Complaint electricity customer clustering method based on electricity big data

Beibei Zhao, Lu Li1, Jiajia Zhang, Jingyu Feng, Lipeng Xia, Jingye Liu and Yongxing Yu
State Grid Fuxin Electric Power Supply Company, Fuxin, China

1 E-mail: 2951231292@qq.com

Abstract. In the fierce domestic market competition, the improvement of service quality is an important goal of electricity companies. This paper proposes the k-means algorithm based on the improved wolf pack algorithm, which is suitable for the power company complaints, there are many factors that affect user complaints. Firstly, the PCA method is used to reduce the dimension of the complaint user's behavior influencing factors, so as to improve the accuracy of clustering. Secondly, due to the shortcomings of falling into local optimal solution and low clustering accuracy, this paper proposes an improved wolf pack algorithm, which proposes the interactive strategy of walking behavior, calling behavior and an adaptive siege strategy for behavior, the clustering accuracy and convergence speed are improved. This algorithm overcomes the shortcomings of the original algorithm. Moreover, it improves the efficiency and clustering quality of the algorithm and it can realize the accurate classification of users.

1. Introduction
In the text, as an important bargaining chip in the decisive market, people have been seeking an efficient and accurate algorithm to classify Complaint users to prevent complaints from electricity users in a timely manner, and to improve the service quality of electricity users [1, 2]. Literature [3] proposes to construct an index system through clustering analysis method and establishing a grouping model. Then divide the customer groups into multi-level and two-dimensional groups and propose a customer grouping system. In the context of big data, it is particularly prominent for the status of unsupervised clustering algorithm. In recent years, the research on clustering algorithms has made better progress.

Literatures [4] and [5] proposed multiple clustering methods such as partition clustering, hierarchical clustering, artificial neural network clustering, nuclear clustering, sequence data clustering, complex network clustering, and intelligent search clustering. Nuclear clustering requires a higher definition of the radius of the high-dimensional spherical region, and different choices will have different heights. Density clustering has higher requirements for clustering density. When the clustering density is different, the performance will be inferior to other algorithms. And the calculation amount of hierarchical clustering is relatively large compared with other algorithms. The k-means algorithm is relatively suitable for the problem of dividing the risk level of electricity users. However, k-means algorithm has the shortcomings of low clustering accuracy and it is easy to fall into local optimal solutions. In response to this problem, Literature [6] proposed an efficient method for power user classification that combines reduction and clustering on the basis analyzing characteristics of power big data. The Spark platform is used to achieve the integration of reduction and clustering,
which has a greater improvement in accuracy and running time compared with traditional methods. Literature [7] introduces the current local optimal solution into the improved artificial bee colony (IABC) algorithm, which improves the minimizing ability of the algorithm; In Literatures [8] and [9], chaotic search process is added to particle swarm algorithm and the improved k-means clustering algorithm is proposed. The result shows that this algorithm can achieve better clustering results.

Literature [10] proposed the improved wolf pack algorithm and its application fields; The chaotic optimization idea in the Literature [11] constructs the calculation framework of the improved wolf pack algorithm, and compares it with the traditional WPA method to verify the proposed algorithm. Based on the respective advantages of the wolf pack algorithm and the k-means algorithm, as well as the proposed interactive strategies for wandering behavior and calling behavior, and adaptive siege strategy, this paper proposes an improved k-means algorithm on the improved wolf pack algorithm to enhance the convergence speed. And then use it to optimize the position of the cluster center in the k-means algorithm to eliminate the influence of the initial cluster center and the possibility of being classified as the local optimal solution.

2. Electricity big data preprocessing

2.1. Data acquisition and processing
The goal of this paper is to classify the risk of complaint power customers, extract the characteristics of electricity consumption of different users, that is, through the users of information data, according to the personal information of various users and the characteristics of electricity consumption in different time periods, and to identify different electricity consumption characteristics of customers are used to classify customer groups with complaint risk levels.

This paper conducts missing value analysis and outlier analysis on user electricity consumption information data, and analyzes the law and outlier value of power user complaint risk characteristic data. Through the observation of the data, it is found that the dimensions of the collected data are too large and difficult to analyze. It is necessary to extract the characteristics of user complaints and extract some key indicators reflecting the nature data to achieve the purpose of clustering and improve the accuracy of clustering.

