

Article

Construction of Conceptual Prospecting Model Based on Geological Big Data: A Case Study in Songtao-Huayuan Area, Hunan Province

Chang Liu ^{1,2}, Jianping Chen ^{1,2,*}, Shi Li ³ and Tao Qin ^{1,2}

- ¹ School of Earth Sciences and Resources, China University of Geosciences (Beijing), Beijing 100083, China; 3001190108@cugb.edu.cn (C.L.); 2101190057@cugb.edu.cn (T.Q.)
- ² Beijing Key Laboratory of Research and Exploration Information of Land Resources, Beijing 100083, China
- ³ School of Information, Beijing Wuzi University, Beijing 101149, China; lishi@bwu.edu.cn
- * Correspondence: 3s@cugb.edu.cn

Abstract: With the era of big data, the prediction and evaluation of geological mineral resources have gradually entered into a new stage from digital prospecting to intelligent prospecting. The theoretical method of big data mining can contribute to deep mineral resource prediction and evaluation. This paper extracts ore-causing and ore-caused anomaly information based on text intelligent mining technology, and constructs a regional conceptual prospecting model based on geological prospecting big data. First, we set up a corpus based on text big data discovery and preprocessing technology. Second, we used CNN multiple scale text classification technology to analyze geological text data from the two main aspects: ore-causing anomalies and ore-caused anomalies. Third, we used a statistical method to analyze the semantic links between content-words, and we constructed chord diagrams and ternary diagrams to visualize the content-words and their links. Finally, we constructed a regional conceptual prospecting model based on the knowledge graphs.



Citation: Liu, C.; Chen, J.; Li, S.; Qin, T. Construction of Conceptual Prospecting Model Based on Geological Big Data: A Case Study in Songtao-Huayuan Area, Hunan Province. *Minerals* **2022**, *12*, 669. <https://doi.org/10.3390/min12060669>

Academic Editors: Xiancheng Mao and Yosoon Choi

Received: 19 March 2022

Accepted: 24 May 2022

Published: 26 May 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: geological big data; prospecting information; text mining; Songtao-Huayuan manganese deposits

1. Introduction

The prospecting model can be classified as a knowledge-driven model, data-driven model, and hybrid-driven model [1–4]. The development of mineral resources prediction technology rooted in metallogenic regularity has always been the primary task of quantitative prospecting prediction. The intelligent mining method of big data such as deep learning is a data-driven model [5]. Its advantage is that it can extract useful features from large amounts of data and form a predictive model. However, there are always some defects, such as only considering the correlation rather than the causality, only focusing on the results rather than the process. How to better integrate expert knowledge with intelligent mining methods is undoubtedly the key to carry out intelligent metallogenic prediction under the background of big data. In the study of quantitative prospecting prediction, geoscience literature is an important carrier of achievements in geoscience research. Extracting key information from it and visualizing it are very important tasks, because academic literature and mineral exploration reports contain much key prospecting information. In the era of big data, the main issue of geological informatization will be mining, organizing, analyzing, and visualizing the key semantic information of massive geological scientific literature intelligently by computer.

The concept of text mining was proposed by Feldman [6], which mainly includes four parts: data set acquisition, text pre-processing, data mining, and visualization. It is an inevitable trend to apply the technical methods of text mining to the mining of geological text information for finding prospecting information. This technology can not only help us

obtain key prospecting information from the multi-dimensional massive heterogeneous geological text big data, but also provide directions for the automatic construction of prospecting models. Wang et al., proposed a CRF-based word segmenter and analysis method for the problem of word segmentation in geological dictionaries [7]. In 2006, the concept of deep learning was put forward, providing a new direction for text data mining [8]. Convolutional Neural Network (CNNs), as one of the most robust deep learning classification algorithms [9], has been successfully applied to natural language processing [10]. Bengio et al., used CNNs to construct language models and measure the similarity between words by their vector distance [11]. As Yoon Kim et al., adopted CNNs for sentence classification [12], Liang Jun et al., discussed the feasibility of CNNs in Chinese Weibo sentiment analysis [13]. Sun Songtao et al., also successfully used the CNN model for supervised multi-emotion classification learning, and completed the multi-label sentiment classification of Weibo [14]. Subsequently, the RM-CNN algorithm proposed by Feng S successfully performed multi-label sentiment detection [15].

A knowledge graph is a form or product of data organization that expresses entities, concepts, and their semantic relationships by means of directed graphs, providing new ideas for the visual expression of natural language processing [16,17]. It is essentially a semantic network [18], and it has been widely used in the field of geosciences [19,20]. Wang et al., successfully applied the knowledge graph to the key information representation of geological unstructured text, demonstrating the application potential of natural language processing and knowledge graph technology in geoscience research [7]. In 2018, Y.L. Wu proposed a prospecting model construction method based on the key technology of geological text big data discovery and mining for the demand of mineral prospecting prediction, and verified the feasibility of the technical method with a typical example [21]. A workflow is proposed by Li Shi et al., to extract key prospecting information from geoscience text data by text mining based on CNN classification, which realizes the intelligent extraction of geological prospecting information [22].

2. Workflow

With the increasing difficulty of prospecting, breakthroughs in mineralization theories and exploration techniques have become the most important factors of scientific and technological innovation. The accurate and comprehensive construction of the prospecting model is of great significance to quantitative prospecting prediction and evaluation. Based on previous studies, we propose a conceptual prospecting model construction technology based on big data. It mainly includes two parts: (1) prospecting information extraction based on CNN and (2) conceptual prospecting model construction based on machine learning. The workflow is shown in Figure 1.

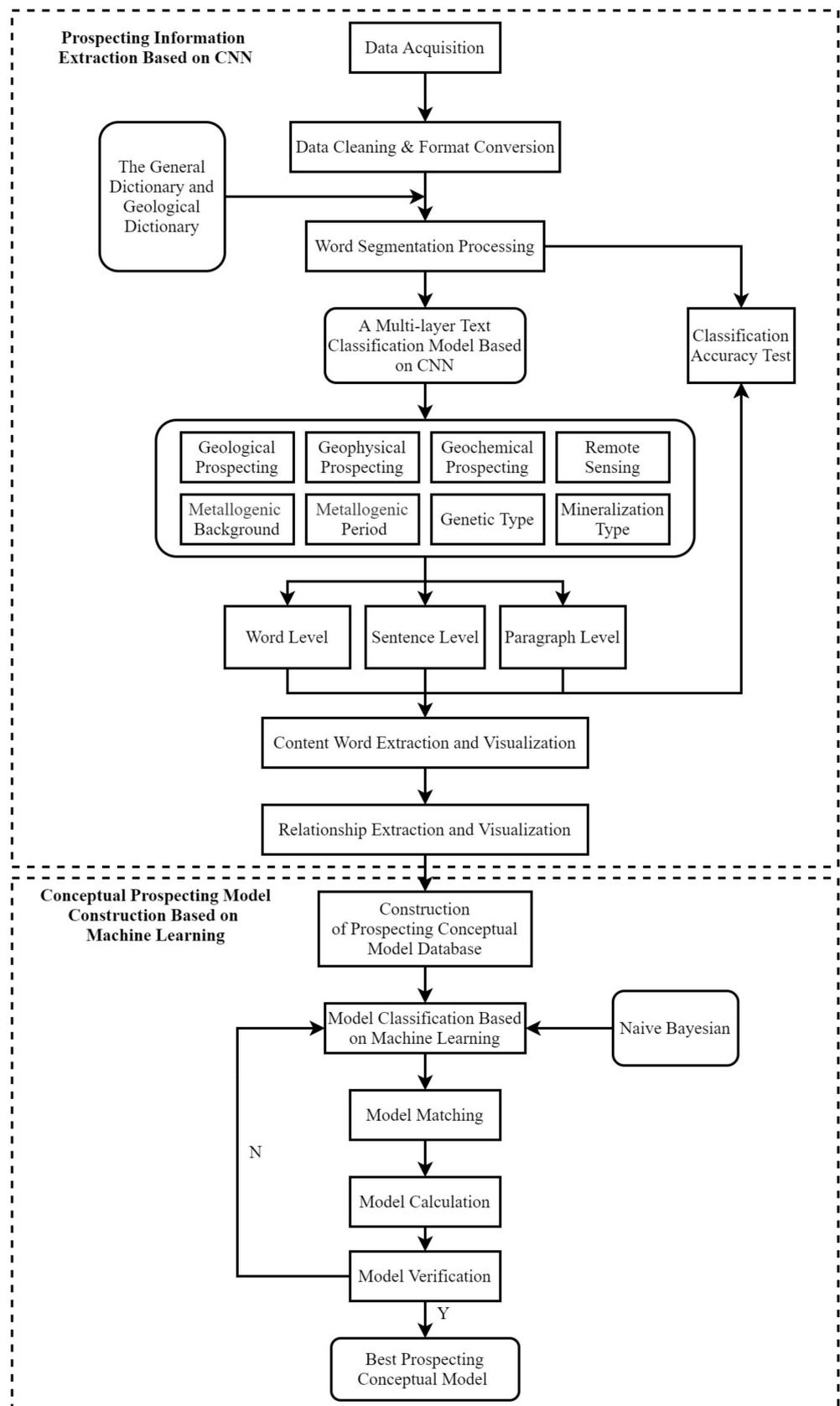


Figure 1. The workflow of conceptual prospecting model construction technology based on big data.

