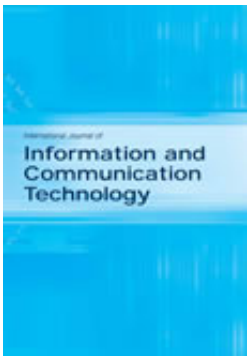


International Journal of Information and Communication Technology

ISSN online: 1741-8070 - ISSN print: 1466-6642
https://www.inderscience.com/ijict

Analysing university student pension insurance using the K-prototypes algorithm and logistic regression model

Qingqing Wei

DOI: 10.1504/IJICT.2024.10063679

Article History:

Received:	29 December 2023
Last revised:	29 January 2024
Accepted:	02 February 2024
Published online:	12 May 2024

Analysing university student pension insurance using the K-prototypes algorithm and logistic regression model

Qingqing Wei

Department of Architecture Engineering,
Shi Jia Zhuang University of Applied Technology,
Shijiazhuang, Hebei, 050081, China
Email: 2014010695@sjzpt.edu.cn

Abstract: Addressing the pressing issue of 'five insurances and one pension' for financially constrained university students, this study employs the K-algorithm in student pension insurance. This algorithm not only safeguards the fund but also evaluates individuals, offering advantages such as tailored financial planning aligned with consumption capacity and risk tolerance. By integrating factor analysis, a refined evaluation index system for student pension insurance is devised. The study utilises the K-B optimisation method to simulate a student's participation and proposes strategies to enhance investment returns. This approach aims to ensure capital safety, maximise insured interests, and foster societal stability through sustainable development.

Keywords: logistic regression model; K-prototypes algorithm; university students' pension insurance; analysis; impact; application; suggestions.

Reference to this paper should be made as follows: Wei, Q. (2024) 'Analysing university student pension insurance using the K-prototypes algorithm and logistic regression model', *Int. J. Information and Communication Technology*, Vol. 24, No. 6, pp.92–102.

Biographical notes: Qingqing Wei studied at Hebei University of Economics and Trade from 2009 to 2012 and received her Master's degree in 2012. Since 2014, she has been working as a college teacher in Shijiazhuang Institute of Vocational Technology and has published more than ten papers in Chinese journals. Her main research areas are social security, innovation and entrepreneurship education, employment guidance, and ideological and political education research.

1 Introduction

K-prototypes algorithm is a combination of K-means algorithm and K-modes algorithm, which is mainly used to solve the problem of handling mixed types of data (Wang and Liu, 2012; Chen et al., 2014). K-prototypes algorithm is still essentially a division-based clustering algorithm, and the difference with the traditional division-based clustering algorithm is that the iterative process of calculating the distance from data objects to prototypes needs to take into account In addition (Wang et al., 2012; Zhang et al., 2013),

the prototypes of the categorical attributes are updated in a frequency-based manner after a new set of clusters has been generated, i.e., the most frequent attribute value of the categorical attributes is taken as the new prototype, and the completion of the iterative process still depends on the clustering criterion function E (Dai et al., 2016; Li and Zhang, 2016; Chen and Tang, 2019).

In this paper, the distance between the attribute values of the subtypes is defined as 0 for two objects with the same attribute value and 1 for two objects with different attribute values. As objects are added to the clustering set in subsequent iterations, the calculation of the dissimilarity of the categorical attributes will follow the distribution of the attribute values of the categorical attributes (Zhu and Xie, 2011; Zhang, 2019). Here we first choose to assume that the distribution of attribute values for a given categorical attribute is (yellow, red, red, blue, blue, blue), with blue being the centre because it has the largest distribution of attribute values, the distance between yellow and blue can be written as d(yellow, blue) = 5/12, between red and blue as d(red, blue) = 4/12, and between blue and blue as d(blue, blue) = 3/12. This distance calculation can be explained in a more intuitive way using the yellow attribute has the lowest attribute value in the colour distribution (Zhang, 2015; Liu, 2016), which means that it is the least likely to be the centre of a colour cluster and therefore the most distant from the centre of an existing cluster, while blue is the most likely to be the centre of a colour cluster as it is the centre of a colour cluster and therefore the least distant from the centre of an existing cluster. The key aspect of this method is to calculate the distance between any attribute value and the cluster centre, so in the K-prototypes algorithm, after a number of initial cluster centres have been selected, the distance of the attribute part will be changed according to the new objects added to each cluster set, i.e., the distance of the attribute part is the distance between the object This means that the distances in the categorical attributes section take into account the distances between the objects and the whole cluster set rather than just the distances between individual objects, which is different from the calculation of the numerical attributes section (Yu et al., 2015; Ouyang et al., 2015).

