International Journal of Information Technology & Decision Making
Vol. 20, No. 1 (2021) 261–285

DOI: 10.1142/S0219622020500479

Data-Driven Evidential Reasoning Method for Evaluating e-Government Performance

Ying Yang[*,§], Rui-Xue Lu[*,¶], Min Xue[*,‖], Zhi-Qin Shou[*,**],
Jian-Bo Yang[†,††] and Lei Fu[‡,‡‡]

[*]*School of Management*
Hefei University of Technology
Hefei, 230009, Anhui, P. R. China

[†]*Alliance Manchester Business School*
The University of Manchester
Manchester M15 6PB, UK

[‡]*College of Economics and Management*
Anhui Agricultural University
Hefei 230061, Anhui, P. R. China
[§]*yangying@hfut.edu.cn*
[¶]*lrx@mail.hfut.edu.cn*
[‖]*xuehfut@163.com*
[**]*szqychut@mail.hf.ah.cn*
[††]*freelyfreely@qq.com*
[‡‡]*jian-bo.yang@mbs.ac.uk*

Published 19 January 2021

The construction of electronic government (e-government) systems is a process of continuous improvement. It is necessary to evaluate the performance of e-government systems regularly to improve the services provided by government agencies and enhance the exchange of information between governments and citizens. Evaluating e-government performance based on citizens' experience is a multiple criterion decision making (MCDM) problem under uncertainty, where assessments are qualitative, and many e-government system users are involved. Deriving criterion weights from a large amount of evaluation data is rarely discussed in previous MCDM studies. This paper proposes a data-driven evidential reasoning (DDER) method for evaluating e-government performance. A criteria framework from the citizens' experience perspective, including service guide clarity, site usability, information sharing, documentation, and the availability of e-services, is proposed. Belief structures are used to portray uncertain assessments from e-government system users. The criterion weights are learned from the data by minimizing the dissimilarity between the aggregation assessments of the alternatives on each criterion and citizens' historical observations on a whole. A case study is conducted in 16 cities of Anhui province in China to evaluate the performance of e-government systems. The ranking results verified the applicability and effectiveness of the proposed method.

Keywords: E-government performance evaluation; data-driven; evidential reasoning; weight learning.

[§]Corresponding author.

1. Introduction

Electronic government (or e-government) is the application of information and communication technologies to facilitate government functions and procedures. E-government has enhanced the work efficiency and reduced management costs in recent years.[1] The development of e-government needs continuous improvement by regularly evaluating its performance from citizens' perspectives. Meanwhile, the popularization of smart devices has significantly increased internet penetration and brought much more convenience than ever for citizens to access e-government services directly. As the users of e-government systems, citizens' demands for e-government services gradually increase. Experiencing failures with the systems may affect users' acceptance of e-government and reduce their satisfaction. Therefore, it is essential to evaluate e-government performance considering users' experience and satisfaction.

Previous studies on the approaches for evaluating e-government performance can be categorized into two kinds: one is the data envelopment analysis (DEA) method, and the other is the multiple criteria decision-making (MCDM) method. The DEA method, as a nonparametric programming approach, is frequently used to evaluate the relative efficiency of e-governments. Antonić and Šegota used the DEA method[2] to evaluate the performance of e-government in the Republic of Croatia by considering eight inputs indices and three outputs indices. Wu and Guo[3] adopted the nonradial slack-based efficiency measurement in DEA to evaluate the government performances of official websites in China. The influential factors related to e-government performance are identified from the input–output perspective by the DEA method. However, their relative importance for evaluating and ranking e-government performance is neglected. The influential factors related to e-government systems are identified as criteria and their relative importance are defined as the criterion weights in the MCDM methods. Byun and Finnie used the analytic hierarchy process (AHP) method[4] to assess the e-government website usability of Australian state governments. Nilashi *et al.* analyzed and ranked the government portal websites in Malaysia based on citizens' desires by using the technique for order preference by similarity to an ideal solution (TOPSIS) method.[5] To handle the uncertainty and ignorance in evaluating processes, Hossain *et al.* developed a belief rule-based expert system with evidential reasoning (ER) to evaluate the e-government of Bangladesh.[6] Some other studies evaluated the e-government performance by using fuzzy MCDM methods.[7,8]

However, the process of evaluating e-government performance is complex and involves multiple criteria with a qualitative nature. These qualitative assessments come from the subjective judgements of e-government system users and are expressed with uncertainty in different linguistic forms, such as yes or no and satisfied or extremely satisfied. Moreover, different sets of evaluation grades may be defined for the qualitative criteria to facilitate data collection in the evaluation process. It is essential to establish a general framework for evaluating e-government performance using subjective assessments under uncertainty.

The ER approach has been developed for MCDM under uncertainty in an integrated way. The qualitative assessments under different grades can be represented by a belief structure. In recent years, the ER approach has been applied to address practical decision problems in organizational assessment. Kong *et al.*[9] used the ER approach to evaluate the medical quality of hospitals. Fu *et al.*[10] proposed a new MCDM method in the ER context to select strategic partners for tax information systems. Determining the weights of the evaluation criteria is vital in the ER method. Due to the increasing development of information technologies, the evaluation circumstances have changed. Many citizens participate in decision-making processes and classical decision patterns encounter challenges. How to utilize the large amount of historical evaluation data and derive criterion weights from the data has become a research focus in the ER approach. However, it is rarely discussed in previous studies.

The motivation of this study is to develop a data-driven decision-making method in the ER context for evaluating e-government performance. A great deal of evaluation data is collected from citizens' responses via a survey. The criterion weights are derived from the data using an optimization model. The contributions of this paper are summarized as follows. (1) A criterion system for evaluating e-government performance that considers users' experience is proposed. (2) A data-driven evidential reasoning (DDER) method is proposed to generate the final ranking of different cities' e-government performance. An optimization model that minimizes the difference among the aggregated assessments of the alternatives to each criterion and decision makers' historical observations as a whole is developed to learn the weights of criteria. (3) A case study on evaluating the e-government performance of 16 cities in Anhui province, China is conducted.

The rest of this paper is organized as follows. Section 2 reviews the existing criterion systems and methods for evaluating e-government performance. Section 3 introduces the proposed DDER method for evaluating e-government performance. In Sec. 4, a case study is conducted in 16 cities in Anhui province, China. Comparative analysis is conducted in Sec. 5 to illustrate the effectiveness of the proposed method. Finally, the paper is concluded in Sec. 6.

2. Literature Review

2.1. *Criteria for evaluating e-government performance*

Many researchers have developed different criterion systems to evaluate the e-government performance of different countries. These criteria frameworks are listed in Table 1 according to the methodologies used, theoretical grounds, and countries.

From the input–output perspective of the DEA method, Wu and Guo[3] measured the e-government performance of provincial government websites in China from seven aspects, including economic, human, equipment maintenance, information

Table 1. Some criteria frameworks.

Criteria framework	Methodologies	Theoretical groundings	Countries
Economic factor, human factor, equipment maintenance factor, information public factor, communication factor, and public service factor	The DEA method	Input–output perspective	China[3]
Information and communication technologies, human potential, information security, information and documentation base, and quality of e-services	The DEA method	Input–output perspective	The Republic of Croatia[2]
Project construction, information security management, special construction, transparency of government affairs, and informationized ability	Empirical study	Sociotechnical model and stakeholder theory	China[11]
Information quality, security, communication, website aesthetics, website design, and access	Empirical study	Information system success model	The South Africa[12]
Trust, accessibility, awareness of e-services, quality of e-services, computer anxiety, customer expectations, and security/privacy	Empirical study	Expectation confirmation theory	Pakistan[13]
Accessibility, citizen engagement, development of trust, responsiveness, dialogue, and quality of information and services	Empirical study	Public values delivery perspective	Sub-Saharan Africa[14]

publicity, communication, public service factors, and the employment service index, by using the slack-based efficiency measurement. Antonić and Šegota[2] selected eight input factors and three output factors to evaluate the local e-government performance of the Republic of Croatia. These factors covered five aspects about e-government performance. They are information and communication technologies, human potential, information security, the information and documentation base, and the availability of e-services.

Based on the sociotechnical model and stakeholder theory, Shan *et al.*[11] proposed an evaluation model with five dimensions, including project construction, information security management, special construction, the transparency of government affairs, and informationized ability, to evaluate the local e-governments in China's eastern coastal areas.

Based on the information system success model, Kaisara and Pather[12] proposed a criteria framework, which includes information quality, security, communication, website aesthetics, website design, and access, for evaluating e-government service quality in South Africa.

