



Predicting the locations of missing persons in China by using NGO data and deep learning techniques

Anning Dong, Yatao Zhang, Zijin Guo, Peng Luo, Yao Yao, Jialyu He, Qia Zhu, Ying Jiang, Kailu Xiong & Qingfeng Guan

To cite this article: Anning Dong, Yatao Zhang, Zijin Guo, Peng Luo, Yao Yao, Jialyu He, Qia Zhu, Ying Jiang, Kailu Xiong & Qingfeng Guan (2024) Predicting the locations of missing persons in China by using NGO data and deep learning techniques, International Journal of Digital Earth, 17:1, 2304076, DOI: [10.1080/17538947.2024.2304076](https://doi.org/10.1080/17538947.2024.2304076)

To link to this article: <https://doi.org/10.1080/17538947.2024.2304076>



© 2024 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



Published online: 17 Jan 2024.



Submit your article to this journal [↗](#)



Article views: 1046



View related articles [↗](#)



View Crossmark data [↗](#)



Citing articles: 2 View citing articles [↗](#)



Predicting the locations of missing persons in China by using NGO data and deep learning techniques

Anning Dong^a, Yatao Zhang^b, Zijin Guo^a, Peng Luo^c, Yao Yao^{ib a,d}, Jialyu He^e, Qia Zhu^a, Ying Jiang^a, Kailu Xiong^a and Qingfeng Guan^a

^aSchool of Geography and Information Engineering, China University of Geosciences, Wuhan, People's Republic of China; ^bFuture Resilient Systems, Singapore-ETH Centre, ETH Zurich, Singapore, Singapore; ^cChair of Cartography and Visual Analytics, Technical University of Munich, Munich, Germany; ^dCenter for Spatial Information Science, The University of Tokyo, Chiba, Japan; ^eDepartment of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University, Hung Hom, Hong Kong

ABSTRACT

Missing person crimes can seriously affect the well-being of Chinese families, and missing person destination prediction can help to solve this problem. Using nongovernmental organization (NGO) data to predict the locations of missing persons by random forest (RF) model has made progress. However, studies using these data have ignored the mass of oral information. Recent studies have demonstrated the effectiveness of oral information in detecting missing persons, but the impact on destination prediction remains unexplored. Therefore, this study proposes a missing person prediction (MP-Net) framework to incorporate oral information into missing person destination prediction and quantitatively describe the effect of different word properties on the prediction. The results show that compared to the baseline RF model, the proposed framework achieves a higher recall rate (87.18%) in the location prediction of missing persons. According to a quantitative word analysis, verbs and nouns in oral information significantly contributed to location prediction. After adjectives that might cause adverse effects were removed, the stability of the model was improved considerably. Overall, the findings of the proposed model and quantitative word analysis can help police or NGOs collect descriptive information in a targeted manner and make more accurate predictions about the whereabouts of missing persons.

ARTICLE HISTORY

Received 31 January 2023

Accepted 6 January 2024

KEYWORDS

Missing persons; location prediction; oral information; natural language processing; quantitative vocabulary analysis

1. Introduction

Human trafficking, also known as modern-day slavery, is a severe human rights violation (Judge et al. 2018) that often results in severe and lasting health consequences for its victims (Kiss and Zimmerman 2019). Since the significant types of population crimes usually faced by different countries are inconsistent, various countries have different definitions of human trafficking and different directions of research (Farrell and De Vries 2020). Globally, the most common types of crimes involving human trafficking are sexual and labor exploitation (Cockbain and Bowers 2019). In China, missing persons are often accompanied by illegal adoption crimes,

CONTACT Yao Yao ✉ yaoy@cug.edu.cn, yaoy@csis.u-tokyo.ac.jp School of Geography and Information Engineering, China University of Geosciences, Wuhan, Hubei Province 430078, People's Republic of China; Center for Spatial Information Science, The University of Tokyo, Chiba 277-8568, Japan

© 2024 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

implying that the group of missing persons consists mainly of minors (Xiong 2023). To combat these crimes, exploring the causes of disappearance cases with regional characteristics and constructing models to predict the whereabouts of missing people can be considered. These efforts are crucial for police departments to dispatch police officers, control crime and maintain social stability.

Previous studies have been conducted to explore the causes of missing cases (Cui et al. 2018; Shen, Antonopoulos, and Papanicolaou 2013). Initially, many studies employed public reports and cooperation with the Public Security Bureau to assess the impact of geographic characteristics and the economic environment on people disappearances (Li, Ren, and Zhang 2013; Shen, Antonopoulos, and Papanicolaou 2013; Wang 2015). These studies revealed an apparent spatial clustering phenomenon of the places people are abducted from and into, suggesting that the transfer of missing persons is organized. However, due to the limitations of data sources and scopes, the research area was confined mainly to a particular region, and experiments on a macro level for the whole country failed to be conducted. Compared to the authoritative data-based studies on missing persons in other countries, the number of such studies on missing persons in China is usually low due to the lack of publicly available government data (Shen, Antonopoulos, and Papanicolaou 2013; Sweileh 2018).

With the development of the Web and the efforts of related organizations, an increasing number of researchers tend to collect a mass of information from the Web and conduct research on a national scale. Using publicly available data from the China Judgments Online (CJO) website, Huang and Weng (2019) constructed five trafficking networks of missing persons by province, city, county, and gender. The research found that network density was tighter for males than for females, and the importance of noncapital cities with respect to trafficking crimes was growing. In addition, Wang et al. (2018) conducted a study on missing person transfer paths and analyzed possible criminal groups along the paths in a clustering manner to visualize regional hotspots and critical paths for child trafficking. For missing person prediction, Yao et al. (2021) used four random forest models to develop a prediction model that integrates the province, block, distance, and location of missing person transfer; the prediction model achieved a high top-10 recall of 82.99%. However, the study was limited by methodological constraints and failed to introduce natural language data, which contain descriptions of missing population status, to aid in predicting missing populations.

Oral information can be viewed as a form of eyewitness testimony or a police-recorded statement. It comes from detailed descriptive information obtained through communication and questioning of eyewitnesses, relatives, or other relevant persons. The collection and credibility (Behrman and Davey 2001; Wells et al. 2000) of testimonies have always been an important part of the investigation of cases. It can provide key clues about the physical characteristics, behavioral habits, and last known time and location of missing persons. These statements may aid in establishing potential search areas for locating the missing persons (Fyfe, Stevenson, and Woolnough 2015). Therefore, using this data and building models to help in the recovery of missing persons can be of help to the police.

The effectiveness of natural language data has been verified to aid in predicting missing populations. For example, Dubrawski et al. (2015) and Portnoff et al. (2017) used large amounts of natural language information from community advertisements to help identify differentiated human trafficking activities. Gomes et al. (2023) also designed a platform to automatically collect and standardize information on missing children in Brazil by using natural language techniques. This study illustrated the importance of basic information in predicting the status of missing children, including age, residence status, skin color, and hair color, which have been shown to correlate with whether or not a missing person is found. However, concerning missing persons, the studies under-explored in terms of predicting the whereabouts of missing persons mentioned above mostly have not taken full advantage of enriched clues provided by oral information. The missing person data on the *Baby Coming Back Home* (www.baobeihuijia.com) website contains considerable oral

information, which is a kind of natural language information. These oral data comprise the physical characteristics of missing persons, such as the time of the missing person's disappearance, places the missing person may have been, and things the missing person may have seen. Combining this oral information with other basic information enables researchers to form a three-dimensional (time, place, and experience before disappearance) description of the missing person. However, how to incorporate the three-dimensional description into the missing person prediction problem is still in the exploration stage.

Based on previous research on missing person relocation and the recognition of the importance of oral information in law enforcement investigations, we realized that previous studies primarily focused on constructing models using discretized basic information while neglecting the value of oral information. Oral information is typically presented in the form of natural language text, and recurrent neural networks (RNNs) are widely recognized as effective models for processing natural language information (Lipton, Berkowitz, and Elkan 2015). Their memory capabilities enable them to capture correlations and dependencies between text sequences. Therefore, we simplified the task to a classification model using models related to RNNs, the output of the model is the probability of the missing population arriving in the city. Additionally, RNNs exhibit flexibility when handling variable-length input sequences. In predicting the whereabouts of missing persons, the oral information of different samples may have varying lengths and structures. The flexibility of RNNs allows them to adapt to different lengths of text data and extract key features.

Therefore, aimed to enhance the recall of the prediction model by incorporating oral information, we proposed a framework called MP-Net (Missing Persons Network), which includes a deep learning approach based on TextRNN. We utilized the oral information, along with other basic attributes, as input. The natural language information was processed through embedding and encoding techniques to extract valuable clues about missing persons. We evaluated the impact of oral information on nationwide missing person predictions using data provided by the 'Baby Coming Back Home' website, which has been utilized in numerous national studies examining missing persons and has yielded important conclusions (Wang et al. 2018; Yao et al. 2021). Additionally, we partitioned the oral information into words based on part-of-speech (POS) tags to further analyze the influence of different types of oral information on the results. Finally, we utilized the results from the hidden layers of the network to quantify the weights of different word types and observed which words have a more significant impact.

