Construction of rural land use knowledge map

Cheng Peng[abc], Cheng Chen[abc], Huaji Zhu[*abc]
[a]Information Technology Research Center, Beijing Academy of Agriculture and Forestry Sciences, Beijing 100097, China; [b]National Engineering Research Center for Information Technology in Agriculture, Beijing 100097, China; [c]Key Laboratory of Digital Village Technology, Ministry of Agriculture and Rural Affairs, Beijing 100097, China

ABSTRACT

The rational use of land in villages and towns is an important prerequisite for the sustainable development of villages and towns and rural revitalization. According to the demand of intelligent knowledge service in the field of land use management in villages and towns, taking the sharing and reusing big data knowledge of land use in villages and towns, the multi-dimensional and hierarchical knowledge system in the field of land use management in villages and towns is built based on multi-source heterogeneous land use data firstly. Then, the related terms and methods are extracted by the rule-based extraction method and the land use entities of villages and towns in semi-structured and unstructured chemistry knowledge texts are extracted by the entity extraction method based on deep learning. The labeling accuracy of extracted basic knowledge entity objects is over 83% and the business knowledge is over 78%. Finally, the basic knowledge, business knowledge and subject knowledge in the field of land use in villages and towns are integrated and correlated. The modeling and expression of knowledge provides intelligent knowledge services for land use and management in villages and towns.

Keywords: Land use in villages and towns, land use entity, knowledge map, bidirectional long-term and short-term memory neural network, conditional random field

1. INTRODUCTION

With the continuous enrichment of technical means of earth observation and land survey, a large number of laws, regulations, standards, technical regulations, survey reports and other data have been accumulated in the field of land use and management in villages and towns. The data has high utilization value, but the lack of correlation and organization among data makes it difficult to use relevant knowledge efficiently.

Knowledge map is a semantic network in essence, which integrates scattered and fragmented information and displays it visually, which can provide the ability to mine the huge relationship among subjects and has been widely used in various fields such as finance, industry, transportation[1,2].

At present, in the field of land resources utilization and management, the research on knowledge map is still in its infancy. Some researchers try to build geographical knowledge map by combining the knowledge of geoscience, geography, soil science and other disciplines. Zhou proposed an adaptive expression model and construction framework of geoscience knowledge map integrating geoscience elements such as map, text and number[3]. Liu put forward the technical process of knowledge map construction based on spatial relationship, focusing on the extraction and representation of spatial relationship and the fusion of multi-source geospatial data[4]. Other scholars use multi-source and heterogeneous data to construct knowledge maps around soil pollution, natural disasters, natural resources and other fields, which formed a large-scale knowledge map[5-7].

In general, the existing research is not combined with the business field of rural land use, and lacks the analysis of the relationship between multiple elements of data such as survey, evaluation, planning and construction of rural land use, which makes it difficult to apply specific knowledge. Facing the demand of intelligent knowledge service in the field of land use management in villages and towns, this paper analyzes and integrates the basic, business and subject knowledge data of village and town land use, and puts forward the method of identifying named entities based on deep learning. The knowledge map of rural land use is formed and provides intelligent knowledge support for scientific and efficient village and town land management.

* zhuhj007@nercita.org.cn

Third International Conference on Computer Science and Communication Technology (ICCSCT 2022)
edited by Yingfa Lu, Changbo Cheng, Proc. of SPIE Vol. 12506, 125064G
 · 0277-786X · doi: 10.1117/12.2662736

2. DESCRIPTION OF LAND USE IN VILLAGES AND TOWNS KNOWLEDGE SYSTEM

Villages and towns are areas designated for the development needs of villages and market towns, and their utilization includes planning, construction and transformation of land to meet the needs of people's life and development[8,9].

According to the knowledge demand in the field of land use and management in villages and towns, the pattern layer of knowledge map of rural land use is designed. The comprehensive knowledge system of land use in villages and towns includes basic knowledge, business knowledge and subject knowledge. Basic knowledge includes natural objects, social objects and subjects involved in land use in villages and towns; Business knowledge is divided into land status survey, land use planning, land development and remediation, etc. Subject knowledge includes concepts, principles, laws, methods, technical means, etc. and usually takes standards, documents and books, technical regulations as carriers. Figure 1 shows the knowledge description system.

