*Forest Chemicals Review*
*www.forestchemicalsreview.com*
*ISSN: 1520-0191*
*May-June 2022 Page No. 1946-1975*
*Article History: Received: 24 February 2022, Revised: 05 April 2022, Accepted: 08 May 2022, Publication: 30 June 2022*

# Construction and Urban Analysis of Smart City Construction Level Evaluation Index System in Jilin Province

**Xingmei Xu, Ning Wang, Yuan Long, Chuang Wang***
College of Information Technology, Jilin Agricultural University, Changchun, China
*Corresponding author.

***Abstract:***

Around the world, due to the development level of cities and the development direction is not consistent, so did not have a set of suitable for wisdom level evaluation index of the city, each city development experiment based on the provincial project of jilin province science and technology development plan project tasks, this paper studies the purpose is to establish a set of suitable for jilin province evaluation standard wisdom urban construction level evaluation index system. This paper takes the policy documents published by jilin Provincial People's Government as the data source, and the time range is set from June 2016 to June 2021. A series of data mining techniques are adopted, based on word frequency and hierarchical clustering algorithm analysis, combined with experts in this field and relevant authoritative literatures. The evaluation index system of smart city construction level in Jilin Province is proposed, which contains 3 first-level indexes, 11 second-level indexes and 44 third-level indexes, and the evaluation index system of smart city construction level with Jilin province characteristics is successfully constructed. In addition to apply evaluation index system of practice, this paper USES the fuzzy comparison matrix, index weight and consistency check method is analyzed, in jilin province bureau of statistics released in 2021 the city of changchun, jilin, four in the GDP, siping, matsubara, a reference to review relevant literature data, the experimental results is unanimously recognized experts.

***Keywords****: Data mining; Jilin province; Wisdom city*

## I. INTRODUCTION

With the rapid development of smart city, it is urgent to construct a complete index evaluation system of smart city. Due to different development conditions, different city sizes, different urban development centers and other factors, it is difficult to develop a common standard to guide and measure the development of smart cities in the world or even in the same country.

## II. RELEVANT OVERVIEW

### 2.1 Smart city index system at home and abroad

In domestic, there are more representative of the national literature of the new evaluation indicators (2016), wisdom city, on the basis of perfecting the new evaluation indicators (2018), wisdom city, urban development and the design institute of Shanghai pudong wisdom wisdom urban evaluation index system of 1.0 and 2.0, the evaluation system, the index has certain representativeness, Data are better collected, but some indicators are still too subjective. All the existing smart city evaluation systems are defective, and it is difficult to guide and evaluate the development of smart city in Jilin Province. Therefore, when formulating the evaluation index system of smart cities in Jilin Province, regional, timeliness, scientific, representative and guidance should be taken into comprehensive consideration. Therefore, this paper chooses the policy documents issued by the People's government of Jilin Province as the data source for analysis.

At present, in China, Su Jingqin, Li Xiaogang et al discussed the relationship between national and local technological innovation policies by using co-word analysis method, and obtained the internal relationship between central and local technological innovation policies[1];Freitas and Nick in foreign countries classified s&T innovation policies and established a coding framework of policy planning containing 46 projects, which reflected the scientific nature and applicability of the research on s&T innovation policy structure[2];Junseop Shim et al. studied the nuclear energy policies of six countries, including the United Kingdom and the United States, and constructed the policy framework with the help of semantic network analysis[3].

Combined with the research of the above scholars, this experimental scheme is based on the relevant research on the policy documents issued by the People's Government of Jilin Province, and collects relevant literature and expert suggestions, aiming to establish a scientific evaluation index system of the standard and local characteristics of the smart city construction level in Jilin Province.

### 2.2 Research status of feature selection algorithms in text classification

Text classification generally includes text preprocessing, text model representation, feature selection, classification model training and performance evaluation. The text preprocessing process mainly includes word segmentation and the removal of stop words in the text data set. The main feature extraction algorithm TFIDF will be analyzed below.

The TFIDF feature weight depends on the term's contribution to the document containing the term. Due to its simple mathematical formula, the overall complexity of the algorithm has lower accurate feature selection, become the most commonly used text feature selection methods, but due to the original algorithm does not consider the difference between data sets text category and class information between text vector distribution in the class, so the tilt of the dataset class classification effect can't get the ideal effect.

In view of the above shortcomings, domestic and foreign scholars put forward a variety of improved TFIDF algorithm. AtanuDey et al. improved TFIDF algorithm by constructing N-gram emotional features, which solved the problem that traditional TFIDF only evaluated feature vectors of unary model Unigram or N-Gram, resulting in poor classification effect. The improved algorithm firstly extracted emotional words and their enhancement or negation words from the review data to construct N-gram emotional features, and then combined with TFIDF and N-Gram emotional features to weight the feature words[4].Manny Rayner et al. proposed a dynamic programming method to optimize TFIDF algorithm by matching grammatical strings, which improved the classification effect of the algorithm in the sparse domain of complex speech data[5]. Such as sky L TFIDF no attention to the key words of text contained in the semantic information, by introducing a term meaning the LSA to improve TFIDF analysis method, the improved algorithm using singular value decomposition feature vector, and then through calculation results decomposition row vector cosine to identify keywords, selection of the similarity between feature, make up for the deficiency of the TFIDF[6]Mohamad Irfan et al. improved the traditional TFIDF algorithm by introducing the fuzzy C-means method and proposed an improved feature weighting algorithm c-TFIDF based on the C-means method. The improved algorithm uses c-means method to weight the result sentences. The high-weight sentences and low-weight sentences are divided into two groups, and the contribution degree of each sentence in the document is determined, so as to improve the accuracy of feature extraction[7]Aiming at TFIDF algorithm, Hao Jianlin et al. did not consider the imbalance of distribution between and within characteristic word classes in the classification system, and reduced the influence of word frequency on the model by limiting word frequency, that is, to optimize TFIDF algorithm by improving the probability of word frequency selection. The algorithm proposes two assumptions: the word frequency of function words is not less than α times the sum of all function words, and the total number of function words and documents is not more than β times the total number of documents. Although this algorithm improves the accuracy of feature extraction, it is not good for feature extraction of text category skew data sets[8]. Zhao Shenghui et al. optimized the traditional TFIDF algorithm by correcting the distribution differences of feature items among document categories. The category coefficient is introduced to indicate the classification degree of feature words, and the weight value of feature words with strong classification ability is increased, while the weight value of fresh words and noisy words is reduced[9]However, this algorithm still cannot reflect the distribution differences between and within text classes, and has some limitations.

Yuan Na et al., from Wuhan University of Technology, fully compensated the two major problems of unbalanced distribution between classes in TFIDF feature weighting algorithm and differences in distribution between classes and within classes in Chinese vector classification system by introducing distribution factors between classes, between classes and within classes.Therefore, this paper introduces this algorithm to study.

## III DATA PROCESSING

### 3.1 Data Collection

Under the guidance of the document released by the state, municipal governments will further divide the focus and details of smart cities according to the content and requirements of the document as well as the local characteristics of each city. It can be seen that the documents issued by local people's governments are actually the concretization and regionalization of national policies. Therefore, the policy guidance of our province on smart city can be better seen on the websites of people's governments at all levels.

Although there are many policies on the Internet of Jilin Provincial People's Government, they are all in the form of links and cannot be directly used as experimental data. Therefore, this paper adopts the web crawler technology to collect the links of smart city-related policy documents, capture the policy text through the links, and save it into an easy-to-process format. Through the web crawler, the smart city application is summarized from Table I under the government affairs open list policy document database of Jilin Provincial People's Government [4]. According to the crawling demand and policy characteristics of Jilin Provincial government, 21 keywords related to smart city construction are selected and summarized, which is also the smart city application involved in this paper. Table II shows the crawler crawler object.