The rest of this paper is organized as follows. In Section 2, this study introduces the process of data collection, data processing, and the proposed MP-Net framework for solving missing person prediction. Then, we present our experimental results in Section 3 by analyzing the effectiveness of oral information, vocabulary impacts, and model stability. Sections 4 and 5 discuss and conclude this study, respectively.

2. Data and methods

Figure 1 demonstrates the flowchart to solve the missing population destination prediction problem by combining oral information with other basic attributes. The solution comprises three main parts: data collection, the proposed MP-Net framework, and vocabulary weight analysis. (1) In data collection, a crawler is used to obtain nongovernmental organization (NGO) missing person data and classify the missing categories according to the information provided by the website. (2) Employing the proposed MP-Net framework to train a network for missing person prediction. The internal structure of the MP-Net framework is shown in Figure 1; this structure includes three steps. First, the framework extracts the encoded dataset from missing people information text utilizing regular matching and part-of-speech. Then, to investigate the effect of oral information and its different word categories on the prediction results, this study combines oral and other basic information into different datasets by making multiple permutations. Finally, after model training and recall evaluation, we qualitatively analyze the impact of data composition on model prediction.

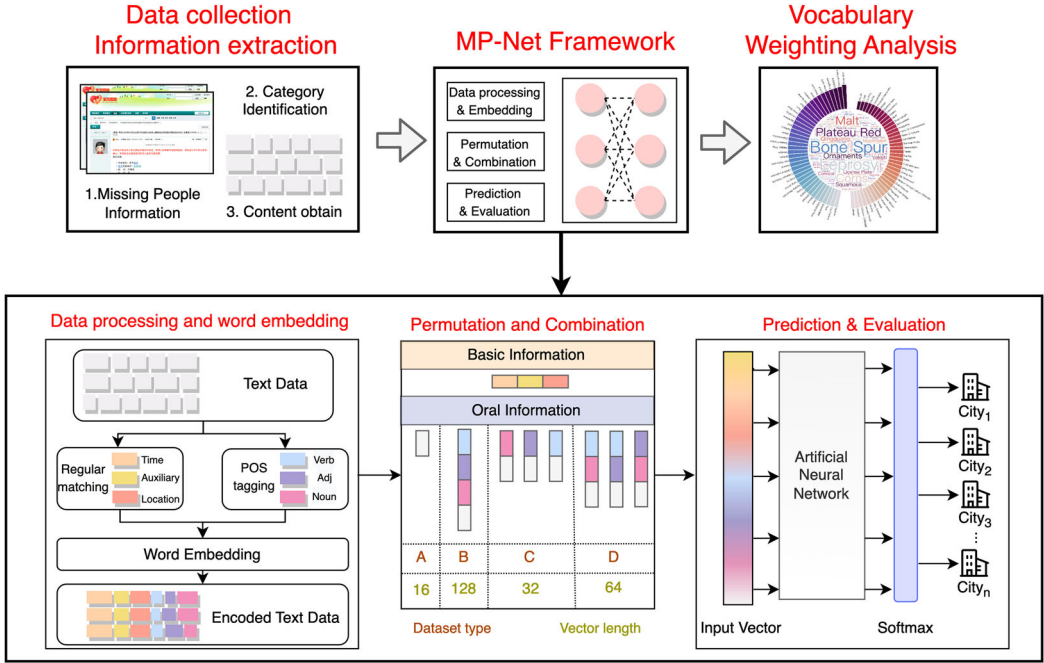


Figure 1. The flowchart for missing population destination prediction. The solution comprises three main parts: data collection, the MP-Net framework, and vocabulary weight analysis.

(3) By calculating the weight of each vocabulary, we can quantitatively analyze the influence of vocabulary type on promoting or suppressing prediction results.

2.1. Data introduction

2.1.1. Data collection

The missing person data used in our study was obtained from the ‘Baby Coming Back Home’ website, which is a widely influential nonprofit organization in China. With the assistance of hundreds of thousands of website volunteers and law enforcement agencies, over 10 thousand families have successfully found their missing children through the publication of search information on the website, including basic and oral information (source: news on the ‘Baby Coming Back Home’ website).

Our web crawler program started with ‘Baby Coming Back Home’ as the initial URL and focused on two major sections, namely ‘Baby Seeking Home’ (Baby Looking for Original Family) and ‘Family Seeking Baby’ (Family Looking for Lost Children) to collect information on missing persons. We collected a total of 83,767 raw data instances, including the post title, posting time, basic information about the missing person, detailed descriptions (oral information), and follow-up information on the search. After filtering out instances with missing locations or detailed descriptions, we obtained a set of 73,691 cleaned data instances.

Among all the data, only 30,294 instances have the missing and destination locations, which will be used as input and output of the model. These instances are not necessarily successful cases of finding missing persons. Many of the data instances originate from trafficked or adopted children or families. For trafficked children, after growing up and discovering their origins, they may turn to the platform for assistance in finding their families. There are also cases where families, for various reasons, sent their children for adoption or trafficking to specific locations and recorded the information on the platform after a long period of time when they want to find them back. Therefore,

these pieces of information, which are recorded by the missing persons or their families, are authentic.

During the data collection process, we strictly adhered to principles of privacy and legal compliance. We only accessed publicly available information and operated in accordance with the terms of use and privacy policy of the ‘Baby Coming Back Home’ website. We did not collect any sensitive personal information and ensured that no infringement on personal privacy occurred. Additionally, we implemented strategies such as setting request frequency limits to reduce the load and interference of the crawler program on the website.

2.1.2. Data processing and word embedding

The missing person information includes both basic information and oral information. In this study, we employed different processing methods based on the characteristics of the data.

For the basic information, we adopted a method used in previous studies (Yao et al. 2021). We applied the Jenks method to cluster continuous data such as height and days. We used rule-based matching for time information and divided location information into blocks. Subsequently, we categorized the basic information into three major types: time information, representing the date or season of birth and disappearance, with a data instance size of 9; auxiliary information, representing gender and height, with a data instance size of 2; and location information, representing blocks, provinces, and cities, with a data instance size of 3, as shown in Table 1. The continuous data through the clustering method, to get the category, before being spliced into the basic information. For example, ‘year of birth classification – 5’ and ‘missing time interval classification – 1’. Data that do not need clustering can be directly spliced into the basic information according to their own category attribute. For example, the result of gender and season is ‘gender – male’, and ‘season of birth – spring’.

To maximize the use of oral information, this study used the Chinese word division model HanLP (He and Choi 2021) to eliminate stop words, auxiliaries, and other words without real meanings from the oral information. The oral information is then divided into three major

Table 1. Sample data of missing person information extracted from the original webpage, including time, auxiliary, location, and oral information

Data information				Time information			
ID	Birth year	Birth month	Birth day	Missing year	Missing month	Missing day	Time interval
295130	1988	1	1	1988	10	17	16 day
422903	1980	9	3	1997	9	3	4572 days
263361	1993	1	1	1993	3	11	69 days
...

ID	Time information		Auxiliary information		Location information		
/	Birth season	Missing season	Gender	Height	Missing block	Missing province	Missing city
295130	Spring	Autumn	Female	15cm	16	Hebei	Chengde
422903	Autumn	Autumn	Female	175cm	30	Henan	Luoyang
263361	Spring	Spring	Male	30cm	33	Hunan	Hengyang
...

ID	Oral information
295130	'... I only remember walking from my home, in the waiting room of Qingdao Railway Station, possibly living near Jimo Road. Her adoptive parents have gone to the countryside, and she also has a sister. Her grandmother seems to be from Lijia Village, west of Jiaozhou ...'
422903	'... My daughter went out to work, and it was said that she went to a clothing factory in Guangzhou, had conflicts with her colleagues, and then left the factory. During this period, I received a letter from my daughter from Chenzhou, Hunan Province ...'
263361	'... I have been particularly fond of eating pepper since I was young, and I love the spicy taste. I suspect it's from Sichuan. According to my adoptive father, I picked it up from a camphor tree on Jianshe Road, Xintang Town, Hengdong County, Hengyang, Hunan Province ...'
...	...

categories based on part-of-speech tags: verbs (average embedding length in 28 bits), adjectives (average embedding length in 5 bits), and nouns (average embedding length in 91 bits).

For word embedding, this study uses a randomly initialized lookup table as the word-embedding approach. Previous studies have demonstrated that using pretrained embedding word vectors benefits the construction of natural language processing tasks (Qi et al. 2018). However, in this experiment, the processed basic information, for example, ‘missing days classification-5’ (5 is the result of Jenks clustering), is not present in the pretraining word list and will be treated as an ‘Unknown’ field, seriously affecting the original meaning of the words. Therefore, this study uses a random initialization approach to learn and update the word vectors in training. The data processing and word vector encoding process are illustrated in Figure 2. Since all the data have complete basic information, consisting of 14 fields in total, the length of the basic information remains fixed after being transformed into word vectors. On the other hand, the length of the oral information varies for each text.