According to expectation confirmation theory, Babur *et al.*[13] proposed seven key factors from the technical, behavioral, and economical dimensions for evaluating citizens' e-satisfaction in Pakistan. Security and privacy and accessibility are technical factors. Computer anxiety, customer expectations, trust, and awareness of e-government public services are behavioral factors. The quality of e-government public services is an economical factor.

From the public values' perspective of e-government websites, Verkijika and De[14] proposed an evaluation criteria system with six aspects, including accessibility, citizen engagement, the development of trust, responsiveness, dialogue, and the quality of information and services, for evaluating the e-government websites in sub-Saharan Africa. The evaluation results show that users were highly unsatisfied with the performance of the e-government websites when public values delivery was the focus of e-government development.

In conclusion, the criterion systems are different according to the methodologies and the theoretical grounds. However, they are mostly concerned with technical and social evaluation aspects such as the quality of e-services, the security and privacy of services, their ease of use, users' expectations, and so on. These evaluation aspects provide a basis for evaluating the e-government performance of local governments in China.

2.2. *Approaches for evaluating e-government performance*

2.2.1. *MCDM approaches for evaluating e-government performance*

The evaluation of e-government performance can be considered as an MCDM problem where the best alternative should be selected among a set of decision alternatives according to multiple conflicting criteria.

Byun and Finnie[4] adopted the AHP method to assess the website usability of the portals of Australian state governments. Six main criteria and 59 sub-criteria are synthesized to assess the overall priority of the criteria and portals. The weights of the criteria are obtained through pairwise comparisons between criteria according to the judgement of experts. Since the judgements may be inconsistent, measuring the consistency is a concern of the AHP method.

Nilashi *et al.* analyzed and ranked the government portals in Malaysia by using the TOPSIS method.[5] The portals are ranked according to the distances both from an ideal solution and a negative-ideal solution. A set of factors influencing Malaysian government portals based on citizens' desires is identified.

To handle the uncertainty and ignorance in evaluations, Hossain *et al.* developed a belief rule-based expert system to evaluate the e-government of Bangladesh.[6] The ER approach is adopted to aggregate the belief rules activated by the input factors of the system. Based on both Dempster–Shafer theory and decision theory, the ER approach used belief structures to depict both quantitative and qualitative attributes under uncertainties.[6] The approach can be used for information aggregation to obtain the rank of alternatives.

Belief structures are constructed based on a set of grades and can describe uncertain evaluations in MCDM problems. Both the cardinal and ordinal evaluations of alternatives can be well represented by belief structures.[15] Belief structures have been used for information fusion and decision making with uncertainties to reduce information loss. Therefore, the ER method with belief structures is adopted in this study to evaluate e-government performance and belief decision matrices are developed to represent uncertain assessments from e-government system users.

2.2.2. *Weight determination approaches*

The weights of criteria play a very significant role in the decision-making process. They reflect the relative importance of different criteria and significantly affect the evaluation results. In addition to the criteria framework required for evaluating e-government, the weights of criteria are essential for obtaining a rational evaluation result. Objective, subjective, and hybrid methods have been proposed to determine the criterion weights in classical decision patterns.

Subjective methods, including the ranking order method,[16] the direct rating method,[17] the eigenvector method,[18] the Delphi method[19] and others, depend on the preferences of decision makers. Since there are many criteria for evaluating e-government performance, it is difficult for a decision maker to present criterion preferences subjectively.

In objective methods, criterion weights are determined using the intrinsic information contained in assessments. The entropy method,[20] the standard deviation (SD) method,[21] the criteria importance through intercriteria correlation (CRITIC) method,[21] and the correlation coefficient and standard deviation (CCSD) method[22] are objective methods. They mainly model the difference among the assessments of alternatives on each criterion. Distance metric learning methods are also used for weighting criteria.[28,29] Kou *et al.* compared different weighting approaches in feature selection using classification performance and stability measures.[30] The weights are obtained from the information in individual or collective decision matrices using mathematical models, which can generate more rational and objective results than those of subjective methods.

Hybrid methods employ both decision-makers' subjective preferences and objective decision matrix synthetically to produce criterion weights. Liu *et al.* integrated the statistical variance, simple additive weighting, and Delphi-AHP to determine the integrated weights of attributes to select cloud computing service vendors. The hybrid methods not only overcome the randomness of subjective methods, but they also avoid the one-sidedness of objective methods.[23]

These weighting approaches mainly concern the discriminating power of each criterion contained in assessments. The whole observations on alternatives from decision-makers, such as a whole assessment for an alternative, are often neglected. When a great number of citizens are engaged in evaluating e-government performance, the observations may reveal either explicitly or implicitly information about criteria. Therefore, the large amount of observations in evaluation processes should be used to obtain the most distinctive criteria through supervised learning algorithms.

3. A Data-Driven ER Method for Evaluating e-Government Performance

In this section, a DDER method for evaluating e-government performance is proposed. It includes four procedures. First, belief structures are used to model the MCDM problem. Then, the assessments from e-government system users of each criterion are transformed from different assessment frames into a unified format with the same grades. Third, an optimization model is designed for learning the criterion weights. Last, the assessments of each criterion are aggregated into final assessments using the obtained weights via the ER algorithm. A flow diagram that explains the procedures is shown in Fig. 1.

3.1. *Problem modeling using belief structures*

Suppose that there are K different cities in local e-governments, denoted by C_k $(k = 1, \ldots, K)$. These cities are to be assessed in terms of T first-level criteria denoted by E_j $(j = 1, \ldots, T)$. Each first-level criterion includes Q second-level criteria denoted by e_{ji} $(i = 1, \ldots, Q)$. These criteria form a three-level hierarchy for e-government performance assessment, as shown in Fig. 2. All the relevant notations in this paper are shown in Table 1.

The relative weights of first-level criteria are symbolized by W_j $(j = 1, \ldots, T)$, $0 \leq W_j \leq 1$ and $\sum_{j=1}^{T} W_j = 1$. The relative weights of second-level criteria are denoted by w_{ji} $(i = 1, \ldots, Q)$, where $0 \leq w_{ji}(i = 1, \ldots, Q) \leq 1$ and $\sum_{i=1}^{Q} \omega_{ji} = 1$. To evaluate the e-government performance of the K cities on each criterion, belief structures are used to represent the assessments from e-government system users in different cities, and these assessments are based on a frame of discernment denoted by $\Omega = \{H_1, H_2, H_3, \ldots, H_N\}$. The frame of discernment includes a set of grades that are ordered in increasing order from worst to best. Each grade is mutually exclusive and collectively exhaustive. The utilities of grades denoted by $u(H_n)$ $(n = 1, \ldots, N)$ satisfy the constraint: $0 \leq u(H_1) \leq u(H_2) \leq \cdots \leq u(H_N) \leq 1$.

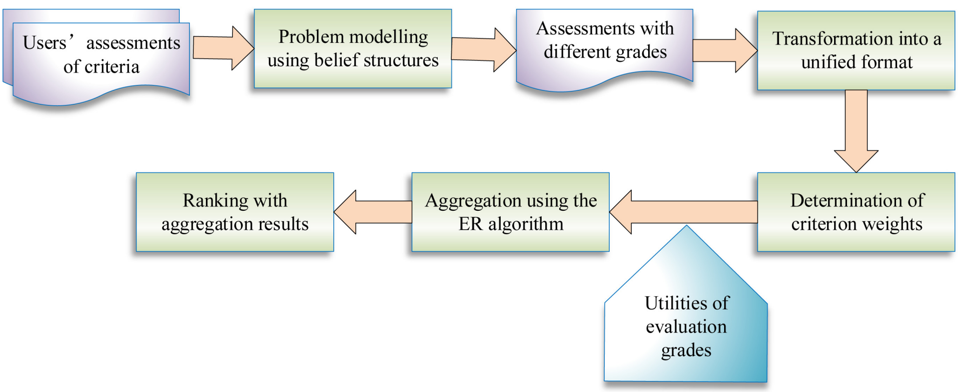

Fig. 1. The procedures of the proposed DDER method.

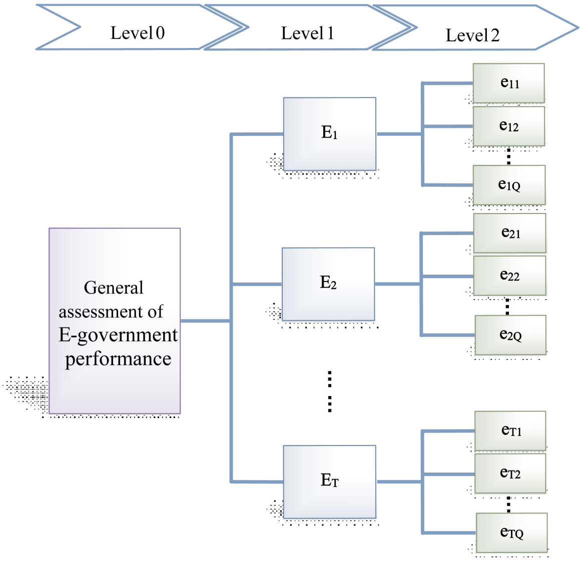

Fig. 2. A three-level hierarchy for evaluating e-government performance.

