's consumption ability and consuming behaviors have increased in many countries, including China. Consumers with more abundant disposable time and energy and sufficient economic conditions have stronger consumption power [3].

University is a microcosm of a society. College students are participating in the transition to independence on attention to a highlighted set of features from the social character [4]. College students' consumption ability is the basis of social consumption behavior. When individuals do not have or have insufficient consumption ability, even if they have consumption intentions, they will eventually lack corresponding consumption behaviors. According to the results of the interviews, it is found that the individual's lack of consumption connotation knowledge and action knowledge, economic ability constraints, individual inherent habits and preferences such as cynicism, lazy and troublesome personality barriers hinder consumer behavior [5]. Therefore, investigation on college students' consumption property help classify them as from rich or relative poor family, thus to distinguish the students who are in urgent need for government's economic support.

The habitual consumption behavior of the young adults in universities can reflect the economic level and the unbalanced development of the society in the next few years. College students' consumption behavior is on concern. Eating consumption is recognized as universal consumption among many consumption methods [6, 7]. Thus the eating habits of the college students are the key factors that can directly reflect the consumption ability of the students' families. Generally, the students' eating habits in colleges and universities are mainly reflected in their canteen consumption records [8]. The characteristics of canteen consumption will release some prominent marker to tell the economic difference of the students' families. Consequently, we can use the consumption characteristics of college students spending in the college canteen to classify and discriminate the students' family economic situation.

In recent years, research on college students' consumption ability has been actively carried out. For example, Kim et al. discussed the students' behavior difference about clothing consumption in comparison between the clothing majors and non-majors [9]. Bruce et al. used conventional statistical methods (such as the p-value and some simple descriptive indicators) to study the male first-year college students' consumption change on in sugar-sweetened beverage [10]. Gao et al. performed an empirical study on the mechanism of college students' green consumption in China [11]. They used questionnaires and make statistical analysis, mainly accomplished by using the SPSS software and lack of novel design in methodologies. There are relatively few studies concerning on data mining by investigating machine learning methods for qualitative or quantitative analysis [12, 13]. A review article discussed the possibility of artificial intelligence technologies be analyzed for enhancing the educational systems for semantic web-based education [14]. Cloud computing technology was used to analyze the requirements of academic and administrative affairs, education affairs, and affiliated institutions in the colleges and universities [15]. However, in the era of big data, the consumption data is presented as large volume, low value density, diverse forms, and fast growth [16, 17]; machine learning strategy plays an important role in data analysis of the trends in college student consumption. More specifically, the students' canteen data simply records the consumption amount at each time the students spend. The data is fragmented and dynamically increasing, which is bond to raise some challenging issues that probably hinder the computational model optimization [18]. Then some multi-variate statistical methods should be introduced into the design of the machine learning algorithmic flow.

Clustering is one of the important machine learning methods in many communities for data classification, discrimination, and pattern recognition [19]. Basically, it aims to group

data points into different clusters based on their similarity or density [20]. In the past decades, many clustering algorithms have been proposed such as the K-means clustering [21], spectral clustering [22], min-max cut [23], and multi-view clustering [24]. A parallel fractional lion algorithm is recently reported for data analysis, the algorithm is tested on six standard databases [25], but it is not examined by practical collected data. The fact that it operates only based on the MapReduce framework limits its further application.

Among the existing clustering methods which are suitable for flexible application, the most popular one is the K-means clustering algorithm because its algorithmic flow is simple and efficient, and its operation only needs to learn a limited number of centroids to minimize the interclass data distances. However, the clustering performance of K-means is much related to its random initialization of the centroids [26]. Spectral clustering is able to prevent the uncertain interference of the randomness. It characterizes the data connection with an appropriate graph whose vertices represent the data points and the weights represent the connection between data pairs, and tries to partition the vertices into different clusters by minimizing the cut information [27]. For this consideration, we jointly perform spectral clustering in the K-means process, in the way of adjusting the spectral rotation to derive the underlying data connection. The adjusting process is detailedly to perform eigen-decomposition on the graph Laplacian matrix transformed from the original data. The graph Laplacian matrix help to amplify the property of the data, and the eigen-decomposition is functional to extract the features [28]. After the adjusted operation, the normal K-means clustering is applied. The newly proposed methods is defined as adjusted K-means clustering.

Although the adjusted K-means method is advanced for classifying the students' canteen consumption data, the data is raw and the feature information of the data is ambiguous. To this end, we design a broad learning neural network architecture to extract the feature variables that are rich in data information. Broad learning architecture is built up based on the broad learning system (BLS) established on the idea of a fully connected neural network (FCNN) accompanied with random vector functional link [29, 30]. Specifically, the input data is transformed into numbers of feature nodes by proper mapping functions, and then these features are concatenated together to generate a series of enhancement nodes randomly, which will be used to expand the network architecture in a broad sense. All the features nodes and the outputs of the enhancement nodes are connected together to feed into the output layer. The desired output weights are to be determined by a fast ridge regression of the pseudo-inverse of the system equation [31]. The weights and biases in BLS are all randomly generated, and further to be trained in data-driven manner. The incremental learning algorithm is incorporated into the BLS which makes the network can be remodeled fast in the broad expansion without a retraining process.

Based on the principle of BLS, the broad learning neural network (BLNN) architecture is constructed as a black box model in our study for the analysis of the college student's canteen consumption. We designed the architecture to be trained in data iteration for optimal observation of the linkage weights. On the other hand, the BLNN training bias is not considered in the model training process, we designed a fuzzy-transformed strategy to catch the feature information of the data under the condition of a fuzzy interval range. What the fuzzy-transformed strategy does is to make a fault-tolerant mode to ensure a relative large amount of candidate feature variables for BLNN training. The fuzzy-transformation is in fusion design with the principal component analysis (PCA) method. Specifically, the data features were extracted by PCA algorithm before the data was input to the BLNN architecture, and the PCA procedure is improved by using the fuzzy rules of the triangular membership function. The fuzzy transformed PCA technique is able to inhibit the original noise interference and make the explicit characteristics of the principal components much remarkable [32].

We proposed an adjusted K-means clustering method for the discrimination of the college students based on their canteen consumption data, thus to classify the students into different economic levels. Also, a broad learning neural network is designed to extract the exact data features to improve the discriminant accuracy, and a fuzzy transformed PCA technique is employed to reduce the model uncertainty and thus to help reduce the classification errors. The remainder of this article is organized as follows. Section 2 introduced the adjusted clustering method, the BLNN architecture and the fuzzy transformed PCA technique as well as the specific algorithmic modification. Section 3 describes the data collection and present the descriptive statistical properties. Section 4 applies the fusion of the proposed methods to establish the combined discrimination models in a data-driven manner, to fulfil the model optimization, validation and evaluation. Section 5 makes conclusions.

2. Methodologies

2.1. The adjusted clustering algorithm

For unsupervised data analysis models, clustering analysis is the simplest classification method. Among the existing clustering methods, the most popular one is K-means clustering due to its good adaptiveness and high efficiency. It is used to tune the cluster centroids by each newly-added testing sample, to minimize the within cluster data distances [33].

The principle of K-means is understandable. Given a data matrix $X = [x_1, x_2, \dots, x_p] \in \mathbb{R}^{d \times p}$, the K-means clustering aims to partition X into k clusters ($\{C_s | s = 1, 2, \dots, k\}, 1 \leq k \leq p$), to achieve that the within-cluster sum of squared distances can be minimized and the sum of squared distances between different clusters can be maximized. The objective function of K-means clustering is

$$\min_{\{C_s\}} \sum_{s=1}^k \sum_{x_i \in C_s} \|x_i - \mu_s\|^2, \tag{1}$$

where $\|\cdot\|$ applies the Euclidean norm operation, x_i is the i -th sample in the C_s cluster set, μ_s is the centroid of C_s . To optimize the K-means results, the membership of each data point and the centroid of each cluster are alternately updated.

However, the data available for clustering in this work is some designated statistical data which points to some specific properties of the raw data. In this point of view, we proposed an adjusted K-means algorithm as the improvement of the traditional K-means method, so that the cluster operation is much suitable to deal with the comprehensive statistic data. Given a built graph, we employ a two-step spectral clustering procedure. The first step is to perform eigen-decomposition on the graph Laplacian matrix to obtain the scaled cluster indicator matrix; The second step aims to make discretization based on the matrix, to get the final cluster assignment [28].

