



Research on urban 3D geological modeling based on multi-modal data fusion: a case study in Jinan, China

Can Zhuang¹ · Henghua Zhu² · Wei Wang² · Bohan Liu² · Yuhong Ma² · Jing Guo² · Chunhua Liu² · Huaping Zhang² · Fang Liu² · Liangliang Cui³

Received: 17 May 2022 / Accepted: 5 November 2022 / Published online: 26 November 2022
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2022

Abstract

Urban underground 3D geological modeling can accurately express various geological phenomena and provide a decision-making basis for urban planning and geological analysis. The construction of smart cities has put forward new requirements for the automation and intelligence of urban geological 3D modeling. Geological survey reports are important reference data for urban geological 3D modeling. However, a large number of geological maps, geophysical data, and other geographic quantitative data of geological science surveys have been buried in geological survey literature and have not been effectively used. Currently, the development of data mining and information extraction technology provides the possibility to integrate these data into 3D geological modeling. Therefore, this study designed the workflow of 3D geological modeling using a geological survey report. First, after the geological survey report was deconstructed, the geological text information was recognized and extracted using geological dictionary matching and pattern rule matching, and the integration of knowledge was provided in the form of a knowledge graph. Then, the drilling information and table data in the drilling histogram are automatically extracted. Through these methods, the unstructured geological survey report can be transformed into structured data and integrated into the 3D geological modeling process. Finally, the 3D geological modeling of the Bridge Group in Jinan based on the Jinan urban geological survey report was taken as an example to verify the feasibility of the proposed method and demonstrate the potential of text mining and information extraction of geological survey reports for 3D geological modeling, which provides geological data support for the transformation of old and new kinetic energy and the construction of major projects of government departments.

Keywords 3D geological modeling · Geological survey report · Information extraction · Borehole log

Introduction

3D geological modeling technology is an interdisciplinary technology integrating geology, computer science, and other disciplines, and the 3D visual representation of geological bodies can intuitively simulate the shape of the geological structure, geological phenomena, and the spatial distribution

of geophysical and geochemical data (Chen et al. 2020; He et al. 2020). Its application field has expanded from the earliest oil and gas exploration (Li et al. 2016) to engineering geology, hydrogeology, urban geological exploration (Xiong et al. 2018), underground space management, and other fields that play a supporting role in the national economy. In recent years, with the rapid development of urbanization, geological problems have become increasingly prominent. The urban underground 3D geological model can not only accurately describe underground geological information, but also provide a decision-making basis for urban environmental monitoring, underground engineering planning, disaster prevention and mitigation. It has become an important topic of smart city research, and it is also the foundation and supporting technology of many major scientific research projects (Wu et al. 2021). To meet the application requirements and support the evaluation and decision-making of

Communicated by: H. Babaie

✉ Chunhua Liu
chunhua_liu321@126.com

¹ Institute of Architectural Engineering, Shandong University of Technology, Zibo, China

² Shandong Institute of Geological Survey, Jinan, China

³ Jinan Zhongan Digital Technology Co., Ltd, Jinan, China

geological resources and the geological environment, the 3D geological model should not only accurately describe the relevant geological structures, but also fill in the model with various types of material components and characteristic attributes, namely the geological attribute model, which refers to the spatial distribution characteristic model of geological attribute fields including density, resistivity, salinity, porosity and other geophysical and geochemical parameters. Currently, 3D geological modeling mainly adopts a data-driven modeling method (Wu et al. 2021), and the data source mainly comes from high-precision original data obtained from direct observation or exploration, such as drilling data, profile data, geological attribute data, etc. The precision of a 3D geological model depends on the accuracy of the data. Limited by actual engineering projects, the amount of data used for modeling is limited and unevenly distributed. Even the geological model automatically drawn by the interpolation algorithm cannot well support the construction of the complex geological model. Therefore, the integration of geological knowledge and the guidance of expert knowledge are indispensable steps in practical production.

In China, many years of geological survey work have accumulated a large amount of geological data, which is multi-modal (Ma et al. 2018; Wu et al. 2017). In addition to the traditional structured geological spatial database and vector plane geological map, it also has a large amount of unstructured document data (Wu et al. 2015). The geological survey report contains rich and valuable modeling data. For instance, the text information described in natural language can help researchers understand the regional geological structure framework; the borehole log, as a comprehensive geological map, can be used to describe the characteristics of stratum name, stratum thickness, and lithology; and the geophysical data in tables can be used for the construction of geological attribute models. However, the multi-modal data in a large number of geological survey reports have not been effectively used (Zhang and Liu 2019), which makes the service ability of these rich geological data low and the application rate extremely low. The sparsity of 3D geological modeling data is a bottleneck of 3D geological modeling work. The description information in the existing geological reports contains many valuable modeling data, but it cannot be directly used for geological modeling. How to fully mine these data, accurately understand, and effectively construct a 3D model need to be further studied. Fully mining geological survey report information for 3D geological model construction can provide a new data source for geological modeling, while reducing some manual work.

To improve the utilization efficiency of geological survey reports and make the 3D geological model more reasonable, this study planned to take the bridge group as the research area, which is located in the starting area of the

transformation of old and new kinetic energy in Jinan. In addition, we proposed an information processing workflow to extract text, figures and table data from geological survey reports. These will help further construct the 3D geological structure model and attribute model, intuitively reflect the underground space resources in the study area, and provide geological data support for the transformation of old and new kinetic energy and the construction of major projects of government departments.

The remainder of this research is structured as follows: Sect. 2 introduces the related work in advances in 3D geological modeling and previous research of information extraction in the geological context; Sect. 3 provides details on the workflow for the construction of a 3D geological model based on multimodal data; Sect. 4 shows the experimental results of urban 3D geological modeling using the proposed method in this paper; Sect. 5 presents the discussion; and conclusions and suggestions are put forward for future improvement in Sect. 6.

Related work

Advances in the 3D geological model

Since Houilding (1994) put forward the concept of 3D geoscience modeling, developed countries such as Britain and the United States have formulated and implemented plans related to 3D geological surveys. In a scientific strategy report titled "Gateway to the Earth", the British Geological Survey proposed building a nationwide 3D geological model and expanding it to offshore areas by 2023 to meet the growing demand for Earth exploration (Council 2014). The USGS also plans to develop and apply new 3D geological models with spatiotemporal characteristics in its research plan report "Center of Excellence for Geospatial Information Science Research Plan 2013–18" (Userly 2013). In China, with the gradual completion of the construction of the geological cloud platform and the improvement of geological survey informatization, 3D geological modeling has also made some progress in some fields. Shandong Province has carried out the "Perspective of Shandong" project to ascertain the urban geological structure and the status of underground space, and build a high-precision 3D visual geological structure model, which lays the technical foundation for the exploitation of underground space and the siting of construction projects.

Currently, according to the different data sources used in modeling, 3D geological modeling methods can be divided into section based, borehole based, geophysical data based, multi-source data based, etc. The section-based can be used for complex geological body modeling, which is relatively mature. The precision of the 3D geological model is

constrained by the number of sections. The more sections there are, the more accurate the model, and the greater the workload. Hao et al. (2021) constructed a three-dimensional geological model of the Kangyangju area using the plane section modeling method with multiple constraints. Chen et al. (2018) presented a locality-based multiple-point statistics approach to reconstruct 3D geological models using 2D cross sections instead of an entire training image. The borehole-based approach is suitable for modeling layered geological bodies, but it is difficult for complex geological bodies. Jiskani et al. (2018) effectively established a 3D solid seam model to produce the spatial distribution maps for coal seam thickness. The modeling method based on multi-source data comprehensively uses plane geological maps, borehole data, profile data, geophysical or geochemical data, attribute data and so on. These data complement and constrain each other, coordinate and unify together, and improve the accuracy of the model. Zhang et al. proposed a collaborative analysis method using multi-source geological data and an interpolation algorithm for geological body modeling (Zhang and Zhu 2018). In recent years, due to the rapid development of machine learning algorithms, many scholars have begun to explore research on 3D geological modeling based on machine learning (Guo et al. 2021; Jia et al. 2021; ZHOU et al. 2019). Bai and Tahmasebi (2020) proposed a hybrid algorithm that combines the cross-correlation simulation method (CCSIM) and CNN, CCSIM adopts a cross-correlation function to calculate the similarity between the reproduced patterns and the training image to obtain initial predictions. Then, the convolutional neural network (CNN) is used to improve these initial values to increase the accuracy of hard data reproduction. Gonçalves et al. (2017) proposed a machine learning approach to the potential-field for implicit modeling of geological structures. More relevant to this study is the work of Zhang et al. (2020), which utilizes information extraction technology to convert paper borehole log data into structured data that can be directly integrated into the 3D geological modeling process.

