

Article

The Application of the Analytic Hierarchy Process and a New Correlation Algorithm to Urban Construction and Supervision Using Multi-Source Government Data in Tianjin

Shaoyi Wang ¹, Zhongjie Sheng ¹, Yuliang Xi ^{2,3}, Xiangyuan Ma ^{2,3}, Huihui Zhang ^{2,3}, Mengjun Kang ^{2,3} , Fu Ren ^{2,3}, Qingyun Du ^{2,3,*}, Ke Hu ¹ and Zhenbiao Han ¹

¹ Tianjin Institute of Surveying and Mapping, Tianjin 300381, China; wshaoy100@126.com (S.W.); suxiu0815@163.com (Z.S.); huketj@163.com (K.H.); gu.han@163.com (Z.H.)

² School of Resources and Environmental Science, Wuhan University, Wuhan 430079, China; yuliangwhu@163.com (Y.X.); maxiangyuan@whu.edu.cn (X.M.); Zhanghuih@whu.edu.cn (H.Z.); mengjunk@whu.edu.cn (M.K.); renfu@whu.edu.cn (F.R.)

³ Key Laboratory of Geographic Information Systems, Ministry of Education, Wuhan University, Luoyu Road 129, Wuhan 430079, China

* Correspondence: qydu@whu.edu.cn; Tel.: +86-27-6877-8842; Fax: +86-27-6877-8893

Received: 1 December 2017; Accepted: 1 February 2018; Published: 5 February 2018

Abstract: As the era of big data approaches, big data has attracted increasing amounts of attention from researchers. Various types of studies have been conducted and these studies have focused particularly on the management, organization, and correlation of data and calculations using data. Most studies involving big data address applications in scientific, commercial, and ecological fields. However, the application of big data to government management is also needed. This paper examines the application of multi-source government data to urban construction and supervision in Tianjin, China. The analytic hierarchy process and a new approach called the correlation degree algorithm are introduced to calculate the degree of correlation between different approval items in one construction project and between different construction projects. The results show that more than 75% of the construction projects and their approval items are highly correlated. The results of this study suggest that most of the examined construction projects are well supervised, have relatively high probabilities of satisfying the relevant legal requirements, and observe their initial planning schemes.

Keywords: analytical hierarchy process; correlation algorithm; construction and supervision; multi-source government data; multi-criteria decision aiding

1. Introduction

With the development of information technology and the popularization of the Internet, large amounts of data are constantly being generated. Thus, big data, which was defined by Viktor Mayer-Schonberger in 2008 and is a product of this era, has attracted increasing attention from researchers [1,2]. Big data can be defined as large volumes of highly varied data that are generated, captured, and processed at high velocity and have five traits that begin with the letter V, specifically: volume, velocity, variety, value, and veracity [1,3,4]. Big data has gained significant impetus as a breakthrough technological development in nearly all fields, such as government administration, commercial operations, and scientific research [5–7].

Many countries such as the USA, the UK, Canada, and Japan are committed to the development and application of multi-source government big data, which represents an important field of application of big data techniques [8,9]. Such data have both a wide range of data sources and

are supplied in various data formats [10,11]. These data can be analyzed and explored to produce valuable knowledge, which can help to improve governance. Furthermore, multi-source government big data is the foundation of the development of smart cities [11–13]. However, the application of these data is presently insufficient. Specifically, most multi-source government big data are only browsed and queried because an efficient means of processing large amounts of unstructured data, which include text-based, spatial, and multimedia data, is lacking [4]. In addition, there is a shortage of applications of data processing and analysis techniques. Moreover, government big data are scattered among different departments without centralized management, organization, or fusion [14].

In view of these three problems, many studies have been carried out. First, big data processing is a complete technical chain that includes data storage, calculation, and analysis [15,16]. Thus, previous studies have focused mainly on these three directions. In terms of data storage, the relational database management system (RDBMS) was proposed early (E.F. Codd, 1970); the NoSQL (Not Only Structured Query Language) is a currently popular approach for large and distributed data management and database design that first appeared in 1998 and was proposed a second time by Eric Evans in 2009; and a new type of distributed database (HBase) has been introduced to store and manage structured and unstructured big data (Fay Chang, 2006). In terms of calculations using data, Hadoop MapReduce, which is a distributed parallel computing framework for big data, was used in early studies (Apache Software Foundation, 2005). With the promotion of big data applications and increased diversity in the modes of computing used (which mainly involve stream calculations, graph calculations, and interactive calculations), many newly distributed parallel computing frameworks, such as Spark (UC Berkeley AMP lab, in the USA, 2009) and GraphLab (CMU, Select, 2010), have been developed [15]. These computing frameworks broaden the modes of computing used and their efficiency is much greater. In terms of data analysis and mining, many machine learning and parallel deep learning algorithms (Hinton, 2006) have been developed to solve problems that arise in practical applications, including image recognition, natural language processing, and social network analysis. Additionally, some government data analysis methods are being applied in various departments. Many scholars use decision trees to assist in land management governance, use cluster analysis to explore the trends of urban economic development, and use statistical regression or machine learning to reveal and predict the spatiotemporal distributions of air quality (e.g., PM 2.5). Furthermore, government data management platforms have been established in many cities. These platforms can aggregate data resources from various municipal departments, so as to realize centralized management and data sharing. However, compared with the massive amounts of government data, the current application research remains inadequate. This situation leads to the wasted data sources [4,5].

