A study of marking utility models based on principal component analysis and integer programming

Nanning Chen [1,*], Zikai Dong [2], Huiyuan Yang [2], Fuchun Li [3], Xinyu Ma [4]

[1] College of Agriculture, Guangxi University, Nanning 530004, China

[2] College of China-ASEAN Financial Cooperation, Guangxi University, Nanning 530004, China

[3] College of Civil Architecture and Engineering, Guangxi University, Nanning 530004, China

[4] College of Light Industry and Food Engineering, Guangxi University, Nanning 530004, China

* Corresponding Author Email: 1283616986@qq.com

Abstract. The establishment and analysis of a marking utility model is of great importance in achieving paperless marking and improving the accuracy and efficiency of marking. Through the application of modern technology, students' examination results can be evaluated more objectively and provide an important basis for determining the final effect of talent training. This paper provides a systematic analysis and modelling of the marking process, and uses a variety of techniques and methods to address the issue of marking staff arrangements. Firstly, the impact of the proportion of markers per question, the proportion of the average mark for each question to the total mark for that question, the proportion of blank papers and the error rate on the speed of marking was explored by analysing marker-related indicator variables and the basic requirements for completing the marking process. Then, principal component analysis was used to reduce the dimensionality of the four factor variables affecting marking speed, and two principal components with strong correlation with marking speed were obtained to provide a more systematic and comprehensive analysis of the influence of marking speed. Finally, an integer programming model and simulated annealing algorithm were used to solve for the optimal arrangement of markers to ensure that the requirement of simultaneous completion of workload for each question was met. In this paper, the optimal allocation of markers is analysed and the time required for all markers to complete the marking process at the same point in time is calculated to be 3.3 days. Two optimal allocations of markers are proposed for each question type per day and for each question type per hour in a day.

Keywords: Principal component analysis, integer programming, simulated annealing algorithms, non-linear fitting, characteristic indicators.

1. Introduction

The development and analysis of a marking utility model is of great importance for the implementation of paperless marking. Paperless marking has become a popular trend, but the speed of marking is often influenced by a number of indicators, such as the marking criteria of different markers, changes in the number of people marking a single question, the difficulty of the question and so on. A marking speed model can be developed to determine which factors play an important role in marking speed by analysing the degree of influence of several relevant variables, so that marking speed and staffing arrangements can be predicted and optimised to maximise the efficiency of the marking process.

Modern technology techniques such as machine learning and big data analytics can help build more objective and accurate models of marking speed. By using supervised learning algorithms to train the models and using large amounts of historical marking data as input, it is possible to predict the marking results of different markers for different questions. In this way, the impact of various factors on marking speed can be understood, while improving the accuracy and efficiency of marking. Data mining and visualisation techniques can also be used to analyse the marking speed and results of different questions to determine which questions require more markers to be involved and how to allocate marking tasks appropriately to improve the efficiency of the overall marking process.

Professor Zhang Wei's team [1] at the Chinese Academy of Sciences has carried out a project entitled 'Deep Learning-based Automatic Scoring Model'. Using deep learning techniques and natural language processing methods, they designed an algorithmic model capable of automatic grading. The model has been validated on multiple datasets and has achieved high accuracy rates, providing technical support for paperless marking. Professor Singh, R and his team [2] at Stanford University have developed an online grading system called 'GradeScope'. GradeScope is already being used in many universities and educational institutions. Professor Arvind Ramanathan and his team [3] at the Indian Institute of Technology have worked on a paperless marking method based on semantic similarity. They used natural language processing techniques and similarity calculation methods to automatically mark student scripts. This method ensures fairness and accuracy of marking while reducing the time spent on manual marking.Dr Hannah Rohde[4] from the Max Planck Society in Germany is focusing on the problem of marking long answers in intelligent marking. Using machine learning and natural language processing techniques, she has developed an algorithm that can quickly mark long answers. This approach improves the speed of marking while maintaining high marking quality. Professor Koichi Yamamoto and his team [5] at the University of Tokyo, Japan, investigated intelligent marking methods based on knowledge graphs. By constructing a subject knowledge map, they analysed the knowledge of students in their scripts to achieve fast and accurate marking. This method has been applied in practice in several schools in Japan.