Suppose a city's e-government performance C_k is assessed using a second-level criterion e_{ji} with a grade H_n with a degree of belief $\beta_{n,ji}(C_k)$. Then, a belief structure represented by

$$B(e_{ji}(C_k)) = \{(H_n, \beta_{n,ji}(C_k)), n = 1, \ldots, N; (\Omega, \beta_{\Omega,ji}(C_k))\},$$

where $\beta_{n,ji}(C_k)$ satisfies $0 \leq \beta_{n,ji}(C_k) \leq 1$ and $\sum_{n=1}^{N} \beta_{n,ji}(C_k) \leq 1$, is used to profile the assessment. In addition, $\beta_{\Omega,ji}(C_k) = 1 - \sum_{n=1}^{N} \beta_{n,ji}(C_k)$ represents the degree of global ignorance.[10]

3.2. *Transformation of different assessment grades*

Different criteria may be assessed on different sets of grades based on the requirements in a real decision problem. These different sets of assessment grades are needed to be transformed into a unified format before the ER algorithm is applied to aggregate assessments. Suppose that an original assessment is collected and denoted as $\{H_{d,ji}, \beta_{d,ji}\}$ $(d = 1, \ldots, N)$. The original assessment $\{H_{d,ji}, \beta_{d,ji}\}$ can be transformed to an equivalent expectation $\{H_t, \beta_{t,ji}\}$ $(t = 1, \ldots, N)$ by a utility-based transformation method[24] as follows:

$$\beta_{t,ji} = \begin{cases} \sum\limits_{n\in\pi_t} \beta_{d,ji}\tau_{t,n}, & t = 1, \\ \sum\limits_{n\in\pi_{t-1}} \beta_{d,ji}(1-\tau_{t-1,n}) + \sum\limits_{n\in\pi_t} \beta_{d,ji}\tau_{t,n}, & 2 \leq t \leq N-1, \\ \sum\limits_{n\in\pi_{t-1}} \beta_{d,ji}(1-\tau_{t-1,n}), & t = N, \end{cases} \tag{3.1}$$

$$u(H_t) \le u(H_{d,ji}) \le u(H_{t+1}), \quad \tau_{t,n} = \frac{u(H_{t+1}) - u(H_{d,ji})}{u(H_{t+1}) - u(H_t)}, \tag{3.2}$$

$$\pi_t = \begin{cases} \{n|u(H_t) \le u(H_{d,ji}) < u(H_{t+1}), n = 1, \ldots, N_i\}, & t = 1, \ldots, N-2, \\ \{n|u(H_t) \le u(H_{d,ji}) \le u(H_{t+1}), n = 1, \ldots, N_i\}, & t = N-1. \end{cases} \tag{3.3}$$

Then, the individual assessments $B(e_{ji}(C_k))$ $(j = 1, \ldots, T,\ i = 1, \ldots, Q,\ k = 1, \ldots, K)$ can be combined using the ER algorithm to generate the aggregated assessment $B(E_j(C_k)) = \{(H_n, \beta_{n,j}(C_k)), n = 1, \ldots, N;\ (\Omega, \beta_{\Omega,j}(C_k))\}$, where $\beta_{\Omega,j}(C_k)$ represents the aggregated global ignorance. Each aggregated assessment is considered as the assessment of first-level criterion $E_j (j = 1, \ldots, T)$. Then, these aggregated assessments are further aggregated by the ER algorithm with the weights of the first-level criteria W_j to generate a final assessment $B(Y(C_k)) = \{(H_n, \beta_n(C_k)), n = 1, \ldots, N;\ (\Omega,\ \beta_\Omega(C_k))\}$.

3.3. *Determination of criterion weights*

Generally, when a survey involves a large numbers of e-government system users, the users are required to present not only their evaluation information on individual criteria but also their historical observations on a whole. It means that the users' satisfaction with the second-level criteria can be validated by their overall satisfaction on the corresponding first-level criteria. This makes it possible for the weights of the second-level criteria to be obtained by learning from the difference between the overall and individual preferences of users.

Due to the relationship between the historical observations and corresponding assessments of second-level criteria, the historical observations can be used to balance the relative importance of second-level criteria. Therefore, an optimization model that minimizes the difference between the citizens' historical observations on a whole and the aggregation of individual assessments on each criterion is developed. The process of determining the criterion weights is shown in Fig. 3.

3.3.1. *An optimization model for learning criterion weights*

In the data-driven context, the criterion weights that achieve rational and objective results can be obtained by learning from the evaluation information instead of the subjective preferences of decision makers. To conduct the learning process of second-level criteria, the assessments should satisfy the following assumption.

Assumption 3.1. For each first-level criterion $E_j (j = 1, \ldots, T)$, the aggregated assessments $B^a(y_j(C_k))$ of user $a (a = 1, \ldots, M)$ reflect the user's overall evaluation of all second-level criteria $e_{ji} (j = 1, \ldots, T, i = 1, \ldots, Q)$ under E_j.

The assessments of second-level criteria $e_{ji} (j = 1, \ldots, T, i = 1, \ldots, Q)$ represent the evaluations of first-level criterion E_j from different perspectives. Therefore, the aggregated assessment of E_j should reflect the comprehensive evaluation of all second-level criteria $e_{ji} (j = 1, \ldots, T, i = 1, \ldots, Q)$ related to E_j. In other words, the

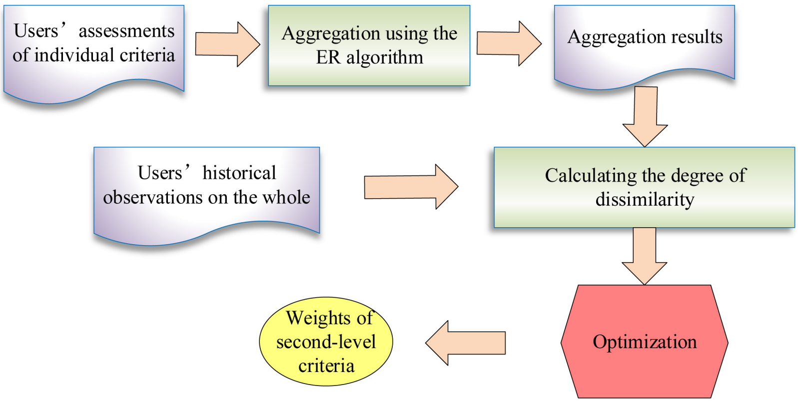

Fig. 3. The process of determining criterion weights.

aggregated assessment of $E_j(j = 1, \ldots, T)$ plays the role of verifying the consistency between the aggregated assessment of second-level criteria and users' historical observations on the corresponding first-level criterion.

Under Assumption 3.1, the assessments of second-level criteria $e_{ji}(j = 1, \ldots, T, i = 1, \ldots, Q)$ can be given independently. The independence among assessments is defined as follows.

Definition 3.1. For each first-level criterion $E_j(j = 1, \ldots, T)$, the independence among the individual assessments of each second-level criterion $e_{ji}(j = 1, \ldots, T, i = 1, \ldots, Q)$ means that the assessment of one second-level criterion will not be changed no matter whether the assessments of the other second-level criteria are known or not. As such, the assessment of the second-level criterion e_{ji} of the first-level criterion E_j will not be changed regardless of whether the assessments of other first-level criterion E_j are known or not.

Under Assumption 3.1 and Definition 3.1, the individual assessments of second-level criteria $e_{ji}(j = 1, \ldots, T; i = 1, \ldots, Q)$ can be given. They can be combined to obtain an aggregated result. Since the difference between the aggregated assessment of second-level criteria and the historical observation of the corresponding first-level criteria should be close as possible, it can be minimized to determine the appropriate weights of second-level criteria.

For each first-level criterion $E_j(j = 1, \ldots, T)$, $B^a(e_{ji}(C_k))$ $(a = 1, \ldots, M)$ is the assessment provided by user $a(a = 1, \ldots, M)$ of second-level criterion e_{ji} with respect to the kth city's e-government performance C_k. $B^a(S_j(C_k)) = \{(H_n, \beta_n^S(C_k)), n = 1, \ldots, N; (\Omega, \beta_\Omega^S)\}$ is the historical observation provided by user a on E_j with respect to the kth city's e-government performance C_k. Each user only evaluates one city's e-government performance.