To broaden the applications of multi-source government data, this paper focuses on those applications in construction project management using various kinds of data fusion methods and the multi-criteria decision aiding (MCDA) analytic hierarchy process (AHP). Currently, given the impetus supplied by electronic government affairs, the correlation of attributes of government data is implemented in most e-government systems [17,18]. However, the correlation of text-based and spatial data is still in its infancy at the application level. The MCDA is also widely used in other fields: Shirley Chua Jin Lin explored the procurement strategies of the government (2015); Eylem Koc and Hasan Arda Burhan explored the location selection of commercial circles (2015); Lakshmi Tulasi and Ramakrishna Rao explored resource allocation in project scheduling (2014); Li Hongxian and Mohamed Al-Hussein explored risk identification and the assessment of modular construction (2013); and Wang Xiaodi and Peng Xiuyan explored the evaluations of educational equipment efficiency factors (2011) [19–23]. However, MCDA has rarely been applied in urban construction and supervision.

To address this problem, this paper examines a practical application of the fusion of multi-source big government data to urban construction and supervision in Tianjin, China. Various data from different departments, such as the urban planning and urban construction departments and the environmental protection bureau, are used to assess the degree to which construction projects satisfy the relevant legal requirements. Specifically, this study combines the three modes of data fusion.

We use a text correlation method, correlated attributes, and a spatial correlation method to calculate the degrees of correlation of a project name and location description, construction unit, and spatial intersection. In addition, this study applies the multi-criteria decision aiding (MCDA) analytic hierarchy process (AHP) to calculate the correlations of every construction segment within individual projects. We note that a construction project can be separated into several segments, according to its degree of completion [24]. Moreover, we develop a new method that is based on the results of the AHP to calculate the correlations between different construction projects. If the correlation is lower than some threshold, which is set based on experience, it means that the construction project or segment does not satisfy the relevant legal requirements. The application of multi-source government data in construction supervision and the quantitative calculation of data correlations will broaden the horizons of this field and provide key insight for use by the construction supervision department.

2. Data and Methods

2.1. Study Area

Tianjin is a coastal city that is located in the northern part of China and north of the Yellow River Estuary. This city is located in the northern temperate zone and southeast of Beijing, the capital of China. Tianjin occupies a longitudinal range of $116^{\circ}43'$ E to $118^{\circ}04'$ E and a latitudinal range of $38^{\circ}34'$ N to $40^{\circ}15'$ N, and its total area is approximately $11,946 \text{ km}^2$.

Tianjin, which has a coastline that is 153 km in length, is rich in marine resources and has excellent ports and abundant fisheries. Tianjin is one of the municipalities of China. With its rapid economic development and urban expansion, Tianjin occupies a position of importance in China. Tianjin presently contains a total of 16 districts. Figure 1 shows the location of Tianjin in China and the names of its districts.

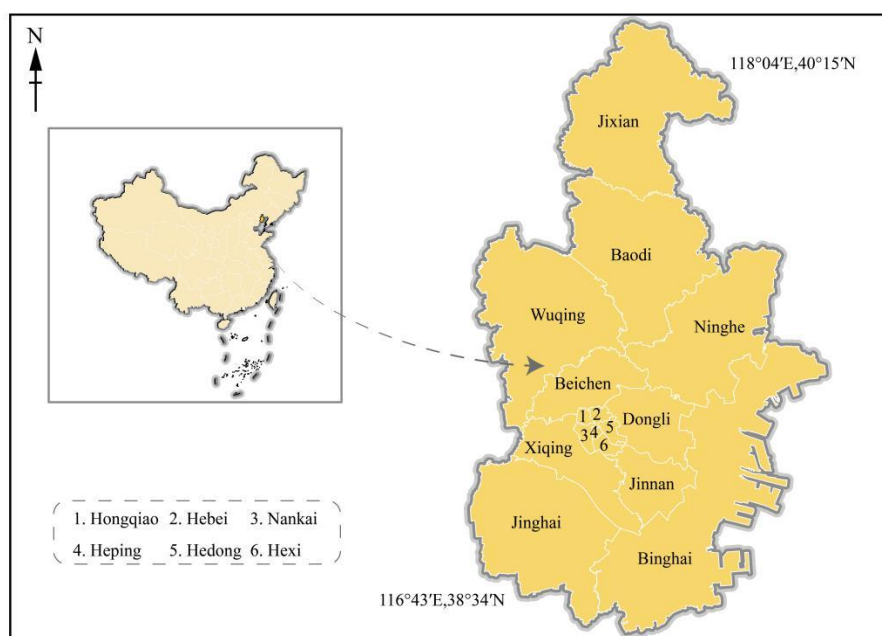


Figure 1. The study area.

2.2. Study Data

2.2.1. Summary of Construction Project Approval Items

The approval of construction projects plays an important role in the field of urban construction management in Tianjin. The competent organizations for examination and approval include the

municipal development and reform commission, planning bureau, construction bureau, land and housing bureau, environmental protection bureau, and the municipal air defense office.

Generally, there are approximately 20 approval items in a construction project, four of which are major items that play important roles in the approval processes: the land supply scheme, land use permit, construction permit, and environmental impact assessment. The data of the four approval items are relatively complete and come from different government organizations. It is beneficial to measure their degrees of correlation. Specifically, the land supply scheme is a land supply plan formulated by the land administration department according to the construction project conditions, which includes a claim of the land area, functional zoning, applications, land supply modes, evaluation prices, land use years, and so on. The land use permit is the legal certificate for the construction organizations to confirm that the locations and scopes of construction projects are in line with urban and rural planning before applying for the requisition and allocation of the land with the land management department. The construction permit is the license stating a construction organization has met various kinds of construction conditions and is also a legal certificate for construction units to carry out engineering construction. Environmental impact assessment is a way of monitoring and includes the analysis, prediction, and assessment of the environmental impacts that may be caused by the implementation of planning and construction projects and proposes countermeasures and measures to prevent or alleviate adverse environmental impacts.

