An Overseas Sentiment Analysis of Xi'an City Based on Deep Learning and NOW Corpus

Jingyu Guan* and Xiangqian Che

Abstract—With the development of corpus linguistic and the inter-disciplinary application of deep learning, the quantitative research on the cross-cultural communication of city image is becoming more and more popular recently. The purpose of this study is to reveal the current communicative situation of Xi'an City on English online news medias and overseas audiences. This paper uses the real-time English online news Corpus NOW and public opinion texts drawn from Twitter as the research data. Firstly, this study analyzes the text data from the perspective of Word Frequency, Keywords, Key Word In Context (KWIC) and Collocation. Then sentiment analysis based on a pre-trained Natural Language Processing (NLP) model Bidirectional Encoder Representations from Transformers (BERT) is used to analyze overseas audiences' attitude toward Xi'an City. The results find that Xi'an City is less reported by overseas online new medias than other 14 new first-tier cities in China. Southeast and Central Asian countries are the main reporters of Xi'an City. International society mainly concentrates on Xi'an City's national defense industry (especially military aircraft), followed by its long history and economy. Besides, 46.2% of overseas audiences keeps neutral toward Xi'an, 30.4% have a negative attitude, and another 23.4% have a positive attitude. According to the research result, this study proposes three practical suggestions to help enhance Xi'an City's international influence effectively.

Index Terms—Deep learning, sentiment analysis, NOW corpus, city image, Xi'an

I. Introduction

City image is the citizen's overall cognition and impression to a city's natural geography, industry structure, economic level, social security, architectural landscape, governance mode, historical traditions and cultural values. A City's overseas image is mainly disseminated through online mass media on 21st century [1]. The rapid development of big data and the prevalence of online news medias have made online intercultural communication more important than ever before. The overseas communicative study of City image is becoming a hot research spot in media studies and corpus linguistics [2].

This paper sets Xi'an City as its research subject. As the largest city in northwestern China and a world-wide historical city selected by UNESCO, Xi'an City is the ancient capital of China for Six Dynasties. Xi'an City is famous for its rich culture resources such as the Epang Palace, the Terracotta Army, the Great Wild Goose Pagoda, Wei-yang Palace, ChangLe Palace, the DaXing Wall and the DaMing Palace. To help Xi'an become more influential in international society, it is necessary to study Xi'an City's overseas dissemination situation. This study can help promote Xi'an

City's international popularity and strengthen its communication to oversea audience.

This paper use sentiment analysis and corpus method to analyze Xi'an City's overseas influence. As an important application of Natural Language Processing (NLP) based on deep learning, sentiment analysis is a significant method in obtaining and analyzing massive internet users' public opinions, thus providing suggestions for governments and enterprises. Sentiment analysis is mainly realized by three methods: sentiment dictionary, machine learning and deep learning. In the field of NLP, sentiment analysis based on deep learning is the most accurate and advanced method so far in achieving sentiment classification of the targeted corpus. It uses multilayer neural networks to extract language features and ensures its high accuracy through pre-trained algorithm models. Compared with traditional methods, it can identify special semantic phenomena such as polysemy, irony and metaphor.

Considering the important status of Xi'an City in China and the necessity of filling the research gap which study city image with the help of NLP, this study aims at analyzing Xi'an City's overseas dissemination situation using corpus method and sentiment analysis technique. Three practical suggestions are proposed at the end of this paper to help enhance Xi'an City's overseas influence.

II. Literature Review

According to CNKI, there are a total of 1920 studies related to city image dissemination, and the number of those studies is increasing year by year. Among them, representative empirical studies include Li's study [3] on the international image dissemination of Jiangsu Province based on NOW corpus; Liu's study [4] on the international image of China's new first tier cities based on NOW corpus; and Li's study [5] on the international image of Guangzhou city from the perspective of English media discourse based on NOW corpus. However, most studies on city image are qualitative, represented by Tang's study [6] on the image of Shenzhen city from the perspectives of overseas expert opinions and public perception. Du's research [7] on the city images of Chongqing City, Xi'an City and Chengdu City on the short video App Tiktok.

To sum up, the existing research on city images is mainly qualitative, and most of its data come from some small corpus collected by researchers individually, and the analysis process heavily relies on the researcher's personal experience and reasoning, lacking empirical support, which will lead to subjective results. Therefore, the corpus-driven quantitative research combined with sentiment analysis can effectively overcome the subjective deficiency of traditional qualitative research on city images, which is also an innovation of this

research.