For each first-level criterion $E_j(j = 1, \ldots, T)$, $B^a(e_{ji}(C_k))$ $(i = 1, \ldots, Q)$ can be aggregated by the ER algorithm to generate the final assessment of all second-level criteria $e_{ji}(i = 1, \ldots, Q)$, which is symbolized by $B^a(y_j(C_k))$. To learn the second-level criterion weights from the aggregated assessment $B^a(y_j(C_k))$ $(a = 1, \ldots, M)$ and the historical observation $B^a(S_j(C_k))$ $(a = 1, \ldots, M)$, a dissimilarity measure is defined as follows.

Definition 3.2. Suppose that the difference between the aggregated assessments $B^a(y_j(C_k))$ and the historical observation $B^a(S_j(C_k))$ is defined as[25]

$$\mathrm{GD}^a(y_j(C_k)) = \{(H_n, \bar{\beta}_n(C_k) = |\beta_n(C_k) - \beta_n^s(C_k)|), n = 1, \ldots, N\}, \tag{3.4}$$

where $\bar{\beta}_n(C_k)$ represents the absolute value of the difference among the belief structures of two assessments. Then, a dissimilarity measure between the two assessments is defined as

$$D^a(y_j(C_k)) = \sum_{n=1}^{N-1} \sum_{m=n+1}^{N} \bar{\beta}_n(C_k) \cdot \bar{\beta}_m(C_k) \cdot (u(H_m) - u(H_n)). \tag{3.5}$$

The dissimilarity measure includes both the difference between assessments and the difference between utilities. It reflects the overall difference between the aggregated assessment $B^a(y_j(C_k))$ and the historical observation $B^a(S_j(C_k))$ in terms of the weighted utility difference between each pair of different grades. $\bar{\beta}_n(C_k) \cdot \bar{\beta}_m(C_k)$ is the weight. Since $0 < u(H_1) < u(H_2) < u(H_3) < u(H_4) < u(H_5) < 1$, the dissimilarity measure is more reasonable for characterizing the overall differences between two assessments than the existing distance measures.

Following the idea of minimizing the average dissimilarity between the aggregated assessments and historical observations,[26] an optimization model is developed to learn the weights of criteria by using the dissimilarity measure.

$$\mathrm{MIN}\ P = \frac{\sum_{a=1}^{M} D^a(y_j(C_k))}{M}, \tag{3.6}$$

$$\text{s.t. } 0 \leq \omega_{ji}^* \leq 1, \tag{3.7}$$

$$\sum_{i=1}^{T} \omega_{ji}^* = 1. \tag{3.8}$$

Here, ω_{ji}^* $(j = 1, \ldots, T,\ i = 1, \ldots, Q)$ is the decision variables that represent the weights of the second-level criteria. The premise of applying this model is that the independence requirement should be satisfied, as presented in Definition 3.1. Then, the optimization model is solved by using a differential evolution (DE) algorithm.

3.3.2. *Differential evolution algorithm*

The DE algorithm is a parallel direct search method proposed by Storn and Price.[27] It is a practical minimization method with a good ability to handle nonlinear and nondifferentiable cost functions. It has parallelizability, usability, and good convergence properties. The DE algorithm has three operations: mutation, crossover, and selection. The DE algorithm is used to solve the optimization model mentioned above.

(1) Mutation

The mutation operation is used to generate new parameter vectors by adding the weighted difference between two population vectors to a third vector. Multiple Q-dimensional parameter vectors $x_{m,G}(m = 1, \ldots, M)$ are utilized as an initial population for each generation G. The initial vectors are generated randomly, which can cover all parameter spaces. Here, the vector $x_{m,G}(m = 1, \ldots, M)$ represents the decision variables $\omega_{ji}(j = 1, \ldots, T, i = 1, \ldots, Q)$ in the optimization model showed in Eqs. (3.6)–(3.8). M represents the population size and Q represents the dimension of a parameter space.

For each vector $x_{m,G}(m = 1, \ldots, M)$, the mutant vector $v_{m,G+1}(m = 1, \ldots, M)$ is obtained as follows:

$$\begin{gathered} v_{m,G+1} = x_{r1,G} + F \cdot (x_{r2,G} - x_{r3,G}), \\ m \neq r1 \neq r2 \neq r3, r1, r2, r3 \in \{1, 2, \ldots, M\}, \end{gathered} \tag{3.9}$$

where r_1, r_2 and r_3 are random indices that are mutually different. They differ from the running index m. In addition, F is a scale factor and it varies in the range from 0 to 2.[27]

(2) Crossover

The mutated vector $v_{m,G+1}(m = 1, \ldots, M)$ is then mixed with the target vector $x_{m,G}(m = 1, \ldots, M)$ to generate a trial vector $u_{m,G+1}(m = 1, \ldots, M)$ in the crossover operation. The trial vector is established at the end of the operation by the crossover rule as follows:

$$u_{m,G+1} = \begin{cases} v_{m,G+1}, & \text{if } (\text{rand}(0,1) \leq \text{CR}) \text{ or } z = r(i) \\ x_{m,G}, & \text{otherwise} \end{cases}, \quad z = 1, 2, \ldots, Q, \tag{3.10}$$

where CR represents the crossover constant in the range of [0,1]. Here, $r(i)$ denotes the randomly chosen integer in the set of $\{1, 2, \ldots, Q\}$ and z denotes the relevant component vector.

(3) Selection

According to the objective function in the optimization model, a cost function can be determined. To decide whether the trial vector can replace the target one in the following generation, the cost function value $P(u_{m,G+1})(m = 1, \ldots, M)$ of the trial vector is compared with the cost function value $P(x_{m,G})(m = 1, \ldots, M)$ of the target vector by using the greedy criterion as follows[24]:

$$x_{m,G+1} = \begin{cases} u_{m,G+1}, & \text{if } P(u_{m,G+1}) < P(x_{m,G}) \\ x_{m,G}, & \text{otherwise} \end{cases}. \tag{3.11}$$

Here, the cost function value represents the value P in Eq. (3.6). If the trial vector achieves a lower value than the target vector does, then the trial vector replaces the target vector in the next generation. Otherwise, the target vector is retained.

After that, the initial vectors are changed using these steps and regarded as the target vectors in the next generation. Then, these steps are repeated until the optimal solution is generated.

After determining the criterion weight $\omega_{ji}(j = 1, \ldots, T, i = 1, \ldots, Q)$ of all second-level criteria under each first-level criterion $E_j(j = 1, \ldots, T)$, all the assessments provided by users for each second-level criterion can be combined by the ER algorithm into the aggregated assessment, which is symbolized by $B(e_{ji}(C_k))$ $(j = 1, \ldots, T,\ i = 1, \ldots, Q,\ k = 1, \ldots, K)$. Then, $B(e_{ji}(C_k))$ can be aggregated to obtain the assessment $B(E_j(C_k))$ $(j = 1, \ldots, T,\ \ k = 1, \ldots, K)$ of first-level criterion $E_j(j = 1, \ldots, T)$. A new assessment matrix that includes K cities and T attributes can be established in the following equation:

$$[B(E_j(C_k))]_{K\times T} = \begin{pmatrix} B(E_1(C_1)) & \cdots & B(E_T(C_1)) \\ \vdots & \ddots & \vdots \\ B(E_1(C_K)) & \cdots & B(E_T(C_K)) \end{pmatrix}_{K\times T}. \tag{3.12}$$

Then, the CCSD method is applied to determine the weights of first-level criteria $E_j(j = 1, \ldots, T)$. The CCSD method can generate more comprehensive and convincing weights of attributes than those generated by the entropy method and the SD method. Furthermore, the method is clearer for modelling decision-making problems than the CRITIC method. The detailed process of CCSD method is shown in Appendix A.

3.4. *Fusion of evaluation information*

After obtaining the weights of the criteria, the assessments are aggregated by the ER algorithm to generate the final assessment$B(E_j(C_k))$ $(j = 1, \ldots, T,\ k = 1, \ldots, K)$ of each city.

