International Journal of
Geo-Information

MDPI

Article

Spatio-Temporal Relevance Classification from Geographic Texts Using Deep Learning

Miao Tian [1], Xinxin Hu [2], Jiakai Huang [3], Kai Ma [2], Haiyan Li [2], Shuai Zheng [2], Liufeng Tao [4,5] and Qinjun Qiu [4,5,*]

1 Key Laboratory of Geological Survey and Evaluation of Ministry of Education, China University of Geosciences, Wuhan 430074, China; tianmiao@ctgu.edu.cn
2 College of Computer and Information Technology, China Three Gorges University, Yichang 443002, China; huxinxin@ctgu.edu.cn (X.H.); makai@ctgu.edu.cn (K.M.); lihaiyan@ctgu.edu.cn (H.L.); zhengshuai@ctgu.edu.cn (S.Z.)
3 Hubei Geological Survey, Wuhan 430034, China; jiake_huang@hubgs.com
4 School of Computer Science, China University of Geosciences, Wuhan 430074, China; taoliufeng@cug.edu.cn
5 Ministry of Nature Resources Key Laboratory of Quantitative Resources Assessment and Information Technology, Wuhan 430074, China
* Correspondence: qiuqinjun@cug.edu.cn

Abstract: The growing proliferation of geographic information presents a substantial challenge to the traditional framework of a geographic information analysis and service. The dynamic integration and representation of geographic knowledge, such as triples, with spatio-temporal information play a crucial role in constructing a comprehensive spatio-temporal knowledge graph and facilitating the effective utilization of spatio-temporal big data for knowledge-driven service applications. The existing knowledge graph (or geographic knowledge graph) takes spatio-temporal as the attribute of entity, ignoring the role of spatio-temporal information for accurate retrieval of entity objects and adaptive expression of entity objects. This study approaches the correlation between geographic knowledge and spatio-temporal information as a text classification problem, with the aim of addressing the challenge of establishing meaningful connections among spatio-temporal data using advanced deep learning techniques. Specifically, we leverage Wikipedia as a valuable data source for collecting and filtering geographic texts. The Open Information Extraction (OpenIE) tool is employed to extract triples from each sentence, followed by manual annotation of the sentences' spatio-temporal relevance. This process leads to the formation of quadruples (time relevance/space relevance) or quintuples (spatio-temporal relevance). Subsequently, a comprehensive spatio-temporal classification dataset is constructed for experiment verification. Ten prominent deep learning text classification models are then utilized to conduct experiments covering various aspects of time, space, and spatio-temporal relationships. The experimental results demonstrate that the Bidirectional Encoder Representations from Transformer-Region-based Convolutional Neural Network (BERT-RCNN) model exhibits the highest performance among the evaluated models. Overall, this study establishes a foundation for future knowledge extraction endeavors.

Keywords: spatio-temporal text classification; geographical knowledge; spatio-temporal relevance; deep learning; geographical text

Citation: Tian, M.; Hu, X.; Huang, J.; Ma, K.; Li, H.; Zheng, S.; Tao, L.; Qiu, Q. Spatio-Temporal Relevance Classification from Geographic Texts Using Deep Learning. *ISPRS Int. J. Geo-Inf.* **2023**, *12*, 359. https://doi.org/10.3390/ijgi12090359

Academic Editors: Wolfgang Kainz, Peng Peng, Shu Wang, Maryam Lotfian, Feng Lu and Yunqiang Zhu

Received: 5 July 2023
Revised: 24 August 2023
Accepted: 30 August 2023
Published: 1 September 2023

1. Introduction

Spatio-temporal data are a rich information resource and production factor in contemporary society, and play an important role in national information construction and socialization applications [1–3]. With the advent of the era of big data, spatio-temporal data services are facing the prominent contradiction of "massive data, information explosion and lack of knowledge" [4], which has given rise to the transformation of traditional information services into knowledge services. Spatio-temporal knowledge can be regarded as all relevant knowledge with the characteristics of a spatio-temporal location and dynamic change [5], which is embedded in different modes and types of spatio-temporal data, with the characteristics of multi-grain size, heterogeneity, high-dimensional and low density, and sparse samples [5]. Currently, research on tuple combination adaptive expression models for different spatio-temporal object types primarily focuses on organizing spatio-temporal knowledge of various object types using ontologies and graphs. This approach helps in constructing tuple combinations to create expression models that offer semantic support for analyzing, storing, modeling, intelligent computation, mining, as well as spatio-temporal reasoning and prediction [6–9]. Among them, the graph-oriented spatio-temporal knowledge expression model mainly expresses spatio-temporal knowledge in the form of triples, and spatio-temporal information is often associated in the form of attributes, resulting in low efficiency and low precision in query and retrieval of spatio-temporal knowledge. For example, "*In 2008, the president of the United States is Barack Obama*" contains a triple (Obama, is, president), which may be wrong knowledge at the current time due to the lack of temporal information association, because the current president of the United States is Joe Biden. Therefore, it is very necessary to analyze the degree of association between spatio-temporal elements and tuples in sentences. The purpose of spatio-temporal correlation classification is to determine the degree of spatio-temporal correlation in sentences containing spatio-temporal information, that is, given a sentence containing spatio-temporal information, to determine the degree of correlation between spatio-temporal information in the sentence and the extraction of tuples, thereby accurately describing the spatio-temporal elements in tuple knowledge. The classification of the spatio-temporal correlation degree can provide important support for rapid retrieval, rule judgment, and reasoning of a spatio-temporal knowledge graph [10].

In this paper, we aim to enhance the clarity of the relationship between geographical knowledge and spatio-temporal information. We categorize sentences containing spatio-temporal information into four types: strong relevance, moderate relevance, weak relevance, and non-relevance. We accomplish this by extracting corresponding entity relationship triples from sentences in the dataset, determining the spatio-temporal correlation of the sentences based on knowledge types and annotation rules. Subsequently, we accurately map the spatio-temporal correlation to the entity relationship triple, resulting in the formation of a spatio-temporal knowledge quadruple. Finally, a text classification method based on deep learning is employed for classification purposes. The specific implementation is shown in Table 1.

Table 1. Serval examples of classification of spatial and temporal relevance.

Sentence	Triple	Spatio-Temporal Quadruple	Spatio- and Temporal Relevance
Brown previously received degrees of Ph.D.	(Brown, received, degrees)	(Brown, received, degrees, previously)	Temporal weak relevance
Brown was elected to presidency of Cincinnati Wesleyan College.	(Brown, elected, presidency)	(Brown, elected, presidency, Cincinnati Wesleyan College)	Spatio-weak relevance
Brown went to New York City In 1877.	(Brown, went to, New York City)	(Brown, went to, New York City, 1877)	Temporal strong relevance
She accepted offer.	(She, accepted, offer)	-	Spatio-temporal no relevance
Brown was present in May 1869 at caucus for national organization of temperance party.	(Brown, caucus, national organization of temperance party)	(Brown, caucus, national organization of temperance party, in May 1869)	Temporal moderate relevance

Text classification, also known as automatic text classification, is the process by which a computer maps a text containing information to a pre-given category or categories of topics [11]. There are generally rule-based methods, machine learning-based methods, and deep learning-based methods [12–14]. Rule-based methods refer to the classification of text into different categories using a set of predefined rules and require in-depth domain knowledge [15,16]. Hua et al. [17] took abstracts of scientific and technical papers as the object of study and classified abstract sentences into four types according to the contextual information of the sentences, namely background knowledge, topic, research method, and experimental results, and completed the initial classification of abstract sentences with some heuristic rules; Asghar et al. [18] proposed a rule-based framework for sentiment-based sentiment classification of sentences, combining cognitive sentiment theory and computer techniques based on a sentiment analysis to detect and classify sentiment from natural language texts; Tan et al. [19] described a rule-based sentiment analysis approach for polarity classification of fused news articles. However, rule-based algorithms, although they can build classifiers quickly, require a lot of human involvement and are not effective for long classified texts. Machine learning-based methods have emerged that can learn rules automatically, usually with algorithms such as plain Bayes, support vector machines, and KNN algorithms based on statistical models [20]. Regarding their ability to automate the learning of features and rules, however, traditional machine learning models have a simpler structure, rely more on manually acquired text features, and have relatively small parameters that make the models difficult to apply to sufficiently large datasets.