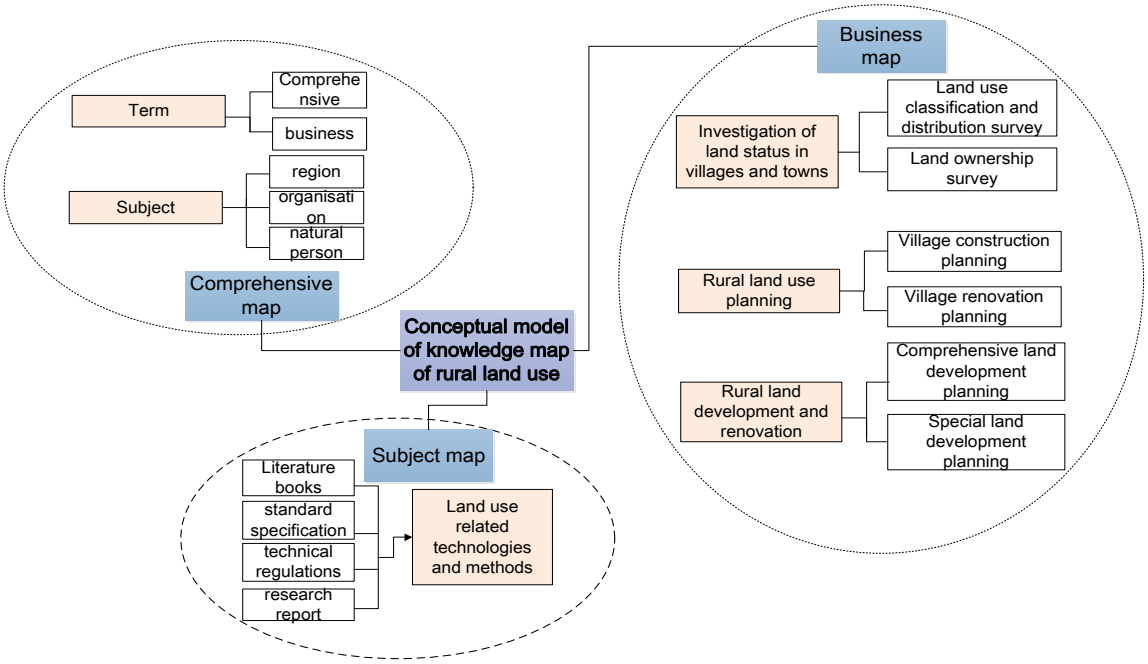


Figure 1. Knowledge description system of land use in villages and towns.

3. CONSTRUCTION OF LAND USE KNOWLEDGE MAP IN VILLAGES AND TOWNS

3.1 Construction framework of land knowledge map in villages and towns

This paper adopts the combination of top-down and bottom-up methods. Firstly, based on the data with high quality and easy structure, such as standards, technical regulations and professional documents of land use in villages and towns, the entity extraction is carried out, and the logical framework and proprietary dictionary of knowledge map are constructed. Then, from unstructured data such as research reports, planning texts and technical data, through natural language processing such as word segmentation, annotation and feature extraction, relevant entity information is extracted, and the entities are processed and stored in the knowledge base. At the same time, entity relationships are summarized, linked and the logical framework is updated.

3.2 Data sources and processing

The relevant data of land use and management in villages and towns mainly come from laws and regulations, standards, regulations, technical regulations, documents, land use planning texts, land use assessment/evaluation reports, land use survey databases, field industry websites, Baidu Encyclopedia and project results, etc. According to the issuing organization and scope of action, the relevant documents are divided into 7 categories: laws and regulations,

departmental rules, national standards, industry standards, local standards, local regulations and local documents, as shown in Table 1.

Table 1. Classification of relevant documents of land use in villages and towns.