Suppose that the assessments of the jth first-level criterion of city k are denoted by $B(E_j(C_k)) = \{(H_{n,}\beta_{n,j}(C_k)),\ n = 1, \ldots, N;\ (\Omega, \beta_{\Omega,j}(C_k))\}$ and the assessments of the $(j+1)$th first-level criterion of city k are denoted by $B(E_{j+1}(C_k)) = \{(H_{n,}\beta_{n,j+1}(C_k)),\ n = 1, \ldots, N;\ (\Omega, \beta_{\Omega,j+1}(C_k))\}$. The relative weights of first-level criteria are symbolized by W_j $(j = 1, \ldots, T)$. The ER algorithm used to aggregate assessments is shown as follows:

$$m_{n,j} = W_n\beta_{n,j}, \tag{3.13}$$

$$m_{H,j} = 1 - \sum_{n=1}^{N} m_{n,1} = 1 - W_j\sum_{n=1}^{N}\beta_{n,1}, \tag{3.14}$$

$$m_{H,j} = \bar{m}_{H,j} + \tilde{m}_{H,j}, \tag{3.15}$$

$$\bar{m}_{H,j} = 1 - W_j, \quad \tilde{m}_{H,j} = W_j\left(1 - \sum_{n=1}^{N}\beta_{n,1}\right), \tag{3.16}$$

$$m_{n,I(j+1)} = k_{I(j+1)}[m_{n,j}m_{n,j+1} + m_{H,j}m_{n,j+1} + m_{n,j}m_{H,j+1}], \tag{3.17}$$

$$\tilde{m}_{H,I(j+1)} = k_{I(j+1)}[\tilde{m}_{H,j}\tilde{m}_{H,j+1} + \bar{m}_{H,j}\tilde{m}_{n,j+1} + \tilde{m}_{n,j}\bar{m}_{H,j+1}], \tag{3.18}$$

$$\bar{m}_{H,I(j+1)} = k_{I(j+1)}[\bar{m}_{H,j}\bar{m}_{H,j+1}], \tag{3.19}$$

$$k_{I(j+1)} = \left[1 - \sum_{t=1}^{N}\sum_{\substack{q=1\\ q\neq t}}^{N} m_{t,j}m_{q,j+1}\right]^{-1}, \quad j = 1, 2, \ldots, T-1. \tag{3.20}$$

Using Eqs. (3.13)–(3.20), the belief structure of the final assessment of city k is obtained as follows:

$$\beta_n(C_k) = \frac{m_{n,I(T)}}{1 - \bar{m}_{H,I(T)}}, \quad n = 1, 2, \ldots, N. \tag{3.21}$$

To compare the final rankings of each city exactly, the final assessments obtained in Eq. (3.21) should be transformed into the scores using the utilities of grades $u(H_n)$ ($n = 1, \ldots, N$). According to expected utility theory,[24] the expected utilities $u(C_k)$ ($k = 1, \ldots, K$) are calculated using Eq. (3.22) to generate the scores of all cities related to e-government performance.

$$u(C_k) = \sum_{n=1}^{N} \beta_n(C_k)u(H_n). \tag{3.22}$$

The scores are the obtained expect utilities and can be used for ranking.

4. Case Study

In this section, the proposed data-driven ER method is used to evaluate the e-government performance of 16 cities in Anhui province, China. A survey was conducted from November to December in 2018 to collect data. The data include responses from 16,718 citizens in Anhui Province. The proposed method translates the responses into the assessments with belief structures, determines the criterion weights and generates the final rankings of different cities related to e-government performance.

4.1. *A criterion framework for evaluating e-government performance*

The criteria used in this study come from the questionnaire of a survey conducted by an e-government institute. The institute has been authorized to evaluate citizens' satisfaction with the government portal websites of Anhui province in recent years. The first-level criteria are eight specific fields that citizens have being focused on, such as *census, housing, transportation, education, fertility, social insurance, employment & entrepreneurship, and business.*

The second-level criteria in each specific field include five aspects. They are generally service guidance clarity, site usability, information sharing, documentation, and e-service quality, as shown in Table 2. Citizens can perceive, use, navigate,

Table 2. Relevant notations.

Notations	Definitions
K	The total number of cities in the local government
T	The total number of first-level criteria
Q	The total number of second-level criteria under each first level
N	The total number of evaluation grades
M	The total number of users participating in the survey
C_k	The kth city's e-government performance ($k = 1, 2, \ldots, K$)
E_j	The jth first-level criterion ($j = 1, 2, \ldots, T$)
e_{ji}	The ith second-level criterion under the jth first-level criterion ($j = 1, 2, \ldots, T$; $i = 1, 2, \ldots, Q$)
ω_{ji}	The weight of second-level criterion $e_{ji}(j = 1, 2, \ldots, T, i = 1, 2, \ldots, Q)$
W_j	The weight of first-level criterion E_j ($j = 1, 2, \ldots, T$)
H_{nn}	Evaluation grades for evaluation criterion ($n = 1, 2, \ldots, N$)
$\beta_{n,ji}$	The degree of belief to which a city's e-government performance is assessed with grade H_n on the second-level criterion e_{ji} ($j = 1, 2, \ldots, T$, $i = 1, 2, \ldots, Q$, $n = 1, 2, \ldots, N$)

and interact with the government portals. Service guide clarity and site usability reflect the accessibility of e-government systems. Information sharing and documentation refer to the informationization of the e-government systems. The documentation and information related to e-government can be shared among the government portals of public service sectors. The e-service quality refers to the functional aspects of the e-government systems. Mutual communication, response speed, website effectiveness, and website contents belong to the function dimension of e-service quality. The five aspects in each specific field are measured by different items according to the users' requirements for the field service. The details about the measurement items for the second-level criteria are presented in Appendix B.

4.2. *Assessments in a unified format*

The eight first-level criteria are denoted by $E_j(j = 1, \ldots, 8)$ and their second-level criteria are denoted by $e_{ji}(i = 1, \ldots, 5; j = 1, \ldots, 8)$. Some second-level criteria and the overall assessments of first-level criteria are assessed on a scale with five grades, which is denoted by $\Omega = \{H_1, H_2, H_3, H_4, H_5\}$ = {Extremely Unsatisfied, Unsatisfied, Neutral, Satisfied, Extremely Satisfied} = {EU, U, G, S, ES}. Meanwhile, some second-level criteria are assessed on a scale with two grades, which is denoted by $\Omega = \{\overline{H}_1, \overline{H_2}\} = \{\text{Yes}, \text{No}\}$. Take the transportation field as an example. The values of its corresponding second-level criteria are shown in Table 2.

The assessments of each second-level criterion and the historical observations of each first-level criterion should be transformed into a unified format by a utility-based transformation method. For example, suppose that city k is assessed as having "Satisfied" criterion e_{ji}. Then, the assessment can be denoted by $B(e_{ji}(C_k)) = \{(H_1, 0), (H_2, 0), (H_3, 0)(H_4, 1), (H_5, 0)\}$. If city k is assessed as "Yes" on criterion e_{ji}, then the assessment can be denoted by $B(e_{ji}(C_k)) = \{(H_1, 1), (H_2, 0)\}$. To allow the proposed data-driven ER method to combine assessments and determine criterion weights, the assessments should be transformed into a uniform format by using the utility-based transformation method, which is mentioned in Sec. 3.4. After the transformation, all the 16,718 assessments are represented by a belief structure with five grades.

4.3. *Determining the weights of criteria*

The second-level criterion weights are learned by minimizing the average dissimilarity between the aggregated assessments of the second-level criteria and the corresponding historical observations of the first-level criteria provided by all users, as mentioned in Sec. 3.3.1.

Take the first-level criterion *transportation*, denoted by E_3, as an example to illustrate the process of learning the weights of the second-level criteria. The criterion E_3 includes five second-level criteria denoted by $\{e_{31}, e_{32}, e_{33}, e_{34}, e_{35}\}$. The assessments satisfy Assumption 3.1 and the independence in Definition 3.1. Suppose that the weights of five second-level criteria are denoted by ω_{3i} $(i = 1, 2, 3, 4, 5)$.

For the assessments of criterion *transportation*, the difference between the aggregated assessments of five second-level criteria, denoted by $B(y_4(C_k))$, and the historical observations on criterion *transportation*, denoted by $B(S_4(C_k))$, should be as small as possible. Using the optimization model shown in Sec. 3.3.1, the average dissimilarity is minimized to determine the weights of five second-level criteria. After solving the model by using the DE algorithm, the average dissimilarity is 0.0079 with $\omega^*_{3i}(i = 1, 2, 3, 4, 5) = (0.5336, 0.4660, 0.0001, 0.0002, 0.0001)$.

Similarly, all second-level criterion weights are obtained using the same procedures. Eight optimization models are developed for eight first-level criteria. The average dissimilarities after optimization are listed in Table 3. The obtained weights of second-level criteria are shown in Appendix B.

Table 3. The second-level criteria in the transportation field.