2. Analysis of the combined factors associated with the speed of marking

2.1. Analysis of the effect of principal component analysis and non-linear fitting on marking speed

There are four variables in the marking speed model: number of markers x_1, mean score x_2, number of blank papers x_3, and error rate x_4. These four variables affect marking speed to varying degrees and may be correlated with each other, and there is a degree of overlap between the four variables in terms of the information they reflect on the impact on marking speed. In order to find the correlation of the 4 variables, this paper finds their Pearson correlation coefficients based on the data in Appendix 1 as shown in Table 1 below.

$$r = \frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{X_i - \bar{X}}{S_X}\right)\left(\frac{Y_i - \bar{Y}}{S_Y}\right) \tag{1}$$

Table 1. Results of correlation analysis of marking speed impact indicators

Indicators	Number of marked papers	Number of marked papers	Number of blank volumes	Number of blank volumes
Number of marked papers	1			
Number of marked papers	0.335	1		
Number of blank volumes	-0.936	-0.363	1	
Number of blank volumes	-0.193	0.321	0.449	1

As can be seen from Table 1 above, the absolute values of the correlation coefficients of x_1 and x_3 are close to 1, and there are also certain correlations between other variables. In order to reduce the number of variables in the analysis of the influencing factors of marking speed and to get more information, and to analyse the influencing factors of marking speed more comprehensively and

systematically, we choose to use principal component analysis to reduce the dimensionality of the four variables x_1, x_2, x_3, x_4 and get two comprehensive indicators, i.e. principal components F_1 and F_2, the two principal components are linear combinations of the four original variables x_1, x_2, x_3, x_4, and the functional relationship between them is obtained by fitting the data to the quadratic function equation,i.e.

$$y = aF_1 + bF_2^2 + c \tag{2}$$

The principal component retains most of the information of the four original variables such as number of markers, mean score, blank papers, error rate and etc. Through the correlation of the original variables, two relevant objects are found to replace the original variables to facilitate a more comprehensive analysis of the impact analysis of marking speed, and then construct a marking speed model.

Firstly, the four indicator variables were subjected to principal component analysis. According to the question there are 5 research objects A、B、C、D、E,4 indicator variables x_1, x_2, x_3, x_4, the $i-th$ object $j-th$ indicator takes the value of a_{ij}, then constitute the data matrix $A=(a_{ij})_{5\times4}$. The original four indicators are standardized to obtain the standardized indicator variables

$$y_j = \frac{x_j - \mu_j}{s_j},\ j = 1,2,3,4 \tag{3}$$

$$\mu_j = \frac{1}{n}\sum_{i=1}^{n} a_{i,j} \tag{4}$$

$$s_j = \sqrt{\frac{\sum_{i=1}^{n}(a_{ij} - u_j)^2}{n-1}} \tag{5}$$

The corresponding normalised data matrix $B=(b_{ij})_{5\times4}$ is obtained, where

$$b_{ij} = \frac{a_{ij} - \mu_j}{s_j}, i = 1,2,3,4,5,\ j = 1,2,3,4 \tag{6}$$

The correlation coefficient matrix $R=(r_{ij})_{4\times4}$,is derived from the normalized data matrix B, there

$$r_{ij} = \frac{\sum_{k=1}^{n} b_{ki}b_{kj}}{n-1}, i, j = 1,2,3,4,5. \tag{7}$$

The four eigenvalues of the correlation coefficient matrix R satisfying λ_1, λ_2, λ_3, λ_4 were calculated by Python as $[1.3376\times10^{-1}, 3.7145\times10^{-2}, 6.3061\times10^{-5}, 8.2243\times10^{-3}]$

And their corresponding standard orthogonalized eigenvectors are respectively

$$\begin{cases} u_1 = [0.03846, 0.24335, 0.74197, -0.62352] \\ u_2 = [-0.20359, 0.88036, 0.07607, 0.42156] \\ u_3 = [-0.12188, -0.38697, 0.66243, 0.62973] \\ u_4 = [-0.97067, -0.12642, -0.06974, -0.19219] \end{cases}$$