In the adjusted K-means details, we firstly construct a graph affinity matrix $A_{p \times p}$ according to certain similarity measures to depict the relationship in the data. Let a_j be the j -th vector of the matrix $A = \{a_j | j = 1, 2 \dots p\}$, which corresponds to the cluster indicator vector for x_i . Each element of vector a_j is defined as 0–1 binary values, 1 means that $x_i \in C_s$, and 0 means otherwise. By defining the scaled cluster indicator matrix as $Y = A(A^T A)^{-1/2}$ whose column vectors are given by

$$y_{col} = \left[\underbrace{0, \dots, 0}_{\sum_{j=1}^{col-1} p_s}, \underbrace{1, \dots, 1}_{p_s}, \underbrace{0, \dots, 0}_{\sum_{j=col+1}^k p_s} \right]^T / \sqrt{p_j}, \tag{2}$$

where p_s is the number of data in the s -th cluster. Then, the objective function of adjusted

clustering can be formulated as

$$\min_Y \text{Tr}(Y^T L Y), \tag{3}$$

where L is a kind of Laplacian matrix defined as $L = D - A$, in which D is a diagonal matrix with its j -th diagonal element calculated by the numerical summation of the vector a_j . The operator $\text{Tr}(\cdot)$ is to calculate the trace of the matrix.

Since $Y^T Y = (A^T A)^{-1/2} A^T A (A^T A)^{-1/2} = I$, the embedding Y can be obtained by stacking the eigenvectors of L corresponding to its k smallest eigen values. However, Y is a real-valued matrix by default, then spatial rotation is necessary to perform discretization for the adjusted K-means clustering [34].

Suppose R is an arbitrary orthogonal matrix, for any solution Y , the matrix product YR is a new generated solution from Y by applying the corresponding spatial rotation. Therefore, spatial rotation aims at finding a proper orthogonal and normalized R such that the resultant YR is closer to the discrete indicator than Y in K-means. Mathematically, it aims to minimize the distance of YR and A , namely,

$$\begin{aligned} & \min_{A,R} \|YR - A\|^2, \\ & \text{s.t. } A \cdot 1_k = 1_p \text{ and } R^T R = I, \end{aligned} \tag{4}$$

where 1_k and 1_p are the all-one column vectors with sizes of $k \times 1$ and $p \times 1$, respectively. We can easily solve the function by using the alternative optimization method, therefore the final cluster assignment can be obtained. By the computation of formulae (2), (3) and (4), we can estimate that the adjusted K-means procedure is designed to run within an effort of $O[d^6 \times p^2]$.

2.2. The broad learning architecture

The broad learning neural network architecture is built up on the basis of a fully connected neural network (FCNN). An FCNN model is for training an adaptive machine learning model by increase input of data information [35]. For instance, a basic single-hidden-layered FCNN structure is formed as shown in Fig 1. The original data $\mathbf{x} = (x_1, \dots, x_d)^T$ is taken as the input, delivered to the hidden layer with activation mapping on the linear summation of \mathbf{x} . Further the generated hidden variables are moved into a softmax unit, fitting a calibration model with multi-variate methods, to generated a discriminant/predictive results as the output data. The FCNN computation is formulated as

$$\mathbf{y} = \text{softmax}(f_2(w_2^T(f_1(w_1^T \mathbf{x} + b_1) + b_2))), \tag{5}$$

where f_v ($v = 1,2$) is the activation functions, w_v ($v = 1,2$) is the linkage weights, and b_v ($v = 1,2$) is the constant thresholds. Subscript $v = 1$ points to the calculation from the input layer to the hidden layer, $v = 2$ means that from the hidden layer to the softmax unit, and finally to obtain the discriminant/predictive results in the output layer. On some loose definition, the softmax unit is taken as part of the output layer [36].

BLNN is an adaptive learning architecture designed based on the FCNN structure. It has similar data training function as a deep learning framework. BLNN has a widened pseudo input layer that can help to take one pre-clustering feature screening in one network structure. But in contrast to deep learning, broad learning is able to skip the process of long-time consuming on training abundant parameters in different filters and layers [29, 37]. The parameters in BLNN could be determined by random projection or by generating pseudo neural nodes to form a wide enough layer for fast tuning. Moreover, the training process can be

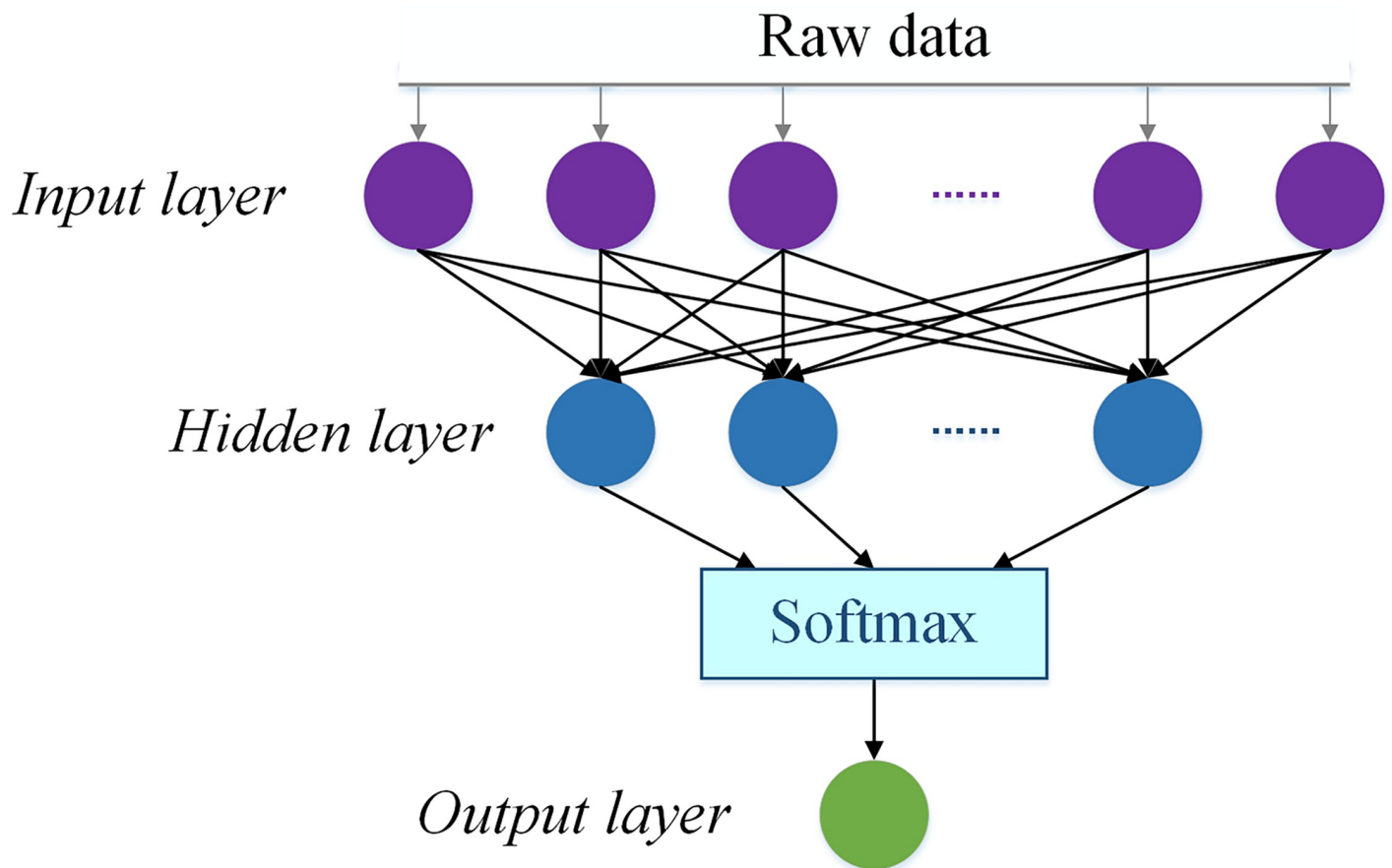


Fig 1. The FCNN structure.

<https://doi.org/10.1371/journal.pone.0276006.g001>

changed to an incremental learning mode when the network structure is extended in the width direction [38]. Specific BLNN training procedure is shown as follows,

Considering the general supervised learning task, we are given the training dataset $X_{p \times d}$ containing p sample data of d properties, and the target $Y_{p \times k}$ marking the p samples as k discriminant classes. Each row data in X and Y denotes the original sample dimensional information and target clustered makers, respectively. The training samples are taken as the original input data with d properties.

Also, the input variables are applied to generate and connect the layer of enhancement nodes. To speed up the training process, the enhancement nodes are obtained in N_a groups. Any feasible algorithmic flow can be designed to observe the outputs of the groups of enhancement nodes.

Furthermore, the neurons in different enhancement layers are taken as added input nodes to widen the range of the input layer. The BLNN mechanism is to gather all of the enhancement nodes as well as the original input nodes, and flatten them as a brand new extended input layer.

$$X^{\text{ext}} = [X, T_1, T_2, \dots, T_{N_a}], \quad (6)$$

where T_i represents the i -th group of enhancement nodes obtained by a same transformation rule.