III. METHODOLOGY

Taking text data set from NOW corpus and Twitter as the research data, this study analyzes Xi'an City's overseas image shaped by English online news media and overseas audience's attitude toward Xi'an City through online corpus and sentiment analysis techniques. Then this paper proposes three suggestions for improving Xi'an City's overseas image. This study mainly focuses on the following two questions: What is the overall situation of Xi'an City's overseas image from 2010 to 2023 according to the research data? How can Xi'an City promote its international influence in international society? Based on the two research questions, the research procedure of this study is shown in the Table I.

TABLE I: RESEARCH PROCEDURE

Procedure	Content
Data Collection	This paper collects research data about Xi'an City from NOW Corpus and public opinion texts data set from Twitter.
Data Analysis	1. Wordlist analysis: this paper set Xi'an City as the search subject. With the help of "chart" function on NOW corpus, this paper firstly analyzes the annual changes of news reports about Xi'an City from 2010 to 2023 in diachronic level. Then this paper compares the amount of news reports about Xi'an City with China's other 14 new first-tier cities in synchronic level. Finally, this paper studies these online news reports' source countries to reveal the influence of Xi'an City in different countries or regions. 2. Collocation analysis: this paper firstly reveals some fields that English online news medias care most through the top 30 collocations with Xi'an City. Then this paper randomly selects about 300 KWICs to analyze Xi'an City's overseas image from two perspectives.
Sentiment Analysis	Based on a pre-trained NLP model BERT, this paper performs a sentiment analysis on the data set (collected from Twitter) in Python environment to reveal the overseas audience's sentiment towards Xi'an City.
Making suggestion	According to the research results, this paper proposes three practical suggestions to promote the overseas image of Xi'an City.

IV. DATA ANALYSIS AND DISCUSSION

A. Corpus Analysis Based on NOW

NOW corpus (News on the Web) is an English online news corpus developed by Brigham Young University in the United States, and is currently the most widely covered and real-time English online news corpus in the world. As of May 27, 2023, the NOW corpus contains 16.9 billion news from English online news medias (including famous newspapers, magazines, and social media) around the world, and continues to expand its capacity by adding 180-220 million words each month. Compared with small traditional corpora built individually by earlier researchers, NOW corpus can reflect the real-time focus of the current world reported in English online news medias. This part analyzes Xi'an City's overseas image on English online news medias from two perspectives: word list and collocation.

1) Wordlist analysis

Word frequency is an indicator of the overseas influence and popularity of Xi'an City in English online news medias. This part reveals the image of Xi'an City shaped by English online news media in synchronic and diachronic level. From 2010 to 2023, according to NOW corpus, the word frequency of "Xi'an City" and other 14 China's new first-tier cities is shown in Table II. Diachronically, the news reports about Xi'an City from those medias have annual changes, and is embodied in Table III.

TABLE II: CHINA'S 14 NEW FIRST-TIER CITIES' WORD FREQUENCY

City	Word frequency
Wuhan	118,099
Tianjin	17,528
Chengdu	16,359
Hangzhou	16,016
Chongqing	13,021
Nanjing	11,692
Qingdao	8,908
Suzhou	5,957
Xi'an	5,718
Zhengzhou	4,523
Ningbo	3,955
Shenyang	3,630
Changsha	3,288
Dongguan	2,975
Foshan	1,868

TABLE III. ANNUAL CHANGES OF THE NEWS REPORTS ABOUT XI'AN CITY

Year	Freq	Rate
2010	80	null
2011	57	−0.29
2012	94	0.65
2013	121	0.29
2014	152	0.26
2015	155	0.02
2016	487	2.14
2017	523	0.07
2018	672	0.28
2019	844	0.26
2020	952	0.13
2021	949	−0.01
2022	1,418	0.49
2023	115	null

In terms of word frequency of Xi'an City and other fourteen new first-tier cities in China, we can see that Xi'an City ranks ninth in Table II. It indicates that compared with cities of the same level in China, Xi'an City's international influence is not quite enough although it has abundant historical resources and transportation advantages. Some efforts are still needed to enhance Xi'an City's international popularity.

It can be seen from Table III that the word frequency about Xi'an City shows a general upward trend from 2010 to 2022, and reach its peak at the end of 2022. The year 2012, 2016 and 2022 have witnessed a highest growth rate. It may be attributed to Xi'an City's "Silk and Road International Tourism Fair" on 2016, the convening of China's 2022 National Olympic Games and the sudden outbreak of covid-19 at the beginning of this year. In summary, Xi'an City's overseas influence and popularity in those online medias has shown a fluctuating but overall increasing trend in recent years.

