Datawhale

Lattice CNNs for Matching Based Chinese Question Answering

团队：Datawhale 深度学习团队
汇报人： 杨夕
2019.11.23

目录

01 Lattice CNNs ——动机

短文本匹配受分词效果影响;

融合词级和字级信息的方法受到原有词序结构的影响；

直接将汉字与相应的词组合起来可能会失去这些字所能表达的意义。 由于顺序输入的原因，他们要么在处理字符序列时丢失字级信息，要么不得不做出分词选择。

01 Lattice CNNs ——动机

特定的任务，如问题回答(QA)可以提出进一步的挑战，短文本匹配。

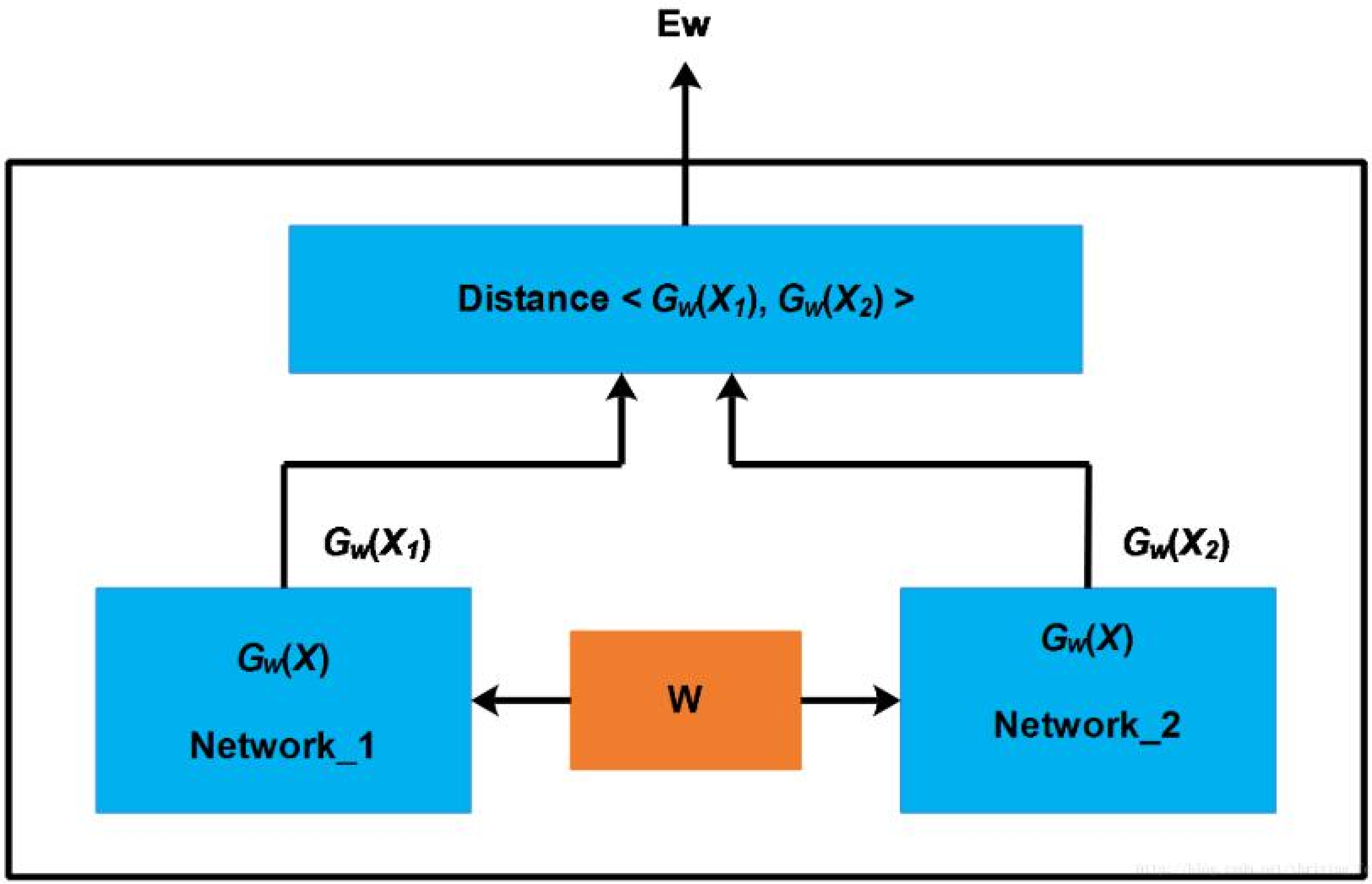
基于文档的问答系统(DBQA)。匹配度反映对一个给定的问题，一个句子是他的回答的概率，问题和回答来源不同，因此会存在风格和句法结构都不同的问题。