3. Methods

3.1. Prospecting Information Extraction Based on CNN

The prospecting information extraction technique used in this paper refers to the method proposed by Li Shi et al. in 2018 [22]. On the basis of this method, the classification of key prospecting information is refined. In this paper, key prospecting information will be extracted from both ore-causing geological anomalies and ore-caused geological anomalies to lay the foundation for the next step of semi-automatic construction of prospecting model.

3.1.1. Data Acquisition and Pre-Processing

We obtained the corpus from both local area network (LAN) and wide area network (WAN). The Everything software was redeveloped based on C# platform and combined with MySQL relational database approach to search and filter the LAN data. Then, regional geological reports that cannot be accessed by public networks were obtained. For WAN data, a dual iterative approach based on keywords and URLs was proposed. Based on the expert knowledge system, especially the knowledge of quantitative prediction of mineral resources, a corresponding logical structure tree was established. The initial URL seed site was generated by searching the keywords in the knowledge system using a search engine, and its data content was analyzed and extracted to generate new keywords and added to the structure tree. Through machine learning of URL link seeds and structure trees, new URLs and keywords were continuously discovered to form the URL structure trees. The two branches were mutually iterative to form comprehensive search encirclement in two directions.

The data pre-processing mainly included data cleaning, format conversion, and word segmentation processing. Data cleaning means de-tagging the text data acquired by big data discovery, removing wrong or duplicate URLs, as well as preliminary screening of local literature. Format conversion refers to the batch conversion of file formats of the geological literature to be processed. Files were divided into three levels: related news, related literature, and regional reports for weighting operations, and finally formed a mixed corpus. Word segmentation refers to the process of segmentation firstly according to the geological dictionary and then the general dictionary in the jieba dictionary before performing statistics.

3.1.2. Text Classification Based on CNN

In this paper, a CNN algorithm based on the open source TensorFlow architecture was chosen to train the word vectors of the geological content-words by using massive geoscience text sample sets. After constructing a word vector lookup table W^e , supervised classification learning of CNN models was performed in combination with the training set. The geological text data of the study area was used as the test set for the final discriminative output. When the input test set was at word, sentence, and paragraph levels, the output of the model was the classification results of the text of different levels. The CNN text classification model (Figure 2) trained by a large number of geological samples in this study refers to the algorithm proposed by Yoon Kim in 2014 [7].

Sample sets of geological text of sentence and paragraph levels are used as initial inputs. Feature information is transferred to a multi-layer model for computation. In each layer, a different convolution is used to extract the important features of the text. The deeper the level becomes, the more abstract the extracted features. Each feature extraction layer is connected to a pooling layer used to find the local average or maximum value. This feature extraction structure enables the model to have good sample transformation error tolerance and feature recognition capabilities. The training process of the classification model based on the paragraph sample set can be regarded as the model training process for multiple long sentences, which is roughly the same as the process above.

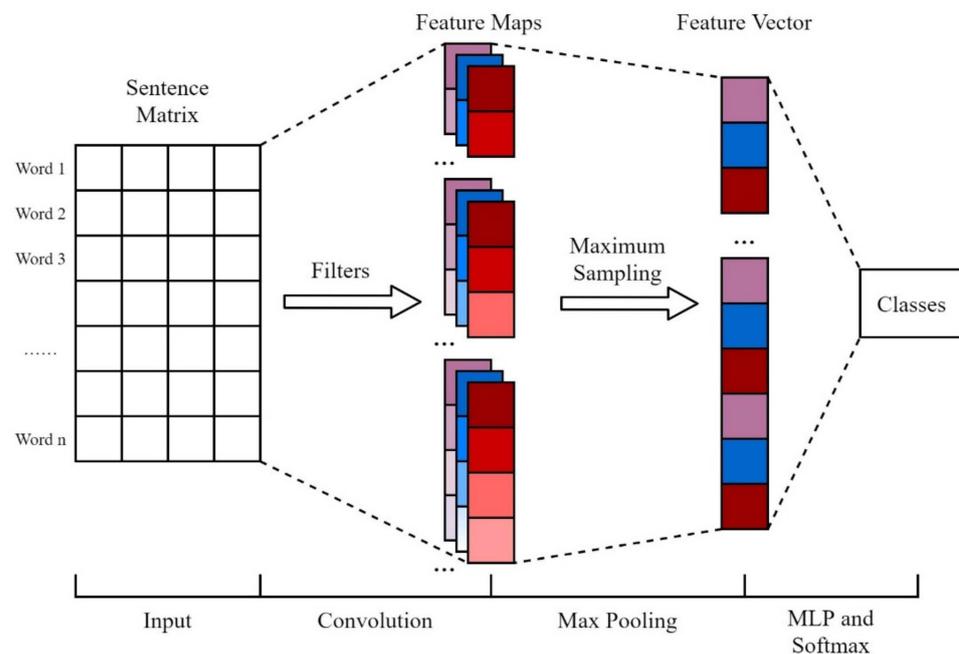


Figure 2. The structure of CNN text classification model used in this work.

3.1.3. Statistics Analysis and Visualization

We performed statistical analysis and visual expression of text data classified by the CNN model, so that the prospecting information in the massive text data can be presented more clearly, and it can better serve the geological research. It mainly included content word extraction and semantic relationship extraction, and finally it was visually expressed through word clouds and knowledge graphs. Among them, knowledge graphs included ternary diagrams drawn by Netdraw and Ucinet6, chord diagrams drawn by Gephi, and word clouds drawn by R language, etc.

Content Word Extraction and Visualization

This study divided geoscience texts into eight categories, namely geological prospecting, geophysical prospecting, geochemical prospecting, remote sensing and metallogenic background, metallogenic period, genetic type, and mineralization type. Content words represent the prospecting information contained in geoscience texts, mainly including terminologies, techniques, data processing, and descriptive texts [7].

Documents containing valuable information often contain high-frequency content words [23], as well as words that are informative but not frequent [24]. If a word or phrase appears frequently in an article and rarely in other articles, it is considered to be a good category differentiator and a good keyword for classification. TF-IDF (Term Frequency-Inverse Document Frequency) is a method to extract low-frequency words that are informative to assess a word's significance to a certain document in a data set.

The term frequency (TF) refers to how often a given word appears in a given document. For a word in a particular document, its importance can be expressed as:

$$tf_{i,j} = \frac{n_{i,j}}{\sum_k n_{k,j}} \quad (1)$$

In the above equation, the numerator is the frequency of occurrences of the word in the document, while the denominator is the sum of the occurrences of all words in the document.

relations in terms of nodes and edges. These diagrams are constructed based on the three variables “from”, “to”, and “weight”. “From” denotes the starting word and “to” denotes the ending word. “From” and “to” are defined based on the order of the substantive words. “Weight” is determined by the co-occurrence frequency of two substantive words in the corpus. Figure 4a is an example of a ternary diagram that illustrates the key information in the text data. The arrow on the edge indicates the word order, pointing to the latter substantive word, and the thickness of the edge indicates the co-occurrence frequency of the substantive words in the text. Figure 4b is an example of a chord diagram showing the interrelationship between substantive words in a geological text. The width of the chords is scaled according to the co-occurrence frequency of the substantive words.

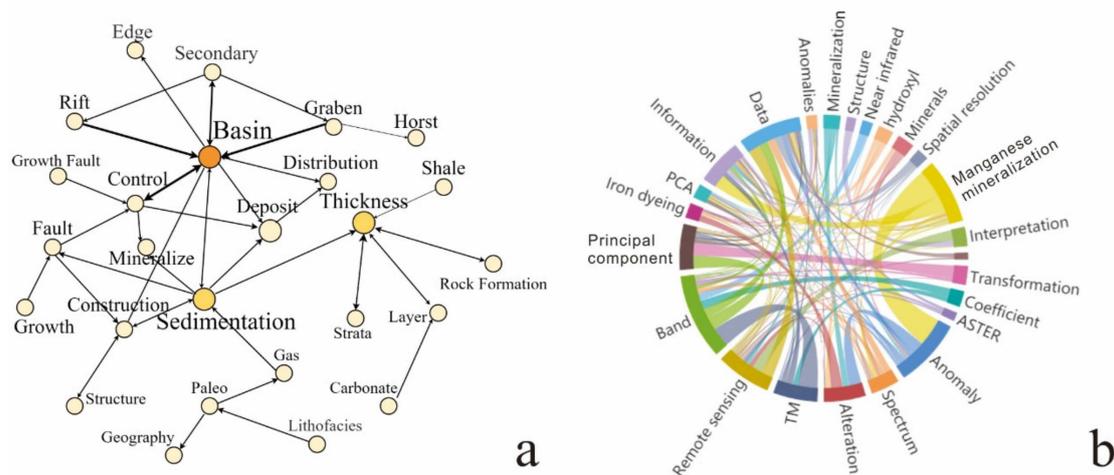


Figure 4. The examples of ternary diagram and chord diagram: (a) ternary diagram of metallogenic background; (b) chord diagram of remote sensing.