Deep learning-based approaches have emerged to automatically acquire basic features and combine them into advanced ones, in addition to deep learning to make better use of features in word order. For example, Sochar et al. [21] proposed a method combining recurrent neural networks and autoencoders to form an unsupervised sentence transformation model and applied it to the prediction of sentiment label classification. Iyyer et al. [22] proposed an MLP-based deep averaging network, DAN, which classifies the output by computing the average of the embedding vectors of all words that are fed into the MLP network. Kim [23] first proposed the use of a CNN classification model for text classification, which is applicable to most text classification tasks compared to traditional methods. Johnson et al. [24] proposed a deep pyramidal CNN, which has lower complexity and better classification results. Graves [25] first used a BiLSTM model for text classification tasks and achieved its purpose. Liu et al. [26] combined a BERT pre-training model with multi-task learning to obtain better results. Qin et al. [27] proposed a BERT-CNN model combining a BERT model and a CNN model for Word2Vec, GloVe, etc., which could not fully extract semantic information. Chung et al. [28] compared LSTM and gated recurrent units for a text classification task, and experimental results showed that LSTM and gated recurrent units outperformed RNNs. In summary, deep learning-based text classification models have emerged to automatically learn the dependencies between data and labels. Devlin et al. [29] used the BERT model to perform bidirectional learning and processing of text using the encoder part of the transformer structure. They predicted word vectors by masking target words and learned the relationships between words using a self-attention mechanism, thus better incorporating sentence-level semantic information. Lai et al. [30] proposed a recurrent convolutional neural network classification method, which maximizes the capture of contextual information by drawing on the common advantages of RNN and CNN, greatly improves the accuracy of the classification ground, and makes the classification of the best; Orosoo et al. [31] proposed a CNN based on a typical English text where it is easy to feature fuzzy elements and other shortcomings, to improve the classification accuracy and adaptability; Kong et al. [32], based on the data of 12,345 government hotlines in Zhejiang Province from 2017 to 2021, constructed a fine-grained three-level classification system of livelihood issues from the perspective of residents, and utilized the BERT pre-training model to construct a text classification model, which transformed the text of the residents' demands into the labels of livelihood issues; Wang et al. [33] proposed a Chinese short-text classification model based on the ERNIE-RCNN model for the difficulties of few feature words, poor normality, and a large amount of the data size in Chinese short text. Li et al. [34] proposed a microblog rumor detection model based on the BERT-RCNN model in response to the traditional rumor detection model that requires a large number of features as well as the difficulty in achieving timely detection, and verified the effectiveness of the model.

Geographic knowledge refers to the cognitive results of human understanding of the spatial distribution, evolutionary process, and interaction of geographical phenomena or things [35]. In recent years, scholars have carried out a lot of research on the expression of geographical knowledge, and propose a model of spatio-temporal knowledge expression with the uniqueness of time and space, but the correlation between geographical knowledge and spatio-temporal information is not considered, and the adaptive correlation and expression of geographical knowledge and spatio-temporal information cannot be realized [36–38]. To address the aforementioned issues and facilitate the integration of geographic knowledge triples with time, space, and spatio-temporal knowledge, this paper proposes a novel solution, which involves converting the spatio-temporal association degree to incorporate time and space information into the general triple. This process results in the formation of a quadruplet that effectively expresses temporal or spatial information. Finally, a text classification model is employed to map the time and space information to the triple. The BERT_RCNN spatio-temporal text classification model can accurately express semantic information to improve retrieval precision.

The main contributions of this study are summarized as follows:

(1) This paper proposes a deep learning-based BERT_RCNN spatio-temporal text classification model and compares it with mainstream deep learning-based text classification models. Experimental results show that using the BERT_RCNN model significantly improves the performance of spatio-temporal text classification.
(2) Using the Wikipedia English corpus as the data source, this paper screens a series of geographically relevant texts, subsumes and extracts triples, and annotates the dataset according to the knowledge types and annotation rules, and finally constructs a spatio-temporal classification dataset.

The remaining sections of this article are organized as follows: Section 2 introduces the process of dataset construction and introduces commonly utilized methods for text categorization; Section 3 introduces the experimental parameters and analyses the experimental results; Section 4 analyses the text classification model from the perspective of deep learning principles; and finally, Section 5 presents the conclusions drawn and future research work.

2. Materials and Methods

This paper focuses on a deep learning-based spatio-temporal text classification model approach (see Figure 1). The method consists of three main steps: dataset construction, the spatio-temporal text classification model based on deep learning, and the experimental validation and analysis. The input is an annotated Wikipedia English dataset, and the final text classification model accurately maps temporal and spatial information to triples to accurately represent semantic information.

2.1. Dataset Construction

2.1.1. Data Acquisition

In this paper, we employ the English corpus zhwiki-latest-pages-articles.xml.gz (https://dumps.wikimedia.org/enwiki/ accessed on 29 February 2023) acquired from Wikipedia as our primary data source. This corpus encompasses knowledge from various domains, including sports, history, and individuals, thereby providing an extensive and comprehensive information repository on diverse topics. Numerous tools are readily available for directly extracting the corpus. Initially, we utilize the WikiExtractor.py tool to extract English article passages from the corpus. Subsequently, we apply a filtering process to select passages that possess geographically relevant content. Next, we perform sentence separation to ensure that each line within the resulting text files represents a distinct English utterance. With the .txt files obtained, we make use of the Stanford OpenIE tool to extract entity–relationship triples from individual sentences. We evaluate the spatio-temporal relevance of these triples and subsequently generate a text categorization dataset based on their spatio-temporal relevance, as depicted in Figure 2.

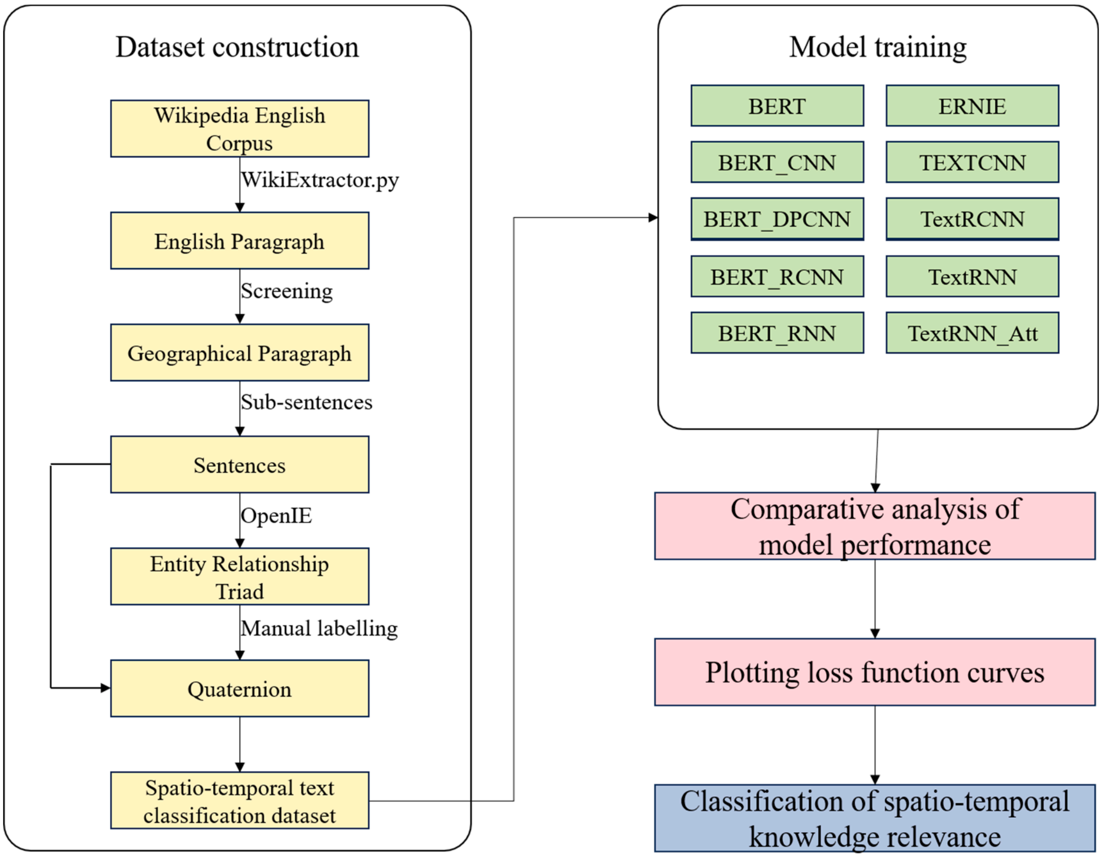

Figure 1. The overall presented framework, which is divided into three main modules: dataset construction, model training, and experimental validation. The yellow part is the dataset construction, the green part indicates the model used in this experiment, and the pink part indicates the experimental validation phase.

```
She arrived 22 December 0
unit has performed at Cherry Blossom Festival in Washington 2
Haiti   declared its independence on New Year's Day 1804    0
committee met again in Berlin   2
He came to New York 2
She recommissioned Lt. on 29 November   0
the Kingdom of England was formed in the early ninth century    1
survey will will analogous to Sloan Digital Sky Survey of Northern hemisphere sky   1
the Western Jin Dynasty ended in 316    1
she returned in 2011    0
it was reported On 14 March 2013    0
The period 1861-1865 was the American Civil War 1
Oleksandr Omelchenko mayor from 1999 to 2006    0
He returned to Chelsea  1
The reign of Zhenguan lasted from 627 to 649    0
The region with the largest oil reserves, production and output in the world is the Middle East 1
winner receives coveted title   0
scheduled airline service is available to Portland  1
Union Rule was trivial fraction 0
the English Navy defeated the Spanish Armada in 1588    1
Brown served as vice-president  0
South America is separated by the Drake Passage Antarctica  3
Delegates form International Yacht Racing Union 0
The second most populous country in the world is India  2
The southeast coast of Brazil is the most densely populated 3
Calhoun served team for 44 years    0
teams and players were constant changing during the 1920s   0
Borough was renamed to Redcar in 1996   0
The early Jurassic period is about 208 million years old    0
The Renaissance movement in Europe lasted from the 14th to the 16th centuries   1
```

Figure 2. Some samples of the spatio-temporal correlation text classification dataset.

