A Recurrent Attention Network for Judgment Prediction

Ze Yang [§] Pengfei Wang[§] Lei Zhang [§] Linjun Shou [†] Wenwen Xu [‡]
[§] School of Computer Science, Beijing University of Posts and Telecommunications, Beijing, China
{yangze01, wangpengfei, zlei}@bupt.edu.cn

Problem Definition

Table 1: An example of the judgment case, including a fact and two articles, where article 263 is the one that the fact violated.

Fact: *At 0:00 on October 9, 2011, the defendant Shi Jiliang, after a prior negotiation, was driven by Wei Mouyi to drive a BYD car...*
Articles: ***Article263: Anyone who robs public or private property*** *in a large amount or who has been robbed several times shall be sentenced to fixed-term imprisonment....* ***Article264: Anyone who robs public or private property*** *by violence, coercion or other means shall be sentenced to fixed-term imprisonment...*

Given the fact, we aim to predict the articles related to the fact.

Method

Encoder Layer With the Bi-LSTM, we obtain the hidden representation of the i-th word by concatenating the hidden state of two directions, $h_t = [\overrightarrow{h_t}; \overleftarrow{h_t}]$, then the fact x and the label l are mapped into continuous representations $H_e = [h_1, h_2, \ldots, h_m]$, $H_a = [h_1, h_2, \ldots, h_n]$, respectively. we use self-attention to get the weighted fact and label representations H_{es} and H_{as}.

Recurrent Layer In the judicial judgment, a judge carefully reads the fact to obtain the important information, and select relevant articles as candidates, then a detail analysis of semantics between fact and the candidate articles are applied to decide final result. We design a recurrent attention block to model the judge's repeated reading behavior.

we calculate the matching score matrix M between label representation and fact's word-level representations as follows:

$$\mathbf{M}(j, k) = H_m(j) \cdot H_{es}(k) \tag{1}$$

We propose a recurrent structure. Intuitively, this operation is continuously looped to learn the important mutual semantic information. The calculation process is shown as follows:

$$\begin{aligned} H_{es} &= H_{es} + H_{es}\mathbf{W}^{\alpha'}\alpha' \\ H_m &= H_m + H_m\mathbf{W}^{\beta'}\beta' \end{aligned} \tag{2}$$

Where $\mathbf{W}^{\alpha'}$ and $\mathbf{W}^{\beta'}$ is dimension transformation matrix. This process will be repeated several times as described above, after which we will get H_{er} and H_{ar}, representing H_{es} and H_m of the last circulation.

Output Layer To integrate the fact and global label information, we use both fact side and label side feature to predict the final result for a given instance in the output layer. The probability distribution over all labels is calculated as follows:

$$\vec{v}_{er} = g(H_{er}) = \frac{1}{n}\sum_{t=1}^{n} \vec{v}_{er}(t)$$

$$\vec{v}_{ar} = g(H_{ar}) = \frac{1}{n}\sum_{t=1}^{n} \vec{v}_{ar}(t)$$

$$\vec{v}_f = \vec{v}_{er} \oplus \vec{v}_{ar}$$

$$\vec{v}_o = \mathbf{W}^o \vec{v}_f + b^o$$

Here, g is the operation of average, $\vec{v}_{er}$ represent the context representation of fact, $\vec{v}_{ar}$ is the averaged representation of all labels, which represent global label feature. $\oplus$ represent concatenate operation, $\mathbf{W}^o$ and $\mathbf{b}^o$ are learnable parameters.