The classification of power users is mainly to classify the typical complaint characteristics of power users, rather than simply using numerical values for classification, so normalization is necessary. In this paper, the maximum normalization is used to normalize the data, and the value is normalized to the interval [0,1].

The characteristics of user's complaint are affected differently by the power consumption data. The scope of application of each information data index is not fixed, and has their own emphasis. The choice of information data has a great impact on the characteristics of user complaints.

The correlation degree can reflect the degree of correlation. Its value is between -1 and 1. The closer it is to 1, the stronger the positive correlation, the closer it is to -1, the stronger the negative correlation, and the closer it is to 0, the lower the correlation. Correlation analysis of age, address, power consumption category and power consumption data of power user complaint information is carried out, and the correlation coefficient Ra is obtained, which represents the relationship between the complaint risk and various influencing factors.

First calculate the covariance of different electricity usage information data:

\[
\text{cov}(X,Y) = E \left( (X - \mu_X)(Y - \mu_Y) \right)
\]  

(1)

The variance of electricity usage information is:

\[
\sigma_X = \sqrt{E(X^2) - E^2(X)}
\]  

(2)

\[
\sigma_Y = \sqrt{E(Y^2) - E^2(Y)}
\]  

(3)
The correlation coefficient is used to express the attribute correlation between them, and the correlation coefficient \( R_a \) can be obtained:

\[
R_a = \frac{\text{cov}(X,Y)}{\sigma_x \sigma_y} = E\left(\left(X - \mu_x\right)\left(Y - \mu_y\right)\right) / \sigma_x \sigma_y
\]  
(4)

\( R_a \) is used to represent the corresponding degree of correlation of each electricity information data. The greater the correlation coefficient \( R_a \), the greater the influence factor on the risk of complaints.

For data with a missing rate of 30\%, a simple deletion process is performed; for data with a missing rate of less than or equal to 30\%, because user information data has certain trend characteristics, the imputation method can make the imputed value more accurate, so this article The interpolation method is used to fill in the existing data. First, obtain the polynomial function \( L(x) \) according to the existing data. The polynomial function \( L(x) \) is calculated as follows:

Calculate the degree polynomial of the known points:

\[ y = a_0 + a_1 x + a_2 x + \cdots + a_{n-1} x^{n-1} \]  
(5)

The Lagrangian interpolation polynomial can be obtained as:

\[
L(x) = y_1 \frac{(x-x_2)(x-x_3)\cdots(x-x_n)}{(x_1-x_2)(x_1-x_3)\cdots(x_1-x_n)} + y_2 \frac{(x-x_1)(x-x_3)\cdots(x-x_n)}{(x_2-x_1)(x_2-x_3)\cdots(x_2-x_n)} + \cdots + y_n \frac{(x-x_1)(x-x_2)\cdots(x-x_{n-1})}{(x_n-x_1)(x_n-x_2)\cdots(x_n-x_{n-1})}
\]  
(6)

Put the point corresponding to the missing value into the interpolation polynomial to get the approximate value \( L(x) \) of the missing value, and the data is complete.

2.2. Power consumption big data reduction

The main characteristics of power user complaint risk clustering include the type of user complaints, number of complaints, user age, user address, annual electricity consumption, monthly electricity consumption, and information such as whether the content of the complaint acceptance is reasonable, and user satisfaction with return visits. And for the final clustering results, these influencing factors are superimposed, and it is important to reduce these features for clustering. The principal component analysis method has the advantages of retaining the main information of the original data and eliminating the mutual influence of the original data. It can effectively decrease the quantity of data calculation and improve the accuracy of clustering.

The principal component analysis method is used for reduction processing, and multiple mutual influences are reduced. Non-independent data sets are transformed into mutually independent data sets. These mutually independent data sets are the principal components of the original mutually independent data sets. The original data is linearly combined to obtain the principal components, and the principal components are independent of each other, which can ensure that the data characteristics of the original data are retained in the principal components, and its independence is guaranteed.

