Check for updates

Evaluation of Digital Transformation in Chinese Government from Data Mining Perspective

Ping Lan(✉)

Faculty of Humanities and Social Science, Macao Polytechnic University, Macao, China
p1909810@mpu.edu.mo

Abstract. Currently, science and technology are changing rapidly, digitalization and intelligence are developing deeply, profoundly affecting economic development trends and social operation laws, and the digital era is coming. Governments around China are in full swing in the process of digital transformation of government. However, few studies have been conducted to evaluate the results of digital transformation of local governments. In order to solve the above problems and gain practical insights about the digital transformation of governments, this paper firstly constructs a government digital transformation evaluation system containing six evaluation indicators. Second, this paper uses a clustering algorithm called K-Means++ to evaluate the digital transformation outcomes of 31 provincial local governments in China. Based on the panel data obtained in 2020, we find that the optimal number of clusters is 4 (the silhouette coefficient is 0.372). This results in four levels of government digital transformation evaluation: high, medium-high, medium-low, and low. The results of this paper provide meaningful theoretical insights for local governments at all levels to enact digital transformation policies and develop relevant government strategies.

Keywords: Digital transformation · Clustering · Data mining

1 Introduction

At present, science and technology are changing rapidly, and digitalization and intelligence are developing deeply, profoundly affecting economic development trends and social operation laws, and the digital era is coming (Agarwal et al. 2010). The 2020 United Nations E-Government Survey Report shows that the pace of digital transformation of governments is accelerating globally. Countries around the world are competing to develop digital transformation strategies, and government digital transformation is gaining widespread attention (Gil-Garcia et al. 2018, Hinings et al. 2018). Government digital transformation is a systematic, collaborative and leading the reform of ideological innovation, business process innovation, organisational innovation and information technology innovation (Janowski et al. 2015). The digital transformation of the government is to carry out the creative digital transformation of the entire government operation model, focusing on business synergy, data integration and technology intensification, promoting

G. Guan et al. (Eds.): ICBDSS 2022, AHCS 8, pp. 134–144, 2023.
https://doi.org/10.2991/978-94-6463-064-0_15

the digitisation of government work processes and work results, and then changing government operation processes, governance methods and even organisational structures (Klievink et al. 2009, Kretschmer 2020, Majchrzak et al. 2016).

In recent years, the Chinese government has taken the construction of a smart city as an opportunity to promote innovation in the public service model, break down barriers between organisations, platforms and data, promote inter-departmental, cross-system and cross-discipline collaboration in government affairs, and launch the "GOV.MO Account (unified public services for every household)" to continuously improve administrative management methods (Matt et al. 2015, Mergel et al. 2019). And make government decision-making more scientific, supervision more appropriate and services better. However, there are still some practical problems such as insufficient collaboration among government departments, insufficient data interconnection and interoperability, delayed reengineering of digital processes in government affairs, hidden dangers of digital security and cyber security, etc. (Pittaway and Montazemi 2020, Tangi et al. 2021, Vial 2019). It is crucial to figure out how to further strengthen the construction of digital society and digital government and enhance the digital intelligence of public services and social governance?

Therefore, this paper adopts a cluster analysis method to evaluate the digital transformation achievements of 31 provincial governments in China, which provides instructive and practical solutions for China's digital transformation and also provides useful theoretical exploration for the top-level design and implementation of digital government strategies.

The reset of this paper is organized as follows: Sect. 2 provides a detailed discussion of the methodology of this paper. The obtained results are discussed in Sect. 3. Finally, Sect. 4 summarizes the full paper and suggests future research directions.

2 Methods and Materials

2.1 Digital Transformation Indicators

In order to better evaluate the digital transformation achievements of the Chinese government, this paper selects the latest relevant data of 31 Chinese provinces in 2020 from sources such as the Statistical Yearbook, the White Paper on China's Digital Government Development, and the White Paper on China's Digital Government Construction, as shown in Table 1.

Table 1. Indicators of digital transformation

Indicator	Symbols
Regional Gross Domestic Product	x_1
Disposable income per capita	x_2
Digital Economy Development	x_3
Digital Government Construction	x_4
Digital Government Development	x_5
Administrative Service Capability	x_6

Table 2. Statistical description of the research data

	Mean	Std	Min	Max
x_1	32578.63	26657.79	1902.70	11115160
x_2	32100.48	12656.40	20335.00	72232.00
x_3	29.62	13.34	8.00	65.30
x_4	66.36	82.15	22.90	504.00
x_5	57.42	10.63	39.60	76.70
x_6	85.60	5.86	73.15	95.38

A statistical description of the data used in this paper is presented in Table 2. Obviously, the data obtained in this paper have a large discrepancy. In order to obtain more reasonable results, later we will first pre-process the data in order to normalize them to [0,1], as shown in Eq. (1).

$$x_{ij}^{'} = \frac{max\{x_{ij}, \ldots, x_{nj}\}}{max\{x_{ij}, \ldots, x_{nj}\} - min\{x_{ij}, \ldots, x_{nj}\}}, \tag{1}$$

where x_{ij} represents the value of the j-th indictors of the i-th sample.