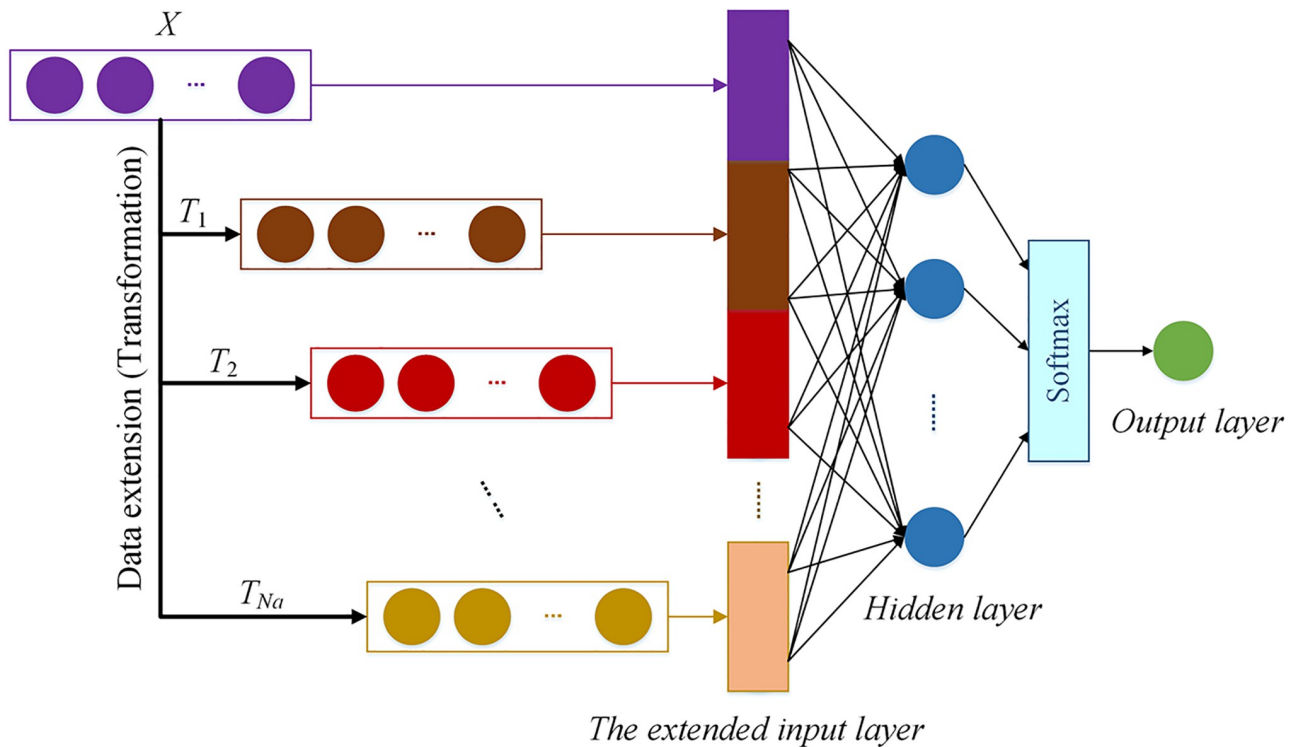


Fig 2. The BLNN architecture.

<https://doi.org/10.1371/journal.pone.0276006.g002>

Thus the BLNN architecture is built up by re-constructing the FCNN with the new input data X^{ext} , and the hidden variables are obtained in vector as

$$H = f_1(w_1^T X^{ext} + b_1), \tag{7}$$

where $H = H_{p \times m}$. The hidden variables are moved through the softmax unit and give discriminant/predictive results in the output layer, i.e.

$$\hat{Y} = \text{softmax}(f_2(w_2^T H + b_2)). \tag{8}$$

The BLNN architecture is shown in Fig 2. Practically, the BLNN architecture enables to test different number of hidden nodes (m), and adaptively tune the linkage weights (w_1 and w_2) for searching the optimal broad learning network structure. Before model training, we need to select suitable activation functions performing as f_1 and f_2 in advance. Concerning on the BLNN feature extraction by obeying the equations of (6), (7) and (8). The BLNN architecture runs in the complexity of $O[N_a^2 \times d \times p^2]$.

2.3. Fuzzy PCA transform technique

Fuzzy-transformed design is carried out under the condition of interval range of information data to be considered. It is used as the preprocessing algorithm for preventing the impact of the single type data [39]. It means to ensure that the fuzzy transform is functional to find the features from the consumption data that are not merely telling the consumption properties, then the model can be improved as conceiving a diversity of data characteristics. The fuzzy transform design of the classical PCA is able to emphasize the explicit informative features of the principal components so that the unnecessary noise can be suppressed.

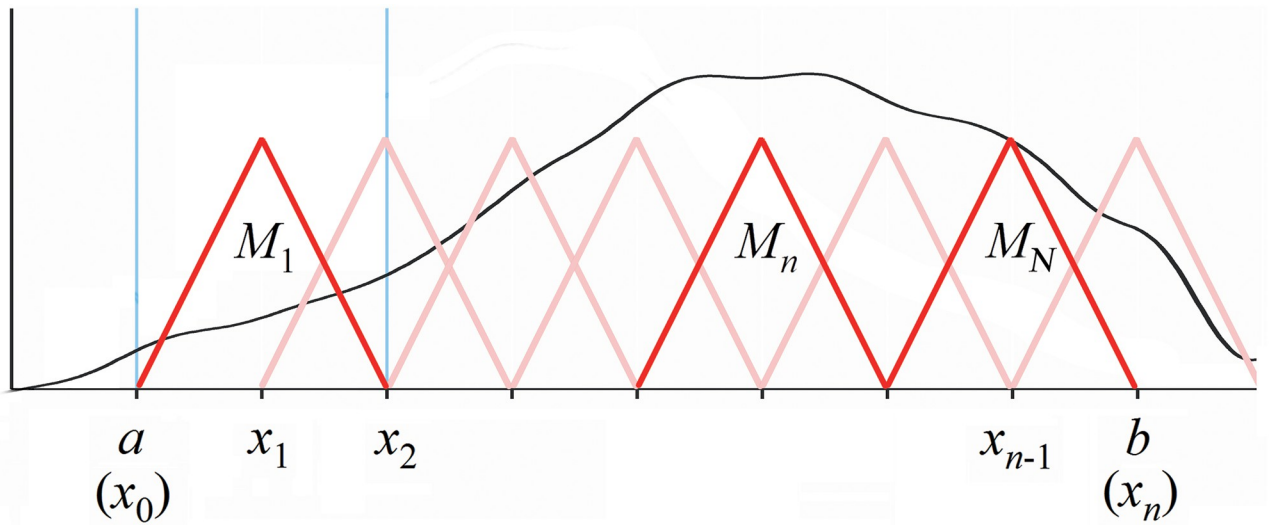


Fig 3. The triangular function for fuzzy transform.

<https://doi.org/10.1371/journal.pone.0276006.g003>

The algorithm is proceeded by the fuzzy partition of the universe fuzzification. This depends on the feature contribution of the measured data. There are several membership functions can be used or modified to fit a specific target, such as the monotonic, triangular, trapezoidal and bell-shaped functions [40]. Specifically, the triangular function (see Fig 3) can much reduce the computational overhead of fuzzy transform due to simpler operation. The triangular fuzzy transform is briefly outlined as follows,

First to calculate the membership function M_n ,

$$M_n(x) \stackrel{k=1,2,\dots, \lfloor \frac{n+1}{2} \rfloor}{\leftarrow} \begin{cases} M_{2k}(x) = 1 - \frac{\|x - x_{2k}\|}{\|c_{2k}\|}, & \text{for } \|x_{2k-1}\| < \|x\| < \|x_{2k}\| \\ M_{2k-1}(x) = \frac{\|x - x_{2k-1}\|}{\|c_{2k-1}\|}, & \text{for } \|x_{2k-2}\| < \|x\| < \|x_{2k-1}\| \\ 0, & \text{otherwise} \end{cases} \quad (9)$$

where the operator $\lfloor \cdot \rfloor$ is to get the floor integer; $\|\cdot\|$ applies the Euclidean norm operation; $x_{2k-1} = \{x_1, x_3, x_5, \dots\}$ and $x_{2k} = \{x_2, x_4, x_6, \dots\}$ are the original dataset; $c_{2k-1} = \{c_1, c_3, c_5, \dots\}$ and $c_{2k} = \{c_2, c_4, c_6, \dots\}$ are the principal components extracted by PCA.

Next to perform the fuzzy transform F_n of the n -th principal component.

$$F_n = \frac{\int_a^b f(x)M_n(x)dx}{\int_a^b M_k(x)dx}, \quad n = 1, 2, \dots, N, \quad (10)$$

where the integral interval $[a, b]$ piece-wise equals to $[\|x_{2k-1}\|, \|x_{2k}\|]$.

The fuzzy PCA transform algorithm is evident to generate a group of enhancement features in machine learning procedures [41]. Thus we designed to apply fuzzy PCA into the BLNN model to search for brand new network nodes, so that the data features are increasingly emphasized, but the computational complexity is increased by $O[N \times k \times p]$.