Model Architecture

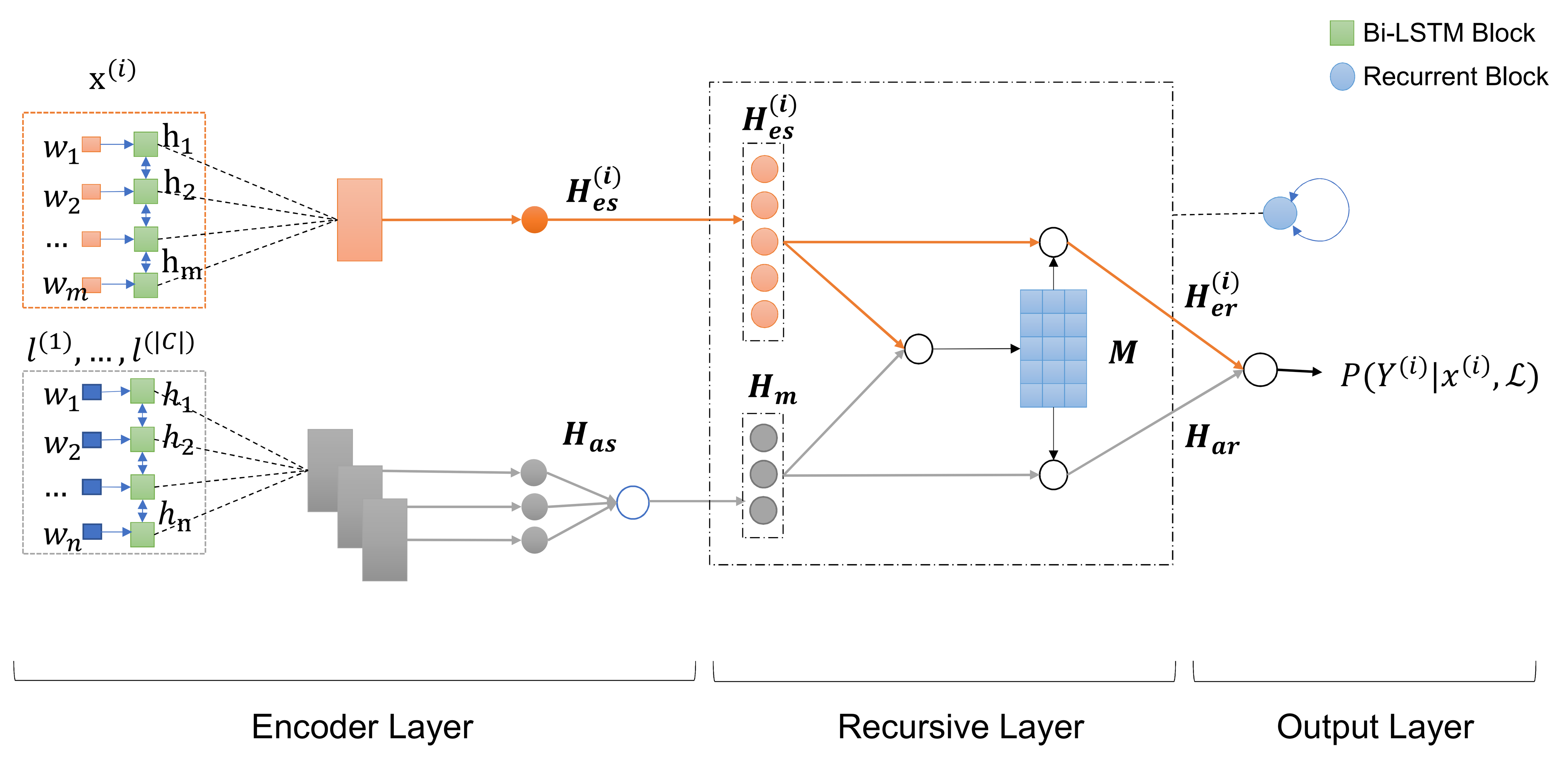

Figure 1: The overall architecture of the proposed Recurrent Attention Model (RAN).

Result

Table 2: Comparison between our method and all baselines on three datasets.

Dataset	metrics	Shallow Model				Neural Network Based Model		Attention Based Model	Our Model
		KNN	BR	CC	SVM	CNN	BiLSTM	DPAM	RSAN
CJO	MP	59.49	74.28	72.33	67.68	78.53	78.81	79.39	**81.52**
	MR	32.14	50.84	53.22	51.37	54.16	54.96	55.60	**55.75**
	MF	38.85	57.41	58.60	55.77	61.40	62.17	62.79	**63.34**
	JS	53.25	79.40	82.02	83.55	80.25	80.40	80.76	**80.96**
CAIL small	MP	31.75	41.59	42.12	43.07	78.32	79.93	80.35	**81.23**
	MR	20.11	30.23	32.49	39.66	54.73	57.77	62.03	**64.90**
	MF	22.93	33.57	35.58	40.14	61.35	63.98	67.42	**69.49**
	JS	38.85	59.74	62.59	71.98	74.12	75.09	76.00	**77.42**
CAIL 2018	MP	28.88	40.42	38.91	40.82	80.83	82.94	82.78	**84.01**
	MR	16.59	26.95	28.86	31.53	56.66	56.08	57.15	**57.52**
	MF	19.68	30.65	31.59	34.01	63.51	63.36	64.44	**64.92**
	JS	70.28	88.34	90.57	90.92	94.61	94.61	94.39	**94.68**

Conclusion

In this paper, we propose an Recurrent Attention Network that can simulate the repeated reading behavior of judge, which method can utilize the semantic mutual information between evidence and article, Extensive experimental results show that the proposed model outperform the baselines. Further analysis demonstrates that our model not only obtain label correlation information, but also capture the multiple informative attention with the recurrent block.