2.2 Clustering

2.2.1 K-Means Algorithm

K-means is an efficient clustering algorithm that is widely used in many research areas (Rezaee 2021). Its main idea can be expressed as follows:

Step-1: Set the number of clusters (K).

Step-2: K initial clustering centres are randomly generated.

Step-3: Each data object is assigned to the cluster with the highest similarity.

Step-4: Find the reasonable cluster centres by iteration.

Step-5: Form clusters around the clustering centres.

The algorithm has the advantages of easy understanding, simple implementation, fast convergence, and good adaptability to sparse matrix data. However, a well-known drawback of this method is the random selection of initial clustering centres. This can easily lead to multiple initial clustering centres in the same category, especially when the data is complex. In addition, this initial point selection makes it difficult to find the globally optimal cluster centres as well. This causes the algorithm to easily converge to a local optimum, resulting in unsatisfactory clustering results.

2.2.2 K-Means++ Algorithm

To overcome the above-mentioned drawbacks of the K-Means algorithm, the researchers further proposed an improved method called the K-Means++ algorithm (Wu 2021).

Compared with the traditional K-Means method, the K-Means++ method makes the distance between the initial clustering centres as large as possible to achieve the globally optimal clustering result. Its optimization strategy for the initial clustering centres is simple. In the first step, a data object is randomly selected from the dataset as the first clustering centre C_1. The second step is to select the initial clustering centres according to the probability formula until the initial clustering centres are obtained. The flow chart of the K-Means++ algorithm is shown as Algorithm 1.

2.2.3 Optimal K Value

In the K-Means++ algorithm, the number of clusters (i.e., K) is a hyperparameter that needs to be set in advance, and the size of this parameter greatly affects the clustering performance. Therefore, before applying the K-Means++ algorithm to the data in this paper, we need to choose the optimal K value. A common method is the silhouette coefficient method (Xin 2021). For sample i, its silhouette $s(i)$ is obtained as:

$$s(i) = \frac{b(i) - a(i)}{max\{a(i), b(i)\}}, \tag{2}$$

where $a(i)$ is the distance between sample i and other samples in the cluster, and $b(i)$ is the average distance between sample i and the samples in the nearest cluster. By averaging the silhouette of all samples in the data set, the silhouette coefficient is shown in Eq. (3):

$$s = \frac{\sum_{i=1}^{N} s(i)}{N} \tag{3}$$

where N is the total number of samples in the dataset. The closer the silhouette coefficient is to 1, the better the clustering effect is; the closer it is to -1, the worse the clustering effect is.

Further, this paper utilizes the Euclidean distance as a measure of sample similarity, which is expressed as shown below:

$$d(x, y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2} \tag{4}$$

where $d(x, y)$ is the similarity of sample x and sample y, x_i and y_i is the vector of sample x and sample y, respectively.

3 Results and Discussions

3.1 Pre-processing

Since the distance formula used is more sensitive to the order of magnitude between the sample data. Therefore, in this paper, the samples were normalized by the method mentioned above before embarking on further research. Table 3 shows the results of the statistical description of the sample after normalization.

Table 3. Statistical description of the pre-processed data

	Mean	Std	Min	Median	Max
x_1	0.28	0.24	0.00	0.21	1.00
x_2	0.23	0.24	0.00	0.15	1.00
x_3	0.38	0.23	0.00	0.32	1.00
x_4	0.09	0.17	0.00	0.06	1.00
x_5	0.48	0.29	0.00	0.51	1.00
x_6	0.56	0.26	0.00	0.54	1.00

Algorithm 1: K-Means++

```
Input: K, dataset
Output: center, indicator
  Function CLUSTERING
   center = null
   indictor = null
   m = the number of data samples in the dataset
   for i=0 → k do
    if i == 0 then
      temp =random(m)
      center[i]=dataset[temp]
    else
    for j=0 → m do
      for h=0 → i do
        distances[j][h]=cosine distance of the dataset[j] and center[h]
      end for
     end for
    Min[j]=min(distance[j])
    Sum+=Min[j]
    Random_number=random (Sum)
    for j=0 → m do
     Random_number=Random_number-Min[j]
     if Random_number<0 then
      cluster[i]=dataset[j]
     end if
    end for
    Proceed as with the standard K-Means algorithm
   end if
  end for
  return center, indictors
  end function
```

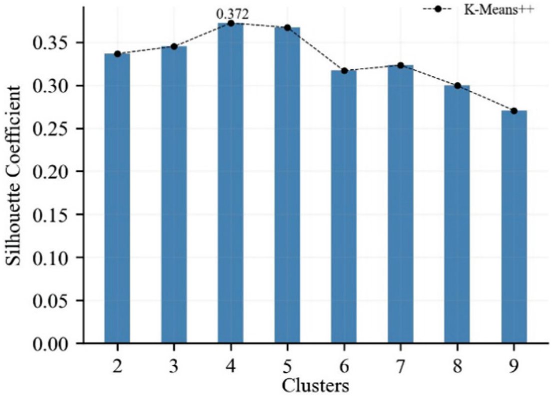

Fig. 1. Clusters.

