sustainability

MDPI

Article

Research on the Classification of New Energy Industry Policy Texts Based on BERT Model

Qian Li [1], Zezhong Xiao [1,*] and Yanyun Zhao [2]

[1] School of Information, Beijing Wuzi University, Beijing 101149, China; liqian03067@163.com
[2] School of Statistics, Renmin University of China, Beijing 100872, China; cas-kriu@ruc.edu.cn
* Correspondence: zezhongxiao@163.com; Tel.: +86-152-3839-2244

Abstract: The existing means for classifying new energy industry policies are mainly based on the theory of policy instruments and manual encoding, which are highly subjective, less reproducible, and inefficient, especially when dealing with large-scale policy texts. Based on the theory of policy instrument, the research tried to apply the automatic classification model based on BERT to new energy industry policies to improve its classification efficiency and accuracy. A new energy industry policy classification model was established to train on policy texts and to compare the policy classification effects with the other two commonly used text classification models. The model comparison results show that the BERT model achieves higher precision, recall, and F1 score, indicating a better classification effect. Furthermore, adding topic sentences to training texts can effectively improve the classification effect of the BERT model. The policy classification results show that environmental policies are the most prevalent in new energy industry policies, while demand-side policy instruments are underutilized. Among the 11 types of subdivided policies, the application of goal planning policies is overflowing.

Keywords: new energy; industry policy; policy instruments; text classification; BERT model

check for updates

Citation: Li, Q.; Xiao, Z.; Zhao, Y. Research on the Classification of New Energy Industry Policy Texts Based on BERT Model. *Sustainability* **2023**, *15*, 11186. https://doi.org/10.3390/su151411186

Academic Editors: Seung-Hoon Yoo, Ruipeng Tan and Xinju He

Received: 22 May 2023
Revised: 11 July 2023
Accepted: 17 July 2023
Published: 18 July 2023

Copyright: © 2023 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/).

1. Introduction

In the context of global energy system change [1] and economic transformation [2], promoting the high-quality development of new energy industry plays an important role in ensuring energy security, achieving the low-carbon goal and creating a sustainable economy [3–5]. However, the new energy industry in China started late and was weak in competitiveness, and the amount of capital required for industrial development was large, requiring government guidance and support [6]. The Chinese government attached great importance to the new energy industry. Since the 10th Five-Year Plan period, the State Council, the National Development and Reform Commission, the National Energy Administration and other departments have issued a series of policies for the development of a new energy industry. The 14th Five-Year Plan for Modern Energy System proposes that 'increase financial and policy support in key core technology areas such as advanced renewable energy power generation', which provides a guarantee for the high-quality development of a new energy industry; the Action Plan for Carbon Peaking by 2030 clearly pointed out that they hope to 'vigorously develop new energy, comprehensively promote the large-scale and high-quality development of wind power and solar power, and accelerate the construction of a new power system in which the proportion of new energy gradually increases'.

With the increasing number and types of new energy industry policies released, the demand for refined research on policy content in government decision-making process is getting stronger [7,8]. However, due to the different lengths of policy texts, inconsistent classification system and high information density, it is difficult to achieve fast and accurate classification of policies through traditional policy text classification methods. Therefore,

based on the theory of policy instrument, this paper used the BERT model in the text mining method to automatically classify the policy texts of new energy industry to improve the efficiency and accuracy of policy text classification, which in turn helps the government clarify different types of new energy industry policies issued, provides a decision basis for subsequent policy formulation, adjustment and optimization to promote the healthy and orderly development of the new energy industry.

The contributions of this paper are as follows: Firstly, the text mining method based on the BERT model was introduced into the study of new energy industry policy classification, which provided a new feasible method to achieve a fast and accurate classification of large-scale policy texts in the new energy industry, aiding decision-makers to grasp the structure of the policy system more quickly and systematically, thus, improving the efficiency of government decision-making. Secondly, a classification system based on the theory of policy instruments was built to make the classification more systematic and objective. Thirdly, this paper fused the policy titles and policy topic sentences as the input features of the model, effectively improving the classification effect of the Bert model.

2. Literature Review

2.1. Research on New Energy Industry Policy Classification

From a macro perspective, the new energy industry includes the new energy power generation industry and the new energy vehicle industry [6,8], where new energy power generation refers to the conversion of new energy sources such as solar, wind, biomass, ocean, and geothermal energy into electricity using developed technologies [9]. New energy vehicles refer to vehicles with new technologies, new structures and advanced technical principles that use unconventional vehicle fuels as a power source [10]. New energy industrial policies refer to policies implemented to compensate for the immature market development of the new energy industry and to promote the development of the industry. In this paper, the new energy generation policies and new energy vehicle policies were analyzed together as new energy industry policies.

Research on new energy industry policies initially focused on qualitative interpretation and summary of policy contents [11,12]. Due to the high information density and uneven distribution of connotations in policy texts [13], scholars have gradually carried out policy classification studies. The existing research methods of new energy policy classification are mainly based on policy instrument theory and manual encoding. Policy instruments are a collection of various instruments and measures used by government departments to achieve policy goals [14], and there are various classification criteria. One of the most classic classification methods is the supply side, demand-side and environmental policy instrument classification proposed by Rothwell et al. [15]. Rothwell et al. introduced policy instruments into innovation policy analysis at an early stage and classified policy categories according to the different levels of impact of policy instruments on technological innovations. This classification method reduced the dimensionality of the complex innovation policy system from the perspective of instruments and measures, which is of significant intra-dimensional aggregation and inter-dimensional differentiation validity. Meanwhile, this combination of policy instruments emphasized the regulatory role of the supply of policy subjects, the demand for policy objects and the macro external environment, which is strongly targeted and content-directed. Therefore, it is widely applied in policy research in the new energy industry and other fields [16], such as wind energy industry policy [17], photovoltaic industry policy [18,19], and new energy vehicle industry policy [20,21]. Wang and Yu [22] evaluated the implementation effects of different types of policy instruments based on China's photovoltaic industry policies from 2010 to 2020 based on the scoring of policy codes. Zhang and Zhou [23] constructed a quantitative evaluation framework to subdivide 33 new energy vehicle subsidy policies into 11 categories based on the classification theory of policy instruments and analyzed the existing problems. In addition, some scholars classify the text of new energy industry policies based on other policy instrument classification criteria. Zeng and Hu [24] constructed a policy analysis unit encoding table to