2.2.2. Data Sources

This paper examines all aspects of the construction project approval data that correspond to the two years of 2015 and 2016; in total, 186,962 records are considered. There are 11 approval items that include spatial data and 9 approval items that include non-spatial data, for which 154,992 and 31,970 records, respectively, are examined. There are 3 approval items that correspond to 2169 records in which the data can be used to determine spatial positions in terms of latitude and longitude. Moreover, in order to simplify the calculations, four typical approval items (the land supply scheme, the land use permit, the construction permit, and the environmental impact assessment), for which a total of 30,996 records are examined, are selected from individual construction projects. These four items describe the approval process to the maximum extent and reflect the application of multi-source government data in the supervision of construction projects.

2.2.3. Data Correlation Factors

Data correlation factors are introduced to measure the degree of correlation, which reflects whether construction projects are supervised well or observe the rules of planning, construction, and environmental protection, as determined from the many approval items associated with construction projects. Specifically, there are five typical correlation factors: pre-post relationships, spatial position relationships, the project name, the project department, and the location description.

1. Pre-post relationships: There are clear pre-post relationships among the construction project approval items. Previous approvals are generally required for later approvals. For example, the planning and site selection of each project can only be checked after the project has been established, and the administration department then organizes the land use according to the planning conditions. Thereafter, the construction management department gives out a construction permit, which states that the builder can carry out the process of construction. After the acceptance is complete, the registration of property rights can be carried out. Thus, the prerequisite information is critical because it links each part of a construction project with the other parts.
2. Spatial relationships: Each approval item is linked to a spatial location. Spatial consistency plays an important role in the process of project approval, land supply, planning, construction, and acceptance. Typical construction projects should be strictly consistent with the spatial scope

in each part of the approval process; however, some factors, including the approximate scope defined by the management department in the approval stage, are not precise, and the phasing construction of the part of the project that determines the scope of the location in the construction stage is less than the scope in the overall design stage. Moreover, differences in the coordinate systems used lead to errors in the data conversion process. Coincidence, inclusion, intersection, adjacency, and other spatial relationships can be equal to incidence relationships.

3. Project name: A consistent or semantically similar name should be used in all the items involved in construction projects.
4. Construction unit: Construction projects should be examined, approved, and constructed by the same construction unit for all the items.
5. Location description: Although there is no standardized data format for the spatial locations used to assess some of the approval items, the formats used should be consistent or similar.

Note that the degree of correlation of pre-post relationships, construction units, and location descriptions are measured according to the textual semantics of the construction approval items. However, the degrees of correlation of the spatial relationships and project names are measured according to the spatial positioning data and attributes, respectively.

2.3. Study Method

2.3.1. Analytic Hierarchy Process (AHP)

Multi-criteria decision analysis (MCDA) is a discipline that helps decision makers make decisions when several conflicting criteria need to be evaluated. When facing a decision problem, the first task of a decision maker is to identify the type of problem (Ishizaka et al. 2012) [25]. The analytic hierarchy process (AHP) is one of the MCDA methods (Saaty, 1977) [26]. This method is a systematic procedure for solving complex multi-criteria decision problems by giving systematic prioritized rankings (Anderson et al. 2008) [27–30]. This procedure then organizes the basic rationalities by breaking down a problem into its smaller constituent parts and calling for only simple pairwise comparison judgments to develop a priority hierarchy. The AHP provides a comprehensive framework to cope with the intuitive, rational, and irrational judgments at the same time and does not even require that judgments be consistent or even transitive. The degree of consistency of the judgment is revealed at the end of the AHP process (Saaty, 1982). A predominant aspect of the AHP is that knowledgeable individuals who supply judgments for the pairwise comparisons usually also play a prominent role in specifying the hierarchy. Although a hierarchy for uses in resource allocation will tend to have the vertical stratification indicated above, it can also be much more general. The only restriction is that any element on a higher level must serve as a governing element or have at least two elements on the immediately lower level (Saaty and Erdener, 1979) [31]. The analysis process is based on the following three stages.

1. Establishing a Hierarchical Structure Model

The priority is to stratify the factors according to their interrelationships and subordinate relationships, and the factors are gathered into different levels [32]. In accordance with the target layer, several relevant rule layers and scheme layers are arranged in the form of a multilevel analytical structural model. The problem to be solved is finally distilled to the relative weights of the different levels from the lowest to the uppermost (which represents the general objective) or the order of merits.

2. Constructing Judgment Matrices

To compare the effects of n elements ($X = \{x_1, x_2, \dots, x_n\}$) on a factor Z , a method is proposed for comparing pairs of elements through the establishment of a pairwise comparison matrix. For each pair of elements (x_i, x_j) , a_{ij} is the ratio of the magnitudes of the influence of x_i and x_j on Z . All of the comparison results are represented by a matrix $A = (a_{ij})_{n \times n}$, where A is the pairwise comparison judgment matrix of Z - X . This matrix is referred to as the judgment matrix in this study.

The judgment matrix is used to make pairwise comparisons and represents the relative importance of elements between the present level and the previous level. Through pairwise comparisons, the relative importance of factors at a higher level and a lower level is obtained.