3. Data collection and descriptive statistics

Experimental data was from colleges and universities located in Guangdong province, China. We collected a total of 307,968 students' records from the year 2019 to 2021. The records include students' personal information such as his/her age, majority and hometown. The properties of students' consumption behaviors include card recharges, canteen expenses, utility bills, library borrowing, expenses for training and certification, etc. Commonly, every college student goes for meals in the college canteen, their expense on canteen meals is the key factor to tell their family backgrounds and economic status. Thus we targeted on the record of canteen expense data. There are some missing data and abnormal data in the 307,968 sample records. we removed them as they account for a small proportion in the total; then we have 265,972 samples for modeling analysis.

In details, the canteen expense data is large and complex. It shows a long list of students' consuming records in time orders. There appear heavy repeats of students' personal information, the meal menus and the price. To reduce the influence of overlapping information, we design to use five common descriptive statistics to summarize the data features. They are the annual-averaged consumption cash amount (yCA), the month-averaged consumption cash amount (mCA), the averaged cash amount by consumption frequency (fCA), the month-averaged consumption frequency (mCF) and the day-averaged consumption frequency (dCF). These indicators are socially meaningful. A large value of yCA, mCA and fCA represents it is economically supported that the student is capable of enjoying good meals in canteen level. On the contrary, a small value of mCF and dCF means that the student is seldom expense on canteen meals, which indicates that he/she is economically supported to have high quality meals in good restaurants other than the college canteen. Fig 4 shows the distribution of the five quantified descriptive statistics.

Additionally, some students' personal information inherently influenced their consuming behaviors. On one hand, we surveyed the students' majorities and summarized them into the subject category. The distribution by subjects is shown in Fig 5. There are relatively more students involved in the subjects of Engineering, Science, Economics, Medicine and Management, and relatively less students are in Agriculture, Art, Education and Literature. On the

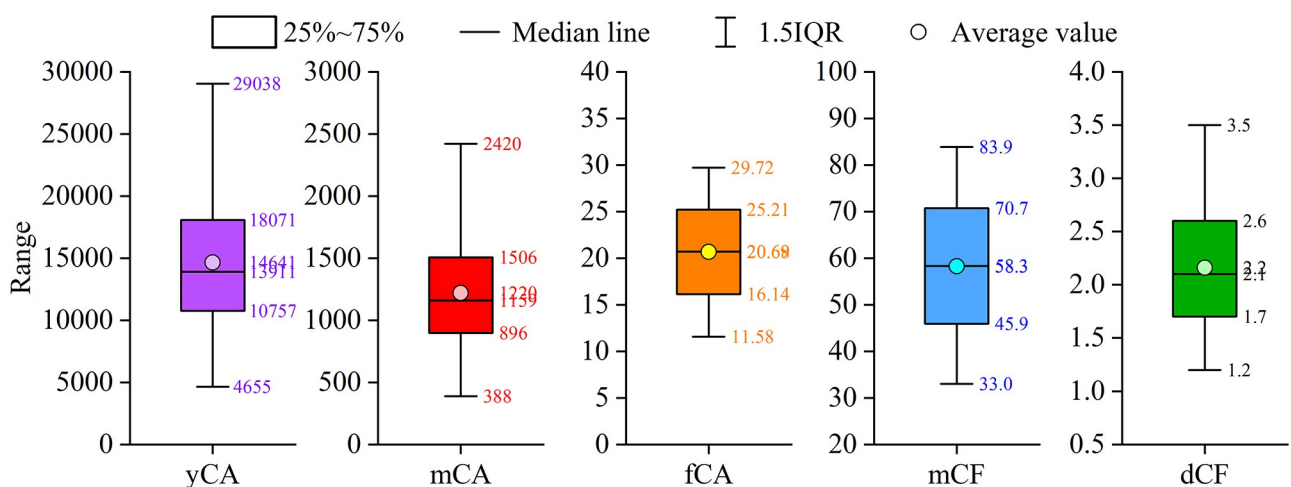


Fig 4. Distribution of the five quantified descriptive statistics.

<https://doi.org/10.1371/journal.pone.0276006.g004>

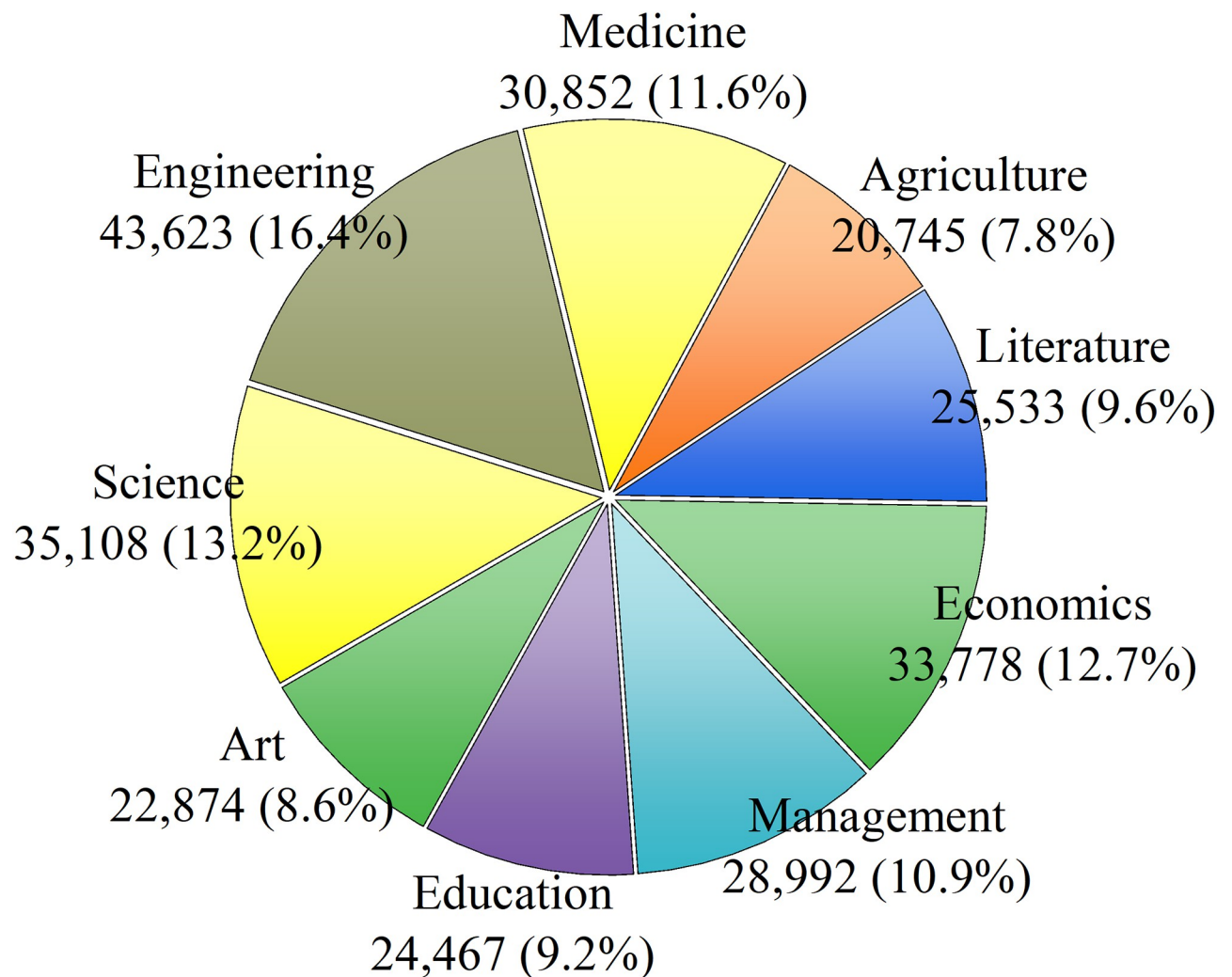


Fig 5. The distribution of students belonging to the 9 subject categories.

<https://doi.org/10.1371/journal.pone.0276006.g005>

other hand, we also performed a statistical analysis of their hometown information. The 21 cities in Guangdong province are grouped into 7 regional segmentations according to their geographical locations. The numbers of students coming from each city group are listed in [Table 1](#).

Table 1. The numbers and the percentages of students coming from each city group.

| | Number of students | Percentage | In-group cities |
|----------|--------------------|------------|-------------------------------------|
| Region 1 | 57,449 | 21.6% | Foshan, Guangzhou |
| Region 2 | 48,672 | 18.3% | Dongguan, Huizhou, Shenzhen |
| Region 3 | 40,961 | 15.4% | Jiangmen, Zhongshan, Zhuhai |
| Region 4 | 32,183 | 12.1% | Maoming, Yangjiang, Zhanjiang |
| Region 5 | 27,129 | 10.2% | Qingyuan, Yunfu, Zhaoqing |
| Region 6 | 28,993 | 10.9% | Heyuan, Meizhou, Shaoguan |
| Region 7 | 30,585 | 11.5% | Chaozhou, Jieyang, Shantou, Shanwei |

<https://doi.org/10.1371/journal.pone.0276006.t001>

4. Model designs and result discussions

4.1. Adjusted K-means modeling for discriminant metrics

For all of the 265,972 targeting samples, we have the five important descriptive statistic indicators for modeling, i.e. yCA, mCA, fCA, mCF and dCF. The former three represent the consuming cash amount while the latter two are the consuming frequency. They are of different types and the data range vary a lot, thus we have to make a 0–1 normalization before modeling. Besides, the students' consuming ability is certified by a large consuming cash amount reflected in yCA, mCA and fCA, and also by a small consuming frequency represented in mCF and dCF. As for modeling simplicity, we defined the consuming property vector as $v = [yCA, mCA, fCA, 1/mCF, 1/dCF]^T$ for each target student sample, so that the sorting discrimination is totally uniform as from the smallest to the largest on each element. Then we applied the adjusted K-means method to the data for classification of the students' consuming abilities.