Furthermore, the classification of spatio-temporal relevance is divided into four main categories: temporal strong relevance, temporal moderate relevance, temporal weak relevance, and no relevance. Similarly, spatial relevance is categorized into spatial strong relevance, spatial moderate relevance, spatial weak relevance, and no relevance. In addition, spatio-temporal relevance is classified into spatio-temporal strong relevance, spatio-temporal moderate relevance, spatio-temporal weak relevance, and no relevance. To evaluate the degree of association, we perform a statistical analysis on the dataset, and the corresponding statistical results are presented in Table 2.

Table 2. Distribution of spatio-temporal correlation data.

Type of Association	Number	Proportions	Label	Example
Temporal strong relevance	869	33%	3	He was rector in 1989–2002.
Temporal moderate relevance	424	16%	2	University named alumnus of year in 2003.
Temporal weak relevance	126	5%	1	CSM joined Federation of European Mineral Programs in 2003.
Temporal no relevance	1215	46%	0	CSM joined Federation of European Mineral Programs.
Spatial strong relevance	242	9%	3	Switzerland is bordered by France to the west.
Spatial moderate relevance	363	14%	2	Pope Paul VI separated territory from Lansing Diocese.
Spatial weak relevance	895	34%	1	Candidates completing programmers in Camborne School Mines.
Spatial no relevance	1134	43%	0	Selection is based on achievement.
Spatio-temporal strong relevance	362	14%	3	He was from 1990 to 1996 chair of board of governors of University of Calgary.
Spatio-temporal moderate relevance	794	30%	2	October 26 has been designated as the National Olympic Day.
Spatio-temporal weak relevance	1183	45%	1	It signed a contract with Hungary.
Spatio-temporal no relevance	295	11%	0	Countries raced yachts Prior to ratification in 1907.

2.1.2. Spatio-Temporal Correlation Classification Labeling Rules

According to the process of knowledge development, this paper divides knowledge into factual knowledge, conceptual knowledge, and rule-based knowledge [4]. Factual knowledge refers to the factual laws derived from the events that have already happened or have happened many times before after being summarized by experts and scholars, which mainly include entity facts, phenomenon facts, and event facts [39]; conceptual knowledge refers to a more abstract and general, organized knowledge [40,41]; rule-based knowledge refers to most mathematical formulas and other knowledge [29,42,43]; the types of knowledge are shown in Table 3.

Based on the above classification of a knowledge base, we defined the relevance type. Temporal relevance is categorized into temporal strong relevance, temporal moderate relevance, temporal weak relevance, and temporal no relevance. Spatial relevance is classified as spatial strong relevance, spatial moderate relevance, spatial weak relevance, and spatial no relevance. Spatio-temporal relevance is divided into spatio-temporal strong relevance, spatio-temporal moderate relevance, spatio-temporal weak relevance, and spatio-temporal no relevance.

Table 3. Classification of knowledge types.

Type of Knowledge	Example
Factual knowledge [4]	(1) The first mass extinction of species occurred at the end of the Ordovician period, 440 million years ago. (2) 210 years ago, Qin Shi Huang's imperial tours died in Xingtai dunes on his way to the east.
Conceptual knowledge [4]	(1) The Ordovician Period is the second period of the Paleozoic, which began in 500 million BC and lasted for 65 million years. (2) Sandstorm is a general term for both sandstorms and dust storms. It refers to the severe sandstorm weather phenomenon where strong winds blow up a large amount of sand and dust on the ground, causing the air to be particularly turbid and the horizontal visibility to be less than 1 km.
Rule-based knowledge [4]	Quartz sandstone: feldspar content < 10%, rock debris content < 10%.

In the process of dataset annotation, corresponding annotation specifications were developed according to the dependent syntax of the sentences and the type of knowledge, and the temporal and spatial relevance of the sentences were annotated according to the above classification in three aspects: temporal, spatial, and spatio-temporal, as shown in Tables 4 and 5.

Table 4. Temporal correlation text classification annotation rules.

Type of Knowledge	Marking Specification	Example
Factual knowledge	Explicit temporal information generally acts as a gerund and acts as an entity when the subject of an event occurs only at a certain time, and when the time changes, the subject of the event is affected and thus changes, considering the time to be strongly associated.	Brown went to New York City In 1877.
	When an event does not change with the event, time is considered unrelated.	She withdrew her name.
	A sentence is considered to be temporal weak relevance when there is a time in it, but not a subject event that depends on the time, but a local subject event or a secondary event that depends on that time.	She shortly thereafter was called to headship of order in State of Ohio.
	• A sentence can be extracted to obtain a triple, but if there is a time in a sentence and this time is information about an attribute of an entity, it can be judged as a moderate association of time. • After excluding temporal strong relevance, temporal weak relevance, and temporal no relevance, the other is temporal moderate relevance.	Brown was elected Grand Vice-Templar after founding in August 1874.
Conceptual knowledge	Judging from the semantic information, the emphasis on time can be judged as a strong association with time.	Pomegranate trees usually bloom in May.
	• When explaining abstract conceptual knowledge, the link between time and space can be judged to be temporal no relevance. • Judgement based on semantic information, and the sentence does not emphasize time, can be judged as temporal no relevance.	FBI investigates crime in most cases.
	The main meaning expressed in the sentence is not related to time, it only modifies small fragments of knowledge in the main triad; then, it is judged to be weakly associated with time.	A region in the northwest experiences dust storms throughout the year.

Table 4. *Cont.*

Type of Knowledge	Marking Specification	Example
Rule-based knowledge	If the time information does not modify the main clause in the triple, but only the minor clauses in entity 1 or entity 2, and does not modify the content of the main clause, the highest degree of association is judged to be moderate.	Their tribe had expanded into what is now France, Belgium, northern Italy, Spain, and the vast Rhine Valley.
	For simple structural triples that exist in regular triples and are essentially unrelated to time.	International rule was created for the measuring and rating of yachts.

Table 5. Spatial relevance text classification annotation rules.

Type of Knowledge	Marking Specification	Example
Factual knowledge	An event space is considered strongly associated when it can only be performed in a specific space and not in other spaces.	Pomegranates are suitable for cultivation on both sides of the three river basins.
	When all events expressed in a triad are unaffected by space, the knowledge is judged to be unrelated to space.	The Mohe gold mine is mentioned in the historical records of the Qing Dynasty.
	When spatial terms appear in a sentence but the occurrence of an event is less dependent on space, the knowledge is judged to be unrelated to space.	He was influenced by Columba Marmion.
	After excluding strong spatial association, weak spatial association, and no spatial association, the other is moderate spatial association.	By the second century BC their tribe had expanded into what is now France, Belgium, northern Italy, Spain, and the vast Rhine Valley.
Conceptual knowledge	Explanation of abstract concepts, disconnected from space, generally unconnected to space.	The Confederacy dissolved In 1866.
	Conceptual knowledge that explains entities and phenomena in which geospatial emphasis is placed and therefore can be judged to be strongly associated with space.	Pomegranates are suitable for cultivation on both sides of the three river basins.
	The main meaning expressed in the sentence is not spatially relevant and it only modifies small fragments of knowledge in the main triad; then, it is judged to be spatially weakly relevant.	It formed the Austro-Hungarian Empire In 1867.
Rule-based knowledge	If the spatial information does not modify the main sentence in the triple, but only the secondary sentence in entity 1 or entity 2, and does not modify the content of the main sentence, the highest degree of association is judged to be moderate.	Chiasso is located in the state of Ticino.
	For simple structural triples that exist in regular triples and are essentially unrelated to time.	The sum of the interior angles of a triangle is 180.

In addition to the cases summarized above, there are still some triples (entity$_1$, relationship, entity$_2$) with a slightly more complex structure, where the spatio-temporal related information does not modify the main sentence in the triple, but acts on the smaller sentences in entity$_1$ and entity$_2$. As long as it does not modify the content of the main sentence, the highest degree of association in this case is judged as moderate in this paper.

In addition, the annotated dataset in this paper tries to ensure that 80% of the data are meaningful data. Triples such as real people's life trivia are not meaningful triples in this paper and are discarded in their entirety in the annotated dataset in this paper. For example, the sentence "*Small trees like to grow in a warm and well-lit environment*" has a strong spatial association, and the sentence needs to be judged by considering the semantic and spatial association of the triple. Ultimately, the temporal and spatial correlations are labelled by combining the classification criteria of temporal and spatial correlations.

2.2. Sentence Classification Methods

The current mainstream deep learning-based text classification models are the FastText model [44], TextCNN model [23], TextRNN model [45], TextRCNN model [30], TextRNN_Att model [46], BERT model [47], ERNIE model [48], BERT_CNN model [49], BERT_RNN model [50], BERT_DPCNN model [51], and BERT_ RCNN model [34]. Since the FastText model is an open-source word vector computation and text classification model with its structure and its simplicity, in this paper, we choose the remaining ten deep learning-based text classification models for comparison experiments.

TextCNN model: It comprises a word embedding layer, a multi-branch convolutional layer, a multi-branch global maximum pooling layer, and a fully connected classification layer. Unlike the FastText network model, which disregards word order information, the TextCNN model employs CNN to extract n-gram-like key information from sentences. This enables it to accurately capture locally relevant information.