classify solar energy industry policies into three policy instrument types: structural coercive tools, contractual inducement tools, and interactive influence tools to analyze the macro support characteristics of the industry. Wang et al. [25] classified 253 new energy vehicle policy texts according to the economic instruments, regulations, and information to analyze their policy content. Using policy instruments and policy encoding methods to classify new energy industrial policies is beneficial for the government to have a comprehensive understanding of industrial policies. However, the method is highly subjective and easily influenced by human factors when classifying policy instruments and encoding policy units. Moreover, for a large and complex industry policy system, the manpower loss is large, and the replicability of the method is low.

2.2. Text Classification Methods

With the generation and development of technologies such as cloud computing and big data, text classification methods were gradually applied in various fields of text content research. The commonly used text classification methods can be divided into three categories. The first category is rule-based text classification methods, including decision trees [26], association rules [27], etc. Liu and Beldona [28] used association rules to study customers' comment texts of urban hotels in the U.S. to assess consumers' revisit intentions. The rule-based classification method requires a clear category definition of the target text and experts with specialized background knowledge to propose effective classification rules, which is difficult to replicate in other text scenarios. The second category is shallow learning models, including plain Bayesian, k-nearest neighbor algorithm (KNN), and support vector machine (SVM). Gao et al. [29] constructed and improved the distributed naive Bayes automatic classification system and achieved better classification results. Jiang et al. [30] proposed an improved KNN algorithm to reduce text similarity computation in the face of surging text data. Zhang [31] realized the automatic classification of public security information based on SVM. The shallow learning models can achieve better performance in text classification accuracy and speed but rely on manual acquisition of effective text features, which have some limitations. The third category is deep learning models, including convolutional neural networks (CNNs), recurrent neural networks (RNNs), long short-term memory networks (LSTMs) [32], attention mechanisms [33], etc. In 2006, Hinton et al. [34] proposed that artificial neural networks with multiple hidden layers have excellent feature learning abilities, which started the wave of deep learning. Subsequently, Kim [35] used CNNs for text classification and built the TextCNN model to verify its effectiveness in short text classification; Liu et al. [36] established the TextRNN model with an integrated multi-task framework to tackle the problem of insufficient training data for a single task. In 2018, Google released the bi-directional encoding representation based on the transformer model (BERT), which is able to extract semantic information through a bidirectional transformer structure and obtain richer text features compared with other deep learning models [37]. BERT model has been widely used in text classification studies such as news text [38], sentiment analysis [39,40] and government administration [41]. Some scholars have applied the BERT model to policy text classification [42], but there are fewer studies related to its application to the new energy industry policy text classification.

By combining the literature, it can be seen that the existing research on new energy industry policy classification mainly classified policy texts manually from the perspective of policy instruments, which is more subjective and less reproducible, and is inefficient especially when dealing with large-scale policy texts. The BERT model based on deep learning can better handle large-scale text content and achieve automatic text classification while ensuring the accuracy and efficiency of classification. Some scholars have tried to apply the BERT model to the text classification of science and technology policies [13], but no relevant research has been conducted for new energy industry policies. Therefore, the paper took the new energy industry policy as the research object and built the new energy industry policy classification system based on the policy instrument theory, using the BERT model to realize the automatic classification of the new energy industry policy text.

3. Research Methodology

3.1. Research Framework

In this research, the BERT model was used to classify the texts from the new energy industry policies. The research framework is displayed in Figure 1, which includes four parts: data acquisition and preprocessing, model training, model comparison, and result analysis. First, using crawler technology to obtain new energy industry policies, perform data cleaning and other preprocessing, then manually label the data according to the classification system built based on the policy instrument theory to form a corpus of new energy industry policies. Second, dividing the data into the training set, test set, and validation set, build a BERT model to train the policy texts and select FastText and TextCNN models as baseline models. Third, compare the classification effect of the BERT model with the baseline models and select the model with a better classification effect. Finally, systematically analyze the classification results of new energy industry policy texts.

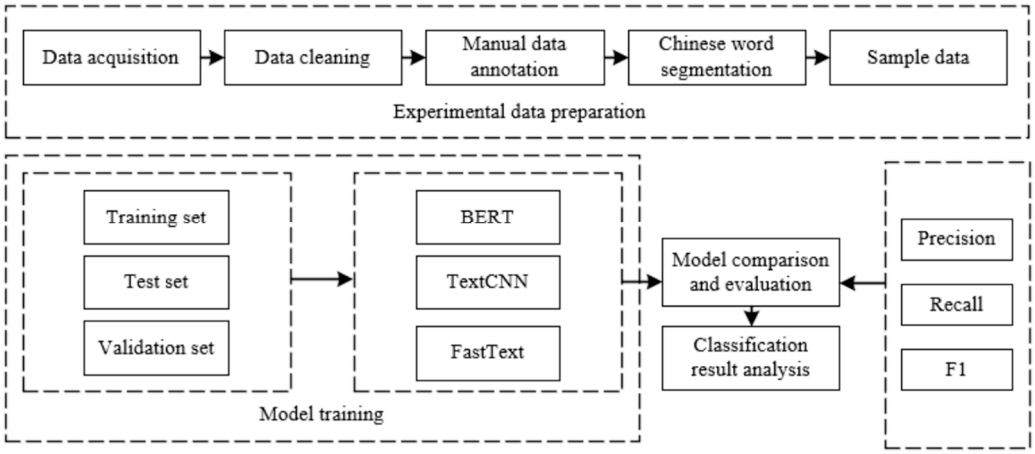


Figure 1. Framework for text classification research on new energy industry policies.

