Learning to Predict Charges for Judgment with Legal Graph

Si Chen†, Pengfei Wang†, Wei Fang†, Xingchen Deng† and Feng Zhang‡

†Beijing University of Posts and Telecommunications
‡Information Science Academy of China Electronics Technology Group Corporation

Background

The task of automatic charge prediction aims to train a machine judge to determine the final charges (e.g., theft, robbery or traffic offence.) of the defendants in criminal cases which is an important application of intelligent legal judgment system. This capability, if developed, can either largely assist the tedious daily routine work performed by human judges, or give legal guidance and assistance to people without legal background knowledge. Significant progress has been made in recent years by the development of traditional machine learning methods, such as k-Nearest Neighbor (KNN) and Support Vector Machine (SVM). Most existing works formalize this task under the text classification framework. For big successes of deep learning method in other areas, such as image and text summarization, this method has been explored for charge prediction. We give an example of the judgment case as shown in Fig.1.

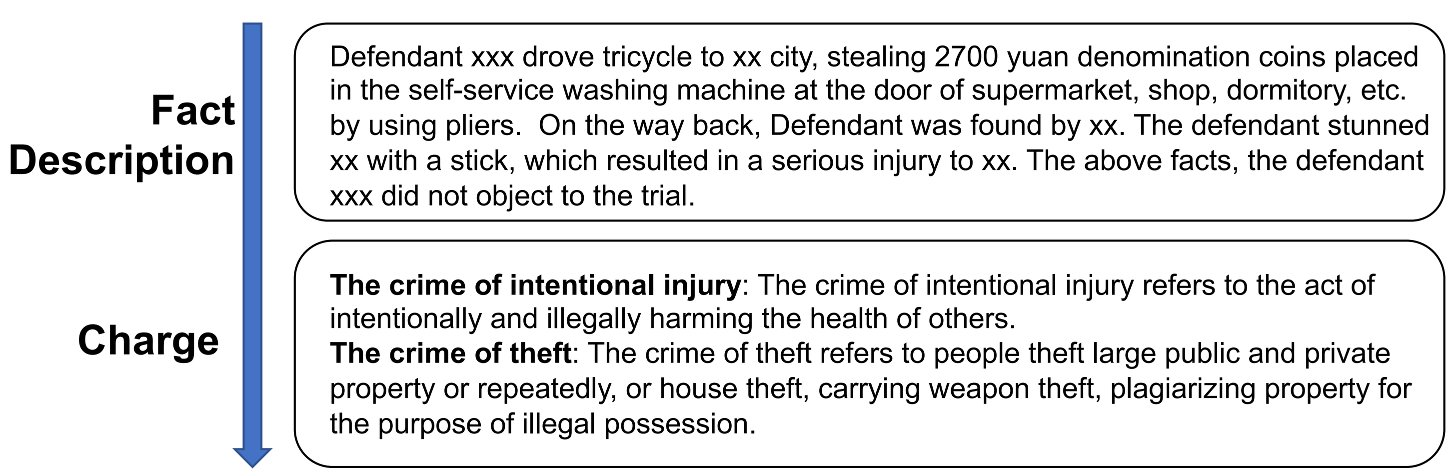

Fig. 1: An example of the judgment case, including the fact description, two charges violated.

Model

We first learn the label representations. Specially, we define two types of entities and four types of relations for the legal graph, where the entities include labels and words. The word is the meaningful keyword extracted from the articles. A label is described by a clear article. Relations include adjacency, co-group, co-occurrence, component-to-charge. Based on the generated legal graph, we exploit the local context information of the labels by exploiting the label node context information in the graph. Next, we propose a attention-based model to predict the labels by introducing label embeddings learned from legal graph. The overall architecture of LGN has been shown in Fig.2.