We used the elbow method to select the best number of clusters k . The elbow method defines a distortion quantified index on the objective function. a small value of distortion represents the samples clustered tightly, while a large value means the samples are loose [42]. We test several candidate k values for different mode of clustering, and find the objective function $\min Tr$ for each k . the comparative plot is shown in Fig 6. The curve indicated that $k = 4$ is the best choice before the clustering becoming loose. Therefore, we decide to classify the samples into 4 classes according to their consuming property indicators. The 4 classes are generated by the second norm value of the vector v (i.e. $\|v\|$). The total of 265,972 samples are sorted in the ascending order of $\|v\|$. To some extent, the sorting list is used to estimate the students' consumption level. Then we can easily observe 4 piece-wise segments by dividing the minimum-to-maximum range by the points of Q1 (the 25% quantile), Q3 (the 75% quantile) and the median values. The segment from Minima to Q1 is marked as the cluster C_1 , the segment from

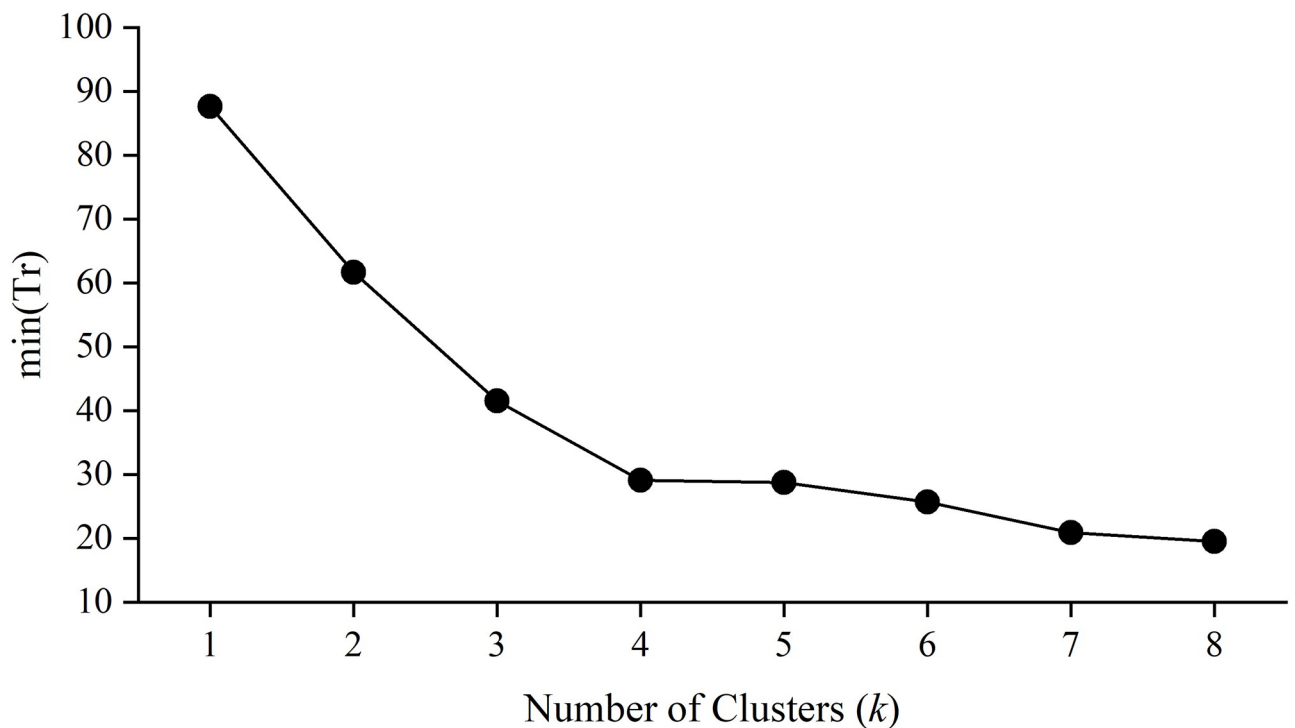


Fig 6. The elbow test for determination of the best number of clustering.

<https://doi.org/10.1371/journal.pone.0276006.g006>

Table 2. The definition of the 4 clustering classes by the quantile data segmentation.

| | Segment 1 | Segment 2 | Segment 3 | Segment 4 |
|------------------|---------------------------|---------------------------|---------------------------|---------------------------|
| $\ v\ $ | $\in [\text{Minima}, Q1]$ | $\in [Q1, \text{Median}]$ | $\in [\text{Median}, Q3]$ | $\in [Q3, \text{Maxima}]$ |
| Clustering class | C_1 | C_2 | C_3 | C_4 |
| No. of samples | 49,562 | 75,519 | 83,757 | 57,134 |
| Regarded as | Poor | Frugal | Normal | Affluent |

<https://doi.org/10.1371/journal.pone.0276006.t002>

Q1 to the Median is marked as the cluster C_2 , the segment from the Median to Q3 is marked as the cluster C_3 , and the segment from Q3 to the Maxima is marked as the cluster C_4 (see Table 2). Thus we have the number of samples partitioned in each cluster set. The students classified in C_1, C_2, C_3 and C_4 are regarded as poor, frugal, normal and affluent, respectively, in the analysis of college canteen consuming behavior. The classification markers are further used to train and evaluate the discrimination models established by BLNN architecture.

4.2. Data extended by fuzzy PCA technique

Of the total of 265,972 samples, 186,180 samples (~70%) are used to train the model and 79,792 samples (~30%) used for model test. The test samples are not involved in the model training process. The training sample data X is used as the raw input data for BLNN training. The raw data is extended to produce one enhancement layer by using the fuzzy transform PCA technique, and then the enhancement neural nodes and the raw input data nodes are flattened as a novel extended input layer for broad network learning.

Concerning the raw data, the network architecture generates 5 neural nodes for receiving the 5-dimension normalized property variables $[yCA, mCA, fCA, 1/mCF, 1/dCF]^T$. The 5-dimension property variables were globally converted into common principal components (PCs) sorted in the order of descending contribution rate. Then the PCs were further carried out for the fuzzy transform strategy, in which we used the triangular membership function. There we tuning the number of PCs from 1 to 5, to determine the best number of PCs for fuzzy optimization. The generated variables acquired by fuzzy transformed PCA was used to train the FCNN model. With adaptively searching the network linkage weights, appreciate discriminant results were obtained based on the 186,180 training samples. There we observed the best FCNN model with aided transformation by the fuzzy-transform PCA technique. The optimal discrimination model was obtained by using 5 PCs, and the confusion matrix of the training samples was showed in Table 3. We have the discrimination accuracy as 89.9% from the confusion matrix. Accordingly, we set 5 extended enhancement neural nodes to build up the BLNN network architecture.

Table 3. The discriminant confusion matrix conducted based on the optimal FCNN classification model trained with the fuzzy-PCA-transformed 5 PCs.

| | | Classification by FCNN model with fuzzy PCA transform | | | | Summation by available marker |
|-----------------------------------|-------|---|---------------|---------------|---------------|-------------------------------|
| | | C_1 | C_2 | C_3 | C_4 | |
| Available marker | C_1 | 30,713 | 1,027 | 859 | 2,094 | 34,693 |
| | C_2 | 1,733 | 47,651 | 1,962 | 1,517 | 52,863 |
| | C_3 | 1,532 | 1,794 | 53,873 | 1,431 | 58,630 |
| | C_4 | 1,274 | 1,596 | 1,977 | 35,147 | 39,994 |
| Summation by classification model | | 35,252 | 52,068 | 58,671 | 40,189 | |

<https://doi.org/10.1371/journal.pone.0276006.t003>

4.3. Data training by BLNN architecture

The BLNN training model is established by using the above-introduced BLNN architecture. We construct the raw input layer as having 5 neural nodes, for correspondence to receive the training sample data X which have 5 different properties. By testing different number of fuzzy PCs in Section 4.2, we finally decide to use 5 different transformed variables as the enhancement neural nodes for broad learning. The newly generated 5 PCs were included in one group denoted as $T_1 = \{PC_1, PC_2, PC_3, PC_4, PC_5\}$. Then the BLNN model is driven by flattening the original 5 input neurons (X) and the extended 5 enhancement neurons (T_1) into a novel broadened layer. The broaden layer is taken as the refreshed input layer for network training. Accordingly, the original data X are extended to be X^{ext} , which have a total of 10 different variables as follows,

$$X^{\text{ext}} = [X, T_1]^T = [yCA, mCA, fCA, 1/mCF, 1/dCF, PC_1, PC_2, PC_3, PC_4, PC_5]^T$$

The refreshed input layer is await for the data input of X^{ext} . Then X^{ext} was delivered to the hidden layer by the algorithmic flow introduced in Section 2.2, generating a series of hidden nodes by activation function mapping of the summation of the input variables. The ReLU function [43] was applied for the hidden layer activation. The linkage weights (w_1) was adaptively trained for iterative optimization, while the number of hidden nodes (m) is preset for tuning in the integer range of $m \in [1, 2 \dots 16] \subset \mathbb{Z}^+$.