With the help of "Chart" function on NOW corpus, this part retrieves these online media's source countries, and visualized the results in Table IV.

As can be seen in Table IV, Philippines, the United States,

Singapore and Malaysia are the main reporters of Xi'an City. In addition, Canada, the United Kingdom, Ireland, India and Australia also show great interest in Xi'an City. In summary, it can be seen that Southeast Asian countries pay the most attention to Xi'an City, while European, African and South American countries care the least. Xi'an City's international popularity and influence are imbalanced on international society, and there is an urgent need to enhance its international popularity outside Southeast Asia.

TABLE IV. SOME REGIONS/COUNTRIES' REPORT OF XI'AN CITY

Countries/regions	Freq.
The United States	1919
Canada	341
Britain	559
Ireland	322
Australia	304
New Zealand	191
Singapore	991
Philippines	231
Malaysia	186
Hong Kong	374
South Africa	160
Nigeria	96
Ghana	50
Kenya	72
India	384
Pakistan	160
Bangladesh	42
Tanzania	5

2) Collocations analysis

Collocation refers to those words that frequently co-occur with the target word, which can be found in the context of the target word. Collocation is one of the most important methods in the corpus-driven research paradigm. In this part, Xi'an City is set as the target word, and we choose collocations that are within 5-character units (span) of the target word from NOW corpus. The top 30 collocations with the highest frequency of co-occurrence with "Xi'an City" are selected, and the result is visualized and shown in Table V.

From the top 30 collocations co-occur with Xi'an City on English online news media, it can be seen that the words "capital", "city", "China" and "province" objectively describe the fact that Xi'an City is the capital city of Shaanxi Province in China. In addition, the high frequency collocations such as "aircraft", "bomber" and "Y-20" reveal that Xi'an City is closely associated with national defense and military weapons in the eyes of overseas online media. The industry of military aircraft manufacturing has attracted those media's extensive attention in particular. Besides, collocations such as "ancient", "Gong", "terracotta", "warriors" and "Tang" show the long history of Xi'an City and its rich cultural tourism resources. The collocations such as "Northwestern" and "transport" show that Xi'an City, as the largest city in northwestern China, is an important transportation center connecting the East and West of China. The collocations like "outbreak", "infections" and "lockdown" refer to the sudden outbreak of COVID-19 in Xi'an City at the beginning of 2022. Finally, the collocations such as

"manufacturing" and "holdings", which are related to economic and industrial development, are less frequently co-occurring with Xi'an City. In summary, from the news reports of overseas online news media, Xi'an City is best known as the capital city of Shaanxi Province in China. military aircraft and national defense activities in Xi'an City has also attracted wide attention. Xi'an City's long history, culture, tourism resources, its economic development and urban governance is the least noticed. In addition, the sudden outbreak of COVID-19 in Xi'an City at the beginning of 2022 has also led to a lot of reports in the oversea online news media.

TABLE V. TOP 30 COLLOCATIONS CO-OCCURRING WITH XI'AN

Word	Freq
city	379
China	311
Chinese	193
capital	102
province	87
Gaza	64
northwestern	62
aircraft	59
ancient	53
lockdown	52
actor	50
bomber	48
terracotta	43
Y-20	38
chip	33
transport	33
outbreak	31
enter	31
consultancy	29
strategic	29
Gong	25
Tang	24
infection	24
manufacturing	24
tourism	24
village	24
warriors	23
slow	21
disrupt	21
holdings	21

Keywords in the Context (KWIC) refers to those words that usually appear in the context of the target word. Compared with collocations which are the nearest ones to the target word, KWICs are more distant from the target word. Based on NOW corpus, this study sets "Xi'an City" as the keyword and searches for its nearby context (within 20 words). Finally, nearly 300 KWICs are randomly selected by NOW corpus, and are categorized into three groups according to their frequency. The research result is classified and shown in Table VI.

Removing all the functional words, the remaining words are mainly nouns and verbs. Nouns usually perform the linguistics function of conveying information. Therefore, the nouns in Table 6 show that oversea online news media mainly focus on Xi'an City in the following areas: country, government, officials, research, women, tourism, provincial capital, media, lifestyle, northwest China, Terracotta Warriors, epidemic, history and culture, transportation, rural migration. It can also be concluded that there are little adjectives in these KWICs, which shows that most of the news reports from English online news medias are only descriptive of facts without strong emotions. These KWICs are further classified according to their characteristics in

describing a city, and the result is shown in Table VII.