Serial number	File type	Examples of partial files
1	Laws and regulations	Land Administration Law of the People's Republic of China (Order No. 41 of the President of the People's Republic of China)
2	Departmental regulations	Notice of the Ministry of Natural Resources and the Ministry of Agriculture and Rural Affairs on Strengthening and Improving the Protection of Permanent Basic Farmland (Natural Resources Regulation [2019] No.1)
3	National standards	Classification of Land Use Status (GB/T 21010-2017)
4	Industry standard	Technical Regulations for the Third National Land Survey (TD/T 1055-2019)
5	Local standards	Technical Regulations on Urban and Rural Planning Management in Hunan Province
6	Local laws and regulations	Regulations of Hunan Province on Overall Land Use Planning
7	Local documents	Master Plan of Land Use in Xiangyin County (2006-2020)

3.3 Extraction of land use knowledge in villages and towns

3.3.1 Rule-Based Entity Extraction Method. First of all, the extraction rules are formulated and the information is extracted into entities by the text matching. Generally, it is applicable to unstructured data with clear description characteristics, such as the terms in the standard specifications and technical regulations of land use in villages and towns[10,11]. Generally, the terms to be explained are uniformly explained at the beginning of the standard specification, or there is a separate chapter to define the terms to explain the meaning of specific words used in the standard specification[11]. Terms are expressed as "< term > [,] [is] means < explanation >" or "< item number > [item number alias] < term > [English] [English alias] < explanation >".For example, in the Classification of Land Use Status (GB/T 21010-2017), "2.3 land utilization; land use the process that human beings use the attributes of land to meet their own needs through certain activities". According to the extraction rules constructed by term features, we can extract: item number "2.3", term "land use", term alias "land use", English name "land utilization", English alias "land use", explaining "the process that human beings use land attributes to meet their own needs through certain activities"[12].

3.3.2 Entity Extraction Method Based on Deep Learning. In view of the unstructured data such as technical regulations, planning texts, survey reports, Bi-directional Long Shot-Term Memory Neural Network (BiLSTM) and Conditional Random Fields (CRF) model are used to extract rural land use entities[13,14]. The specific process is shown in Figure 2. Due to the lack of annotated corpus resources in the field of rural land use, the above-mentioned terms extracted by rules are used to sort out relevant entries and carry out manual text annotation. As a result, an annotation set and a special dictionary for rural land use are constructed. The BiLSTM-CRF model inputs word vectors and outputs the predicted sequence annotations of each word.

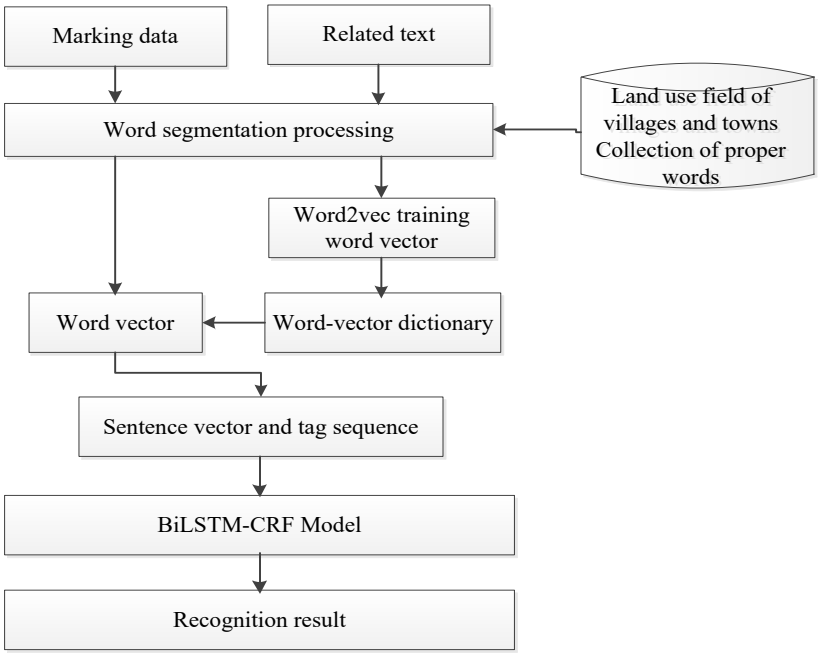


Figure 2. Extraction process of land use entities in villages and towns.1

3.3.3 Relation Extraction. In the rule-based term extraction, some clauses in the standard specification contain terms related to terms, and there are proper names with inclusion relationship and equivalent relationship. For example, “land consolidation is the general name of land development, land consolidation, land reclamation and land restoration”. The relationship among land consolidation, land consolidation, land reclamation and land restoration can be extracted. For unstructured text data, the pattern matching method is adopted. According to the characteristics of rural land use data, the language patterns with high frequency are summarized, and then the entity text is matched with the patterns.