Of which $u_j = [u_{1j}, u_{2j}, u_{3j}, u_{4j}]^T$,The four new indicator variables comprising the eigenvectors are:

$$\begin{cases} F_1 = u_{11}x_1 + u_{21}x_2 + u_{31}x_3 + u_{41}x_4, \\ F_2 = u_{12}x_1 + u_{22}x_2 + u_{32}x_3 + u_{42}x_4, \\ F_3 = u_{13}x_1 + u_{23}x_2 + u_{33}x_3 + u_{43}x_4, \\ F_4 = u_{14}x_1 + u_{24}y_2 + u_{34}x_3 + u_{44}x_4, \end{cases} \tag{8}$$

Where F_1 is the 1st principal component, F_2 is the 2nd principal component, F_3 is the 3rd principal component and F_4 is the 4th principal component.

The contribution of the principal component and the cumulative contribution were calculated, and the contribution of the principal component F_j was

$$\omega_j = \frac{\lambda_j}{\sum_{k=1}^{m} \lambda_k}, j = 1,2,3,4 \tag{9}$$

The resulting data for the four principal component contributions of F1,F2,F3,F4 were obtained as $(7.46462 \times 10^{-1}, 2.07290 \times 10^{-1}, 3.51914 \times 10^{-4}, 4.58961 \times 10^{-2})$

The cumulative contribution of the first j principal components is

$$k_j = \frac{\sum_{k=1}^{i} \lambda_k}{\sum_{k=1}^{m} \lambda_k} \tag{10}$$

The calculated data for the cumulative contribution of the first 1, first 2, first 3 and first 4 principal components, respectively, were therefore obtained as follows:

$\left(7.46462 \times 10^{-1}, 9.53750 \times 10^{-1}, 9.54100 \times 10^{-1}, 1.00001\right)$

In this study, according to the magnitude of the cumulative contribution, the first two values of the larger cumulative contribution are taken as the corresponding F_1 and F_2 principal components.

To further investigate the relationship between marking speed y and principal components F_1, F_2 . The relationship between marking speed y and principal components F_1, F_2 was described by making graphs respectively, as shown in Figure 1.

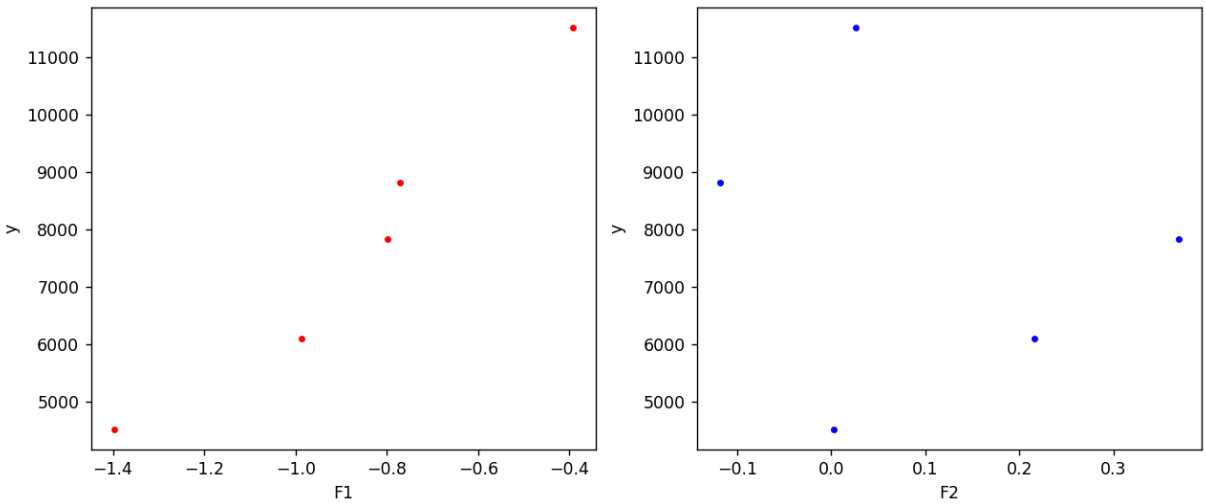


Figure 1. Scatter plot of y versus the two principal components (left panel shows y versus the first principal component, right panel shows y versus the second principal component)