Next, the hidden variables are delivered to the output layer by the similar operation of variable summation and activation function mapping, where we used the Sigmoid function [44] for output activation. The linkage weights (w_2) was also set for adaptively iterative training. Afterwards, the output variables were involved in the softmax unit, and we used the adjusted K-means method to deduce the discriminant results. As for parameter tuning, different number of hidden nodes released different BLNN training results. We compared observations of model discrimination accuracy for using a changing value of m (see Fig 7). It is seen from the figure that we observed the best discrimination accuracy as 93.5% when $m = 14$. This indicated that the BLNN structure is best optimized by using 14 hidden nodes. Besides, the network structures built up with the numbers of from 11 to 17 hidden nodes are able to reach nearly appreciating discrimination results (see the green points in Fig 7). Their discrimination accuracies are also over 92%. For the comparison to the FCNN model, the discriminant confusion matrix of the optimal BLNN model was showed in Table 4. Although the BLNN model is able to improve the total accuracy, we can see in details that the better results are noticeable for the C_1 class and the C_2 class, while the C_3 and C_4 classes are not so obvious improved. The reasons are probably instinct to the segments for dividing the minimum-to-maximum range by the Q1, Q3 and the median values. The C_1 and C_2 classes are regarded as the poor and frugal level; It means that the students who consumes less in college canteen are relatively more easy to discriminated by the BLNN model. In contrast, the normal and the affluent level students (i.e. the C_3 and C_4 classes) can be classified to an equal extend by using the BLNN or using the FCNN model.

4.4. Model evaluation

There are 79,792 samples (~30% of the total) reserved before model training, which are excluded and totally independent from the model training process. The test samples were utilized to evaluate the best BLNN training model, as well as the FCNN model for comparison.

The optimized BLNN is built up with 5 original input nodes and 5 pseudo input nodes for broad learning, thus the extended input layer contains 10 neurons, for accepting the raw data variables ($yCA, mCA, fCA, 1/mCF, dCF$) and the generated comprehensive variables ($PC_1,$

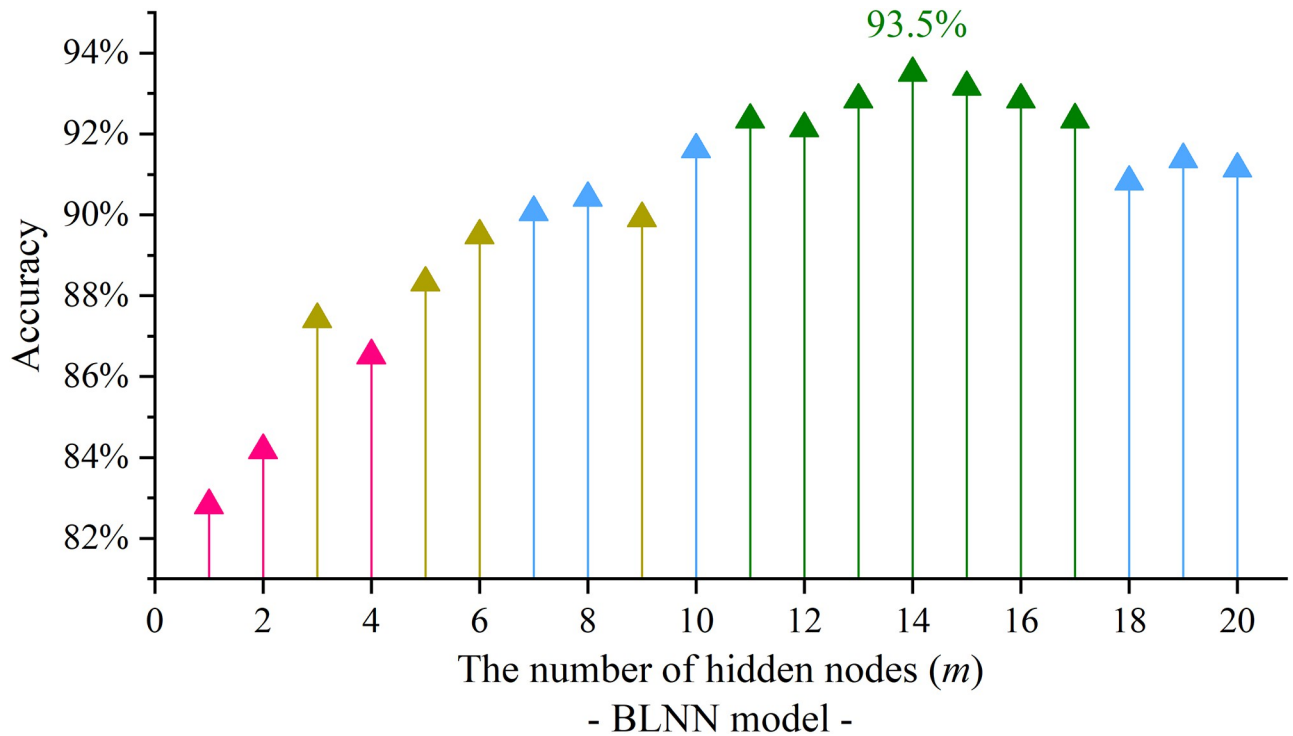


Fig 7. The tuning of the number of hidden neuron nodes for BLNN training.

<https://doi.org/10.1371/journal.pone.0276006.g007>

PC₂, PC₃, PC₄, PC₅) transformed by the fuzzy PCA technique. The optimal network was identified as constructed with 14 hidden nodes by the abovementioned training fulfillment. The linkage weights were automatically tuning by the network adaptive training mechanism. The ReLU function was used as the activation functions for each neural perceptron. The softmax unit was designed to launch the adjusted K-means clustering algorithm for data classification.

The data of the 79,792 test samples were input to test the optimized BLNN model. As the model training process is only to identify the optimal network structure, thus the model evaluation needs to re-train the network linkage weights for any new input. We recorded 20 of the changing classification accuracies during the adaptive iteration on the network linkage weights (showed in Fig 8), and the iterative discrimination results of the FCNN model were also noted in the figure for comparison. Fig 8 shows that the BLNN model observed the best discrimination accuracy of 91.4% for the test set samples. In contrast, the FCNN model was not able to reach up to 90%, its highest data only reached 88.2%.

Table 4. The discriminant confusion matrix conducted based on the adaptively optimized BLNN model built up with 14 hidden neural nodes.

| | | Classification optimized BLNN model with 14 hidden neurons | | | | Summation by available marker |
|-----------------------------------|----------------|--|----------------|----------------|----------------|-------------------------------|
| | | C ₁ | C ₂ | C ₃ | C ₄ | |
| Available marker | C ₁ | 32,174 | 726 | 621 | 1,172 | 34,693 |
| | C ₂ | 1,134 | 49,424 | 1,486 | 819 | 52,863 |
| | C ₃ | 988 | 1,209 | 55,346 | 1,087 | 58,630 |
| | C ₄ | 669 | 951 | 1,229 | 37,145 | 39,994 |
| Summation by classification model | | 34,965 | 52,310 | 58,682 | 40,223 | |