4. EXAMPLE AND APPLICATION OF KNOWLEDGE MAP CONSTRUCTION

4.1 Experimental data set

Using the above methods, the relevant data of land use in villages and towns are extracted. The unstructured data such as the technical regulations, planning texts, investigation reports, documents and materials of land use in villages and towns are labeled manually at first. Then 80% of label text is used as training set and 20% as verification set. As a result, 2218 entities and 3716 relationships are obtained. The entity and relationship data are stored in Neo4j database. F1 is used to evaluate the effect of entity recognition. The calculation method is shown in equations (1)-(3), where TP is the correct positive example, FP is the wrong positive example, and FN is the wrong negative example. The labeling accuracy is shown in Table 2.

$$Recall = \frac{TP}{TP + FN} \times 100\% \tag{1}$$

$$Precision = \frac{TP}{TP + FP} \times 100\% \tag{2}$$

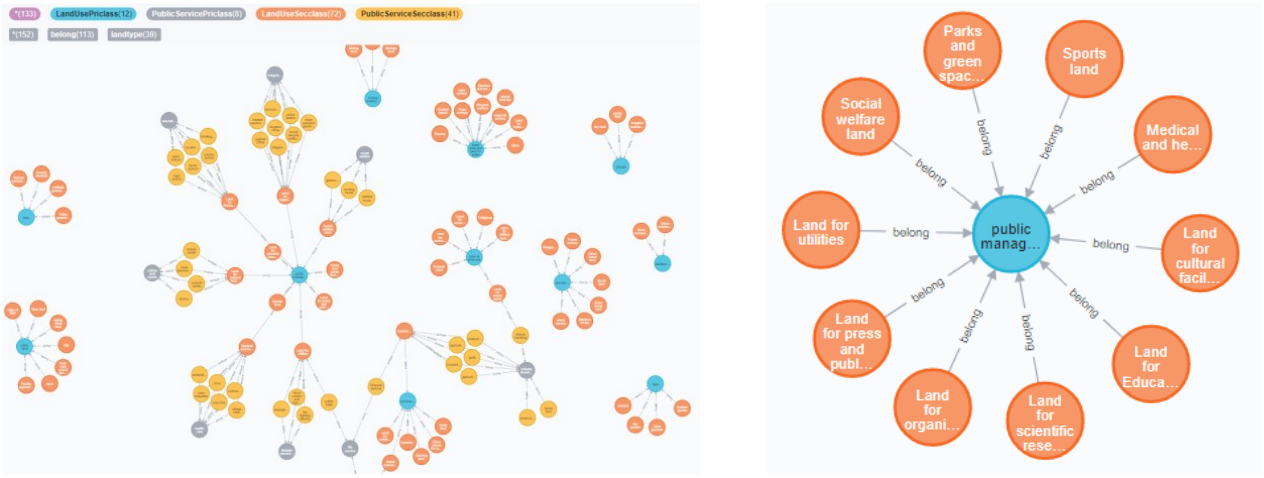
$$F1 = \frac{2Precision * Recall}{Precision + Recall} \times 100\% \tag{3}$$

Table 2. Labeling accuracy of various entity types.

Entity type	F1 (%)	Entity type	F1 (%)
Land use planning-institutions	81.6	Land use planning-method	80.6
Land status survey-organization	82.3	Investigation of land status-method	82.1

4.2 Knowledge map generation

Neo4j supports the large amount of knowledge storage and visual display. So the Neo4j is used to store the village land use knowledge map database, thus creating a visual semantic network. Figure 3a shows part of the knowledge map and the result of querying the relationship between public facilities in villages and towns and the types of public management land is shown in Figure 3b.