3.2. Conceptual Prospecting Model Construction Based on Machine Learning

In the quantitative prediction and evaluation of mineral resources, the construction of a conceptual prospecting model has a very important guiding role [26]. The general idea of this study is to obtain the information of ore-causing anomalies (metallogenic background, metallogenic period, genetic type, mineralization type) and ore-caused anomalies (geological prospecting, geophysical prospecting, geochemical prospecting, and remote sensing) in the study area from the text data through data mining technology, select the metallogenic keywords and ore-controlling factors, match with the conceptual prospecting model database, and construct the conceptual prospecting model of this type of deposit based on machine learning.

The generalized conceptual prospecting model defines the metallogenic background and metallogenic period of the deposit formation, determines the genetic type of the deposit, and fully summarizes the combination of various ore-controlling factors of the same type of deposit in the same geological setting. On this basis, various ore-controlling factors that can be reflected by the actual data collected in the study area were screened out, then the conceptual prospecting model was constructed. Combined with the two-dimensional or three-dimensional digital model of the study area, the cube (or grid unit) was given different characteristic variables to form the digital model of the study area. Based on the digital model, combined with the known ore bodies (or ore points), the favorable metallogenic conditions were analyzed and extracted.

The method of conceptual prospecting model construction used in this article was proposed by Wu in 2017, and the following is a brief description of how it works [21].

3.2.1. Construction of Conceptual Prospecting Model Database

Conceptual prospecting model database is the foundation of conceptual prospecting model construction. The construction of conceptual prospecting model database based on

big data requires the establishment of a unified comprehensive and universal data structure of prospecting model, so as to form a database composed of massive typical deposits. The data organization of the database mainly includes two aspects. One is the name of the model, and the other is the ore-controlling factor. There is a many-to-many relationship between prospecting model and ore-controlling factors (as shown in Figure 5). The sources of the models are mainly evaluation and prediction models of mineral resources potential in China, the existing prospecting models in the laboratory and the prospecting models sorted out from the relevant literature.

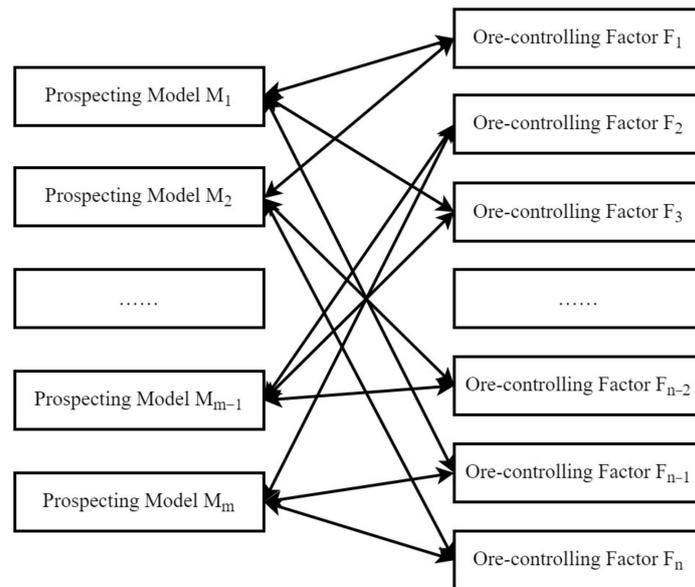


Figure 5. The many-to-many relationship between prospecting model and ore-controlling factors (adapted with permission from Ref. [21]. 2017, Wu, Y.L).

At present, 88 prospecting models in China, 248 metallogenic models for typical deposits in China, and 1521 ore-controlling factors have been collected in this database. An entity description table of the conceptual prospecting model database is shown in Table 1.

Table 1. Entity description table of the conceptual prospecting model database.

Num	Entity	Description
1	Prospecting model	Model number, name, reference, typical deposit, description information, creation time, modification time
2	Ore-controlling factors	Factor number, name, factor category, factor type, creation time, modification time
3	Intermediate table	Model number, factor number

3.2.2. Determination of Prospecting Model

The main process of machine learning approach to determine the best conceptual prospecting model is as follows. The conceptual prospecting model database mentioned in Section 3.2.1 is used as a training set, and the prospecting information obtained in Section 3.1 is used as the data to be processed. The naive Bayesian probability of each model in the database is calculated. Through the calculation of Bayesian probability, deposits similar to the target deposit can be analyzed and the corresponding ore-controlling factors can be extracted. The importance and utilization rate of the factors are calculated, and weighted with the overall Bayesian probability of the model. The weighted calculation results are ranked, and the top-ranked ore-controlling factors are selected to construct the best prospecting conceptual model. The main principles of this method are as follows:

Suppose there are m conceptual prospecting models y_1, y_2, \dots, y_m , denoted as Y ; n ore-controlling factors are collected based on data mining, denoted as X , so there are:

$$Y = \{y_1, y_2, \dots, y_m\} \quad (4)$$

$$X = \{x_1, x_2, \dots, x_n\} \quad (5)$$

The method of model classification is to calculate the probabilities of ore-controlling factors in the study area classified as a conceptual prospecting model. It is to solve for the probability value $P = \{p_1, p_2, \dots, p_m\}$ of $X = \{x_1, x_2, \dots, x_n\}$ in the sample category set $Y = \{y_1, y_2, \dots, y_m\}$, where p_i is the probability that X belongs to the category Y_i .

Assume that there are k_i ore-controlling factors in each conceptual prospecting model. Therefore, there are H ore-controlling factors in m prospecting models:

$$H = \sum_{i=1}^m k_i \quad (6)$$

According to the formulas above, the prior probability $p(Y_i)$ corresponding to each conceptual prospecting model is:

$$P(y_i) = \frac{k_i}{H} \quad (7)$$

According to Bayes theorem, it is obtained that:

$$p(y_i|X) = \frac{p(X|y_i)p(y_i)}{p(X)} \quad (8)$$

The probability that the j -th ($1 \leq j \leq n$) mineral control factor in the study area is in the i -th ($1 \leq i \leq m$) conceptual prospecting model is noted as $p(x_j|y_i)$. Because the individual mineral control factors are conditionally independent, there are:

$$p(X|y_i)p(y_i) = p(x_1|y_i)p(x_2|y_i) \cdots p(x_n|y_i) = p(y_i) \prod_{j=1}^n p(x_j|y_i) \quad (9)$$

Therefore, the Bayesian probability of a certain model in the database is:

$$p_i = \frac{p(y_i) \prod_{j=1}^n p(x_j|y_i)}{p(X)} = \frac{k_i}{H} \prod_{j=1}^n p(x_j|y_i) \quad (10)$$

The Bayesian probability calculation results are sorted according to the probability. On this basis, the keywords extracted after text mining are used as the target keywords of the study area to match the names of high-Bayesian-probability models, then the model matching results are obtained. The ore-controlling factors extracted after text mining are matched with those in the model matching results, and m prospecting models M_1, M_2, \dots, M_m are selected. For a certain model, the ore-controlling factors are divided into c_i different categories according to the ore-controlling geological conditions in the data cleaning process. The number of ore-controlling factors corresponding to each type is $Num_{i1}, Num_{i2}, \dots, Num_{ic_i}$ ($1 \leq i \leq m$), then in the category of the first model, the importance of each ore-controlling factor is:

$$I_{ij} = \frac{1}{Num_{ij}} \quad (11)$$

The range of i is $[1, m]$, and that of j is $[0, c_i]$.

Since one ore-controlling factor may appear in multiple models, the final importance index of any ore-controlling factor in the study area is obtained by adding its importance in each model. The calculation formula is:

$$I = \sum I_{ij} \quad (12)$$

According to the selected m conceptual prospecting models, the number of ore-controlling factors of all these models is H (without deleting the duplicate ore-controlling factors), and the utilization rate of a certain ore-controlling factor can be calculated as follows:

$$f_i = \frac{L}{H} \quad (13)$$

Among them, L refers to the times of occurrences of the ore-controlling factor in m prospecting models.