Observe the original data, and the original data matrix is shown in formula:

\[
X = \begin{bmatrix}
x_{11} & x_{12} & \cdots & x_{1m} \\
x_{21} & x_{22} & \cdots & x_{2m} \\
\vdots & \vdots & \ddots & \vdots \\
x_{n1} & x_{n2} & \cdots & x_{nm}
\end{bmatrix} = (x_{1}, x_{2}, \cdots, x_{m})
\]  
(7)

Among them, \( x_{1}, x_{2}, \cdots, x_{m} \) represent the complaint characteristics of the i-th user, and \( x_{i} = \{x_{i1}, x_{i2}, \cdots, x_{im}\} \) represents the user's complaint event type, the number of complaints, and the sample size of the monthly electricity consumption.

In view of the difference in the maximum value of the data in each user information, the user electricity consumption information data is standardized to obtain a data matrix for cluster analysis.

The PCA can ensure that the data characteristics of the original data are retained in the principal components, and its independence is guaranteed. It can improve the accuracy of clustering [12].
The average value of the sample matrix is:

\[ \bar{x}_j = \frac{1}{n} \sum_{i=1}^{n} x_{ij} \]  

(8)

\( \bar{x}_j \) is the average value of each column.

The variance of the sample matrix **X** is:

\[ S_j^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_{ij} - \bar{x}_j)^2 \]  

(9)

\( S_j \) is the variance of each column in the sample matrix.

The formula for data standardization is as follows:

\[ x_{ij}^* = \frac{(x_{ij} - \bar{x}_j)}{S_j} \]  

(10)

For an orthogonal matrix **U** such as \( U^TYU = \Lambda \), \( \Lambda = (\lambda_1, \lambda_2, \ldots, \lambda_m) \) and \( \lambda_1 > \lambda_2 > \ldots > \lambda_m \), \( \lambda_1, \lambda_2, \ldots, \lambda_m \) is the characteristic value, \( \alpha_1, \alpha_2, \ldots, \alpha_m \) is a feature vector corresponding to \( \lambda_1, \lambda_2, \ldots, \lambda_m \) among them.

The variance contribution rate of the i-th principal component is:

\[ \alpha_i = \frac{\lambda_i}{\sum_{j=1}^{m} \lambda_j} \]  

(11)

\( \alpha_i \) is the variance contribution rate of the i-th principal component. The greater the value of the variance contribution rate \( \alpha_i \) of the principal component, the stronger the correlation with the sample.

And the influencing factors of the principal components are arranged in descending order according to the size of the characteristic root.

The cumulative variance contribution rate of the first i-th principal component components is:

\[ \beta_i = \frac{\sum_{j=1}^{i} \lambda_j}{\sum_{j=1}^{m} \lambda_j} \]  

(12)

\( \beta_i \) is the cumulative variance contribution rate of the first i-th principal component, and the number of principal components depends on the cumulative variance contribution rate \( \beta_i \).

In this paper, the PCA method is used to extract several principal components of the complaint user information, and the improved k-means algorithm of the wolf pack clusters based on the results realizes the effective classification of the complaint user level. This method not only reduces the dimensionality of the clustering, but also retains the important information in the original variables and improves the accuracy of the clustering.

3. **K-means algorithm based on improved WPA**

Electricity consumption information data has the characteristics of large data volume and many types of data. K-means algorithm is an unsupervised learning algorithm, which has the advantages of simple, fast and effective processing of large data sets. It is suitable for classification (category) data and can realize fast and efficient classification of power user complaints.

The traditional K-means clustering algorithm has obvious limitations in clustering analysis of power users. K-means clustering algorithm has the problems of being sensitive to the initial value and low clustering accuracy. The wolf group algorithm is used to optimize K-means clustering because the wolf group algorithm is not interfered by the initial value and has the advantages of fast iteration speed. These advantages ensure that it can be obtained in the entire search space to make the clustering objective function small cluster centers as possible. The hybrid clustering algorithm that combines the wolf group algorithm and K-means algorithm uses the strong global search ability of the wolf group algorithm to find the center points that make the value of the clustering objective function minimize, which weakens the original K-means to a certain extent. means clustering over-reliance on the initial center point. However, there is a problem that the wolf pack algorithm is not easy to break away the
local solution, so that the cluster center point may not be optimal. In the three search stages of the implementation of the wolf pack algorithm, there are problems such as slow convergence, easy to fall into the local optimum, and unsatisfactory interaction of artificial wolves [13]. Therefore, this paper proposes the improved wolf pack algorithm based on the search strategy.

