Chinese Relation Extraction with Multi-Grained Information and External Linguistic Knowledge

Ziran Li, Ning Ding, Zhiyuan Liu, Hai-Tao Zheng, Ying Shen

ACL2019

Problem

- Different from English, word segmentation in Chinese is harder.
- NLP models in Chinese usually suffer from segmentation errors.
- This paper introduces a character-based model for Relation Extraction in Chinese where the character representations are enriched by:

  + Word segmentation information

  + Word senses retrieved from an external knowledge base.

Overview

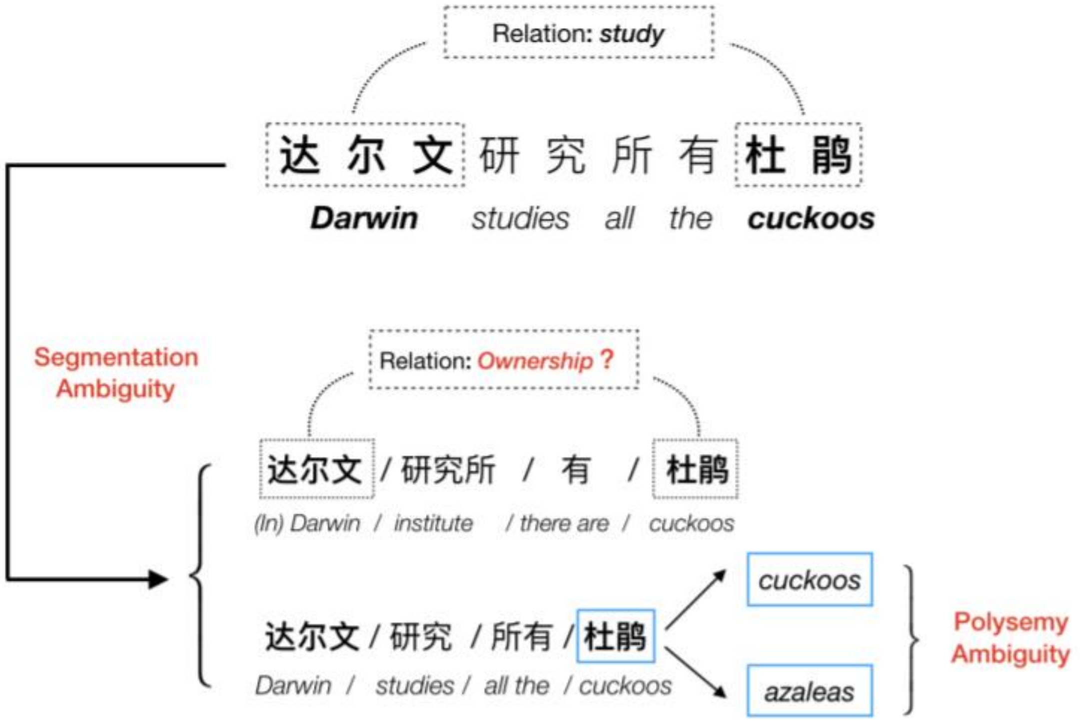


Figure 1: An example of segmentation ambiguity and polysemy ambiguity in Chinese RE.

