NLP based Research on Traditional Energy Trade of the "Belt and Road" Energy Cooperation Partnership Countries

Jun Gong [1], Duoyong Sun [1*], Liang Feng [1]

[1] National University of Defense Technology, Changsha, Hunan, 410073, China
*duoyongsun@sina.com

Abstract. With the establishment of the "Belt and Road" energy cooperation partnership in 2019, the opportunities for improving China's energy supply structure are broader. Taking the traditional energy trade of partner countries as the entry point, relevant public data is collected, and natural language processing (NLP) technology is used to analyze and process it using the PEGASUS text summary model. The main conclusions include: (1) Currently, there is a structural imbalance in China's traditional energy trade. In the partnership, countries with abundant oil and gas resources such as Iraq, Kuwait, and Venezuela have become the main targets of China's traditional energy trade; (2) The fluctuation of multilateral exchange rates has a significant inhibitory effect on China's traditional energy imports, with the strongest inhibitory effect being the imbalance of the unilateral exchange rate of the RMB; (3) The energy itself has a significant import promoting effect on China's traditional energy imports, and the construction of international energy pipeline connectivity also has a significant promoting effect on this. On the one hand, it provides a foundation for the improvement of China's traditional energy trade policy, and on the other hand, this study is a beneficial exploration of using NLP technology to analyze social science issues, providing new ideas and insights for the research of China's traditional energy trade.

Keywords: NLP, The Belt and Road, Energy Cooperation Partnership, Traditional Energy Trade.

1 INTRODUCTION

1.1 A Subsection Sample

In 2022, China's total energy consumption reached 5410 million tons of standard coal, of which coal, oil and natural gas accounting for 82.5%. [1] In 2023, China imported a total of 3624.4 billion yuan of traditional energy such as coal, oil, and natural gas, accounting for 20.15% of the total import volume for the year. [2][3] With the continuous development of the Chinese economy, the demand for energy is further increasing, and the peak demand for primary energy will arrive around 2030. [4] China's energy industry has long faced low domestic supply capacity and high external depend

V. Vasilev et al. (eds.), *Proceedings of the 2024 5th International Conference on Management Science and Engineering Management (ICMSEM 2024)*, Advances in Economics, Business and Management Research 306,
https://doi.org/10.2991/978-94-6463-570-6_52

ence; the overseas energy source is single, and the potential supply risk is high; A series of problems such as low transportation capacity and lack of security in transportation channels. [5]

In April 2019, during the second "the Belt and Road" Summit Forum, China and 29 countries jointly launched "the Belt and Road" Energy Cooperation Partnership (hereinafter referred to as the "Partnership") in Beijing. [6] The establishment of the partnership provided new opportunities for strengthening foreign energy trade and safeguarding national energy security. Studying the traditional energy trade issues of partner countries is of great significance for improving China's energy trade policies, optimizing the energy supply structure, and ensuring national energy security.

Natural Language Processing (NLP) is one of the branches of computer linguistics and an important branch of Artificial Intelligence (AI). It mainly studies how to enable computers to understand, process, and generate human natural language. With the rapid development of various Internet platforms and the further maturity of big data and artificial intelligence technology. NLP technology has had a wide and far-reaching impact in various fields of humanities and social sciences in the context of disciplinary integration.

This article attempts to use NLP technology to process and analyze relevant texts on traditional energy trade in partner countries from a large amount of open-source data, summarize the issues of traditional energy trade in partner countries, in order to provide a basis for the improvement of China's traditional energy trade policies. At the same time, introducing NLP technology into the analysis of traditional energy trade issues in partner countries provides new research ideas and insights for interdisciplinary research in China's energy trade field.

2 NLP TECHNOLOGY

2.1 The Development History of NLP

In his book "Speech and Language Processing," Daniel Jurafsky, a renowned professor at Stanford University's School of Computer Science, divides the development of NLP into six stages(Table 1):

Basic research. From the 1940s to the 1950s; Two camps. From 1957 to 1970, NLP developed into two camps: Symbolists and Randomists; Four paradigms. From 1970 to 1983, there were four research paradigms in NLP: random paradigm, logic based paradigm, natural language understanding paradigm, and discourse model paradigm. The revival of empiricism and finite state models. From 1983 to 1993, NLP returned to empiricism and finite state models. The convergence of different fields. From 1994 to 1999, research in some subfields of NLP gradually began to merge; The rise of machine learning. Starting from the 21st century, empiricism has accelerated its development. [7]

