



From Graph to Word Bag: Introducing Domain Knowledge to Confusing Charge Prediction

Ang Li¹, Qiangchao Chen², Yiquan Wu¹, Ming Cai¹, Xiang Zhou², Fei Wu¹, Kun Kuang¹

¹College of Computer Science and Technology, Zhejiang University ²GuangHua Law School, Zhejiang University



01 Introduction Contents 02 Method 03 Experiment

04 Conclusion







Introduction

Confusing Charge Prediction



01 Introduction: Task Definition



Confusing Charge prediction:

- Input: Fact Description
- Output: Charge Label
- subset of charge prediction

Charge	Fact description
Snatch	Defendant XX forcibly pulled off the gold necklace on Moumou's neck while the victim Moumou was not prepared, and the market value of the robbed gold necklace was RMB 63,202.
Robbery	The defendant XX beat the victim Moumou and robbed the victim of a gold necklace. The market value of the robbed gold necklace was RMB 1,060.91.
Fraud	Defendant XX in the name of communication, after gaining trust, defendant XX fabricated various reasons to defraud victim Moumou of an Apple computer, whose market value is RMB 5,450
Theft	The defendant XX stole an Apple computer of the victim Moumou. The market value of the Apple computer is RMB 5,000.
all ch	narges share some charges share unique words of each charge

Confusing charges in real legal cases. Red words indicate the words all charges share, blue words indicate the words some charges share, and the yellow highlighted words indicate the unique words of each charge.





From Graph to Word Bag (FGWB)



02 Method: Overview of FGWB





Overview of FGWB

02 Method: Word Bag former





Construction and utilization of expert knowledge

- Constituent elements of fraud are:
- (1) The suspect commits a fraudulent act;(2) The victim falls into a wrong understanding
- Knowledge graph
 We use a set of charges as nodes and employ the constituent
 elements as the connections between nodes.
- Graph-based Keyword Selection

$$\frac{\sum_{w \in C_i', r \in R_i} sim(f_{\theta}(w), f_{\theta}(r))}{|R_i|} > \eta \to w \epsilon C_i$$









Experiment



03 Experiment: Dataset

We use the following dataset for experiments:

- collect our data from **12309 China Procuratorate** Website[1]
- Common property charges, including Theft, Fraud, Snatch,

and Robbery

• Common drug charges, containing Drugs Selling (DS),

Providing Venues For Drug Users (PVFDU), Illegal Possession

Of Drugs (IPOD) and Drugs Transportation (DT) charges

Evaluation Metrics

 we calculate the Macro precision (Ma-P), Macro recall(Ma-R), and Macro F1 score (Ma-F) and accuracy (Acc). Ma-F* is used to represent the macro-F1 score tested on the balanced test set

Туре	Property Set	Drug Set
# Train set	30929	9591
# Valid set	3436	1066
# Test set	4000	2000
Avg. # Tokens in Fact	339	347

Table 1: Data Set Collection

Property Set	Number	Drug Set	Number
Theft	28535	DS	5345
Fraud	8426	PVFDU	4789
Robbery	1048	IPOD	1668
Snatch	356	DT	855

Table 2: Label Distribution



03 Experiment: Experiment Results



Mathad		Р	roperty S	Set				Drug Set	t	
Method	Ma-P	Ma-R	Ma-F	Acc	Ma-F*	Ma-P	Ma-R	Ma-F	Acc	Ma-F*
TextRNN	0.789	0.784	0.769	0.783	0.769	0.913	0.910	0.909	0.909	0.895
LSTM	0.857	0.815	0.808	0.814	0.798	0.914	0.901	0.909	0.924	0.914
TextCNN	0.846	0.799	0.776	0.800	0.743	0.903	0.880	0.878	0.884	0.908
DPCNN	0.882	0.864	0.856	0.865	0.896	0.931	0.934	0.927	0.937	0.947
C3VG	0.882	0.862	0.860	0.868	0.985	0.920	0.920	0.920	0.914	0.939
Electra	0.903	0.889	0.881	0.888	0.892	0.936	0.930	0.928	0.928	0.943
Topjudge	0.891	0.877	0.874	0.878	0.886	0.926	0.923	0.922	0.917	0.940
LADAN	0.905	0.893	0.892	0.895	0.908	0.933	0.930	0.929	0.940	0.966
NeurJudge	0.907	0.897	0.902	0.905	0.917	0.939	0.940	0.935	0.950	0.968
R-Former	0.905	0.895	0.894	0.901	0.918	0.931	0.948	<u>0.94</u> 1	<u> 6.951</u>	0.970
FGWB (LSTM)	0.888	0.880	0.876	0.880	0.880	0.934	0.933	0.930	0.929	0.942
FGWB (Electra)	0.923	0.925	0.924	0.925	0.928	0.957	0.955	0.956	0.955	0.979