TextRNN model: While the TextCNN model performs well in text classification tasks, it has limitations regarding longer sequence information modeling, and adjusting the filter_size hyperparameter can be cumbersome. On the other hand, the TextRNN model excels in expressing contextual information.

TextRNN_Att model: Both CNN and RNN are effective in text categorization tasks, but they lack interpretability. The attention mechanism, commonly used in natural language processing, addresses this limitation by providing a clear understanding of how each word contributes to the overall result, thus enhancing long-term memory modeling.

TextRCNN model: The TextRCNN model leverages forward and backward RNNs to obtain contextual representations of each word. It connects the word vectors with the forward and backward context vectors, incorporating a convolutional layer and pooling layer similar to TextCNN. The only difference is that the filter_size of the convolutional layer is set to one.

BERT model: The BERT model primarily employs an encoder in a Transformer for encoding. In comparison to the aforementioned models, BERT utilizes masked language modeling and next-sentence prediction to more effectively capture word-level, sentence-level, and relationship features between sentences.

ERNIE model: ERNIE employs multiple T-encoders, similar to BERT, which are trained with input token embeddings to capture information from text sequences. Additionally, multiple K-encoders are used to "splice" entity embedding inputs from the text with the T-encoder inputs. The ERNIE model further incorporates the knowledge embedding model TransE and multi-attention mechanism to extract information. Like BERT, the ERNIE model is utilized for text categorization. Input sentences are segmented into words, and the output corresponding to [CLS] is employed for categorization.

BERT_CNN model: While the BERT model provides a straightforward interface for downstream tasks and can be directly used for text categorization, its application as a document vector has been relatively unexplored. Moreover, solely using the CNN model does not adequately address the problem of semantic polysemy. To address these issues, the BERT_CNN model combines BERT and CNN. BERT is fed into the CNN model as an embedding layer, enabling effective text classification.

BERT_RNN model: The BERT_RNN model consists of three main parts. Firstly, it obtains the semantic representation of each text by training the BERT model to obtain vector representations of words. Secondly, the word vectors are inputted into an RNN model to further analyze and extract semantics. Finally, the obtained word vectors are connected to a softmax layer for text classification.

BERT_DPCNN model: The DPCNN compresses words into a low-dimensional semantic space, which can result in words with similar meanings having the same word vector. To address this issue, BERT generates word vectors more efficiently, and the use of a Transformer enables the mining of context-based information and long-distance dependencies. Additionally, the DPCNN model can extract long-distance textual dependencies by deepening the network continuously.

BERT_RCNN model: The BERT_RCNN model takes sentences as input to the BERT model for training and transforms the text into word vectors. The output of the last layer of BERT training, combined with the weights, is then fed into the RCNN model. In the RCNN model, deeper semantic features of geographic text are extracted to obtain potential semantic vectors. These results are then inputted into a pooling layer to obtain the most representative key features using max-pooling. Finally, the output of the fully connected layer is passed through a softmax function for text classification. The main model structure is depicted in Figure 3.

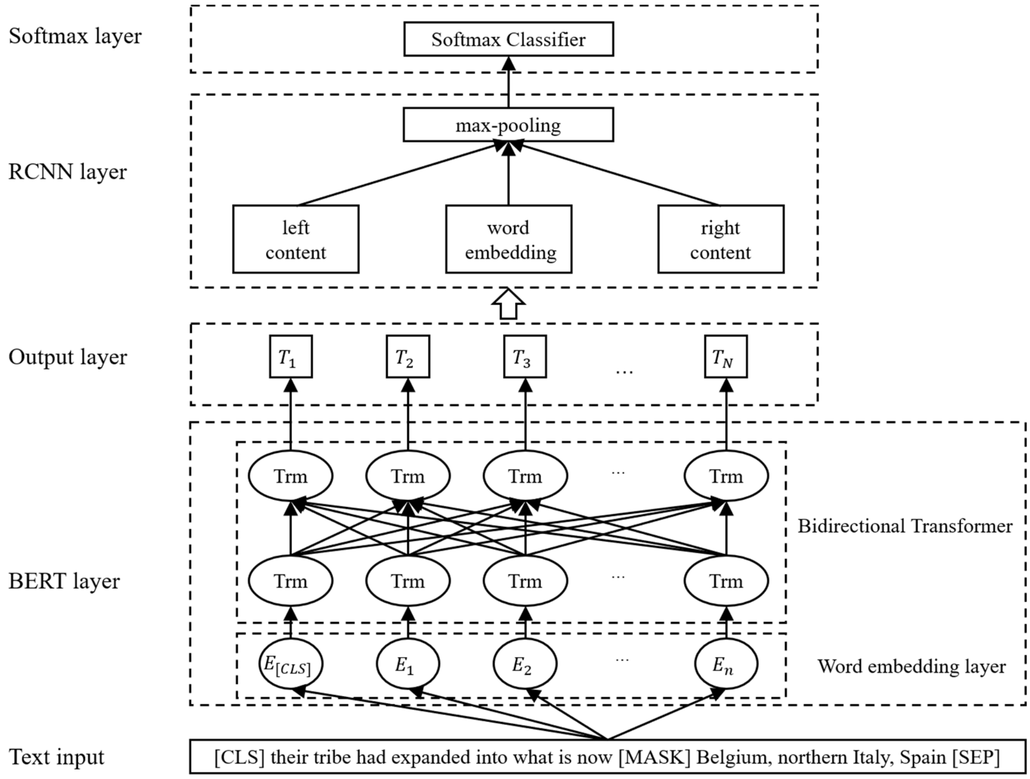

Figure 3. Structure of BERT_RCNN text categorization model.

The RCNN model in the BERT_RCNN model replaces the convolutional layer of traditional CNNs with a recurrent structure using inner recurrent convolutional layers. This establishes a deep network structure based on feed-forward connections. The structure of the recurrent convolutional layer is illustrated in the following Figure 4. The introduction of the RCNN model in the BERT_RCNN model allows for more even acquisition of contextual information from words, overcoming the drawback of recurrent neural networks where later words have more influence than previous words. Additionally, it avoids the need to set the degree of context dependence through the window size, as required in CNNs.

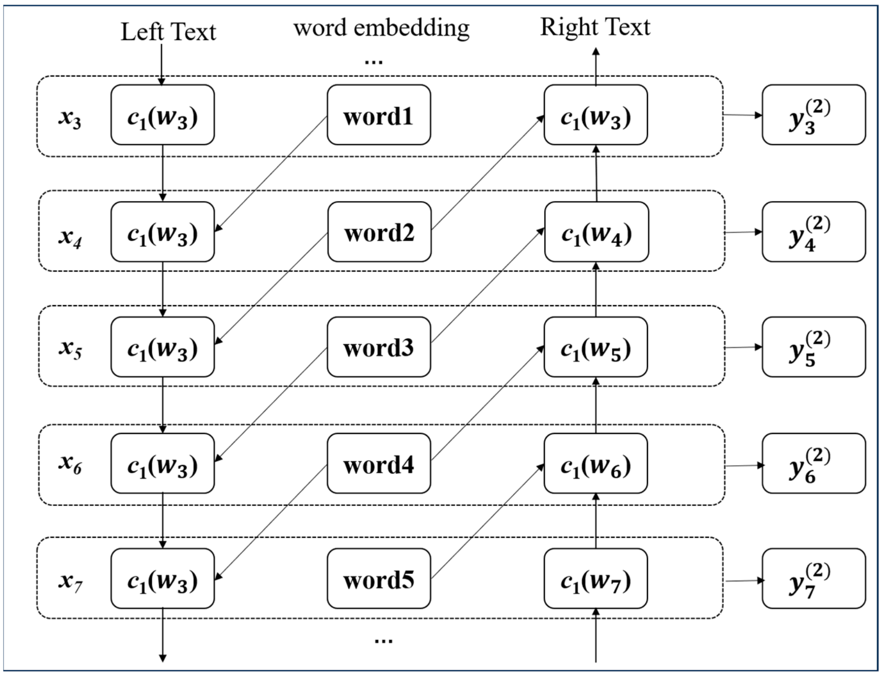

Figure 4. Structure of circular convolutional layer.

3. Experiments and Results

3.1. Parameter Setting

All experiments in this study are based on NVIDIA GTX 2080TI GPU and the Linux operating system. A variety of different experiments were set up to verify the effectiveness of temporal, spatial, and spatio-temporal text classification, respectively. In this experiment, to ensure the precision of the classification model, certain guidelines should be followed to reduce the training time and improve the precision of the model training, so an algorithm with a learning rate of 5×10^{-5} was used for training. In this experiment, the Adam optimizer was chosen as it is widely regarded as an effective optimizer for optimizing the final loss function [52,53]. Furthermore, since this experiment involves a multiclassification problem, the loss function selected is Categorical_Crossentropy. The experimental parameters are shown in Table 6.

Table 6. Experimental parameter settings.

Batch_Size	Epoch	Optimizer	Learn Rate	Loss
32	10	Adam	5×10^{-5}	Categorical_crossentropy

We employed a 10-fold cross-validation methodology to assess the performance of our model. The dataset was randomly partitioned into 10 mutually exclusive subsets, with each subset serving as the testing set once while the remaining nine subsets were used for training. The reported performance metric was the average score derived from these 100 runs, providing a comprehensive evaluation of the model's robustness and generalization ability.