3.2 Optimal K Value via Silhouette Coefficient

After completing the normalized evaluation of the samples, we further select the optimal number of clusters (i.e., K) using a method based on the contour coefficients. Specifying that K belongs to a positive integer between [2,10], we obtain the silhouette coefficient results as shown in Fig. 1.

From Fig. 1, we can clearly see that the contour coefficient keeps rising when K is taken between [2,4], indicating that the clustering effect becomes gradually better at this time. And when the contour coefficient is greater than 4, the silhouette coefficient gradually decreases, indicating that the clustering effect decreases at this time instead. When $K = 4$, the silhouette coefficient reaches the highest value (i.e., 0.372). This indicates that the 4 clusters are the most compatible in the research scenario of this paper. Therefore, in the later analysis, we set the number of clusters to 4.

3.3 Clustering Based on K-Means++

Through the above analysis, we know that the optimal clustering effect can be achieved when the number of clusters is 5. Therefore, in this paper, we set the number of clusters to 5 for the evaluation of government digital transformation. The clustering results we obtained are further shown in Table 4.

As can be seen in Table 4, cluster 2 has the highest number of provinces (13) and cluster 4 has the lowest number of provinces (1). In addition, cluster 1 has the highest government service capacity (i.e., x_6) with a performance of 0.96, which is much higher than the other clusters. Cluster 2 has a more average performance in all aspects. Cluster 3 has better digital government transformation results, performing better in both x_5 and x_6.

Therefore, we can denote them as High, Medium-High, Medium-Low, and Low government digital transformation, respectively, based on the clustering results. We further expressed the above results in Table 5.

Table 4. Clustering results

	Cluster 1	Cluster 2	Cluster 3	Cluster 4
x_1	0.63	0.13	0.29	0.38
x_2	0.68	0.10	0.18	0.15
x_3	0.80	0.20	0.39	0.43
x_4	0.10	0.04	0.07	1.00
x_5	0.86	0.20	0.62	0.61
x_6	0.96	0.35	0.61	0.67
Count	5	13	12	1

Table 5. Government digital transformation evaluation result.

Level	Province
High	Beijing; Shanghai; Jiangsu; Zhejiang; Guangdong
Medium-High	Hubei
Medium-Low	Tianjin; Anhui; Fujian; Jiangxi; Shandong; Henan; Hunan; Guangxi; Hainan; Chongqing; Sichuan; Guizhou
Low	Hebei; Shanxi; Inner Mongolia; Liaoning; Jilin; Heilongjiang; Yunnan; Tibet; Shaanxi; Gansu; Qinghai; Ningxia; Xinjiang

From the results in Table 5, it can be found that most of the provinces with better economic development get better government digital transformation results (e.g., Beijing, Shanghai, Zhejiang, etc.), which may be due to better fiscal revenues and more government funds to invest in the digital transformation process.

In contrast, some inland, more economically underdeveloped provinces (e.g., Tibet and Gansu) fall into the lowest government digital transformation ratings. These provinces inevitably fall behind in supporting the digital transformation of local governments because of their inland location, harsher economic development conditions, and lower local fiscal revenues than coastal provinces.

3.4 Policy Suggestions

Based on the results of our research, we further proposed the following policy suggestions:

(1) In an effort to improve the results of digital transformation of local governments, cross-level and cross-departmental collaboration should be strengthened to build a comprehensive, advanced and reliable digital collaboration and governance system. It is crucial to strengthen top-level design and overall planning, adhere to the principles of system governance, interconnection and collaborative governance, break

down departmental boundaries, and break the management and service approaches of different functional departments that operate separately and in isolation. The government should also open channels for the orderly flow, sharing and exchange, and secure application of data between government departments. It is necessary to deepen government business applications, incorporate the processes of government decision support, economic regulation, statistical analysis, official document flow, command and dispatch, and information disclosure into the digital track, and steadily improve the level of online, mobile, visualization and integration of government operations.