The scatter plot shows that y is roughly a primary function of F_1 and y is roughly a quadratic function of F_2 .Therefore, it is assumed that $y = aF_1 + bF_2^2 + c$. Using the obtained principal components F_1, F_2 to analyse the factors influencing markers' marking speed, the fitted values of a, b, c were obtained by fitting them to the data as

$\{7202.295, -7010.3642, 14291.3574\}$

And make a functional relationship between the two principal components and the speed of marking i.e.

$$y = \mathrm{f}(F_1, F_2) = 7202.295F_1 - 7010.3642F_2^2 + 14291.3547 \tag{11}$$

The graphs of the three-dimensional relationship between the marking speed y and the principal components F_1, F_2 are then made separately, as shown in Figure 2 below.

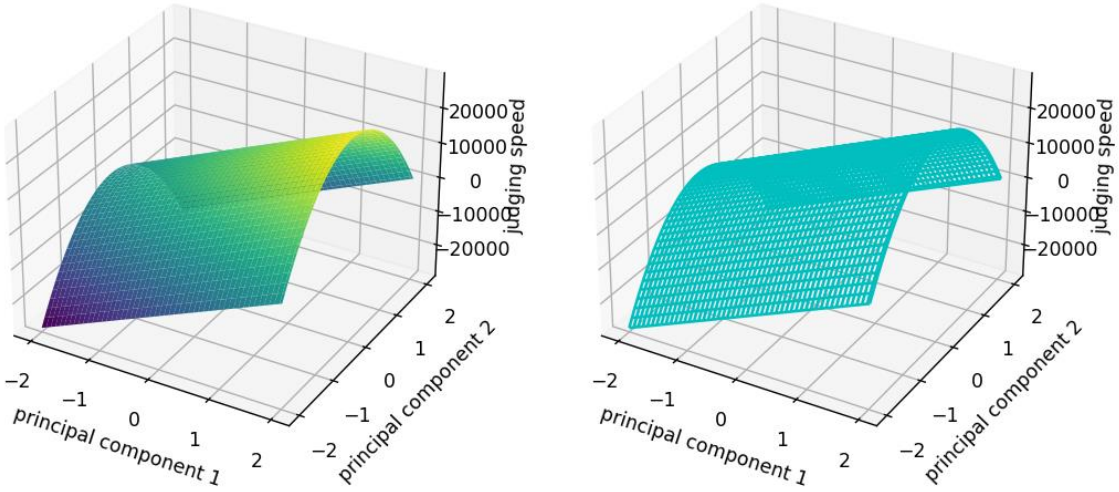


Figure 2. Surface and grid plots of y versus the two principal components

Then by putting $\begin{cases} F_1 = 0.0385x_1 - 0.2036x_2 - 0.1219x_3 - 0.9707x_4 \\ F_2 = 0.2433x_1 + 0.8804x_2 - 0.3870x_3 - 0.1264x_4 \end{cases}$ substituting into the expression for y yields a quadratic equation, i.e.

$$y = \mathrm{f}(F_1, F_2) = \mathrm{f}(x_1, x_2, x_3, x_4) \tag{12}$$

$$\begin{aligned} y = {} & 0.05919x_1^2 + 0.7751x_2^2 + 0.1498x_3^2 + 0.01598x_4^2 + 0.4284x_1x_2 \\ & - 0.1883x_1x_3 - 0.6814x_2x_3 - 0.06151x_1x_4 - 0.2226x_2x_4 + 0.0978x_3x_4 \end{aligned} \tag{13}$$

3. Marking utility model based on integer programming

3.1. Optimal and rational distribution of the number of markers for each question when the workload is completed simultaneously for each question

In the establishment of the marking speed model, this paper based on the four original variables such as the number of markers for each question, the average mark for each question, the number of blank papers for each question, the error rate for each question, etc. After dimensionality reduction, two variables, namely two principal components F_1, F_2, were obtained and used to analyze the influencing factors of marking speed, by fitting the function images between the principal components F_1, F_2 and the marking speed y. The marking speed model was established based on the marking data of markers for five questions such as A, B, C, D and E in one day, i.e. the relationship between the marking speed of markers and the proportion of the number of markers for each question,

the average mark for each question, the proportion of blank papers for each question and the error rate for each question was

$$y = f(F_1, F_2) = 7202.29 \times F_1 - 7010.36 \times F_2^2 + 14291.36 \tag{14}$$

$$\begin{cases} F_1 = u_{11}x_1 + u_{21}x_2 + u_{31}x_3 + u_{41}x_4, \\ F_2 = u_{12}x_1 + u_{22}x_2 + u_{32}x_3 + u_{42}x_4, \end{cases} \tag{15}$$