First-level criteria	Second-level criteria	Evaluation grades	Criterion values
Transportation	Service guidance clarity	$H1$–$H5$	{Extremely Unsatisfied, Unsatisfied, Neutral, Satisfied, Extremely Satisfied}
	Site usability	$H1$–$H5$	{Extremely Unsatisfied, Unsatisfied, Neutral, Satisfied, Extremely Satisfied}
	Information sharing	$H1$, $H2$	{Yes, No}
	Documentation	$H1$, $H2$	{Yes, No}
	E-service quality	$H1$, $H2$	{Yes, No}
	Historical observations on the whole	$H1$–$H5$	{Extremely Unsatisfied, Unsatisfied, Neutral, Satisfied, Extremely Satisfied}

After obtaining the weights of 40 second-level criteria, a new assessment matrix is established, as shown in Eq. (3.12). Then, the weights of eight first-level criteria are determined by using the CCSD method, as shown in Table 3. The procedures of the CCSD method are listed in Appendix A.

4.4. *Generating final rankings*

To generate the final rankings of the e-government performance of 16 cities, the assessments of eight first-level criteria are aggregated by using the ER algorithm. The final assessments aggregated by the ER algorithm are belief distributions with five grades. For example, the e-government performance of the city C1 is {(EU, 0.1648), (U, 0.0068), (G, 0.0263), (S, 0.0708), (ES, 0.7313)}, as shown in Table 4.

For comparison, the utility values of five grades are used to calculate the final scores. Then, the final assessments are transformed into scores using Eq. (3.22). A probability assignment approach is used to set the utility values of the five grades. According to expected utility theory,[24] given that $u(H_5) = 1$ and $u(H_1) = 0$, $u(H_n)$ $(n = 2, 3, 4)$ is determined to be limited to the range [0, 1] and the utility of each grade is obtained as $u(H_n) = \{u(H_1) = 0, u(H_2) = 1/4, u(H_3) = 2/4, u(H_3) = 3/4, u(H_1) = 1\}$. The final score of city C1 is 0.7992. The results and a ranking order of 16 cities are shown in Table 4.

4.5. *Result analysis*

The evaluation of local e-government performance is to promote the services provided by government websites and to bring real benefits to users. Thus, the results derived by the proposed method can provide a reference to local e-government performance promotion. According to the results of the criterion weights shown in Table 3, it is evident that users are mainly concerned about three aspects of e-government services including *service guidance clarity*, *site usability*, and *information sharing*. This means that users are sensitive to these criteria. If the performances of these aspects are increased a little, the whole performance of e-government systems can be improved greatly. Therefore, e-government development should focus

Table 4. The weights of first-level criteria and average dissimilarities.

First-level criterion	Weights	Average dissimilarities
Census	0.0629	0.0152
Housing	0.1443	0.0064
Transportation	0.0811	0.0079
Education	0.0873	0.0070
Fertility	0.232	0.0086
Social security	0.1094	0.0286
Employment	0.0629	0.0083
Commercial service	0.2201	0.0137

on the accuracy of service guidance information, the management of information and the development of self-service functions for e-government websites.

Among the first-level criteria, *fertility*, *business*, and *housing* contribute more to the evaluation results than the other criteria. This means that these three fields received the most concern from e-government users. From the final assessments presented in Table 5, the ranking of the e-government performance of the 16 cities in Anhui province is $C_6 \succ C_{13} \succ C_{12} \succ C_5 \succ C_{15} \succ C_4 \succ C_{16} \succ C_{10} \succ C_8 \succ C_7 \succ C_2 \succ C_3 \succ C_{14} \succ C_1 \succ C_{11} \succ C_9$. This means that city C6 has the highest performance according to the criterion weights obtained from the proposed DDER method.

5. Comparative Analysis

In this section, comparative analyzes are conducted to compare the classical objective methods for weight determination and the proposed data-driven ER method.

5.1. *Comparison of the average differences*

To illustrate the rationality of the DDER method, comparative analysis is conducted in this section between the DDER method and four classic objective methods, including the entropy method, the SD method, the CRITIC method, and the CCSD method. According to the weights obtained by the different methods, the average differences between the aggregated assessments of second-level criteria and the historical observations of the corresponding first-level criterion provided by all users should be small as possible. Therefore, the average difference is defined as

$$\tilde{P} = \frac{\sum_{a=1}^{M} \tilde{D}^a(y(C_k))}{M}, \tag{5.1}$$

where $\tilde{D}^a(y(C_k))$ represents the dissimilarity measure between the aggregated assessments of second-level criteria and the historical observations of the corresponding first-level criteria provided by user a. The average differences of eight fields are calculated using the criterion weights obtained from different methods and are shown in Fig. 4.

In terms of the proposed e-government performance criterion system and Assumption 3.1, the smaller the average difference $\tilde{P}$ is, the closer that the aggregated assessment of the individual second-level criteria is to the historical observation of the corresponding first-level criteria. It is obvious that the DDER method has the minimum average differences among all eight criteria. Therefore, the second-level criterion weights generated by the proposed DDER are more rational than those generated by the other four classic objective methods.

5.2. *Comparison of evaluation scores and rankings*

In this section, the comparative analysis of the final scores and rankings between the different methods is conducted to illustrate the effectiveness of the DDER method.

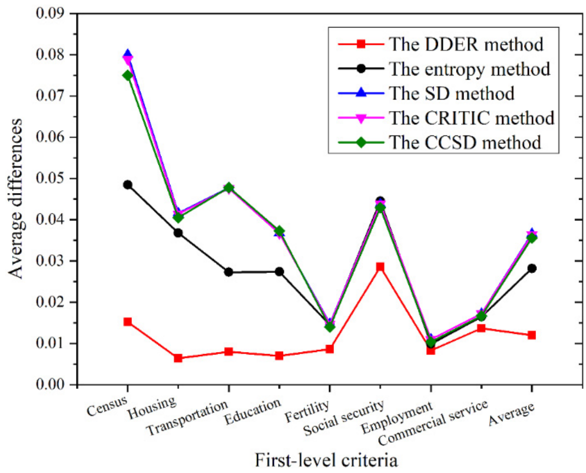

Fig. 4. The average differences obtained from different methods.

The results of the scores and rankings ($R.$) generated using the five methods are shown in Table 5.

When using the proposed DDER method, the assessments of the 40 second-level criteria are categorized into eight dimensions and then the assessments of the eight dimensions are obtained. When using the other four classic objective methods, the assessments of the 40 basic criteria are regarded as one level. Therefore, the utilities obtained from the five methods are different, as shown in Table 6.

It is obvious that the ranking orders generated by using the SD method, the CRITIC method, and the CCSD method are consistent with each other.

Table 5. Final assessments and rankings of 16 cities.

City ID	Final assessment					Scores	Rankings
	EU	U	G	S	ES		
C1	0.1648	0.0068	0.0263	0.0708	0.7313	0.7992	14
C2	0.1387	0.0053	0.0117	0.0487	0.7955	0.8392	11
C3	0.1245	0.0092	0.0375	0.0624	0.7663	0.8342	12
C4	0.1116	0.0018	0.0130	0.0488	0.8248	0.8684	7
C5	0.0704	0.0092	0.0246	0.0494	0.8463	0.8980	4
C6	0.0594	0.0027	0.0273	0.0651	0.8454	0.9086	1
C7	0.1260	0.0025	0.0252	0.0542	0.7921	0.8460	10
C8	0.1248	0.0030	0.0192	0.0439	0.8090	0.8523	9
C9	0.2042	0.0116	0.0646	0.0530	0.6666	0.7416	16
C10	0.1123	0.0025	0.0257	0.0556	0.8038	0.8590	8
C11	0.1619	0.0038	0.0510	0.0800	0.7034	0.7898	15
C12	0.0713	0.0054	0.0138	0.0538	0.8556	0.9043	3
C13	0.0733	0.0021	0.0155	0.0538	0.8554	0.9040	2
C14	0.1243	0.0138	0.0204	0.1022	0.7393	0.8296	13
C15	0.0844	0.0057	0.0244	0.0581	0.8274	0.8846	5
C16	0.1076	0.0020	0.0143	0.0597	0.8164	0.8688	6

Table 6. The scores and rankings obtained from the different methods.