Previous information extraction in the geological context

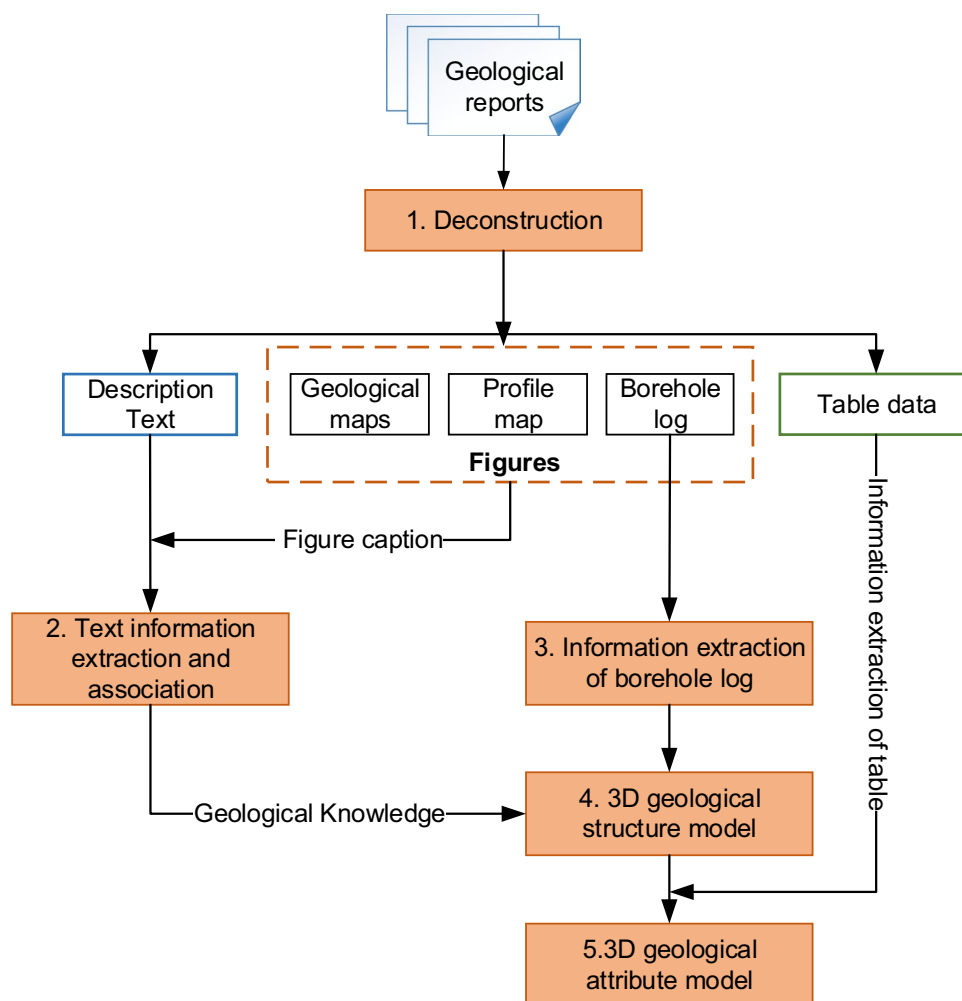
In the field of geosciences, many scholars are exploring the research of geological text mining, which focuses on using NLP (Natural Language Processing), statistical analysis, and data mining techniques to find patterns, structures, or trends in unstructured natural language texts. Information extraction based on original unstructured data is the basis and core of text data mining. Peters et al. (2014) developed a statistical machine learning system called the PaleoDeepDive (PDD), which automatically identifies and extracts information about fossils from the scientific literature. Experiments show that the data quality generated by

PDD can be comparable to that generated by humans, even when only a small amount of training data is available. Wang and Stewart (2015) built a disaster ontology and combined it with a gazetteer to process semantic information of disaster events from online news reports, extract spatiotemporal information, and obtain spatiotemporal patterns of disaster-related events to track the occurrence and evolution of natural disasters. Holden et al. (2019) developed a geological document analysis system called GeoDocA, which uses automatic text analysis technology to help geologists browse and search geological content in large document libraries.

Recently, there have been some explorations of Chinese text mining research in the field of geosciences. Wang et al. (2015) extracted spatiotemporal information from online news reports with the help of a geological disaster event ontology. Wu et al. (2017) divided each document in the geological archives into content fragments through content splitting, extracted thematic and spatiotemporal features from each fragment, and stored these features in a NoSQL database that is more suitable for unstructured data management, making it suitable for content retrieval and data mining. Qiu et al. (2019) developed an unsupervised deep learning model, that can automatically extract and identify geological named entities from complex geological report texts. Shi et al. (2018) proposed a text mining method based on CNN, which can learn and classify four types of geoscience text data of geology, geophysics, geochemistry, and remote sensing, and automatically extract prospecting information. Qiu et al. (2020) introduced the workflow of extracting spatiotemporal and semantic information from geological reports.

To effectively extract information from Chinese texts, many scholars have studied word segmentation methods for Chinese geological texts to segment meaningful words from continuous Chinese texts without separators (Qiu et al. 2018a, 2018b). Huang et al. (2015) developed a segmenter, named GeoSegmenter, which specializes in the field of geoscience and can effectively identify geoscience terms and general terms. In addition, research on the construction of professional domain ontology has also made many achievements (Mantovani et al. 2020; Xu et al. 2014; Zhong et al. 2017). The introduction of domain ontology can help to improve the accuracy of information extraction. Garcia et al. (2020) proposed a general core ontology model for the geological industry. Li et al. (2019) constructed the geological event ontology, which integrates the event information in the geological field to meet the particularity of the geological field. Hou et al. (2018) designed the Chinese Geological Time Scale Ontology regarding the temporal features within datasets, which can better resolve the semantic ambiguity of geological concepts and data.

Fig. 1 3D geological modeling flow based on multi-modal data



Methodology

The main data source of this study is the geological survey report in PDF format, which is a summary report compiled by geological professionals who integrate field investigation results and indoor research data. It has the characteristics of both figures and texts. In addition to the text information described in natural language, it also has a large number of geological maps and attribute tables. These figures or tables are usually explicit or implicit representations of geological knowledge (Qiu et al. 2020). For example, the plane profile can fully reflect the stratum and geological structures at a certain depth underground; the borehole log is the basic data source for constructing a 3D geological structure model; and the table data record extremely valuable geophysical or geochemical analysis or experimental data. Therefore, this study explored the use of multimodal data fusion to construct a 3D geological model by extracting information from geological survey reports. The workflow mainly consists of five parts (Fig. 1): 1) the deconstruction of geological reports; 2) extraction and association of text information;

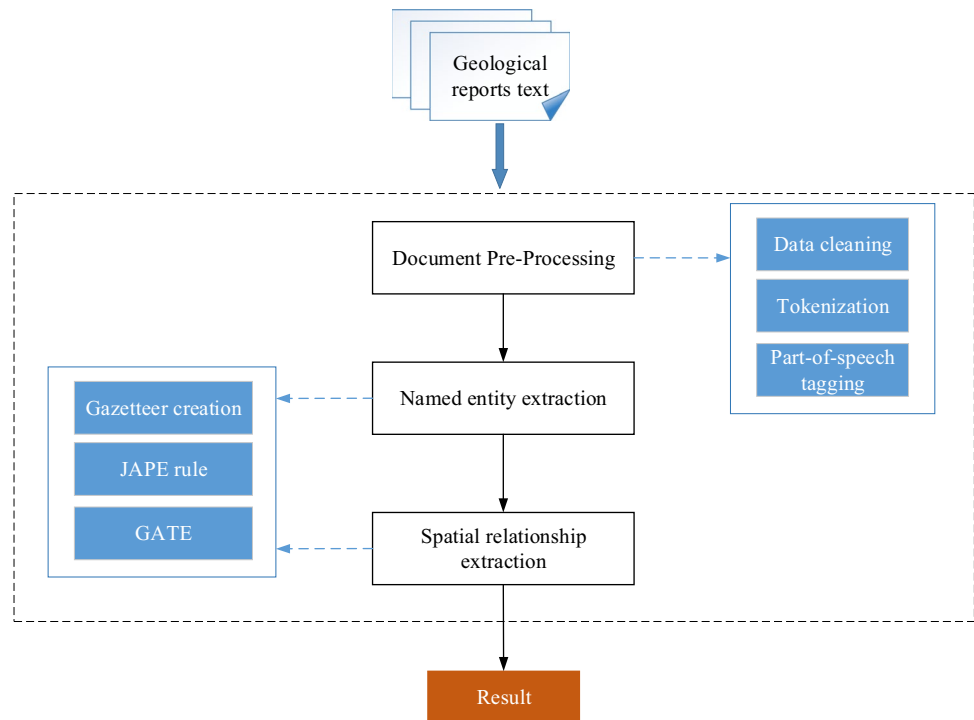
3) information extraction of borehole logs; 4) construction of a 3D geological structure model; 5) construction of 3D geological attribute model. The details of each part are described in the following sections.