Bringing the eigenvectors u1, u2, u3 and u4 into the above equation (14) yields:

$$\begin{cases} F_1 = 0.03486x_1 - 0.20359x_2 - 0.12188x_3 - 0.97067x_4, \\ F_2 = 0.24335x_1 + 0.88036x_2 - 0.38697x_3 - 0.12642x_4. \end{cases} \tag{16}$$

Where y is the speed of marking (number of copies/day), F_1 is principal component 1, F_2 is principal component 2, x_1, x_2, x_3, x_4 are the number of markers for the question, the mean mark for the question, the number of blank papers for the question, and the error rate for the question, respectively. In this paper, the proportion of the number of marked scripts, the mean score, the rate of blank scripts and the error rate of A, B, C, D and E are substituted into $y = f(x_1, x_2, x_3, x_4)$ to obtain the relationship between the speed of marking (number of copies/day) and the number of people marking that question for A, B, C, D and E is obtained as $f_1(x_1)$, $f_2(x_1)$, $f_3(x_1)$, $f_4(x_1)$, $f_5(x_1)$.

In modelling the utility of marking, it is necessary to rationalise the distribution of headcount on each day so that the marking time is minimised when the error rate is also guaranteed. In this paper, two model scenarios are proposed for the allocation of the number of people in the simultaneous completion of the workload, they are:

Option 1: Assume that the number of people marking each question each day does not change during the day and that the marking is completed simultaneously on the last day. In this case, the number of days needs to be minimised by selecting a suitable distribution of numbers.

Option 2: Assuming the same marking schedule for each question each day, rationalise the hours of marking work each day to ensure maximum efficiency each day, so that the overall efficiency is maximised, i.e. the total number of days worked is minimised.

3.1.1 Solution of an integer programming model with the number of markers for each question on different days as the decision variable

Assuming that the number of markers for each question type is fixed in a day, the problem is solved by building an integer programming model in order to rationalise the number of markers for each question each day. The total number of markers for a particular question on a given day is assumed to be constant, assuming that each person is only marking one question per day. The objective function of the planning is the marking time used by the markers. Based on the above paper, an integer programming model is developed for the workload required to complete the marking of scripts simultaneously at the final time point. Let the decision A_{ij} be the number of people scheduled for question j on day i.

$$\text{objective function: min } n$$

$$s.t. \begin{cases} \sum_{i=1}^{n} f_j(a_{ij}) \geq m_i, j = 1,2...5. \\ \sum_{j=1}^{5} a_{ij} = 117 \\ 0 \leq a_{nj} \leq 117, j = 1,2...5 \end{cases} \tag{17}$$

Where n is the point in time when all marking is completed and a_{ij} is the number of people scheduled for the $j-th$ question on day i. where m_i is the number of non-blank papers for each question and $f_j(a_{ij})$ is the speed function when a_{ij} is reading question j at the same time.

By solving the integer programming model for completing the marking workload within a specified time, the feasibility of the integer programming model developed in this paper for markers to complete all marking work simultaneously at a uniform time point can be verified.