Solving the pension problem of university students in China is conducive to alleviating the impact of an ageing population and safeguarding the long-term harmonious and stable development of the country while ensuring that individuals have a stable old age. The thesis revolves around the issues that constrain university students' choice to participate in basic pension insurance (Jia and Song, 2020; Tao and Feng, 2010). Through a combination of qualitative and quantitative methods, it focuses on the influencing factors that restrict university students' choice to participate in the insurance, and provides a feasible reference basis for improving the pension insurance system for university students in China based on the current situation of the research.

2 The K-prototypes algorithm and its improvements

The distance measure formula of the K-prototypes algorithm needs to be divided into two parts, i.e., the numerical attribute part and the subtype attribute part. The distance measure for the numerical attribute part is based on the squared Euclidean distance, and the distance measure formula for this part is defined as follows.

$$d_1\left(X_i, V_j\right) = \sum_{l=1}^{q} \left| x_{il} - v_{jl} \right|^2 \tag{1}$$

$$d_2\left(X_i, V_j\right) = \sum_{l=q+1}^{m} \delta\left|x_{il}, v_{jl}\right|^2 \tag{2}$$

$$\delta\left(x_{il}, v_{jl}\right) = 0 x_{il} = v_{jl} 1 x_{il} \neq v_{jl} \tag{3}$$

$$d\left(X_i, V_j\right) = d_1\left(X_i, V_j\right) + \theta d_2\left(X_i, V_j\right) \tag{4}$$

Current improvements to the distance measure of the K-prototypes algorithm are often based on the Hemming formula of the K-modes algorithm, which calculates the distance between attributes of a subtype if the two attributes are the same, If the two attributes are the same, then the distance between them is recorded as 1; if the two attributes are different, then the distance between them is recorded as. The specific calculation formula is as follows.

$$\delta\left(x_{il}, v_{jl}\right) = \begin{cases} 0, & x_{il} = v_{jl} \\ 1, & x_{il} \neq v_{jl} \end{cases} \tag{5}$$

This formula is simple to understand and simple to calculate, however there are obvious drawbacks. Firstly, since the distance between any two types of attributes is only 0 or 1, some objects cannot be accurately classified into clusters because the distances to the cluster centres are the same. Secondly, the distances between subtypes of attributes are recorded as 0 or 1, ignoring the distribution of the values of each attribute under the corresponding attribute and the variation in the values of the subtypes of attributes for the overall population. Only a more detailed division of the distances between the sub-types of attributes can facilitate the measurement of distances between objects and their accurate classification. Therefore, the author proposes a new method for calculating the distance between sub-typed attributes with the following formula.

$$d_2\left(X_i, V_j\right) = \sum_{l=q+1}^{m} \delta\left(x_{il}, v_{jl}\right) \tag{6}$$

$$\delta\left(x_{il}, v_{jl}\right) = \frac{D_{lp}}{D_l} \tag{7}$$

The distance formula of the improved K-prototypes algorithm is still divided into two parts, i.e., the numerical attribute part and the typed attribute part. The distance measure for the numerical attribute part is based on the squared Euclidean distance and the distance measure formula for this part is defined as follows:

$$d_1\left(X_i, V_j\right) = \sum_{l=1}^{q} \left|x_{il} - v_{jl}\right|^2 \tag{8}$$

$$d_2\left(X_i, V_j\right) = \sum_{l=q+1}^{m} \delta\left(x_{il}, v_{jl}\right) \tag{9}$$

$$\delta\left(x_{il}, v_j\right) = \frac{D_{lp}}{D_l} \tag{10}$$

Combining the distance formulae for the numerical attribute part and the typed attribute part, the distance measure formulae and the clustering criterion functions for the modified K-prototypes algorithm can be derived as follows:

$$d\left(X_i, V_j\right) = \sum_{i=1}^{q} \left| x_{il} - v_{jl} \right|^2 + \theta \sum_{l=q+1}^{m} \delta\left(x_{il} - v_{jl}\right) \tag{11}$$

$$J(X, V) = \sum_{i=1}^{n} \sum_{j=1}^{k} w_{ij} d\left(X_i, V_j\right) \tag{12}$$

$$\sum_{i=1}^{n} w_{ij} = 1\left(w_{ij} \in \{0, 1\}\right) \tag{13}$$

When the value of w_{ij} is 1, it means that the i^{th} object belongs to the j^{th} class. Then, k objects are randomly selected as the initial cluster centres, and the distance between all the objects and the cluster centres is calculated, and the object with the smallest distance is merged into the corresponding cluster set, thus forming k clusters (Jiang et al., 2015; Shi and Wang, 2017). The k clusters are then updated to determine the new cluster centres by calculating the mean of the numerical attributes of all objects in the cluster as the centres of the numerical attributes, while the frequency-based approach is chosen to determine the centres of the sub-types of attributes. The improved K-prototypes algorithm proposed in this paper proceeds as follows.

Input A set X with n data objects and the number of classes *k*.

Output *k* disjoint classes.

Step1 Randomly select *k* objects from the dataset *X* as the initial clustering centres.

Step2 Calculate the dissimilarity between all objects and the initial clustering centres according to the distance measure given in equation (11), and assign the objects to the classes represented by their closest clustering centres according to the minimisation principle.

Step3 Update the clustering centres of all classes, with the mean value of the numerical attributes being used as the centre of the numerical attributes, while the frequency-based approach, i.e., the attribute value with the most occurrences under the categorical attributes, is chosen as the centre for the sub-typed attributes.

Step4 Repeat step2 and step3 until the clustering criterion function converges.

3 Validation of the improved K-prototypes algorithm

In order to verify the effectiveness of the improved K-prototypes algorithm proposed in this paper, two metrics, clustering accuracy and clustering purity, are used to measure the effectiveness of the algorithm. The expressions of clustering accuracy and clustering purity are as follows:

$$AC = \frac{1}{n} \sum_{i=1}^{k} a_i \tag{14}$$

$$PR = \frac{1}{k} \sum_{i=1}^{k} \frac{a_i}{a_i + b_i} \tag{15}$$

In this formula, n represents the number of objects, a_i represents the number of objects correctly classified into the corresponding clusters, b_i represents the number of objects incorrectly classified into the corresponding clusters, and k represents the number of clusters. The higher the value of these two metrics, the closer the classification of the dataset by the clustering method is to the classification of the dataset itself. In this paper, we first select the credit approval dataset from UCI, which is a dataset of 690 objects and 15 attributes related to the credit card application of users. Among all the attributes, there are nine attributes of type and 6 attributes of numerical type; the dataset is divided into two categories, with positive signs indicating good credit and negative signs indicating poor credit. The improved K-prototypes algorithm proposed in this paper was compared with the original algorithm to obtain Table 1.

Table 1 Comparison of the accuracy of clustering under the dataset credit approval

	AC price	*PR price*
K-prototypes algorithm	0.8231	0.8242
Improved algorithm	0.8436	0.8436

The TAE dataset from the UCI was then selected, which contains the assessment results of a university branch statistics department for three regular semesters and two summer semesters, with 151 teaching assistant tasks, classified into three numerical attributes and two sub-types of attributes; the dataset was divided into three categories, low, medium and high, to represent the categories of performance assessment (Mao, 2017; Zheng, 2018). The improved K-prototypes algorithm proposed in this paper was compared with the original algorithm to obtain Table 2.