2.2. MP-net framework

2.2.1. Permutation and combination of datasets

To investigate the performance of oral information and qualitatively analyze the impact of different data on prediction, this study organized and combined the data into four datasets based on the following rules. The classification results, which are shown in Table 2, are based on the following rules.

A: Basic information. This dataset contains three types of word vectors: time, auxiliary, and place information, which are used to explore the learning ability of the neural network model for basic information.

B: Basic information + oral information. This dataset contains all six types of word vectors, which are used to explore the effect of oral information on the prediction results.

C: Basic information + oral information of different lexicalities. This dataset contains three basic word vectors and one oral information word vector.

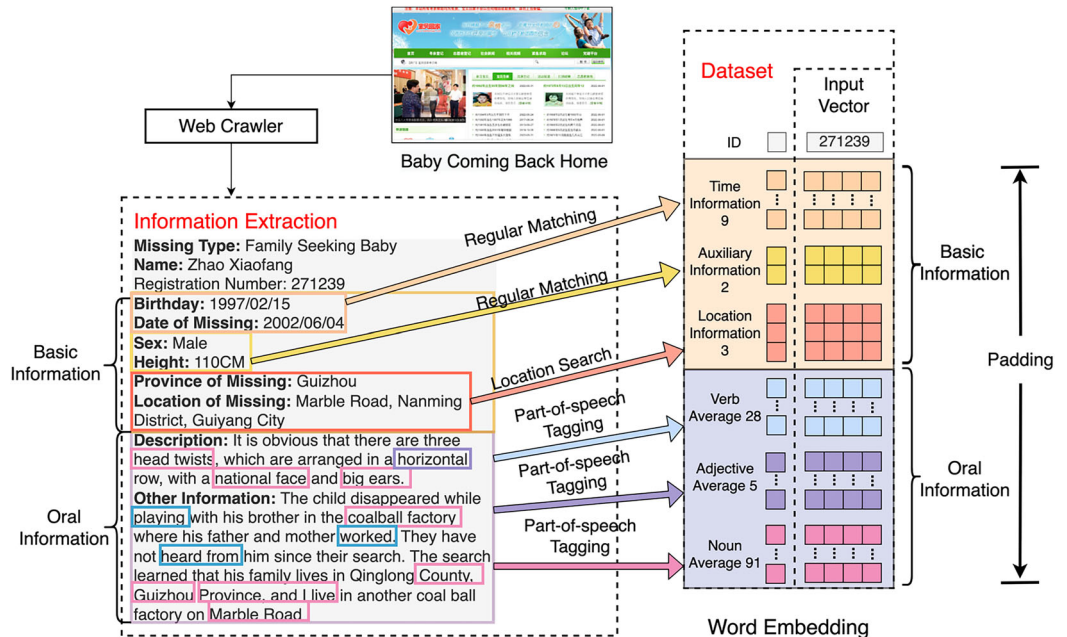


Figure 2. Extracting valuable data from basic and oral information through diverse methods and word vector encoding.

Table 2. Four dataset strategies: data combinations and padding-filling selection.

Dataset name	Information type						Quartile of text length (Q1, Q2, Q3)	Padding length	
	Basic information			Oral information					
	Time	Auxiliary	Location	Verb	Adjective	Noun			
A	✓	✓	✓	–	–	–	16,16,16	16	
B	✓	✓	✓	✓	✓	✓	103,114,128	128	
C	C1	✓	✓	✓	–	–	37,41,46	64	
	C2	✓	✓	✓	–	✓	20,21,23	32	
	C3	✓	✓	✓	–	–	✓	78,86,96	96
D	D1	✓	✓	✓	✓	–	✓	97,107,122	128
	D2	✓	✓	✓	✓	✓	–	41,45,51	64
	D3	✓	✓	✓	–	✓	✓	82,91,102	128

D: Basic information + oral information (removed a word type). The purpose of this dataset is to explore whether the model recall improves after removing data that may negatively affect the experimental results.

It should be noted that the text length is inconsistent across different datasets. When the average length of different datasets varies significantly, using the same padding length may lead to overfilling of certain samples. Overfilled data cannot provide useful information; instead, it increases the computational burden and reduces model performance (Ang, Dhingra, and Wu Wills 2022). In addition, there are also studies exploring the use of different padding lengths for datasets of different lengths in experiments (Liu, Liang, and Yu 2020). For this reason, we use the Quartile of each dataset as a reference (Q1, Q2, Q3), and select a padding length slightly larger than Q3 to ensure that most texts can be fully represented. The category names, types of information contained in the dataset, and padding length of the dataset we constructed are all listed in Table 2.

In addition, this study mixes the two main types of data obtained from the forum to create a dataset. This is because if the data are separated, as shown in Table 3, 3931 training samples for ‘Family Seeking Baby’ are assigned to more than 300 cities in China, which would result in not enough samples for each city. Too few samples per category will result in the classification model not obtaining good results (Beleites et al. 2013). In addition, both categories contain common missing person disappearances, such as common adoption and abductions, and the data formats are consistent and do not differ significantly. Therefore, this study mixes missing data from two sections to investigate the overall destination prediction of missing persons.

2.2.2. Introduction of TextRNN structure

Considering the model’s effectiveness and future application requirements, this study adopts TextRNN as the classification model when designing the MP-Net framework. This is based on the following two considerations: on the one hand, TextRNN performs well in Chinese classification tasks (Hu et al. 2019), and the Bi-LSTM module it uses also performs well in Chinese classification tasks, and there are many improved tasks based on this model (Pang, Xiao, and Zhao 2020); on the other hand, compared with the traditional Bi-LSTM, TextRNN keeps the same network structure, but with an additional multi-task layer. In Chinese text categorization tasks, multi-task learning

Table 3. Classification of datasets by type of missing person.

Data type	Dataset segmentation introduction	Segmentation results (train:test: validation)
Family Seeking Baby	Families, for various reasons, send their children for adoption or trafficking to specific locations and record the information on the platform after a long period of time when they want to find them back.	3931 : 1123 : 562
Baby Seeking Home	Trafficked or adopted children who learned about their origins after growing up and wanted to find their families.	17274 : 4936 : 2468
Mixed	Missing persons consist of both types, including common missing person disappearances such as common abduction and adoption.	21205 : 6059 : 3030

has been proven to be effective in improving model performance (Liu, Qiu, and Huang 2017; Zhang et al. 2020), so the module of multi-task learning can enhance the scalability of the framework.

TextRNN was proposed by Liu, Qiu, and Huang (2016), compared with the typical RNN structure (Hochreiter and Schmidhuber 1997), which cannot identify the correlation between long-distance information of natural language sequences, TextRNN adopts Bi-LSTM structure based on LSTM to establish long-distance information dependency to solve this shortcoming effectively. In addition, the structure of TextRNN can capture the shared information of all input tasks, enriching the meaning of words and improving the model recall. TextRNN has also been proven effective on Chinese datasets (Hu et al. 2019), making it suitable for the destination prediction problem, which could be regarded as a Chinese text classification problem in this study.

To gain a more intuitive understanding of the influence of each word on the results, we used the weight analysis method of TextRNN to analyze the output of the model in each time step. In the section on visualizing weights by using TextRNN, Liu, Qiu, and Huang (2016) found that the behavior of neurons can be understood by observing the activation of global gates $g(x)$ inside the network. To explain this concept, we need to detail the content of the TextRNN model. The structure of TextRNN is illustrated in Figure 3(A), where a shared layer is introduced to capture shared information across related tasks, enabling joint learning of the tasks. Where ‘This Task’ denotes the missing person prediction task of this study, and ‘Related Tasks’ denotes the hidden layer parts extracted by the model after modeling in other related domains. However, given the complexity of the missing persons problem, it is difficult to find suitable tasks for learning. Therefore, in the current task, this module is proposed as an extension of the model.

The basic unit of TextRNN is improved from LSTM, as shown in Figure 3(B). LSTM maintains a memory storage unit C and introduces the concept of ‘gates’ to control adding or eliminating information from the storage unit. Slightly different from the traditional LSTM, in the TextRNN model, the LSTM operation s are shown in Formulas (1) to (6):

$$i_t = \sigma(W_i x_t + U_i h_{t-1} + V_i C_{t-1}) \quad (1)$$

$$f_t = \sigma(W_f x_t + U_f h_{t-1} + V_f C_{t-1}) \quad (2)$$

$$o_t = \sigma(W_o x_t + U_o h_{t-1} + V_o C_t) \quad (3)$$

$$\tilde{C}_t = \tanh(W_c x_t + U_c h_{t-1}) \quad (4)$$

$$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t \quad (5)$$

$$h_t = o_t \odot \tanh(C_t) \quad (6)$$

where i , f , and o denote the input gate (how much new content needs to be remembered), the forget gate (how much old content needs to be forgotten), and the output gate (how much content needs to be output), respectively. The subscript t denotes the current time step, and h denotes the hidden layer. W is the input-to-hidden weight matrix, U is the state-to-state recurrent weight matrix, and W , U , and V are the weight matrices of the input layer, hidden layer, and cell state, respectively. \odot denotes elementwise multiplication.