**TABLE I APPLICATION SUMMARY OF SMART CITY**

| APPLICATION | PROVENANCE |
|---|---|
| Smart citizens, smart life, smart governance, smart scripture<br>Economy, intelligent transportation, intelligent environment | Rudolf Giffinger[10] (2007) |
| Smart transportation, smart public utilities, smart water resources, smart buildings, smart public security, smart management,<br>Urban Smart Center | IBM (2009). |
| Smart government, smart healthcare, smart education, smart transportation, smart security, smart energy, smart enterprise, smart community, smart public services, smart logistics | Wu Shengwu and Yan Guoqing (2010) [11] |
| Public service class (government, medical wisdom, wisdom, culture education wisdom) industry development class (wisdom industry, logistics, e-commerce, wisdom) social management class (intelligent community, the wisdom of public safety, environmental protection, wisdom, food and drug safety) municipal facilities class (energy saving) transportation, municipal administration of wisdom, and wisdom | Wu Yulong and AI Haojun(2011) [12] |

| | |
|---|---|
| Smart transportation, smart electricity, smart finance, smart cars, smart hospitals, smart schools, smart enterprises, smart farms, smart families and smart communities, smart cities | Qian Zhixin [6] (2011)<br>[30] |
| Smart water management, public safety management, smart transportation, chronic disease management of smart medical treatment, professional talent training program of smart education, smart business, smart port and navigation, smart building, etc | Yue Meiying [7] (2012)<br>[31] |
| Smart infrastructure, smart governance, smart livelihoods, smart economy, smart environment, and smart planning and construction | Gu and Qiao[8] (2012) |
| Broadband areas, smart transportation, smart health, smart education, smart management, smart energy, energy efficiency and natural resources | Ernst and Young[9] (2016) |
| Smart infrastructure, smart government, smart livelihood, smart production and innovation driven | Ren Liang, Zhang Haitao et al. [13] (2019) |

TABLE II OBJECTS CRAWLED BY CRAWLERS

| **Crawl object** | **Crawl content** |
|---|---|
| keywords | Smart city, smart community, smart business circle, smart pension, smart medical care, smart campus, smart property, smart logistics, smart hotel, smart transportation, smart government, smart tourism, smart scenic spot, smart security, smart parking, smart agriculture, smart environmental protection, smart ecological, smart urban management, smart education Wisdom energy |
| Name of Government Website | Changchun People's Government Jilin People's Government Siping People's government<br>Liaoyuan People's Government Tonghua People's Government Baicheng People's Government Yanbian Prefecture people's government<br>Changbai Mountain Management Committee People's Government Changchun New District Management Committee |
| | People's Government of Jilin Province |

A total of 16,501 policy documents were captured, covering June 2016 to June 2021.Policy documents are the documents of jilin Provincial People's government, excluding announcements and departmental documents. Table III shows the smart city policies issued by Jilin Province and relevant departments

**TABLE III SMART CITY RELATED POLICIES ISSUED BY JILIN PROVINCE AND RELEVANT DEPARTMENTS**

| SERIAL NUMBER | KEYWORDS | ARTICLE NUMBER |
|---|---|---|
| 1 | Wisdom city | 1180 |
| 2 | Intelligence community | 955 |
| 3 | Wisdom business circle | 2741 |
| 4 | Wisdom endowment | 475 |
| 5 | smart medical care | 584 |
| 6 | Wisdom campus | 246 |
| 7 | Wisdom agriculture | 640 |
| 8 | Intellectual property | 203 |
| 9 | Intelligent transportation | 696 |
| 10 | Wisdom government affairs | 537 |
| 11 | Wisdom of tourism | 588 |
| 12 | Wisdom scenic spot | 201 |
| 13 | Intelligent security | 2741 |
| 14 | Wisdom parking | 840 |
| 15 | Wisdom hotel | 80 |
| 16 | Wisdom green | 444 |
| 17 | Ecological wisdom | 804 |
| 18 | Wisdom of the watch | 923 |
| 19 | Education wisdom | 1313 |
| 20 | Wisdom energy | 310 |
| A total of |  | 16501 |

## 3.2 Optimization of feature weighting algorithm

TFIDF is the most commonly used feature weighting method. The basic algorithm idea is that the weight value of a feature item in the text is proportional to the frequency of the feature item in the current text, and inversely proportional to the number of texts containing the feature item in the text data set. To sum up, TFIDF algorithm has the advantages of simple mathematical calculation formula and low operation complexity. However, it also has certain limitations, mainly including the following two points:

(1) The unbalanced distribution among categories of data sets is not considered

In reality, the classification and distribution of data sets are mostly unbalanced, and different classification and distribution often have certain differences. However, TFIDF algorithm does not take this difference into account, and the weights of the feature vectors calculated are only based on the number of documents.

When the category distribution of data sets differs greatly, especially the weak category distribution, the calculated weight value will be very small, affecting the classification accuracy and failing to reflect the distribution difference of text vectors between various data sets.

(2) The difference of text vector distribution between classes and within classes in the classification system is not reflected correctly

The distribution of the text vector of the class is the problem that needs to be considered. When the Tfij value of the word frequency of the function word is great in class Ti Cj, but the Tfij value of the word frequency of other classes is very small, the function word should also reflect the degree of difference of the text category and should be placed in a high weight.

When it appears in most categories, and the proportion in each category is not significantly different, a lower weight should be given. The intra-class distribution of text vectors needs to be considered. When the feature term Ti is in the Cj class, the feature term with relatively consistent distribution in the class should be given higher weight. However, when feature item only appears in a few documents of the same kind and rarely appears in other documents of the same kind, it may be a special term that cannot well reflect the category information of the text and should be given a low weight.

Therefore, this section adopts the improved TFIDF algorithm fdCD-TFIDF based on word frequency distribution factor and category distribution factor.

Interclass distribution factor α:

Represents the distribution of feature items across document classes. Calculate the quotient between the number of documents containing feature term Ti in Cj class AIj and the number of documents containing feature term Ti in non-CJ class CI, and then take the logarithm to get. The formula is as follows:

$$\alpha = \log\left(\frac{a_{ij}}{c_i}\right)$$

Intra-class distribution factor β:

Represents the distribution of feature items across document classes. Calculate the quotient between the number of documents containing feature term Ti in Cj class AIj and the number of documents containing feature term Ti in non-CJ class CI, and then take the logarithm to get. The formula is as follows:

$$\beta = \log\left(2 + \frac{a_{i,j}}{1 + b_j}\right)$$

Quasi-distribution factor γ:

Represents distribution information for each category of the document. Calculate the quotient of the total number of documents in dataset N divided by the total number of documents contained in category Cj nj, and then take the logarithm, which is defined as follows:

$$\gamma = \log\left(\frac{N}{n_j}\right)$$

Therefore, the improved weight calculation formula is as follows:

*Article History: Received: 24 February 2022, Revised: 05 April 2022, Accepted: 08 May 2022, Publication: 30 June 2022*

$$\mathrm{FDCD} - \mathrm{TFIDF} = \mathrm{tf}_{\mathrm{i,j}} \times \mathrm{idf}_{\mathrm{i}} \times \alpha \times \beta \times \gamma$$

Comparative experimental design

The comparison data is from the open Sogou Laboratory text classification corpus, from which nine categories such as finance and economics are selected for comparison experiments, and 1000 files are selected from each category for experiment. Original TFIDF algorithm, FD-TFIDF algorithm[14] Fdcd-tfidf algorithm uses SVM classifier and KNN classifier respectively to classify class-balanced data sets. The classification results of the three algorithms are shown in TABLE IV and TABLE V. Fig 1 and Fig 2 show the f1-score comparison results of the three algorithms for the nine classification documents in the category-balanced data set.