The deconstruction of geological reports

To further extract and analyze different modal data, this study first deconstructed the geological survey report. Through document preprocessing, the text, figures, and tables in the report were divided, classified, and marked, and irrelevant information (e.g., invalid characters), references, catalogs, and other contents were removed at the same time. Inspired by Holden et al. (2019), we also adopted the open source software PDF Figures 2.0 to identify and distinguish text information and graphic content from geological reports. For a given geological report, PDF Figures 2.0 will extract the figures in JPG format, and the document hierarchy was stored in JSON file format, which records the location and content of each chapter title, the location, and the caption of figures and tables (Clark and Divvala 2016).

Table 1 Examples of geological named entities

Category of named entity	Example
Geological structure	Cultural bridge fault, Bucun syncline, Luxitai anticline
Geological age	Paleozoic, Cenozoic, Ordovician
Geological stratification	Cambrian-Ordovician strata, East Huangshan formation
Rock	Dolomite, Gabbro, Granite, Quartz sandstone

Fig. 2 Text information extraction workflow based on GATE

Second, the Camelot¹ tool is used to extract high-value tabular data, and the data are exported to various file formats, e.g., JSON format, Excel, HTML file, or SQLite database, which can be used for the construction of geological attribute models. In particular, information extraction for borehole logs will be detailed in Sect. 3.3.

Extraction and association of geological text information

In this study, the purpose of geological text context information extraction was to identify named entities and spatial constraints of geological entities from geological reports. Here, geological named entities refer to concepts or terms in the geological field, including but not limited to place names, geological stratification, geological structure, rock-soil mass, minerals, rocks, etc. (Table 1). This paper uses the General Architecture for Text Engineering (GATE)²

tool to perform named entity recognition tasks, which is a mature, open-source natural language processing framework that integrates many well-known NLP libraries (e.g. Stanford CoreNLP, LingPipe, OpenNLP, UIMA, etc.), is being widely used by corporations, research labs and universities around the world (Maynard et al. 2020; Song et al. 2021; Van Erp et al. 2021). Figure 2 shows the entire workflow of text information extraction, which is mainly composed of three steps: text preprocessing, named entity recognition, and spatial constraint relation extraction. Text preprocessing is the process of data cleaning, word segmentation, and part-of-speech tagging. Named entity recognition is the basis and premise of spatial constraint relation extraction because the output of named entity recognition is the input of spatial constraint relation extraction.

The geological report text has obvious domain characteristics and contains a large number of geological domain concepts or terms. The direct use of the gazetteer extracted from GATE and the rules of Java Annotations Pattern Engine (JAPE) (Cunningham et al. 1999) is not ideal, mainly in: 1) The Chinese gazetteer usually contains only general domain

¹ <https://github.com/atlanhq/camelot>.

² <https://gate.ac.uk/>.

Table 2 Keywords of entity constraint relationship

No.	Relation Terms	Example
1	The attribute information of geometric features	a total length of...km, the thickness of
2	Deterministic spatial position	Latitude and latitude
3	Neighboring or vague topology relationship	Located on ...
4	Orientation relation of spatial position	To the east of
5	Deterministic orientation information	Position 165°
6	Contact relation between geological entities	Angular conformity, Fault contact

terms, and does not involve terms in the field of geology; 2) The JAPE rules written based on the characteristics of English grammar cannot effectively support the recognition and extraction of Chinese entities and constraints. To this end, to use GATE for named entity recognition and constraint relationship extraction, it is necessary to work on the above problems: 1) create a professional gazetteer in the geological domain; 2) according to the characteristics of geological report texts, write two-stage JAPE extraction rules.

Gazetteer creation To effectively identify the proprietary terms in the geological domain, we intend to create a gazetteer with the help of the terms in the geological domain ontology (Zhuang et al. 2021) constructed in previous research. Geological domain ontology is further subdivided into geological thematic ontology, geological temporal ontology, and geological toponym ontology. Specifically, geological toponym ontology is mainly Chinese traditional place names and words related to spatial relations. The geological thematic ontology is constructed according to the geological dictionary released by the Geological Cloud Platform of the China Geological Survey (GeoCloud³). As of April 2022, a total of 122,923 geological terms have been included, covering multiple disciplines such as geology, geophysics, geochemistry, rock and mineralogy, hydraulic environment, and paleontology, etc.

Named entity recognition GATE is based on JAPE rules for named entity recognition, and the result mainly depends on the ability of JAPE matching patterns. JAPE rules are defined in files with.jape, and each.jape file defines the extraction rules of an entity type. Concretely, to use GATE for named entity recognition, we need to perform the following steps (Susanna et al. 2018):

- (1) Analyze the expression patterns of named entities in geological reports;
- (2) Define corresponding JAPE syntax rules for each pattern;
- (3) Load.jape files into GATE as a new JAPE transducer.

GATE also provides a visual interface. Once the named entity is matched, the identified named entity will be highlighted in the editor and assigned a related type. Missing or incorrect annotations can also be manually repaired.

Spatial constraint relation extraction The description of geological report text is usually carried out under certain standards. The words have obvious domain characteristics, and the expression patterns can be easily extracted. Combined with a large number of the actual corpus, many different types of keywords that can reflect the constraint relationship of geological entities can be summarized (Table 2). Similarly, based on this relatively normative information, combined with the annotated geological named entities, we can define JAPE rules and use the JAPE transducer to extract spatial constraints. The spatial constraint relationship expresses the relative positions between spatial objects. Take the sentence "contact between Granite gneiss and Cambrian Zhangxia Formation limestone" as an example, which expresses the contact relationship between geological entities. There may be many kinds of contact relationship predicates, such as "< geological entity > and < geological entity > [contact type]", "< geological entity > [contact type] in < geological entity > ", etc. These relationship predicates will be expressed according to certain rules, which will be used to construct the rule base for geological feature and constraint relationship identification. After pattern matching based on the rule base, the knowledge is extracted according to the predefined predicate relationship "< geological entity > and < geological entity > [contact type]" to obtain the geological entity constraint relationship "fault contact (Granite gneiss, Cambrian zhangxia formation limestone)", to extract and express the geological entity relationship between geological entities. Similarly, temporal relationship extraction adopts similar methods and rules as the above-mentioned spatial relationship extraction.

Information association Geological reports are divided into knowledge units such as figures, named entities, and constraints through information extraction. These knowledge units present the characteristics of diversification and fragmentation. Due to the lack of data associations, it is difficult to exert the potential value of data. In this study, the named

³ <https://geocloud.cgs.gov.cn/>.