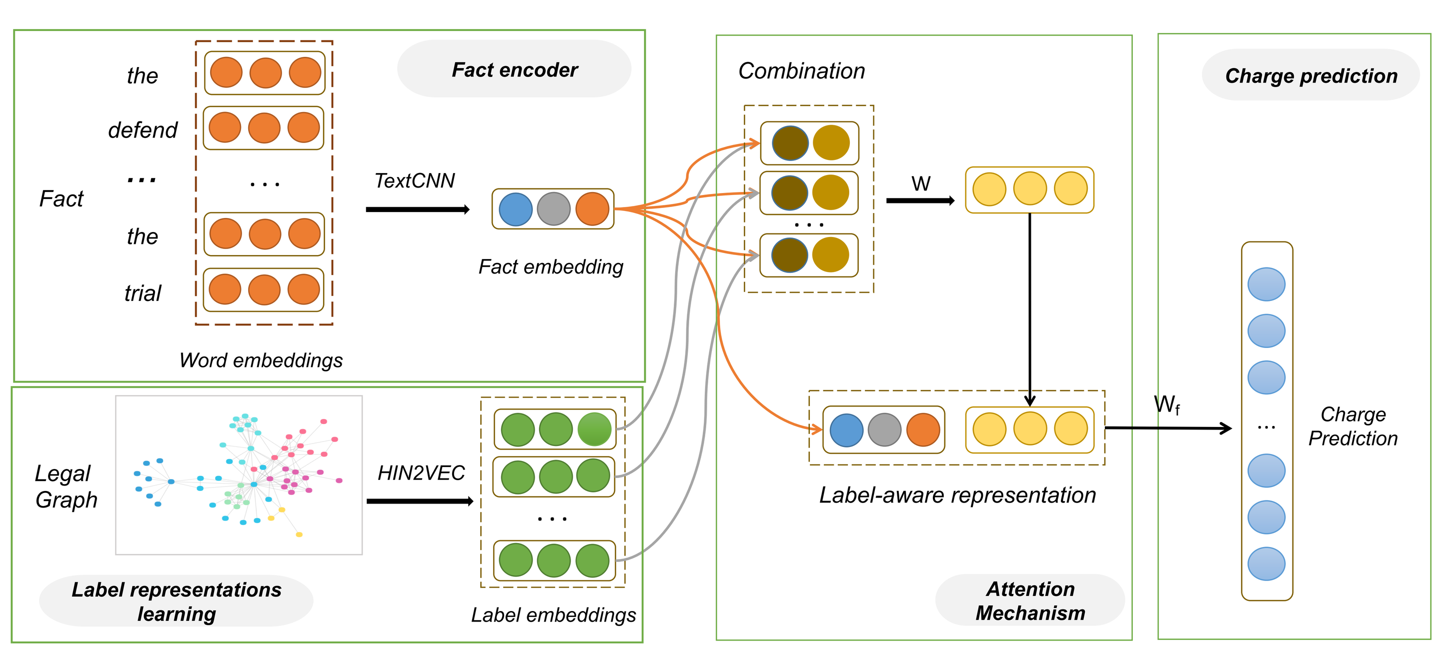

Fig. 2: The overall architecture of the proposed Legal Graph Network (LGN). It is divided into four parts: label representations learning, fact encoder, label-aware attention mechanism and charge prediction.

Results

We collect and construct three different charge prediction datasets for our experiments, including PKU, HLS, and CAIL. Then we adopt the five representative methods of text classification as baselines to compare with our model, including traditional learning methods and deep learning methods. As shown in the Fig. 3, we can see that our model significantly and consistently outperforms all the baselines and has a huge advantage over three datasets and achieves promising improvements (1.26% $\sim$ 1.92%, 1.37% $\sim$ 3.2%, and 2.12% $\sim$ 3.32% absolutely on three metrics respectively).

Datasets	PKU			HLS			CAIL		
Metrics	MP	MR	MF	MP	MR	MF	MP	MR	MF
MLKNN	62.85	32.23	39.52	55.91	27.67	33.87	-	-	-
SVM	63.13	53.75	56.56	67.77	54.03	56.79	76.96	52.40	59.06
FastText	87.45	55.23	64.81	63.83	56.01	60.22	62.64	56.03	58.05
BiGRU-ML	87.45	59.25	68.01	82.39	52.15	60.62	84.29	56.23	64.87
TextCNN-ML	87.88	60.37	69.35	82.04	54.26	61.99	84.19	54.85	63.20
Our model	**89.14**	**63.57**	**72.28**	**84.31**	**57.38**	**65.31**	**85.55**	**58.60**	**66.99**
Improvement	1.26	3.2	2.93	1.92	1.37	3.32	1.26	2.57	2.12
Deviation	0.02	0.12	0.14	0.03	0.15	0.21	0.05	0.21	0.31

Fig. 3: Charge prediction results of three datasets. "—" means that the result can not be trained due to device memory limitations. Marked black denotes the best results on each dataset. Deviation means the difference of predicting charges performance for several times on each dataset.