**TABLE IV CLASSIFICATION RESULT ALGORITHM OF SVM USING ORIGINAL TFIDF ALGORITHM, FD-TFIDF ALGORITHM AND FDCD-TFIDF ALGORITHM**

| ALGORITHM | PRECISION (%) | RECALL (%) | F1-SCORE (%) |
|---|---|---|---|
| TFIDF+SVM | 86.090 | 85.569 | 85.632 |
| FD-TFIDF+SVM | 88.730 | 88.356 | 88.539 |
| FDCD-TFIDF+SVM | 89.930 | 89.632 | 89.600 |

**TABLE V KNN CLASSIFICATION RESULTS OF ORIGINAL TFIDF ALGORITHM, FD-TFIDF ALGORITHM AND FDCD-TFIDF ALGORITHM**

| ALGORITHM | PRECISION (%) | RECALL (%) | F1-SCORE (%) |
|---|---|---|---|
| TFIDF+KNN | 86.790 | 86.240 | 85.247 |
| FD-TFIDF+KNN | 89.693 | 89.176 | 88.193 |
| FDCD-TFIDF+KNN | 90.930 | 90.654 | 90.693 |

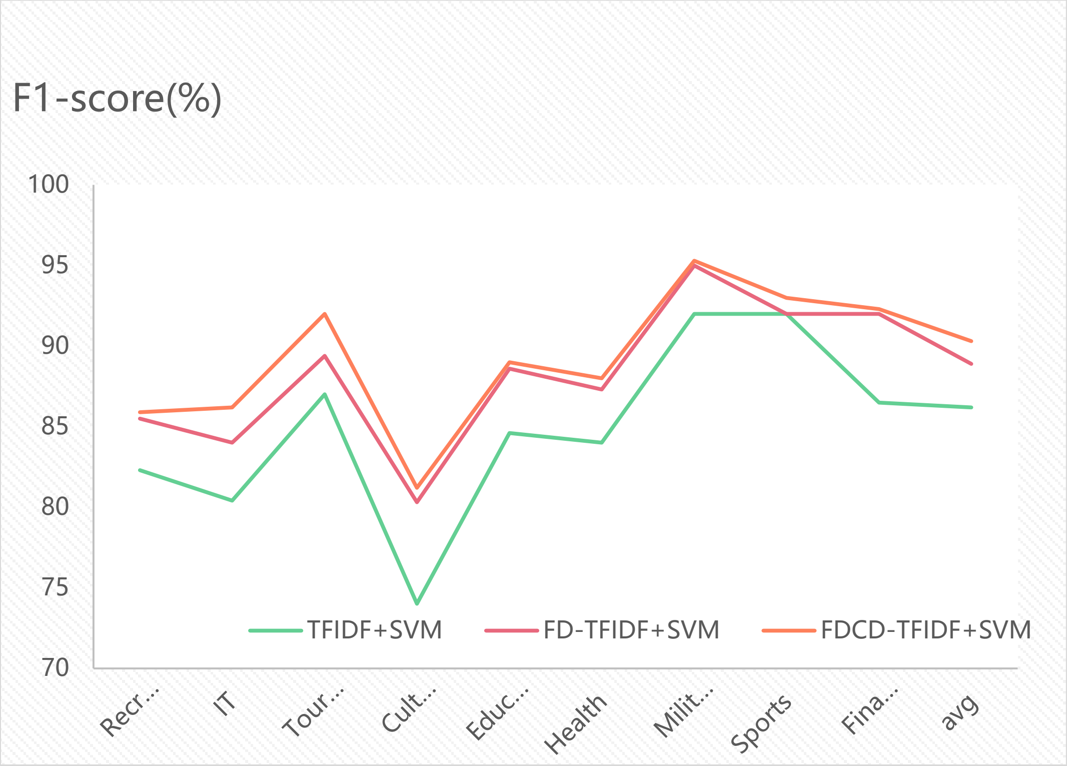

Fig 1 F1 score comparison results of three algorithms using SVM classification

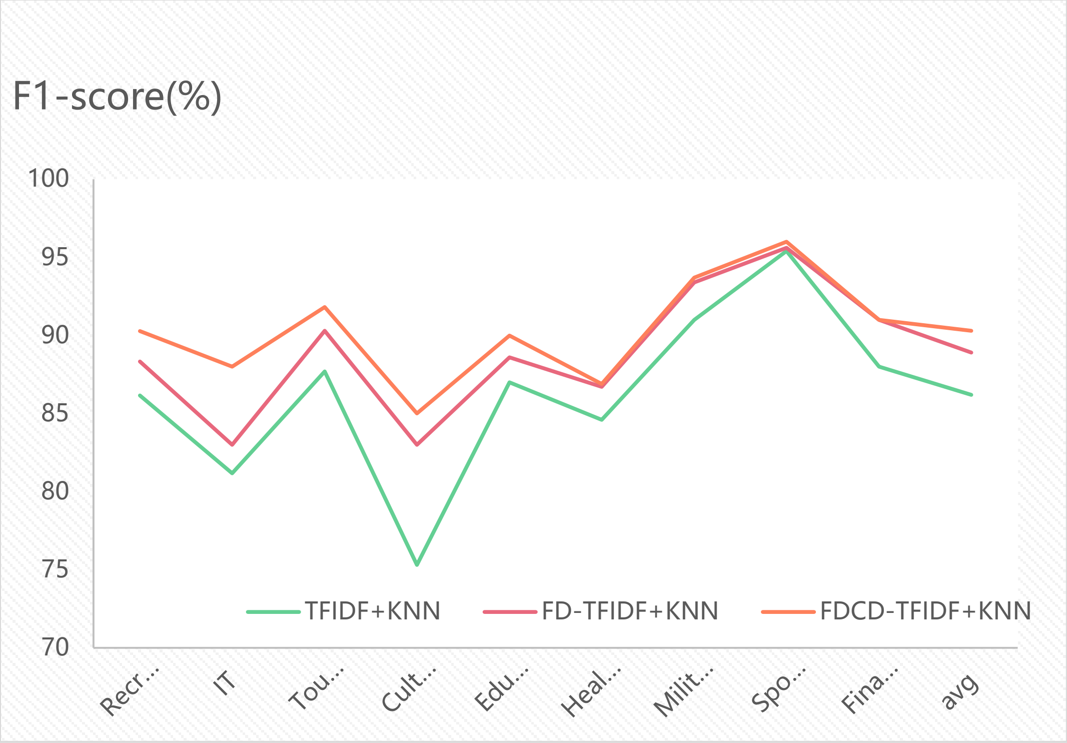

Fig 2 classification results of KNN used by original TFIDF algorithm, fd-tfidf algorithm and fdcd-tfidf algorithm

*Article History: Received: 24 February 2022, Revised: 05 April 2022, Accepted: 08 May 2022, Publication: 30 June 2022*

By comparing the average accuracy of Table IV and Table V, FD-TFIDF algorithm performs best when SVM classification is adopted, with its accuracy reaching 89.584%, 3.535% higher than TFIDF algorithm and similar to FD-TFIDF algorithm. The figure rose 0.754%.When KNN classification was adopted, FDCD-TFIDF algorithm also had the highest accuracy, reaching 90.829%, 4.059% higher than TFIDF algorithm and similar to FD-TFIDF algorithm. By comparing the average Recall value of the two tables, it can be found that when SVM classification is adopted, the Recall value of FDCD-TFIDF algorithm reaches 89.218%, which is higher than that of TFIDF algorithm and close to that of FD-TFIDF algorithm. When KNN classification was adopted, FDCD-TFIDF algorithm also performed best, and Recall was also higher than TFIDF algorithm, which was similar to FD-TFIDF algorithm. According to f1-Score comparison results in Fig 1 and Fig 2, FDCD-TFIDF algorithm and FD-TFIDF algorithm performed best when SVM classification was adopted, both far higher than F1-Score of TFIDF algorithm. When KNN classification is adopted, FDCD-TFIDF algorithm performs best, FOLLOWED by FD-TFIDF algorithm, and TFIDF algorithm performs worst. As can be seen from f1-score comparison results of all classifiers, f1-Score of FDCD-TFIDF algorithm is higher than that of original TFIDF algorithm in 9 classifiers regardless of which classifier SVM and KNN adopt, and its performance is similar to FD-TFIDF algorithm.