(a) Visual display of partial knowledge map (b) Result of querying land types

Figure 3. Example of land use knowledge map in villages and towns.23

5. CONCLUSION

Aiming at sharing and reusing big data knowledge of land use in villages and towns, this paper adopts the mixed construction method of top-down and bottom-up. Focusing on the core businesses such as land status investigation, land use planning and land development and remediation in villages and towns, the conceptual framework of building comprehensive knowledge map of land use in villages and towns is designed and the description method of comprehensive knowledge systems is put forward. The key technologies such as entity recognition and relationship extraction of rural land use from unstructured texts are explored using BiLSTM as feature extractor and sequence labeling with CRF. By annotating the corpus set, the entities such as basic knowledge, business knowledge and subject knowledge in the field of rural land use are identified by naming entity categories and extracted from relationships. Finally the knowledge map of rural land use is formed, which can provide technical support for the knowledge organization, management and application of rural land use.

In the future work, data sources can be appropriately broadened, such as land use related professional websites, forums, blogs, etc., so as to expand knowledge association and be more conducive to the discovery of hidden knowledge. It can further improve the knowledge extraction algorithm, improve the accuracy of entity recognition and relationship extraction, and provide comprehensive and efficient knowledge support for rational land use and intelligent governance in villages and towns.

ACKNOWLEDGMENTS

This work is supported by National Key R&D Program (2020YFD1100204).

REFERENCES

[1] Zhang, J. X., Zhang, X. S., Wu, C. X., et al., "Survey of knowledge graph construction techniques," Computer Engineering 48(3), 23-37 (2021).
[2] Hang, T. T., Feng, J. and Lu, J. M., "Knowledge graph construction techniques: Taxonomy, survey and future directions," Computer Science 48(2), 175-189 (2021).
[3] Zhou, C., Wang, H., Wang, C., Hou, Z., Zheng, Z., et al., "Prospects for the research on geoscience knowledge graph in the big data era," Science China Earth Sciences 51(7), 1070-1079 (2021).
[4] Liu, J. N., Liu, H. Y., Chen, X. H., et al., "The construction of knowledge graph towards multi-source geospatial data," Journal of Geo-information Science 22(7), 1476-1486 (2020).
[5] Huo, Z., Wu, Y. B., Tian, J. P., et al., "Construction of knowledge graph from multi-source heterogeneous soil pollution data," Journal of Hebei Agricultural University 44(2), 137-141 (2021).
[6] Du, Z. P., Li, Y., Zhang, Y. T., et al., "Knowledge graph construction method on natural disaster emergency," Geomatics 45(9), 1344-1355 (2020).
[7] Meng, L. and Wang, G. F., "Framework for knowledge graph-driven construction of natural resources big data mining model in Guangdong Province," Geomatics & Spatial Ecomatics & Spatial Information Technology 43(6), 91-94 (2020).
[8] He, L., et al., [Rural Land Use], China Agricultural Press, Beijing, chapter 6, 23-27 (2009).
[9] Meng, Q. X., Tian, H. W., Zhang, Y. L., et al., "Coupled research on town land use plan and village-town plan," China Population Resources and Environment 20(5), 67-69 (2010).
[10] Liu, G. S., Wang, H. M., Hu, Y. M., et al., "Infrastructure of standard system for land use engineering in China," Transactions of the Chinese Society of Agricultural Engineering 31(13), 257-264 (2015).
[11] Wang, L., [Research on Intelligent Knowledge Support for Urban Rail Transit Construction Safety Management], China University of Mining and Technology, Doctor's Thesis, 84-85 (2019).
[12] GB/T 21010-2017. Classification of land use status.
[13] Cui, J. X., Tan, Q., Zhang, C. X., et al., "A novel framework of graph Bayesian optimization and its applications to real-world network analysis," Expert System with Applications 170, 1-13 (2021).
[14] Duan, H., Han, K., Zhao, H. L., et al., "Research on water conservancy comprehensive knowledge graph construction," Journal of Hydraulic Engineering 52(8), 948-958 (2021).