We further compare our model to other baseline methods on charge prediction for the in-depth analysis. Compared against traditional models, manually defined feature extraction methods cause the extracted features limited. Hence, it is not surprising to see that traditional learning methods obtain the worst or the second worst performance in terms of all the evaluating indicator. Deep learning models are superior to the traditional learning methods in the charge prediction task. FastText has faster speed and smaller memory usage as the characteristics it has. However, its performance on various datasets is unstable. BiGRU and TextCNN ignore the representation information of labels, so they are not better than our model. Therefore, drawing conclusion from the above analysis, our method has its own unique advantages.

Ablation Test and Case Study

Our method is characterized by the incorporation of various relations to explore the label space. Thus, we design an ablation test to investigate the effectiveness of these relations in the legal graph. We first construct four graph structures. The latter graph structure considers one more relation than the former. For example, the first graph structure only considers co-occurrence relation, then we add one other kind of relation to the first graph to construct the second graph structure. After inputting both the label representations learned from the various graph structure and fact representations into our proposed model, we see the charge prediction results. If the prediction performance is better, we can prove the validity of the newly added relation for charge prediction. We employ the HLS dataset for experiment verification. The experiment results are shown in the Fig.4. We conclude that for each newly added relation, there is a positive impact on the charge prediction.The improvement of performance is most obvious when the co-occurrence relation is added, increased by 0.8% $\sim$ 2% . Although the growth rate of other relations are not as obvious as the co-occurrence relation, the extent of the cooperation improvement of the three relations is obvious, increased by 0.5% $\sim$ 1.3%.

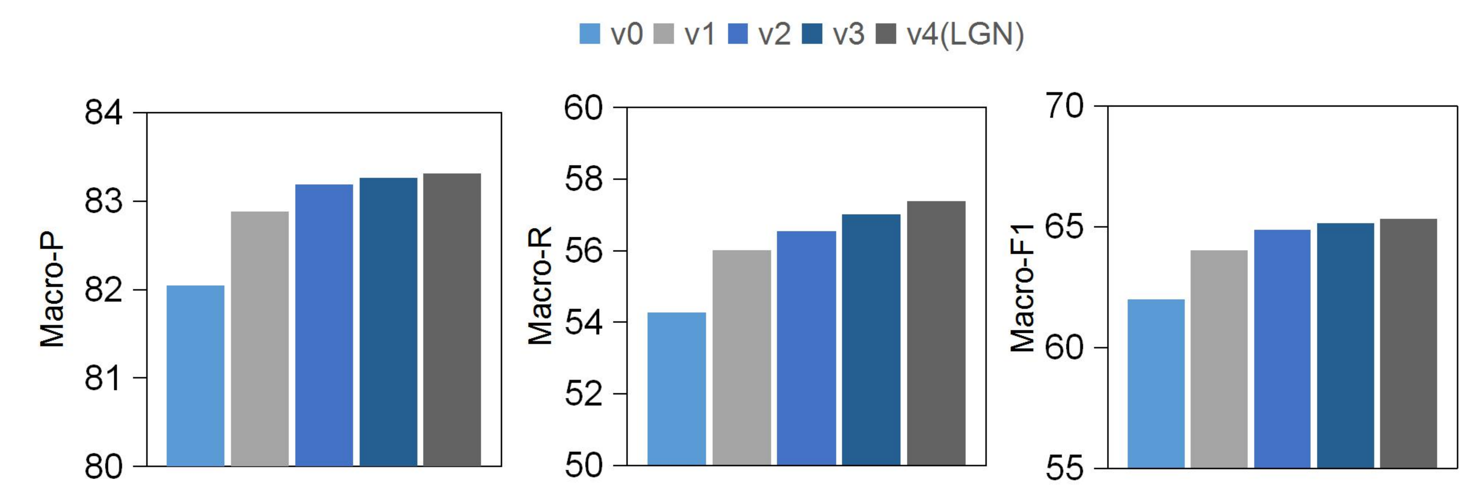

Fig. 4: Verification of the validity of each relation in the legal graph. v0 represents we just employ TextCNN without introducing label information, v1 represents we employ the graph consisting only of co-occurrence relation, v2 represents we employ the graph consisting of co-occurrence relation and component-to-charge relation, v3 represents we employ the graph consisting of co-occurrence relation, component-to-charge relation and adjacency relation, and v4 represents we employ the graph consisting of all relations.