Based on the above analysis, FDCD-TFIDF improves the feature selection by introducing the word frequency distribution factor, which is superior to the traditional TFIDF algorithm in data set. The classification result is similar to FD-TFIDF algorithm, and the word frequency is improved. At the same time, the improved algorithm improves the accuracy of support vector machine and KNN in text classification. Experimental results show that the word frequency distribution factor introduced by the improved algorithm is reasonable, and the effectiveness of the improved algorithm is verified.

## IV DATA ANALYSIS

### 4.1 Weight statistical analysis based on FDCD-TFIDF algorithm

All policy documents of Jilin Province were retrieved according to the smart city application summary table. FDCD-TFIDF algorithm is used to calculate the weight of the retrieved files. Finally, use the WordCloud tool library to perform the visual display in WordCloud form, as shown in TABLE VI.

**TABLE VI JILIN PROVINCE SMART CITY DOMAIN THESAURUS**

| **THESAURUS KEYWORDS** | | | | | | | |
|---|---|---|---|---|---|---|---|
| Intelligent | technology | information | data | epidemic | party | intelligent | platform |
| Science | project | plan | Intelligent | plan | agriculture | media | stage |
| Epidemic | vector | evaluation | traditional | system | standard | planning | effect |

| **THESAURUS KEYWORDS** | | | | | | | |
|---|---|---|---|---|---|---|---|
| Internet | security | ecological | model | state | hardware | industry | chain |
| Report | synergetic | safety | quality | time | assisted | relationship | quality |
| Personnel | Processx | outcomes | practice | overall | content | street | health care |
| Service | municipal | government | elements | scale | direction | national | development |
| Automobile | central | artificial | node | opportunity | terminal | population | city |
| …… | | | | | | | |

Some feature words and their weights in the thesaurus are shown in TABLE VII below.

**TABLE VII SECTION FEATURES AND WEIGHT**

| **KEY WORDS** | **WEIGHT VALUE** | **KEY WORDS** | **WEIGHT VALUE** |
|---|---|---|---|
| wisdom | 0.18349648081224758 | information | 0.071219267225482267 |
| technology | 0.17417335917247557 | smart | 0.060011178696416932 |
| data | 0.12191311638625407 | platform | 0.062175692672899021 |
| epidemic | 0.10526541756926788 | project | 0.061882654176791535 |
| central | 0.08360297619406842 | Science and technology | 0.049315225448729636 |
| …… | | | |

It can be found from the tab that smart city construction hotspots in Jilin province cover a wide range, including multiple dimensions from macro to micro and from top to bottom, which can better reflect the overall development of smart city construction in Jilin Province from the side.In the word cloud, words such as "wisdom", "technology" and "data" have the highest weight and are also the hot topics of concern, indicating that the government has put forward higher requirements for technology, intelligence and data issues in smart city construction. Words with high weights such as "science and technology", "industry", "information" and "enterprise" also reflect the importance jilin Provincial government attaches to the development of science and technology.Science and technology will gradually become the main driving force for the rapid development and stable operation of smart cities.In addition, words such as "epidemic situation", "hospital" and "medical treatment" have high weights due to the outbreak of COVID-19, which is also testing the governance capacity of smart cities at the present stage.I believe that after the test of the epidemic, the construction of smart cities will be more urgent than ever, and cities will become more "smart".

### 4.2 Cluster analysis based on hierarchical clustering algorithm

Cluster analysis is a means to understand internal relations and explore things. The concept of "birds in groups" is the most simple and intuitive reflection. Its purpose is to divide a data set into several clusters and make data objects have high similarity in the same cluster, while data objects of different clusters have low similarity.

Therefore, through hierarchical clustering algorithm, cluster analysis is carried out on some smart city applications (20 in total except "smart city"), and smart city applications with different similarity are clustered. According to the clustering results , the current smart city construction in Our province can be divided into 3 categories and 11 subcategories (TABLE VIII).

**TABLE VIII SMART CITY APPLICATION CLASSIFICATION**

| CATEGORIES OF LABEL | SMALL CLASS | SMART CITY APPLICATION |
|---|---|---|
| Digital economy of Jilin Province | The travel industry | Smart scenic spots and smart tourism |
| | Commercial construction | Wisdom business circle |
| | Hotel service | Wisdom hotel |
| Infrastructure construction in Jilin Province | Transport system | Smart parking, smart transportation, smart logistics, smart security |
| | Life safeguard | Smart medical care, smart community, smart elderly care, smart urban management |
| | Environmental protection energy | Smart energy, smart environmental protection |
| | Realty service | Intellectual property |
| | Education popularization | Smart education, smart campus |
| | Government affair management office | Wisdom government affairs |
| Humanistic Ecology of Jilin Province | Ecological management | Ecological wisdom |
| | Agricultural level | Wisdom agriculture |

Can be seen from table 4, the first major categories are digital economy in jilin province, it can be divided into three classes, including tourism scenic spots and wisdom, the wisdom of the wisdom of the commercial building business circle, the wisdom of the hotel services, strengthen urban construction and application of wisdom, improve the living standards of the people, rich people's amateur life, increase happiness life in the city, at the same time, also can promote consumption, Increase consumption power, which is conducive to promoting the development of urban economy. The second is the jilin province infrastructure construction, it can be divided into six categories, respectively is the intelligent traffic system, parking, transportation, wisdom logistics, security, wisdom, intelligence community life, wisdom, intelligence, wisdom urban pension, environmental protection energy, environmental protection, intellectual property services, wisdom education popularization education, intelligence, wisdom, campus e-government affairs. The construction of such smart city application is conducive to urban transportation infrastructure construction, medical security, logistics management, government administration, etc., and the improvement of smart city infrastructure construction is conducive to improving the convenience of urban life and improving residents' satisfaction. The third category is human ecology in Jilin Province, which can be divided into two subcategories: intelligent ecology of ecological governance and intelligent agriculture of agricultural level. The construction of such smart city application is conducive to improving people's awareness of environmental protection and promoting the harmonious coexistence and development of man and nature.

## V. CONSTRUCTION OF SMART CITY CONSTRUCTION LEVEL EVALUATION INDEX SYSTEM

### 5.1 Principles and ideas of evaluation index system construction

Wisdom city construction in our country is still at an early stage, the intelligence level of the index system of urban construction survey also at the exploration stage, in the existing evaluation index system, authority is low, but also exists many problems, analyze the principle of index system construction, and combining with index system based on the corresponding authority of the state and relevant departments to build standard, The construction principles of this index evaluation system are summarized as follows:

(1) Systematicness: The evaluation index system of smart city construction is a circulable system, requiring each index to be logical and systematic to a certain extent.

(2) Scientific: With the development concept of smart city as the core, strictly grasp the scientific nature of the index evaluation system. The acquisition of evaluation indexes should be reasonable, grounded and not imaginary, so as to ensure the scientific, effective, normative and practical evaluation indexes obtained.

(3) Guidance: The evaluation index system should not only analyze the current situation of smart city construction, but also provide guidance for subsequent construction. Therefore, the selection of indicators should have guiding significance, that is, realistic guiding significance, which can standardize and guide the next development of smart city construction.