To make each task share learning information during training to enhance model performance, Liu, Qiu, and Huang (2016) defined a shared layer and the concept of a global gate g , which enables the model neurons to accept or reject the information passed by neurons in the shared layer. The model implements this function by redefining Formula (4), as shown in Formula (7):

$$\tilde{C}_t^{(m)} = \tanh(W_c^{(m)} x_t + g^{(m)} U_c^{(m)} h_{t-1}^{(m)} + g^{(s \rightarrow m)} U_c^{(s)} h_t^{(s)}) \quad (7)$$

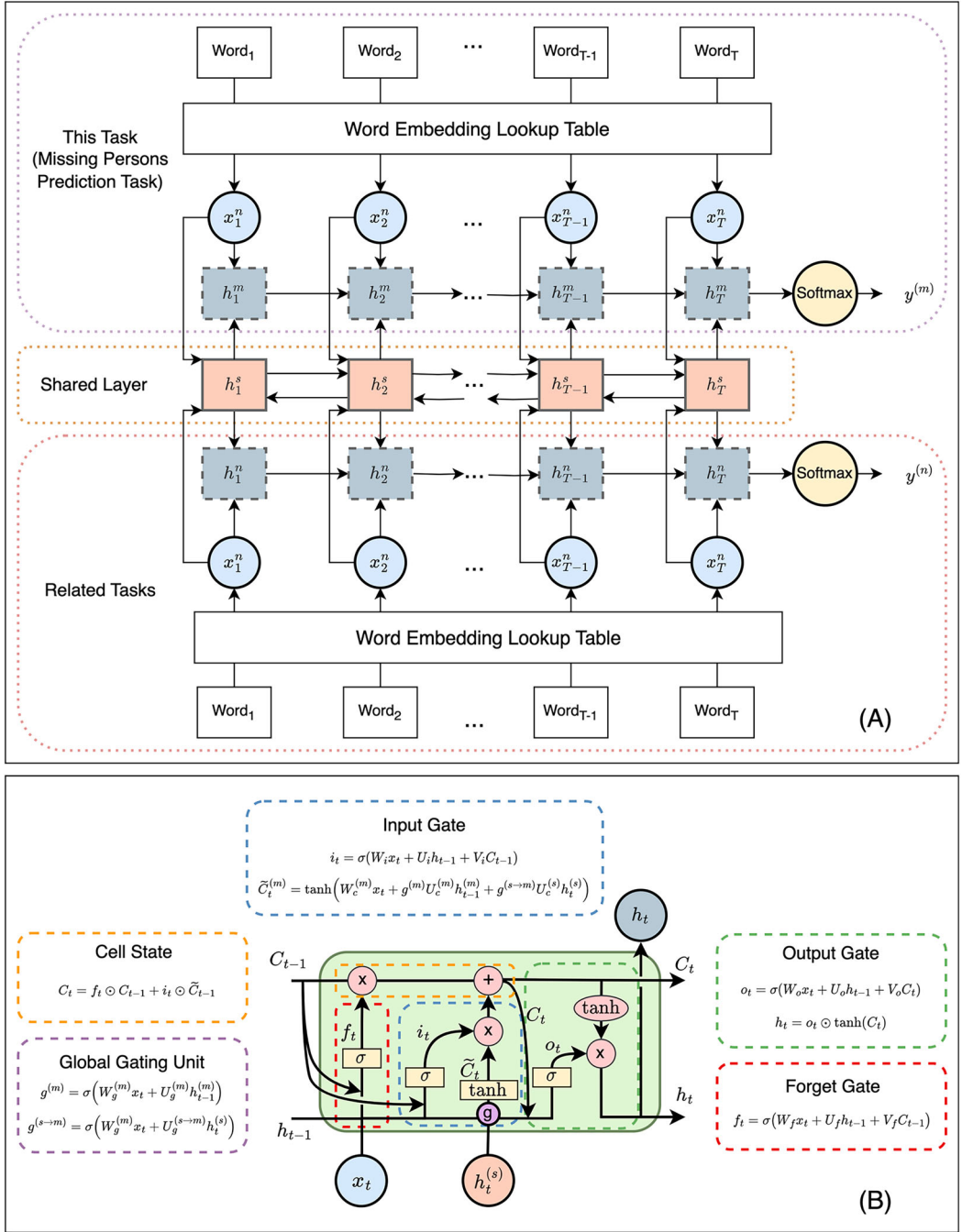


Figure 3. Introduction to the TextRNN model (A) The overall structure of the model (B) Internal structure of the neuron unit.

where s denotes the shared layer, m denotes the current input task, and the global gate g is defined as

$$g^{(m)} = \sigma(W_g^{(m)} x_t + U_g^{(m)} h_{t-1}^{(m)}) \quad (8)$$

$$g^{(s \rightarrow m)} = \sigma(W_g^{(m)} x_t + U_g^{(s \rightarrow m)} h_t^{(s)}) \quad (9)$$

The upper corner marker $s \rightarrow m$ represents the index of the associated set of parameters for the transition from the shared layer to task m in the current time step.

Since the TextRNN uses the Bi-LSTM structure, the shared hidden layer $h^{(s)}$ consists of forward and backward LSTMs, as shown in Formula (10), where \oplus denotes a merging operation:

$$h_t^s = \overrightarrow{h}_t^s \oplus \overleftarrow{h}_t^s \quad (10)$$

Based on that, this study uses the global gate $g(x)$ proposed by Liu, Qiu, and Huang (2016) and the hidden layer weights to calculate the activation weights of the neural network for each word encoding. If there is no multitask learning, then $g(x)$ can be considered as 1. Relying on word categories, this study analyzes the effect of each category of vocabulary on the results.

2.2.3. Model evaluation

In keeping with previous research, we evaluate the model using recall, but include both both Micro and Macro scales. The difference is that the Micro scale calculates the proportion of each category in the total data, so the calculation will favor categories with more data. In contrast, the macro metrics assign the same weight to each category in order to better understand the overall predictive level of the model in the case of sample imbalance (Gowda et al. 2021). Since missing persons tend to cluster in certain cities or provinces (Xin and Cai 2018; Yao et al. 2021), the imbalance of missing person cases across different regions is a common phenomenon. Therefore, it is essential to investigate the model's overall performance using macro indicators. The recall for each city is shown in Formula (11), where TP and FN denote true positives and false negatives, respectively. The formula for recall at the Micro and Macro scales is shown in (12) and (13), where c_{city_i} is the sample size for the i th city. In this study, the Micro scale is compared with the baseline random forest model, and the Macro scale will be used to assess the stability of the model.

$$R_{city} = \frac{TP}{TP + FN} \quad (11)$$

$$Micro - Recall = \sum_{i=1}^N \frac{R_{city_i} * c_{city_i}}{N} \quad (12)$$

$$Macro - Recall = \sum_{i=1}^N \frac{R_{city_i}}{N} \quad (13)$$

3. Results

3.1. Prediction results of MP net

3.1.1. Effectiveness of oral information

In this study, the dataset was divided into training, testing, and validation sets in a ratio of 7:2:1, as shown in Table 3. We maintained this ratio for all the datasets in Table 2. The total amount of data is 30,294, including 21,205 data in the training set, 6059 data in the testing set, and 3,030 data in the validation set. The lookup table which maps all basic and oral information to vectors, comprises 66,573 words, each with a 300-dimensional embedding. Among these words, there are 1,029 pieces of basic information in the format of 'missing time interval classification -5'; and 65,544 pieces of oral information, all of which are the results of natural language word segmentation based on part-of-speech. In this study, different padding lengths are used for different datasets, as shown in Table 2. Adam was used as the optimizer in the model. Combining the machine performance (Intel(R) Xeon(R) W-2123 CPU @ 3.60 GHz) and the VRAM size (NVIDIA GeForce GTX 1080 Ti, 11G), we chose 128 as the batch size with 100 iterations, and the learning rate of the model training

Table 4. Performance comparison of MP-Net and random forest models in predicting the top1, 3, 5, 10 recall of missing persons' destinations.

Type	Dataset				Model	Micro Scale			
	Basic	Verb	Adj.	Noun		Recall@1	Recall@3	Recall@5	Recall@10
A	✓	–	–	–	Random Forest	39.50%	63.56%	73.02%	82.99%
A	✓	–	–	–	MP-Net	45.56%	60.13%	66.68%	75.57%
B	✓	✓	✓	✓	MP-Net	42.75%	63.04%	73.41%	86.21%

was set to 0.0001. When the losses of the validation set are all stabilized the best model is recorded and training is stopped.

After 100 iterations of training, the recall results of the model under basic information (Dataset Type A) and oral information (Dataset Type B) are presented in Table 4, The symbol '@n' in the table indicates that the correct result appears in the first n predicted cities. The experimental results revealed that the TextRNN model can help recognize the important information hidden behind missing person data. In addition, incorporating oral information into the missing person prediction task can significantly improve recall. The top-10 recall rate of the class A dataset trained by the random forest model is higher than that of TextRNN. The results after adding the oral information are illustrated in the third row of Table 4. Although TextRNN's performance in the 1st and 3rd recall is slightly lower than that of the model using only the basic information (we will discuss it in the next section), TextRNN outperforms the integrated random forest model in the top-5 and top-10 recall with an improvement of 3.22% in the top-10 recall.