Fig. 3 Examples of borehole logs

层位深度 (m)	层位标高 (m)	单层厚度 (m)	岩石岩性描述	岩性照片	岩性描述
1	28.81	1	砂质粉砂岩		砂质粉砂岩 (1018.8.8)：中等厚层状中-细粒砂岩，夹泥质胶结层...
3.02	21.79	2.82	砂质粉砂岩		砂质粉砂岩 (1018.8.8)：中等厚层状中-细粒砂岩，夹泥质胶结层...
3.81	20.81	0.80	砂质粉砂岩		砂质粉砂岩 (1018.8.8)：中等厚层状中-细粒砂岩，夹泥质胶结层...
5	19.61	1.20	砂质粉砂岩		砂质粉砂岩 (1018.8.8)：中等厚层状中-细粒砂岩，夹泥质胶结层...
7.49	17.12	2.49	砂质粉砂岩		砂质粉砂岩 (1018.8.8)：中等厚层状中-细粒砂岩，夹泥质胶结层...
8.42	16.21	0.80	砂质粉砂岩		砂质粉砂岩 (1018.8.8)：中等厚层状中-细粒砂岩，夹泥质胶结层...
10.36	14.28	1.99	砂质粉砂岩		砂质粉砂岩 (1018.8.8)：中等厚层状中-细粒砂岩，夹泥质胶结层...
11.78	12.83	1.47	砂质粉砂岩		砂质粉砂岩 (1018.8.8)：中等厚层状中-细粒砂岩，夹泥质胶结层...
13.7	11.91	0.92	砂质粉砂岩		砂质粉砂岩 (1018.8.8)：中等厚层状中-细粒砂岩，夹泥质胶结层...
15.67	10.94	0.97	砂质粉砂岩		砂质粉砂岩 (1018.8.8)：中等厚层状中-细粒砂岩，夹泥质胶结层...
17.3	7.21	3.05	砂质粉砂岩		砂质粉砂岩 (1018.8.8)：中等厚层状中-细粒砂岩，夹泥质胶结层...
18.33	6.26	1.06	砂质粉砂岩		砂质粉砂岩 (1018.8.8)：中等厚层状中-细粒砂岩，夹泥质胶结层...
18.95	5.66	0.60	砂质粉砂岩		砂质粉砂岩 (1018.8.8)：中等厚层状中-细粒砂岩，夹泥质胶结层...
20.7	2.01	0.76	砂质粉砂岩		砂质粉砂岩 (1018.8.8)：中等厚层状中-细粒砂岩，夹泥质胶结层...
22.77	2.04	0.77	砂质粉砂岩		砂质粉砂岩 (1018.8.8)：中等厚层状中-细粒砂岩，夹泥质胶结层...
24.3	1.44	0.79	砂质粉砂岩		砂质粉砂岩 (1018.8.8)：中等厚层状中-细粒砂岩，夹泥质胶结层...
26.8	1.03	0.2	砂质粉砂岩		砂质粉砂岩 (1018.8.8)：中等厚层状中-细粒砂岩，夹泥质胶结层...
28	-0.39	3.1	砂质粉砂岩		砂质粉砂岩 (1018.8.8)：中等厚层状中-细粒砂岩，夹泥质胶结层...
30.6	6.19	4.8	砂质粉砂岩		砂质粉砂岩 (1018.8.8)：中等厚层状中-细粒砂岩，夹泥质胶结层...
34	-9.39	3.2	砂质粉砂岩		砂质粉砂岩 (1018.8.8)：中等厚层状中-细粒砂岩，夹泥质胶结层...
36.3	-10.89	1.3	砂质粉砂岩		砂质粉砂岩 (1018.8.8)：中等厚层状中-细粒砂岩，夹泥质胶结层...
38.34	-11.72	0.84	砂质粉砂岩		砂质粉砂岩 (1018.8.8)：中等厚层状中-细粒砂岩，夹泥质胶结层...
38.29	-12.06	1.92	砂质粉砂岩		砂质粉砂岩 (1018.8.8)：中等厚层状中-细粒砂岩，夹泥质胶结层...
41.1	-16.49	2.84	砂质粉砂岩		砂质粉砂岩 (1018.8.8)：中等厚层状中-细粒砂岩，夹泥质胶结层...
42.60	-18.05	1.55	砂质粉砂岩		砂质粉砂岩 (1018.8.8)：中等厚层状中-细粒砂岩，夹泥质胶结层...
43	-19.39	1.33	砂质粉砂岩		砂质粉砂岩 (1018.8.8)：中等厚层状中-细粒砂岩，夹泥质胶结层...
45.8	-20.38	1.7	砂质粉砂岩		砂质粉砂岩 (1018.8.8)：中等厚层状中-细粒砂岩，夹泥质胶结层...
48.9	-22.29	1.9	砂质粉砂岩		砂质粉砂岩 (1018.8.8)：中等厚层状中-细粒砂岩，夹泥质胶结层...
50.7	-24.59	3.3	砂质粉砂岩		砂质粉砂岩 (1018.8.8)：中等厚层状中-细粒砂岩，夹泥质胶结层...
57.45	-27.84	7.35	砂质粉砂岩		砂质粉砂岩 (1018.8.8)：中等厚层状中-细粒砂岩，夹泥质胶结层...
63.53	-28.30	0.55	砂质粉砂岩		砂质粉砂岩 (1018.8.8)：中等厚层状中-细粒砂岩，夹泥质胶结层...
64.71	-30.68	2.17	砂质粉砂岩		砂质粉砂岩 (1018.8.8)：中等厚层状中-细粒砂岩，夹泥质胶结层...
67.75	-33.14	1.58	砂质粉砂岩		砂质粉砂岩 (1018.8.8)：中等厚层状中-细粒砂岩，夹泥质胶结层...
69	-34.39	1.6	砂质粉砂岩		砂质粉砂岩 (1018.8.8)：中等厚层状中-细粒砂岩，夹泥质胶结层...
69.5	-35.09	1.2	砂质粉砂岩		砂质粉砂岩 (1018.8.8)：中等厚层状中-细粒砂岩，夹泥质胶结层...
65.15	-40.54	3.05	砂质粉砂岩		砂质粉砂岩 (1018.8.8)：中等厚层状中-细粒砂岩，夹泥质胶结层...
66	-41.39	0.85	砂质粉砂岩		砂质粉砂岩 (1018.8.8)：中等厚层状中-细粒砂岩，夹泥质胶结层...
66.75	-43.64	3.75	砂质粉砂岩		砂质粉砂岩 (1018.8.8)：中等厚层状中-细粒砂岩，夹泥质胶结层...
69.5	-44.89	1.75	砂质粉砂岩		砂质粉砂岩 (1018.8.8)：中等厚层状中-细粒砂岩，夹泥质胶结层...
71	-46.39	1.3	砂质粉砂岩		砂质粉砂岩 (1018.8.8)：中等厚层状中-细粒砂岩，夹泥质胶结层...
72.3	-47.07	1.3	砂质粉砂岩		砂质粉砂岩 (1018.8.8)：中等厚层状中-细粒砂岩，夹泥质胶结层...
75	-50.30	2.7	砂质粉砂岩		砂质粉砂岩 (1018.8.8)：中等厚层状中-细粒砂岩，夹泥质胶结层...
76	-51.09	1.1	砂质粉砂岩		砂质粉砂岩 (1018.8.8)：中等厚层状中-细粒砂岩，夹泥质胶结层...
76.75	-51.39	0.6	砂质粉砂岩		砂质粉砂岩 (1018.8.8)：中等厚层状中-细粒砂岩，夹泥质胶结层...
79.4	-54.79	2.90	砂质粉砂岩		砂质粉砂岩 (1018.8.8)：中等厚层状中-细粒砂岩，夹泥质胶结层...
79.9	-55.79	0.9	砂质粉砂岩		砂质粉砂岩 (1018.8.8)：中等厚层状中-细粒砂岩，夹泥质胶结层...

entities extracted from the figure caption and text are used as the feature marks, and the knowledge association system between different knowledge units is constructed through the spatial constraint relationship and similarity relationship between the feature marks, which is conducive to geological modelers integrating multiple knowledge units into the process of 3D modeling of urban geology. The association model can be seen in previous papers (Zhuang et al. 2021).

Information extraction of borehole logs

The borehole log is one of the most important diagrams in geological reports. It can be used to describe stratigraphic units, stratum thickness, lithological characteristics, geological structure and contact relationships (Zhang et al. 2020), as shown in Fig. 3. As an important reference for the visualization of underground exploration information, the borehole log is the basis of constructing a three-dimensional urban geological model and plays an important role in the analysis and decision-making of various underground projects. In the past, people needed to manually collect data from papyry data to complete the extraction of drilling information. The urban geological 3D model built based on this method has the problems of low efficiency and poor accuracy (Zhang et al. 2020). To this end, we hope to have an approach of automatic and intelligent identification and extraction of

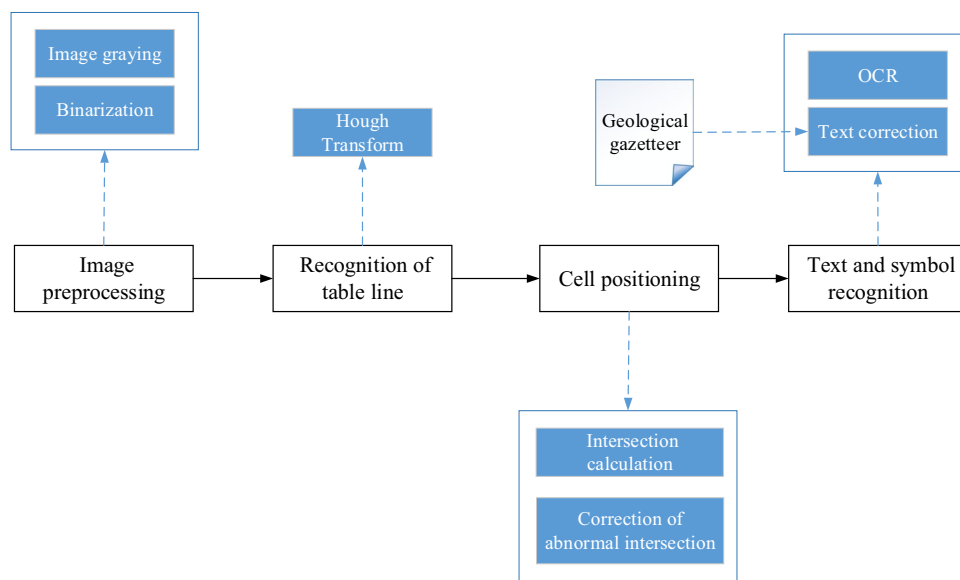
borehole logs, which has precedent to follow. In this study, we employ the approach of Zhang et al. (2020) to identify and extract the borehole log. The entire process is shown in Fig. 4.

Image preprocessing After the borehole logs are separated from the geological reports, gray processing is used to transform the color images into gray images to enhance the image. Then, the appropriate threshold value is adopted for the binary image, and only the useful information (e.g. table lines and text) is retained to reduce the influence of irrelevant information.

Recognition of table line An effective method to quickly identify table lines in the image is the Hough transform (Hassanein et al. 2015) algorithm, which is a feature extraction algorithm in image processing and is widely used to identify various geometric shapes in images. Among them, the Hough lines transform is an algorithm that identifies the straight lines in the image. The Houghlines function can be called in OpenCV⁴ to perform the Hough lines transform, and the interference lines generated by the image can be filtered according to the threshold to obtain the line frame of the table.