3.2. BERT Model

The BERT model is a pre-trained language model, released by Google in 2018, which uses bidirectional encoding representations [37]. As the first representation model based on fine-tuning, it has refreshed the best results of 11 tasks in the field of natural language processing (NLP) and is one of the major breakthroughs in current natural language processing research. The BERT model structure consists of multi-layer Transformer stacks, as shown in Figure 2. In order to improve the ability of semantic representation, BERT designed two new unsupervised prediction tasks: masked language model (MLM) and next sentence prediction (NSP), enabling it to reflect semantic information more accurately at the word block, sentence or even text level.

The BERT model learns via two processes: pre-training and fine-tuning. BERT pre-trains on a large number of document-level general corpora to learn the essential features of language and general semantic information. Fine-tuning is the process of modifying the pre-trained initialization parameters so that the parameters can be adapted to different datasets. BERT can be fine-tuned on a specific corpus to obtain new semantic features required in specific tasks, truly realizing migration learning in the NLP field and providing great convenience for related research. The specific algorithm process can be roughly divided into three steps. ① Data preparation. Use the tokenizer to perform preprocessing operations such as data cleaning and word segmentation on the collected original text data, and then store the data in the document in a specific form; ② Vector conversion. Read the prepared data in the BERT embedding layer, convert the data into a multi-feature vector form through Token Embeddings, Segment Embeddings and Position Embeddings, and convert the data into TFrecords Data format; ③ Model construction. Load the BERT pre-trained model, fine-tune the model according to downstream tasks, and finally run the model and output the results.

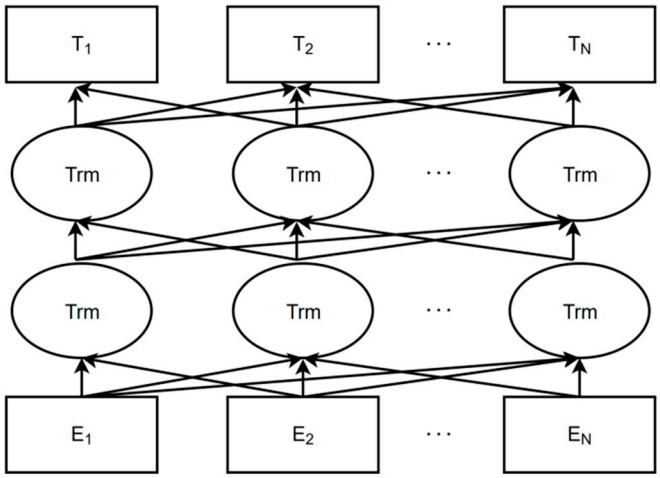


Figure 2. Structure of BERT model.

3.3. Samples and Data Processing

The paper conducted text classification research based on the new energy industry policies publicly released by government departments in China at all levels from 2000 to 2022, with data mainly obtained from the National Development and Reform Commission, the National Energy Administration, provincial and municipal government websites at all levels, and web pages, such as Legal Star and Beida Fabao. First, using crawler technology to obtain the new energy industry policies queried by keyword search, and the data fields included the policy title, release time, release department, and full text of the policy. Then, the new energy industry policy data were preprocessed. Through manual screening and sorting, duplicate information and information not related to new energy industry policies were removed, and finally, 12,987 valid sample data were obtained. Finally, one of the researchers labelled the texts manually according to the classification system built based on the policy instrument theory. The labelled data were then re-labelled by another researcher on a sample basis. The agreement rate of the labelling results reached over 90%, reflecting the high reliability of the ground truth label of this experiment. The experimental data were divided into training sets, test sets, and validation sets in the ratio of 8:1:1.

3.4. Experimental Environment and Parameter Settings

In this research, the model was established using python3.7 programming language based on version 1.5.0 of the pytorch deep learning framework to implement the classification task of new energy industry policy texts. The experimental environment is displayed in Table 1.

Table 1. Experimental environment.

Name	Configuration
CPU	Intel (R) Core (TM) i7 CPU @ 2.8 GHz
Memory	16 GB
GPU	Intel Iris Pro 1536 MB
System	MacOs 11.6.8

The experimental model used is the Chinese pre-training model of BERT-Base, which consists of 12 layers of Transformer, the hidden layer size is 768, and the number of multi-head attention is 12. The maximum text length is 512. The loss function adopted the cross-entropy loss function. The activation function is ReLU. During the model training process, the hyperparameters of the model were adjusted by observing the trend of the loss rate on the validation set. Initially, the learning rate was set to a smaller value of 1×10^{-5}. The number of rounds was gradually increased from 1 to 10. It was found that as the

number of rounds went from 1 to 5, the loss value of the validation set decreased, and the model was in an underfitting state. From round 6 to 10, the loss value increased, and the model training started to overfit, so the epochs were set to 5, at which point the loss value was 0.041. Next, the learning rate was adjusted to observe the effect of the model learning rate on the loss value. When increasing the learning rate from 1×10^{-5} to 5×10^{-5}, it was found that the loss decreased faster and did not appear unstable. When the learning rate was increased from 5×10^{-5} to 1×10^{-4}, the loss graph oscillated, and the model did not converge. Therefore, the learning rate was set to 5×10^{-5}, at which point the loss value was 0.022 at the time. As shown in Figure 3, the decline in loss of the BERT model slowed down significantly when the number of iterations reached 103, at which point the model stabilized and could harvest the best effect. The final hyperparameter values and the related training details of this model are depicted in Table 2.