On this basis, the weighted coefficient of each factor is calculated. Assuming that there is an ore-controlling factor F , whose importance is I and utilization rate is f , it belongs to conceptual prospecting models M_1, M_2, \dots, M_a in the database, the Bayesian probability of the corresponding conceptual prospecting model is p_1, p_2, \dots, p_a . Then, the formula is on this basis, the weighted coefficient of each factor is calculated. Assuming that there is an ore-controlling factor F , whose importance is I and utilization rate is f , it belongs to conceptual prospecting models M_1, M_2, \dots, M_a in the database, the Bayesian probability of the corresponding conceptual prospecting model is p_1, p_2, \dots, p_a . Then, the formula is

$$W = \sum_{i=1}^a p_i I f \quad (14)$$

Based on the calculated weighting factor W , the ore-controlling factors are ranked from highest to lowest according to their W values, and the top-ranked ones are selected to form the best conceptual prospecting model.

In order to verify the correctness of machine learning, several ore-controlling factors in the model are removed. The machine learning results are reliable if there are still the removed factors in the matching results.

4. Experiment

The Songtao-Huayuan area was selected as this study area. A conceptual prospecting model of Songtao-huayuan area will be constructed based on big data.

The study area is located in the adjacent area of Hunan and Guizhou provinces, and it is the most important manganese resource accumulation area in China [27]. In recent years, great progress has been made in manganese ore prospecting [28]. As one of the important metallogenic belts in China, there are many large and medium-sized deposits in northwestern Hunan, such as the Chatian mercury deposit, Minle manganese deposit, Limei lead-zinc deposit, and Naopo lead-zinc deposit. By comparing the geological data of the study area and other favorable areas of sedimentary manganese deposits, it is not difficult to find that the adjacent areas of Hunan and Guizhou have a favorable environment for sedimentary manganese mineralization. Therefore, the geological structure pattern and paleo-sedimentary environment in the adjacent areas of Hunan and Guizhou provinces have gradually become a research hotspot. The study area has also become an important potential area for further exploring the "Datangpo" sedimentary manganese deposits in South China [29,30].

4.1. Prospecting Information Extraction Based on CNN

4.1.1. Data Acquisition and Pre-Processing

The establishment of text data is the basis of text analysis, and the acquisition sources mainly include two aspects. On the one hand, the relevant text data were automatically discovered and crawled from the WAN. On the other hand, the regional reports were discovered from LAN, etc. The statistics are shown in Table 2 below.

Table 2. Statistics of the corpus collection.

Type of Text	Domestic	Foreign	Study Area	Total
Related News	54,308	33,238	74	87,620
Related Literature	9327	4876	111	14,314
Regional Reports	57	0	8	65
Total	63,692	38,114	193	101,999

4.1.2. Text Classification Based on CNN

In this paper, the corpus collected was standardized and labeled with eight different tags, namely geological prospecting, geophysical prospecting, geochemical prospecting, remote sensing, metallogenic background, metallogenic period, genetic type, and mineralization type.

In this study, we extracted some corpora to form training sets, including sentence level and paragraph level. The sentence level contained 7766 labeled samples, and the paragraph level contained 4771 labeled paragraph samples. A total of 1000 samples were randomly selected from each training; 70% of the samples were used as the training set and 30% as the validation set. The testing set was processed by screening the corpus related to “Songtao-Huayuan Manganese Mine”, and 800 words, 800 sentences and 800 paragraphs were randomly selected (Table 3).

Table 3. Statistics describing the numbers of each category and levels in the training and testing sets.

Levels		Geological Prospecting	Geophysical Prospecting	Geochemical Prospecting	Remote Sensing	Metallogenic Background	Metallogenic Period	Genetic Type	Mineralization Type	Total
Training Set	Sentence	1659	506	522	331	3873	171	262	442	7766
	Paragraph	1128	450	452	267	1652	196	288	338	4771
	Word	100	100	100	100	100	100	100	100	800
Testing Set	Sentence	100	100	100	100	100	100	100	100	800
	Paragraph	100	100	100	100	100	100	100	100	800

The best combination of parameters was adjusted by experimental comparison and analysis. The maximum length of corpus samples was set to 216 bytes, the maximum length of paragraph samples was set to 766 bytes, the word vector dimension was set to $d = 128$, and the filter window size was set to $d_{win} = 3,4,5$. The random gradient descent algorithm was used to update the weight according to the set number of cycles of 500, and the model verification was carried out every 10 iterations. Finally, the optimal classification model was trained.

From the overall performance of the model, the CNN classification model based on multi-scale training set had better classification effect on geological texts (Figure 6). The training accuracies of CNN models for ore-caused anomalies and ore-causing anomalies in the classification model of sentence-level data set and paragraph-level data set were 93.6%, 98.4%, 91.2%, and 97.3%, which all achieved the optimal results. The validating accuracies of these four models reached the peak at 500 iterations, which were 99.2%, 89.6%, 99%, and 83.9%, respectively, and also reached the approximate optimal results. The classification errors of the models converged with the increase of iterations and the models did not show over fitting. The training losses were 0.217, 0.033, 0.450, and 0.074, and the validating losses were 0.019, 0.330, 0.042, and 0.772. The classification validating loss increased slightly as the text length of the dataset increased. The classification results for the ore-causing anomaly model were slightly worse compared to the ore-caused anomaly model. This is due to the fact that the categories of ore-causing anomalies were biased towards geological language description and therefore much more difficult to classify and identify.

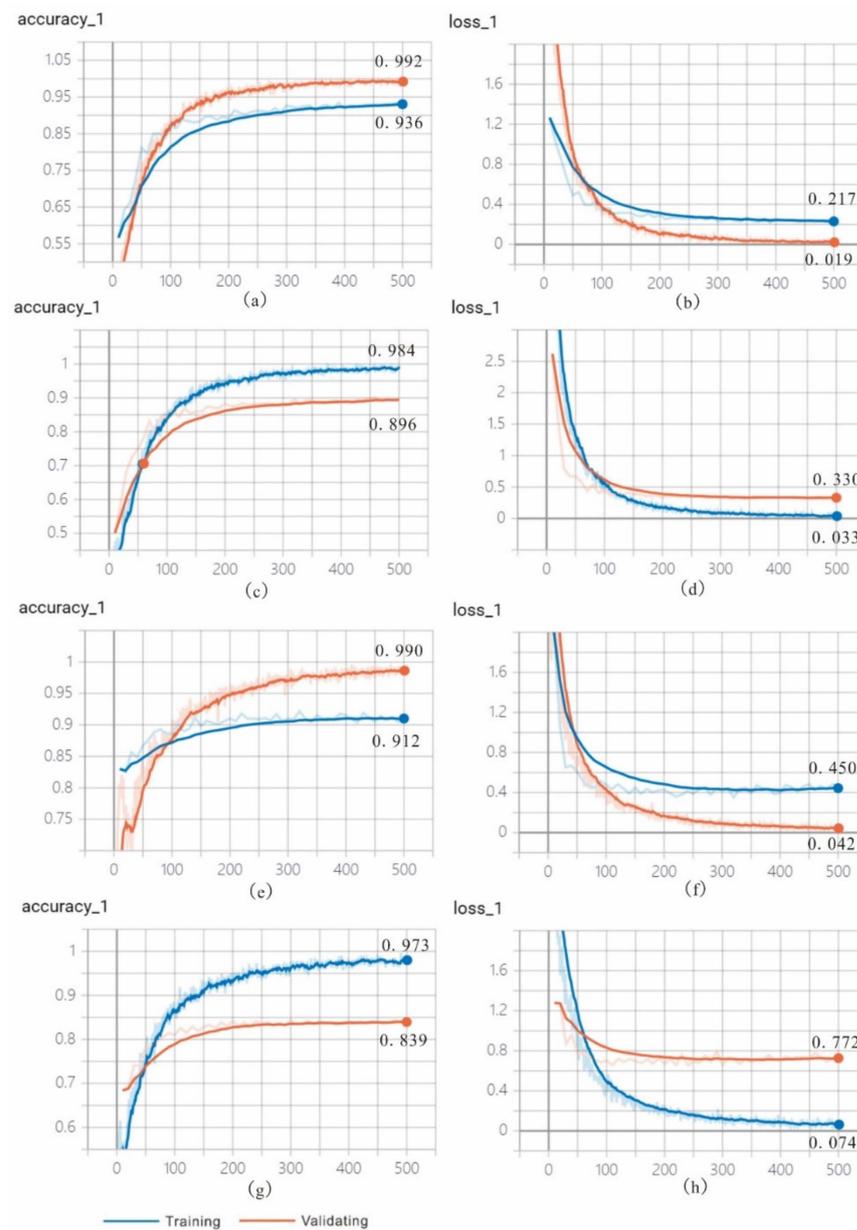


Figure 6. Comparison of training accuracy and loss of CNN model based on multi-scale training set: (a,b) shows the training model of ore-caused anomaly based on sentence training set; (c,d) shows the training model of ore-caused anomaly based on paragraph training set; (e,f) shows the training model of ore-causing anomaly based on sentence training set; (g,h) shows the training model of ore-causing anomaly based on paragraph training set.

