Ya QIN, Guo-wei SHEN, Wen-bo ZHAO, Yan-ping CHEN, Miao YU, Xin JIN, 2019. A network security entity recognition method based on feature template and CNN-BiLSTM-CRF. *Frontiers of Information Technology & Electronic Engineering*, 20(6):872-884. https://doi.org/10.1631/FITEE.1800520

A network security entity recognition method based on feature template and CNN-BiLSTM-CRF

Key words: Network security entity; Security knowledge graph; Entity recognition; Feature template; Neural network

Corresponding author: Guo-wei SHEN
E-mail: gwshen@gzu.edu.cn
ORCID: https://orcid.org/0000-0002-2685-3445

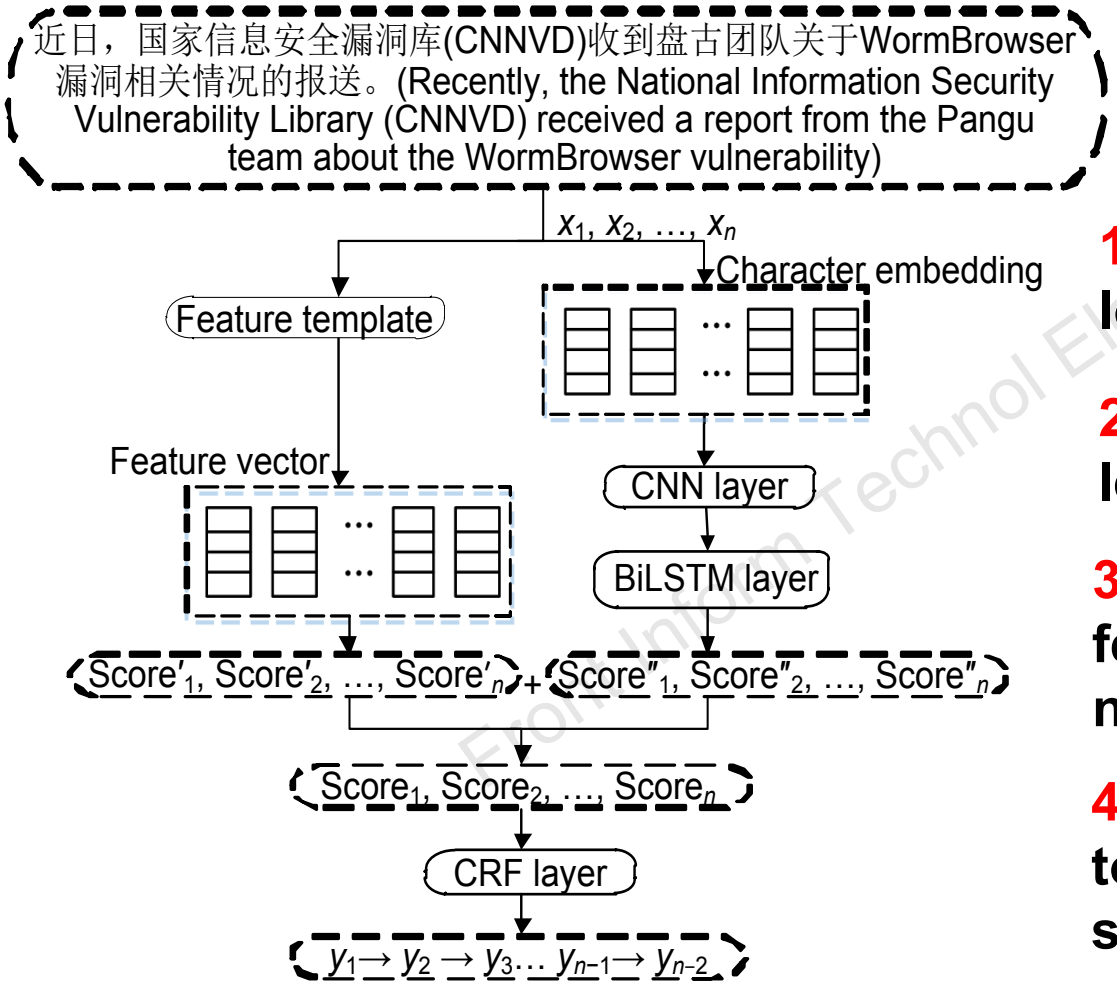
Motivation

- With the development of information technology, network security has become the focus of attention around the world. To ensure the security of cyberspace, countries deploy security services at key locations to monitor network security threats. These services produce much network security data.
- Due to the fragmentation and magnanimity of cyber security data, there is a lack of certain links among the data. Therefore, we build a network security knowledge graph (SKG) to effectively correlate, analyze, and mine a massive amount of security data.

Main idea

- Main technique for constructing a network SKG: security entity recognition.
- Security entity recognition models:
  - Feature template (FT)
  - CNN
  - BiLSTM
  - BiLSTM-CRF
- Experimental results showed that the proposed method outperforms other methods.

Security entity recognition model architecture



1.Feature template: extract local context features

2.CNN: extract character-level feature sequences

3.BiLSTM: extract the global feature vector sequence of network security text

4.CRF: mark each character to obtain the best label sequence

Corpus and annotation patterns

- We recognize mainly six types of security entities, and use the BIO annotation model.

Corpus statistics

Dataset	Number		
	Training	Validation	Testing
Sentences	15 090	1989	2697
Labels	70 324	9181	16 669
Person	1139	230	249
Location	975	441	112
Organization	4059	374	1216
Software	4236	666	1794
Relevant term	59 849	7428	13 102
Vulnerability ID	66	42	196

Major results

- To verify the performance of our security entity recognition model, we compare different algorithm models.

Experimental results of different models, evaluated by accuracy, precision, recall, and *F*-value (%)

Model	Acc	*P*	*R*	*F*
CRF	91.50	84.26	73.34	78.42
LSTM	92.36	83.75	80.62	82.16
LSTM-CRF	92.95	86.17	82.07	84.07
BiLSTM-CRF	92.83	84.70	85.18	84.94
CNN-BiLSTM-CRF	93.10	86.47	84.07	85.25
FT-LSTM-CRF	93.03	87.09	82.78	84.88
FT-BiLSTM-CRF	90.87	88.19	82.19	85.08
FT-CNN-BiLSTM-CRF	93.31	88.45	83.68	86.00

Network security entity extraction instance

- We visualize six types of extracted network security entities.

Security entity word cloud

Conclusions

- We present a novel neural network approach for security entity recognition in the network security domain by introducing a feature template, which allows the model to extract local context features.
- In future work, we will explore the issue of extracting the local features of entity. We will use the attention mechanism to extract local features and focus on the features that have important influence on security entity recognition.