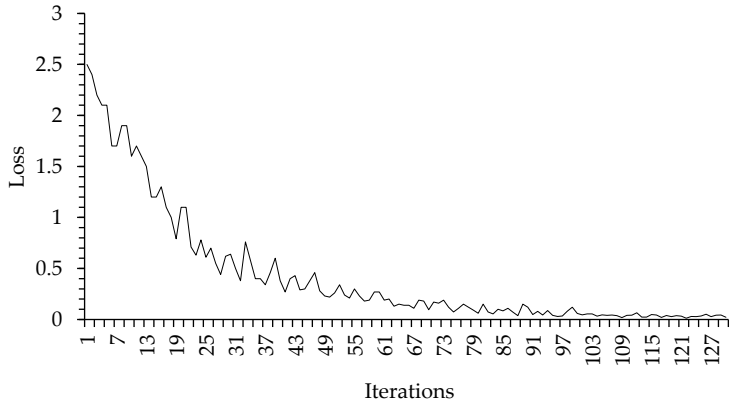


Figure 3. Change in loss of the BERT model on the validation set.

Table 2. Hyperparameter settings.

Hyperparameter	Value
Optimizer name	Adam
Batch size	512
Epochs	5
Learning rate	5×10^{-5}
Dropout rate	0.1

3.5. Evaluation Indices

This study adopted the universal evaluation criteria, namely, Precision, Recall, and F1, as evaluation indices to evaluate the performance of different classification models. The evaluation metrics are defined as follows. First, we calculated the Precision, Recall, and F1 for each class according to the following equations and then weighted the average separately as the final experimental results to evaluate the comprehensive model performance.

$$\text{Precision} = \text{TP}/(\text{TP} + \text{FP}) \tag{1}$$

$$\text{Recall} = \text{TP}/(\text{TP} + \text{FN}) \tag{2}$$

$$\text{F1} = 2\,(\text{Precision} \cdot \text{Recall})/(\text{Precision} + \text{Recall}) \tag{3}$$

where TP is the number of samples that are positive and judged as positive by the classifier. FP is the number of samples that are negative but judged as positive by the classifier. FN is the number of samples that are positive but judged as negative by the classifier.

4. Results

4.1. New Energy Industry Policy Classification System

The classification of supply side, environmental, and demand-side policy instruments proposed by Rothwell et al. is the most commonly used industry policy classification method [15]. Among them, supply side policy instruments are mainly manifested as a driving force for the industry, which means that the government promotes industrial development by expanding or improving the supply of production factors such as funds, talents, and facilities; environmental policy instruments are mainly manifested as indirect effects on industrial development. It means that the government establishes a benign institutional environment and social environment for industrial development by setting normative and leading documents, etc., improving the enthusiasm of relevant industrial entities; demand-side policy instruments are mainly manifested in stimulating market demand, which means that the government, through government procurement, outsourcing, etc., stimulates market demand and reduces the uncertainty in industrial development. The paper built a new energy industry policy classification system based on the policy instrument theory proposed by Rothwell et al. However, due to the increasing number and types of new energy industry policies, the policy dimensions of supply side, environmental, and demand-side can no longer fully reflect the policy characteristics. Therefore, the paper further refined a new energy industry policy classification based on the policy instrument theory. In order to ensure the rationality of the classification, the researchers referred to and summarized the predecessors' research [14,20,21,43], as well as combined the collected policy data to develop the classification criteria in a comprehensive manner. The new energy industry policies are divided into 11 sub-policy instrument types such as financial support, technical support and project construction, etc. The detailed classification and basis are shown in Table 3.

Table 3. Categories of new energy industry policy texts.

Instrument Type	Subdivision of Policy Instrument	Definition and Examples
Supply side	Funding support	The government promotes industry development through financial investment, including special funds for renewable energy, subsidies for the purchase of new energy vehicles, and subsidies for electricity prices
	Technical support	The government provides support for the innovative development of new energy-related technologies, including talent training, action plans for the innovative development of new energy industries, and integrated platforms for industry, academia, and research
	Project construction	The government promotes the development and utilization of new energy by means of project construction, including the construction of photovoltaic power plant projects and decentralized wind power projects
	Infrastructure construction	The government improves the supporting infrastructure required by the new energy industry, including the construction of urban and rural distribution grids, new energy vehicle charging piles, hydrogen refueling supporting infrastructure construction, etc.
Environmental	Tax incentives	The government has issued tax relief policies related to the new energy industry, including a 50% instant refund of VAT and corporate income tax
	Financial support	The government provides a good financing environment for new energy enterprises through direct or indirect means, including renewable energy stock project tariff subsidy confirmation loans, green bonds, green credit, etc.

Table 3. *Cont.*

Instrument Type	Subdivision of Policy Instrument	Definition and Examples
Environmental	Regulatory control	The government strengthens the supervision and management of new energy industry activities, including new energy development and utilization management methods, product quality regulation, etc.
	Goal planning	The government makes overall layout and goal guidance for new energy industry, including renewable energy development planning, solar energy development planning, etc.
Demand-side	Promotion and application	The government encourages the application and promotion of achievements related to the new energy industry, including the promotion and application of photovoltaic power generation and building integration, and the application of biomass energy in agricultural production and rural life
	Government procurement	The government promotes the consumption of technology products and services related to the new energy industry through purchases, including new energy vehicle procurement and public sector vehicle electrification programs
	Trade control	The government uses import and export trade controls for new energy products to stimulate demand, including product tariffs, trade agreements, etc.

4.2. Model Comparison

The full texts of the new energy industry policies are too long, which is not conducive to classification model training. Since the maximum length of a single text allowed by BERT model is 512 characters, the fusion of the new energy industry policy title and topic sentence were used as training texts in this research. First, the semantic similarity of each sentence and title in the full text of the policy was calculated, and the first two sentences with higher similarity to the policy title were selected as the topic sentences. Then, the title and the policy topic sentences were spliced as the input features for model training. Some examples of input–output types for model training are displayed in the Table 4 below.

Table 4. Examples of input–output types for model training.