TABLE VI: CLASSIFICATION RESULT 1 OF KWICS

Low Freq			
adventure	cosmic	yuan	sub-variant
astrophysicist	cousins	outbreaks	talent
aligns	flare-up	precisely	translated
anxiety	galaxies	province	updated
ambassador	imposed	provoke	verifying
betting	Infections	psychologist	warrants
brand	lifestyle	repay	well-rounded
chores	newsletters	quarantined	whenever
clinical	Northwestern	statues	willingness
Mid Freq			
travel	camera	event	media
ability	capital	extending	picture
arrest	chosen	famous	rays
host	claim	Immediately	slip
camp	contain	invited	target
billion	dancing	justice	titled
	despite	launched	tough
High Freq			
city	during	lot	probably
more	found	love	read
reported	friend	million	research
able	government	next	social
added	head	noted	speak
against	how	officials	stay
areas	kids	outside	understood
case	leave	people	watch
Country	line	play	woman
drive	local	present	year

TABLE VII: CLASSIFICATION RESULT 2 OF KWICS

Modern Metropolis (110 words)	Historical City (133 words)
Developed Trade Industry (17)	Historical Sites (78)
Internationalization (8)	Traditional Costume (10)
Convenient Transportation (34)	Food Culture (34)
High Technology (33)	Folk Art (11)
Youthfulness (18)	

It can be seen from Table VII that the image of Xi'an City is a composite of "modern metropolis" and "historical city". The characteristics of modern metropolis can be reflected in the KWICs that relates to commerce, international popularization, convenient transportation, high technology and vitality. The feature of a historical city can be reflected through KWICs like historical relics, traditional costumes, food culture and folk art. These multi-dimensional characteristics of Xi'an City allows it to maintain its old historical city image while becoming an international modern city with good developing prospects.

B. Sentiment Analysis Based on Deep Learning

The development of artificial intelligence has triggered a revolution in almost all the disciplines including traditional humanities. Nowadays, the majority of AI researches are focusing on Computer Vision (CV) and natural language Processing (NLP). ChatGPT is the newest and the best-known application of NLP. Both CV and NLP are mainly realized by deep learning technology. As the main method to realize artificial intelligence, deep learning is an improvement of machine learning based on artificial neural networks, which can be supervised, semi-supervised or unsupervised. Natural language processing has several research directions such as knowledge graph, text classification, information extraction and sentiment analysis. Among them, sentiment analysis can greatly help linguistics researches and media studies. The main models of sentiment

analysis include: Convolution Neural Network (CNN), recurrent neural network, Long Short-Term Memory (LSTM), Transformer, BiLSTM and attention mechanism.

This study uses the model BERT to realize sentiment analysis. The coding environment is Python 3.8.1 and packages such as datasets, transformers, scikit learn, numpy, and pytorch are used. This research uses 1000 users' comments about Xi'an City from Twitter as the data set. After multiple times of training, this model can achieve about 76% precision for sentiment analysis of the data set, and the research result is displayed in the following code:

```
Loading cached processed dataset at tweet/train/cache-
bdce5f8f18b3530f.arrow
Loading cached processed dataset at tweet/test/cache-
e6df6939fe44d9aa.arrow
Loading cached processed dataset at tweet/
validation/cache-Oa4ec67050c65f78.arrow
Counter({1:462, 0:304, 2:234})
{'F1 Score':0.6663991641156145,
'Recall':0.6789574451244639,'Precision':0.768}
```

As a result, Xi'an City's image on the reports of English online news media is shown in Table VIII.

TABLE VIII: ATTITUDES TOWARD XI'AN CITY FROM OVERSEAS Audience

Attitude	Percentage
neutral	46.2%
negative	30.4%
positive	23.4%

As the most widely used and well-known online social software and micro-blog in the world, Twitter has millions of users worldwide, which can reflect all the current hot topics in the world. Therefore, Twitter users' online comments are a valuable data set for analyzing overseas audience's public opinion toward Xi'an City. This paper uses a deep learning model to realize sentiment analysis. This model will automatically choose and learn some language characteristics of inputted similar text data set and train itself so as to improve its accuracy and precision to predict the target texts. The data in Table 7 shows that about half of the overseas audiences using Twitter have a neutral attitude toward Xi'an City, about 30% have a negative attitude, and about 20% maintain a positive attitude. It reveals that majority of the overseas audiences keeps objective toward Xi'an City, but their negative emotion toward Xi'an City is slightly more than their positive one.

V. CONCLUSION

Having studied Xi'an City's overseas influence based on NOW corpus and sentiment analysis, this paper has found some important findings.