3.2. Evaluation Metric

In this experiment, the precision rate (P), recall rate (R), and F1 score (F1) are used as the evaluation metrics. The precision rate is the proportion of correct predictions in the spatio-temporal text classification to the number of predicted positive cases; the recall rate is the proportion of predicted correct cases to the number of actual positive cases; and the F1 score is the summed average of the precision rate and recall rate. The F1 score takes into account both precision and recall to achieve the highest balance between the two, and is calculated as shown.

$$P = \frac{TP}{TP + FP} \tag{1}$$

$$R = \frac{TP}{TP + FN} \tag{2}$$

$$F1 = \frac{2 * P * R}{P + R} \tag{3}$$

where TP is the number of correct matches and FP is the number of incorrect matches; FN is the number of correct matches not found; and TN is the number of correct non-matches.

3.3. Temporal Correlation Classification

3.3.1. Comparative Analysis of Sentence Classification Models

For temporal classification, we chose ten current mainstream text classification models for our experiments, and the experimental results are shown in Table 7. As shown in Table 7, it can be seen that for temporal no relevance, the BERT_CNN model achieves the highest precision, recall, and F1 values of 0.98, 0.98, and 0.97, respectively; however, for temporal weak relevance, temporal moderate relevance, and temporal strong relevance, the BERT_RCNN model is more stable, where in temporal weak relevance classification, its precision, recall, and F1 values are 0.89, 0.91, and 0.78, respectively; in the temporal moderate relevance classification, its precision, recall, and F1 values are 0.70, 0.79, and 0.71, respectively; in the temporal strong relevance classification, its precision, recall, and F1 values are 0.88, 0.82, and 0.85, respectively. In summary, the BERT_RCNN model is more effective in classifying temporal relevance, mainly because the semantic representation of each text can be obtained using the BERT model to obtain a vector representation of the word; secondly, compared to the traditional CNN and RNN models, the RCNN uses a cyclic recurrent neural network to reduce noise and uses a maximum pooling layer to effectively capture the most critical features in a sentence.

Table 7. The experimental results for temporal text classification.

Model	**No Relevance**			**Weak Relevance**			**Moderate Relevance**			**Strong Relevance**		
	P	**R**	**F1**	**P**	**R**	**F1**	**P**	**R**	**F1**	**P**	**R**	**F1**
BERT	0.98	0.95	0.96	0.20	0.20	0.20	0.52	0.62	0.57	0.81	0.77	0.79
BERT_CNN	**0.98**	**0.98**	**0.97**	0.42	0.50	0.45	0.69	0.69	0.68	0.85	0.85	0.85
ERNIE	0.49	0.53	0.64	0.56	0.42	0.48	0.52	0.57	0.62	0.44	0.39	0.55
BERT_DPCNN	0.95	0.97	0.97	0.68	0.71	0.42	0.55	0.79	0.64	0.85	0.79	0.64
BERT_RCNN	0.98	0.97	0.96	**0.89**	**0.91**	**0.78**	**0.70**	**0.79**	**0.71**	**0.88**	**0.82**	**0.85**
BERT_RNN	0.94	0.98	0.97	0.68	0.79	0.76	0.56	0.39	0.43	0.62	0.67	0.76
TextCNN	0.79	0.93	0.85	0.67	0.21	0.31	0.41	0.21	0.28	0.73	0.77	0.75
TextRNN	0.63	0.88	0.73	0.41	0.52	0.66	0.20	0.23	0.42	0.62	0.59	0.61
TextRCNN	0.80	0.87	0.83	0.40	0.20	0.27	0.50	0.21	0.30	0.60	0.73	0.66
TextRNN_Att	0.77	0.91	0.83	0.67	0.20	0.31	0.52	0.29	0.37	0.73	0.76	0.74

3.3.2. Loss Function Analysis Based on Different Classification Models

The loss function is a measure of how similar the predicted values are to the actual values. In order to observe the change of the loss function during model training, we plotted the loss function curves of these ten temporal text classification models, as shown in Figure 5. As shown in Figure 5, we can see that the loss function curves are decreasing sharply at the beginning of the model training phase, and eventually level off. The loss function curves of the four models, BERT, BERT_RCNN, BERT_CNN, and BERT_DPCNN, eventually converge to 0. The convergence of the BERT_RCNN model is the best, while the loss functions of the remaining six models are still high. This is mainly due to the fact that in the process of conducting the experiment, the training set, test set, and validation set were generated in a random way, and there may be problems such as unbalanced datasets in the random generation process; secondly, there may be problems such as overfitting in the training process, so the loss function convergence effect is not satisfactory.

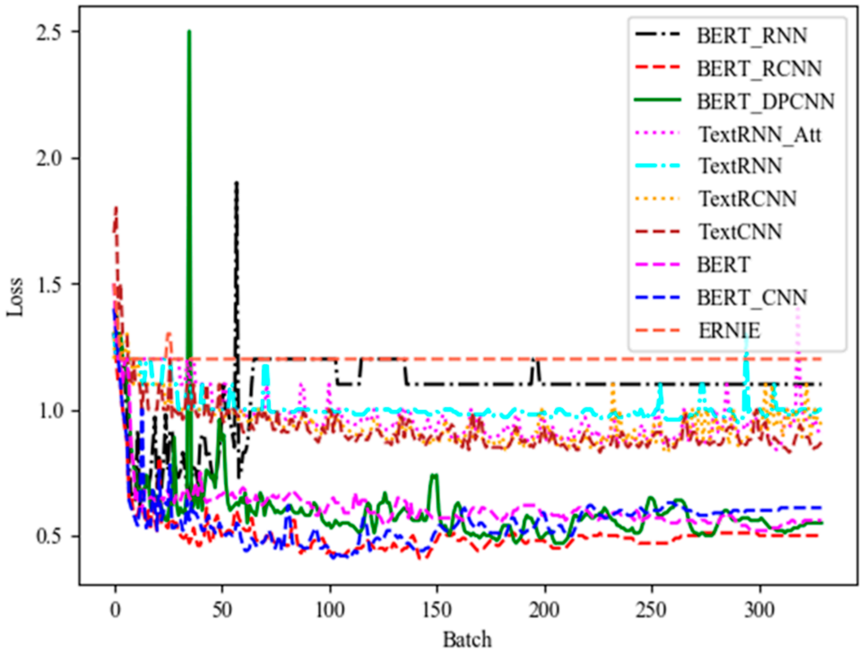

Figure 5. Loss function graph for temporal text classification.

3.4. Spatial Correlation Classification

3.4.1. Comparative Analysis of Sentence Classification Models

For spatial classification, ten current mainstream text classification models were selected for this experiment, and the experimental results are shown in Table 8. As shown in Table 8, it can be seen that the precision, recall, and F1 values of spatial classification when using the BERT model are generally higher than those of other text classification results without the BERT model. Similar to the temporal classification results, the BERT_CNN model achieved the highest precision and recall in the spatial no relevance classification results, with 0.94 and 0.93, respectively. However, for the spatial weak relevance, spatial moderate relevance, and spatial strong relevance classification results, the BERT_RCNN model achieved better results. Among them, its precision, recall, and F1 values reached 0.75, 0.69, and 0.71, respectively, in the spatial weak relevance classification results, which were more than 4% better than other classification models; its precision, recall, and F1 values were 0.57, 0.77, and 0.65, respectively, in the spatial moderate relevance classification results; in the spatial strong relevance classification results, its precision, recall, and F1 values were 0.87, 0.88, and 0.90, respectively, which are more than 2% higher than those of the BERT_CNN model in spatial strong association classification. In summary, the BERT_RCNN model significantly improved its performance in spatial classification.

Table 8. The experimental results for spatial text classification.

Model	No Relevance			Weak Relevance			Moderate Relevance			Strong Relevance		
	P	R	F1	P	R	F1	P	R	F1	P	R	F1
BERT	0.92	0.80	0.86	0.67	0.76	0.71	0.60	0.54	0.57	0.74	0.88	0.80
BERT_CNN	**0.94**	**0.93**	0.89	0.42	0.50	0.45	0.69	0.69	0.69	0.85	0.85	0.86
ERNIE	0.83	0.84	0.84	0.72	0.69	0.70	0.51	0.54	0.53	0.75	0.75	0.75
BERT_DPCNN	0.88	0.85	0.87	0.69	0.73	0.71	0.38	0.67	0.48	0.42	0.38	0.47
BERT_RCNN	0.91	0.88	**0.89**	**0.75**	**0.69**	**0.71**	**0.57**	**0.77**	**0.65**	**0.87**	**0.88**	**0.90**
BERT_RNN	0.89	0.84	0.87	0.68	0.69	0.689	0.56	0.49	0.53	0.70	0.91	0.79
TextCNN	0.75	0.78	0.77	0.63	0.67	0.65	0.41	0.44	0.43	0.76	0.50	0.60
TextRNN	0.61	0.70	0.65	0.45	0.52	0.48	0.48	0.36	0.41	0.50	0.22	0.30
TextRCNN	0.75	0.68	0.71	0.56	0.71	0.62	0.53	0.41	0.46	0.70	0.59	0.64
TextRNN_Att	0.73	0.70	0.72	0.53	0.72	0.61	0.55	0.28	0.37	0.64	0.50	0.56

3.4.2. Loss Function Analysis Based on Different Classification Models

For the spatial text classification experiments, we also plotted its loss function, as shown in Figure 6. It can be seen that at the beginning of the model, its loss function decreases rapidly, but as the amount of data increases and the number of training sessions increases, the loss value tends to level off. The BERT_CNN model has the best convergence, followed by the BERT_RCNN model. However, as shown in Figure 6, it is evident that the TextRNN and BERT_DPCNN models exhibit the poorest loss curves. The main reason for this analysis is that the process of randomly assigning the dataset leads to an unbalanced data distribution; in addition, the noise present in the dataset can be learned during the training of the above two models, so the experimental results are not satisfactory. The results of the combined graph and spatial text classification clearly demonstrate that our selected model, BERT_RCNN, achieves the best experimental outcomes and exhibits the most consistent performance.