Input			Output
Title	**Topic Sentence 1**	**Topic Sentence 2**	**Policy Type**
Interim Measures for the Management of Financial Subsidy Funds for Solar Photovoltaic Building Applications	Strengthening the management of financial assistance funds for solar photovoltaic building applications	The subsidy rate for solar photovoltaic buildings is in principle set at RMB 20/Wp	Funding support, Supply side
The 12th Five-Year Plan for the Development of New Energy in Shanghai	Strengthening new energy planning guidance	Main tasks for the development of new energy in the 12th Five-Year Plan	Goal planning, Environmental
Implementation plan for the purchase of new energy vehicles by government agencies and public institutions	Regulating the management of new energy vehicle procurement	Inclusion of new energy vehicles in the centralized government procurement catalogue and priority procurement	Government procurement, Demand-side

In order to verify the classification effect of the BERT model, the TextCNN model and the FastText model were used as the baseline models for comparative experiments. The classification effects achieved by different text classification models are listed in Table 5. Comparing the experimental results of the three classification models, the results for

precision rate, recall rate, and F1 value of the BERT model are higher than those of the FastText and TextCNN models, and its classification effect is the best, with accuracy, recall, and F1 values of 86.92%, 88.60%, and 87.74%, respectively. Compared with FastText model, the F1 value of BERT improved by 6.67%; compared with TextCNN model, the F1 value of BERT improved by 2.78%. Therefore, it is possible to derive that the BERT model has better classification ability in dealing with the large-scale new energy industry policy tasks.

Table 5. Comparative experimental results.

Model	Precision	Recall	F1
FastText	79.21%	83.15%	81.07%
TextCNN	84.01%	86.13%	84.96%
BERT	86.92%	88.60%	87.74%

Table 6 shows the classification results for each policy category in the BERT model trained using policy titles plus topic sentences, of which project construction, goal planning, and trade control policies were better classified, with F1 values above 90%. The six policy categories of funding support, technical support, infrastructure construction, tax incentives, promotion and application, and government procurement were the next most effective, with F1 values exceeding 85%. Financial support and regulatory control were less effective, with F1 values of 78.28% and 81.61%, respectively. Further analysis of the policy sample revealed that the similarities between some of the financial support and funding support policies reduce the effectiveness of their classification to a certain extent. The regulatory control policies have a broader scope and tend to overlap with other policy categories, thus leading to misclassification. In general, the BERT-based classification model can automatically classify new energy industry policy texts more accurately and can quickly transform key policy contents into quantitative information, facilitating decision-makers to classify and summarize new energy industry policy measures from the perspective of policy instruments, thus systematically grasping the policy structure and improving government decision-making efficiency.

Table 6. Classification effect of the BERT model on the text set.

Policy Category	Precision	Recall	F1
Funding support	87.12%	86.23%	86.67%
Technical support	85.76%	92.25%	88.89%
Project construction	90.23%	91.31%	90.77%
Infrastructure construction	85.21%	86.22%	85.71%
Tax incentives	87.13%	86.95%	87.04%
Financial support	76.12%	80.56%	78.28%
Regulatory control	79.33%	84.02%	81.61%
Goal planning	88.42%	91.81%	90.08%
Promotion and application	89.73%	87.27%	88.48%
Government procurement	87.35%	84.25%	85.77%
Trade control	89.01%	92.29%	90.62%

4.3. Ablation Study

In order to evaluate the effectiveness of fusing policy titles and topic sentences as input features in the BERT-based new energy policy text classification task, this paper conducted ablation experiments with the following settings.

① Ablation experiment 1: using policy titles as input features.

② Ablation experiment 2: using policy topic sentences as input features.

③ This paper proposed to splice policy titles and topic sentences as input features.

The results of the ablation experiments are shown in Table 7. The method in this paper achieved better accuracy, recall, and F1 value than the ablation experiments 1 and 2. Compared with the ablation experiment 1, the F1 value of this method was improved by

about 0.99%; compared with the ablation experiment 2, the F1 value was improved by about 1.16%. Therefore, it can be concluded that adding the policy top sentences to the training texts can further improve the performance of the BERT model when classifying new energy industry policy texts.

Table 7. Results of ablation experiments.

Ablation Experiment	Precision	Recall	F1
Ablation experiment 1: policy titles	85.93%	87.70%	86.75%
Ablation experiment 2: policy topic sentences	86.24%	87.13%	86.58%
Ours: policy titles + topic sentences	86.92%	88.60%	87.74%

4.4. Text Classification Result Analysis

The policy classification results of the BERT model trained using policy titles and topic sentences are shown in Figure 4. As can be seen from Figure 4, the new energy policies cover three types of policy instruments: supply side, environmental, and demand-side. The number of environment-side and supply side policies is higher, accounting for 44.24% and 39.89%, respectively, while demand-side policies are less, accounting for only 15.87%.