⁴ <https://opencv.org/>.

Fig. 4 The flow of information extraction of borehole logs



Cell positioning On the basis of table line recognition, the four marked points of the table cell are further obtained through staggered calculation. However, the table lines of the borehole log are not completely standardized, and there are some abnormal intersections, which need to be adjusted to the normal position.

Text and symbol recognition Text and symbol recognition is a process of using optical character recognition (OCR) to convert words into text format in an image, which will greatly reduce manual entry. Additionally, for some difficult words, the optimal word can be selected as the recognition result by combining with the geological thesaurus, which can improve the accuracy of identification to a certain extent.

According to the corresponding national standards and industrial standards, with the help of an information system and combined with the knowledge and experience of professionals, the borehole data will be standardized through classification, data quality inspection, and stratum standardization.

Construction of the 3D geological structure model

The purpose of constructing a 3D geological model is to reflect the real geological situation as accurately as possible based on respecting the existing geological data. Due to the sparsity or inaccuracy of input data and the error of the modeling process, a 3D geological model inevitably has uncertainty (Hou et al. 2021; Olierook et al. 2021). Consequently, to ensure that the constructed model can more accurately restore the geological structure of the modeling area, this study established geological profiles based on standardized borehole data to assist in stratigraphic analysis, which can

more accurately and intuitively simulate the structure of the underground space. In this process, the 2D section still needs to be corrected to assist the modeling, which mainly includes the following:

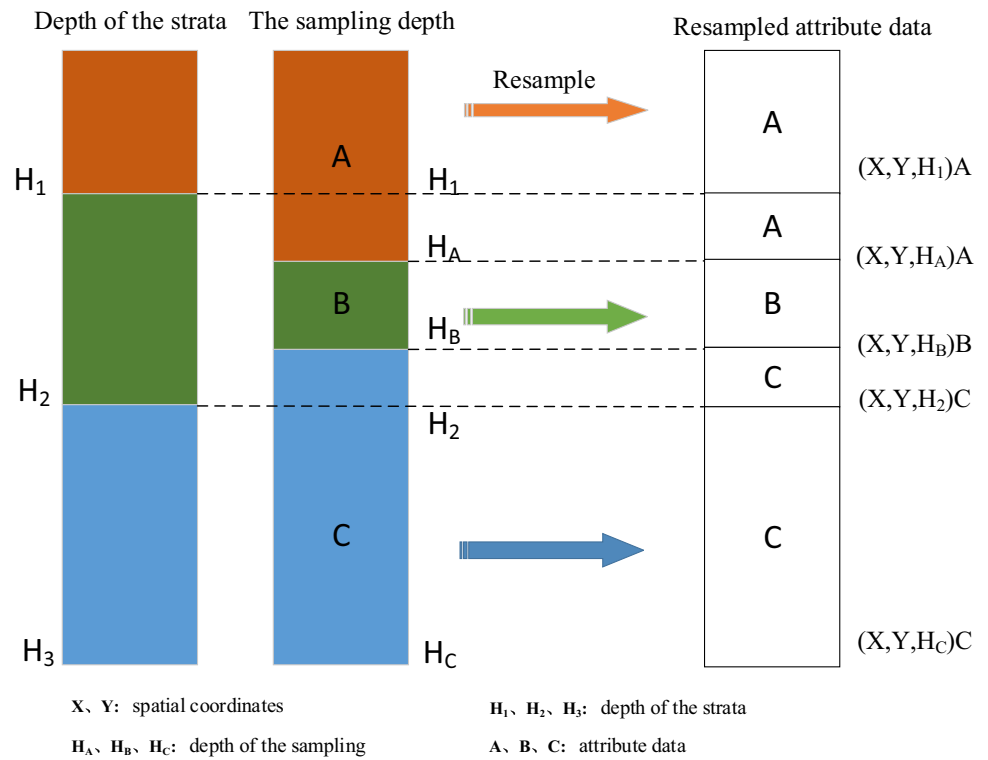
Structurization According to the requirements of 3D geological modeling, the attributes of the profile are structured, and the necessary geological attribute and control point information is added, such as assigning attributes to strata and borehole trajectory lines to form standard 3D section data with colors, textures, and attributes.

Topology reconstruction Following the requirements of the section of the 3D geological model, according to the drilling information on the section, the horizontal and vertical coordinate values of the inflection point and endpoint of each section are obtained, the drilling trajectory line of the section is drawn, and the geological boundary and topological relationship of the geological area on the section are reconstructed.

Consistency check In the case of inconsistent layering or dislocation in the section, the underground position of the stratum can be determined by referring to various data source information such as adjacent boreholes, geological sections, ground floor panels, and plane geological maps (most of the data have been extracted and correlated in Sects. 3.1 and 3.2), and the correct layers can be drawn by taking the terrain line as the benchmark and checking from top to bottom until all cross-sections are consistent with the information of the intersection layer.

3D modeling is established on the basis of a 2D section. This model takes the imported cross-geological section as

Fig. 5 Resampling of attribute data



the main modeling data, which controls the frame structure of the entire modeling area. The 2D section is converted to the 3D section after being imported. If the stratigraphic line cannot be connected, adjustment treatment must be performed on the section. The modeling area was divided into multiple cells through mesh subdivision, and each cell was the smallest unit of modeling. Import the 3D section and geological surface, and transfer the surface boundary line to generate the cell stratigraphic line. The auxiliary line is reasonably drawn, the encapsulation line is used to build the surface, the body is further built, and the encapsulation of the geological body is finally checked. The geological surface in all cells is merged to automatically construct the geological surface model of each stratum in the entire region. According to the attribute and topological relationship of the stratum, all the cell blocks are merged, and the geological model of the whole region is built by combining the 3D surface.

Construction of 3D geological attribute model based on stratigraphic constraints

As mentioned in Sect. 3.1, we employed Camelot to extract the test parameters of engineering and hydrological boreholes in the geological report. Optimize the attribute data mathematically, and eliminate the noise data in light of the geological characteristics of the stratum. The attribute data

are resegmented according to the standardized depth of the strata. If there is a situation where the sampling depth of the geotechnical test intersects with the depth of the strata, the corresponding geotechnical test data will be extracted segmentally based on the depth of the top or bottom of the strata, as shown in Fig. 5.

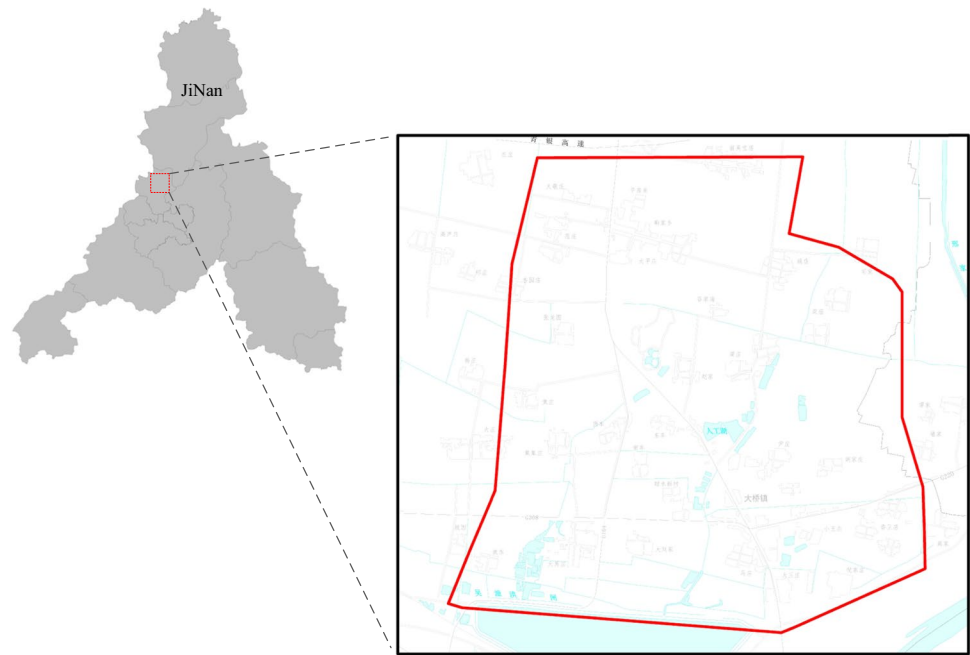
Taking each stratum of the structural model as the constraint, attribute interpolation is carried out layer by layer to build a 3D geological attribute model, which more finely depicts the distribution form of stratigraphic attributes in this area.

Implementation

Taking the bridge cluster in Jinan city as the modeling area, we adopted the urban 3D geological modeling data processing flow based on the geological survey report proposed in this study and employed MapGIS10.2, which is a mature 3D geological modeling software, to establish the final urban geological model.