To investigate how different words in the basic information dataset affect prediction, this study calculates the weights for each vocabulary in the TextRNN model. Figure 4 shows the calculated weights of our randomly selected data instances (7% from the total data). The vocabulary with a greater impact on the results is the location class, whose impact is approximately 1.32 times the average weight of the temporal information and 1.35 times the average weight of the auxiliary information. This outcome is consistent with previous findings obtained by using the random forest approach (Yao et al. 2021), demonstrating that the model can autonomously learn connection between missing persons and cities from basic information during training. By counting the average weight of basic and oral information, the value of the basic information was 79.05, while the value of

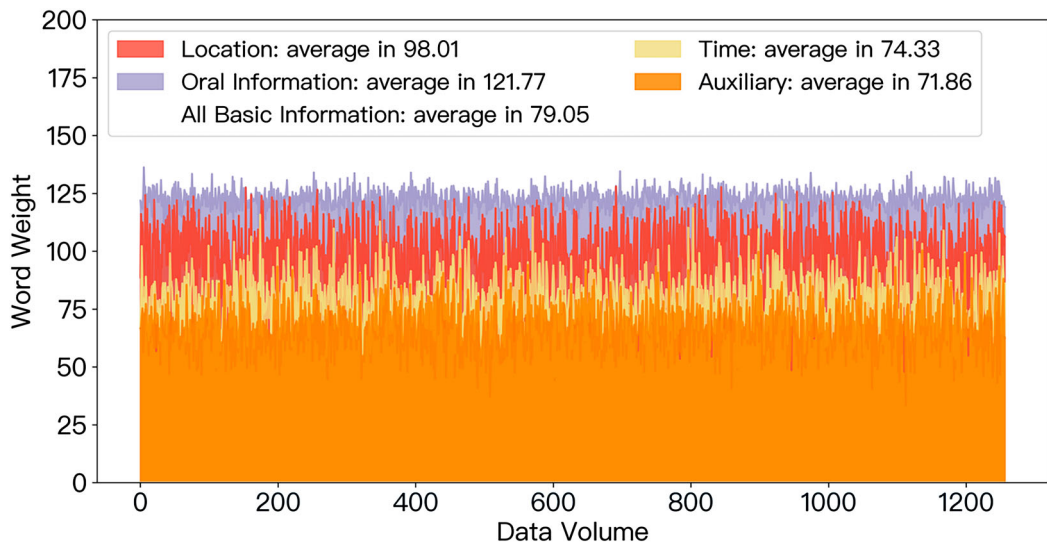


Figure 4. Weight calculation for each word of basic and oral information in datasets.

oral information was 121.77. This result proves that the influence of oral information is greater than the weight of basic information on the predicted result, proving the effectiveness of oral information.

3.1.2. Analysis differences in recall rate performance

In Table 4, we additionally observed that the baseline model performed well in terms of recall@3, and the MP-Net model with only basic information outperformed the model using both basic and oral information in terms of recall@1. To analyse the causes of these findings, we focused on two influencing factors: the number of samples per province and the distance of population transfer between provinces, and analyzed their correlation with the prediction results for each province. To measure the correlation between two variables, we introduced the Pearson correlation coefficient, which provides a statistical measure of the strength of the linear relationship between the variables. The larger the absolute value of the coefficient, the stronger the correlation between the variables.

Our experimental results demonstrated that both the baseline model and our proposed MP-Net model were positively influenced by the number of samples, indicating that higher sample sizes led to higher recall. However, the impact of the sample size was relatively small for both models ($\rho_{\text{recall@3, Sample size}}$ for Random Forest = 0.087; MP-Net = 0.103). This suggests that the models exhibited stable performance on imbalanced samples.

However, in provinces with shorter distances of population transfer, the baseline model showed higher recall ($\rho_{\text{recall@3, Average provincial transfer distance}} = -0.52$), while the MP-Net model had a smaller impact from the transfer distance ($\rho_{\text{recall@3, Average provincial transfer distance}} = 0.17$). Therefore, although the baseline model exhibits a higher recall rate, it is important to note that this performance is achieved under specific conditions. In more general scenarios, the MP-Net model may demonstrate greater competitiveness. In order to present this phenomenon more visually, we created a bar chart as shown in Figure 5. From the figure, it can be seen that regardless of the transfer distance, the performance of MP-Net is relatively balanced across provinces. However, in the region with a longer transfer distance, which nears the left side of the image, the Random Forest model shows a significant drop in predicting performance.

The result may be attributed to the inherent algorithmic properties of Random Forests. As a probabilistic model, Random Forest constructs multiple decision trees during the classification process, taking into account the significance of features and the data distribution. Subsequently, it combines the outcomes of each tree to assign scores to the categories. If the majority of instances in the original dataset pertain to short-distance trafficking or adoption, the Random Forest model may exhibit a tendency to accurately predict this kind of data. In fact, several studies have confirmed that missing person cases in China predominantly take place within the same province or in adjacent provinces (Huang and Weng 2019; Xin and Cai 2018), demonstrating a pattern of short distances, which may potentially cause the higher recall rate achieved by the random forest model in

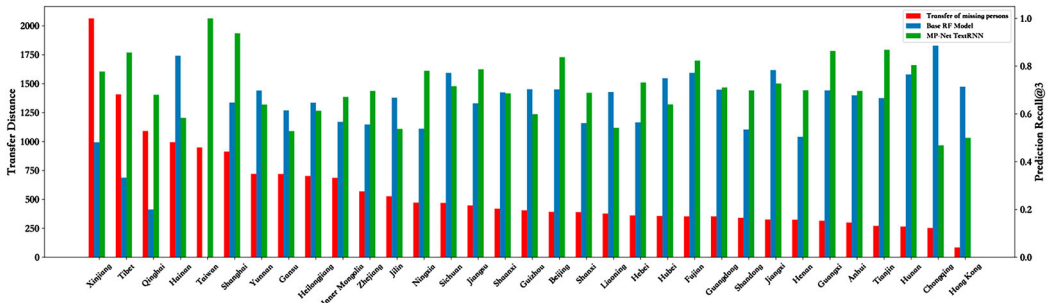


Figure 5. Performance of two models in recall rates in different cities.

such regions. Conversely, neural networks possess the ability to establish intricate nonlinear relationships, rendering them less susceptible to the influence of specific data types, such as short-distance missing person transfer.

Since basic information has been proven to be effective in previous studies (Yao et al. 2021), this could explain the performance difference between the model using basic information (MP-Net based on dataset A) and the model using both basic and oral information (MP-Net based on dataset B) on recall@1. However, for broader recall metrics such as recall@3, 5, and 10, the models achieved better performance because the basic and oral information incorporated richer background knowledge and provided more comprehensive information for the classification task. In addition, models with oral information have more complex structures and contextual relationships, which may contribute to their stronger generalization ability and overall prediction performance.

3.2. Quantitative analysis of vocabulary impact

By dividing the data into datasets containing different vocabulary categories, this study finds that adjectives inhibit the prediction task. In contrast, nouns and verbs help significantly improve model prediction recall. In this part, verbs, adjectives, and nouns in the oral information were mixed with basic information for prediction separately, and the results are shown in Table 5. As the table shows, when the adjective type of oral information is added (C2), the indicators of the model are even lower than the prediction using only the basic information (A). After statistical analysis, we found that most of the oral adjectives consisted of words with less information content, such as ‘unique’, ‘basic’, ‘free’, and ‘unknown’. Only a few adjectives, such as the description of color, have specific meanings and might positively impact the results. Therefore, it is concluded that adding such adjectives after the basic information fails to increase the information obtained from the model, thereby affecting the model’s training and leading to worse model results.

The evaluation indices of the model trained by basic information (A) were both improved after oral verbs (C1) and oral nouns (C3) were added. The outcomes indicate that these two vocabulary categories can improve the predictive ability of the model. In this part, the influence ranking was obtained by calculating the mean value of the weights of several categories of vocabulary performed in the model, as shown in Figure 6. The average weight of verbs slightly exceeds that of nouns. However, the prediction result of C1 in Table 5 is slightly lower than that of C3, presumably due to the difference in the number of data. In each oral message, there are 91 nouns and 28 verbs on average, so the effect of nouns with more bits will outweigh the effect of verbs on recall. This effect is reflected in the difference in recall of the models trained using the C1 and C3 datasets. In the sentence structure of natural language, verbs are often crucial and may be the most critical information used to understand the meaning of a sentence (Anton-Mendez 2020). In multiple NLP tasks and applications, VerbNet, an English verb repository built by (Schuler 2005), has been shown to be helpful in supporting multiple NLP tasks and applications (Chiu et al. 2019), this finding presumably explains the large weighting of verbs in this study.