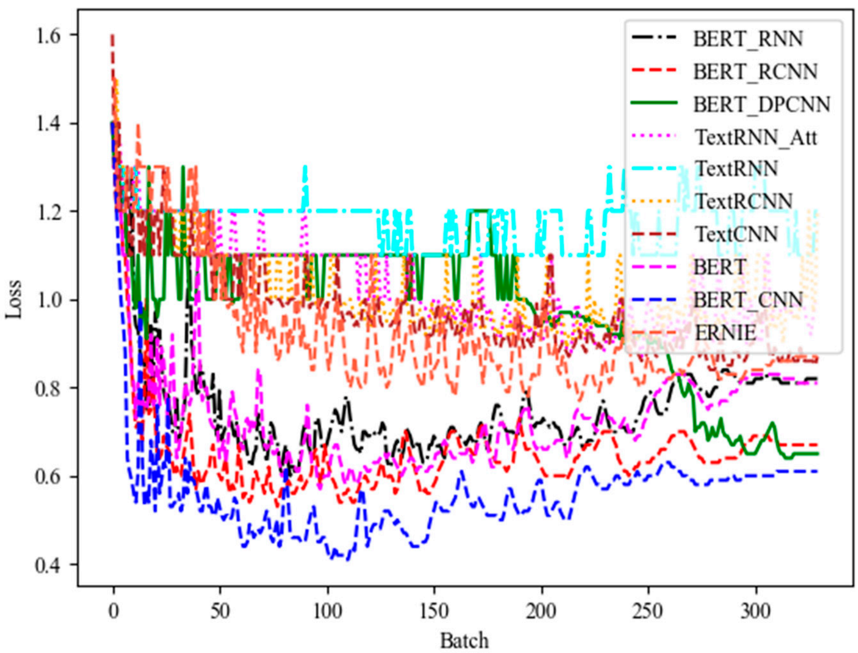

Figure 6. Spatial text classification loss function graph.

3.5. Classification of Spatial and Temporal Relevance

3.5.1. Comparative Analysis of Sentence Classification Models

For the spatio-temporal integrated classification, ten mainstream deep learning text classification models were selected for this experiment, and the experimental results are shown in Table 9. As shown in Table 9, the BERT_RCNN model achieves the highest precision, recall, and F1 values for the spatio-temporal integrated classification results in the uncorrelated, weakly correlated, moderately correlated, and strongly correlated classification results. In the spatio-temporal no relevance classification results, the precision, recall, and F1 values of the BERT_RCNN model are 0.93, 0.73, and 0.81, respectively; in the spatio-temporal weak relevance classification results, the precision, recall, and F1 values of the BERT_RCNN model are 0.75, 0.87, and 0.80, respectively; in the spatio-temporal moderate relevance classification results, the precision, recall, and F1 values of the model are 0.70, 0.55, and 0.62, respectively, while the precision, recall, and F1 values of the model are 0.73 in the spatio-temporal strong relevance classification model for spatio-temporal classification. Secondly, compared to the RNN and CNN models, the RCNN model uses a bidirectional cyclic structure to obtain contextual information, which can preserve a large range of word order compared to CNN in learning text expressions, and secondly, it uses a maximum pooling layer to obtain important parts of the text and automatically discriminates which features play an important role in text classification.

Table 9. The experimental results of spatio-temporal text classification.

Model	No Relevance			Weak Relevance			Moderate Relevance			Strong Relevance		
	P	R	F1	P	R	F1	P	R	F1	P	R	F1
BERT	0.88	0.77	0.82	0.74	0.85	0.80	0.71	0.55	0.62	0.69	0.78	0.73
BERT_CNN	0.89	0.83	0.86	0.74	0.83	0.78	0.66	0.54	0.62	0.63	0.73	0.67
ERNIE	0.63	0.67	0.65	0.67	0.64	0.65	0.51	0.40	0.45	0.40	0.62	0.49
BERT_DPCNN	0.56	0.58	0.66	0.55	0.78	0.65	0.42	0.51	0.46	0.51	0.44	0.62
BERT_RCNN	**0.93**	**0.73**	**0.81**	**0.75**	**0.87**	**0.80**	**0.70**	**0.55**	**0.62**	**0.73**	**0.73**	**0.73**
BERT_RNN	0.91	0.73	0.81	0.73	0.84	0.78	0.68	0.56	0.59	0.57	0.70	0.64
TextCNN	1	0.33	0.50	0.56	0.82	0.67	0.55	0.40	0.46	0.69	0.49	0.57
TextRNN	0.46	0.20	0.28	0.50	0.65	0.57	0.40	0.41	0.40	0.44	0.22	0.29
TextRCNN	0.79	0.50	0.61	0.60	0.80	0.69	0.57	0.46	0.51	0.72	0.49	0.58
TextRNN_Att	0.72	0.43	0.54	0.58	0.72	0.64	0.51	0.43	0.46	0.59	0.54	0.56

3.5.2. Loss Function Analysis Based on Different Classification Models

In the training process of the spatio-temporal classification model, we paid close attention to the changes in the loss function (see Figure 7). We plotted the loss function curve. Initially, the loss function value dropped rapidly at the beginning of the model, but as the amount of data and the number of training iterations increased, the loss function eventually stabilized. Among the models, TextRNN_Att, TextRCNN, and TextCNN had slower convergence speeds. On the other hand, BERT_RCNN exhibited a sharp decrease in the loss function at the beginning of training, and as the data volume and number of training iterations increased, the loss value gradually decreased and converged. Considering the experimental results of the aforementioned spatio-temporal classification models, the BERT_RCNN model demonstrated the best performance.

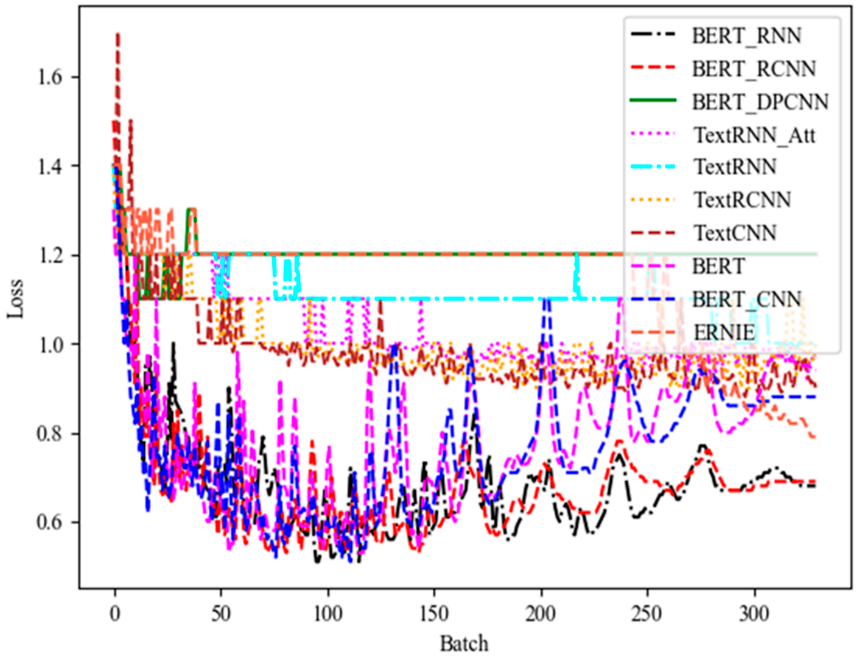

Figure 7. Spatio-temporal text classification loss function graph.

3.6. Text Categorization Prediction

In order to verify the accuracy of the presented model, we used real sentences for our experiments, where Table 10 shows the results obtained using the BERT_RCNN model for text categorization, showing examples of correctly or incorrectly categorized matches and mismatches. Although we only used a small number of examples for our experiments, we also performed a series of analyses on the results and found that although we can classify some sentences with spatio-temporal relevance using deep learning-based text classification, there are still mismatches. For example, the sentence "*Switzerland was conquered by Caesar's army In 58 BC*" contains both temporal and spatial words; moreover, the deep learning-based text categorization model has poor interpretability, and it is not possible to understand the specific role of the model in text categorization, and therefore it is not possible to accurately judge the temporal and spatial relevance of the sentence.