Model:Input Representation

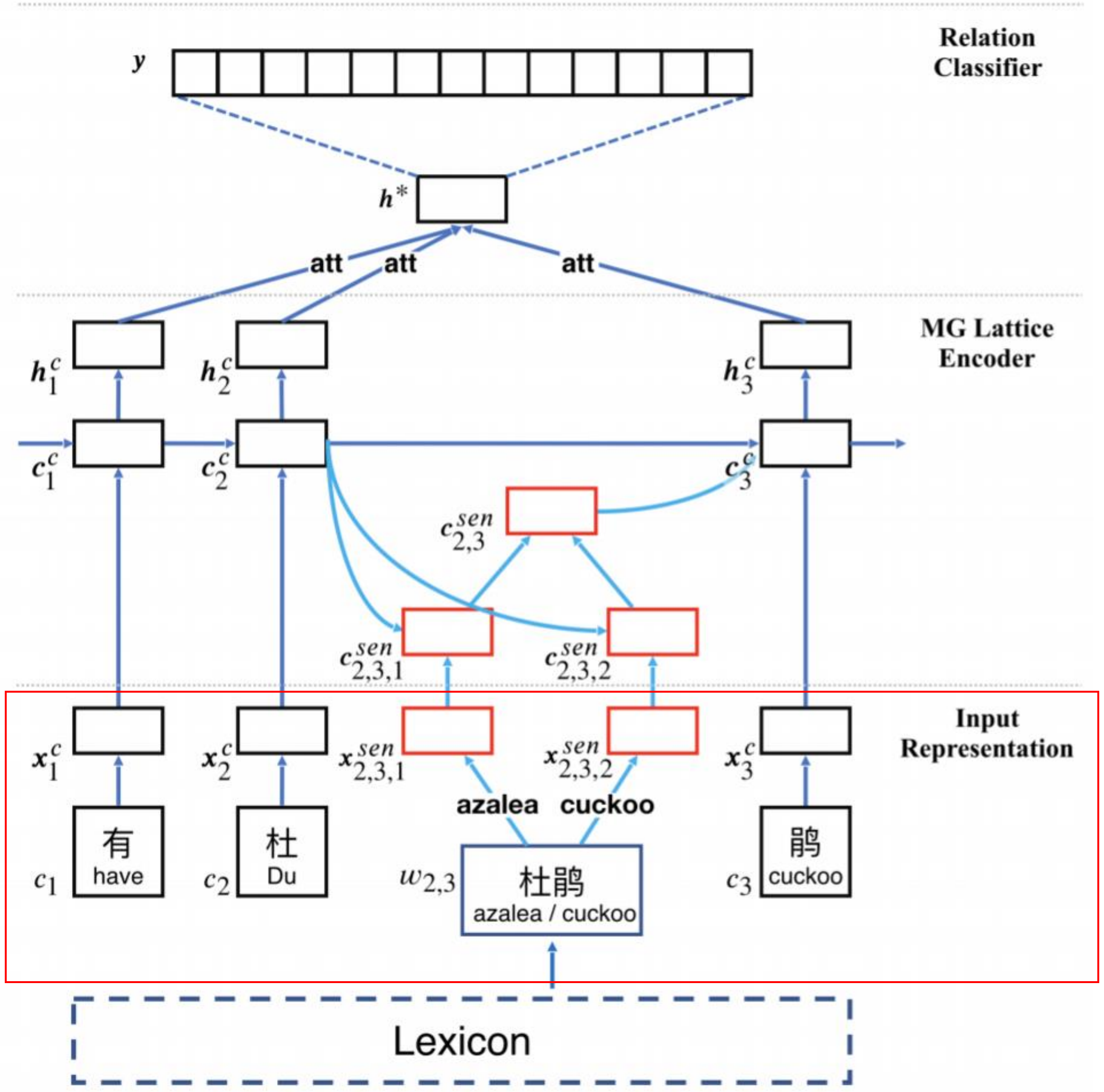
- Given two entities:

  + (b1 , e1)

  + (b2 , e2)
- Character-level embedding:

$$\boldsymbol{x}_i^c = [\boldsymbol{x}_i^{ce}; \boldsymbol{x}_i^{p_1}; \boldsymbol{x}_i^{p_2}]$$

$$p_i^1 = \begin{cases} i - b^1 & i < b^1, \\ 0 & b^1 \leq i \leq e^1, \\ i - e^1 & i > e^1, \end{cases}$$