The processing of the geological survey reports was conducted on a PC with a 2.50 GHz Intel Core i9-11,900 CPU with 16.0 GB of RAM, and to ensure that it works properly, the minimum requirements of the system are 2G memory. The modeling process is conducted on a PC loaded with discrete graphics.

Fig. 6 The study area in Jinan City



Overview of the study area

Jinan is located on the northern edge of the middle mountainous region of Shandong. It is adjacent to Mount Tai in the south and the Yellow River in the north. The terrain is high in the south and low in the north. In this paper, the study area Bridge Group is located in the starting area of new and old kinetic energy conversion in Jinan, with a total area of approximately 51 square kilometers. As shown in Fig. 6, the Bridge Group is located on the North Bank of the Yellow River, northeast of the Tianqiao District, bordering Cuizhai Town, Jiyang County in the east, facing Huashan Town, Licheng District across the Yellow River in the south, adjacent to Sangzidian town in the west and Sungeng Town, Jiyang County in the north. Four national highways 104, 308, 309, and 220, and provincial highway 001 run through the territory, and are the transportation hubs leading to Beijing, Tianjin, and Northern Shandong.

Introduction of used data

This study used the PDF format "Jinan Urban Geological Survey Report" as the research data source. The report includes approximately 388,555 words and more than 500 figures and tables. The modeling area is the Bridge Group in the starting area of Jinan's new and old kinetic energy conversion. The stratigraphic structure of this area is relatively simple. There were 12 boreholes in the study area, and the drilling depth was about 100 m. The bottom of the model was a plane, and the modeling depth was from the ground down to—100 m.

The processing flow of the geological survey report

Taking the "Jinan Urban Geological Survey Report" as the input data, according to the deconstruction method in Sect. 3.1, the geological report was divided into three types of data: text fragments, figures, and tables, as shown in Fig. 7. Geological named entities are indicators that reflect the information and knowledge in the report. In this study, after the extraction of geological text information, we used Neo4j graphics to visualize the diverse relationships between geological named entities, which reflects the basic knowledge of geological survey reports, as shown in Fig. 8. In the figure, the orange circle represents the geological named entity, the blue circle represents the figure and the red circle represents the table.

In addition, there are abundant borehole logs and tabular data in geological survey reports. According to the flow in Sect. 3.3, after graying the picture, the Hough transform is used to identify the table grid, extract the table corners, obtain the divided cells, and identify the contents of the cells. According to the "DD2015-04 Urban Geological Survey Database Structure Specification", the borehole data are standardized, and the attribute table data are extracted by Camelot and stored in the database for the construction of the 3D geological model.

Modeling process

After loading the borehole data from the database, the same stratum of different boreholes was connected to form a

Fig. 7 An illustrative example of a geological report split into figures, tables, and descriptive text