In this paper, when building the marking speed model, the data of the original variables such as the number of markers x_1 for each question, the average mark x_2 for each question, the number of blank papers x_3 with the error rate x_4 for each question are dimensioned down to obtain two influencing factors related to the markers' marking speed, i.e. principal component 1 and principal component 2, represented by F_1 for principal component 1 and F_2 for principal component 2, and we put the data of questions A, B, C, D and E number of markers, mean score, number of blank papers and error rate into the speed function $y = f(x_1, x_2, x_3, x_4)$ to obtain the speed of marking (copies/day) for A, B, C, D and E in relation to the number of markers for the question to obtain $f_1(x_1), f_2(x_1), f_3(x_1), f_4(x_1), f_5(x_1)$

In order to simplify the integer programming model, it is assumed in this paper that the speed of marking per person per day is only related to the question itself being marked. From the data in Appendix 1, the speed per person per day of marking a particular question is solved for as shown in the following table:

Table 2. Speed of marking for each question at simplification

Question	Simplified marking speed (number of questions/(person-day))
A	237.316
B	380.875
C	195.725
D	575.700
E	400.909

Take the simplified marking speed and multiply it by the number of people assigned to a question per day to equal the total number of markers per question to complete the marking of that question. That is

$$\begin{cases} Y = \left[\dfrac{4509}{4509+6106},\ldots,\dfrac{8820}{8820+8569}\right] * 43800 \\ X = \left[\dfrac{4509}{19},\ldots,\dfrac{8820}{22}\right] = [237.315,380.875,195.725,575.700,400.910] \\ Z = \dfrac{Y}{X} \\ J = \sum_{t=1}^{5} Z \Big/ 117 \end{cases} \tag{18}$$

Therefore, in this paper The number of days required to complete the marking process when all questions are completed simultaneously at the same point in time is 3.3 days.

Therefore, if the marking speed of each person on a question is only affected by the question, the time taken by the marker to complete all the marking work is 3.3 days, which means that the optimal marker arrangement takes about 4 days to complete the marking scheme. By fixing the marking time n (number of days), the following integer programming model is developed to explore the feasibility of the marking staff arrangement, and the results show that this paper is feasible for the marking staff arrangement scheme.

In this paper, the number of marking days for markers, n is taken as 3, 4, 5, 6, 7 and 8, by fixing the number of days used for marking and considering that the number of markers for each question can change at different points in time and in different time periods, but still requiring the same amount of work to be completed at the end of each question. Therefore, the number of markers per day for each question is used as the decision variable, and the objective function is the sum of the marking workload over all marking days, without limiting the total workload. In other words, the objective is to investigate how to achieve the maximum number of scripts marked at a fixed number of days. The integer programming model with hourly time intervals is thus extended to obtain the following model:

$$\text{objective function: } \max \sum_{i=1}^{n}\sum_{j=1}^{5} f_i(a_{ij})$$
$$s.t.\begin{cases} \sum_{j=1}^{5} a_{ij} = 117, i = 1,2...n \\ \sum_{i=1}^{n} f_i(a_{ij}) = m_i, j = 1,2...5 \\ 0 \le a_{nj} \le 117, j = 1,2...5 \end{cases} \tag{19}$$

That is, the total number of markers for questions A, B, C, D and E is 117, and by setting a fixed number of marking days, it is calculated that

$$\sum_{i=1}^{5} f_i(a_{ij}) = m_i, j = 1,2...5 \tag{20}$$

This is the number of days of marking time used by each question marker to complete the marking of all non-blank scripts. In this paper, the minimum n for which a feasible solution is obtained is chosen as the optimal choice of days and the arrangement of the number of markers corresponding to that number of days is taken as the optimal solution for the decision variable.

As a result, the optimal number of markers for each question is obtained on day 4, when the markers have marked all the papers at the same time, and the optimal number of markers for each question, such as A, B, C, D and E, is obtained on that 4 days.

Table 3. Marking schedule for Option 1

Day Question	A	B	C	D	E
Day 1	20	17	55	10	15
Day 2	44	19	32	5	17
Day 3	6	16	57	26	12
Day 4	23	15	47	13	19

3.1.2 Solution of an integer programming model with the allocation of personnel to each question per hour as the decision variable

In the analysis of Scenario 2, an integer programming model with hourly time intervals is developed. In this paper, the number of markers per question per hour is arranged in such a way that the overall efficiency of marking is maximised, while ensuring that each question progresses at the same rate each day and that each marker has to mark only one question type in each hour. Since the efficiency of marking can be assumed to be the same every day under this assumption, we only need to solve for the maximum single-day efficiency.