(4) Integrity: Smart city construction is a complex project with multi-domain and multi-dimensional design. Therefore, the selection of index evaluation should be comprehensive, systematic, multidimensional and complete, reflecting all aspects of smart city construction as far as possible.

(5) Mutual exclusion: While ensuring full diversity of evaluation indexes, there is no duplication or interleaving among evaluation indexes.

(6) Operability: The evaluation index system should have strong operability, requiring that all indicators are simple and intuitive, and can be quantified and summarized in a standardized way, so as to facilitate the actual operation and implementation of the index evaluation system.

Based on the results of data mining and following the construction principle of the evaluation index system, the index framework of the evaluation index system of smart city construction level is designed into three levels. Firstly, the first-level and second-level indexes are determined according to the results of hierarchical cluster analysis. Then, on the basis of literature research, similar words were extracted based on Word2vec, and three-level evaluation indexes were extracted combining with authoritative evaluation indexes.

### 5.2 Preliminary construction of evaluation index system

Different related words are extracted from word vectors trained by Word2vec.Taking "smart government affairs" as an example, extract the 30 words with the highest correlation, as shown in TABLE IX.

TABLE IX TOP 30 WORDS RELATED TO "SMART GOVERNMENT AFFAIRS"

| WORDS | SIMILARITY | WORDS | SIMILARITY |
|---|---|---|---|
| E-government services | 0.5234 | The matters | 0.3585 |
| The pipes suit | 0.4406 | Government governance capacity | 0.3557 |
| Handle affairs hall | 0.4173 | The government | 0.3529 |
| The electronic government affairs | 0.4164 | tax | 0.3455 |
| Wisdom city | 0.4155 | Free net phone | 0.3407 |
| The crowd | 0.3972 | Social governance | 0.3389 |
| The government department | 0.3955 | The people's livelihood | 0.3365 |
| Open government | 0.3875 | Integrated platform | 0.3254 |
| Intelligent transportation | 0.3830 | And better to delegate | 0.3238 |
| Administrative examination and approval | 0.3790 | Sharing platform | 0.3203 |
| The judicial | 0.3756 | The public service | 0.3155 |
| Administrative information | 0.3710 | efficiency | 0.3140 |

| | | | |
|---|---|---|---|
| Smart City construction | 0.3703 | For the convenience of | 0.3110 |
| The border | 0.3690 | Information synergy | 0.3110 |
| Service-oriented government | 0.3612 | Service network | 0.3077 |

As can be seen from the above table, the extracted related words are multi-dimensional, covering macro to micro, and comprehensively reflect the construction hot spots and demands of "smart government affairs". Combined with some authoritative evaluation index system, select and conclude the evaluation index which is easy to be quantified, easy to collect and universal. According to the international standard proposal "Smart City ICT Reference Framework", "Smart City ICT Evaluation Indicators" and "Smart City EVALUATION Model and Basic Evaluation Indicator System", Part 4: Construction Management (GB/T 34680.4-2018), the three level indicators are refined. The three indexes corresponding to "smart government" are: citizens' satisfaction with one-stop management and e-government, government coverage, government service openness, and government information coordination. Other three-level indicators adopt this method in turn, and finally get the preliminarily selected indicator system of smart city construction, as shown in TABLE X.

**TABLE X PRELIMINARY SELECTION OF SMART CITY CONSTRUCTION INDEX SYSTEM**

| **LEVEL INDICATORS** | **THE SECONDARY INDICATORS** | **SMART CITY APPLICATION** | **LEVEL 3 INDICATORS** |
|---|---|---|---|
| Digital economy of Jilin Province | The travel industry | Wisdom scenic spot | Per capita park green area |
| | | Wisdom of tourism | Information service satisfaction in tourism industry |
| | Commercial construction | Wisdom business circle | Online shopping satisfaction<br>Proportion of online commodity retail<br>The richness of online consumption platform satisfaction,<br>Satisfaction with convenient payment<br>Public resource trading platform |
| | Hotel service | Wisdom hotel | Information service satisfaction in the hospitality industry |

| | | | |
|---|---|---|---|
| Infrastructure construction in Jilin Province | Transport system | Wisdom parking | Parking satisfaction |
| | | Intelligent transportation | Accuracy of real-time traffic information<br>Smart light pole construction level<br>Rush hour traffic jams |
| | | Wisdom logistics | Warehouse unmanned management degree<br>Logistics delivery satisfaction level |
| | | Intelligent security | Video resource collection and coverage rate<br>Network sharing of video resources |
| | Life safeguard | smart medical care | Satisfaction with payment of medical fees on self-service online platform<br>Online reservation and electronic medical record popularity satisfaction<br>Social security service satisfaction in different places<br>Coverage of basic medical insurance |
| | | Intelligence community | Level of community service information |
| | | Wisdom endowment | Coverage of basic old-age insurance |
| | | Wisdom of the watch | Digital urban management situation |
| | Environmental protection energy | Wisdom energy | New energy utilization rate<br>Energy efficiency |
| | | Wisdom green | Harmless treatment rate of household garbage<br>Comprehensive utilization and disposal rate of industrial solid waste<br>Ambient air quality good rate<br>Sewage treatment rate<br>Rate of disposal of environmental problems |

| | | | |
|---|---|---|---|
| | | | Environmental protection information disclosure rate of enterprises and institutions |
| | Realty service | Intellectual property | Property Service Satisfaction |
| | Education popularization | Education wisdom | The development level of network education<br>Ease of access to online education resources |
| | | Wisdom campus | School multimedia classroom penetration rate<br>School wireless network coverage |
| | Government affair management office | Wisdom government affairs | Public satisfaction with one-stop handling<br>E-government Coverage<br>Government services openness<br>Level of government information coordination |
| Characteristic ecological and humanistic development in Jilin Province | Ecological management | Ecological wisdom | Satisfaction with urban environmental quality early warning<br>Satisfaction with urban environment improvement<br>Afforestation coverage rate of built-up area |
| | Living level | Intelligent life | Degree of digitization of residence |

### 4.3 Screening and revision of evaluation index system

After determining the initial indicator system of smart city, we are still one step away from the final determination, namely, the screening and revision of indicators. In accordance with the principles and ideas of the paper the evaluation index system construction on the six principles, you need to choose questionnaire to collect expert advice for further screening on in this paper, the evaluation system, eliminate wisdom in urban development in jilin province is not so important, or low correlation index, at the same time to supplement indicators in deficiency, Thus, an index system with an appropriate number of indicators is obtained, which is sufficient to measure the level of smart city construction in Jilin Province for comprehensive evaluation.

This paper follows the basic principles and requirements of questionnaire design and sets up expert questionnaire. Nineteen experts were invited to give scores, including civil servants who have been engaged in smart city for many years, scholars in the field of smart city with vice-high title or above, and various personnel in Jilin Province. Questionnaires were distributed to experts through E-mail and other means, and the experts assigned values for the importance of each indicator.

Membership degree analysis was conducted on the returned questionnaires, and the indexes with low membership degree of the original initial indexes were removed. The membership degree analysis steps were as follows:

(1) Let {X} be a set of first, second and third level indexes in the initial index system, $X_k$It's the KTH index in the set X, $R_k$Is the index X in the index system$_k$ The membership degree of $M_k$Is that $X_k$Indicator number of indispensable experts, $T_k$Is the total number of experts participating in the survey, and the formula is as follows:

$$R_k = \frac{M_k}{T_k}$$

Type in the $R_k$The value of is between 0 and 1, and the closer the value is to 1, it represents $X_k$The more important.

(2) less than membership degree index of critical value of 0.4, combined weights above statistics are then step screening, and according to the questionnaire of supplementary indicators of open questions, further investigation of questionnaire survey, after careful analysis and discussion, finally identified a set consists of 4 first-level indicators, 11 secondary index, 44 three-level index, shown in the TABLE XI.