City ID	Entropy		SD		CRITIC		CCSD		The DDER method	
	Scores	*R.*	Scores	*R.*	Scores	*R.*	Scores	*R.*	Scores	*R.*
C1	0.4321	1	0.4300	1	0.4306	1	0.4294	1	0.7992	14
C2	0.4301	2	0.4289	2	0.4294	2	0.4284	2	0.8392	11
C3	0.4165	16	0.4246	6	0.4253	6	0.4241	8	0.8342	12
C4	0.4189	13	0.4250	4	0.4255	4	0.4245	4	0.8684	7
C5	0.4181	15	0.4248	5	0.4254	5	0.4244	5	0.8980	4
C6	0.4204	12	0.4253	3	0.4258	3	0.4248	3	0.9086	1
C7	0.4184	14	0.4235	16	0.4239	16	0.4232	16	0.8460	10
C8	0.4233	9	0.4241	13	0.4244	13	0.4237	13	0.8523	9
C9	0.4227	11	0.4239	15	0.4243	15	0.4235	15	0.7416	16
C10	0.4233	10	0.4241	14	0.4244	14	0.4237	14	0.8590	8
C11	0.4244	5	0.4246	7	0.4249	7	0.4241	9	0.7898	15
C12	0.4241	6	0.4245	10	0.4249	8	0.4241	10	0.9043	2
C13	0.4248	3	0.4246	8	0.4248	10	0.4242	6	0.9040	3
C14	0.4248	4	0.4246	9	0.4249	9	0.4242	7	0.8296	13
C15	0.4237	8	0.4245	11	0.4248	11	0.4241	11	0.8846	5
C16	0.4238	7	0.4244	12	0.4246	12	0.4240	12	0.8688	6

In particular, the rankings generated by both the SD method and the CRITIC method are almost the same. This also verified that the CRITIC method, the SD method, and the CCSD method are developed based on the concept of the contrast intensity.

Although the final scores of e-government performances of the 16 cities generated by the proposed method are different from those obtained by using other four classic objective methods, the scores obtained by the proposed method have the largest value ranges and no duplicate values, which reflect its good discriminatory power. The distinction among the e-government performances of the 16 cites can be perceived by the proposed method.

In practice, the results of the final assessment provide valuable references for the management of local e-government performance. The good differentiation ability of the proposed method can provide an orientation for improvement. The weights of the criteria determined by the optimization model indicate the focus of the improvement for local e-government management.

6. Conclusions

The application of information and communication technologies stimulates the wide development of e-government. The web-based applications in e-government systems can provide faster, easier, and more efficient access and delivery of services to citizens. In this study, a criterion framework for evaluating e-government performance from the perspective of citizens' experience is constructed and a DDER method for evaluating e-government performance is proposed. The e-government performances of 16 cities in Anhui province of China are evaluated and ranked by their final utilities.

The criterion weights are determined by learning from data in this study. The construction of e-government is a continuous process and the e-government performance is assessed regularly. The vast amount of historical assessment data can be utilized to obtain the relative importance of the criteria. Therefore, decision makers' subjective preferences for criteria can be reduced in the next evaluation process. The empirical results of this study show that the service guidance clarity criterion is very important in most fields. This is consistent with previous studies. Babur *et al.*[7] state that web accessibility has a significant positive influence on citizens' degree of e-satisfaction in Pakistani. Web accessibility includes user guidance, the ease of navigation, and the user interface. According to the empirical results, the local government in Anhui province should pay more attention to the problems about service guidance clarity, documentation, and e-service quality in the fertility, business, and housing fields, which have the highest weights generated by the proposed method.

Due to the uncertainties existing in the process of evaluating e-government performance, we develop a DDER method to aggregate the assessments of criteria. The proposed DDER method for evaluating e-government performance can handle the uncertainties. The various kinds of uncertain assessments can be represented by a unified format using belief structures. Since the proposed method has good discriminatory power, the final scores and rankings provide a valuable reference for local e-government practitioners to identify the gaps and find insights for improvement.

Since there are many weight-measurement approaches among the various MCDM methods, we can hardly compare all methods. Only four kinds of objective methods are selected for the comparison analysis with the proposed data-driven method in this paper. Meanwhile, the criterion weights obtained in this paper are learned from the data from a survey. When the citizens involved in the survey change, the relative importance of the criteria may be different. These are the limitations of this study. In future research, the proposed method can be extended by adding different dissimilarity measures to the optimization model in different contexts. Moreover, machine learning methods, such as the random forest, have a good ability to learn criterion weights. The aggregation of a machine learning algorithm and MCDM may facilitate the decision-making process of evaluating e-government performance.

Acknowledgments

This research is partially supported by the National Nature Science Foundation of China (Nos. 72071061, 71573071, 71571060, 71671057, 71771077, 71801108, 71471054, and 91646111). The authors would like to express their appreciation to the anonymous reviewers for their valuable comments that helped improve the paper greatly.

Appendix A. The Process of the CCSD Method

The process of implementing the CCSD method to generate the weights of first-level criteria is as follows.

(1) The first step is to generate the aggregated assessment of all alternatives C_k $(k = 1, \dots, K)$ on all first-level criteria except the jth one, denoted by E_j $(j = 1, \dots, T)$. The aggregated assessment is calculated by

$$B^{j-}(C_k) = (\beta_1^{j-}(C_1), \dots, \beta_N^{j-}(C_k)), \quad k = 1, \dots, K. \tag{A.1}$$

The degree of global ignorance of $B^{j-}(C_k)$ is symbolized by $\beta_\Omega^{j-}(C_k)$ satisfying $\beta_\Omega^{j-}(C_k) = 1 - \sum_{n=1}^{N} \beta_n^{j-}(C_k)$.

(2) Then, the correlation coefficient between $B^{j-}(\cdot)$ and $B(E_j(\cdot))$, where $B^{j-}(\cdot)$ represents the set of $B^{j-}(C_k)(k = 1, \dots, K)$ and $B(E_j(\cdot))$ represents the set of $B(E_j(C_k))(k = 1, \dots, K)$, can be calculated by

$$R_j = \frac{C(B(E_j(\cdot)), B^{j-}(\cdot))}{\sqrt{E(B(E_j(\cdot))) \cdot E(B^{j-}(\cdot))}}, \quad j = 1, \dots, T, \tag{A.2}$$

where

$$C(B(E_j(\cdot)), B^{j-}(\cdot)) = \frac{1}{2}\sum_{k=1}^{S}\left(\sum_{n=1}^{N} \beta_{n,j}(C_k) \cdot \beta^{j-}(C_k) + \beta_{\Omega,j}(C_k) \cdot \beta_\Omega^{j-}(C_k)\right),$$
$$E(B(E_j(\cdot))) = \frac{1}{2}\sum_{k=1}^{S}\left(\sum_{n=1}^{N} \beta_{n,j}(C_k)^2 + \beta_{\Omega,j}(C_k)^2\right),$$

$E(B^{j-}(\cdot)) = \frac{1}{2}\sum_{k=1}^{S}(\sum_{n=1}^{N} \beta_n^{j-}(C_k)^2 + \beta_\Omega^{j-}(C_k)^2)$, and R_j is limited to [0, 1].

(3) The third step is to calculate the contrast intensity of the criteria. The proposed SD of the minimal satisfaction of alternatives on each criterion can quantify the contrast intensity of criteria.[22]

Let $V(E_j(C_k))$ be the minimal satisfaction of E_j on C_k, which is defined as

$$V(E_j(C_k)) = u_{\min}(E_j(C_k)) - \max_{m \neq k}\{u_{\max}(E_j(C_m))\}, \tag{A.3}$$

where $u_{\min}(E_j(C_k)) = \sum_{n=2}^{N} \beta_{n,j}(C_k)u(H_n) + (\beta_{1,j}(C_k) + \beta_{\Omega,j}(C_k))u(H_1)$ and $u_{\max}(E_j(C_k)) = \sum_{n=1}^{N-1} \beta_{n,j}(C_k)u(H_n) + (\beta_{N,j}(C_k) + \beta_{\Omega,j}(C_k))u(H_N)$ represent the minimum and maximum expected utilities of alternative C_k on criterion E_j, respectively.

(4) By using Eq. (A.3), the contrast intensity of criteria can be calculated as follows:

$$\sigma_j = \sqrt{\frac{1}{S}\sum_{k=1}^{S}\left(\bar{V}(E_j(C_k)) - \frac{\sum_{m=1}^{S}(\bar{V}(E_j(C_m))}{S}\right)^2}, \tag{A.4}$$

where $\bar{V}(E_i(C_k)) = (V(E_i(C_k)) - (-1))/2$. $\bar{V}(E_i(C_k))$ is a normalization of $V(E_j(C_k))$.

(5) The weight of the first-level criterion E_j can be determined by

$$W_j = \frac{\sigma_j \sqrt{1 - R_j}}{\sum_{k=1}^{T} \sigma_k \sqrt{1 - R_k}}, \quad j = 1, \ldots, T, \tag{A.5}$$

where $\sum_{j=1}^{T} W_j = 1$ and $W_j \geq 0$.

After the whole process of the CCSD method, the weights of the first-level criterion E_j $(j = 1, \ldots, T)$ are derived.