Table 10. Illustrative examples for the results with the method based on deep learning.

Type	Sentence	Real Classification Result	Predicted Classification Result
No correct classification	Switzerland was conquered by Caesar's army in 58 BC.	Temporal moderate relevance	Temporal weak relevance
Correct classification	Switzerland is located in central Europe.	Spatio-strong relevance	Spatio-strong relevance
Correct classification	It became an independent state in the mid-12th century under the Babenburg family.	Temporal strong relevance	Temporal strong relevance
No correct classification	Klingenbeck attended Allegany College of Maryland State University.	Spatio-moderate relevance	Spatio-weak relevance
Correct classification	Michael James Todd QPM was Chief Constable of Greater Manchester Police from October 2002 until his death.	Spatio-temporal strong relevance	Spatio-temporal strong relevance

4. Discussion

In this experiment, we used ten current mainstream deep learning models for spatio-temporal text classification, and through the above series of experimental validations and analyses, we can see that the performance of the models is BERT_RCNN > BERT_CNN > BERT > BERT_DPCNN > BERT_RNN > ERNIE > TextRCNN > TextRNN_Att > TextCNN > TextRNN.

From the above experimental results, we can find that the performance of the text classification model using the BERT model for preprocessing is generally better than the experimental performance without the BERT model, mainly because the BERT model has better semantic understanding ability, the BERT model can learn the contextual information in the text, so as to better understand the meaning of the text, thus improving the text classification precision. Secondly, the input to the BERT model can be a single sentence or two sentences, which enhances the robustness of the model and allows it to process multiple types of text; thirdly, the BERT-based text classification model can be adapted to different text classification tasks by fine-tuning the BERT model, thus having better generalization capabilities.

The RCNN model can extract features from the BERT input word vector, the convolutional layer uses convolutional kernels of different sizes to extract feature values of different sizes, and then the pooling layer is used for mapping processing, and finally the classification is performed with softmax, which can improve the performance of Chinese short-text classification. Firstly, the BERT pre-processed vector is used as the input to the CNN module, then the text features are captured by performing a convolution operation on the input matrix using a convolution kernel, which is a matrix that can be moved around and convolved with the input matrix; thirdly, a pooling operation can be used to reduce the dimensionality of the output matrix of the convolution layer, which slides over the matrix and selects the maximum or average value as the value of the new matrix; finally, the captured features are fed to the output layer for classification by concatenating the fully connected layers. However, compared with CNN and RNN, RCNN uses a bidirectional cyclic structure to obtain contextual information and can preserve word order to a large extent when learning textual expressions; thus, the BERT_RCNN model works best in this experiment.

5. Conclusions and Future Work

In this paper, we propose a spatio-temporal relevance classification method based on a deep learning method, which aims to establish the relevance between geospatial and spatio-temporal big data based on geographic knowledge with the support of unified spatio-temporal ontology, and accurately mapping time and space to the original triple, which can accurately realize the semantic expression and thus improve the retrieval precision. This paper provides a summary as follows: Firstly, the Wikipedia English corpus is utilized as the data source. Passages containing geographic knowledge are obtained through screening and undergo sentence processing before being written into a text file. Then, the OpenIE tool is employed to extract triples. Finally, temporal or spatial information is manually annotated and added to the triads, resulting in the creation of a text classification dataset with spatio-temporal correlation. Secondly, a deep learning-based BERT_RCNN text classification model is proposed. Compared with the current mainstream deep learning text classification models, the performance of the BERT_RCNN text classification proposed in this paper was significantly improved. This study can lay the foundation of deep learning methods for future spatio-temporal relevance classification of geographic knowledge, and we will continue to carry out in-depth work on geographic knowledge spatio-temporal knowledge mining and spatio-temporal knowledge graph construction in the next step.

While this paper examines the role and level of relevance of spatio-temporal information in various forms of geo-scientific knowledge, its interpretability is hindered by limited data and the absence of established conventions or citation descriptions for representing spatio-temporal features and relationships in the deep learning-based spatio-temporal relevance textual categorization approach. As a result, there is a lack of comprehensive documentation regarding the versions and variations of geo-scientific knowledge itself, which further compounds the issue. In response to the above shortcomings, we also have many ideas for future work. For example, we can try to analyze the characteristics of different types of knowledge to establish a more complete spatio-temporal feature and classification system; in addition, we can continue to expand the spatio-temporal relevance classification dataset to map spatio-temporal knowledge into entity–relationship triples more accurately, so as to provide important support for rapid retrieval of spatio-temporal knowledge maps, rule judgments and reasoning, and so on.

Author Contributions: Experiment proposal, Qinjun Qiu; funding acquisition, Qinjun Qiu, Kai Ma and Liufeng Tao; preliminary research, Miao Tian and Jiakai Huang; data collection, Miao Tian, Xinxin Hu, Haiyan Li and Shuai Zheng; experimental design and analysis, Miao Tian and Xinxin Hu; writing the original manuscript, Miao Tian; writing—review and editing, Miao Tian and Qinjun Qiu. All authors have read and agreed to the published version of the manuscript.

Funding: This work is supported by the Deep-time Digital Earth (DDE) Big Science Program. This study was financially supported by the National Key R&D Program of China (No. 2022YFB3904200), the Natural Science Foundation of China (No. 42301492),the Natural Science Foundation of Hubei Province of China (No. 2022CFB640), the Open Fund of Hubei Key Laboratory of Intelligent Vision Based Monitoring for Hydroelectric Engineering (No. 2022SDSJ04), the Open Fund of Key Laboratory of Urban Land Resources Monitoring and Simulation, Ministry of Natural Resources (No. KF-2022-07-014), and the China Postdoctoral Science Foundation (No. 2021M702991).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: This study was financially supported by the National Key Research and Development Program (No. 2022YFB3904200), the Open Fund of Key Laboratory of Urban Land Resources Monitoring and Simulation, Ministry of Natural Resources (No. KF-2022-07-014), the Opening Fund of Key Laboratory of Geological Survey and Evaluation of Ministry of Education (No. GLAB 2023ZR01), the Hubei Geological Bureau Science and Technology Project (KJ2023-12), the Fundamental Research Funds for the Central Universities, the Natural Science Foundation of Hubei Province of China (No. 2022CFB640), and the Open Fund of Hubei Key Laboratory of Intelligent Vision Based Monitoring for Hydroelectric Engineering (No. 2022SDSJ04).

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

References

1. Ruan, Y.Z.; Jia, D. Some thoughts on basic surveying and mapping production service system under the new system. *Surv. Mapp. Sci.* **2020**, *45*, 178–182.
2. Chen, J.; Liu, W.Z.; Wu, H.; Li, Z.; Zhao, Y.; Zhang, L. Basic issues and research agenda of geospatial knowledge service. *Geomat. Inf. Sci. Wuhan Univ.* **2019**, *44*, 38–47.
3. Liu, J.N.; Guo, W.F.; Guo, C.; Gao, K.; Cui, J. Rethinking ubiquitous mapping in the age of intelligence. *J. Surv. Mapp.* **2020**, *49*, 403–414.
4. Zhang, X.; Zhang, C.; Wu, M.; Lv, G. Spatio-temporal features based geographical knowledge graph construction. *Sci. Sin. Inform.* **2020**, *50*, 1019–1032.
5. Lu, F.; Zhu, Y.Q.; Zhang, X.Y. Spatio-temporal knowledge graph: Advances and perspectives. *J. Geo-Inf. Sci.* **2023**, *25*, 1091–1105.
6. Brodt, A.; Nicklas, D.; Mitschang, B. Deep integration of spatial query processing into native RDF triple stores. In Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems, San Jose, CA, USA, 2–5 November 2010; pp. 33–42.
7. Liagouris, J.; Mamoulis, N.; Bouros, P.; Terrovitis, M. An effective encoding scheme for spatial RDF data. *Proc. VLDB Endow.* **2014**, *7*, 1271–1282. [CrossRef]
8. Wang, D.; Zou, L.; Feng, Y.; Shen, X.; Tian, J.; Zhao, D. S-store: An engine for large rdf graph integrating spatial information. In Proceedings of the Database Systems for Advanced Applications: 18th International Conference (DASFAA 2013), Wuhan, China, 22–25 April 2013; Proceedings, Part II 18. Springer: Berlin/Heidelberg, Germany, 2013; pp. 31–47.
9. Lu, F.; Yu, L.; Qiu, P.Y. On geographic knowledge graph. *J. Geo-Inf. Sci.* **2017**, *19*, 723–734.