3.1. Interactive wandering behavior
In the WPA algorithm, detective wolf explores in n directions, the larger the value of n is, the higher the optimization accuracy is, but the optimization speed of the algorithm will decrease, and the optimal clustering center point will easily fall into the local optimum. If r the value of n is too small, the cluster center point will be inaccurate, and it may even be impossible to find the cluster center point. The reason for the above situation is that it lacks information interaction between the wolf hunters and the inability to know the information of the companions, which affects the global search ability of the wolf hunters. In order to increase the interaction between wolves and improve the ability to find the best.

\[ y_{i,d} = x_{i,d} + \alpha_{i,d}(x_{\text{best}} - x_{i,d}) + \beta_{i,d}(x_{i,d} - x_{k,d}) \] (13)

Among them: \( \alpha_{i,d} \) is the random number of [0, 1], \( \beta_{i,d} \) is the random number of [-1, 1], k≠i≠j. The first half of the formula enhances the local optimization ability of the wolf pack, and the second half enhances the global search ability of the wolf pack, which well balances the global search ability of the wolf pack and the local optimization ability, so that the center point of the cluster is sought, more accurate.

3.2. Interactive calling behavior
The basic summoning behavior can make the wolf fully explore the search space, but it will cause the algorithm to be too complicated and easy to fall into the local optimal cluster center point. Therefore, this article adopts the summoning strategy of "rounding up" the sought cluster center point in one raid of the wolf. In the group algorithm, the communication between groups is an important part of the algorithm. Choose a better cluster center point \( Y_i \) to move forward, update the wolf position \( X_i \), and select the wolf with the best cluster center point position as the head wolf.

3.3. Adaptive siege behavior
The siege behavior requires the wolf to have a strong local optimization ability, but the wolf has randomness and uncertainty. Adding an adjustment mechanism to the algorithm is a better direction for improvement, in order to make the siege behavior have an adaptive adjustment ability, so that the cluster center points sought are more accurate. In this paper, the random step size \( \lambda \) is changed to an adaptive step size formula that linearly changes with the increase in the number of algorithm iterations.

\[ x_{id}^{k+1} = x_{id}^k + w(1 - \theta t/t_{\text{max}}) \times \text{step}^d \times [\alpha_{id}^k - x_{id}^k] \] (14)

Among them: \( \theta \) is the factor, taken as the random number within (0, 1), w is the random integer, taken as the random number within (-1, 1). The purpose of the value of \( \theta \) within (0, 1) is to ensure \( w(1 - \theta t/t_{\text{max}}) \) that it avoids approaching zero at the later stage of the algorithm iteration, resulting in no change in the cluster center point. The function of \( w \) is to ensure that the search range is not limited to the direction, and can search the area near \( x_{id} \) more comprehensively.

The algorithm analyzes the principal component results of power complaint customers, adopts the interactive strategy of wandering behavior and calling behavior, and adaptive siege strategy, and uses the improved wolf pack algorithm to search and optimize the clustering center globally. Compared with traditional k-means algorithm, this method optimizes the location of the cluster centers, thereby quickly achieving global convergence. It improves the efficiency of the algorithm and makes the clustering results more accurate. Figure 1 shows the algorithm flow chart.
4. The improved algorithm implementation process
The implementation process is as follows:

Step 1: Firstly, adopt the principal component analysis method to reduce the information data, and filter out important information.

Step 2: Initialize the wolf pack, set the artificial wolf position $X_i$, the number of iterations $k$, the scale factor $\alpha$, the number of walks $T_{\text{max}}$, and the number of clusters $N$, compute the fitness function of the wolf pack and choose the optimal solution $X_{\text{best}}$, except for the head The best artificial wolf outside the wolf is the wolf detection.

Step 3: Perform interactive walking behavior until the prey odor concentration $Y_i$ detected by a certain wolf is better than the prey odor concentration $Y_{\text{lead}}$ perceived by the wolf or reaches the maximum number of wandering $T_{\text{max}}$.

Step 4: The wolf rushes towards the prey according to the interactive summoning behavior. If the perceived prey odor concentration $Y_i > Y_{\text{lead}}$ on the way, then $Y_{\text{lead}} = Y_i$, replacing the wolf to start the summoning action.