<https://doi.org/10.1371/journal.pone.0276006.t004>

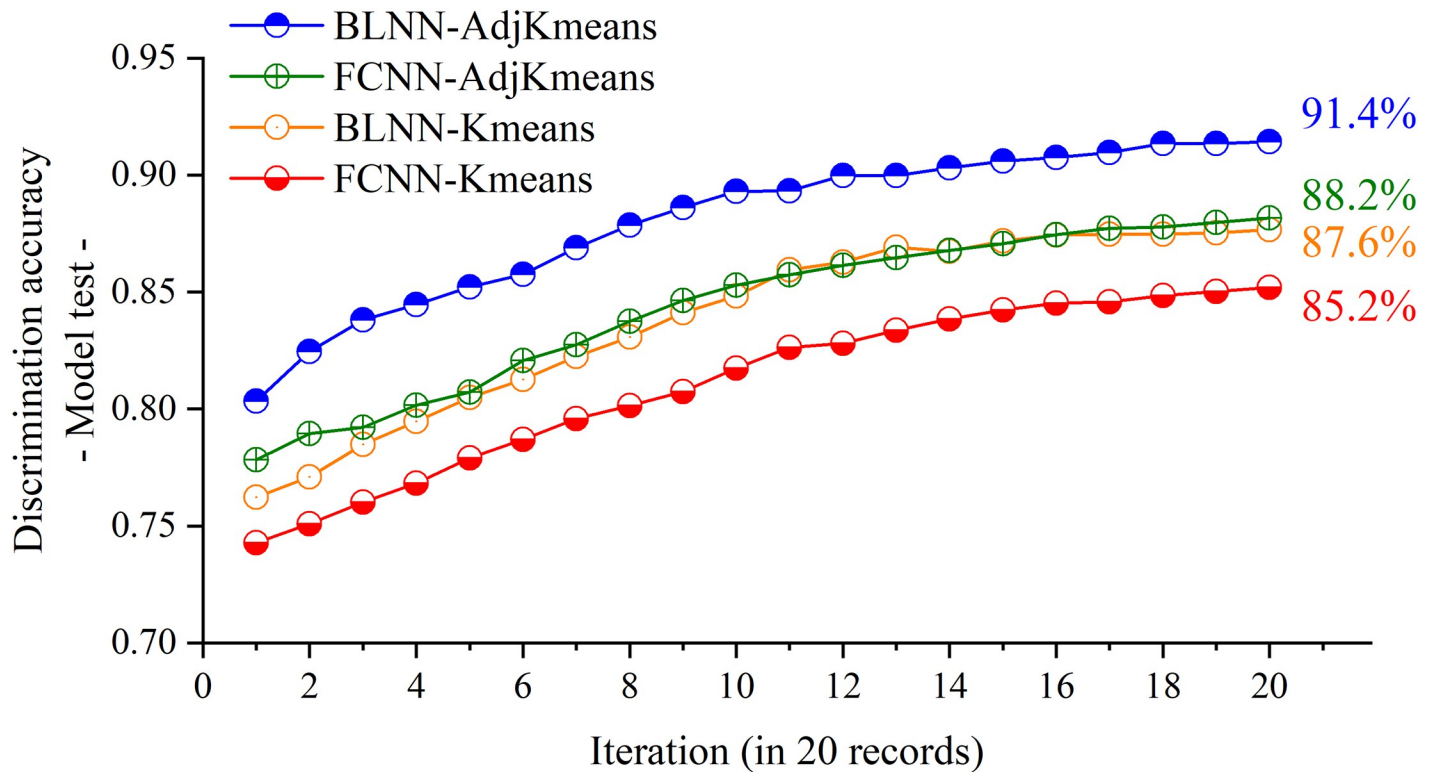


Fig 8. The 20 iterative records of the model improvement for the optimized BLNN-AdjKmeans model (blue) and the FCNN-AdjKmeans model (green), in comparison to the BLNN-Kmeans (yellow) and FCNN-Kmeans models (red). (Note that the label AdjKmeans represents the adjusted K-means method).

<https://doi.org/10.1371/journal.pone.0276006.g008>

As for comparison, the BLNN and FCNN extracted data features were delivered to establish the common K-means classification model. results were also illustrated in Fig 8. The comparative results indicated that the proposed BLNN architecture combined with the adjusted K-means clustering is functional to improve the model discrimination accuracy in data analysis for the classification of college student's canteen consumption levels.

5. Conclusions

A broad learning clustering architecture was built up in this paper for the qualitative analysis of college students' canteen consumption data. The fuzzy transformed PCA technique was applied to find extended data features before the data was input to the network. the adjusted K-means clustering method was fused in the network softmax unit for optimizing the classification model. Confronting the data of Guangdong college students, the BLNN structure is constructed by an extended input layer, a node-number-tuning hidden layer and an adaptive clustering softmax. The extended input layer was formed by a set of pseudo neural nodes for fast tuning. There identified 10 neural nodes, of which the former 5 are to receive the 5 descriptive statistical features of the raw data, and the latter 5 pseudo nodes are available for taking the 5 fuzzy PCs generated by fuzzy PCA transform technique, in which a triangle membership function was applied. Then the extended input data was trained through the hidden layer, and delivered to a softmax unit for establishing the adjusted K-means clustering models, thus to analyze the students' canteen consumption data. The set-up BLNN modeling architecture is applied to classify the college students into 5 different economic level, by analyzing their canteen consumption data.

As to find the most optimal BLNN model for the discrimination, the hidden layer of the BLNN architecture was to tune the number of hidden nodes, and to test the model discrimination accuracy. The most optimal BLNN training model is structured as having 14 hidden nodes, the corresponding best training discrimination accuracy is high up to 93.5%. except for the best model, the BLNN models built up with from 11 to 17 hidden nodes were also able to get appreciating discrimination results over 92%. After that, the optimal training model was evaluated by using the test set samples. By 20 iteration model tuning, we observed the best testing discrimination accuracy as 91.4% by the BLNN model, which is better than the conventional FCNN model.

These modeling and comparing results indicated that the designed BLNN architecture is reasonable to get a more accurate discrimination results for the statistical evaluation of college students' economic level, by analyzing the students' canteen consumption recorded data. The design of BLNN framework in fusion use of fuzzy transformed PCA technique and the adjusted K-means method is an advanced machine learning mode. It is feasible to enhance the model effects for classifying college students into different economic levels, and thus we are able to make more accurate distinguishment of the students who are in need for financial aid. Further, the framework is prospectively expected to be applied in solving some challenging data mining issues in other fields.

Supporting information

S1 Appendix.
(DOCX)

Author Contributions

Conceptualization: Shaoyong Hong.

Data curation: Chun Yang.

Formal analysis: Hongwei Wen.

Funding acquisition: Shaoyong Hong.

Investigation: Darui Jiang.

Methodology: Chun Yang, Hongwei Wen, Shaoyong Hong.

Resources: Shaoyong Hong.

Software: Hongwei Wen, Lijuan Xu.

Validation: Darui Jiang, Lijuan Xu.

Visualization: Lijuan Xu.

Writing – original draft: Chun Yang.

Writing – review & editing: Darui Jiang, Shaoyong Hong.

References

1. Duffy J, Li Y. Lifecycle consumption under different income profiles: Evidence and theory. *Journal of Economic Dynamics and Control*. 2019; 104: 74–94. <https://doi.org/10.1016/j.jedc.2019.05.006>
2. Goh KH, Bockstedt JC. The Framing Effects of Multipart Pricing on Consumer Purchasing Behavior of Customized Information Good Bundles. *Information Systems Research*. 2013; 24: 334–351. <https://doi.org/10.1287/isre.1120.0428>