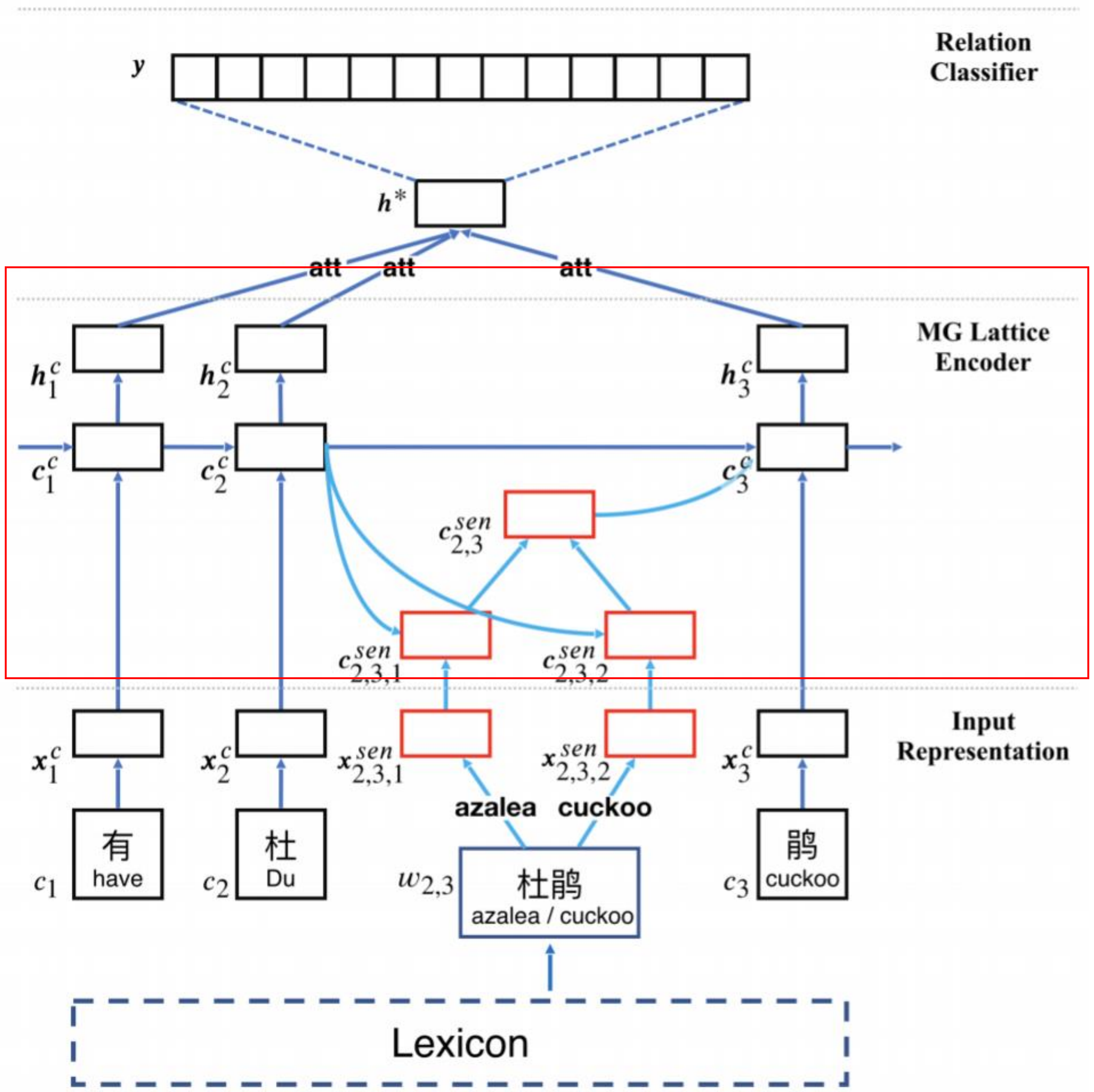


Model:Multi-grained lattice Encoder

$$\begin{cases} \boldsymbol{i}_j^c = \sigma(W_i \boldsymbol{x}_j^c + U_i \boldsymbol{h}_{j-1}^c + \boldsymbol{b}_i), \\ \boldsymbol{o}_j^c = \sigma(W_o \boldsymbol{x}_j^c + U_o \boldsymbol{h}_{j-1}^c + \boldsymbol{b}_o), \\ \boldsymbol{f}_j^c = \sigma(W_f \boldsymbol{x}_j^c + U_f \boldsymbol{h}_{j-1}^c + \boldsymbol{b}_f), \\ \tilde{\boldsymbol{c}}_j^c = \tanh(W_c \boldsymbol{x}_j^c + U_c \boldsymbol{h}_{j-1}^c + \boldsymbol{b}_c), \end{cases}$$

$$\boxed{\boldsymbol{c}_j^c = \boldsymbol{f}_j^c \odot \boldsymbol{c}_{j-1}^c + \boldsymbol{i}_j^c \odot \tilde{\boldsymbol{c}}_j^c,}$$