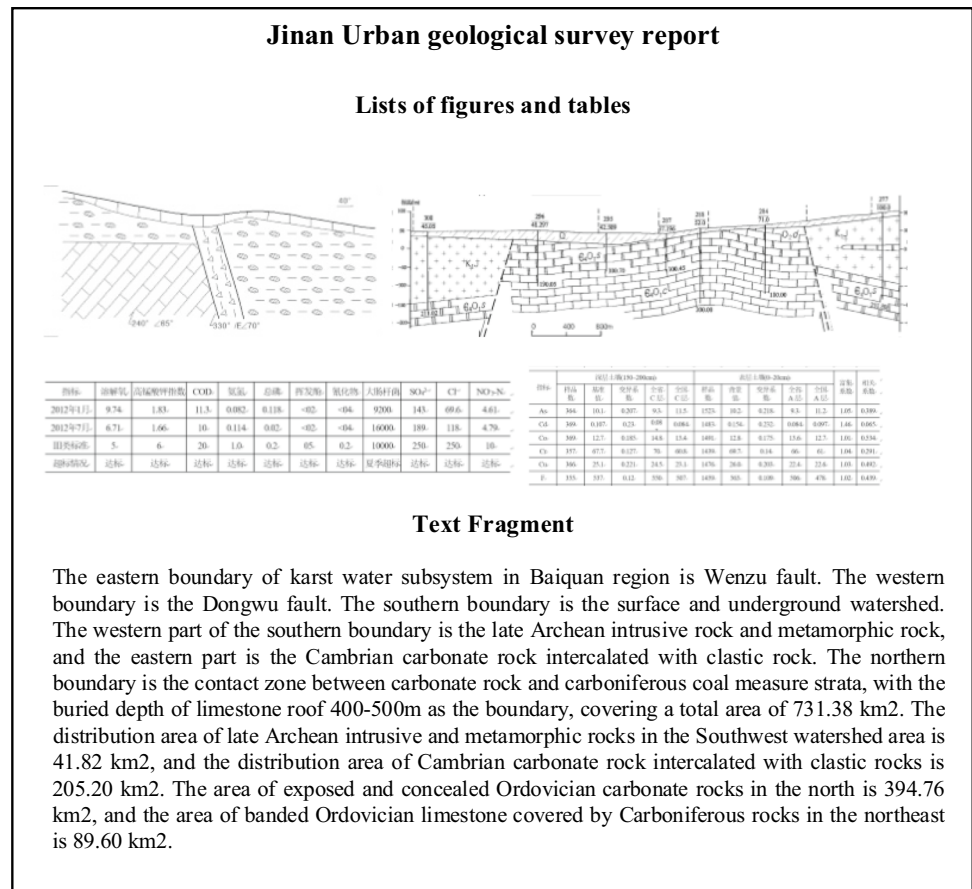


Fig. 8 Entity Association expression using neo4j

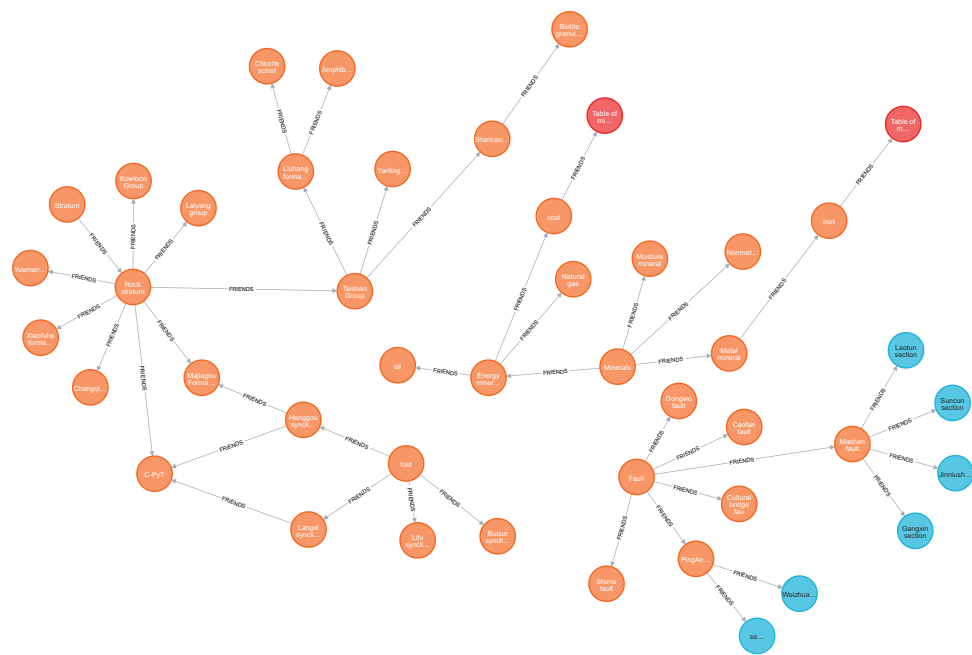


Fig. 9 Example of a partial section

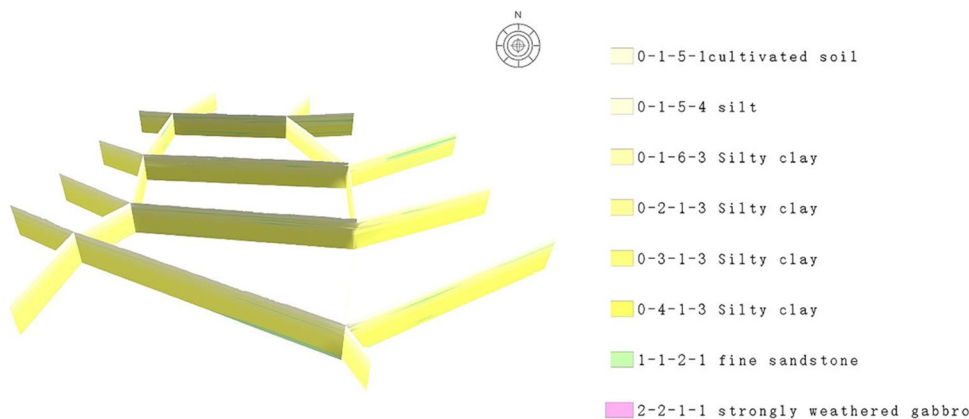
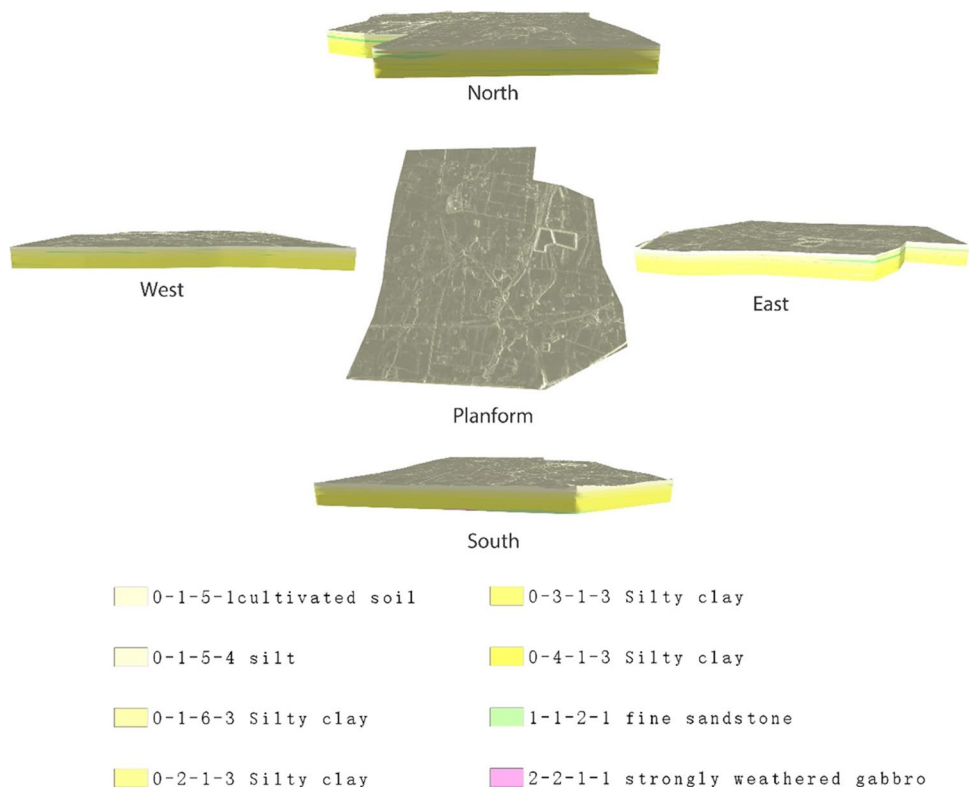


Fig. 10 The final geological structure model of the study area from five perspectives



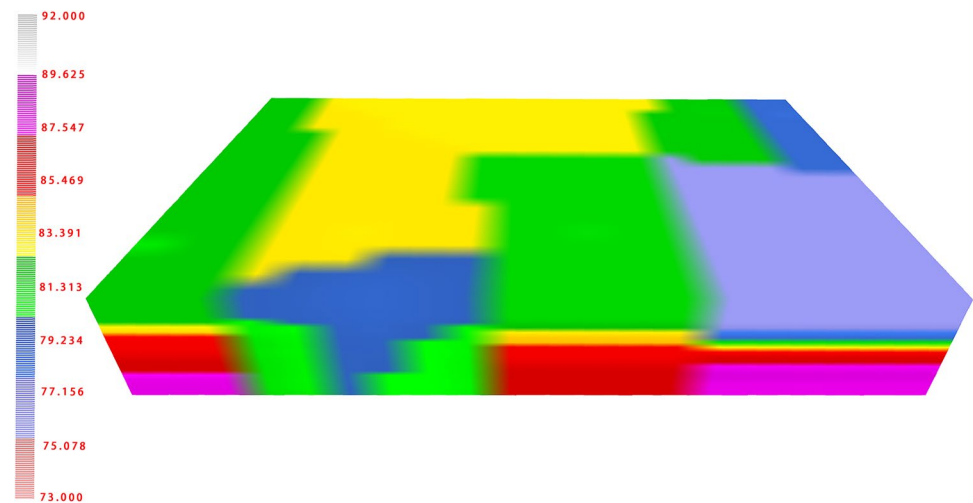
complete regional stratum. According to the knowledge association provided by the geological text, the section was processed or manually modified, and it is easy to obtain the section at a given position, as shown in Fig. 9. The fine geological structure model was constructed by crossing and splicing sections to connect the geological boundary with the corresponding boundary points on the section. The line was used to build the surface, and the surface was used to build the body. Finally, it has been checked and corrected. In addition, we use DEM combined with remote sensing to simulate the surface model, which has an intuitive and realistic show and was closer to the actual situation (Fig. 10). MapGIS software provides a variety of interpolation algorithms. We compared the results of different interpolation algorithms and selected the appropriate algorithm

based on geological cognition. Finally, taking the stratum as the constraint, we used the inverse distance interpolation method to complete the attribute assignment and finally realized the three-dimensional geological attribute model. Figure 11 shows the saturation geological attribute model we built.

Discussion

In this study, we mainly introduced the workflow of building a 3D geological model by extracting information from geological survey reports. In the process of information extraction, the errors of manual processing are reduced, and the efficiency and accuracy are improved.

Fig. 11 The geological attribute model (saturation)



Information extraction of geological text Due to the lack of domain knowledge and large differences in target domains, the recognition ability of specific terms in the geological domain is insufficient, and the feature extraction algorithm in the general domain cannot effectively extract the theme features of the text content. With the help of domain-specific terms provided by the geological domain ontology, we have created a gazetteer to provide a common understanding of geological knowledge and a clear definition of the relationship between geological concepts at different levels, to improve the recognition ability of geological domain specific terms.

Information extraction of borehole logs There are few studies in this area at present. In the past, people needed to manually collect data from geological survey reports to achieve the structure of drilling data. Our work will reduce the workload of manual processing to a certain extent, especially when the amount of data is large, which will be more efficient. Of course, there are still many problems to be solved in this area, such as the recognition and extraction of irregular tables.

Uncertainty In addition, since geological survey reports are written by different people and geologists, each person may have different interpretations of the same matter, which may lead to some differences between different reports and uncertainty of information extraction. In this study, this uncertainty has not been addressed for the time being, but it will be the direction of future work. For text information extraction, we can consider adding synonyms to the gazetteer to solve the uncertainty of named entities. For the information extraction of borehole logs, the existing reports are basically expressed in the form of tables, which is a more standardized form. It mainly solves the identification problem of some

irregular tables. Later, the uncertainty can be further eliminated through standardization.

Conclusion and future work

Using text mining and natural language processing technology, this study designs a new method to construct a three-dimensional geological model through the information extraction of a geological survey report. Geological text information provides knowledge guidance for geologists; Using the automatic extraction and standardization of borehole data, as well as a certain degree of profile editing and adjustment, the three-dimensional geological structure model was constructed; the three-dimensional geological attribute model is constructed using the extraction of geophysical table data. The entire process makes the creation of a 3D geological model more intelligent and automatic, reduces errors caused by manual treatment, and exerts the data value of geological survey reports.

However, there are still many details to be improved upon. In the process of borehole data and tabular data extraction, the data must be relatively standardized. For some irregular data, the extraction effects are poor, and professional guidance is needed. However, the study area in this paper is the Bridge Group in Jinan, which does not contain a large number of faults, landslides, and rock solubilities. The geological structure is relatively simple, so it is necessary to further establish a more complex urban geological model for verification.

The current research work treats the figures as a whole without more fine-grained analysis and exploration. In the future, we will try to objectify and extract more fine-grained information from the plane geological profile through the deep learning algorithm, and establish the connection between the text and the geological profile through the legend.

Acknowledgements We would like to thank the anonymous reviewers for carefully reading this paper and their very useful comments. We thank the Shandong Institute of Geological Survey for providing data support. We thank the Jinan Zhongan Digital Technology Co., Ltd for providing technology support.

Author contributions All authors contributed to the study conception and design. Material preparation, data collection and analysis: Can Zhuang, Henghua Zhu, Wei Wang, Bohan Liu, Yuhong Ma, Jing Guo, and Chunhua Liu; Performed the experiments: Can Zhuang, Henghua Zhu, Wei Wang, Bohan Liu, Yuhong Ma, Jing Guo and Liangliang Cui; Analyzed the data: Huaping Zhang and Fang Liu; Wrote the paper: Can Zhuang, Henghua Zhu, Wei Wang, Bohan Liu, Yuhong Ma, and Jing Guo. All authors reviewed the final manuscript.

Data availability The data in this manuscript have not been published elsewhere.

Declarations

Ethical approval and consent to participate Not applicable.

Consent for publication Written informed consent for publication was obtained from all participants.

Competing interests The authors declare they have no competing interests.