Step 5: Update the location of the wolf and execute the siege behavior.
Step 6: Update the position of the head wolf according to the "winner is king" rule of head wolf generation, and then update the group according to the wolf pack update mechanism of "survival of the strong", and calculate the new cluster center according to the latest position of improved wolf pack optimization.

Step 7: Stop when the end condition is reached, otherwise, return to step 3.

5. Improved clustering algorithm application results
This paper selects 1000 complaining customers’ data in Northeast China from October 2020 to October 2021, including the total electricity consumption, number of complaints, voltage level, electricity consumption variance, house price and user’s age. This paper uses SPSS data analysis tool to process the data and the PCA to reduce the influencing factors.

Table 1. KMO and bartlett test.

| Ingredient | Total | Percentage of variance | Accumulation% |
|------------|-------|------------------------|---------------|
| X1         | 2.747 | 45.781                 | 45.781        |
| X2         | 1.122 | 18.705                 | 64.486        |
| X3         | 0.848 | 14.139                 | 78.625        |
| X4         | 0.716 | 11.929                 | 90.554        |
| X5         | 0.545 | 9.076                  | 99.630        |
| X6         | 0.022 | 0.370                  | 100.000       |

Results in Table 1 show that test result of Bartlett's spherical test is 7605.724, and the KMO sampling suitability is 0.646 greater than the critical value 0.5. Therefore, the power complaint data is suitable for statistical analysis using factor analysis.

Table 2. Total variance interpretation.

| Ingredient | Initial | Concentrated extraction |
|------------|---------|-------------------------|
| X1         | 1.000   | 0.450                   |
| X2         | 1.000   | 0.453                   |
| X3         | 1.000   | 0.814                   |
| X4         | 1.000   | 0.572                   |
| X5         | 1.000   | 0.828                   |
| X6         | 1.000   | 0.652                   |

Results in Table 2 show that the number of extracted factors is 6, the variance contribution rate is 100.0%, there is 100.0% of the index information interpretation ability. The purpose of factor analysis is to condense information. Each influencing factor has a variance explanation rate, and the variance explanation rate represents the degree of information extraction of the factor for power complaint data.

Table 3. Commonness of original.

| Ingredient | Initial | Concentrated extraction |
|------------|---------|-------------------------|
| X1         | 1.000   | 0.450                   |
| X2         | 1.000   | 0.453                   |
| X3         | 1.000   | 0.814                   |
| X4         | 1.000   | 0.572                   |
| X5         | 1.000   | 0.828                   |
| X6         | 1.000   | 0.652                   |

Results in Table 3 show that the lowest information retention rate among the original variables is 45.0% and the remaining indicators retain more than 45% of the information. It indicates that the all indicators information has been maintained.
The gravel diagram displays the characteristic roots in graphical form, and is mainly used to assist in judging the number of factors. It can be determined how many factors are extracted to be applicable to the power user complaint feature set. Figure 2 shows it is more reasonable to extract 6 factors of power user complaint features. Therefore, this paper selects 6 characteristic factors for cluster analysis of power user complaints, and use principal component analysis to reduce the dimensionality of power user data.

**Figure 2.** Gravel diagram.

![Gravel diagram](image)

Results in Table 4 imply that the PCA method is used to reduce the dimensionality of six-dimensional data to two-dimensional data, and the principal component coefficients of F1 and F2 can be obtained., and then the K-means clustering algorithm of improved wolf pack optimization is used for clustering analysis.

This paper is based on matlab software to realize the implementation of the K-means clustering algorithm on improved wolf pack optimization. The algorithm settings are as follows: the cluster number k=5, the wolf number `wolfnum`=50, maximum number of iterations `maxgen`=300, number of wolves `alfa`=3, the maximum number of wanderings `T_max`=30, the number of wolf detection `beta`=10, and the maximum search direction `h_max`=15.

**Table 4.** Component matrix.

| Ingredient | Ingredient 1 | Ingredient 2 |
|------------|--------------|--------------|
| X1         | 0.551        | -0.383       |
| X2         | 0.474        | -0.479       |
| X3         | 0.952        | 0.092        |
| X4         | 0.671        | 0.350        |
| X5         | 0.809        | -0.041       |
| X6         | 0.194        | 0.783        |
Figure 3. Clustering results distribution map.