10. Qiu, Q.; Xie, Z.; Ma, K.; Tao, L.; Zheng, S. NeuroSPE: A neuro-net spatial relation extractor for natural language text fusing gazetteers and pretrained models. *Trans. GIS* **2023**, *27*, 1526–1549. [CrossRef]
11. Wang, R.W. *Python and Data Science*; East China Normal University Press: Shanghai, China, 2015; p. 267.
12. Maron, M.E. Automatic indexing: An experimental inquiry. *J. ACM* **1961**, *8*, 404–417. [CrossRef]
13. Santos, R.; Murrieta-Flores, P.; Calado, P.; Martins, B. Toponym matching through deep neural networks. *Int. J. Geogr. Inf. Sci.* **2018**, *32*, 324–348. [CrossRef]
14. Wang, J.; Hu, Y. Enhancing spatial and textual analysis with EUPEG: An extensible and unified platform for evaluating geoparsers. *Trans. GIS* **2019**, *23*, 1393–1419. [CrossRef]
15. Minaee, S.; Kalchbrenner, N.; Cambria, E.; Nikzad, N.; Chenaghlu, M.; Gao, J. Deep learning based text classification: A comprehensive review. *arXiv* **2020**, arXiv:2004.03705. [CrossRef]
16. Tao, L.; Xie, Z.; Xu, D.; Ma, K.; Qiu, Q.; Pan, S.; Huang, B. Geographic Named Entity Recognition by Employing Natural Language Processing and an Improved BERT Model. *ISPRS Int. J. Geo-Inf.* **2022**, *11*, 598. [CrossRef]
17. Hua, X.L.; Xu, F.; Wang, Z.Q.; Li, P.F. Fine-grained classification method for abstract sentence of scientific paper. *Comput. Eng.* **2012**, *38*, 138–140.
18. Asghar, M.Z.; Khan, A.; Bibi, A.; Kundi, F.M.; Ahmad, H. Sentence-level emotion detection framework using rule-based classification. *Cogn. Comput.* **2017**, *9*, 868–894. [CrossRef]
19. Tan, L.; San Phang, W.; Chin, K.O.; Patricia, A. Rule-based sentiment analysis for financial news. In Proceedings of the 2015 IEEE International Conference on Systems, Man, and Cybernetics, Hong Kong, China, 9–12 October 2015; pp. 1601–1606.
20. Zhang, M.; Wang, J. Automatic Extraction of Flooding Control Knowledge from Rich Literature Texts Using Deep Learning. *Appl. Sci.* **2023**, *13*, 2115. [CrossRef]
21. Socher, R.; Perelygin, A.; Wu, J.; Chuang, J.; Manning, C.D.; Ng, A.Y.; Potts, C. Recursive deep models for semantic compositionality over a sentiment treebank. In Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, Seattle, WA, USA, 18–21 October 2013; pp. 1631–1642.
22. Iyyer, M.; Manjunatha, V.; Boyd-Graber, J.; Daumé, H., III. Deep unordered composition rivals syntactic methods for text classification. In Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), Beijing, China, 26–31 July 2015; pp. 1681–1691.
23. Kim, Y. Convolutional neural networks for sentence classification. *arXiv* **2014**, arXiv:1408.5882.
24. Johnson, R.; Zhang, T. Deep pyramid convolutional neural networks for text categorization. In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Vancouver, BC, Canada, 30 July–4 August 2017; pp. 562–570.
25. Graves, A.; Schmidhuber, J. Framewise phoneme classification with bidirectional LSTM and other neural network architectures. *Neural Netw.* **2005**, *18*, 602–610. [CrossRef]
26. Liu, X.; He, P.; Chen, W.; Gao, J. Multi-task deep neural networks for natural language understanding. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, Florence, Italy, 28 July–2 August 2019; pp. 4487–4496.
27. Qin, Q.; Yi, J.K. A BERT-CNN model for text classification. *J. Beijing Univ. Inf. Sci. Technol.* **2023**, *38*, 69–74.
28. Chung, J.; Gulcehre, C.; Cho, K.H.; Bengio, Y. Empirical evaluation of gated recurrent neural networks on sequence modeling. *arXiv* **2014**, arXiv:1412.3555.
29. Devlin, J.; Chang, M.W.; Lee, K.; Toutanova, K. BERT: Pre-training of deep bidirectional transformers for language understanding. *arXiv* **2018**, arXiv:1810.04805.
30. Lai, S.; Xu, L.; Liu, K.; Zhao, J. Recurrent convolutional neural networks for text classification. In Proceedings of the AAAI Conference on Artificial Intelligence, Austin, TX, USA, 25–30 January 2015; p. 29.
31. Orosoo, M.; Govindasamy, S.; Bayarsaikhan, N.; Rajkumari, Y.; Fatma, G.; Manikandan, R.; Bala, B.K. Performance analysis of a novel hybrid deep learning approach in classification of quality-related English text. *Meas. Sens.* **2023**, *28*, 100852. [CrossRef]
32. Kong, X.F.; Dong, B.; Xu, K.; Tan, Y.L. A text classification model for livelihood issues based on BERT—An example of Zhejiang provincial government hotline data. *J. Peking Univ.* **2023**, *59*, 456–466.
33. Wang, H.C.; Sun, M.Z. Chinese short text classification based on ERNIE-RCNN model. *Comput. Technol. Dev.* **2022**, *32*, 28–33.
34. Li, Y.C.; Qian, L.F.; Ma, J. Research on early detection of microblog rumours based on BERT-RCNN model. *Intell. Theory Pract.* **2021**, *44*, 173–177.
35. Qiu, Q.; Xie, Z.; Ma, K.; Chen, Z.; Tao, L. Spatially oriented convolutional neural network for spatial relation extraction from natural language texts. *Trans. GIS* **2022**, *26*, 839–866. [CrossRef]
36. Du, S.; Feng, C.C.; Guo, L. Integrative representation and inference of qualitative locations about points, lines, and polygons. *Int. J. Geogr. Inf. Sci.* **2015**, *29*, 980–1006. [CrossRef]
37. Du, S.; Guo, L. Similarity measurements on multi-scale qualitative locations. *Trans. GIS* **2016**, *20*, 824–847. [CrossRef]
38. Purves, R.S.; Clough, P.; Jones, C.B.; Arampatzis, A.; Bucher, B.; Finch, D.; Fu, G.; Joho, H.; Syed, A.K.; Vaid, S.; et al. The design and implementation of SPIRIT: A spatially aware search engine for information retrieval on the Internet. *Int. J. Geogr. Inf. Sci.* **2007**, *21*, 717–745. [CrossRef]
39. Huang, C. Factual knowledge meta-identification and citation in digital resources of digital libraries. *Sci. Technol. Entrep. Mon.* **2020**, *33*, 58–62.

40. Braithwaite, D.W.; Sprague, L. Conceptual knowledge, procedural knowledge, and metacognition in routine and nonroutine problem solving. *Cogn. Sci.* **2021**, *45*, e13048. [CrossRef] [PubMed]
41. Souza, C.; Garrido, M.V.; Horchak, O.V.; Carmo, J.C. Conceptual knowledge modulates memory recognition of common items: The selective role of item-typicality. *Mem. Cogn.* **2022**, *50*, 77–94. [CrossRef]
42. Sha, Z.Y.; Bian, F.; Chen, J.P. Research on spatial reasoning based on rule-based knowledge. *J. Wuhan Univ.* **2003**, *48*, 45–50.
43. Qiu, Q.; Xie, Z.; Wang, S.; Zhu, Y.; Lv, H.; Sun, K. ChineseTR: A weakly supervised toponym recognition architecture based on automatic training data generator and deep neural network. *Trans. GIS* **2022**, *26*, 1256–1279. [CrossRef]
44. Joulin, A.; Grave, E.; Bojanowski, P.; Mikolov, T. Bag of tricks for efficient text classification. *arXiv* **2016**, arXiv:1607.01759.
45. Liu, P.; Qiu, X.; Huang, X. Recurrent neural network for text classificatio with multi-task learning. *arXiv* **2016**, arXiv:1605.05101.
46. Yang, Z.; Yang, D.; Dyer, C.; He, X.; Smola, A.; Hovy, E. Hierarchical attention networks for document classification. In Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, San Diego, CA, USA, 12–17 June 2016; pp. 1480–1489.
47. Sun, C.; Qiu, X.; Xu, Y.; Huang, X. How to fine-tune bert for text classification? In Proceedings of the Chinese Computational Linguistics: 18th China National Conference (CCL 2019), Kunming, China, 18–20 October 2019; Proceedings 18; Springer International Publishing: Berlin/Heidelberg, Germany, 2019; pp. 194–206.
48. Sun, Y.; Wang, S.; Li, Y.; Feng, S.; Tian, H.; Wu, H.; Wang, H. Ernie 2.0: A continual pre-training framework for language understanding. In Proceedings of the AAAI Conference on Artificial Intelligence, New York, NY, USA, 7–12 February 2020; Volume 34, pp. 8968–8975.
49. Lu, X.L.; Ni, B. Research on BERT-CNN multi-level patent classification based on pre-trained language model. *J. Chin. Inf.* **2021**, *35*, 70–79.
50. Lin, D.P.; Wang, H.J. Comparison of news text classification based on BERT and RNN. *J. Beijing Inst. Print.* **2021**, *29*, 156–162.
51. Peng, Y.; Jiang, X.Y. A spam filtering system based on BERT_DPCNN text classification algorithm. *Comput. Knowl. Technol.* **2022**, *18*, 66–69.
52. Zhang, Z. Improved adam optimizer for deep neural networks. In Proceedings of the 2018 IEEE/ACM 26th International symposium on quality of service (IWQoS), Banff, AB, Canada, 4–6 June 2018; pp. 1–2.
53. Sharma, S.; Mehra, R.; Kumar, S. Optimised CNN in conjunction with efficient pooling strategy for the multi-classification of breast cancer. *IET Image Process.* **2021**, *15*, 936–946. [CrossRef]