To do this, we can build an integer programming model

$$\max \sum_{j=1}^{7} f_1^{'}(a_{1j})$$
$$s.t.\begin{cases} \sum_{i=1}^{5} a_{ij} = 117,\ j = 1,2,\ldots 7 \\ \sum_{j=1}^{7} f_1^{'}(a_{1j}) = \sum_{j=1}^{7} f_i^{'}(a_{ij}), i = 2,3,4,5 \\ 0 \le a_{ij} \le 117 \end{cases} \tag{21}$$

Where $f_i^{'}(a_{ij})$ is used as an indicator of the speed of marking and is the number of marks per hour for that question when a_{ij} people mark the $i-th$ question. As only 7 hours are worked per day, so

$$f_i^{'}(a_{ij}) = \frac{1}{7} f_i(a_{ij}) \tag{22}$$

In order to further optimise the solution of the model, this paper uses the simulated annealing algorithm for analysis. The simulated annealing (SA) algorithm is derived from the solid annealing principle and is a probability-based algorithm. The idea of the algorithm is to start with a high initial temperature and gradually reduce the temperature until the thermal equilibrium conditions are met. At each temperature, n rounds of search are performed, with a new solution generated by adding random perturbations to the old solution at each round and accepting the new solution according to certain rules.

The specific procedure of simulated annealing is described as follows:

(1) Let any solution satisfying the initial conditions be $A_{5\times7}(0)$,whose $i-th$ row and $j-th$ column is $a_{ij}(0)$, which is the initial solution of the problem. Let the solution space be all the solutions satisfying the above constraints, and let the initial temperature be T_0 =100°C, so that k =0.

(2) When the solution at this point is $A_{5\times7}^{'}(k)$, choose a new state at random in a neighbouring subset of the solution space $A_{5\times7}^{'}(k)$,take $C(A_{5\times7}) = \sum_{j=1}^{7} f_1^{'}(a_{1j})$ as the evaluation function, If $\Delta C = C(A_{5\times7}(k)) - C(A_{5\times7}(k)) > 0$,then accept this result directly as the new $A_{5\times7}(k)$, otherwise accept this result with probability p, which we take to be $p = e^{-\frac{\Delta C^{'}}{T}}$. The process is repeated until $A_{5\times7}(k)$ converges.

(3) Cool down using the cooling factor α =0.999 and take the new temperature to be $T = \alpha T$ such that $k = k+1$. Repeat step 2.

(4) Stop iterating when $T < 10^{-30}$, and take the final solution as the global optimal solution.

Using Python code to solve the problem, the optimal solution is obtained as shown in Table 4

Table 4. Marking schedule for Option 2

question hours	1h	2h	3h	4h	5h	6h	7h
A	20	44	6	23	22	17	18
B	17	19	16	15	15	33	15
C	55	32	57	47	53	22	53
D	10	5	26	13	8	32	13
E	15	17	12	19	19	13	18

Therefore, the number of markers per hour for each question type is obtained. By using this result to schedule all markers each day, we can maximise the efficiency of each day and therefore the overall marking efficiency. The results show that by the fourth day, all scripts will have been marked.

4. Conclusion

This paper used Pearson's correlation coefficient to obtain the factors affecting marking speed, namely the number of markers for each question, the mean score for each question, the number of blank papers for each question and the error rate for each question, and used principal component analysis with non-linear fitting to reduce the dimensionality of the four main factors of marking speed, and obtained the two principal components F_1 and F_2 for the vast majority of information of these four factors, obtained the relationship between marking speed and the two principal components, and established the The relationship between marking speed and the two principal components was obtained, and a model of marking speed was established.

In this paper, the speed of marking A, B, C, D and E per person per day is 237.316, 380.875, 195.725, 575.700 and 400.909 respectively, and the total time to complete the marking of each question at the same time is 3.3 days. Based on the integer programming model, the optimal arrangement of markers for A, B, C, D, E, etc. was obtained when the marking of all the questions was completed in approximately 4 days; then, by building an integer programming model with the allocation of personnel per hour for each question as the decision variable, and solving it through the simulated annealing algorithm, the optimal arrangement of personnel for each hour of each day for each question type was obtained, so that the marking of its papers could be maximised The marking was completed within 4 days for all question types.

5. Error analysis of the model

In order to maximise the fairness of the examination, the focus should be on how to reduce the error rate of marking while ensuring the efficiency of marking. The error rate is related to a number of factors.