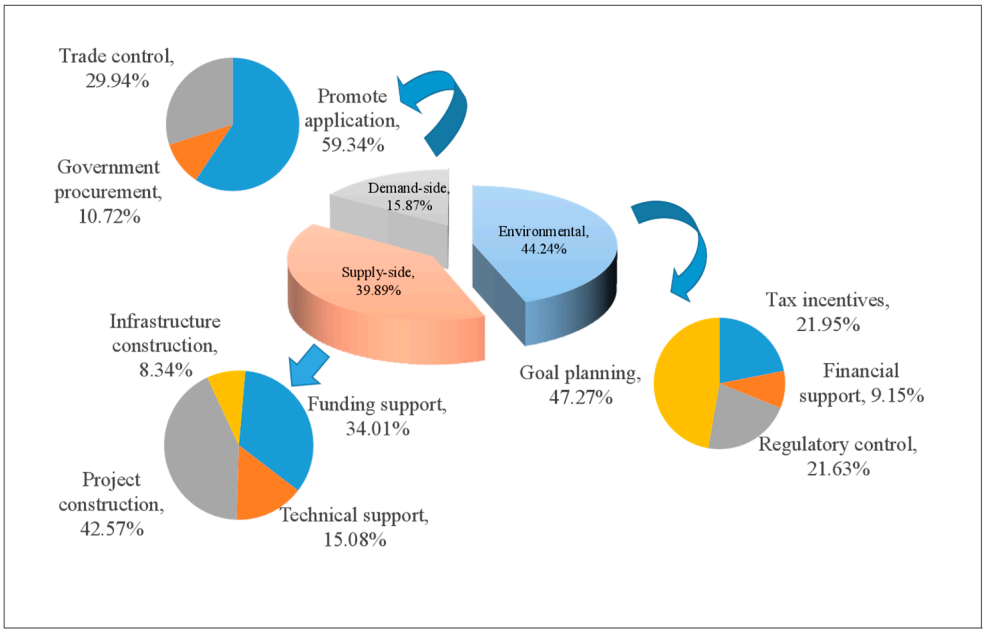


Figure 4. Classification result of new energy industry policies.

Further analysis reveals that among the environmental policies, goal planning policies account for a larger share (47.27%), followed by tax incentives policies (21.95%), and regulatory control policies (21.63%). The use of financial support policy instrument is relatively low (9.15%), which shows that the government attached great importance to the development planning of the new energy industry, dynamically adjusted the development goals according to the current stage of the industry, effectively played a leading role in policy and created a permanent and stable institutional environment for industrial production activities. However, the current goal planning policy instrument accounts for too much. The possible causes of this phenomenon are: On the one hand, after the central government released the new energy industry development plan, local governments formulated and released corresponding policies based on the central government's policies according to the local actual situation [44], resulting in the duplication of goal planning policies to a certain extent. On the other hand, some of the new energy industry policies released due to the former weak implementation, leading to supplementary policies later [45], thus causing policy overflow. For example, since 2012, the Chinese state has issued relevant policies

in response to the emergence of insufficient renewable energy consumption capacity and wind and photovoltaic power curtailment problems, including the *Notice on Doing a Good Job of Wind Power Grid Consumption in 2015*, the *Measures for resolving curtailment of hydro, wind and PV power generation* issued by the National Development and Reform Commission in 2017. In the *13th Five-Year Renewable Energy Development Plan* and *14th Five-Year Renewable Energy Development Plan* constantly emphasized improving the level of consumption and reducing the wind and photovoltaic power curtailment.

On the supply side policies, the project construction and funding support policies are more numerous, accounting for 76.58% of the total. The Chinese government took the construction of major projects as the focus, continuously expanding the scale of installed new energy capacity and promoting the development and utilization of new energy. Meanwhile, the government released the *Interim Measures for the Management of Financial Subsidy Funds for Golden Sun Demonstration Project*, *Interim Measures for the Management of Special Funds for Renewable Energy Development*, as well as the *Subsidy Program for the Promotion of New Energy Vehicles* and the *Notice on the Advance Issuance of Additional Subsidy Funds for Renewable Energy Tariff in 2021*, which provided strong financial support for the development of new energy industry. However, the technical support policies only account for 15.08%. At present, the manufacturing technology level of China's new energy industry has reached the world's top level, but there are still gaps with the international advanced level in the fields of offshore wind power, hydrogen, and fuel cell technology, and there is also a lack of original technology in some of the dominant areas [46,47], so the government should increase investment in innovation and focus on core key technologies for new energy. The proportion of infrastructure construction policies is 8.34%. Since the 12th Five-Year Plan period, the infrastructure to ensure wind power, photovoltaic power generation and other energy supply has been gradually improved, and new infrastructure, such as a smart grid, is also being built. However, as the number of new energy vehicles continues to increase, it is difficult to meet the demand for charging infrastructure [48].

Among the demand-side policies, the promotion and application policies account for the highest percentage, reaching 59.34%, which encourages the promotion of a new energy industry through demonstration and application measures. Especially in recent years, the government subsidies have been reduced, gradually achieving withdrawn and grid parity, indicating China's new energy industry has begun to gradually realize market-oriented transformation [49]. Therefore, the government increased the promotion and application of new energy products and encouraged the consumption of new energy products in order to create a stable market for the new energy industry. Trade control and government procurement policies account for a relatively low percent, 29.94% and 10.72%, respectively. The Chinese government enhanced the consumption of new energy products demand through product tariffs, trade agreements, government procurement and other measures.

5. Conclusions

This paper drew the following conclusions from a text classification study on China's new energy industry policies from 2000 to 2022:

The BERT model can improve the accuracy and efficiency of industry policy text classification and realize the automatic classification of new energy industry policies. This paper constructed a new energy industry policy classification standard based on policy instrument theory and verified the superiority of the BERT model in the field of new energy industry policy text classification by comparing the classification effect of FastText, TextCNN, and BERT. Furthermore, the classification effect of the BERT model can be further improved when using the fusion of policy title and topic sentences for text training. With the continuous development of the new energy industry, the policy system is becoming more and more complex. Deep learning classification models such as BERT can help decision-makers quickly identify policy types from the perspective of policy instruments, systematically grasp the current policy system structure, and provide a basis for subsequent policy formulation and policy optimization, improving decision-making efficiency.

New energy industry policies cover three types of policy instruments: supply side, environmental and demand-side, among which environmental policy instruments are most frequently used, accounting for 44.24%. Among the 11 types of subdivided policies, the application of goal planning policies is overflowing. The government should scientifically adjust and optimize the current policy structure. Demand-side policy instruments are less commonly used in new energy industry policies, accounting for less than 20% of the total. The Chinese government should pay more attention to the role of demand-side policy instruments in driving market demand and provide sufficient policies to meet the market-oriented development needs of the new energy industry at this stage.