Experiment results for property charges and drug charges

Mathad		prop	erty cha	rges			dr	ug charg	es	
Wethod	Ma-P	Ma-R	Ma-F	Acc	Ma-F*	Ma-P	Ma-R	Ma-F	Acc	Ma-F*
FGWB (LSTM)	0.888	0.880	0.876	0.880	0.880	0.934	0.933	0.930	0.929	0.942
w/o SV	0.862	0.839	0.857	0.848	0.872	0.914	0.897	0.911	0.904	0.924
w/o Multi-Attn	0.866	0.842	0.861	0.854	0.862	0.918	0.908	0.917	0.921	0.931
w/o KG	0.874	0.869	0.863	0.865	0.872	0.930	0.923	0.921	0.919	0.943
FGWB (Electra)	0.923	0.925	0.924	0.925	0.938	0.957	0.955	0.956	0.955	0.979
w/o SV	0.901	0.892	0.907	0.897	0.905	0.942	0.904	0.920	0.919	0.957
w/o Multi-Attn	0.907	0.897	0.917	0.902	0.911	0.948	0.927	0.934	0.959	0.965
w/o KG	0.912	0.902	0.910	0.895	0.919	0.950	0.948	0.935	0.948	0.977

Ablation Experiment Results

NeurJudge	Robbery	Snatch	Theft	Fraud
Robbery	709	7	0	2
Snatch	48	836	1	2
Theft	108	41	989	6
Fraud	2	10	10	992

Electra	Robbery	Snatch	Theft	Fraud
Robbery	888	244	1	0
Snatch	4	355	2	0
Theft	101	376	982	10
Fraud	7	25	15	990

FGWB (SV)	Robbery	Snatch	Theft	Fraud
Robbery	872	71	0	0
Snatch	68	862	2	0
Theft	35	34	997	6
Fraud	25	33	1	994

FGWB (MSV)	Robbery	Snatch	Theft	Fraud
Robbery	902	29	0	0
Snatch	181	942	1	0
Theft	45	98	996	3
Fraud	5	37	3	995

Confusing Matrices for Different Models.



Model	LSTM-attn	FGWB (SV)	FGWB (MSV)
Attention distribution	At about 02:02 on June 18, 2019 when the defendant Zhao Moumou went to the toilet in Tuqiao Village, Guandu District, Kunming City, he saw the victim Li Moumou put a red OPPO R17 mobile phone at his feet when he went to the toilet. Defendant Zhao XX took advantage of the victim Li XX's unpreparedness and snatched away the mobile phone that was at his feet. The value of the OPPO R17 mobile phone identified as involved in the case was 2190 yuan and has been returned to the victim.	At about 02:02 on June 18, 2019, when the defendant Zhao Moumou went to the toilet in Tuqiao Village, Guandu District, Kunming City, he saw the victim Li Moumou put a red OPPO R17 mobile phone at his feet when he went to the toilet. Defendant Zhao XX took advantage of the victim Li XX's unpreparedness and snatched away the mobile phone that was at his feet. The value of the OPPO R17 mobile phone identified as involved in the case was 2190 yuan and has been returned to the victim.	At about 02:02 on June 18, 2019, when the defendant Zhao Moumou went to the toilet in Tuqiao Village, Guandu District, Kunming City, he saw the victim Li Moumou put a red OPPO R17 mobile phone at his feet when he went to the toilet. Defendent Zhao XX took advartage of the victim Li XX's unpreparedness and snatched away the mobile phone that was at his feet. The value of the OPPO R17 mobile phone identified as involved in the case was 2190 yuan and has been returned to the victim.
prediction	Theft	Theft	Snatch

Attention distribution from different models for a Snatch case.



Model performance by the coefficient $\boldsymbol{\lambda}$ for attention supervision.





Conclusion





- We investigate the task of confusing charge prediction by taking the domain knowledge into consideration.
- We propose a novel From Graph to Word Bag (FGWB) approach. Specifically, we construct an expert knowledge graph with constituent elements and then form the word bag, combining multi-attention supervision to guide the model in distinguishing between confusing charges.
- We construct the confusing charge dataset from real-world data. Our experiments evaluate the effectiveness of our proposed method.





Thanks for Listening!