$$\boldsymbol{h}_j^c = \boldsymbol{o}_j^c \odot \tanh(\boldsymbol{c}_j^c),$$



Model:Multi-grained lattice Encoder

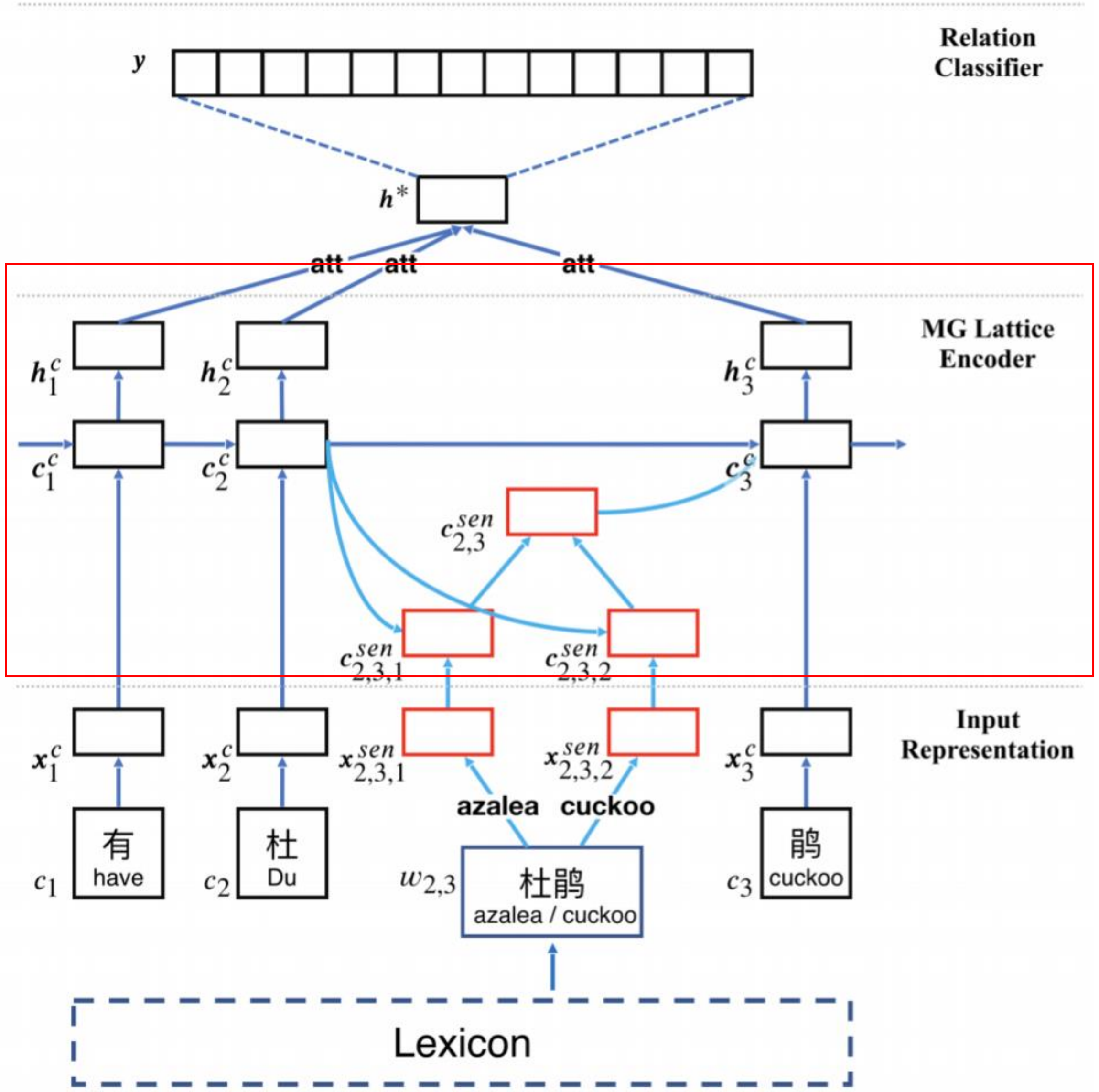
$$\begin{cases} \boldsymbol{i}_{b,e,k}^{sen} = \sigma(W_i \boldsymbol{x}_{b,e,k}^{sen} + U_i \boldsymbol{h}_b^c + \boldsymbol{b}_i), \\ \boldsymbol{f}_{b,e,k}^{sen} = \sigma(W_f \boldsymbol{x}_{b,e,k}^{sen} + U_f \boldsymbol{h}_b^c + \boldsymbol{b}_f), \\ \tilde{\boldsymbol{c}}_{b,e,k}^{sen} = \tanh(W_c \boldsymbol{x}_{b,e,k}^{sen} + U_c \boldsymbol{h}_b^c + \boldsymbol{b}_c), \end{cases}$$