This paper analyzed the proportion of quantity and structure of China's new energy industry policies according to the classification system built based on policy instrument theory and demonstrated the feasibility of applying the BERT model to the classification of new energy industry policies. The follow-up study will try to improve the following three aspects. First, discuss the policy classification features of the new energy generation and new energy vehicle industrial policies, respectively. Explore a more scientific and reasonable set of new energy industry policy classification systems, improving the standardization of industry policy classification. Second, further policy recommendations can be proposed by measuring the strength and effectiveness of different new energy industry policies. Thirdly, apply multi-task and multi-label text classification methods by combining the characteristics of industry policy texts to meet the needs of decision-makers for cutting-edge industry policies and fine-grained policy information. Also, with the rapid development of natural language processing, more text mining models, such as GPT3/4, BARD, etc., have been generated with great improvements. It is of great value to compare the performance of the BERT model with these existing state-of-the-art models.

Author Contributions: Q.L., Z.X. and Y.Z. were responsible for the writing of the full text. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by The National Social Science Fund of China, grant number 22CTJ006.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data were obtained from China's government departments at all levels from 2000 to 2022, mainly including the National Development and Reform Commission, the National Energy Administration, provincial and municipal government websites at all levels.

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

References

1. Chishti, M.Z.; Ahmad, M.; Rehman, A.; Khan, M.K. Mitigations pathways towards sustainable development: Assessing the influence of fiscal and monetary policies on carbon emissions in BRICS economies. *J. Clean. Prod.* **2021**, *292*, 12603. [CrossRef]
2. Yin, L.M.; Liu, J. Impact of Environmental Economic Transformation Based on Sustainable Development on Financial Eco-Efficiency. *Sustainability* **2023**, *15*, 856. [CrossRef]
3. Sinha, A.; Shahbaz, M.; Sengupta, T. Renewable energy policies and contradictions in causality: A case of Next 11 countries. *J. Clean. Prod.* **2018**, *197*, 73–84. [CrossRef]
4. Xu, B.; Lin, B.Q. Do we really understand the development of China's new energy industry? *Energy Econ.* **2018**, *74*, 733–745. [CrossRef]
5. Wang, Q.; Kwan, M.-P.; Fan, J.; Zhou, K.; Wang, Y.-F. A study on the spatial distribution of the renewable energy industries in China and their driving factors. *Renew. Energy* **2019**, *139*, 161–175. [CrossRef]
6. Wu, G.; Zeng, M.; Peng, L.L.; Liu, X.M.; Li, B.; Duan, J.H. China's new energy development: Status, constraints and reforms. *Renew. Sustain. Energy Rev.* **2016**, *53*, 885–896. [CrossRef]
7. Chang, Y.H.; Fang, Z.; Li, Y.F. Renewable energy policies in promoting financing and investment among the East Asia Summit countries: Quantitative assessment and policy implications. *Energy Policy* **2016**, *95*, 427–436. [CrossRef]
8. Wang, Q.Q.; Li, C.B. An evolutionary analysis of new energy and industry policy tools in China based on large-scale policy topic modeling. *PLoS ONE* **2021**, *16*, e0252502. [CrossRef]
9. Lin, B.; Xu, B. How to promote the growth of new energy industry at different stages? *Energy Policy* **2018**, *118*, 390–403. [CrossRef]