Table 2 Comparison of clustering accuracy under dataset TAE

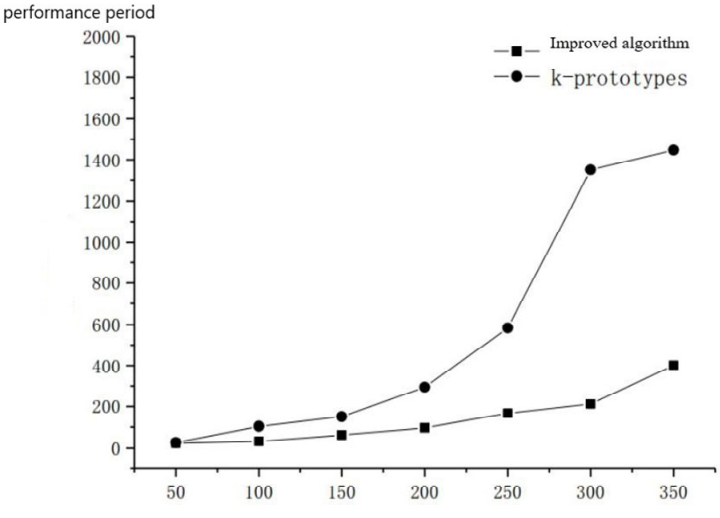
	AC price	*PR price*
K-prototypes algorithm	0.6873	0.7827
Improved algorithm	0.8104	0.7860

According to Table 1 and Table 2, the AC and PR values of the improved algorithm are higher than those of the K-prototypes algorithm, which means that the improved algorithm is better than the K-prototypes algorithm in terms of clustering effectiveness, and also shows that the proposed formula for calculating the dissimilarity can correctly reflect the distance between objects for clustering. Subsequently, in order to compare the difference in clustering time between the improved algorithm and the original K-prototypes algorithm, 50, 100, 150, 200, 250, 300 and 350 objects were randomly selected and clustered using the two algorithms, and the relationship between running time and number of objects is shown in Figure 1.

When the number of objects is increased, the time consumed by the k-prototypes algorithm is much longer than that of the original k-prototypes algorithm, which is due to the fact that after several iterations of the clustering algorithm, the time consumed by the k-prototypes algorithm is much longer than that of the original k-prototypes algorithm. This is due to the fact that after several iterations of the clustering algorithm, the improved algorithm has developed an effective division pattern due to the scientific nature of the division, which allows the objects to be placed into their respective sets

more quickly, and thus the improved algorithm's running time is significantly lower than that of the k-prototypes algorithm after the number of objects reaches 200.

Figure 1 Runtime-number of objects relationship



4 Exploring the factors influencing university students' choice to participate in basic pension insurance and their relationships by integrating logistic regression models

4.1 Analysis of the factors influencing university students' choice to join the insurance

The dependent variable in the regression model, i.e., whether university students choose to participate in basic pension insurance or not, usually results in only two cases, i.e., participation and non-participation, and the sum of the probabilities of these two behaviours is 1. Therefore, for this type of dependent variable, we usually use a binary logistic regression model for analysis. A value of 1 indicates that the event occurred and the university student chose to join the basic pension insurance; a value of 0 indicates that the event did not occur and the university student did not join the basic pension insurance. The probability that a university student chooses to join the insurance is (16), and the following logistic regression model can be obtained.

$$P(Y=1|x_i)=P_i \tag{16}$$

$$P_i=\frac{1}{1+e^{-\left(a+\sum_{i=1}^{m}\beta x_i\right)}}=\frac{e^{\alpha+\sum_{i=1}^{m}\beta x_i}}{1+e^{\alpha\sum_{i=1}^{m}\beta x_i}} \tag{17}$$

$$1-P_i=1-\frac{1}{1+e^{-\left(\alpha+\sum_{i=1}^{m}\beta_i x_i\right)}}=\frac{1}{1+e^{\alpha+\sum_{i=1}^{m}\beta_i x_i}} \tag{18}$$

They are all nonlinear functions consisting of the independent variable *X*. The ratio of the probability of an event occurring to the probability of it not occurring becomes the probability ratio of the event, denoted Odds, which is positive and has no upper limit. The linear model of the logistic regression model can be obtained by logarithmic transformation of the ratio as follows:

$$\ln\left(\frac{P_i}{1-p_i}\right)=\alpha+\sum_{i=1}^{m}\beta_i x_i \tag{19}$$

The selection of independent variables in this paper is summarised in six categories of influence: individual, household, region, policy perception, trust in participation and participation context. Some of these variables were processed and integrated according to the needs of the study. As the source of economic income of individuals in the survey sample is related to the level of local economic development, it is not possible to determine the screening criteria for the economic income situation in the sample, and there is a large income gap between regions in China. Therefore, based on the study of the national university students' choice to participate in the insurance issue, it lacks a certain scientific rigour to analyse the impact on the participation situation by studying the amount of income of the economy. The author tried to measure the relationship between economic status and university students' participation in basic pension insurance by studying other options in the questionnaire. The subjective and non-quantitative question 'Your economic status in the city' was used to determine the economic status of the respondents. In addition, as the questionnaire did not include information about the current pension policy, I used 'how many days in the past week did the respondent know about political news' as a reference, and defined those who did not know for one day and those who knew for only one day as not knowing at all, those who knew for 2–3 days as knowing a little, and those who knew for 4–3 days as knowing a lot. The number of days of awareness is defined as not knowing at all, the number of days of awareness is defined as 2–3 days, the number of days of awareness is defined as partial awareness, and the number of days of awareness is defined as 6–7 days of full awareness. The frequency of knowledge of political news is converted into knowledge of the policy. Satisfaction with the basic pension policy is measured by 'satisfaction with the current national policy', although there is some variation in the content of the survey, all of which is based on individual satisfaction with the current national policy. The results of the descriptive statistical analysis of the selected variables are shown in Table 3.

4.2 Initial modelling attempts and final model construction

A binary logistic regression of the above variables was carried out using SPSS 20.0. Table 4 shows a categorical tabular plot of the regression model predictions after all the independent variables were introduced. The software was used to determine whether or not to participate based on whether or not P was greater than 0.5. The results show that 77.1% of the total was uninsured.

Table 5 outputs the results without the introduction of the independent variable; where B is the estimate of the constant term at –1.213 and the p-value is 0.000, which is somewhat significant; where Exp(B) is the alpha power of e. Its practical significance is the ratio of uninsured to insured for the overall study population 856/2879 = 0.297.

Table 3 Descriptive analysis table

Min		*MAX*	*Mean value*		*Standard deviation*	*Variance*
Statistical quantity		*Statistical quantity*	*Statistical quantity*	*Standard error*	*Statistical quantity*	*Statistical quantity*
Age	1	9	3.92	0.028	1.730	2.992
Sex	1	2	1.45	0.008	0.497	247
Income status	1	5	2.47	0.015	0.905	818
Health condition	1	5	2.64	0.017	1.046	1.095
Province	1	3	1.68	0.013	0.795	633
Registered permanent residence nature	1	2	1.29	0.007	0.455	0.207
Policy awareness	1	4	2.75	0.021	1.280	1.638
Policy satisfaction	1	5	3.56	0.016	0.990	0.981
To the government about trust	1	5	2.68	0.017	1.018	1.036
Evaluation by the municipal government	1	5	2.65	0.013	821	674
Social class	1	5	2.54	0.016	0.956	0.914
Life satisfaction	1	5	3.43	0.017	1.044	1.091

Table 4 Categorical table plot

Observed			*Predicted*		
			Ginseng protect a project		*Percent correction*
			Not insured	*Has participated in the insurance*	
Step 0	Ginseng protect item	Not insured	2879	0	100.0
		Has participated in the insurance	856	0	0.0
	Total percentage				77.1

Table 5 Variables in the equation

		B	*S.E.*	*Wals*	*df*	*Sig.*	*Exp (B)*
Step 0	Constant (quantity)	–1.213	0.039	970.722	1	0.000	0.297

The model was first tested for fit and the regression results showed that the Chi-squared value of the model was 729.957 with a corresponding p-value of 0.000, indicating that the model was significant. The Cox & Snell R^2 and Nagelkerke R^2 were 0.178 and 0.269 respectively, with a –2 log likelihood value of 3,291.053, which indicates that the overall fit of the model was generally, and the regression results can be used to analyse and judge the results of each variable on the independent variables. Based on the results of the initial attempts to build the model, my thesis attempted to remove the variables that failed to pass significance before conducting the regression analysis again. The results of the pooled test of the model coefficients showed that the chi-squared value of the model was

726.201 and the corresponding p-value of 0.000, indicating that the model was significant. The model summary results show that the Cox & Snell R^2 and Nagelkerke R^2 are 0.277 and 0.368 respectively, with a –2 log likelihood value of 3,174.809, indicating that the model fits well overall and that the regression results can be used to analyse and judge the results of each variable on the independent variables.