$$\boldsymbol{c}_{b,e,k}^{sen} = \boldsymbol{f}_{b,e,k}^{sen} \odot \boldsymbol{c}_b^c + \boldsymbol{i}_{b,e,k}^{sen} \odot \tilde{\boldsymbol{c}}_{b,e,k}^{sen},$$

$$\boldsymbol{c}_{b,e}^{sen} = \sum_k \boldsymbol{\alpha}_{b,e,k}^{sen} \odot \boldsymbol{c}_{b,e,k}^{sen},$$

$$\boldsymbol{\alpha}_{b,e,k}^{sen} = \frac{\exp(\boldsymbol{i}_{b,e,k}^{sen})}{\sum_{k'}^{K} \exp(\boldsymbol{i}_{b,e,k'}^{sen})},$$

$$\boldsymbol{c}_e^c = \sum_{b \in \{b' | w_{b',e}^d \in \mathbb{D}\}} \boldsymbol{\alpha}_{b,e}^{sen} \odot \boldsymbol{c}_{b,e}^{sen} + \boldsymbol{\alpha}_e^c \odot \tilde{\boldsymbol{c}}_e^c.$$



Model:Multi-grained lattice Encoder

$$\begin{cases} \boldsymbol{i}^w_{b,e} = \sigma(W_i \boldsymbol{x}^w_{b,e} + U_i \boldsymbol{h}^c_b + \boldsymbol{b}_i), \\ \boldsymbol{f}^w_{b,e} = \sigma(W_f \boldsymbol{x}^w_{b,e} + U_f \boldsymbol{h}^c_b + \boldsymbol{b}_f), \\ \tilde{\boldsymbol{c}}^w_{b,e} = \tanh(W_c \boldsymbol{x}^w_{b,e} + U_c \boldsymbol{h}^c_b + \boldsymbol{b}_c), \end{cases}$$