References

- Bai T, Tahmasebi P (2020) Hybrid geological modeling: combining machine learning and multiple-point statistics. *Comput Geosci* 142:104519. <https://doi.org/10.1016/j.cageo.2020.104519>
- Chen Q, Mariethoz G, Liu G, Comunian A, Ma X (2018) Locality-based 3-D multiple-point statistics reconstruction using 2-D geological cross sections. *Hydrol Earth Syst Sci* 22(12):6547–6566. <https://doi.org/10.5194/hess-22-6547-2018>
- Chen Q, Liu G, He Z, Zhang X, Wu C (2020) Current situation and prospect of structure-attribute integrated 3D geological modeling technology for geological big data. *Bull Geol Sci Technol* 39(4):51–58. <https://doi.org/10.19509/j.cnki.dzqk.2020.0407>
- Clark C, Divvala S (2016) PDFFigures 2.0: mining figures from research papers. In: 2016 IEEE/ACM Joint Conference on Digital Libraries (JCDL). pp 143–152
- Council NER (2014) Gateway to the earth: science for the next decade. British Geological Survey
- Cunningham H, Maynard D, Tablan V (1999) Jape: a java annotation patterns engine
- Garcia LF, Abel M, Perrin M, dos Santos AR (2020) The GeoCore ontology: a core ontology for general use in geology. *Comput Geosci* 135:104387
- Gonçalves ÍG, Kumaira S, Guadagnin F (2017) A machine learning approach to the potential-field method for implicit modeling of geological structures. *Comput Geosci* 103:173–182
- Guo J, Li Y, Jessell MW, Giraud J, Liu S (2021) 3D geological structure inversion from Noddy-generated magnetic data using deep learning methods. *Comput Geosci* 149(7):104701
- Hao M, Li M, Zhang J, Liu Y, Huang C, Zhou F (2021) Research on 3D geological modeling method based on multiple constraints. *Earth Sci Inf* 14(1):291–297
- Hassanein AS, Mohammad S, Sameer M, Ragab ME (2015) A survey on Hough transform, theory, techniques and applications. *Computer Science arXiv preprint arXiv:1502.02160*
- He H, He J, Xiao J, Zhou Y, Liu Y, Li C (2020) 3D geological modeling and engineering properties of shallow superficial deposits: a case study in Beijing, China. *Tunn Undergr Space Technol* 100:103390
- Holden E-J, Liu W, Horrocks T, Wang R, Wedge D, Duuring P, Beardsmore T (2019) GeoDocA – fast analysis of geological content in mineral exploration reports: a text mining approach. *Ore Geol Rev* 111:102919. <https://doi.org/10.1016/j.oregeorev.2019.05.005>
- Holding SW (1994) 3D geoscience modeling: computer techniques for geological characterization, vol 46, no 3. Springer Verlag, pp 85–90
- Hou Z, Zhu Y, Gao Y, Song J, Qin C (2018) Geologic time scale ontology and its applications in semantic retrieval. *J Geo-Inf Sci* 20(1):17–27
- Hou W, Yang Q, Chen X, Xiao F, Chen Y (2021) Uncertainty analysis and visualization of geological subsurface and its application in metro station construction. *Front Earth Sci* 15(3):692–704. <https://doi.org/10.1007/s11707-021-0897-6>
- Huang L, Du Y, Chen G (2015) GeoSegmenter: a statistically learned Chinese word segmenter for the geoscience domain. *Comput Geosci* 76:11–17
- Jia R, Lv Y, Wang G, Carranza E, Chen Y, Wei C, Zhang Z (2021) A stacking methodology of machine learning for 3D geological modeling with geological-geophysical datasets, Laochang Sn camp, Gejiu (China). *Comput Geosci* 151:104754
- Jiskani IM, Siddiqui FI, Pathan AG (2018) Integrated 3D geological modeling of Sonda-Jherruck coal field, Pakistan. *J Sustain Min* 17(3):111–119
- Li C, Zhang J, Li H, Liu C (2016) Application of new geological modeling technology in secondary development in Daqing oil field. In: IOP Conference Series: Earth and Environmental Science, vol 1. IOP Publishing, pp 012086
- Li W, Wu L, Xie Z, Tao L, Zou K, Li F, Miao J (2019) Ontology-based question understanding with the constraint of Spatio-temporal geological knowledge. *Earth Sci Inf* 12(4):599–613
- Ma K, Wu L, Tao L, Li W, Xie Z (2018) Matching descriptions to spatial entities using a siamese hierarchical attention network. *IEEE Access* 6:28064–28072
- Mantovani A, Piana F, Lombardo V (2020) Ontology-driven representation of knowledge for geological maps. *Comput Geosci* 139:104446. <https://doi.org/10.1016/j.cageo.2020.104446>
- Maynard D, Lepori B, Petrak J, Song X, Laredo P (2020) Using ontologies to map between research data and policymakers' presumptions: the experience of the KNOWMAK project. *Scientometrics* 125(2):1275–1290
- Olierook H, Scalzo R, Kohn D, Chandra R, Müller R (2021) Bayesian geological and geophysical data fusion for the construction and uncertainty quantification of 3D geological models. *Geosci Front* 12(1):479–493
- Peters SE, Zhang C, Livny M, Ré C (2014) A machine reading system for assembling synthetic paleontological databases. *PLoS ONE* 9(12):e113523
- Qiu Q, Xie Z, Wu L, Li W (2018a) DGeoSegmenter: A dictionary-based Chinese word segmenter for the geoscience domain. *Comput Geosci* 121:1–11
- Qiu Q, Zhong X, Liang W (2018b) A cyclic self-learning Chinese word segmentation for the geoscience domain. *Geomatica* 72(1):16–26
- Qiu Q, Xie Z, Wu L, Tao L (2019) GNER: a generative model for geological named entity recognition without labeled data using deep learning. *Earth Space Sci* 6(6):931–946
- Qiu Q, Xie Z, Wu L, Tao L (2020) Automatic spatiotemporal and semantic information extraction from unstructured geoscience reports using text mining techniques. *Earth Sci Inform* 13(4):1393–1410. <https://doi.org/10.1007/s12145-020-00527-9>

- Shi L, Jianping C, Jie X (2018) Prospecting information extraction by text mining based on convolutional neural networks — a case study of the Lala Copper Deposit, China. *IEEE Access* 6:52286–52297
- Song X, Petrak J, Jiang Y, Singh I, Maynard D, Bontcheva K (2021) Classification aware neural topic model for COVID-19 disinformation categorisation. *PLoS ONE* 16(2):e0247086
- Susanna A, Stephan M, Lars B (2018) Extraction of spatio-temporal data about historical events from text documents. *Trans GIS* 22(3):677–696
- Usery EL (2013) Center of excellence for geospatial information science research plan 2013–18 U.S. Geological Survey Open-File Report 2013–1189
- Van Erp M et al (2021) Using natural language processing and artificial intelligence to explore the nutrition and sustainability of recipes and food. *Front Artif Intell* 115
- Wang W, Stewart K (2015) Spatiotemporal and semantic information extraction from Web news reports about natural hazards. *Comput Environ Urban Syst* 50:30–40
- Wu L, Xue L, Li C, Lv X, Chen Z, Guo M, Xie Z (2015) A geospatial information grid framework for geological survey. *PLoS ONE* 10(12):e0145312
- Wu L et al (2017) A knowledge-driven geospatially enabled framework for geological big data. *ISPRS Int J Geo Inf* 6(6):166
- Wu X, Liu G, Weng Z, Tian Y, Zhang Z, Li Y, Chen G (2021) Constructing 3D geological models based on large-scale geological maps. *Open Geosci* 13(1):851–866
- Xiong Z, Guo J, Xia Y, Lu H, Wang M, Shi S (2018) A 3D multi-scale geology modeling method for tunnel engineering risk assessment. *Tunn Undergr Space Technol* 73:71–81
- Xu J, Nyerges TL, Nie G (2014) Modeling and representation for earthquake emergency response knowledge: perspective for working with geo-ontology. *Int J Geogr Inf Sci* 28(1):185–205
- Zhang Q, Zhu H (2018) Collaborative 3D geological modeling analysis based on multi-source data standard. *Eng Geol* 246:233–244
- Zhang Q, Liu X (2019) Big data: new methods and ideas in geological scientific research. *Big Earth Data* 3(1):1–7
- Zhang X, Zhang J, Tian Y, Li Z, Zhang Y, Xu L, Wang S (2020) Urban geological 3D modeling based on papery borehole log. *ISPRS Int J Geo Inf* 9(6):389
- Zhong S, Fang Z, Zhu M, Huang Q (2017) A geo-ontology-based approach to decision-making in emergency management of meteorological disasters. *Nat Hazards* 89(9):531–554
- Zhou C, Zhang G, Du Z, Liu Z (2019) Stratigraphic sequence simulation based on machine learning. *J Eng Geol* 27(4):873–879
- Zhuang C, Li W, Xie Z, Wu L (2021) A multi-granularity knowledge association model of geological text based on hypernetwork. *Earth Sci Inform* 14(1):227–246. <https://doi.org/10.1007/s12145-020-00534-w>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.