In addition, we note that the models constructed for oral verbs (C1) and oral nouns (C3) perform better than the full oral information (B) in both recall@1 and recall@3 dimensions. This phenomenon is similar to random forests, and for ease of comparison, we uniformly illustrate it

Table 5. Performance of MP-Net in predicting the destinations of missing persons with different combinations of datasets.

Type	Dataset				Micro Scale			
	Basic	Verb	Adjective	Noun	Recall@1	Recall@3	Recall@5	Recall@10
A	✓	–	–	–	45.56%	60.13%	66.68%	75.57%
B	✓	✓	✓	✓	42.75%	63.04%	73.41%	86.21%
C1	✓	✓	–	–	46.45%	62.02%	69.08%	77.54%
C2	✓	–	✓	–	42.07%	57.80%	64.27%	72.96%
C3	✓	–	–	✓	48.22%	64.78%	71.64%	79.94%

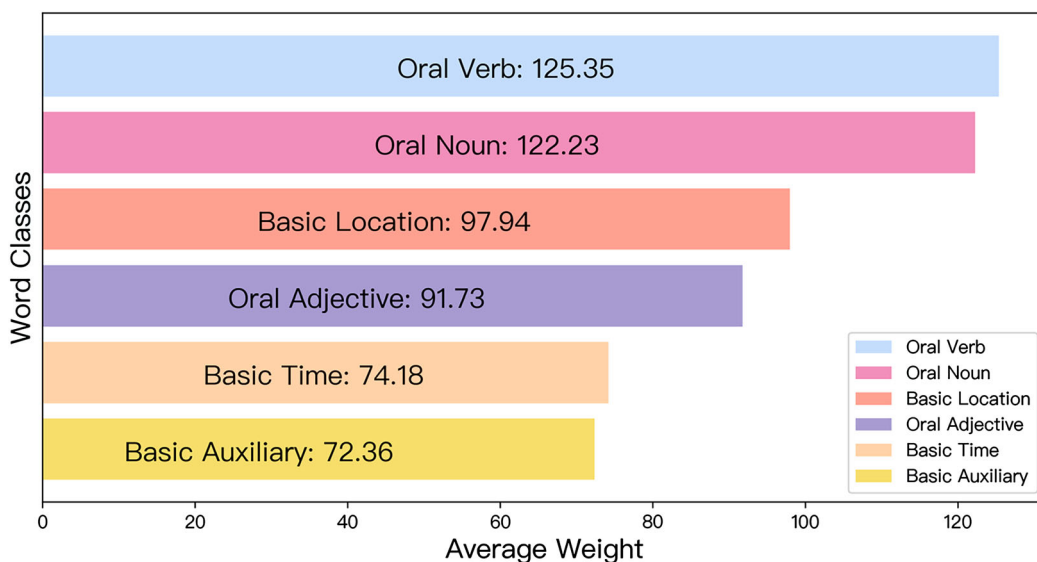


Figure 6. The average weight calculation results of different words in MP-Net.

in the recall@3 dimension. For Dataset C1 (verbs), it is affected by sample size and transfer distance by 0.35 and -0.17 , respectively; while for Dataset C3 (nouns), it is affected by sample size and transfer distance by 0.25 and -0.08 , respectively. This implies that the model performs better in some regions (such as regions where there are a large number of missing persons cases, representing a large sample size) but is less generalizable in other regions. As a result, Dataset B, which is less affected by the number of missing persons and the transfer distance, performs better as the search range increases and achieves the highest value for recall@10. It is also worth mentioning that Dataset C2 (adjectives) is affected by sample size and transfer distance by 0.52 and -0.44 , respectively, which provides another angle of interpretation that the introduction of adjectives has a great impact on model accuracy.

Since one category of nouns that can be clearly classified is location nouns, this study classifies nouns into location nouns and other nouns in Figure 7(A) and (B), respectively. The size of words and the height of bars represent the word weights calculated by MP-Net. Among the location-based

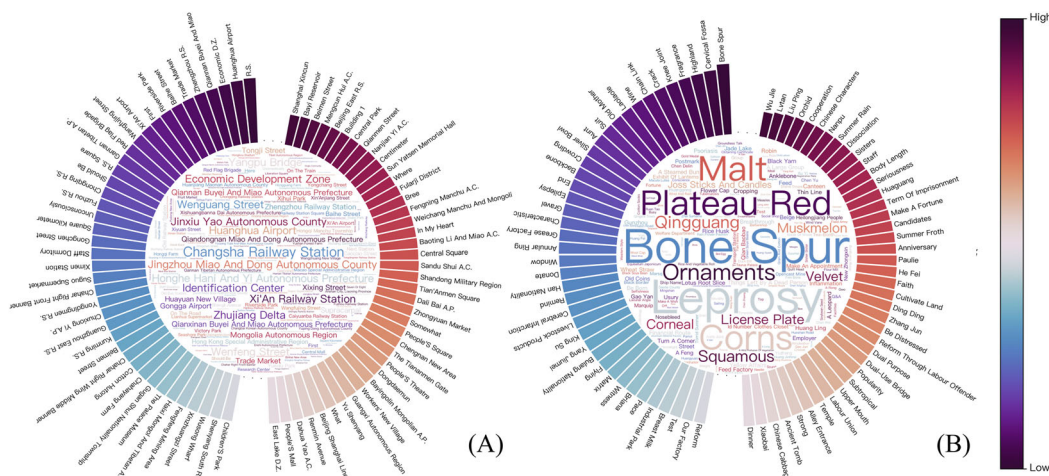


Figure 7. Visual display of noun weights. (A) Location nouns. (B) Other nouns.

nouns, transportation places, such as railway stations, airports, and bridges, have higher weights, indicating that these locations are prone to loss. While focusing on nouns representing provinces and cities, we find that autonomous counties, autonomous prefectures, and economic development zones occupy higher weights because of traditional thinking in these places (Xing, Chen, and Xu 2017) or the rapid development that has led to population mobility (Li et al. 2018), resulting in the prevalence of abduction and trafficking. In Figure 7(B), some terms used to describe the characteristics of missing persons also have high weights; these terms include ‘bone spurs’, ‘plateau red’, and ‘leprosy’. These results indicate that these recognizable characteristics can promote the detection of cases when recovering missing persons.

3.3. Model stability analysis

To explore the impact of removing adjectives on the model’s stability, this study conducts a prediction experiment using only verbs and nouns with basic information (Dataset D1), as shown in Table 6. For the Micro metric, compared to the performance of using the oral information (Dataset B) model, the recall in all dimensions has improved, which demonstrates that the model’s performance improves by removing the adverse impact of adjectives. To better evaluate the stability and validity of the model in the case of unbalanced data, this study evaluated the model using the macro method. Compared to the recall of the model evaluated on a Micro scale, the recall of the model trained with dataset B decreases significantly, with a reduction of 15.16% (from 86.21% to 71.05%), indicating that the recall of the forecast varies greatly from city to city under the influence of adjectives. In contrast, the recall of the model trained on dataset D1 decreased by 4.81% (from 87.18% to 82.37%), indicating that the model became more stable with the help of verbs and nouns, and outperformed the model with adjectives in each city.

In addition, we also explored the effects of verbs combined with adjectives (Dataset D2) and nouns combined with adjectives (Dataset D3) on the model prediction. We observed that compared to the fusion of three oral information types (Dataset B), the fusion of adjectives and verbs resulted in a decrease in the recall rate of the model in all dimensions, and the fusion of adjectives and nouns increased the recall rate of the model in all dimensions. This may be because the composition of the data by verbs and adjectives makes it difficult to form associations between words, while the combination of nouns and adjectives can provide richer contextual information, helping the model better understand the state or feature of the missing person and improving recall rates. However, we found a significant decrease in recall rates in both D2 and D3 on Macro. The finding indicates that neither the combination of adjectives and verbs nor the combination of adjectives and nouns can perform stably in imbalanced samples. In contrast, the collocation of nouns and verbs in the oral information improved the stability of the model, again demonstrating the importance of both types of vocabulary in this task.

Table 6. Macro metrics evaluation for models in the case of a dataset with all oral information (dataset B), and in the case of a dataset with one vocabulary removed (dataset D).

Dataset					Micro scale recall			
Type	Basic	Verb	Adj.	Noun	Recall@1	Recall@3	Recall@5	Recall@10
B	✓	✓	✓	✓	42.75%	63.04%	73.41%	86.21%
D1	✓	✓	–	✓	56.86%	74.03%	80.54%	87.18%
D2	✓	✓	✓	–	44.41%	61.23%	68.19%	76.99%
D3	✓	–	✓	✓	49.55%	68.56%	75.99%	84.05%
Dataset					Macro scale recall			
Type	Basic	Verb	Adj.	Noun	Recall@1	Recall@3	Recall@5	Recall@10
B	✓	✓	✓	✓	31.72%	46.47%	56.18%	71.05%
D1	✓	✓	–	✓	48.67%	65.96%	74.19%	82.37%
D2	✓	✓	✓	–	30.40%	46.81%	54.97%	66.63%
D3	✓	–	✓	✓	38.61%	57.73%	66.92%	77.31%

4. Discussion

4.1. Contribution of the experiment

The analysis and utilization of missing person information are crucial for recovering missing persons and maintaining social peace. Most previous prediction models focus on basic information, such as location, age, and characteristics, but they fail to extract the rich oral information in missing person data. In view of this, this study constructs the MP-Net framework, a missing person prediction model coupled with oral information, to solve the problem with high recall. The model adds natural language information to the missing person destination prediction problem for the first time, and visualizes the words with a high impact on the results by analyzing the internal weights of the model. The proposed MP-Net framework in this study verifies the validity of natural language information. Compared to the previous model, the proposed MP-Net framework improves the top-10 recall metric on Micro scale (from 82.99% to 86.21%). We also validate the conclusion that location information is weighted more heavily in the basic information, illustrating the effectiveness of the model in extracting information.