Table 1. The Six Development Stages of NLP

Time	Stage	Important figures or institutions
1940-1950	Basic research	Turing, Shannon, Chomsky
1957-1970	Two camps	Harris, Newell, Simon
1970-1983	Four paradigms	Thomas J Watson, Baker, Colmerauer
1983-1993	The revival of empiricism and finite state models	Kaplan, Kay, Church
1994-1999	The convergence of different fields	-
2000-	The rise of machine learning	LDC

2.2 NLP Applications

With the development and maturity of NLP technology, the application scope of NLP is becoming increasingly widespread. From the initial stages of machine translation, referential resolution, automatic question answering, text classification and clustering, information extraction, to sentiment analysis, text generation, text summarization, and then to intelligent search, chatbots, speech recognition, and more, the application fields of NLP continue to expand. This study mainly uses text summarization technology.

Text summarization is a concise, accurate, and comprehensive summary of the original text content, and is an important research direction in the field of NLP. It aims to extract key information from the text, remove redundancy and secondary content, form a concise text form, and help readers quickly understand the core content of the original text. The main principle is to extract key information from the text, such as themes, keywords, entities, etc., through processing such as word segmentation, part of speech tagging, syntactic analysis, and semantic understanding. Then, based on certain algorithms or rules, these key information are filtered, sorted, and combined to generate the final summary.

3 DATA ACQUISITION

3.1 Data Sources

The data sources include: Academic research literature; Official report announcement; Think tank research report; News media coverage(Table 2).

Table 2. Data collection source

Category	Chinese	Foreign language
Academic research literature	CNKI, Wanfang, VIP	IEL, WOS, EIsevier
Official report announcement	The National Energy Administration, the Ministry of Commerce, the National Bureau of Statistics	WB,WTO,IMF,UNCD
Think tank re-	China Petroleum Economic and Technological Re-	WRI,BP

search report	search Institute, China Energy Strategy Research Institute	
News media coverage	Xinhua News Agency, People's Daily, Guangming Daily, Economic Daily	CNN,AP,RNA,AFP

3.2 Collection Scope

According to the research object, the scope of data collection will be determined in the text data related to traditional energy trade in the partner countries since the establishment of the partnership in April 2019.

3.3 Collection Method

This study mainly uses the Scrapy web crawler framework for data crawling. Scrapy is a powerful Python crawler framework, whose core components include the Engine, scheduler, downloader, spider, and Item Pipeline(Fig. 1).

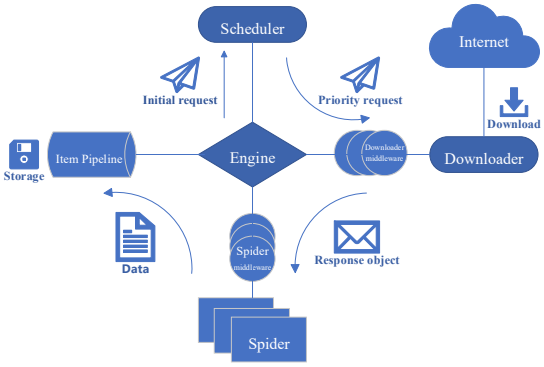

Fig. 1. Scrapy crawler principle

Based on the above principles, this study uses the keywords of "the Belt and Road Energy Partnership", "traditional energy trade", "coal trade", "oil trade", "natural gas trade", etc., and uses Scrapy to crawl the relevant text data of the target data source, including time, title, text, etc.

3.4 Data preprocessing

For text data crawled from the network, it cannot be directly used for text summarization models for calculation. The data needs to be preprocessed, and the specific steps include: translation conversion, data cleaning, cut words, remove stop words.

4 EMPIRICAL ANALYSIS

4.1 Statistical Analysis

Source statistics. The sources of data crawled by the statistics center are as Table 3 shows:

Table 4. Distribution of data source channels

Channel type	Source channel	Quantity	Proportion	Subtotal
Academic research literature	Wanfang	59	18.4%	20.2%
	CNKI	6	1.9%	
Official report announcement	Bureau of Statistics	11	3.4%	8.5%
	National Energy Administration	10	3.1%	
	China's the Belt and Road Network	4	1.2%	
	Belt and Road Cooperation Network	1	0.3%	
	Ministry of Commerce	1	0.3%	
Think tank research report	WRI	24	7.5%	17.1%
	WB	23	7.2%	
	IMF	5	1.6%	
	China Energy Strategy Research Institute	1	0.3%	
	WTO	1	0.3%	
	BP	1	0.3%	
News media coverage	Economic Daily	74	23.1%	54.2%
	Guangming Daily	56	17.5%	
	People's Daily	30	9.4%	
	Xinhua News Agency	11	3.4%	
	AP	1	0.3%	
	CNN	1	0.3%	