Appendix B. The Measurement Items for the Second-Level Criteria in Eight Specific Fields

First-level criteria	Second-level criteria	Evaluation grades	Weights
Census	Service guidance clarity	H1–H5	0.6584
	Proof of identification in household registration transfer transactions	H1, H2	0.0001
	Proof of identification in hotel registration transactions	H1, H2	0.0001
	Service convenience of applying for an identity card	H1–H5	0.3413
	Information sharing among different transactions	H1, H2	0.0001
Housing	Service guidance clarity	H1–H5	0.9838
	Proof of identification for public accumulation funds	H1, H2	0.0095
	Proof of identification in loan transactions	H1, H2	0.0065
	Proof of no-property certificate in housing transactions	H1, H2	0.0001
	Proof of identification in real estate transfer transactions	H1, H2	0.0001
Transportation	Service guidance clarity	H1–H5	0.5336
	Function service of making appointment and payment	H1–H5	0.4660
	Information sharing	H1, H2	0.0001
	Proof of identification in transportation transactions	H1, H2	0.0002
	Function service of transportation transactions	H1, H2	0.0001
Education	Information release	H1–H5	0.5185
	Function service of registering for school	H1–H5	0.4781
	Proof of student certificate in transfer within a province	H1, H2	0.0001
	Proof of student certificate in transfer across provinces	H1, H2	0.0001
	Proof of student certificate of identity card reissue transactions	H1, H2	0.0033
Fertility	Proof of pregnancy certificate in birth registrations	H1, H2	0.0720
	Service convenience of birth registration transactions	H1, H2	0.8918
	Proof of birth certificate in claiming expense transactions	H1, H2	0.0001
	Proof of birth certificate in social insurance transactions	H1, H2	0.0001
	Proof of parents' identification in registered residence transactions	H1, H2	0.0361
Social insurance	Information sharing	H1, H2	0.0129
	Function service of progress query	H1, H2	0.9349
	Service convenience of medical treatment transactions	H1, H2	0.0001
	Function service of medical insurance payment transactions	H1, H2	0.0001
	Function service of endowment insurance transactions	H1, H2	0.0520

(*Continued*)

First-level criteria	Second-level criteria	Evaluation grades	Weights
Employment & entrepreneurship	Service guidance clarity	H1, H2	0.9950
	Information release & security of recruitment information	H1, H2	0.0025
	Function service of personnel files transfer transactions	H1, H2	0.0023
	Function service of employment & entrepreneurship certificate transactions	H1, H2	0.0001
	Function service of training allowance of technical ability transactions	H1, H2	0.0001
Business	Service convenience of online business registration	H1, H2	0.0001
	Service convenience of enterprise registration transactions	H1–H5	0.0001
	Information release & security of declaring enterprises	H1, H2	0.9188
	Function service of stock rights change transactions	H1, H2	0.0001
	Service efficiency of social insurance transactions	H1, H2	0.0809

References

1. L. Ma, Digital divide and citizen use of e-government in China's municipalities, *International Journal of Public Administration in the Digital Age* **5**(3) (2018) 16.
2. J. J. Antonić and A. Šegota, Measuring performance of local e-government in the Republic of Croatia using data envelopment analysis, *Problems and Perspectives in Management* **10**(3) (2012) 35–45.
3. J. Wu and D. Guo, Measuring E-government performance of provincial government website in China with slacks-based efficiency measurement, *Technological Forecasting and Social Change* **96** (2015) 25–31.
4. D. H. Byun and G. Finnie, An AHP method for evaluating usability of electronic government portals, *Electronic Government, an International Journal* **8**(4) (2011) 343.
5. M. Nilashi, K. Bagherifard, O. Ibrahim, N. Janahmadi and H. Alizadeh, A multi-criteria approach to the evaluation of Malaysian government portal, *Journal of Theoretical and Applied Information Technology* **40**(2) (2012) 194–201.
6. M. S. Hossain, P. O. Zander, M. S. Kamal and L. Chowhury, Belief-rule-based expert systems for evaluation of e-government: A case study, *Expert Systems* **32**(5) (2015) 563–577.
7. S. Burmaoglu and Y. Kazancoglu, E-government website evaluation with hybrid MCDM method in fuzzy environment, *International Journal of Applied Decision Sciences* **5**(2) (2012) 163–181.
8. S. Zhang, D. J. Yu, Y. Wang and W. Y. Zhang, Evaluation about the performance of E-Government based on interval-valued intuitionistic fuzzy set, *The Scientific World Journal* **2014** (2014) 1–10.
9. G. Kong, D. L. Xu, J. B. Yang *et al.*, Combined medical quality assessment using the evidential reasoning approach, *Expert Systems with Applications* **42**(13) (2015) 5522–5530.
10. C. Fu, M. Xue and D. L. Xu, Selecting strategic partner for tax information systems based on weight learning with belief structures, *International Journal of Approximate Reasoning* **105** (2019) 66–84.
11. S. Shan, L. Wang, J. Wang, Y. Hao and F. Hua, Research on e-Government evaluation model based on the principal component analysis, *Information Technology and Management* **12**(2) (2011) 173–185.

12. G. Kaisara and S. Pather, The e-Government evaluation challenge: A South African Batho Pele-aligned service quality approach, *Government Information Quarterly* **28**(2) (2011) 211–221.
13. H. M. Babur, S. Cai, G. M. Abdul, G. Noreen and G. Hifza, Evaluating citizen e-satisfaction from e-government services: A case of Pakistan, *European Scientific Journal* **12**(5) (2016) 346–370.
14. S. F. Verkijika and W. L. De, Quality assessment of e-government websites in Sub-Saharan Africa: A public values perspective, *The Electronic Journal on Information Systems in Developing Countries* **84**(2) (2018) e12015.
15. M. A. Boujelben and Y. D. Smet, A choice model with imprecise ordinal evaluations, *International Journal of Approximate Reasoning* **55**(2) (2014) 689–710.
16. R. Roberts and P. Goodwin, Weight approximations in multi-attribute decision models, *Journal of Multi-Criteria Decision Analysis* **11**(6) (2002) 291–303.
17. P. A. Bottomley and J. R. Doyle, A comparison of three weight elicitation methods: Good, better, and best, *Omega* **29**(6) (2001) 553–560.
18. E. Takeda, K. O. Cogger and P. L. Yu, Estimating criterion weights using eigenvectors: A comparative study, *European Journal of Operational Research* **29**(3) (1987) 360–369.
19. W. E. Riggs, The delphi technique: an experimental evaluation, *Technological Forecasting and Social Change* **23**(1) (1983) 89–94.
20. H. Deng, C. H. Yeh and R. J. Willis, Inter-company comparison using modified TOPSIS with objective weights, *Computers & Operations Research* **27**(10) (2000) 963–973.
21. D. Diakoulaki, G. Mavrotas and L. Papayannakis, Determining objective weights in multiple criteria problems: The critic method, *Computers & Operations Research* **22**(7) (1995) 763–770.
22. Y. M. Wang and Y. Luo, Integration of correlations with standard deviations for determining attribute weights in multiple attribute decision making, *Mathematical and Computer Modelling* **51**(1–2) (2010) 1–12.
23. S. Liu, F. T. S. Chan and W. Ran, Decision making for the selection of cloud vendor: An improved approach under group decision-making with integrated weights and objective/subjective attributes, *Expert Systems with Applications* **55** (2016) 37–47.
24. J. Yang, Rule and utility based evidential reasoning approach for multiattribute decision analysis under uncertainties, *European Journal of Operational Research* **131**(1) (2001) 31–61.
25. C. Fu, J. B. Yang and S. L. Yang, A group evidential reasoning approach based on expert reliability, *European Journal of Operational Research* **246**(3) (2015) 886–893.
26. C. Fu, W. Liu and W. Chang, Data-driven multiple criteria decision making for diagnosis of thyroid cancer, *Annals of Operations Research* **293** (2020) 833–862.
27. R. Storn and K. Price, Differential evolution – A simple and efficient heuristic for global optimization over continuous spaces, *Journal of Global Optimization* **11**(4) (1997) 341–359.
28. G. Kou, P. Yang, Y. Peng, F. Xiao and F. E. Alsaadi, Evaluation of feature selection methods for text classification with small datasets using multiple criteria decision-making methods, *Applied Soft Computing* **86** (2019) 105836.
29. T. Li, G. Kou and Y. Peng, Improving malicious URLs detection via feature engineering: Linear and nonlinear space transformation methods, *Information Systems* **91** (2020) 101494.
30. T. Li, G. Kou, Y. Peng and Y. Shi, Classifying with adaptive hyper-spheres: An incremental classifier based on competitive learning, *IEEE Transactions on Systems Man and Cybernetics: Systems* **50**(4) (2017) 1–12.