3. Loschelder DD, Siepelmeyer H, Fischer D, Rubel JA. Dynamic norms drive sustainable consumption: Norm-based nudging helps cafe customers to avoid disposable to-go-cups. *Journal of Economic Psychology*. 2019; 75: 102146. <https://doi.org/10.1016/j.joep.2019.02.002>
4. Fernandez N, Benitez F, Romero-Maltrana D. Social Character of Science and Its Connection to Epistemic Reliability. *Science & Education*, (early access), <https://doi.org/10.1007/s11191-021-00290-3> PMID: 34703080
5. Sun J, Li J, Yang X. Why Consumer's Word is not in Agreement with Their Deed: Study on Factors Impeding Green Consumption Behavior. *Journal of Huazhong Agricultural University (Social Sciences Edition)*. 2015; 119(5): 72–81.
6. Tamburro M, Ripabelli G, Forleo MB, Sammarco ML. Dietary behaviours and awareness of seasonal food among college students in central Italy. *Italian Journal of Food Science*. 2017; 29: 667–680. <https://doi.org/10.14674/IJFS-973>
7. Jung SE, Shin YH, Dougherty R. A Multi Theory–Based Investigation of College Students' Underlying Beliefs About Local Food Consumption. *Journal of Nutrition Education and Behavior*. 2020; 52: 907–917. <https://doi.org/10.1016/j.jneb.2020.07.002> PMID: 32859527
8. Guagliardo V, Lions C, Darmon N, Verger P. Eating at the university canteen. Associations with socio-economic status and healthier self-reported eating habits in France. *Appetite*. 2011; 56: 90–95. <https://doi.org/10.1016/j.appet.2010.11.142> PMID: 21094667
9. Kim H-J, Moon J-W, Han B-R, Choo H-J, Lee K-H. Clothing Knowledge and Clothing Consumption Behavior: Comparison between Clothing Majors and Non-Majors. *International Journal of Costume and Fashion*. 2012; 12: 33–40.
10. Bruce MA, Beech BM, Thorpe RJ, Griffith DM. Racial disparities in sugar-sweetened beverage consumption change efficacy among male first-year college students. *American Journal of Men's Health*. 2016; 10: N168–N175. <https://doi.org/10.1177/1557988315599825> PMID: 26272888
11. Jie Gao, Jae-Hoon Song, Ke Yuan. An Empirical Study on the Mechanism of College Students' Green Consumption in China. *The Journal of Eurasian Studies*. 2019; 16: 211–238. <https://doi.org/10.31203/aeqa.2019.16.1.009>
12. Shen X, Yuan C. A College Student Behavior Analysis and Management Method Based on Machine Learning Technology. *Wireless Communications and Mobile Computing*. 2021; vol.2021:3126347(1–10). <https://doi.org/10.1155/2021/3126347>
13. Appati JK, Nartey PK, Yaokumah W, Abdulai JD. A Systematic Review of Fingerprint Recognition System Development. *International Journal of Software Science and Computational Intelligence (IJSSCI)*, 2022; 14(1), art. 58(1–17). <https://doi.org/10.4018/IJSSCI.300358>
14. Hu B, Gaurav A, Choi C, Almomani A. Evaluation and Comparative Analysis of Semantic Web-Based Strategies for Enhancing Educational System Development. *International Journal on Semantic Web and Information Systems (IJSWIS)*, 2022; 18(1): art. 95(1–14). <https://doi.org/10.4018/IJSWIS.302895>
15. Lee JY, Kang HY, Noh SY. A Case Study of Cloud Computing Service Models for the General Computing Environment in a University. *International Journal of Cloud Applications and Computing (IJCAC)*, 2022; 12(1): art. 96(1–17). <https://doi.org/10.4018/IJCAC.297096>
16. Chen W, Zhang Q, Jin M, Yang J. Research on online consumer behavior and psychology under the background of big data. *Concurrency and Computation: Practice and Experience*. 2019; 31: 1–5. <https://doi.org/10.1002/cpe.4852>
17. Park S, Won D. Analysis of trends in consumption and happiness through the media since the 2000s - Focusing on Big Kinds and Google Trend Analysis. *Review of Culture and Economy*. 2020; 23: 59–84.
18. Shin H. Analysis of Subway Passenger Flow for a Smarter City: Knowledge Extraction from Seoul Metro's "Untraceable" Big Data. *IEEE Access*. 2020; 8: 69296–69310. <https://doi.org/10.1109/ACCESS.2020.2985734>
19. Schwenker F, Trentin E. Pattern classification and clustering: A review of partially supervised learning approaches. *Pattern Recognition Letters*. 2014; 37: 4–14. <https://doi.org/10.1016/j.patrec.2013.10.017>
20. Ubukata S. A unified approach for cluster-wise and general noise rejection approaches for k-means clustering. *PeerJ Computer Science*. 2019; 2019. <https://doi.org/10.7717/peerj-cs.238> PMID: 33816891
21. Sinaga KP, Yang MS. Unsupervised K-means clustering algorithm. *IEEE Access*. 2020; 8: 80716–80727. <https://doi.org/10.1109/ACCESS.2020.2988796>
22. Jia H, Ding S, Xu X, Nie R. The latest research progress on spectral clustering. *Neural Computing and Applications*. 2014; 24: 1477–1486. <https://doi.org/10.1007/s00521-013-1439-2>
23. Leonori S, Martino A, Mascioli FMF, Rizzi A. ANFIS Microgrid Energy Management System Synthesis by Hyperplane Clustering Supported by Neurofuzzy Min-Max Classifier. *IEEE Transactions on*

- Emerging Topics in Computational Intelligence. 2019; 3: 193–204. <https://doi.org/10.1109/TETCI.2018.2880815>
24. Fu L, Lin P, Vasilakos AV, Wang S. An overview of recent multi-view clustering. *Neurocomputing*. 2020; 402: 148–161. <https://doi.org/10.1016/j.neucom.2020.02.104>
 25. Chander S, Vijaya P, Dhyani P. A Parallel Fractional Lion Algorithm for Data Clustering Based on MapReduce Cluster Framework. *International Journal on Semantic Web and Information Systems (IJS-WIS)*, 2022; 18(1): art. 34(1–25). <https://doi.org/10.4018/IJSWIS.297034>
 26. Shen G, Jiang Z. Optimisation of K-means algorithm based on sample density canopy. *International Journal of Ad Hoc and Ubiquitous Computing*. 2021; 38: 62–69. <https://doi.org/10.1504/IJAHUC.2021.119087>
 27. Motallebi H, Nasihatkon R, Jamshidi M. A local mean-based distance measure for spectral clustering. *Pattern Analysis and Applications*. 2022; 25: 351–359. <https://doi.org/10.1007/s10044-021-01040-5>
 28. Sarma TH, Viswanath P, Reddy BE. A hybrid approach to speed-up the k-means clustering method. *International Journal of Machine Learning and Cybernetics*. 2013; 4: 107–117. <https://doi.org/10.1007/s13042-012-0079-7>
 29. Chen CLP, Liu Z. Broad Learning System: An Effective and Efficient Incremental Learning System Without the Need for Deep Architecture. *IEEE Transactions on Neural Networks and Learning Systems*. 2018; 29: 10–24. <https://doi.org/10.1109/TNNLS.2017.2716952> PMID: 28742048
 30. Pratama M, Angelov PP, Lughofer E, Joo Er M. Parsimonious random vector functional link network for data streams. *Information Sciences*. 2018; 430–431: 519–537. <https://doi.org/10.1016/j.ins.2017.11.050>
 31. Jin JW, Philip Chen CL. Regularized robust Broad Learning System for uncertain data modeling. *Neurocomputing*. 2018; 322: 58–69. <https://doi.org/10.1016/j.neucom.2018.09.028>
 32. Bae JS, Oh SK, Pedrycz W, Fu Z. Design of fuzzy radial basis function neural network classifier based on information data preprocessing for recycling black plastic wastes: comparative studies of ATR FT-IR and Raman spectroscopy. *Applied Intelligence*. 2019; 49: 929–949. <https://doi.org/10.1007/s10489-018-1300-5>
 33. Lee E, Schmidt M, Wright J. Improved and simplified inapproximability for k-means. *Information Processing Letters*. 2017; 120: 40–43. <https://doi.org/10.1016/j.ipl.2016.11.009>
 34. Huang W, Peng Y, Ge Y, Kong W. A new Kmeans clustering model and its generalization achieved by joint spectral embedding and rotation. *PeerJ Computer Science*. 2021; 7: 1–22. <https://doi.org/10.7717/peerj-cs.450> PMID: 33834111
 35. Chen H, Liu Z, Gu J, Ai W, Wen J, Cai K. Quantitative analysis of soil nutrition based on FT-NIR spectroscopy integrated with BP neural deep learning. *Analytical Methods*. 2018; 10: 5004–5013. <https://doi.org/10.1039/c8ay01076e>
 36. Ebiaredoh-Mienye SA, Esenogho E, Swart TG. Integrating enhanced sparse autoencoder-based artificial neural network technique and softmax regression for medical diagnosis. *Electronics (Switzerland)*. 2020; 9: 1–13. <https://doi.org/10.3390/electronics9111963>
 37. Ye H, Li H, Philip Chen CL. Adaptive Deep Cascade Broad Learning System and Its Application in Image Denoising. *IEEE Transactions on Cybernetics*. 2021; 51: 4450–4463. <https://doi.org/10.1109/TCYB.2020.2978500> PMID: 32203051
 38. Chang P, Zhang R, Ding C. Dynamic hidden variable fuzzy broad neural network based batch process anomaly detection with incremental learning capabilities. *Expert Systems with Applications*. 2022; 202: 117390. <https://doi.org/10.1016/j.eswa.2022.117390>
 39. Perfilieva I. Fuzzy transforms: Theory and applications. *Fuzzy Sets and Systems*. 2006; 157: 993–1023. <https://doi.org/10.1016/j.fss.2005.11.012>
 40. Pota M, Esposito M, De Pietro G. Transforming probability distributions into membership functions of fuzzy classes: A hypothesis test approach. *Fuzzy Sets and Systems*. 2013; 233: 52–73. <https://doi.org/10.1016/j.fss.2013.03.013>
 41. Chen H, Qiao H, Xu L, Feng Q, Cai K. A Fuzzy Optimization Strategy for the Implementation of RBF LSSVR Model in Vis–NIR Analysis of Pomelo Maturity. *IEEE Transactions on Industrial Informatics*. 2019; 15: 5971–5979. <https://doi.org/10.1109/TII.2019.2933582>
 42. Liu F, Deng Y. Determine the Number of Unknown Targets in Open World Based on Elbow Method. *IEEE Transactions on Fuzzy Systems*. 2021; 29: 986–995. <https://doi.org/10.1109/TFUZZ.2020.2966182>
 43. Klimo M, Tarabek P, Such O, Smiesko J, Skvarek O. Implementation of a Deep ReLU Neuron Network with a Memristive Circuit. *International Journal of Unconventional Computing*. 2016; 12: 319–337.
 44. Costarelli D, Spigler R. Multivariate neural network operators with sigmoidal activation functions. *Neural Networks*. 2013; 48: 72–77. <https://doi.org/10.1016/j.neunet.2013.07.009> PMID: 23973869