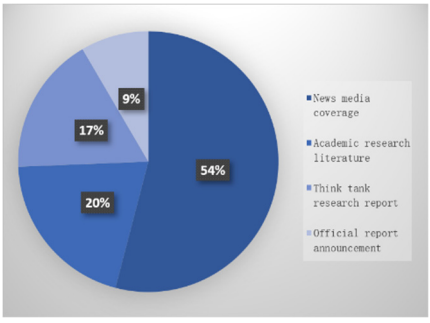

Fig. 2. Distribution of data source types

From Fig. 2, it can be intuitively observed that the majority of research related data comes from news reports, accounting for more than half, reaching 54.2%. Among them, Economic Daily has 74 texts, accounting for 23.1% of the total, and is the larg-

est source of research data. The media still reporting it as an economic issue. At the same time, there are a total of 65 texts in academic research literature, accounting for 20.2%, making it the second largest source of research data. On the other hand, crawling a total of 56 foreign language data, accounting for 17.5%, is significantly less than Chinese data, indicating that the attention of partner countries to traditional energy trade in the international community is not high enough, and there is still a long way to go in promoting its internationalization.

Word frequency count. Using the word cloud generation tool Weiciyun, the preprocessed text data is generated into a word cloud based on the frequency of words in the text, as Fig.3 shown:

Fig. 3. Text word cloud

As can be seen from the figure, the words with high frequency in the text include "the Belt and Road", "energy cooperation partnership", "consensus building", "mutual benefit and win-win", "five links", "China", "trade", "oil", "natural gas", etc., which reflects the characteristics of traditional energy trade of partnership countries, such as taking China as the core, adhering to the principle of equal cooperation, mutual benefit and win-win results.

4.2 Text summary based on PEGASUS model

Model selection. There are currently many models used for text summarization, which are trained based on different text summarization corpora. The trained models have varying effects on different types of target texts. The commonly used method for automatic evaluation is ROUGE (Recall Oriented Study for Gisting Evaluation). The ROUGE evaluation indicators mainly include two indicators: R-N and R-L.

By comparing the training corpora and ROUGE evaluation scores of different models as Table 4 shows, this study chose the PEGASUS model. The score of R-1, R-2, and R-L trained on the Chinese corpus LCSTS can reach 43.46, 29.59, and 39.76, which is better than other models and suitable for this study.

Table 5. Performance comparison of some models on LCSTS

Model	R-1	R-2	R-L
RNN-context[8]	29.90	17.40	27.20
CopyNet[9]	34.40	21.60	31.30
DRGD[10]	36.99	24.15	34.21
RNN+MRT[11]	38.20	25.20	35.40
Seq2Seq+CGU[12]	39.40	26.90	36.50
RTC-WL[13]	39.93	21.58	37.92
PEGASUS[14]	**43.46**	**29.59**	**39.76**

Environmental configuration. This study conducted experiments on Lenovo computers configured with Intel(R) Core(TM) i5-7200U CPU，Intel(R) HD Graphics 620 GPU，8GB of running memory and installed Python 3.12 environment on a 64 bit professional version of Windows 10 operating system.

Model training. This study used the LCSTS from HIT to train the text summarization model. The corpus of this dataset comes from Sina Weibo and consists of over 2 million authentic Chinese short texts, each of which is accompanied by a brief summary by the author. In addition, there are 10666 manually marked relevant short texts and corresponding short summaries.[15]

This study trained the model using data with scores of 4 and 5 in the second part of the LCSTS dataset, and used data pairs with scores of 4 and 5 in the third part as the test set. The evaluation scores of the trained model for R-1, R-2, and R-L were 42.19, 27.64, and 38.41, respectively. Although they did not reach the best score, they were still at a high level, indicating good model performance.