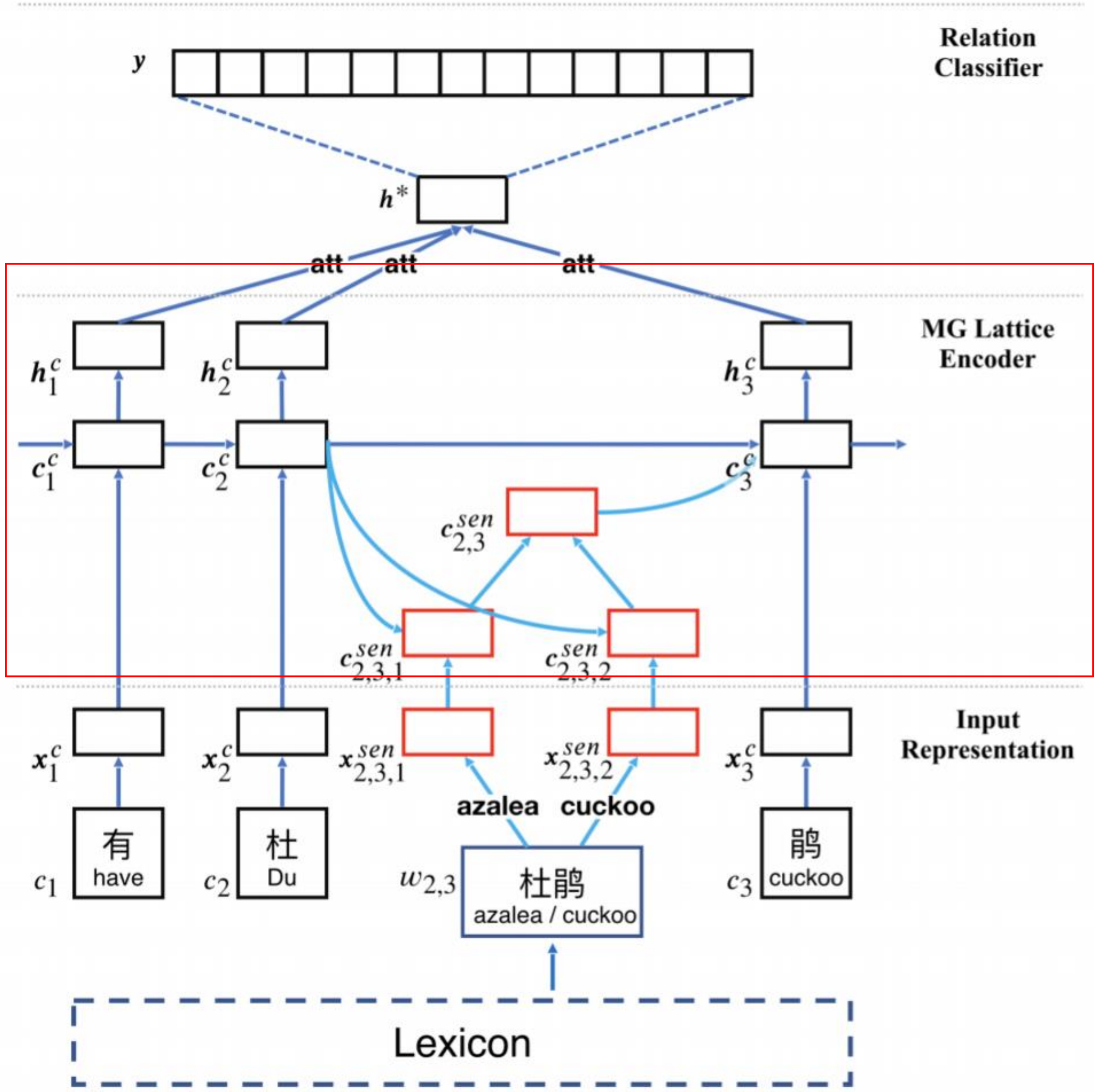
$$\boldsymbol{c}^w_{b,e} = \boldsymbol{f}^w_{b,e} \odot \boldsymbol{c}^c_b + \boldsymbol{i}^w_{b,e} \odot \tilde{\boldsymbol{c}}^w_{b,e},$$

$$\boldsymbol{i}^c_{b,e} = \sigma(W \boldsymbol{x}^c_e + U \boldsymbol{c}^w_{b,e} + \boldsymbol{b}^l).$$

$$\boldsymbol{\alpha}^c_{b,e} = \frac{\exp(\boldsymbol{i}^c_{b,e})}{\exp(\boldsymbol{i}^c_e) + \sum_{b' \in \{b'' | w_{b'',e} \in \mathbb{D}\}} \exp(\boldsymbol{i}^c_{b',e})},$$

$$\boldsymbol{\alpha}^c_e = \frac{\exp(\boldsymbol{i}^c_e)}{\exp(\boldsymbol{i}^c_e) + \sum_{b' \in \{b'' | w_{b'',e} \in \mathbb{D}\}} \exp(\boldsymbol{i}^c_{b',e})}.$$

$$\boxed{\boldsymbol{c}^c_e = \sum_{b \in \{b' | w^d_{b',e} \in \mathbb{D}\}} \boldsymbol{\alpha}^{sen}_{b,e} \odot \boldsymbol{c}^{sen}_{b,e} + \boldsymbol{\alpha}^c_e \odot \tilde{\boldsymbol{c}}^c_e.}$$



Model: Relation Classifier

- Sentence-level representation:

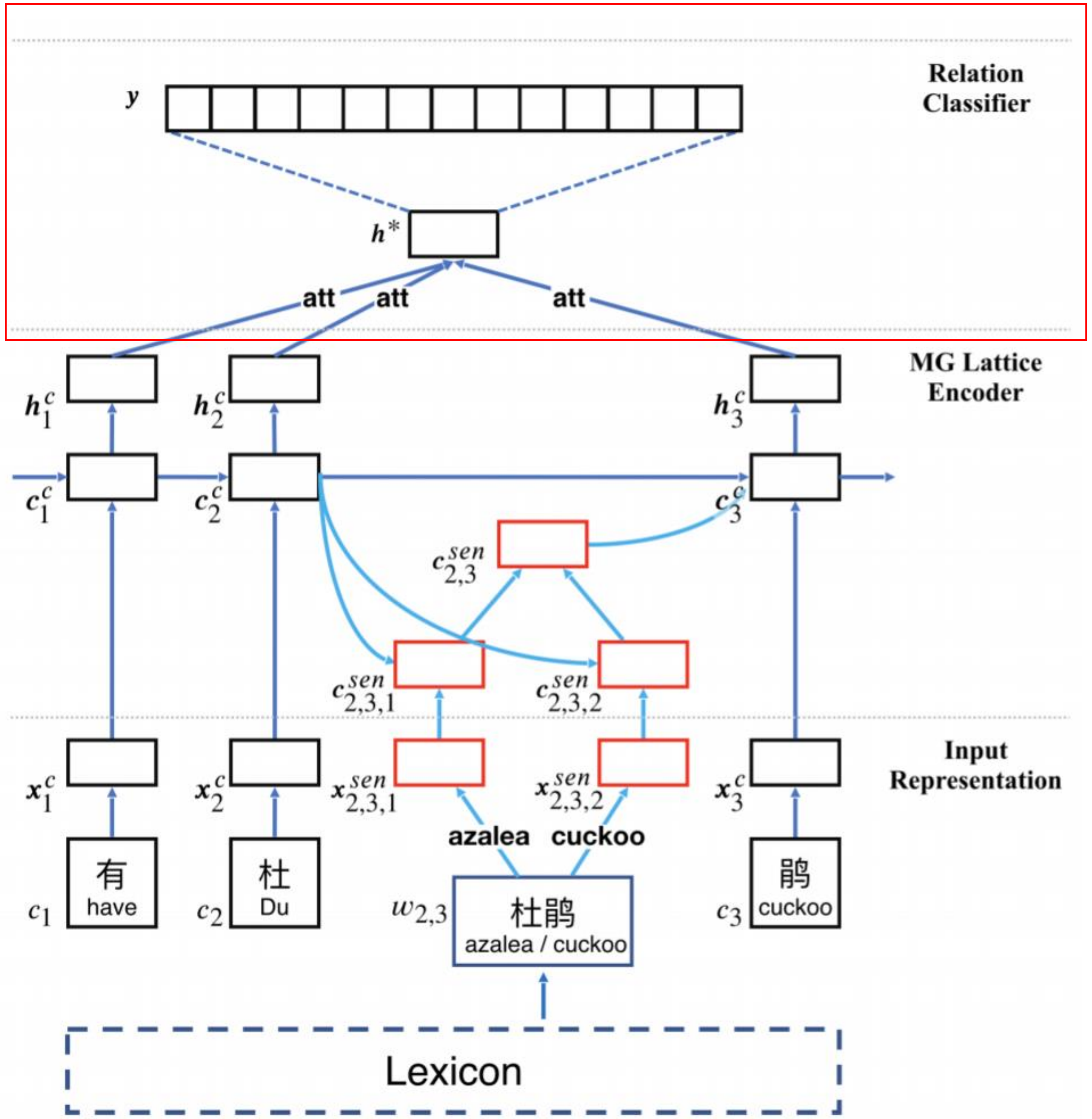
$$\boldsymbol{H} = \tanh(\boldsymbol{h}),$$

$$\boldsymbol{\alpha} = \mathrm{softmax}(\boldsymbol{w}^T \boldsymbol{H}),$$

$$\boldsymbol{h}^* = \boldsymbol{h}\boldsymbol{\alpha}^T,$$

$$\boldsymbol{o} = \boldsymbol{W}\boldsymbol{h}^* + \boldsymbol{b},$$

$$p(\boldsymbol{y}|S) = \mathrm{softmax}(\boldsymbol{o}),$$



Results

Models		FinRE	SanWen	ACE
Word-based	Word-baseline	41.23	54.26	64.43
	+char CNN	41.60	56.62	68.86
	+char LSTM	42.20	57.92	69.81
Char-based	Character-baseline	40.50	60.34	71.52
	+softword	41.42	60.69	69.81
	+bichar	40.52	61.34	71.86
	+softword + bichar	42.03	61.75	72.63
Ours	Basic Lattice	47.41	63.88	77.12
	MG Lattice	**49.26**	**65.61**	**78.17**

Table 2: F1-scores of word-baselines, character baselines and lattice-based models on all datasets.