The elements in the judgment matrix are calculated using the following formula:

$$a_{ij} > 0, a_{ij} = 1/a_{ji}, a_{ii} = 1 \quad (1)$$

The result of comparing two elements at the same level can be expressed by introducing an appropriate scale. The values of the elements are determined in the judgment matrix according to Saaty's suggestion and range from 1 to 9, and the reciprocal of one of these values is referenced using the scale [33] (Table 1). This process provides a foundation for carrying out matrix operations; these values are used to express the relative importance of pairs of factors, and they are transferred into a judgment matrix.

3. Calculating Weight Vectors and Performing Consistency Tests

Weight vectors represent the importance of the elements that lie in the current layer to the upper layer. They can be calculated using the matrix. That is, the eigenvectors corresponding to the maximal eigenvalue of a matrix consistency test are needed to determine whether it can take the hierarchical single order. To ensure that the calculation results are reasonable, the elements within the judgment matrix should be approximately consistent. The formula used in the consistency test is as follows:

$$CR = CI/RI \quad (2)$$

where CR is the consistency ratio and CI is the consistency index; a general formulation is given as follows:

$$CI = \lambda_{\max} - n / (n - 1) \quad (3)$$

where λ_{\max} is the maximum root of the judgment matrix and n is the number of pairs of elements being compared (or the order of the judgment matrix). The random index (RI) is introduced to measure the degree of the CI .

In the judgment matrix consistency test, if the $CI = 0$, complete consistency is indicated; if the $CI \leq 0.1$, the degree of consistency is acceptable. Higher values of the CI indicate greater degrees of inconsistency.

Finally, the consistency ratio of the total order is used to perform a test. If the test is passed, the decision and analysis can be made according to the results of the weight vector of the total order. Otherwise, it is necessary to reconsider the hierarchical model or to re-adjust the comparison matrices with a larger CR .

Table 1. The fundamental scale of absolute numbers.

Intensity of Importance	Definition	Explanation
1	Equal importance	The two activities contribute equally to the objective
3	Moderate importance	Experience and judgment slightly favor one of the activities over the other
5	Strong importance	Experience and judgment strongly favor one of the activities over the other
7	Very strong or demonstrated importance	One activity is favored very strongly over the other; this dominance has been demonstrated in practice

Table 1. Cont.

Intensity of Importance	Definition	Explanation
9	Extreme importance	The evidence favoring one activity over another is of the highest possible order of affirmation
2, 4, 6, 8	2 (Weak or slight), 4 (moderate plus), 6 (strong plus), 8 (extremely strong)	Intermediate values between the two adjacent judgments
Reciprocals of the numbers above	If activity <i>i</i> has one of the above non-zero numbers assigned to it when compared with activity <i>j</i> , then <i>j</i> has the reciprocal value when compared with <i>i</i>	

2.3.2. Degree of Correlation Algorithm

The AHP can be used to measure the degree of correlation among different approval items within a single construction project. It is difficult to measure the degree of correlation among many different construction projects due to the curse of dimensionality [34]. To address this problem, a new approach based on the hierarchical analysis method is introduced to calculate the degree of correlation among different construction projects.

Specifically, the general correlation matching the degree of a complete construction project can be shown as the total score of the five factors in the criterion layer. To quantitatively measure the correlation matching degree of each factor in the criterion layer, the division standard is expressed according to the experience of the experts. Table 2 shows the division standards for the degree of correlation of each factor in the criterion layer.

Table 2. Division standards for the degree of correlation of each factor in the criterion layer.

Associative Matching Degree Factor	Classifying Grade	Quantitative Description of Degree of Correlation (ranked from high to low)
Pre-post relationship	2 grades (consistent or inconsistent)	2, 1
Spatial intersection	5 grades (coincident, intersect >50%, intersect <50%, adjacent, disjoint)	5, 4, 3, 2, 1
Project name	3 grades (all consistent, partly consistent, inconsistent)	3, 2, 1
Project enterprise	3 grades (all consistent, partly consistent, inconsistent)	3, 2, 1
Location description	3 grades (all consistent, partly consistent, inconsistent)	3, 2, 1

Note that if the correlation matching degree is between two adjacent grades, then the correlation matching degree is the median of the adjacent grades.

For example, if there are *n* approval items, a matrix that is based on the AHP model is built as follows:

$$\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix} \tag{4}$$

If $a_{ij} \leq 1$, the degree of correlation is expressed as cd_{ij} , and the number of total annotations is $C_n^2 + n$. The formula that expresses the overall degree of correlation of the individual scoring factors is as follows:

$$S_{CD} = \sum cd_{ij}, a_{ij} \leq 1 \tag{5}$$

The formula that expresses the overall degree of correlation of construction projects is as follows:

$$S = \sum_{i=1}^n S_{CDi} \cdot w_i \tag{6}$$

where w_i is the weight of the scoring factor.

If the degree of correlation is high, the construction project has a greater probability of satisfying the relevant legal requirements and observes the initial planning scheme.

2.3.3. The Calculation Procedure of the Degree of Correlation among Construction Projects

As shown in Figure 2, the process of calculating the degree of correlation among construction projects includes five steps. Of these, establishing the AHP model is the most important step; it is the foundation for the calculation of the degree of correlation and determines the results. Note that the typical approval items can be changed, depending on the realities of the construction projects under consideration.