A total of 800 test data were extracted from the test sample set for classification test. The test accuracy, recall, and F1 of vocabulary, sentence, and paragraph and their average values are compared in Tables 4 and 5 below. It shows that the classification accuracy of sentences was the highest, followed by paragraphs, and the accuracy difference between these two was 3.3%, indicating that the model has a good classification effect on long texts, but the classification accuracy was lowest for words. The reason is that although the word vector trained with the sentence training set preserved the back-and-forth relationship of the text semantics, there may be multiple different categories of words in a sentence. Therefore, ambiguity arises when classifying words with this model, leading to a decrease in classification accuracy. By comparing the ore-caused anomaly classification group with the ore-causing anomaly classification group, it can be seen that the former was significantly

better than the latter, but the classification effects of different scale test sets still maintained the same pattern. However, in general, the classification model based on CNN was effective in multi-scale geological text classification.

Table 4. Comparison table of evaluation indexes between sentence classification model and paragraph classification model.

	Ore-Caused Anomaly Sentence Classification Model			Ore-Caused Anomaly Paragraph Classification Model			Ore-Causing Anomaly Sentence Classification Model			Ore-Causing Anomaly Paragraph Classification Model		
	Word	Sentence	Paragraph	Word	Sentence	Paragraph	Word	Sentence	Paragraph	Word	Sentence	Paragraph
Test Accuracy	0.860	0.939	0.879	0.895	0.919	0.912	0.569	0.795	0.627	0.594	0.785	0.810
Recall	0.860	0.938	0.877	0.895	0.919	0.911	0.569	0.795	0.628	0.594	0.785	0.810
F1	0.860	0.9385	0.878	0.895	0.919	0.9115	0.569	0.795	0.6275	0.594	0.785	0.810

Table 5. Comparison table of average accuracy, recall, and F1 measure of CNN model testing based on multi-scale test set.

	Ore-Caused Anomaly Classification Model			Ore-Causing Anomaly Classification Model		
	Word	Sentence	Paragraph	Word	Sentence	Paragraph
Average Test Accuracy	87.8%	92.9%	89.6%	58.2%	79%	71.9%
Average Recall	87.8%	92.9%	89.4%	58.2%	79%	71.9%
Average F1	87.8%	92.9%	89.5%	58.2%	79%	71.9%

After CNN multi-scale classification, the original data were automatically divided into eight categories (geological prospecting, geophysical prospecting, geochemical prospecting, remote sensing and metallogenic background, metallogenic period, genetic type, mineralization type), and three different levels (word, sentence, and paragraph).

4.1.3. Statistics Analysis and Visualization

Diverse statistical methods and visualization techniques were proposed for classifying texts in different scales of the study area. This paper performed word frequency statistics for word-level text, TF-IDF statistics for segment-level text to achieve content word extraction, and co-occurrence matrix statistics for sentence-level text to achieve relationship extraction. The visual representation was performed by word cloud and knowledge graph (ternary diagram and chord diagram) to deepen text features and semantic associations in a more targeted way.

Conventional word frequency statistics were performed on the word-level classification results of prospecting information (as shown in Figure 7). Facing the massive text data, various types of geological texts were mixed with other words in the previous word frequency statistics, and some of the keywords were easily washed out by other words, lacking a certain degree of relevance. In contrast, the word frequency statistics of the classified word-level documents made the research more targeted and easier to extract the effective finding information of eight aspects in a targeted manner. For example, Figure 7a shows that the Huayuan manganese deposit is mainly controlled by sedimentary basins in the study area, and electrical exploration is the most important geophysical exploration method in the study (Figure 7b), while geochemical exploration focuses on the high-value anomalies of manganese such as manganese carbonate (Figure 7c). Figure 6d shows that the remote sensing metallogenic information of manganese deposits is indirectly searched by hydroxyl and iron stains. Figure 7e shows that the important ore-controlling factor in the study area is the structure controlling the sedimentary strata, and the main metallogenic epoch is the Datangpo stage of the South China Era (Figure 7f). The genetic type of Huayuan manganese deposit is generally considered as sedimentary manganese deposit (Figure 7g), and the most important metallogenic type is rhodochrosite (Figure 6h).

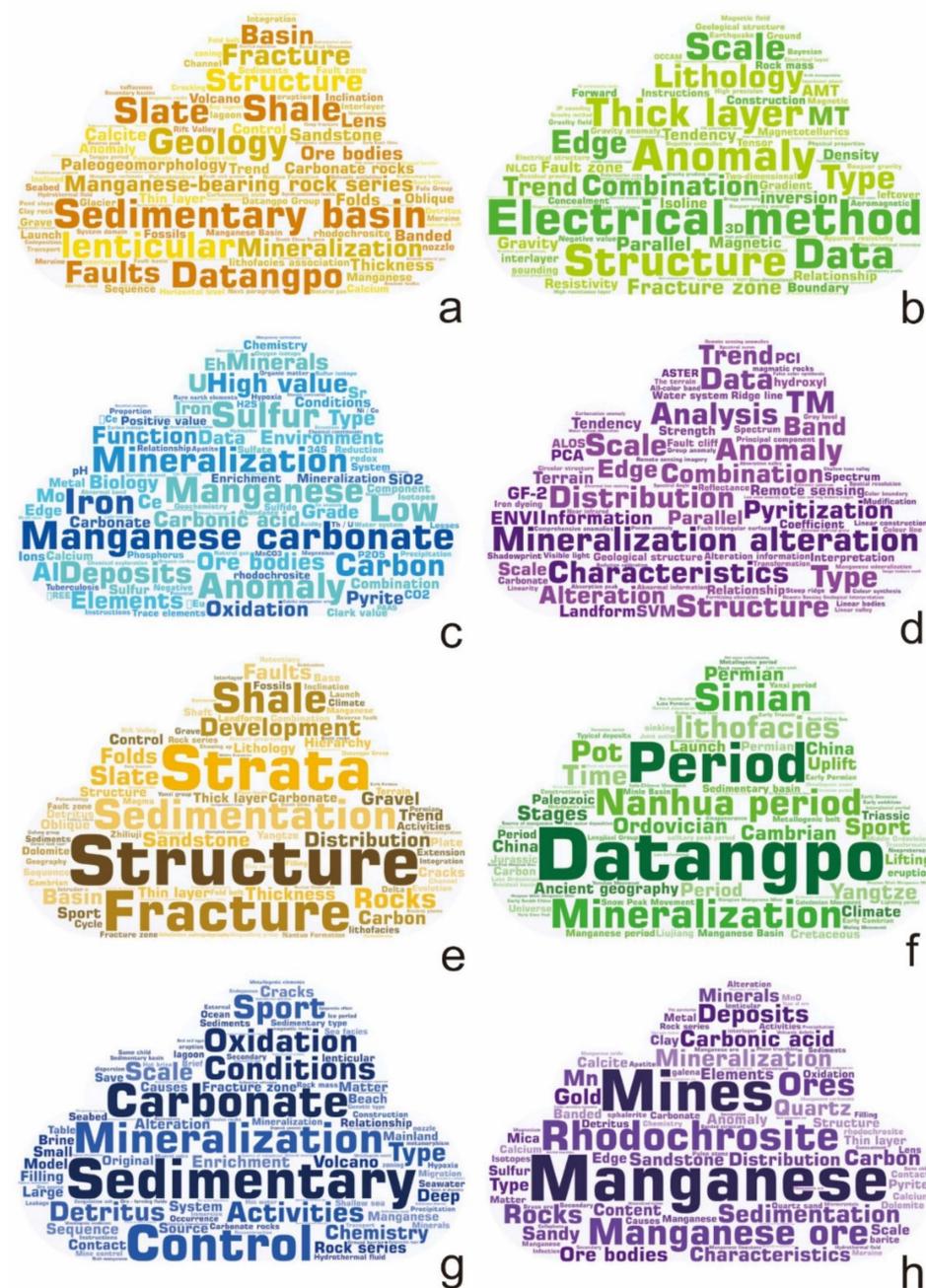


Figure 7. Word clouds of Songtao-Huayuan manganese ore: (a) geological prospecting, (b) geophysical prospecting, (c) geochemical prospecting, (d) remote sensing, (e) metallogenic background, (f) metallogenic period, (g) genetic type, and (h) mineralization type.

The results of paragraph level classification were analyzed by TF-IDF and the top ten substantive terms were obtained (Figure 8). By comparing the word frequency and TF-IDF extraction results, we found that TF-IDF was more targeted and more suitable for geological terminology extraction, such as “graben”, “rift basin”, “synsedimentary fault”, etc. The words extracted by the common word frequency method were more general, such as “geology”, “shale”, “mineralization”, etc. In terms of the extraction of geophysical prospecting substantives, TF-IDF had obvious advantages in the extraction of technical methods and data processing vocabulary, such as “inversion” and “AMT” and so on. For geochemical prospecting, more factors related to mineralization were extracted compared with word frequency statistics. TF-IDF was more effective in extracting descriptive vocabulary of remote sensing, and it also extracted some vocabulary related

to technical methods. This is because the text of remote sensing data was more general, so the substantive vocabulary extracted by the two methods was basically consistent. Similar to the results of remote sensing, the results of vocabulary extraction of metallogenic background by TF-IDF were basically consistent with the results of word frequency statistics. In terms of the metallogenic period, on the basis of specific geological ages such as “early Nanhua era” and “Datangpo”, some geological time units with certain descriptions were added, such as “interglacial”. In terms of genetic types, more professional words were extracted by the TF-IDF method, such as “exogenous”. By comparing the results of TF-IDF and word frequency statistics of metallogenic types, it was found that the results of TF-IDF added more details to word frequency statistics, such as “greinerite”.