Result analysis. By inputting preprocessed data into the trained PEGASUS model and analyzing the output results, the following conclusions can be drawn:

Firstly, there is currently a structural imbalance in China's traditional energy trade. In the partnership, countries with abundant oil and gas resources such as Iraq, Kuwait, and Venezuela have become the main targets of China's traditional energy trade. Secondly, multilateral exchange rate fluctuations have a significant inhibitory effect on China's traditional energy imports, with the strongest inhibitory effect being the imbalance of the RMB unilateral exchange rate. Thirdly, energy itself has a significant import promoting effect on China's traditional energy imports, and the construction of international energy pipeline connectivity also has a significant promoting effect. Fourthly, in the short term, the petrodollar still holds an important position in the international traditional energy trade market, while the fluctuation of the RMB/USD exchange rate remains a key reference in international investment decisions. Fifthly, China should further expand the settlement scope and share of the RMB under the current account in energy trade with partner countries, promote the RMB oil strategic

layout, strengthen energy trade dialogue with countries such as Iraq, Kuwait, and Venezuela, deepen international energy trade cooperation, and promote the diversification of China's energy investment regions.

5 CONCLUSION

Energy has always been an important issue related to national development and people's livelihood. At present, traditional energy still accounts for a large proportion of China's energy consumption. The establishment of the "the Belt and Road" energy cooperation partnership provides a new opportunity to solve the structural contradiction of China's energy supply through energy trade. This study collected public text data since the establishment of the partnership and summarized the traditional energy trade issues of partner countries using the PEGASUS text summary model. It is an exploration of using NLP technology to analyze social science issues. Due to the limited quantity and quality of research object data, as well as the current level of NLP technology, there is still significant room for improvement in the output results. With the further development of AI and the gradual maturity of NLP technology, better performing text summarization models will inevitably emerge; At the same time, data on traditional energy trade among partner countries will gradually accumulate, and subsequent research will achieve better results.

ACKNOWLEDGEMENT

The study is supported by the Key Project of Military Theory Research (Project Approval No. 23-JSLLKY-003).

REFERENCES

1. National Bureau of Statistics, https://data.stats.gov.cn/easyquery.htm?cn=C01, 2024/02/17.
2. The General Administration of Customs of the People's Republic of China, http://www.customs.gov.cn/customs/302249/zfxxgk/2799825/302274/302277/302276/5637019/index.html, 2024/01/18.
3. The General Administration of Customs of the People's Republic of China, http://www.customs.gov.cn/customs/302249/zfxxgk/2799825/302274/302277/302276/5637005/index.html, 2024/01/18.
4. Wang Anjian, Gao Xinrui: Outlook on China's Energy and Important Mineral Resource Demand. Journal of the Chinese Academy of Sciences 35(3),338-344 (2020).
5. Wang Anjian, Wang Chunhui: Challenges and Response Strategies of International Turbulence on China's Energy and Resource Security. Journal of the Chinese Academy of Sciences 38(1), 72-80 (2023).
6. China's Belt and Road Network, https://www.yidaiyilu.gov.cn/p/87664.html, 2019-04-25.
7. Daniel Jurafsky,James H. Martin: Speech and Language Processing. Electronic Industry Press,bejing(2018).

8. Hu Baotian,Chen Qingcai,Zhu Fangze: LCSTS:A large scale Chinese short text summarization dataset. In: Proc of the 54th Annual Meeting of the ACL,pp.1967-1972. ACL, Stroudsburg (2015).
9. Gu Jiatao,Lu Zhengdong,Li Hang: Incorporating copying mechanism in sequence-to-sequence learning. In: Proc of the 54th Annual Meeting of the ACL,pp.1631-1640.ACL, Stroudsburg(2016).
10. Li Piji,Lam W,Bing Lidong: Deep recurrentGenerative decoder for abstractive text summarization. In: Proc of the 2017 Conf on Empirical Methods in Natural Language Processing,pp. 2091-2100.ACL, Stroudsburg(2017).
11. Neural headline generation with sentence-wise optimization, https://arxiv.org/pdf/1604.01904.pdf, 2021/02/10.
12. Lin J,Sun X,Ma S:Global encoding for abstractive summarization. In: Proc of the 56th Annual Meeting of the ACL,pp. 163-169.ACL, Stroudsburg(2018).
13. Li W,Yao J,Tao Y:A reinforced topic-aware convolutional sequence-to-sequence model for abstractive text summarization. In: Proceedings of the 27th International Joint Conference on Artificial Intelligence,pp. 4453-4460. ACM, New York(2018).
14. Mohammad Saleh,Peter J.Liu.PEGASUS:Pre-training with extracted gap-sentences for abstractive summarization, https://arxiv.org/pdf/1912.08777v3.pdf, 2021/02/10.
15. Baotian Hu，Qingcai Chen，Fangze Zhu：LCSTS:A Large Scale Chinese Shoa Text Summarization Dataset. In: Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing,ACL, Lisbon(2009).