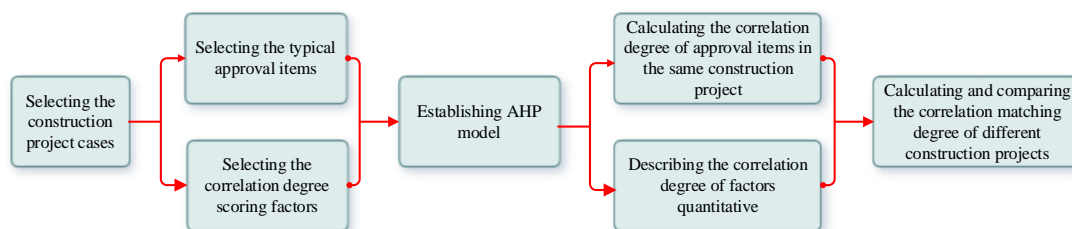


Figure 2. The calculation procedure of the degree of correlation among construction projects.

3. Results and Discussion

3.1. Description of Construction Project Cases

To apply the abovementioned method to actual construction projects using multi-source government data, two construction project cases are introduced to calculate the degree of correlation of different approval items within individual construction projects and between different construction projects. The selected approval items and a qualitative evaluation of the degree of correlation between the factors of two construction projects are listed in Table 3.

Table 3. (A) The approval items and a qualitative evaluation of the degree of correlation among the factors of construction project A. (B) The approval items and a qualitative evaluation of the degree of correlation among the factors of construction project B.

A					
Approval Items	Pre-Post Relationships	Spatial Intersection	Project Name	Project Unit	Location Description
Land supply scheme	Inconsistent	Intersect < 50%	Between inconsistent and partly consistent	Inconsistent	Partly consistent
Land use permit	Consistent	Deviation	Inconsistent	All consistent	Between partly consistent and all consistent
Construction permit	Inconsistent	Containment	All consistent	Inconsistent	Partly consistent
Environmental impact assessment	Inconsistent	Containment	Partly consistent	Inconsistent	Between inconsistent and partly consistent

Table 3. Cont.

B					
Approval Items	Pre-Post Relationships	Spatial Intersection	Project Name	Project Unit	Location Description
Land supply scheme	Consistent	Deviation	Between inconsistent and partly consistent	All consistent	Between partly consistent and all consistent
Land use permit	Consistent	Between intersect >50% and containment	All consistent	All consistent	All consistent
Construction permit	Inconsistent	Deviation	Inconsistent	Inconsistent	Inconsistent
Environmental impact assessment	Between inconsistent and consistent	Between intersect >50% and containment	Inconsistent	Inconsistent	Between inconsistent and partly consistent

3.2. Calculation of the Degree of Correlation of the Approval Items in a Single Construction Project

1. Hierarchical Single Arrangement

The priority is to determine the values of the correlation factors in the criterion layer. These values are determined and screened from the approval data of the construction projects based on the experience of the experts, which has a significant effect on the correlation scores. In this study, five correlation factors (pre-post relationships, spatial intersection, the project name, the project unit, and the location description) are grouped to form the criterion layer in the analytic hierarchy model. Some approval items, such as the land supply scheme, the land use permit, the construction permit, and the environmental protection measures of a project, play major roles in the entire construction project. Thus, these approval items are assigned to elements in the scheme layer. On this basis, these factors are organized into a complete hierarchical structure that conveys the subordinate relationships and correlations. The AHP model is layered. Although the upper layer is affected by the correlation factors of the middle layer and the approval items of the bottom layer, the factors within a given layer are independent. The AHP model for Tianjin is shown in Figure 3.

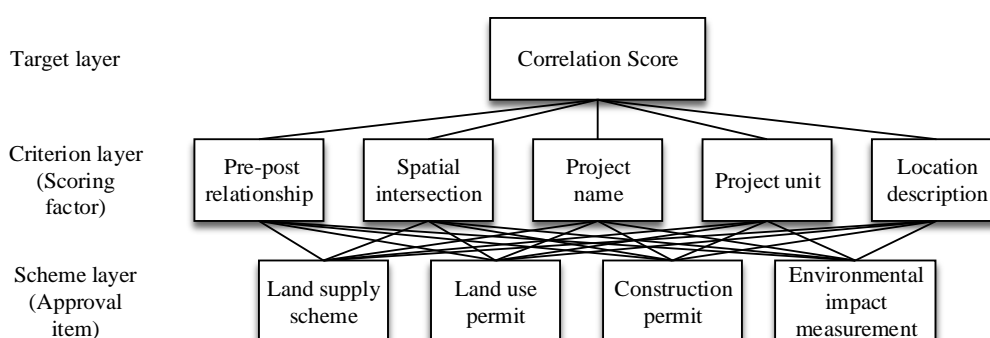


Figure 3. Analytic hierarchy model for the correlation of multi-source data from construction projects.

Note that each correlation factor in the criterion layer is connected with the different approval items of the scheme layer, and the factors in the upper layers are clearly related and subordinated to those of the lower layers. To calculate and quantitatively represent the importance of the elements in the scheme layer to the correlation factors in the criterion layer, pairs of the correlation factors in the criterion layer are compared with each other to construct a comparison matrix. Based on the comparison of pairs of factors, the results are scaled in the judgment matrix. Table 4 shows the judgment matrix that reflects the results of comparing the correlation factors.

Table 4. Judgment matrix for the scoring factors of the criterion layer.

	Pre-Post Relationships	Spatial Intersection	Project Name	Project Unit	Location Description
Pre-post relationships	1	2	3	5	3
Spatial intersection	1/2	1	2	3	2
Project name	1/3	1/2	1	2	1
Project unit	1/5	1/3	1/2	1	1/2
Location description	1/3	1/2	1	2	1

Note that the project unit is the weakest of the five factors, and pre-post relationships represent the strongest factor. The weights of each correlation factor in the criterion layer can be calculated. The normalized weights of the five correlation factors corresponding to the maximum eigenvalue are $w(0.801, 0.465, 0.261, 0.144, 0.232)^T$, and the CR is $0.004 \ll 0.1$. These results demonstrate that the comparison matrix displays complete consistency. Moreover, the pre-post relationships are shown to have the greatest effect on the calculated degree of correlation among the approval items in a construction project. When arranged in descending order according to their weights, the remaining factors are a spatial intersection, the project name, the location description, and the project unit.