Table 6 Comprehensive test of final model coefficients and final model summary

Comprehensive test of final model coefficients				
		Chi-square	*df*	*Sig.*
Step 1	Step	726.201	10	0.000
	Block	726.201	10	0.000
	Model	726.201	10	0.000
Final model summary				
Step	*Log likelihood values*	*Cox & Snell R^2*		*Nagelkerke R^2*
1	3,174.809	0.277		0.368

From the analysis of the empirical results obtained in Table 6, we can analyse that personality characteristics are the deep-seated influencing factors on whether university students participate in the policy, and these factors are mainly related to the age, income status and education level of university students. These factors are mainly related to age, income status and education level. Therefore, the key to promoting university students' choice to participate in basic pension insurance is to combine the characteristics of university students and formulate policy provisions that match their needs, so as to ensure their stable and long-term participation. The middle variables are mainly policy perceptions and regional characteristics, and policy awareness and satisfaction are directly influenced by individual characteristics, which directly change the first level of trust and participation scenarios. The three-level progression of the relationship shows that whether university students will choose to participate in basic pension insurance is closely related to the evaluation of the government and the construction of the participation environment, that is, the performance and image of the government directly affects the effect of pension policy facilities, and the participation situation of university students also affects the final outcome of the policy. In addition, the participation environment is also influenced by regional characteristics. From the analysis of the differences between the eastern, central and western regions of China and the differences between rural and urban areas, it is necessary to pay attention to the construction of the characteristics of each region, rely on the background of the national macro policy, develop feasible policy solutions suitable for the region, create a local characteristics of the participation environment, and mobilise the enthusiasm of college students to participate in insurance in all aspects.

The results have important policy and practical implications. They highlight the need for tailored provisions catering to students' specific needs and situations to boost participation. Policymakers should focus on improving awareness, affordability perception, transferability, and enhancing incentives. Additionally, differentiated regional strategies are needed rather than one-size-fits-all approaches. The performance and trustworthiness of various levels of government also impact program effectiveness. So efforts to increase transparency and outreach to universities and students could aid

expansion. Overall, the advanced analytics provide data-driven insights to diagnose shortcomings and identify high-potential areas to target for increasing pension participation rates among university students. Both methodological innovation and domain knowledge contribution are achieved.

While this study makes several contributions, some limitations provide avenues for further research. The sample only covers students at one university, which may not generalise. Expanding the diversity and representation of participants could strengthen findings. Additionally, measuring students' risk appetites could further enrich the models. Dynamic aspects like changing views over the course of studies could also be examined. On the methodology side, applying other techniques like neural networks or decision trees could provide comparative analytical perspectives. Ensemble models blending multiple approaches may improve predictive accuracy as well. Qualitative interview or focus group data could complement the quantitative results with more contextual insights into thought processes and behaviours. Overall, extensions leveraging bigger, broader samples and more complex tools would be valuable.

5 Conclusions

This study aimed to analyse the factors influencing university students' participation in pension insurance using the K-prototypes algorithm and logistic regression model. The literature review covered these algorithms and models, and how they can be applied to this research problem. The research questions focused on identifying the key factors and relationships affecting students' participation decisions. The above study found that although the overall participation rate of college students in China is gradually increasing, it is still at a relatively low level. Firstly, the awareness of university students is weak, secondly, the basic pension insurance contribution policy is too high and unfair, thirdly, it is difficult for university students to participate in the basic pension insurance and transfer the relationship, and fourthly, the incentive effect of basic pension insurance participation is not obvious. From the research questions, a binary logistic regression model and an explanatory structure model were used to investigate the factors that influence university students' choice of insurance. The results revealed that income, culture, economy, region, policy awareness, trust in the government and life satisfaction have significant influence on university students' decision making behaviour. The results align with the goals outlined in the introduction and literature review of utilising data mining techniques and statistical models to uncover patterns in a complex dataset. Additionally, the hierarchical analysis of influencing factors builds on the conceptual framework to show how personal, policy, regional, and other characteristics interrelate to shape outcomes. This advances theoretical understanding of the multidimensional nature of pension participation. A hierarchical analysis of the influencing factors through the explanatory results model reveals that individual characteristics have a deep influence on trust in participation and participation scenarios, mainly through policy perceptions, while regional characteristics have an influence on the target variables through participation scenarios, and family characteristics directly influence university students' choice to participate in insurance and do not constitute an influential relationship with other variables.