**TABLE XI INDEX SYSTEM OF SMART CITY CONSTRUCTION**

| LEVEL INDICATORS | THE SECONDARY INDICATORS | LEVEL 3 INDICATORS |
|---|---|---|
| Urban digital economy | The travel industry | Per capita park green area<br>Information service satisfaction in tourism industry |
| | Commercial construction | Online shopping satisfaction<br>Proportion of online commodity retail<br>The richness of online consumption platform satisfaction,<br>Satisfaction with convenient payment<br>Public resource trading platform |
| Urban infrastructure construction | Transport system | Accuracy of real-time traffic information<br>Smart light pole construction level<br>Rush hour traffic jams<br>Logistics delivery satisfaction level<br>Video resource collection and coverage rate<br>Network sharing of video resources |

| | | |
|---|---|---|
| | Life safeguard | Satisfaction with payment of medical fees on self-service online platform<br>Online reservation and electronic medical record popularity satisfaction<br>Social security service satisfaction in different places<br>Coverage of basic medical insurance<br>Level of community service information<br>Coverage of basic old-age insurance<br>Digital urban management situation |
| | Environmental protection energy | New energy utilization rate<br>Energy efficiency<br>Harmless treatment rate of household garbage<br>Comprehensive utilization and disposal rate of industrial solid waste<br>Ambient air quality good rate<br>Sewage treatment rate<br>Rate of disposal of environmental problems<br>Environmental protection information disclosure rate of enterprises and institutions |
| | Realty service | Property Service Satisfaction |
| | Education popularization | The development level of network education<br>Ease of access to online education resources<br>School multimedia classroom penetration rate<br>School wireless network coverage |
| | Government affair management office | Public satisfaction with one-stop handling<br>E-government Coverage<br>Government services openness<br>Level of government information coordination |
| Urban human ecology | Ecological management | Satisfaction with urban environmental quality early warning<br>Satisfaction with urban environment improvement |

| | | |
|---|---|---|
| | | Afforestation coverage rate of built-up area |
| | Living level | Degree of digitization of life |
| Urban characteristic development | The characteristic industry | Economic benefits of characteristic industries<br>The construction degree of characteristic industry |

## VI PERFORMANCE INDICATORS OF SMART CITY EVALUATION INDEX SYSTEM IN JILIN PROVINCE

The performance indicators of Jilin Province were drawn based on the China Statistical Yearbook (2020) released by the National Bureau of Statistics and the open data on the Internet. The specific calculation method is as follows:

(1) Objective evaluation indicators

Objective indicators include per capita park green floor area, online commodity retail ratio, video resource collection and coverage rate, etc. The paper is limited. Taking school network coverage rate and per capita green space area as examples, the calculation method is as follows by referring to relevant literature and policies:

$$\text{School Network Coverage} = \frac{\text{Network coverage rate of schools in the province}}{\text{Total value of schools in the province}} \times 100\%$$

$$\text{Per capita public green space} = \frac{\text{Urban public green area}}{\text{Urban non} - \text{agricultural population}} \times 100\%$$

(2) Subjective evaluation indicators

Subjective evaluation indexes include online shopping satisfaction, property service satisfaction, etc. For those with strong subjective will, 100-point questionnaire survey is adopted to take average score for analysis. For more difficult to assess the travel industry information service satisfaction, satisfaction with the richness of network consumption platform, convenient pay satisfaction, self-help network platform to pay satisfaction, online booking and electronic medical records popularization satisfaction, long distance to deal with social security service satisfaction, satisfaction, etc., early warning system for the urban environment quality The coverage rate of each service industry in Jilin Province in the yearbook and the form of questionnaire can be evaluated and assigned.

Thus, the index data of smart city construction in Jilin Province can be obtained in TABLE XII:

**TABLE XII INDEX DATA OF SMART CITY CONSTRUCTION IN JILIN PROVINCE**

| LEVEL 3 INDICATORS | DATA | UNIT | THE DATA SOURCE |
|---|---|---|---|
| Per capita park green area<br>Coverage of information services in the tourism industry | 12.54<br>89 | Square meters<br>% | Jilin Provincial people's government portal data<br>Survey of scenic spots in the province on major social network platforms |
| Coverage of online shopping<br>Proportion of online commodity retail<br>Online consumption platform richness satisfaction<br>Convenient payment coverage<br>Public resource trading platform | 98<br>54.62<br>84<br>93<br>45.46 | %<br>%<br>points<br>%<br>% | People's information<br>Hua Jing Industrial Research Institute<br>Jilin daily<br>China Gilin.com data<br>East Asia Business News |
| Accuracy of real-time traffic information<br>Smart light pole construction level<br>Rush hour traffic jams<br>Logistics delivery satisfaction level<br>Video resource collection and coverage rate<br>Network sharing of video resources | 100<br>6<br>31 (63.47)<br>98<br>97<br>96 | %<br>%<br>Km/h (%)<br>%<br>%<br>% | Jilin Province Public Security Department Traffic Management Bureau<br>China Lighting Electrical Appliance Association<br>General Office of Jilin Provincial People's Government<br>Jilin Province Post Industry Consumer Appeal Center<br>Jilin daily<br>China Gilin.com |

*Forest Chemicals Review*
*www.forestchemicalsreview.com*
*ISSN: 1520-0191*
*May-June 2022 Page No. 1946-1975*
*Article History: Received: 24 February 2022, Revised: 05 April 2022, Accepted: 08 May 2022, Publication: 30 June 2022*

| | | | |
|---|---|---|---|
| Self-service online platform to pay the coverage of medical fees<br>Online reservation and electronic medical record popularity satisfaction<br>Social security service satisfaction in different places<br>Coverage of basic medical insurance<br>Level of community service information<br>Coverage of basic old-age insurance<br>Digital urban management situation | 85.85<br>97<br>100<br>98.2<br>100<br>90<br>100 | %<br>%<br>%<br>%<br>%<br>%<br>% | Hua Jing Industrial Research Institute<br>Jilin Provincial people's government portal<br>The questionnaire survey<br>Jilin Provincial people's government portal<br>China Society Daily<br>Jilin Provincial people's government portal<br>Jilin Provincial Administration bureau of Government Service and Digital Construction |
| New energy utilization rate<br>Energy efficiency<br>Harmless treatment rate of household garbage<br>Comprehensive utilization and disposal rate of industrial solid waste<br>Ambient air quality good rate<br>Sewage treatment rate<br>Rate of environmental problem management<br>Environmental protection information disclosure rate of enterprises and institutions | 98.3<br>97.9<br>90.24<br>73<br>94<br>95.19<br>90.1<br>100 | %<br>%<br>%<br>%<br>%<br>%<br>%<br>% | China Energy News<br>International Grid News<br>View research report network<br>China Industry News Network<br>Jilin Provincial Department of Ecology and Environment<br>View research report network<br>China Gilin.com<br>Jilin Provincial Bureau of Statistics |
| Property Service Satisfaction | 62.59 | points | Jilin Daily all media |
| The development level of network education<br>Ease of access to online education resources<br>School multimedia classroom penetration rate | 94.69<br>100<br>100<br>31.19 | %<br>points<br>%<br>% | Education Department of Jilin Province<br>The questionnaire survey<br>Education Department of Jilin Province<br>Education Department of Jilin |

| | | | |
|---|---|---|---|
| School wireless network coverage | | | Province (by 2020) |
| Public satisfaction with one-stop handling<br>E-government Coverage<br>Government services openness<br>Level of government information coordination | 100<br>100<br>100<br>100 | points<br>%<br>points<br>points | The questionnaire survey<br>China Gilin.com<br>The questionnaire survey<br>Jilin Provincial People's government information disclosure |
| Satisfaction with urban environmental quality early warning<br>Satisfaction with urban environment improvement<br>Afforestation coverage rate of built-up area | 92.3<br>100<br>42.5 | %<br>%<br>% | Jilin Provincial Department of Ecology and Environment<br>The questionnaire survey<br>The People's Government of Jilin Province makes public information |
| Degree of digitization of life | 63 | points | The questionnaire survey |
| Economic benefits of characteristic industries<br>The construction degree of characteristic industry | 60<br>good | %<br>- | Jilin Provincial Bureau of Statistics<br>Jilin Provincial Bureau of Statistics |

6.1 Fuzzy comparison matrix and normalization

In order to facilitate weight analysis of index data with different dimensions and different value ranges at the same level, it is necessary to standardize the original data mentioned above. Most of the smart city evaluation indexes in Jilin Province proposed in this paper have been normalized.