2. Hierarchical Total Arrangement

The purpose of this section is to calculate the degree of correlation among different approval items within a single project. This study selects four approval items, which are mentioned in Figure 2. Specifically, they are the land supply scheme, the land use permit, the construction permit, and the environmental impact assessment. Note that each correlation factor has a fourth-order comparison matrix, and each matrix shows the real correlation among the four approval items. For example, the comparison matrix of the pre-post relationships for construction project A is as follows (the other comparison matrices are provided in the Supplementary Materials):

$$\begin{bmatrix} 1 & 1/2 & 1 & 1 \\ 2 & 1 & 2 & 2 \\ 1 & 1/2 & 1 & 1 \\ 1 & 1/2 & 1 & 1 \end{bmatrix} \quad (7)$$

All the results are shown in Table 5, including the maximum eigenvalues and the corresponding eigenvectors of the two construction projects.

Table 5. The results for the two construction projects, including the maximum eigenvalues and the corresponding eigenvectors.

Construction Project Name	Degree of Correlation Factor	Eigenvalue	Eigenvector (Approval Items)			
			Land Supply Scheme	Land Use Permit	Construction Permit	Environmental Impact Assessment
Construction project A	Pre-post relationships	4.000	0.378	0.756	0.378	0.378
	Spatial intersection	4.043	0.264	0.118	0.677	0.677
	Project name	4.051	0.224	0.136	0.888	0.378
	Project unit	4.000	0.189	0.945	0.189	0.189
	Location description	4.010	0.416	0.777	0.416	0.224
Construction project B	Pre-post relationships	4.010	0.647	0.647	0.201	0.349
	Spatial intersection	4.000	0.088	0.702	0.088	0.702
	Project name	4.028	0.293	0.928	0.164	0.164
	Project unit	4.000	0.693	0.693	0.139	0.139
	Location description	4.048	0.512	0.823	0.130	0.209

The hierarchical total arrangement, that is, the degree of correlation among different approval items in a single construction project, can be calculated according to this table. The results are shown in Table 6.

Table 6. The degree of correlation among the different approval items for the individual construction projects.

Construction Project	Degree of Correlation between the Different Approval Items			
	Land Supply Scheme	Land Use Permit	Construction Permit	Environmental Impact Assessment
A	0.618	1.012	0.973	0.795
B	0.854	1.377	0.295	0.717

Overall, land use is the approval item with the highest degree of correlation for both construction projects. This result indicates that the approval data of the construction projects are relatively complete for this approval item, and this item has a relatively high probability of satisfying the relevant legal requirements. Thus, the supervision of construction projects should be strengthened in the other approval items.

3.3. Calculation of the Degree of Correlation Among Different Construction Projects

The purpose of this section is to calculate the degree of correlation among different construction projects. As a result of the curse of dimensionality, it is impossible to use the AHP to calculate the degree of correlation among different construction projects. The new approach has been introduced to solve the problem, and the details of this approach are provided in Section 2.3.2. This study selects two construction projects and calculates the correlation matching degree of these projects separately. For example, on the basis of the pre-post relationships comparison matrix, the quantitative representation of the correlation matching degree of this factor in construction project A is as follows (the other comparison matrices are shown in the Supplementary Materials):

$$\begin{bmatrix} 1^{(1)} & 1/2^{(1.5)} & 1^{(1)} & 1^{(1)} \\ 2 & 1^{(1.5)} & 2 & 2 \\ 1 & 1/2^{(1.5)} & 1^{(1)} & 1^{(1)} \\ 1 & 1/2^{(1.5)} & 1 & 1^{(1)} \end{bmatrix} \quad (8)$$

The correlation scores of the two construction projects are shown in Table 7, according to all the comparison matrices with quantitative representations of the correlation matching degree and the weight of each factor.

Table 7. The general correlation scores of the two construction projects.

Construction Project	Pre-Post Relationships		Spatial Intersection		Project Name		Project Unit		Location Description		S
	S _{CD}	W	S _{CD}	W	S _{CD}	W	S _{CD}	W	S _{CD}	W	
A	9.6	0.801	12.1	0.465	5.7	0.261	2.6	0.144	5	0.232	35
B	14.4		16.1		5.1		3.7		5.4		44.7

The result shows that project B has higher correlation scores than project A. Pre-post relationships (SB = 14.4 versus SA = 9.6) and spatial intersection (SB = 16.1 versus SA = 12.1) are the main factors that determine their scores. Specifically, the description of pre-post relationships in construction project B is more consistent than that of construction project A. In addition, the spatial relationship is nearly coincident among the four approval items in construction project B, whereas the area of spatial intersection is less than 50% or even less in construction project A.

3.4. The Calculation of the General Degree of Correlation of Typical Approval Items and Complete Construction Projects

This study calculates the degree of correlation of 30,996 typical approval items and 7749 construction projects. Table 8 shows the results, including the matching correlation rate and the matching accuracy, and Table 9 shows the same results for the complete construction projects. Note that the resulting matching correlation rates are calculated using data uploaded by different departments, and the resulting matching accuracies are counted on a per-investigation basis. In addition, the thresholds of the degree of correlation for approval items and construction projects have been set to 0.55 and 28.0, respectively, based on practical experience.