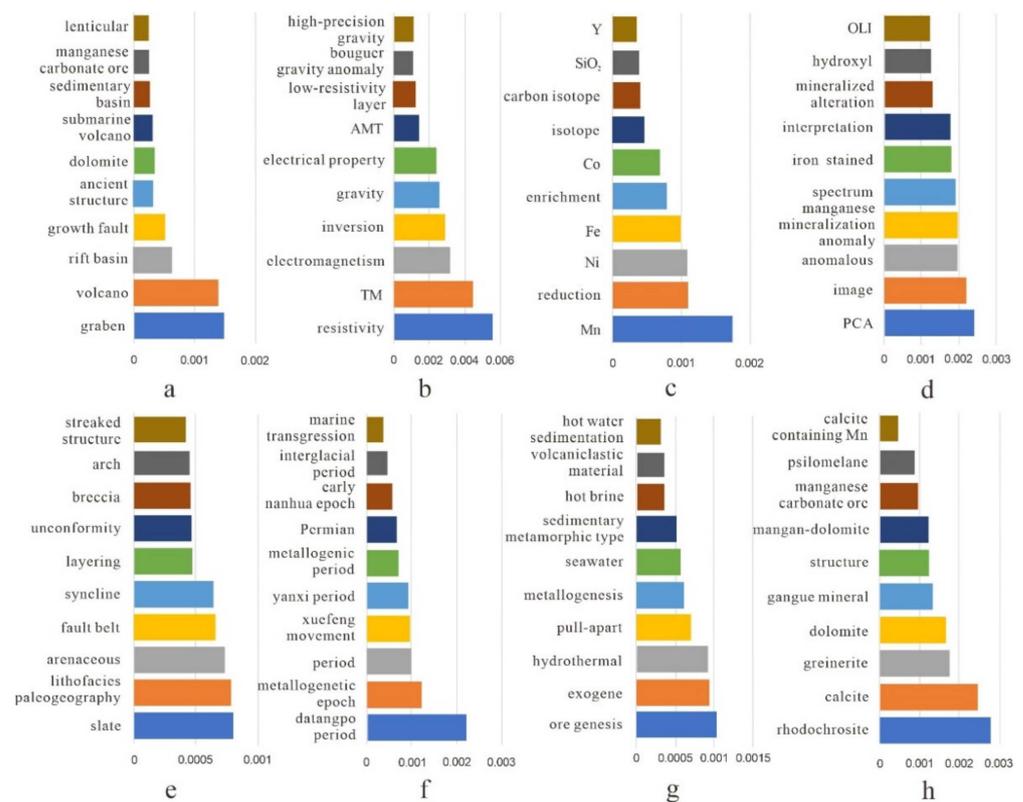


Figure 8. TF-IDF statistical graph of Songtao-Huayuan manganese ore: (a) geological prospecting, (b) geophysical prospecting, (c) geochemical prospecting, (d) remote sensing, (e) metallogenic background, (f) metallogenic period, (g) genetic type, and (h) mineralization type.

The co-occurrence matrix statistics were carried out for the sentence level to extract the semantic relationship, and the knowledge graph of Songtao-Huayuan manganese deposit was drawn as shown in Figure 9. It can be seen from the geological prospecting ternary diagram in Figure 9a that manganese deposits in the study area are mainly controlled by basins, especially secondary graben basins, and most of these basins related to mineralization are controlled by synsedimentary faults. The distribution of manganese ore in the study area is also closely related to the thickness of shale deposition. The geophysical prospecting ternary diagram (Figure 9b) shows that the commonly used methods related to manganese exploration are high precision gravity method and audio-frequency magnetotelluric method (MT, AMT). The geochemical ternary diagram (Figure 9c) reflects the widespread occurrence of iron and manganese in the study area. The main chemical composition of the ore is manganese carbonate, and the main redox environment formed is the reduction environment. The remote sensing chord diagram in Figure 9d shows that the anomaly information of manganese mineralization is closely related to hydroxyl and iron staining alteration, and the remote sensing information of manganese mineralization can

be extracted by the relevant alteration combination. The ternary diagram of metallogenic background (Figure 9e) shows that the formation of manganese ore is related to the distribution of sedimentary basins controlled by faults. The fold structure in the study area, especially the occurrence of anticlines, leads to the exposure of ore-bearing strata, and then manganese ore can be found and exploited. The ternary diagram of metallogenic period indicated that the main metallogenic period of Songtao-Huayuan manganese deposit is the Datangpo period of Early Nanhua era (Figure 9f). From the ternary diagram of genetic type (Figure 9g), it can be seen that experts and scholars have various understandings of the genetic type of Songtao-Huayuan manganese deposit, which is generally sedimentary manganese deposit. However, some scholars believe that it is affected by ancient gas leakage, hot water sedimentation, or volcanic action. These understandings are dominant in the current research of experts and scholars. Figure 9h is the mineralization type diagram of manganese ore in the study area, which indicates that the main types of manganese ore in the study area are rhodochrosite, calcium rhodochrosite and manganese-bearing calcite. In addition to manganese ore, pyrite, galena, and sphalerite are also produced in the study area.

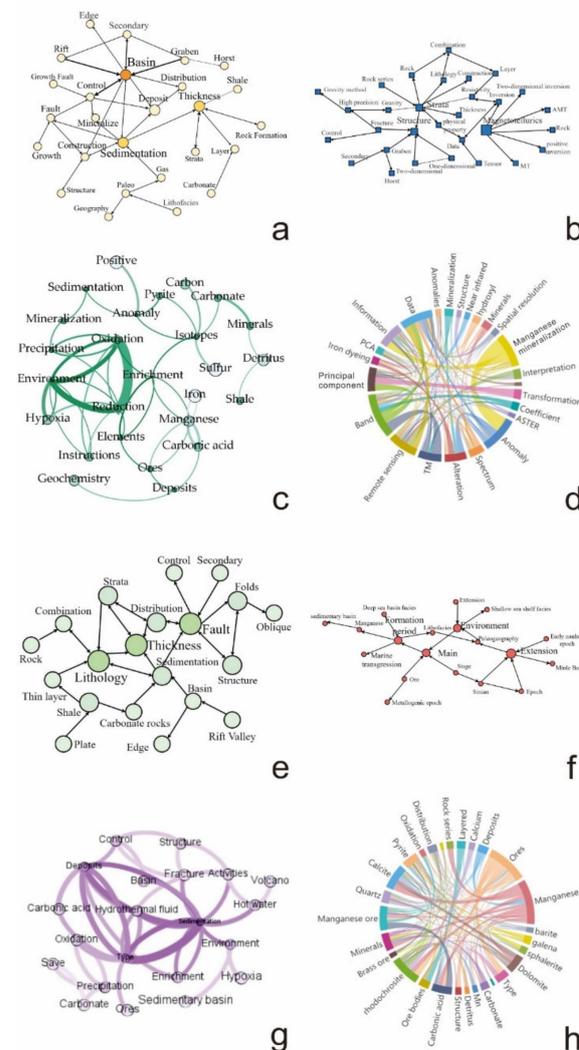


Figure 9. Visualization of semantic relationships extracted: (a) ternary diagram of geological prospecting, (b) ternary diagram of geophysical prospecting, (c) ternary diagram of geochemical prospecting, (d) chord diagram of remote sensing, (e) ternary diagram of metallogenic background, (f) ternary diagram of metallogenic period, (g) ternary diagram of genetic type, and (h) chord diagram of mineralization type.

4.1.4. Generalized Conceptual Prospecting Model Construction

According to the results of big data text analysis (Figures 7–9), the generalized prospecting model in the study area is summarized as follows:

Ore-Caused Anomalies

Geological prospecting: Songtao-Huayuan manganese deposit is mainly controlled by secondary graben basins. Therefore, synsedimentary faults have become relatively important prospecting factors. The sedimentary facies related to manganese deposits in the study area are manganese-bearing shale facies, and the sedimentary formation is black shale formation. The deposit is composed of lenticular ore bodies. The manganese ore in the study area was originally buried deeply, but it was exposed due to tectonic activity. The lower section of Datangpo Formation is the occurrence horizon of manganese ore in this area. The thickness of ore body is basically positively correlated with the thickness of manganese-bearing rock series.