Combined with the weight statistics based on FDCD-TFIDF algorithm above, this paper refers to the fuzzy comparison matrix, index weight and consistency test method proposed by Chen Yi and Lin Baocheng in 19[15], compare the indicators at the same level in pairs, and construct the comparison matrix according to the following formula.

$$P=\begin{bmatrix} 1 & u_{12} & \cdots & \cdots & u_{1n} \\ u_{21} & 1 & \cdots & \cdots & u_{2n} \\ \vdots & \vdots & 1 & \iddots & \vdots \\ \vdots & \vdots & \iddots & 1 & \vdots \\ u_{n1} & u_{n2} & \cdots & \cdots & 1 \end{bmatrix}$$

The uijAs an index piThe pjThe relative importance of, is: uij=1/uji.

The maximum eigenroot λ of the comparison matrix is calculatedmaxAnd the corresponding eigenvector W.

PW = lambda.maxW

The corresponding index P can be obtained by normalization of the obtained feature vector WiThe weight values under the k layer are shown in the following formula.

$$\omega_{i,k} = \frac{w_i}{\sum_{i=1}^{n} w_i}$$

wi∈W, I = 1, 2, 3...n

Index piThe global weight of is the multiplication of layer-by-layer weight, as shown below:

$$\omega = \prod_{k=2}^{m} \omega_{i,k}$$

K = 1, 2,...m

Where is the index P$\omega_{i,k}$iThe weight of the index of the k-th layer to the k-1st layer, m is the total number of layers.

Through the above method, a more accurate comparison matrix P can be obtained by referring to the previous weight statistics and combining with the contents of this paperUThe maximum characteristic root λ ofmax=6.95, the normalized weight vector is ωU={0.12 0.05 0.35 0.07 0.19 0.24 0.11}T, the consistency index is CI=0.039, and the average random consistency index RI is 1.28. According to the consistency ratio formula CR=CI/RI, CR is 0.0<0.1, which meets the consistency test conditions.

By repeating the process for each layer of classification index, the relative weight of each layer of classification index to the upper layer can be obtained.

TABLE XIII shows the weight distribution table of smart city construction index system in Jilin Province

## TABLE XIII WEIGHT DISTRIBUTION TABLE OF SMART CITY CONSTRUCTION INDEX SYSTEM IN JILIN PROVINCE

| LEVEL INDICATORS | THE SECONDARY INDICATORS | LEVEL 3 INDICATORS | THE WEIGHT |
|---|---|---|---|
| Urban digital economy (18.79%) | The travel industry | Per capita park green area<br>Information service satisfaction in tourism industry | 2.16%<br>3.31% |
| | Commercial construction | Online shopping satisfaction<br>Proportion of online commodity retail<br>The richness of online consumption platform satisfaction,<br>Satisfaction with | 3.24%<br>4.31%<br>2.19%<br>3.16%<br>0.42% |

| | | | |
|---|---|---|---|
| | | convenient payment<br>Public resource trading platform | |
| Urban infrastructure construction (51.05%) | Transport system | Accuracy of real-time traffic information<br>Smart light pole construction level<br>Rush hour traffic jams<br>Logistics delivery satisfaction level<br>Video resource collection and coverage rate<br>Network sharing of video resources | 1.32%<br>0.16%<br>1.64%<br>0.32%<br>2.06%<br>3.49% |
| | Life safeguard | Satisfaction with payment of medical fees on self-service online platform<br>Online reservation and electronic medical record popularity satisfaction<br>Social security service satisfaction in different places<br>Coverage of basic medical insurance<br>Level of community service information<br>Coverage of basic old-age insurance<br>Digital urban management situation | 2.93%<br>2.51%<br>0.38%<br>4.73%<br>1.07%<br>2.17%<br>1.64% |

*Forest Chemicals Review*
*www.forestchemicalsreview.com*
*ISSN: 1520-0191*
*May-June 2022 Page No. 1946-1975*
*Article History: Received: 24 February 2022, Revised: 05 April 2022, Accepted: 08 May 2022, Publication: 30 June 2022*

| | | | |
|---|---|---|---|
| | Environmental protection energy | New energy utilization rate<br>Energy efficiency<br>Harmless treatment rate of household garbage<br>Comprehensive utilization and disposal rate of industrial solid waste<br>Ambient air quality good rate<br>Sewage treatment rate<br>Rate of disposal of environmental problems<br>Environmental protection information disclosure rate of enterprises and institutions | 3.89%<br>1.34%<br>2.68%<br>1.34%<br>2.88%<br>0.88%<br>0.26%<br>0.64% |
| | Realty service | Property Service Satisfaction | 1.62% |
| | Education popularization | The development level of network education<br>Ease of access to online education resources<br>School multimedia classroom penetration rate<br>School wireless network coverage | 3.31%<br>0.32%<br>1.28%<br>0.14% |
| | Government affair management office | Public satisfaction with one-stop handling<br>E-government Coverage<br>Government services openness<br>Level of government information coordination | 0.41%<br>3.88%<br>0.88%<br>0.88% |
| Urban human | Ecological | Satisfaction with urban | 0.24% |

| | | | |
|---|---|---|---|
| Ecology (12.96%) | management | environmental quality early warning<br>Satisfaction with urban environment improvement<br>Afforestation coverage rate of built-up area | 4.25%<br>6.40% |
| | Living level | Degree of digitization of life | 2.07% |
| Urban characteristic development (17.20%) | The characteristic industry | Economic benefits of characteristic industries<br>The construction degree of characteristic industry | 10.96%<br>6.24% |

From table 11, it is not difficult to see the policy orientation of Jilin Province in the field of smart city in recent years. Jilin Province is an obvious developing smart province, and it is believed that Jilin Province will be built into a first-class smart province in the near future.

6.2 Application of smart city construction level evaluation index system in Jilin Province

According to the GDP ranking of cities in Jilin province released by Jilin Statistics Network in 2021, this paper selects the top four cities, including Changchun, Jilin, Songyuan and Siping, for relevant analysis and research according to the standardized index evaluation system mentioned above.

The original data obtained in this paper are diversified in different fields, with different properties of indicators and different units of measurement for the original data, which cannot be directly evaluated comprehensively. Therefore, it is necessary to conduct dimensionless processing for the obtained original data so as to carry out subsequent evaluation activities. Dimensionless is a method to transform data of different specifications into the same specification through mathematical transformation, so as to eliminate the dimensional influence of each indicator.

At present, the mature and most widely used dimensionless methods can be summed up into three kinds, namely linear dimensionless, folded dimensionless and curved dimensionless. This paper selects linear dimensionless formula for data processing according to the actual needs and follows the principle of simplicity, and converts it into a percentage index score by comparing the original data with the base value. The specific calculation method is as follows:

$$\mathrm{F_{dj}} = \mathrm{f_{dj}}/\mathrm{G_{dj}} \times 100$$

*Forest Chemicals Review*
*www.forestchemicalsreview.com*
*ISSN: 1520-0191*
*May-June 2022 Page No. 1946-1975*
*Article History: Received: 24 February 2022, Revised: 05 April 2022, Accepted: 08 May 2022, Publication: 30 June 2022*

Fdj is the index score; fdj is the original data of indicators; Gdj is the index reference value; For example, per capita park green area, F is the score of per capita park green area in Jilin Province, F is the per capita park green area in Jilin Province, 12m2, G is the national park per capita green area of 14.8m2, Then F is 81.08 points.