The overall degree of correlation of typical approval items and different construction projects is satisfactory (higher than 70%). Two of the approval items, the land supply scheme and the land use permit, have high matching correlation rates and matching accuracies. The former values are approximately 90%, whereas the latter values even exceed 95% because the data for both of these approval items are complete and are less likely to change than those of the others. Moreover, given the decline in the quality of construction project approval data and the increase in illegal construction projects, the general degree of correlation of complete construction projects in 2016 is lower than that in 2015. These results indicate that the quality of construction project approval data should be improved, and the level of construction project supervision must be increased.

Table 8. The resulting general degrees of correlation for typical approval items.

Year	Items	Land Supply Scheme	Land Use Permit	Construction Permit	Environmental Impact Assessment
2015	Matching correlation rate	88.50%	95.20%	71.20%	85.30%
	Matching accuracy	90.10%	98.10%	85.60%	80.70%
2016	Matching correlation rate	91.20%	97.50%	80.20%	88.30%
	Matching accuracy	90.60%	98.20%	88.30%	85.60%
Total	Matching correlation rate	89.20%	96.20%	77.40%	87.40%
	Matching accuracy	90.10%	97.30%	86.20%	84.50%

Table 9. The resulting general degrees of correlation for complete construction projects.

Year	Number of Construction Projects	Matching Correlation Rate of All Construction Projects	Matching Accuracy of All Construction Projects
2015	4016	82.40%	80.70%
2016	3733	76.80%	75.10%

The experiments reported in Sections 3.1–3.4 demonstrate the core concepts used in the calculation of the degree of correlation; namely, the calculation of the degree of correlation of typical approval items within individual construction projects using AHP and the calculation of the degree of correlation among different construction projects using the new approach presented here. The two forms of the degree of correlation can be calculated using these two methods. The results provide useful suggestions for the supervision of construction projects.

However, our methods have several implicit flaws. The quantification of all of the factors considered in calculating both forms of the degree of correlation is subjective. In addition, this paper does not consider other relevant factors such as project progress and the project principal.

4. Conclusions

The fusion of multi-source government data has been applied to urban construction and supervision to assess the supervision degree and the legitimacy of construction projects in Tianjin. This study broadens the horizons of the fusion of government data and provides some suggestions to the construction supervision department.

The AHP and a new degree of correlation algorithm are used to calculate the correlation scores between the different approval items within individual construction projects and among different projects, respectively. The results show about 80% correlation for the construction projects and their approval items, which is relatively high.

Future work based on this study will include three main parts. First, the degree of correlation algorithm should be modified to improve the accuracy of the degree of correlation matching of construction projects. Second, the fusion of multi-source government data should be extended to fields including environmental protection, urban planning, and public safety. Third, a multi-source big government data computing framework such as Hadoop or Spark should be established to accelerate the calculation process.

Supplementary Materials: The following are available online at <http://www.mdpi.com/2220-9964/7/2/50/s1>.

Acknowledgments: We thank Yongjun Wu, Haodong Meng, Miss Yue Hu, and Ningyezi Peng for their help. In addition, this study was supported by the Special Foundation for Informatization of Tianjin (Project No. 201509108) and was supported by the National Natural Science Foundation of China (Project No. 41571438)

Author Contributions: All of the authors contributed to this study. Specifically, Shaoyi Wang and Zhongjie Sheng pre-processed the data and conducted the primary experiments, mapping, and analysis of the results. Shaoyi Wang organized the content and obtained research funding for the study. Yuliang Xi, Xiangyuan Ma, and Huihui Zhang introduced the research idea and designed the primary experiment. Ke Hu and Zhenbiao Han provided the research data collected in Tianjin. Shaoyi Wang wrote the manuscript. Mengjun Kang, Fu Ren, and Qingyun Du helped to revise the article.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Wang, H.; Xu, Z.; Pedrycz, W. An overview on the roles of fuzzy set techniques in big data processing: Trends, challenges and opportunities. *Knowl.-Based Syst.* **2017**, *118*, 15–30. [[CrossRef](#)]
2. Viktor, M.S.; Kenneth, C. *Big Data: A Revolution That Will Transform How We Live, Work, and Think*; Hodder Export, British: London, UK, 2013; ISBN 9781848547919.
3. Laney, D. *3D Data Management: Controlling Data Volume, Velocity and Variety*; Res. Note 6; META Group: San Francisco, CA, USA, 2001.
4. Chen, C.P.; Zhang, C.Y. Data-intensive applications, challenges, techniques and technologies: A survey on big data. *Inf. Sci.* **2014**, *275*, 314–347. [[CrossRef](#)]
5. Günther, W.A.; Mehrizi, M.H.R.; Huysman, M.; Feldberg, F. Debating big data: A literature review on realizing value from big data. *J. Strateg. Inf. Syst.* **2017**, *26*, 191–209. [[CrossRef](#)]
6. Chen, H.; Chiang, R.H.L.; Storey, V.C. Business intelligence and analytics: From big data to big impact. *MIS Q.* **2012**, *36*, 1165–1188.
7. Fichman, R.G.; Dos Santos, B.L.; Zheng, Z. Digital innovation as a fundamental and powerful concept in the information systems curriculum. *MIS Q.* **2014**, *38*, 329–353. [[CrossRef](#)]
8. Ndou, V. E-Government for Developing Countries: Opportunities and Challenges. *Electron. J. Inf. Syst. Dev. Ctries.* **2004**, *18*, 1–24. [[CrossRef](#)]
9. Morabito, V. Big Data and Analytics for Government Innovation. In *Big Data and Analytics*; Springer International Publishing: London, UK, 2015; pp. 23–45.
10. Alexopoulos, C.; Loukis, E.; Mouzakitis, S. Analysing the Characteristics of Open Government Data Sources in Greece. *J. Knowl. Econ.* **2015**, 1–33. [[CrossRef](#)]
11. Hashem, I.A.T.; Chang, V.; Anuar, N.B. The role of big data in smart city. *Int. J. Inf. Manag.* **2016**, *36*, 748–758. [[CrossRef](#)]
12. Batty, M. Big data, smart cities and city planning. *Dialogues Hum. Geogr.* **2013**, *3*, 274–279. [[CrossRef](#)]
13. Batty, M. Does Big Data Lead to Smarter Cities? Problems, Pitfalls and Opportunities. *J. Law Policy Inf. Soc.* **2015**, *11*, 127.
14. Manda, M.I.; Backhouse, J. An analysis of the barriers to e-government integration, interoperability and information sharing in developing countries: A systematic review of literature. In Proceedings of the African Conference in Information Systems and Technology, Accra, Ghana, 5–6 July 2016.
15. Gu, R. *Research on Techniques and Systems for Big Data Processing*; Nanjing University: Nanjing, China, 2016.