Geophysical prospecting: The main geophysical methods used in the study area are AMT and high-precision gravity method. There are many strata and complex lithology in the study area, but it can be basically divided into the resistivity structure of “high resistivity–low resistivity–high resistivity”. The manganese-bearing rock series is located in the transition zone of the second low resistivity layer and the third high resistivity layer, and has the electrical characteristics of high polarizability and low resistivity. The Bouguer gravity anomaly in the study area is negative, and the distribution of ore-bearing strata in the study area can be inferred by studying the Bouguer residual gravity anomaly.

Geochemical prospecting: Iron and manganese are commonly associated in the study area. The main chemical composition of the ore is manganese carbonate (MnCO_3), which is mainly formed in a relatively anoxic deep water reduction environment. Chemical elements related to the distribution characteristics of manganese deposits in the study area are Mn, P, Fe, etc.

Remote sensing: The abnormal information of manganese mineralization is closely related to the alteration of hydroxyl and iron. The remote sensing information of manganese mineralization can be extracted by the relevant alteration combinations.

Ore-Causing Anomalies

Metallogenic background: The study area is located in the southeast margin of the Upper Yangtze Block, and the stratigraphic area is a transitional zone between the Yangtze Block and the South China Block. The distribution of manganese deposits in the area is strictly controlled by major faults, especially synsedimentary faults. At present, most of the known manganese deposits are distributed in secondary graben basins.

Metallogenic period: The main metallogenic period of Huayuan manganese deposit is Datangpo period. In the extensional environment of the Datangpo period, synsedimentary faults constitute a manganese sedimentary basin. With the advent of the interglacial period, the climate at that time became warm, leading to glacier ablation and a sharp rise in sea level. At the same time, due to the decline of the crust after the Xuefeng Movement, the transgression began, and the shallow-sea shelf deposits formed.

Genetic type: A general understanding of Songtao-Huayuan manganese deposit is that it belongs to sedimentary manganese deposit. However, there are different opinions on its specific genesis, including biochemical sedimentary genesis, carbonate cap sedimentary genesis, volcanic eruption-sedimentary genesis, ancient gas leakage genesis, and hot water sedimentary genesis. The most influential one is sedimentary genesis, but the ancient gas leakage genesis, hot water sedimentary genesis, and volcanic eruption-sedimentary genesis also have certain influence.

Mineralization type: The manganese minerals in the study area are mainly rhodochrosite, followed by manganese calcite, calcium rhodochrosite, etc. The gangue minerals are mainly quartz, dolomite, barite, pyrite, and so on. In addition to manganese ore, lead-zinc deposits

(galena and sphalerite), copper deposits (chalcopyrite), and iron deposits (pyrite) are also developed in the study area.

4.2. Conceptual Prospecting Model Construction Based on Machine Learning

Based on the above big data text analysis and mining technology of Songtao-Huayuan manganese ore, the keywords are: “sedimentary”, “volcano”, “hot water”, “ancient gas”, “manganese ore”; the selected ore-controlling factors are as follows: “deep fault belt”, “early Nanhua epoch Datangpo priod”, “synsedimentary fault”, “rift basin”, “sedimentary basin”, “shallow-sea basin facies”, “shallow-sea carbonate rock”, “shelf edge paleogeographic deposition”, “manganese-bearing rock series (black shale)”, “manganese carbonate”, “rhodochrosite mineralization”, “hydroxyl alteration”, “iron alteration”, “lenticular”, “interglacial period”, “Bouguer gravity anomaly”, “low resistivity”, “high polarizability”, “Mn anomaly”, “P anomaly”, and “Fe anomaly”. The keywords and ore-controlling factors obtained by text mining are matched with the database. Through the calculation of Bayesian probability, the deposit model with high similarity with Songtao-Huayuan manganese deposit can be selected (partly shown in Table 6), and the corresponding ore-controlling factors can be extracted.

Table 6. List of the top 15 prospecting models in Bayesian probability ranking.

Rank	Model ID	Bayesian Probability	Model Name
1	ba0be2863b874b4086a7a359f423b6e4	0.077754	Shallow marine sedimentary manganese deposit in Dounan, Yunnan
2	f1007cc260d442c084ddd69ef09a47da	0.013668	Sedimentary manganese deposit
3	7eff184b0ee24d34aa2ae4ed0fc25d02	0.010934	Sedimentary iron deposit
4	fe0efa84176e4ea78a5b6eb5c4bb2aed	0.010934	Sedimentary manganese deposit in Xialei, Guangxi
5	e827a69faf85494bad054d0d8aed6bb2	0.008639	Layered carbonate lead-zinc-silver ore
6	b7416ae01c3f4bf9a6758af889ec32ab	0.008639	Sedimentary natural pyrite ore
7	33593590864c4a87adff8482d9016b25	0.008639	Sedimentary pyrite ore
8	6b760cc451ae4f4e87dadddfc6ab3cd5	0.004999	Carbonate type potash deposits
9	149c98869b704bf0942c6cab67f4e0d4	0.004999	Marine volcanic eruption sedimentary iron-copper-sulfur deposit
10	81a2d91389e34f258658dbb63bb56fee	0.004665	Weathered crust type manganese ore
11	4a1887ad25ad45c49c98062aa87521bf	0.004116	Hydrothermal antimony polymetallic deposit in clastic rock strata
12	32cf88ba33e844a8abdd0a36136e3daa	0.003888	Layered or hydrothermal veined layer-controlled barite ore
13	2e874ed4a60f41949cc1455d1d5eda1c	0.003499	Marine or volcanic sedimentary rock type copper-silver-gold deposits
14	545b0ff20d904d3183ed1dd4ec0e89bb	0.003499	Hydrothermal antimony deposits in carbonate rocks
15	7bb1b395c8a54c179e97c463bdacf050	0.003499	Continental volcanic type pyrite ore

We calculate the importance and utilization rate of each ore-controlling factor in these models. Combined with the Bayesian probability of each model, we calculate the weighted value W , and rank the values of W from large to small. Table 7 is an example of W value calculation of ore-controlling factors in shallow marine sedimentary manganese deposit in Dounan, Yunnan.

We selected the top-ranked ore-controlling factors to form the best conceptual prospecting model. Finally, the conceptual model of Songtao-Huayuan manganese ore prospecting was formed as in Table 8. By comparing with the names of ore-controlling factors obtained from text mining, it was found that there were more factors not extracted from text mining, such as “Gravity Anomaly Transformation Zone” and “Y anomaly”. We believe that the differences can reflect the additional value created by machine learning methods to some extent.

Table 7. An example of importance, utilization rate, and Bayesian probability weighting calculation.

Deposit Name	Factor Type	Factor Name	Utilization Rate	Importance	Bayesian Probability	W	
Shallow marine sedimentary manganese deposit in Dounan, Yunnan	Stratigraphic Signatures	Middle Triassic Ladinian stage	3	3	0.077754	0.699786	
		Deyoujiang fold belt of South China fold system	2	2	0.077754	0.311016	
		Speculated distribution of manganese-bearing rock series	2	0.67	0.077754	0.10419	
		Calcareous siltstone	2	0.67	0.077754	0.10419	
		Manganese-bearing outcrops	32	16.03	0.077754	39.88469	
	Geochemistry Signatures	Bioclastic limestone intercalated with mudstone	2	0.67	0.077754	0.10419	
		Mn anomaly	2	2	0.077754	0.311016	
		Geophysical Signatures	Aeromagnetic anomaly	9	9	0.077754	6.298074
			Gravity anomaly	3	3	0.077754	0.699786
		Tectonic Signatures	Fault	19	18	0.077754	26.59187
		Ore Body Morphology	Lenticular	21	8.53	0.077754	13.92807
	Interbedded	41	19.53	0.077754	62.25996		

Table 8. Songtao-Huayuan manganese ore conceptual prospecting model.

Deposit Name	Factor Type	Factor Name
Sedimentary manganese ore of "Datangpo style"	Rock conditions	Interglacial period, thick moraine conglomerate
	Ore body morphology	Lenticular, interbedded
	Stratigraphic signatures	Nanhua epoch Datangpo period Manganese-bearing outcrops Speculated distribution of manganese-bearing rock series
	Tectonic signatures	Manganese forming basin Synsedimentary fault Petrographic paleogeography
	Geophysical signatures	Gravity anomaly Gravity Anomaly Transformation Zone
	Geochemistry signatures	Mn anomaly P anomaly Y anomaly

5. Conclusions

A conceptual model construction method of mineral search based on geological big data was innovatively proposed. Taking Songtao-Huayuan manganese ore in Hunan province as an example, this paper completes the intelligent classification and annotation of geological text big data, intelligently extracts the key information of mineral prospecting from the massive geological text data, and constructs the prospecting model of the study area. This method provides a strong basis for the automatic construction of the geological big data-based prospecting model.

