SECaps: A Sequence Enhanced Capsule Model for Charge Prediction

Congqing He[1], Li Peng[1], Yuquan Le[1], Jiawei He[1], and Xiangyu Zhu[2]

College of Computer Science and Electroic Engineering, Hunan University, China[1]
JD Digits, China[2]

Abstract

Automatic charge prediction aims to predict appropriate final charges according to the fact descriptions for a given criminal case. Automatic charge prediction plays a critical role in assisting judges and lawyers to improve the efficiency of legal decisions, and thus has received much attention. Nevertheless, most existing works on automatic charge prediction perform adequately on high-frequency charges but are not yet capable of predicting few-shot charges with limited cases. In this paper, we propose a **S**equence **E**nhanced **Caps**ule model, dubbed as SECaps model, to relieve this problem. Specifically, following the work of capsule networks, we propose the seq-caps layer, which considers sequence information and spatial information of legal texts simultaneously. Then we design an attention residual unit, which provides auxiliary information for charge prediction. In addition, SECaps model introduces focal loss, which relieves the problem of imbalanced charges. Comparing the state-of-the-art methods, SECaps model obtains 4.5% and 6.4% absolutely considerable improvements under Macro F1 in Criminal-S and Criminal-L respectively. The experimental results consistently demonstrate the superiorities and competitiveness of SECaps model.
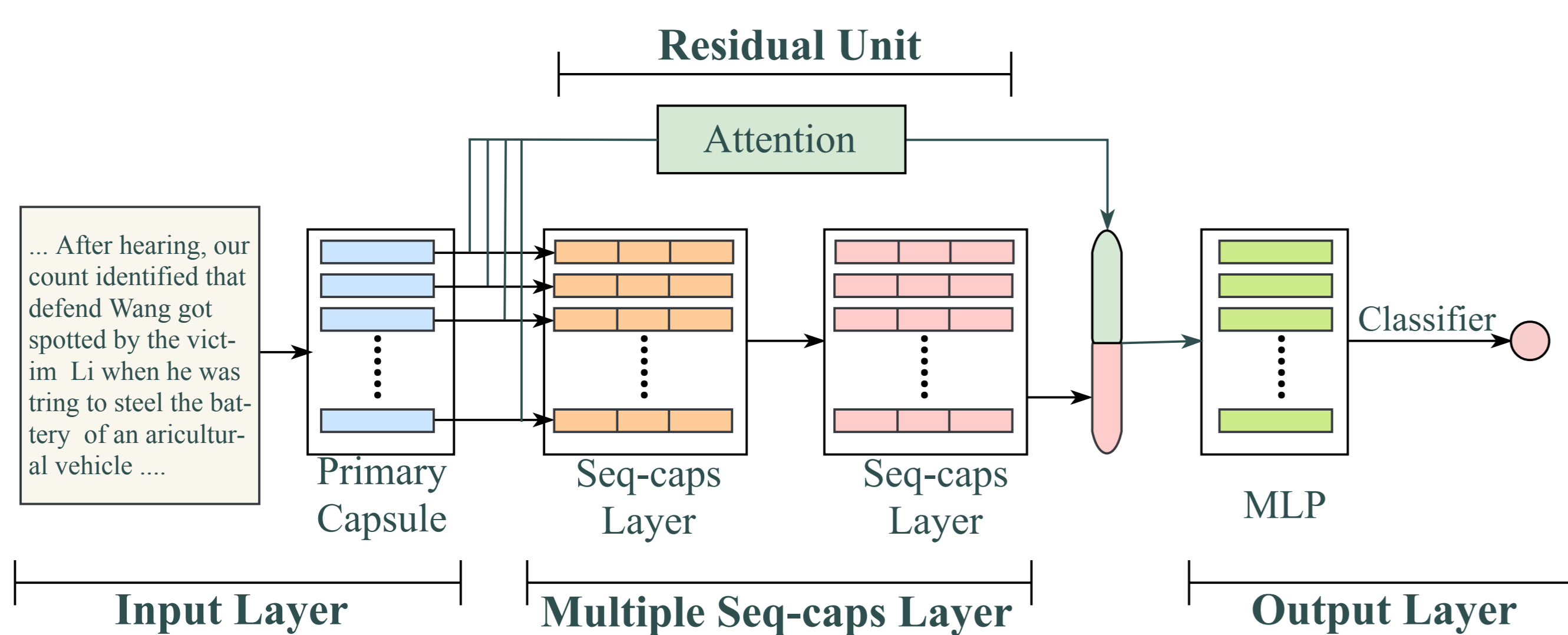
SECaps Model

Figure 1: The architecture of SECaps model, including Input layer, Multiple seq-caps layer, Attention, and Output layer.