基于知识的问题回答(KBQA)。一个关键任务是对知识库的谓词短语来匹配问题的关系表达式。

Lattice CNNs —— 方法介绍

本文提出了一种用于中文问答中短文本匹配的多粒度方法，该方法利用基于 Lattice 的CNN提取单词 Lattice 上的句子级特征。 具体而言，LCN 不再依赖于字符或单词级别序列，而是将单词 Lattice 作为输入，其中每个可能的单词和字符将被平等对待并具有各自的上下文，以便它们可以在每一层进行交互。 对于每层中的每个单词，LCN可以通过合并方法以不同的粒度捕获不同的上下文单词。

Siamese architecture



整体框架介绍

For our models, we use multi-layer CNNs for sentence representation. Residual connections (He et al. 2016) are used between convolutional layers to enrich features and make it easier to train. Then, max-pooling summarizes the global features to get the sentence level representations, which are merged via element-wise multiplication. The matching score is produced by a multi-layer perceptron (MLP) with one hidden layer based on the merged vector. The fusing and matching procedure is formulated as follows:

$$s = \sigma(\boldsymbol{W}_2 \, \mathrm{ReLU}(\boldsymbol{W}_1(\boldsymbol{f}_{qu} \odot \boldsymbol{f}_{can}) + \boldsymbol{b}_1^T) + \boldsymbol{b}_2^T) \quad (1)$$

where $\boldsymbol{f}_{qu}$ and $\boldsymbol{f}_{can}$ are feature vectors of question and candidate (sentence or predicate) separately encoded by CNNs, σ is the sigmoid function, $\boldsymbol{W}_2, \boldsymbol{W}_1, \boldsymbol{b}_1^T, \boldsymbol{b}_2^T$ are parameters, and $\odot$ is element-wise multiplication. The training objective

损失函数介绍

and $\odot$ is element-wise multiplication. The training objective is to minimize the binary cross-entropy loss, defined as:

$$L = -\sum_{i=1}^{N}[y_i log(s_i) + (1 - y_i) log(1 - s_i)] \quad (2)$$

where y_i is the $\{0,1\}$ label for the i_{th} training pair.

Lattice CNNs —— 方法介绍

关键问题： 句子表示可以是原始 CNN，也可以是 Lattice CNN。在原始 CNN 中，卷积核按照顺序扫描每个 n-gram，并得到一个特征向量，该向量可以看作是中心词的表示，并被传递至下一层。但是，每一个词在每一个 lattice 中可能具有不同粒度的上下文词，并且可以被视为具有相同长度的卷积核的中心。因此，不同于原始 CNN，lattice CNN 对于一个词可能产生多个特征向量，这是将标准CNN直接用于lattice输入的关键挑战。