10. Wang, S. Exploring the Sustainability of China's New Energy Vehicle Development: Fresh Evidence from Population Symbiosis. *Sustainability* **2022**, *14*, 10796. [CrossRef]
11. Zeng, M.; Liu, X.M.; Li, N.; Xue, S. Overall review of renewable energy tariff policy in China: Evolution, implementation, problems and countermeasures. *Renew. Sustain. Energy Rev.* **2013**, *25*, 260–271. [CrossRef]
12. Singh, R.; Sood, Y.R. Current status and analysis of renewable promotional policies in Indian restructured power sector: A review. *Renew. Sustain. Energy Rev.* **2011**, *15*, 657–664. [CrossRef]
13. Ma, Y.M.; Huang, J.X.; Wang, F.; Rui, X. Research on multi-label classification of S&T policy content combining BERT and Multi-Scale CNN. *J. Intell.* **2022**, *41*, 157–163. (In Chinese)
14. Huang, C.; Yang, C.; Su, J. Identifying core policy instruments based on structural holes: A case study of China's nuclear energy policy. *J. Informetr.* **2021**, *15*, 101145. [CrossRef]
15. Rothwell, R.; Zegveld, W. *Reindustrialization and Technology*; Logman Group Limited: London, UK, 1985; pp. 83–104.
16. Tu, Q.; Mo, J.L.; Fan, Y. The evolution and evaluation of China's renewable energy policies and their implications for the future. *Chin. J. Popul. Resour.* **2020**, *30*, 29–36. (In Chinese)
17. Wang, X.Z.; Zou, H.H. Study on the effect of wind power industry policy types on the innovation performance of different ownership enterprises: Evidence from China. *Energy Policy* **2018**, *122*, 241–252. [CrossRef]
18. Zhi, Q.; Sun, H.H.; Li, Y.X.; Xu, Y.R.; Su, J. China's solar photovoltaic policy: An analysis based on policy instruments. *Appl. Energy* **2014**, *129*, 308–319. [CrossRef]
19. Zhang, H.M.; Xu, Z.D.; Sun, C.W.; Elahi, E. Targeted Poverty Alleviation Using Photovoltaic Power: Review of Chinese Policies. *Energy Policy* **2018**, *120*, 550–558. [CrossRef]
20. Wang, X.; Wang, J.; Xu, C.; Zhang, K.; Li, G. Electric Vehicle Charging Infrastructure Policy Analysis in China: A Framework of Policy Instrumentation and Industrial Chain. *Sustainability* **2023**, *15*, 2663. [CrossRef]
21. Gao, W.; Hu, X.Y. New energy vehicle policy effect: Does scale or innovation serve as an intermediary? *Sci. Res. Manag.* **2020**, *41*, 32–44. (In Chinese) [CrossRef]
22. Wang, B.J.; Yu, P. Evaluation on the policy efficacy and effect of photovoltaic industry: Quantitative analysis of China's policy texts from 2010 to 2020. *Soft Sci.* **2022**, *36*, 9–16. (In Chinese) [CrossRef]
23. Zhang, Y.A.; Zhou, Y.Y. Policy instrument mining and quantitative evaluation of new energy vehicles subsidies. *Chin. J. Popul. Resour.* **2017**, *27*, 188–197. (In Chinese)
24. Zeng, J.J.; Hu, J.X. Textual and quantitative research of solar industry in China from the perspective of policy tools. *Sci. Technol. Manag. Res.* **2014**, *34*, 224–228. (In Chinese)
25. Wang, X.L.; Huang, L.C.; Daim, T.; Li, X.; Li, Z.Q. Evaluation of China's new energy vehicle policy texts with quantitative and qualitative analysis. *Technol. Soc.* **2021**, *67*, 101770. [CrossRef]
26. Bădulescu, L.A. Data mining classification experiments with decision trees over the forest covertype database. In Proceedings of the 21st International Conference on System Theory, Control and Computing, Sinaia, Romania, 19–21 October 2017. [CrossRef]
27. Awad, N.A.; Mahmoud, A. Analyzing customer reviews on social media via applying association rule. *Comput. Mater. Contin.* **2021**, *68*, 1519–1530. [CrossRef]
28. Liu, Y.; Beldona, S. Extracting revisit intentions from social media big data: A rule-based classification model. *Int. J. Contemp. Hosp. Manag.* **2021**, *33*, 2176–2193. [CrossRef]
29. Gao, H.; Zeng, X.; Yao, C. Application of improved distributed naive Bayesian algorithms in text classification. *J. Supercomput.* **2019**, *75*, 5831–5847. [CrossRef]
30. Jiang, S.Y.; Pang, G.S.; Wu, M.L.; Kuang, L.M. An improved K-nearest-neighbor algorithm for text categorization. *Expert Syst. Appl.* **2012**, *39*, 1503–1509. [CrossRef]
31. Zhang, L. Implementation of classification and recognition algorithm for text information based on support vector machine. *Int. J. Pattern Recogn.* **2020**, *34*, 2053005. [CrossRef]
32. Chen, C.W.; Tseng, S.P.; Wang, J.F. Outpatient text classification system using LSTM. *J. Inf. Sci. Eng.* **2021**, *37*, 365–379. [CrossRef]
33. Liu, M.Q.; Liu, L.Z.; Cao, J.Y.; Du, Q. Co-attention network with label embedding for text classification. *Neurocomputing* **2022**, *471*, 61–69. [CrossRef]
34. Hinton, G.E.; Salakhutdnov, R.R. Reducing the dimensionality of data with neural networks. *Science* **2006**, *313*, 504–507. [CrossRef] [PubMed]
35. Kim, Y. Convolutional neural networks for sentence classification. *arXiv* **2014**. [CrossRef]
36. Liu, P.; Qiu, X.; Huang, X. Recurrent neural network for text classification with multi-task learning. In Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence, New York, NY, USA, 9–15 July 2016. [CrossRef]
37. Devlin, J.; Chang, M.-W.; Lee, K.; Toutanova, K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Minneapolis, MN, USA, 2–7 June 2019. [CrossRef]
38. Chen, X.; Cong, P.; Lv, S. A Long-Text Classification Method of Chinese News Based on BERT and CNN. *IEEE Access* **2022**, *10*, 34046–34057. [CrossRef]
39. Jiang, X.C.; Song, C.; Xu, Y.C.; Li, Y.; Peng, Y.L. Research on sentiment classification for netizens based on the BERT-BiLSTM-TextCNN model. *PeerJ Comput. Sci.* **2022**, *8*, e1005. [CrossRef]

40. Weng, X.; Zhao, J.; Jiang, C.; Ji, Y. Research on sentiment classification of futures predictive texts based on BERT. *Computing* **2021**. [CrossRef]
41. She, X.; Chen, J.; Chen, G. Joint Learning With BERT-GCN and Multi-Attention for Event Text Classification and Event Assignment. *IEEE Access* **2022**, *10*, 27031–27040. [CrossRef]
42. Zhao, J.; Li, C. Research on the Classification of Policy Instruments Based on BERT Model. *Discret. Dyn. Nat. Soc.* **2022**, *2022*, 6123348. [CrossRef]
43. Liao, Z. The evolution of wind energy policies in China (1995–2014): An analysis based on policy instruments. *Renew. Sustain. Energy Rev.* **2016**, *56*, 464–472. [CrossRef]
44. Shen, W. Who drives China's renewable energy policies? Understanding the role of industrial corporations. *Environ. Dev.* **2017**, *21*, 87–97. [CrossRef]
45. Cheng, Q.; Yi, H. Complementarity and substitutability: A review of state level renewable energy policy instrument interactions. *Renew. Sustain. Energy Rev.* **2017**, *67*, 683–691. [CrossRef]
46. Liu, X.; Li, J.; Han, L.; Zhou, B. Empirical analysis of the role of new energy transition in promoting china's economy. *Front. Environ. Sci.* **2022**, *10*, 955730. [CrossRef]
47. Wang, X.; Li, C.; Shang, J.; Yang, C.; Zhang, B.; Ke, X. Strategic Choices of China's New Energy Vehicle Industry: An Analysis Based on ANP and SWOT. *Energies* **2017**, *10*, 537. [CrossRef]
48. Li, X.; Peng, Y.; He, Q.; He, H.; Xue, S. Development of New-Energy Vehicles under the Carbon Peaking and Carbon Neutrality Strategy in China. *Sustainability* **2023**, *15*, 7725. [CrossRef]
49. Qiu, L.; Yang, D.; Hong, K.; Wu, W.; Zeng, W. The Prospect of China's Renewable Automotive Industry Upon Shrinking Subsidies. *Front. Energy Res.* **2021**, *9*, 661585. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.