We demonstrate for the first time that oral information affects the prediction of the likely arrival of a missing person in a city. The results show that verbs and nouns positively affect prediction, while adjectives negatively affect it. This difference is additional strong evidence that verbs usually occupy an important component in natural language processing tasks and provide a new idea for missing person prediction studies. Further disambiguation of the nouns revealed that nouns in the location category and physical condition description category played a more positive role in predicting the results. This finding indicates that in both the basic information and the oral information, words related to location have a positive impact on the successful recovery of missing person. This finding proves that the environment in which disappearances occur more significantly affects the destination of the disappeared. In addition, the contribution of physical description information to the results suggests that more detailed descriptions can effectively help in recovering missing persons. These findings provide guidance to the police in collecting targeted information and speeding up the solving of cases when facing the problem of missing persons.

Through further experiments, we found that the removal of adjectives improves the model effect on the Micro scale (87.18%), thereby indicating that adjectives with less information negatively affect the model. Due to the presence of sample imbalance, using the Macro scale to reflect the overall model would be more unbiased. The experiments using the Macro scale found that the presence of adjectives in this dataset can seriously interfere with the training effect of the model. After removing the adjectives, the recall@10 of the model was significantly improved from 71.05% to 82.37% on Macro scale, the stability has been greatly improved. This result proved the existence of a complex role of oral information in the prediction results.

Our results can help police departments plan for targeted police deployment and support government decision-making. For example, police patrols can be increased in specific locations, such as train stations, where disappearances occur. In addition, child-wandering prevention campaigns can be enhanced in areas where disappearances frequently occur to raise parents' awareness of child protection. In addition, governments in areas with high mobile populations and areas influenced by traditional attitudes should introduce policies to reduce the occurrence of missing person cases. From the perspective of collecting oral information from family tracing, police departments can focus on behavioral actions, and nouns indicating the locations and characteristics of missing persons, to enrich the information description of missing persons and enhance the possibility of recovery. Regarding its application, the framework presented in this study can be employed to make predictions and narrow the search area for a missing person as soon as possible after their disappearance, in order to better recover the missing person.

4.2. Limitations and future works

However, this study also has some limitations. Since the missing person data comes from NGOs, there are missing and unbalanced samples, and assigning enough samples to each city is impossible. Thus, the model explores the prediction of the destination of missing person cases under mixed scenarios rather than separate predictions of adoption and abduction. In addition, NLP data augmentation in previous studies has focused mainly on English data, and the augmentation of Chinese natural language data is less practised (Ma and Li 2020; Wang et al. 2019; Yuan et al. 2020), coupled with the complexity of the issue of missing persons, it is difficult to compensate for the missing data. The study anticipates that more features can be obtained with more comprehensive data, which in turn can build more complete prediction models. This study aims to explore the role of oral information in the task of missing person prediction, however, the impact of information such as urban socioeconomic status on missing person prediction also needs to be explored.

Considering the extensibility of future research, this study designed the module of multi-task learning in the model building part. However, the application of multitask learning may be limited in the current problem of missing person prediction. The dataset of this study contains basic information and oral information. The basic information mainly consists of spliced vocabulary with a structure like ‘Missing Month-1’, which is a format derived from the Random Forest model we referenced (Yao et al. 2021), but has not been used in other tasks, making it difficult to learn knowledge from other domains. Although there are numerous tasks based on natural language that may be useful for the oral information component. However, it is difficult to find similar tasks for urban categorization. Common categorization tasks are mainly based on news data for label categorization or comment data for sentiment analysis, etc. Whether they can be migrated to the missing person prediction domain still needs further experimentation and exploration. Compared with common NLP tasks, multi-task learning for missing person prediction is much more complex, but we believe it is also a meaningful research. Given the multi-task learning capability of TextRNN, future research could choose to quantitatively explore the impact of different domain tasks on the missing person task based on this model or other similar models.

In addition, this study only considered a model construction scheme using randomly initialized lookup tables exclusively and did not explore the use of both pre-trained lookup tables and randomly initialized lookup tables. This is due to the fact that in a context where the dataset consists of both basic and oral information, the approach requires the model structure to be adapted to accommodate mixed inputs, and such model-specific optimization may be beyond the scope of this study. Therefore, future research could also consider how to fuse multiple embedding layers to achieve flexible control over different lookup tables, and to explain the mechanism of the impact of different type of information on the prediction results with multiple embedding layers. Exploring the impact of different lookup tables on the results is also an meaningful research direction, which might include using different sources of lookup tables for training, such as comparing the impact of lookup tables trained on generic data and police databases on the performance of the model in order to further optimizing the model’s performance.

Furthermore, future research could also consider fusing multiple tasks or data sources and exploring model construction methods. We can also work with the police to build a system for synchronizing case records with predictions. This system can save time for family searchers while obtaining the prediction result for the first time and can locate the missing person faster.

5. Conclusion

Targeted prevention tools and prediction methods with high recall are important research elements in missing person recovery. This study aims to design models and input methods to incorporate oral information into the missing person prediction problem. The results show that the MP-Net framework proposed in this study improves the recall of population location prediction, with the highest

top-10 recall rate of 87.18%. Our results demonstrate for the first time that oral information can effectively assist in the missing person destination prediction task and has a complex relationship with predicting urban outcomes. We demonstrate that verbs and nouns can significantly improve model prediction results due to the influence of verbs in NLP tasks and the effective information provided by place and person features in nouns. This study can provide a reference for police deployment and crime location ranking and can provide a guiding direction for police to focus on location- and feature-based words when collecting information. Future research will consider actual police data, multiple sources of spatiotemporal data, and interpretable methods to analyze the destination prediction of missing persons from more perspectives. We can also integrate algorithms into online systems to provide more accurate and efficient presentations of the possible whereabouts of missing persons.

Data availability statement

The data and codes that support the findings of the present study are available on Figshare at <https://doi.org/10.6084/m9.figshare.21971804>.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by the National Key Research and Development Program of China [grant number 2023YFB3906803], the National Natural Science Foundation of China [grant number 42171466]; the “CUG Scholar” Scientific Research Funds at China University of Geosciences (Wuhan) [grant number 2022034], the Key Laboratory of Earth Surface System and Human-Earth Relations, Ministry of Natural Resources of China [grant number LBXT2023YB03] and a Guangdong-Hong Kong-Macau Joint Laboratory Program [grant number 2020B1212030009].

ORCID

Yao Yao  <http://orcid.org/0000-0002-2830-0377>

References

- Ang, P., B. Dhingra, and L. Wu Wills. 2022. *Characterizing the Efficiency vs. Recall Trade-off for Long-Context NLP Models*. Dublin, Ireland: Association for Computational Linguistics.
- Anton-Mendez, I. 2020. “The Role of Verbs in Sentence Production.” *Frontiers in Psychology* 11: 189. <https://doi.org/10.3389/fpsyg.2020.00189>.
- Behrman, B. W., and S. L. Davey. 2001. “Eyewitness Identification in Actual Criminal Cases: An Archival Analysis.” *Law and Human Behavior* 25 (5): 475–491. <https://doi.org/10.1023/A:1012840831846>.
- Beleites, C., Ute Neugebauer, Thomas Bocklitz, Christoph Krafft, and Jürgen Popp. 2013. “Sample Size Planning for Classification Models.” *Analytica Chimica Acta* 760: 25–33. <https://doi.org/10.1016/j.aca.2012.11.007>.
- Chiu, B., Olga Majewska, Sampo Pyysalo, Laura Wey, Ulla Stenius, Anna Korhonen, and Martha Palmer. 2019. “A Neural Classification Method for Supporting the Creation of BioVerbNet.” *Journal of Biomedical Semantics* 10 (1): 1–12. <https://doi.org/10.1186/s13326-018-0193-x>.
- Cockbain, E., and K. Bowers. 2019. “Human Trafficking for Sex, Labour and Domestic Servitude: How Do Key Trafficking Types Compare and What are Their Predictors?” *Crime, Law and Social Change* 72 (1): 9–34. <https://doi.org/10.1007/s10611-019-09836-7>.
- Cui, C., Sha Peng, Cai Li, Chuanchuan Zhang, Zhen Wang, Xiaoling Zhang, and Liyuan Wei. 2018. “Clan Culture, One-Child Policy and Child Trafficking of Illegal Adoptions in China.” *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.3496517>.
- Dubrawski, A., Kyle Miller, Matthew Barnes, Benedikt Boecking, and Emily Kennedy. 2015. “Leveraging Publicly Available Data to Discern Patterns of Human-Trafficking Activity.” *Journal of Human Trafficking* 1 (1): 65–85. <https://doi.org/10.1080/23322705.2015.1015342>.