(2) Accelerate the integration of network, data and services, and build a "digital network cloud" integrated government service platform system is imminent. The government should adhere to system integration, coordination and efficiency, accelerate the migration of various government information systems to the cloud, and achieve cloud coverage of all application system deployment and data storage. We should pay attention to the needs of enterprises and residents as the guide, and constantly upgrade and build the "Government Affairs" APP, which brings together all the approval matters. Household" APP to centralize all approval services, public services and related social services to create a one-stop digital public service "master portal" for enterprises and residents.

(3) Local governments at all levels should strengthen the application of economic and social governance system and build a multi-level and systematic digital application and supervision system. Focusing on the digital economy, social governance and public services, it will continuously expand the application of business scenarios, build a basic mainline application system of digital government, and promote the joint development of digital economy and digital society. It is of inevitable significance for the government to improve laws and regulations on data security, strengthen security checks on data collection and storage, open sharing, circulation and transaction, and confirmation of use, and ensure data security. It is necessary to introduce and implement measures to strengthen the supervision and restraint on the use of citizens' privacy for commercial activities, and effectively protect the rights and interests of citizens' personal information.

4 Conclustions

At present, science and technology are changing rapidly, digitalization and intelligence are developing deeply, profoundly affecting the economic development trend and social operation law, and the digital era is coming. Governments across China are in full swing with the digital transformation of government. However, there is little research on the evaluation of the results of digital transformation of local governments. In order to solve the above problems and gain a practical understanding of government digital transformation, this paper firstly constructs a government digital transformation evaluation system (Regional Gross Domestic Product, Disposable income per capita, Digital Economy Development, Digital Government Construction, Digital Government Development, Administrative Service Capability) containing six evaluation indicators.

Secondly, this paper evaluates the digital transformation outcomes of 31 provincial local governments in China using K-Means++ clustering algorithm. Based on the panel data obtained in 2020, we find that the optimal number of clusters is 4 (with a silhouette coefficient of 0.372). This leads to four levels of government digital transformation evaluation: high, medium-high, medium-low, and low. It shows that Beijing; Shanghai; Jiangsu; Zhejiang and Guangdong are in the highest digital transformation.

The results of this paper provide meaningful theoretical insights for local governments at all levels to enact digital transformation policies and develop relevant government strategies.

References

Agarwal, R., Gao, G., DesRoches, C., & Jha, A. K. (2010). Research commentary—The digital transformation of healthcare: Current status and the road ahead. Information systems research, 21(4), 796-809.

Gil-Garcia, J. R., Dawes, S. S., & Pardo, T. A. (2018). Digital government and public management research: finding the crossroads. Public Management Review, 20(5), 633-646.

Hinings, B., Gegenhuber, T., & Greenwood, R. (2018). Digital innovation and transformation: An institutional perspective. Information and Organization, 28(1), 52-61.

Janowski, T. (2015). Digital government evolution: From transformation to contextualization. Government information quarterly, 32(3), 221-236.

Klievink, B., & Janssen, M. (2009). Realizing joined-up government—Dynamic capabilities and stage models for transformation. Government information quarterly, 26(2), 275-284.

Kretschmer, T., & Khashabi, P. (2020). Digital transformation and organization design: An integrated approach. California Management Review, 62(4), 86-104.

Majchrzak, A., Markus, M. L., & Wareham, J. (2016). Designing for digital transformation. MIS quarterly, 40(2), 267-278.

Matt, C., Hess, T., & Benlian, A. (2015). Digital transformation strategies. Business & information systems engineering, 57(5), 339-343.

Mergel, I., Edelmann, N., & Haug, N. (2019). Defining digital transformation: Results from expert interviews. Government information quarterly, 36(4), 101385.

Pittaway, J. J., & Montazemi, A. R. (2020). Know-how to lead digital transformation: The case of local governments. Government Information Quarterly, 37(4), 101474.

Rezaee, M. J., Eshkevari, M., Saberi, M., & Hussain, O. (2021). GBK-means clustering algorithm: An improvement to the K-means algorithm based on the bargaining game. Knowledge-Based Systems, 213, 106672.

Tangi, L., Janssen, M., Benedetti, M., & Noci, G. (2021). Digital government transformation: A structural equation modelling analysis of driving and impeding factors. International Journal of Information Management, 60, 102356.

Vial, G. (2019). Understanding digital transformation: A review and a research agenda. The journal of strategic information systems, 28(2), 118-144.

Wu, J., Shi, L., Yang, L., Niu, X., et al. (2021). User Value Identification Based on Improved RFM Model and-Means++ Algorithm for Complex Data Analysis. Wireless Communications and Mobile Computing, 2021.

Xin, M. A., Gang-long, D. U. A. N., Jian-ren, W. A. N. G., & Hong-quan, X. U. E. (2021). Research on Airline Customer Clustering Based on Improved Silhouette Coefficient Method. Operations Research and Management Science, 30(1), 140.