References

Chen, J.L. and Tang, Y.J. (2019) 'Research on the phenomenon of flexible employment in the era of mobile internet', *Labor Security World*, Vol. 2019, No. 17, pp.19–20.

Chen, J-M., Tang, Y., Li, J-G. and Cai, Y. (2014) 'Research on personalized recommendation algorithm', *Journal of South China Normal University (Natural Science Edition)*, No. 1004-9398, Vol. 5, pp.8–15.

Dai, S.L., Tang, M.D. and Lu, S.X. (2016) 'Web service selection based on grey correlation analysis', *Computer Engineering and Science*, Vol. 2, No. CN31-1289, pp.297–304.

Jia, Z.Q. and Song, L. (2020) 'A hybrid data clustering algorithm for hybrid data. A k-prototypes clustering algorithm for hybrid data clustering', *Small Microcomputer Systems*, Vol. 41, No. 9, pp.55–62.

Jiang, X., Wei, Y. and Qiu, B. (2015) 'QoS-aware personalized recommendation for web services', *Computer Technology and Development*, Vol. 12, No. 1673-629X, pp.85–90.

Li, Z. and Zhang, T. (2016) 'Study on genetic algorithm for topologically clustered adaptation degree sharing of small habitats', *Journal of Harbin Institute of Technology*, Vol. 48, No. 5, pp.178–183.

Liu, Q. (2016) 'Analysis of the factors that make it difficult to expand the coverage of urban social pension insurance', *Journal of Party and Government Cadres*, Vol. 2016, No. 6, pp.9–10.

Mao, B-B. (2017) 'Weaving a social security net for people in 'informal employment' in a gradual and flexible manner', *People's Forum*, Vol. 2017, No. 8, pp.64–65.

Ouyang, H., Dai, X., Wang, Z. et al. (2015) 'A rough K-prototypes clustering algorithm based on information entropy', *Computer Engineering and Design*, Vol. 2015, No. 5, pp.1239–1243.

Shi, M. and Wang, B. (2017) 'Research on the adjustment model of informal employment labor relations in China', *China Labor*, Vol. 2017, No. 11, pp.22–24.

Tao, C. and Feng, Z. (2010) 'QoS-aware recommendation model for web services', *Computer Application Research*, Vol. 10, No. 1001-3695, pp.3902–3905+3914.

Wang, G. and Liu, H.P. (2012) 'A review of personalized recommendation systems', *Computer Engineering and Applications*, Vol. 7, No. 1002-8331, pp.66–76.

Wang, S., Sun, Q. and Yang, F. (2012) 'Credibility evaluation methods in web service selection', *Journal of Software*, Vol. 6, No. 1000-9825, pp.1350–1367.

Yu, W-L., Yu, J-J. and Fang, J-W. (2015) 'A k-prototypes clustering algorithm for mixed attribute data', *Computer Systems Applications*, Vol. 2015, No. 6, pp.170–174.

Zhang, L. (2015) 'A study on the current situation of labor dispatch development', *China Labor*, Vol. 2015, No. 6, pp.13–17.

Zhang, L. (2019) 'Flexible employment: an important channel to promote employment under the crisis', *China Party and Government Cadres Forum*, Vol. 2019, No. 4, pp.9–12.

Zhang, X., He, K., Wang, J. and Liu, J. (2013) 'A review of personalized recommendation for web services', *Computer Engineering and Science*, Vol. 9, pp.132–140.

Zheng, Y. (2018) *Research on the Participation of University Students in Basic Social Pension Insurance*, Northwestern University.

Zhu, Z. and Xie, Z. (2011) 'Reasons and countermeasures for the low participation rate of university students in pension insurance', *Social Insurance*, Vol. 2011, No. 10, pp.4–10.