Figure 3 shows the power complaint users are divided into five categories. The improved wolf pack algorithm is used to search and optimize the clustering center globally. Compared with traditional k-means clustering algorithm, the improved algorithm can optimize the location of the cluster centers, thereby quickly achieving global convergence. Therefore, it can be considered that the improved wolf pack algorithm is used to optimize the clustering. The center can get the best clustering results.

Table 5. Scoring result.

| Category | 1    | 2    | 3    | 4    | 5    |
|----------|------|------|------|------|------|
| Rating result | 0.0854 | 0.1521 | 0.3211 | 0.5212 | 0.8156 |

Results in Table 5 imply that power complaint users can be divided into 5 levels. Category 1 is the high-risk power complaint users, Category 2 is the low-risk power complaint users, Category 3 is the general risk power complaint users, Category 4 is the good users, and Category 5 is the better users. Therefore, customers are divided into 5 different levels, and an early warning push plan is proposed according to the possibility of user complaints.

6. Conclusions
This paper proposes an improved K-means clustering algorithm on the wolf pack optimization, and uses the PCA method to reduce the dimensionality of the power complaint user data. By obtaining complaint data in the northeast of Liaoning Province, the proposed improved wolf pack optimization K-means clustering algorithm has been experimentally verified. Experiments show that the improved wolf pack optimization K-means clustering algorithm has higher accuracy in the classification of power user complaints, it adopts the interactive strategy of wandering behavior and calling behavior, and adaptive siege strategy, and uses the improved wolf pack algorithm to search and optimize the clustering center globally. It has a good application prospect in power user complaint service marketing and proposing an early warning mechanism based on the risk level of power complaints will be the next research content.

Acknowledgement
This paper was partially supported by Science and Technology Project of State Grid Corporation of China (2021YF-29).
References

[1] Chen Huajun 2015 Big data revolution-exploration of the application of big data in China Southern Power Grid China Electric Power Enterprise Management 17

[2] Olsen R L, Chappell R W and Loftis J C 2020 Water quality sample collection, the data treatment and results presentation for principal components analysis—literature review and Illinois River Watershed case study Water Research 46 9

[3] Paul H Whitfield, Escalante-Sandoval C and L Amores-Rovelo 2017 Regional monthly runoff forecast in Southern Canada using ANN, K-means, and L-moments techniques Canadian Water Resources Journal 42 3

[4] Xiaoli Cui, Pingfei Zhu, Xin Yang, Keqiu Li and Changqing Ji 2014 Optimized big data K-means clustering using MapReduce The Journal of Supercomputing 70 3

[5] Faisal Saeed, Naomie Salim, Ammar Abd and Hentabli Hamza 2013 Using graph-based consensus clustering for combining K-means clustering of heterogeneous chemical structures Cheminformatic 5 1

[6] Ping Li 2013 Study on the Application of K-Means Algorithm Implemented Hadoop Platform to the Library Work in Colleges and Universities International Journal of Technology Management 8

[7] Xiong Hui, Wu Junjie and Chen Jian 2009 K-Means Clustering Versus Validation Measures: A Data-Distribution Perspective IEEE Transactions on Systems 39 2

[8] E W Tyree and J A Long 2008 A Monte Carlo evaluation of the moving method, k-means and two self-organising neural networks Pattern Anal. Apply 2 1

[9] Pierre Hansen, Eric Ngai, Bernard K Cheung and Nenad Mladenovic 2005 Analysis of Global k-Means, an Incremental Heuristic for Minimum Sum-of-Squares Clustering Journal of Classification 22 2

[10] Piyush Kumar Sharma and Gary Holness 2019 Clustering Student's Satisfaction in Complex Adaptive Blended Learning with the Six Value System Using the K-Means Algorithm Universal Journal of Educational Research 9 7

[11] Skyler Thompson, M Emre Celebi and Krizia H Buck 2019 Fast color quantization using MacQueen’s k-means algorithm Journal of Real-Time Image Processing 17 1609-1624

[12] Qiao Wenyu, Li Ye, Liu Haoyu, Lin Yang and Yang Ting 2020 Missing electricity data completion method based on curve similarity and low-rank matrix Electric Power Construction 23 02

[13] Ting Wang, Hailin Tang and Yuebao Yu 2014 Array Antenna Pattern Synthesis Based on Selective Levy Flight Culture Wolf Pack Algorithm Journal of Harbin Institute of Technology 27 5