16. Cheng, X.Q.; Jin, X.L.; Yang, J.; Xu, J. Technological progress and trends of big data. *Sci. Technol. Rev.* **2016**, *36*, 49–58.
17. Zhu, Y.; Wang, L.; Liu, J.P.; Sun, L.J. Application of large RDBMS in E-government. *Sci. Surv. Mapp.* **2005**, *30*, 30–31.
18. Song, S. *The Research and Design of the E-government System Based on NoSQL*; Wuhan University of Technology: Wuhan, China, 2013.
19. Shirley, C.J.L.; Azlan, S.A.; Anuar, B.A. Analytic Hierarchy Process Decision-Making Framework for Procurement Strategy Selection in Building Maintenance Work. *J. Perform. Constr. Facil.* **2015**, *29*, 04014050.
20. Eylem, K.; Hasan, A.B. An Application of Analytic Hierarchy Process (AHP) in a Real-World Problem of Store Location Selection. *Adv. Manag. Appl. Econ.* **2015**, *5*, 41–50.
21. Lakshmi, T.; Ramakrishna, R. Resource Allocation in Project Scheduling Application of Fuzzy AHP. In Proceedings of the International Conference on Technology and Business Management, Dubai, United Arab Emirates, 23–25 March 2015; pp. 512–521.
22. Li, H.X.; Mohamed, A.H.; Lei, Z.; Ziad, A. Risk identification and assessment of modular construction utilizing fuzzy analytic hierarchy process (AHP) and simulation. *Can. J. Civ. Eng.* **2013**, *40*, 1184–1195. [[CrossRef](#)]
23. Wang, X.D.; Peng, X.Y. Application of Analytic Hierarchy Process in Evaluating Education Equipment Efficiency Factors. In Proceedings of the 2011 International Conference on Business Management and Electronic Information (BMEI), Guangzhou, China, 13–15 May 2011; IEEE: Piscataway, NJ, USA, 2011; pp. 96–100.
24. Sheng, Z.J.; Wang, S.Y.; Su, X. Thinking and practice of the construction of the government big data based on the location information in Tianjin. *Sci. Surv. Mapp.* **2016**, *41*, 198–200.
25. Ishizaka, A.; Pearman, C.; Nemery, P. AHPSort: An AHP-based method for sorting. *Int. J. Prod. Res.* **2012**, *50*, 4767–4784. [[CrossRef](#)]
26. Saaty, T. A scaling method for priorities in hierarchical structures. *J. Math. Psychol.* **1977**, *15*, 234–281. [[CrossRef](#)]
27. Anderson, T.J.; Wang, H.T.; Kang, B.S. Effect of bias voltage polarity on hydrogen sensing with AlGaIn/GaN Schottky diodes. *Appl. Surf. Sci.* **2008**, *255*, 2524–2526. [[CrossRef](#)]
28. Sipahi, S.; Timor, M. The analytic hierarchy process and analytic network process: An overview of applications. *Manag. Decis.* **2010**, *48*, 775–808. [[CrossRef](#)]
29. Vargas, L.G. An overview of the analytic hierarchy process and its applications. *Eur. J. Oper. Res.* **1990**, *48*, 2–8. [[CrossRef](#)]
30. Saaty, T.L.; Hu, G. Ranking by eigenvector versus other methods in the Analytic Hierarchy Process. *Appl. Math. Lett.* **1998**, *11*, 121–125. [[CrossRef](#)]
31. Saaty, T.L.; Erdener, E. A new approach to performance measurement, the analytic hierarchy process. *Des. Methods Theor.* **1979**, *13*, 64–72.
32. An, S.; Zhang, X.; Wang, J. Finding Causes of Irregular Headways Integrating Data Mining and AHP. *ISPRS Int. J. Geo-Inf.* **2015**, *4*, 2604–2618. [[CrossRef](#)]
33. Saaty, T.L. Relative measurement and its generalization in decision making why pairwise comparisons are central in mathematics for the measurement of intangible factors the analytic hierarchy/network process. *Rev. Real Acad. Cienc. Exactas Fis. Nat. Ser. A Mat.* **2008**, *102*, 251–318. [[CrossRef](#)]
34. Yoshua, B.; Rejean, D.; Pascal, V.; Christian, J. A neural probabilistic Language Model. *J. Mach. Learn. Res.* **2003**, *3*, 1137–1155.