SECaps model includes four parts: Input layer, Multiple seq-caps layer, Attention, Output layer. Figure 1 shows the architecture of SECaps model.

Input layer: In this part, we treat the fact description of a case as a sequence of words $x = \{x_1, x_2, \cdots, x_n\}$, then, each word of the sequence is transformed to the primary capsule.

Multiple seq-caps layer: This part has two seq-caps layers. We treat the word embeddings as primary capsules, and then transfer primary capsules to higher-level capsules. The seq-caps layer outputs advanced semantic representation which is captured from fact description of a case. Meanwhile, seq-caps layer restores the sequence information of fact description, which is a key factor for charge prediction.

Attention: When the multiple seq-caps layer aggregates primary capsules into higher-level capsules, the model only focuses on the most important legal case's information. Similar to He et al., we propose a novel residual unit to improve the generalization and provide auxiliary information for charge prediction. SECaps model introduces the attention mechanism as the residual unit, to encode the primary capsule which can capture the global context information. Suppose the primary capsules from the input part is $\{t_1, t_2, \cdots, t_n\}$, the residual unit's vector c is computed as follows:

$$e_i = tanh(Wt_i + b) \tag{1}$$

$$\alpha_i = \frac{exp(e_i)}{\sum_j exp(e_j)} \tag{2}$$

$$c = \sum_i \alpha_i t_i \tag{3}$$

where W is a weight matrix and b is the bias.

Output layer: In order to consider prominent features and the global context information together, we first flatten all the feature vectors from the Multiple seq-caps layer, and concatenate with the global context vector c. Then, we use a fully connected network and softmax function to generate the probability $y = (y_1, y_2, \cdots, y_k)$, where k is the number of charges. As for loss function, we apply focal loss to SECaps model. Focal loss is proposed for dense object detection initially, which address the few-shot problem by reshaping the standard cross entropy loss such that it down-weights the loss assigned to well-classified examples. It can be calculate as follows:

$$FL(y_t) = \alpha_t(1 - y_t)^\gamma log(y_t) \tag{4}$$

where y_t is the t-th output of y, $\alpha_t \in [0, 1]$ is weighting factor and γ is the focusing parameter.
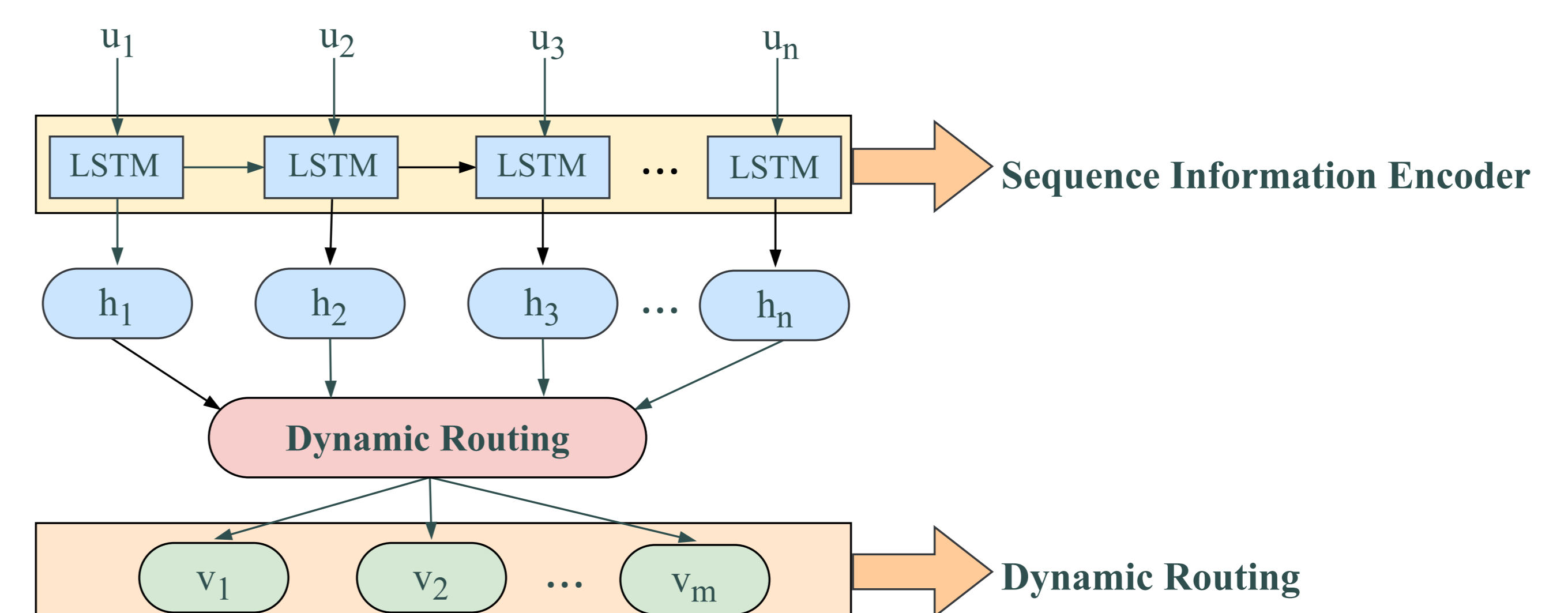
Seq-caps Layer