This study shows that CNN-based classification models are effective in multi-scale geoscience text classification. Different classification models were trained for words, sentences, and paragraphs, with the best classification results for long texts and the second-best results for words. The method effectively avoids the phenomenon that important information is overwhelmed by the massive data due to the mixing of multiple types of text data, and provides a more targeted processing method for the intelligent extraction of prospecting information.

The TF-IDF statistics and co-occurrence matrix statistics were used to extract content words and semantic relationships, and the keywords and semantic associations of ore-causing and ore-caused anomalies were expressed visually through word clouds, ternary diagrams, and chord diagrams, which effectively and purposefully characterized the potential prospecting information in geological literature.

This paper classifies the common ore-controlling factors in the prospecting model into two categories: ore-causing anomalies and ore-caused anomalies, with a total of eight subcategories. For these eight different subcategories, a CNN multi-scale classifier was

used to classify the geological literature. The test accuracy rates of ore-caused anomalies were 87.8%, 92.9%, and 89.6%, respectively; and ore-causing anomalies were 58.2%, 79%, and 71.9%. By applying different statistical and visualization methods to different types of texts, key prospecting information was extracted and machine learning algorithms were used to match the conceptual prospecting model of the study area. This classification method is more detailed than the method we proposed in our previous study, which is of great significance to the prospecting of sedimentary types closely related to ore-causing anomalies such as metallogenic background and metallogenic period.

In our previous work, the keywords used to match the prospecting model database were obtained by reading literature. This paper uses the geological big data method to replace the manual method in the stage of keyword selection, laying the foundation for the intelligent construction of prospecting model in the future.

Author Contributions: J.C. and S.L. designed the project; J.C. conducted the original literature reviews; C.L., S.L. and T.Q. carried out the experiment; C.L. wrote and organized the paper with a careful discussion and revision by J.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was financially supported by the Chinese MOST project “Methods and Models for Quantitative Prediction of Deep Metallogenic Geological Anomalies” (No.2017YFC0601502).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are not publicly available because it’s confidential.

Acknowledgments: The authors acknowledge Yongliang Wu and Zhijie Jia for their technical support.

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

References

1. Bonham-Carter, G.F. Geographic information systems for geoscientists—modeling with GIS. *Comput. Methods Geosci.* **1994**, *13*, 398.
2. Carranza, E.J.M.; Hale, M. Logistic regression for geologically constrained mapping of gold potential, Baguio district, Philippines. *Explor. Min. Geol.* **2001**, *10*, 165–175. [[CrossRef](#)]
3. Zhao, P.D. Quantitative mineral prediction and deep mineral exploration. *Earth Sci. Front.* **2007**, *14*, 1–10.
4. Agterberg, F. *Geomathematics: Theoretical Foundations, Applications and Future Developments*; Springer International Publishing: Berlin, Germany, 2014.
5. Zhao, P.D. Digital mineral exploration and quantitative evaluation in the big data age. *Geol. Bull. China* **2015**, *34*, 1255–1259.
6. Feldman, R.; Dagan, I.; Hirsh, H. Mining text using keyword distributions. *J. Intell. Inf. Syst.* **1998**, *10*, 281–300. [[CrossRef](#)]
7. Wang, C.; Ma, X.; Chen, J.; Chen, J. Information extraction and knowledge graph construction from geoscience literature. *Comput. Geosci.* **2018**, *112*, 112–120. [[CrossRef](#)]
8. Hinton, G.E.; Osindero, S.; Teh, Y.W. A fast learning algorithm for deep belief nets. *Neural Comput.* **2006**, *18*, 1527–1554. [[CrossRef](#)] [[PubMed](#)]
9. LeCun, Y.; Bottou, L.; Bengio, Y.; Haffner, P. Gradient-based learning applied to document recognition. *Proc. IEEE* **1998**, *86*, 2278–2324. [[CrossRef](#)]
10. Lei, T.; Barzilay, R.; Jaakkola, T. Molding CNNs for text: Non-linear, non-consecutive convolutions. *Indiana Univ. Math. J.* **2015**, *58*, 1151–1186.
11. Bengio, Y.; Ducharme, R.; Vincent, P. A neural probabilistic language model. *J. Mach. Learn. Res.* **2003**, *3*, 1137–1155.
12. Kim, Y. Convolutional neural networks for sentence classification. *arXiv* **2014**, arXiv:10.3115/v1/D14–1181.
13. Liang, J.; Chai, Y.; Yuan, H.; Zan, H.; Liu, M. Deep learning for Chinese micro-blog sentiment analysis. *J. Chin. Inf. Process.* **2014**, *28*, 155–161.
14. Sun, S.; He, Y. Multi-label emotion classification for microblog based on CNN feature space. *J. Sichuan Univ.* **2017**, *49*, 162–169. [[CrossRef](#)]
15. Feng, S.; Wang, Y.; Song, K.; Wang, D.; Yu, G. Detecting multiple coexisting emotions in microblogs with convolutional neural networks. *Cogn. Comput.* **2018**, *10*, 136–155. [[CrossRef](#)]
16. Schuhmacher, M.; Ponzetto, S.P. Knowledge-Based Graph Document Modeling. In Proceedings of the WSDM 14: 7th ACM International Conference on Web Search and Data Mining, New York, NY, USA, 24–28 February 2014; pp. 543–552. [[CrossRef](#)]

17. Ji, S.; Pan, S.; Cambria, E.; Marttinen, P.; Philip, S.Y. A survey on knowledge graphs: Representation, acquisition, and applications. *IEEE Trans. Neural Netw. Learn. Syst.* **2022**, *33*, 494–514. [[CrossRef](#)] [[PubMed](#)]
18. Lu, F.; Yu, L.; Qiu, P. On geographic knowledge graph. *J. Geo.-Inf. Sci.* **2017**, *19*, 723–734.
19. Hou, Z.W.; Zhu, Y.Q.; Gao, Y.; Song, J.; Qin, C. Geologic time scale ontology and its applications in semantic retrieval. *J. Geo-Inf. Sci.* **2018**, *20*, 17–27.
20. Xu, J.; Pei, T.; Yao, Y. Conceptual framework and representation of geographic knowledge map. *J. Geo-Inf. Sci.* **2010**, *12*, 496–502. [[CrossRef](#)]
21. Wu, Y.L.; Jia, Z.J.; Chen, J.P.; Zhu, Y.Q. Construction and prediction of prospecting model based on big data intelligence. *China Min. Mag.* **2017**, *26*, 79–84.
22. Shi, L.; Jianping, C.; Jie, X. Prospecting information extraction by text mining based on convolutional neural networks—A case study of the lala copper deposit, China. *IEEE Access* **2018**, *6*, 52286–52297. [[CrossRef](#)]
23. Hovy, E.; Lin, C.-Y. Automated text summarization and the SUMMARIST system. In Proceedings of the TIPSTER 98: A Workshop, Baltimore, MD, USA, 13–15 October 1998; pp. 197–214. [[CrossRef](#)]
24. Piantadosi, S.T. Zipf’s word frequency law in natural language: A critical review and future directions. *Psychon. Bull. Rev.* **2014**, *21*, 1112–1130. [[CrossRef](#)] [[PubMed](#)]
25. Fries, C.C. *The Structure of English: An Introduction to the Construction of English Sentences*; Harcourt, Brace & World: New York, NY, USA, 1952.
26. Yu, P.; Chen, J.; Chai, F.; Zheng, X.; Yu, M.; Xu, B. Research on model-driven quantitative prediction and evaluation of mineral resources based on geological big data concept. *Geol. Bull. China* **2015**, *34*, 1333–1343.
27. Gao, L.; Xu, S.; Hu, X.; Liu, S.; Zhou, Q.; Yang, B. Sedimentary setting and ore-forming model in the songtao manganese deposit, Southwestern China: Evidence from audio-frequency magnetotelluric and gravity data. *Minerals* **2021**, *11*, 1273. [[CrossRef](#)]
28. Zhou, Q.; Du, Y.S.; Yuan, L.J.; Zhang, S.; Yu, W.C.; Yang, S.T.; Liu, Y. The structure of the Wuling Rift Basin and its control on the manganese deposit during the Nanhua period in Guizhou-Hunan-Chongqing Border Area, South China. *Earth Sci.* **2016**, *41*, 177–188.
29. Du, Y.S.; Zhou, Q.; Yu, W.C.; Wang, P.; Yuan, L.; Qi, L.; Xu, Y. Linking the Cryogenian manganese metallogenic process in the southeast margin of Yangtze block to break-up of rodinia supercontinent and sturtian glaciation. *Geol. Sci. Technol. Inf.* **2015**, *34*, 1–7.
30. Zhou, Q.; Du, Y.; Qin, Y. Ancient natural gas seepage sedimentary type manganese metal logenic system and ore-forming model: A case study of “Datangpo type” manganese deposits formed in rift basin of Nanhua Period along Guizhou-Hunan-Chongqing border area. *Miner. Depos.* **2013**, *32*, 457–466.