According to the selected linear dimensionless formula, the benchmark value of Changchun city is compared with the benchmark value of Jilin Province. The excess ratio is the contribution of Changchun city to the development of smart city in Jilin Province, while the insufficient part is the direction of the city's construction and development. The processing results are shown in the Fig 4.

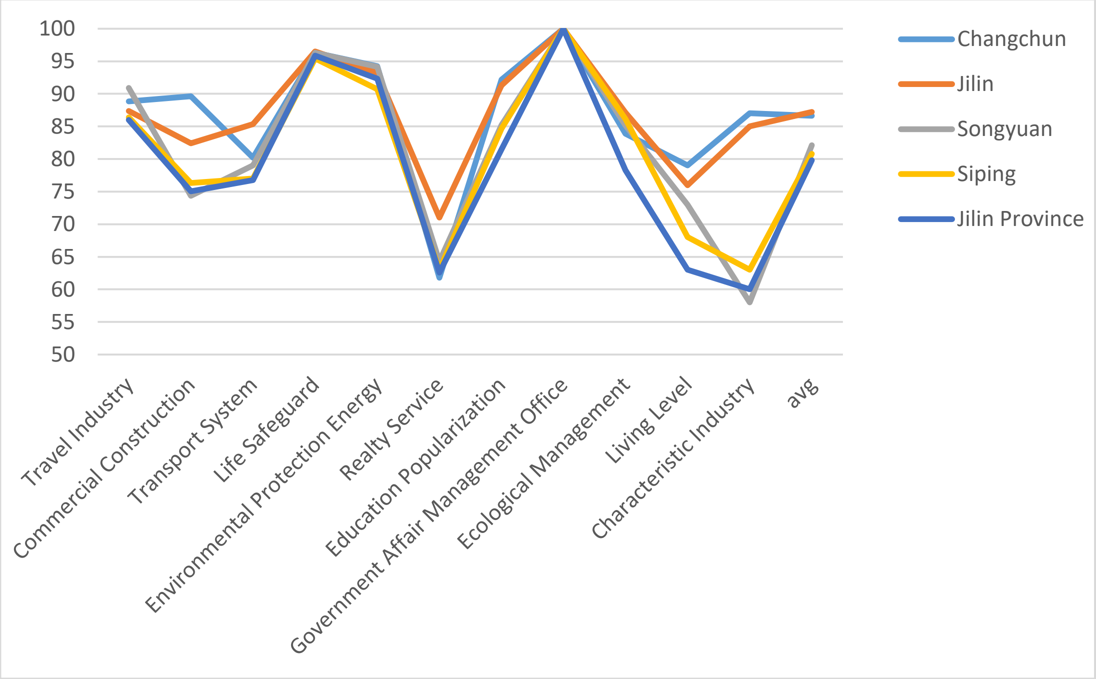

Fig 4 Index score broken line chart of some cities in Jilin Province

In Fig 4, we can clearly understand that changchun in property services, are higher than average, jilin province, at the same time in the city commercial construction of digital economy, urban infrastructure construction of energy and environmental protection education popularization, the urban cultural ecology of ecological living level and the features of the city features the industry are the first four city, But also need to pay attention to the development of urban property services; Jilin City's smart city construction level scores are higher than the provincial benchmark values, and all the data are excellent. Songyuan's commercial construction and characteristic industries are slightly lower than the provincial benchmark value, so it is necessary to promote the construction of commercial public platform and increase the characteristics of commodity categories. Siping city should focus on the utilization rate of new energy, energy utilization rate, harmless treatment rate of domestic waste and sewage treatment rate.

## VII. CONCLUSIONS AND DISCUSSION

In order to solve the problems existing in the jilin province wisdom city construction index system, index system which is different from traditional building method, attempt from the perspective of jilin province government policies, use of the crawler, such as information technology of data mining, and combined with relevant literature review and expert consultation, the wisdom of building a set of scientific and effective evaluation index system of urban construction level. As the government policy documents are used as research data, it has the advantages of strong influence and wide coverage, which makes the evaluation index system constructed more scientific and referential. The experimental results are also in line with expectations, and have been unanimously affirmed by experts.

## ACKNOWLEDGEMENTS

This research was supported by Scientific Research Fund of Jilin Province.

## REFERENCES

[1] Su Jingqin, XU Xinao, Li Xiaang. Research on the policy structure of technological innovation in China based on co-word analysis. Science and technology progress and countermeasures, 2013, 3009:110-115.

[2] Isabel.Freitas, Nick.t. Mapping Public Support for innova-tion: Research Policy, 2008(37): 1446-1464. (in Chinese)

[3] Branco Ponomariov. Government-sponsored University-industry Collaboration and the Production of Nanotechnology Patents in US Universities. Journal of Technology Transfer, 2013, 38(6): 749-767.

[4] Dey A, Jenamani M, Thakkar J J.Lexical TF-IDF: An n-gram feature space for cross-domain classification of sentiment reviews. International Conference on Pattern Recognition and Machine Intelligence. Springer, Cham, 2017:380-386.

[5] Rayner M, Tsourakis N, Gerlach J. Lightweight spoken utterance classification with CFG, tf-idf and dynamic programming. International Conference on Statistical Language and Speech Processing. Springer, Cham, 2017:143-154.

[6] Li Y, Shen B. Research on sentiment analysis of microblogging based on lsa and tf-idf.3rd IEEE International Conference on Computer and Communications,Chengdu,China.2017:2584-2588.

[7] Irfan M, Zulfikar W B. Implementation of Fuzzy C-Means algorithm and TF-IDF on English journal summary. 2nd International Conference on Informatics and Computing, Papua, Indonesia. 2017:1-5.

[8] Hao Jianlin, Huang Zhangjin, Gu Naijie. Automatic music classification method based on user review. Application of computer systems, 2018, 27(01):154-161.64.

[9] Zhao Shenghui, Li Jiyue, Xu Bi, Sun Boyan. Transactions of Beijing institute of technology, 2017, 37(09):982-985.

[10] Zhang Yong 'an, Yan Jin. Internal Structural relations and macro layout of scientific and technological achievements transformation policy based on text mining. Journal of Information science, 2016, 3502:44-49.

[11] Yuan Ye, YU Minmin, TAO Yuxiang, et al. Quantitative Research on China's ARTIFICIAL intelligence Industrial Policy based on text Mining . Journal of China Academy of Electronics Science, 2018, (06):663-668.

[12] Chen K, Zhang Z, Long J,et al. Turning from TF-IDF to TF-IGM for term weighting in text classification. Expert Systems with Applications, 2016, 66: 245-260.

[13] Yuan N. Research and implementation of scenic spot information mining system based on feature weighting and density clustering [D]. Wuhan university of technology, 2019. DOI: 10.27381 /, dc nki. Gwlgu. 2019.001617.

[14] Zhao Ming, DU Huifang, DONG Cuicui, et al. Research on food health Text Classification based on Word2VEC and LSTM . Transactions of the Chinese society for agricultural machinery, 2017, 48(10):202-208.

[15] Chen Yi, Lin Baocheng. Wisdom of radio, film and television city based on analytic hierarchy process (ahp) evaluation index system design. Journal of cable technology, 2019 (10): 23. DOI: 10.16045 / j.carol carroll nki catvtec. 2019.10.009.

[16] Smart City evaluation index System 2.0.