Figure 2: The framework of seq-caps layer, the input capsules $u = \{u_1, u_2, \cdots, u_n\}$ is lower-level capsules and the output capsules of the seq-caps layer is $v = \{v_1, v_2, \cdots, v_m\}$.

Figure 2 shows the framework of seq-caps layer. Suppose the input capsules of the seq-caps layer is $u = \{u_1, u_2, \cdots, u_n\}$, seq-caps layer follows two component:

● **Sequence Information Encoder**. It uses a Long Short-Term Memory (LSTM) encoder as a sub layer to restore sequence information of the input capsules. In this step, we get hidden layer $h_1, h_2, \cdots, h_n = LSTM(u_1, u_2, \cdots, u_n)$.

● **Dynamic Routing**. It transforms the hidden layer to higher-level capsules by using dynamic routing mechanism. In this step, we get higher-level capsules $\{v_1, v_2, \cdots, v_m\}$.

Main Results

Datasets	Criminal-S				Criminal-M				Criminal-L			
Metrics	Acc.	MP	MR	MF	Acc.	MP	MR	MF	Acc.	MP	MR	MF
TFIDF+SVM	85.8	49.7	41.9	43.5	89.6	58.8	50.1	52.1	91.8	67.5	54.1	57.5
CNN	91.9	50.5	44.9	46.1	93.5	57.6	48.1	50.5	93.9	66.0	50.3	54.7
CNN-200	92.6	51.1	46.3	47.3	92.8	56.2	50.0	50.8	94.1	61.9	50.0	53.1
LSTM	93.5	59.4	58.6	57.3	94.7	65.8	63.0	62.6	95.5	69.8	67.0	66.8
LSTM-200	92.7	60.0	58.4	57.0	94.4	66.5	62.4	62.7	95.1	72.8	66.7	67.9
Fact-Law Att.	92.8	57.0	53.9	53.4	94.7	66.7	60.4	61.8	95.7	73.3	67.1	68.6
Attribute-att.	93.4	66.7	69.2	64.9	94.4	68.3	69.2	67.1	95.8	75.8	73.7	73.1
SECaps Model	**94.8**	**71.3**	**70.3**	**69.4**	**95.4**	**71.3**	**70.2**	**69.6**	**96.0**	**81.9**	**79.7**	**79.5**

Table 1 shows the results of SECaps model and baselines on three datasets. Overall, we find that SECaps model outperforms all previous baselines with a significant margin on three datasets.

Few-shot Charges Results

Charge Type	Low-frequency	Medium-frequency	High-frequency
Charge Number	49	51	49
LSTM-200	32.6	55.0	83.3
Attribute-att.	49.7	60.0	85.2
SECaps Model	**53.8**	**65.5**	**89.0**

Table 2 shows the performance of SECaps model with different frequency on Criminal-S, we report the low-frequency, the medium-frequency and the high-frequency results of MF. From the table, we see that the MF of low-frequency is 53.8% which achieves more than 65% improvements than LSTM-200 and obtains a considerable improvement by 4.1% over the state-of-the-art baseline.

Conclusions

In this paper, we focus on the few-shot problem of charge prediction according to the fact descriptions of criminal cases. To alleviate the problem, we propose a SECaps model for charge prediction. In particular, SECaps model employs the seq-caps layer, which can capture characteristics of the sequence and abstract advanced semantic features simultaneously, and then combine with focal loss, which can handle the unbalanced problem of charges. Experiments on the real-world datasets show that SECaps model achieves 69.4%, 69.6%, 79.5% Macro F1 on three datasets respectively, surpassing existing state-of-the-art methods by a considerable margin.

References

[1] Zikun Hu, Xiang Li, Cunchao Tu, Zhiyuan Liu, and Maosong Sun. Few-shot charge prediction with discriminative legal attributes. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 487–498, 2018.

[2] Sara Sabour, Nicholas Frosst, and Geoffrey E Hinton. Dynamic routing between capsules. In *Advances in neural information processing systems*, pages 3856–3866, 2017.