First, the limitations of standard answers and marking rules can cause errors. In the marking of subjective questions, there can be differences in the marking of the same paper due to the subjective factors of the marker, as well as the marker's level of grasp of the marking criteria and the different understanding of the answers by each marker. This is one of the sources of error rates.

Secondly, the marking process may have some influence on the marking of examination papers due to the psychological factors of the markers. When a marker approves of the acceptance of a person's answer is, a subconscious standard is formed, which means that the candidates' answers are influenced by each other. This is known as ranking influence. At the same time, a candidate's paper can influence the marker's judgement. Again, research has shown that as marking proceeds, the speed of marking and the rate of marks awarded by the same marker increases, but there is a tendency for marking to converge. The tendency to score in the middle of the scale refers to the tendency of markers to concentrate their scores in the middle and lower range of the scale as a whole. However, it will directly lead to a reduction in the validity of the assessment, which will affect the presentation of candidates' true standards and lead to marking errors. Also, markers' personal preferences can interfere with judgements and cause marking errors.

Thirdly, the quality and standard of different markers also affects the margin of error.

In conclusion, in order to maximise the quality of marking and the fairness of the examination, it is necessary to develop an error reduction model for this paper. In order to establish the error elimination model, the Kolmogorov test was first used. The test follows the following steps:

(1) Let the full score of a question be S and the sample size be n and sufficiently large. The significance level α is 0.05. Calculate the statistic

$$\hat{\mu} = \bar{x}$$
$$\hat{\sigma} = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2 \tag{23}$$

(2) Calculate the $D_{n,a}$ and D_n and check the table to know the critical value λ_a. Compare the relationships to derive whether the original distribution hypothesis is accepted.

(3) The consistency of markers' results is judged using the Kolmogorkov-Smirnov test, which is the size of the error by whether the observations are consistent. Suppose that the total score for a particular question is S .Now make a judgement on the results of the work of a particular marker by performing the following steps:

(4) Select a particular target marker whose total number of papers marked is c. The number of papers marked by the target marker with a score of i is c_i. The total number of papers marked by the other markers is L and the number of papers with a score of i is l_i. $i = 0,1,2...S$

(5) Define cumulative weighting. Let the cumulative weighting of the target marker be $p_i = \frac{\sum_{j\le i} c_j}{C}$ and the cumulative weighting of the remaining markers $q_i = \frac{\sum_{j\le i} l_j}{L}$.

(6) Calculate the maximum of the absolute value of the difference between the cumulative weight of the target marker and the cumulative weight of the other markers $D = \sup_i \left| p_i - q_i \right|, i = 0,1,2...S$

(7) Calculation of judgement factor $\xi = D\sqrt{\frac{CL}{C+L}}$.

(8) Use the calculation results to make a judgement. When $\xi > \lambda_\alpha \alpha$, it is considered that the target marker's deviates significantly from the results of others and the target marker should stop marking to make adjustments. Conversely, the target marker's results are considered to be in very good agreement with other markers and the error is very small.

References

[1] Zhang W, Wang HB, Zhao Y. (2018) Research on automatic scoring model based on deep learning. Computer Science and Exploration, 12(6), 873-884.

[2] Singh, R., Saleh, M., & Ng, A. (2017) Gradescope: A Fast, Flexible, and Fair System for Scalable Assessment of Handwritten Work. Proceedings of the Fourth ACM Conference on Learning at Scale, 365-368.

[3] Ramanathan, A., & Chakraborty, T. (2020) A Semantic Similarity Based Approach for Automated Paperless Grading. 2020 IEEE 32nd International Conference on Tools with Artificial Intelligence (ICTAI), 71-78.

[4] Rohde, H., & Schlangen, D. (2017) Evaluating the Quality of Automatically Evaluated Long-Text Answers in E-Assessment. Proceedings of the 12th Workshop on Innovative Use of NLP for Building Educational Applications, 146-155.

[5] Koichi Yamamoto, Tomoaki Takagi, Hidetoshi Saito. (2019) An intelligent scoring method based on knowledge graphs. Research Report of the Computer Society of Japan, 2019(9), 1-8.