First of all, it can be seen from word frequency in Now corpus that Xi'an City's overseas influence is not as great as China's other fourteen new first-tier cities. Diachronically, from 2010 to 2023 the English online news reports about Xi'an City show a fluctuating upward trend and reaches its apex in 2023. Besides, the United States, the Philippines, Singapore and Malaysia are the main reporters of Xi'an City who show a great interest. Additionally, the English online

news medias mainly care Xi'an city in the following areas: firstly in its military aircraft industry, then national defense, historical culture resources, and finally its economic development. Ultimately, the result of sentiment analysis based on deep learning shows that nearly half of the overseas audiences' attitudes toward Xi'an City remain neutral, about 30% audiences have negative attitudes, and about 20% audiences have positive ones.

Based on the research results, this study proposes three practical suggestions to help further improve the international popularity and influence of Xi'an City.

Above all, research results show that, compared with other new first-tier cities in China, Xi'an City has less overseas influence and reputation on English online news medias. In response to this issue, the Xi'an City Municipal Government can establish official medias to speak for itself, and can also make creative short videos or city advertising video on Facebook, Twitter, Instagram, Tiktok, Bilibili and other medias. Secondly, Xi'an City can cooperate with overseas social medias to set up representative television programs, such as putting soft advertisements in well-known telecast or talk shows, and hold cross cultural communication activities for foreign nationals living in Xi'an City. Finally, universities in Xi'an City can encourage their international students to become communicators after graduation and returning to their countries.

Besides, the research results also show that English online news medias currently pay more attention on Xi'an City's national defense, military aircraft, historical sites, and economic development. In order to meet these medias interest, Xi'an City could advertise its advanced military research and introduce universities such as Northwest Polytechnic University, which is famous for national military aircraft industry. Secondly, Xi'an City could set up cultural IP and draw inspirations from its local Tang cultures and Three Qin cultures. Xi'an City can learn experiences from the Forbidden City in Beijing to conduct 3D modeling and visualize city landmarks, allowing overseas audiences to immerse themselves in the landscape of Xi'an City through virtual reality techniques. Finally, the government should propagate Xi'an City as the starting city of the ancient "Belt and Road", so that the world can understand Xi'an City's important role in ancient international trade.

Finally, the results indicate that Xi'an City has the most influence in Southeast Asia, and lacks popularity in South America and Africa. In response to this phenomenon, Xi'an City should promote its influence in some developing countries from South America and Africa. Xi'an City should adopt different communicative strategies according to the local culture and values of different countries. For these developing countries, Xi'an City should pay attention to transmits Xi'an City's urban governance experience, advertise its urban landscape and development methodology, and highlight its important status in the "Belt and Road initiative".

This research has great significance both in theory and practice. On the one hand, this study theoretically enriches the research paradigm of cross-cultural communication study about city overseas image driven by corpus and sentiment analysis based on deep learning. On the other hand, it practically proposes three feasible communicative suggestions to help enhance the overseas influence of Xi'an City. As a result, Xi'an City can attract more foreign investment, develop international trade and communicate its traditional culture to the international society. This study will not only help promote the overseas influence of Xi'an City, but also provide some references for other cities that are in similar situation with Xi'an City in the world.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

Jingyu Guan had conducted the research and analyzed the data; Xiangqian Che wrote the suggestion part and polished this paper. All authors had approved the final version.

REFERENCES

[1] A. Broudehoux, *Mega-Events and Urban Image Construction: Beijing and Rio de Janeiro*, Oxford: Taylor and Francis, 2017.
[2] Y. Sun, *Cross-Cultural Communication*, Peking University Press: 21st Century Journalism and Communication Planning Textbook Communication Series, 2021.
[3] C. Li and Y. Pan, "Research on the international image communication of Jiangsu based on NOW corpus," *Foreign Communication*, vol. 12, pp. 73–75, 2019.
[4] J. Liu, "Study on the international image of China's new first tier cities based on NOW corpus," *Youth Journalist*, vol. 12, pp. 44–45, 2020.
[5] X. Li, "Study on the international image of Guangzhou city from the perspective of English media discourse based on NOW corpus," *Journal of Guangzhou City College*, vol. 13, no. 4, pp. 91–96, 2019.
[6] L. Tang, "Study on the image of Shenzhen city from the perspectives of overseas expert opinions and public perception," *Journal of Shenzhen University*, vol. 37, no. 2, pp. 41–49, 2020.
[7] J. Du, "Research on the city image of Chongqing City, Xian City and Chengdu City on the short video App Tiktok," *Media*, vol. 15, pp. 82–84, 2019.