- Farrell, A., and I. De Vries. 2020. "Measuring the Nature and Prevalence of Human Trafficking." In *The Palgrave International Handbook of Human Trafficking*, edited by Winterdyk John and Jones Jackie, 147–162. Cham: Springer International Publishing.
- Fyfe, N. R., O. Stevenson, and P. Woolnough. 2015. "Missing Persons: The Processes and Challenges of Police Investigation." *Policing and Society* 25 (4): 409–425. <https://doi.org/10.1080/10439463.2014.881812>.
- Gomes, J. A. O., Heder Soares Bernardino, Jairo Francisco de Souza, and Enayat Rajabi. 2023. "Indexing, Enriching, and Understanding Brazilian Missing Person Cases from Data of Distributed Repositories on the Web." *AI & SOCIETY* 38: 565–579.
- Gowda, T., Weiqiu You, Constantine Lignos, and Jonathan May. 2021. "Macro-Average: Rare Types are Important Too." *arXiv preprint arXiv:2104.05700*.
- He, H., and J. D. Choi. 2021. "The Stem Cell Hypothesis: Dilemma Behind Multi-Task Learning with Transformer Encoders." In *arXiv preprint arXiv:2109.06939*.
- Hochreiter, S., and J. U. R. Schmidhuber. 1997. "Long Short-Term Memory." *Neural Computation* 9 (8): 1735–1780. <https://doi.org/10.1162/neco.1997.9.8.1735>.
- Hu, W., Zhaoquan Gu, Le Wang Yushun Xie, and Keke Tang. 2019. "Chinese Text Classification Based on Neural Networks and Word2vec." In *2019 IEEE Fourth International Conference on Data Science in Cyberspace (DSC)*, 284–291. Hangzhou, China: IEEE.
- Huang, Z., and W. Weng. 2019. "Analysis on Geographical Migration Networks of Child Trafficking Crime for Illegal Adoption from 2008 to 2017 in China." *Physica A: Statistical Mechanics and its Applications* 528: 121404. <https://doi.org/10.1016/j.physa.2019.121404>.
- Judge, A., Jennifer A. Murphy, Jose Hidalgo, and Wendy Macias-Konstantopoulos. 2018. "Engaging Survivors of Human Trafficking: Complex Health Care Needs and Scarce Resources." *Annals of Internal Medicine* 168 (9): 658–663. <https://doi.org/10.7326/M17-2605>.
- Kiss, L., and C. Zimmerman. 2019. "Human Trafficking and Labor Exploitation: Toward Identifying, Implementing, and Evaluating Effective Responses." *Public Library of Science San Francisco* 16 (1): e1002740.
- Li, G., et al. 2018. "The Evolution of Spatial and Temporal Patterns of Child Trafficking Crime in China and its Influencing Factors." *Human Geography* 33 (02): 26–34.
- Li, C., R. Ren, and X. Zhang. 2013. "Research on the Current Situation of Rescue and Protection of Trafficked Children in China and the Improvement of Countermeasures – Analysis Based on 133 Publicly Reported Cases in Recent Years." *Journal of People's Public Security University of China: Social Sciences Edition* 6 (6): 16–25.
- Lipton, Z. C., J. Berkowitz, and C. Elkan. 2015. "A Critical Review of Recurrent Neural Networks for Sequence Learning." *arXiv preprint arXiv:1506.00019*.
- Liu, T., Y. Liang, and Z. Yu. 2020. *The Influence of Text Length on Text Classification Model*. Edited by J. Wang, et al. Singapore: Springer Singapore, 79–90.
- Liu, P., X. Qiu, and X. Huang. 2016. "Recurrent Neural Network for Text Classification with Multi-Task Learning." *arXiv preprint arXiv:1605.05101*.
- Liu, P., X. Qiu, and X. Huang. 2017. "Adversarial Multi-Task Learning for Text Classification." *arXiv preprint arXiv:1704.05742*.
- Ma, J., and L. Li. 2020. "Data Augmentation for Chinese Text Classification Using Back-Translation." *Journal of Physics: Conference Series* 1651 (1): 12039. <https://doi.org/10.1088/1742-6596/1651/1/012039>.
- Pang, N., W. Xiao, and X. Zhao. 2020. "Chinese Text Classification Via Bidirectional Lattice LSTM." In *International Conference on Knowledge Science, Engineering and Management*, edited by Li Gang, Shen Heng Tao, Yuan Ye, Wang Xiaoyang, Liu Huawen, and Zhao Xiang, 250–262. Springer.
- Portnoff, R. S., Danny Yuxing Huang, Periwinkle Doerfler, Sadia Afroz, and Damon McCoy. 2017. "Backpage and Bitcoin: Uncovering Human Traffickers." In *KDD '17: Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1595–1604. New York, NY, USA: Association for Computing Machinery.
- Qi, Y., Devendra Singh Sachan, Matthieu Felix, Sarguna Janani Padmanabhan, and Graham Neubig. 2018. "When and Why are Pre-Trained Word Embeddings Useful for Neural Machine Translation?" *arXiv preprint arXiv:1804.06323*.
- Schuler, K. K. 2005. *VerbNet: A Broad-Coverage, Comprehensive Verb Lexicon*. Philadelphia: University of Pennsylvania.
- Shen, A., G. A. Antonopoulos, and G. Papanicolaou. 2013. "China's Stolen Children: Internal Child Trafficking in the People's Republic of China." *Trends in Organized Crime* 16 (1): 31–48. <https://doi.org/10.1007/s12117-012-9167-z>.
- Sweilhe, W. M. 2018. "Bibliometric Analysis of Peer-Reviewed Literature on Antimicrobial Stewardship from 1990 to 2019." *Globalization and Health* 17 (1): 1–12. <https://doi.org/10.1186/s12992-020-00651-7>.
- Wang, X. 2015. "The Current Situation of Child Trafficking Crimes and Countermeasures to Curb Them – an Empirical Study of Province F as an Example." *Journal of People's Public Security University of China: Social Sciences Edition* 31 (05): 21–31.
- Wang, X., Yiqiang Sheng, Haojiang Deng, and Zhenyu Zhao. 2019. "CharCNN-SVM for Chinese Text Datasets Sentiment Classification with Data Augmentation." *International Journal of Innovative Computing, Information and Control* 15 (1): 227–246.

- Wang, Z., Liyuan Wei, Sha Peng, Liangchun Deng, and Beibei Niu. 2018. "Child-Trafficking Networks of Illegal Adoption in China." *Nature Sustainability* 1 (5): 254–260. <https://doi.org/10.1038/s41893-018-0065-5>.
- Wells, G., Roy S. Malpass, R. C. L. Lindsay, Ronald P. Fisher, John W. Turtle, and Solomon M. Fulero. 2000. "From the Lab to the Police Station: A Successful Application of Eyewitness Research." *American Psychologist* 55: 581–598. <https://doi.org/10.1037/0003-066X.55.6.581>.
- Xin, Y., and T. Cai. 2018. "Child Trafficking in China: Evidence from Sentencing Documents." *International Journal of Population Studies* 4 (2): 1–10. <https://doi.org/10.18063/ijps.v4i2.817>.
- Xing, H., R. Chen, and F. Xu. 2017. "The Developmental Characteristics of Child Trafficking Crimes – an Analysis of 447 Judgments." *Juvenile Delinquency Prevention Research* 1: 38–43.
- Xiong, W. 2023. "Evidence of Son Preference in the Child Trafficking Market for Illegal Adoption in China." *Journal of Human Trafficking* 9: 242–255.
- Yao, Y., Yifei Liu, Qingfeng Guan, Ye Hong, Ruifan Wang, Ruoyu Wang, and Xun Liang. 2021. "Spatiotemporal Distribution of Human Trafficking in China and Predicting the Locations of Missing Persons." *Computers, Environment and Urban Systems* 85: 101567. <https://doi.org/10.1016/j.compenvurbsys.2020.101567>.
- Yuan, F., Xianbin Hong, Cheng Yuan, Xiang Fei, Sheng-Uei Guan, Dawei Liu, and Wei Wang. 2020. "Keywords-Oriented Data Augmentation for Chinese." In *IEEE 6th International Conference on Computer and Communications (ICCC)*, 2006–2012. Chengdu, China: IEEE.
- Zhang, H., Shaoqi Sun, Yongjin Hu, Junxiu Liu, and Yuanbo Guo. 2020. "Sentiment Classification for Chinese Text Based on Interactive Multitask Learning." *IEEE Access* 8: 129626–129635. <https://doi.org/10.1109/ACCESS.2020.3007889>.