To better verify the performance of our method, we set up a case study experiment to compare the performance of the best three models (BiGRU, TextCNN, LGN). Taking HLS as an example, 130 labels are first sorted according to their frequency of occurrences, then we split the sorted labels into 6 groups where the first five group contains 20 labels in each group and the last group contains 10 labels. In this way, the first group contains the most frequent 20 labels, while the sixth group contains the sparsest 10 labels. Given this, we compare the three models mentioned above on the first group, and we repeat the process five times. Each time we add the next label group into comparison. By this we want to test whether LGN can perform well when it faces with the label imbalance problem. The results are shown in Fig.5.

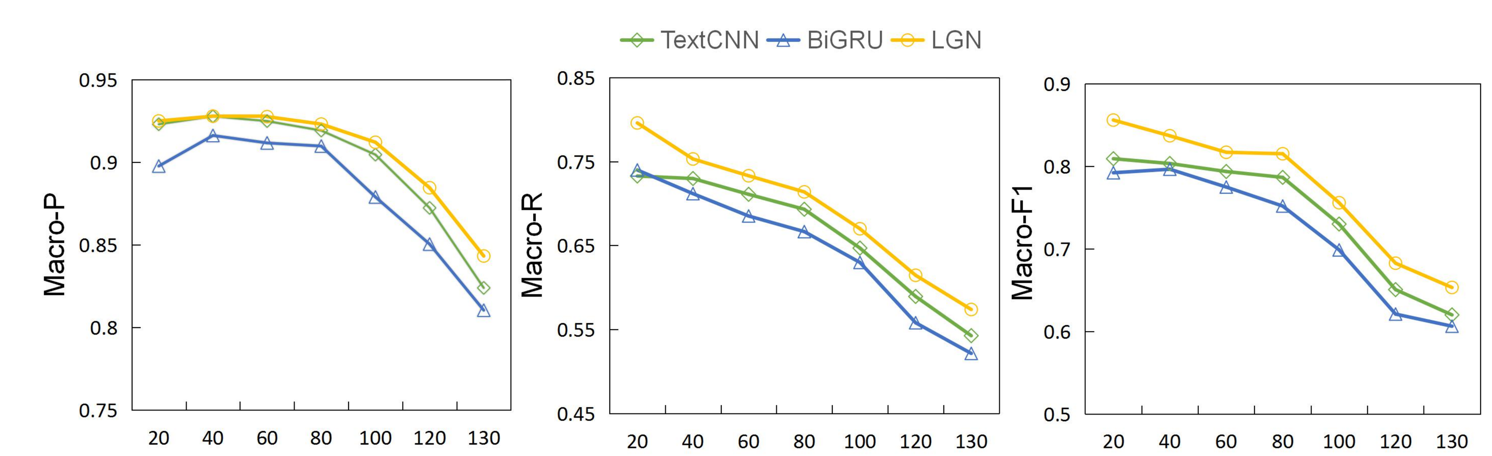

Fig. 5: Performance comparison among different label group size. The x-axis represents the label size modeled, y-axis represents the performance in terms of different evaluations metrics.

Conclusion

For the charge prediction, this paper starts with modeling label space via a multi-relations legal graph. In contrast, existing methods ignore the various label space information. We are the first to focus on this problem and propose to fuse all charge space information into a unified legal graph to solve this problem in charge prediction. Moreover, we propose a novel attention mechanism to learn label-free and label-aware fact representation jointly for charge prediction. Finally, we conduct experiments on three real-world datasets, and verify that our approach can outperform many state-of-the-art baseline methods consistently under different evaluation metrics.