02 Lattice CNNs —— 方法介绍

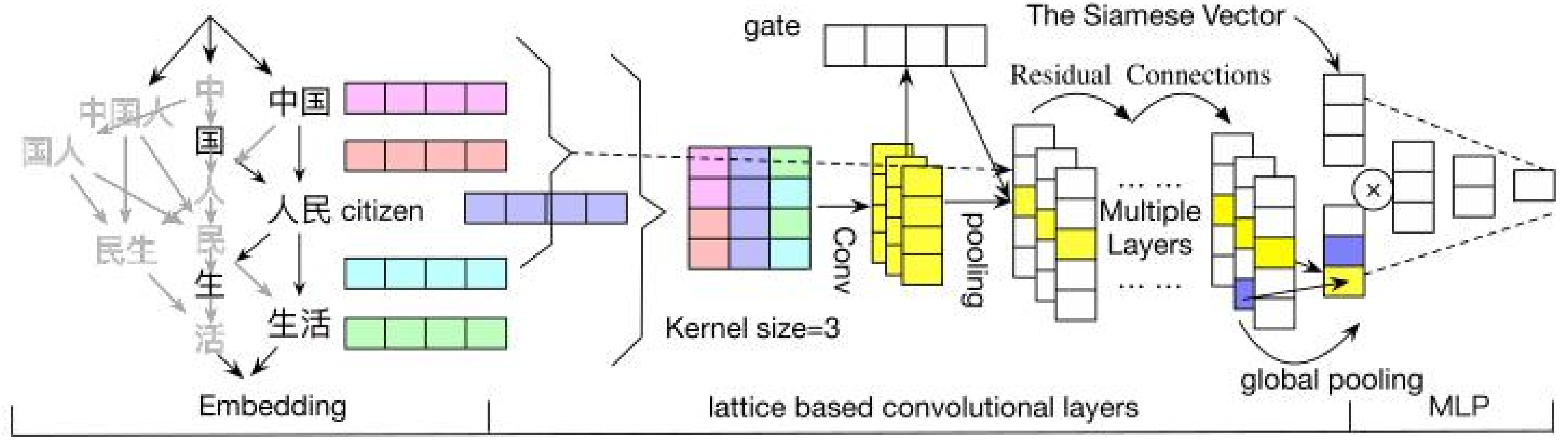


Figure 2: An illustration of our LCN-gated, when “people” is being considered as the center of convolutional spans.

Word lattice

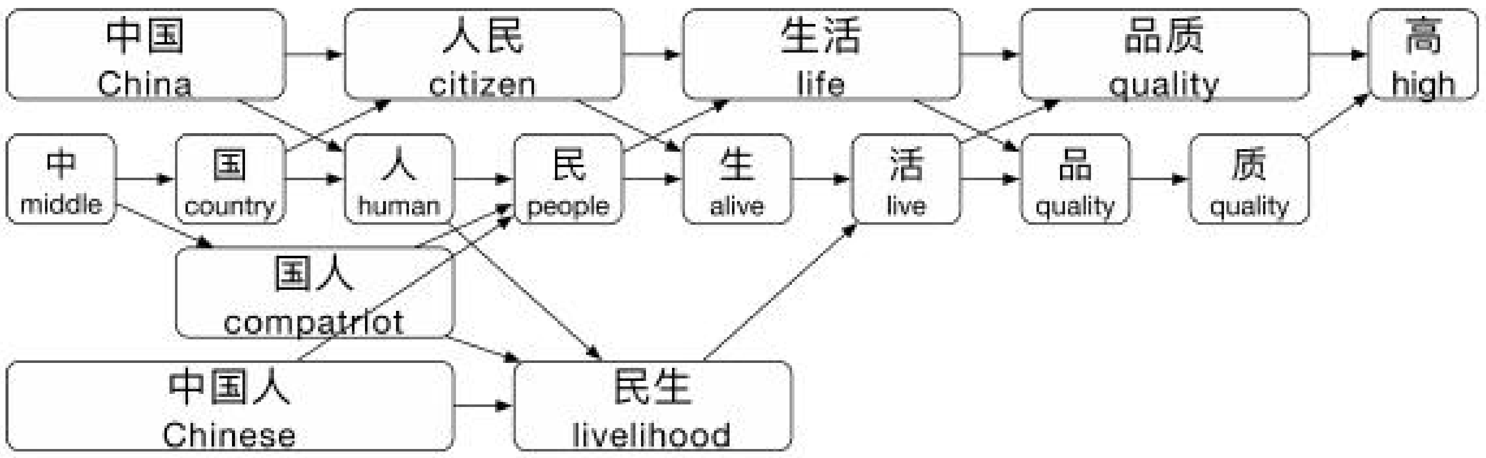


Figure 1: A word lattice for the phrase "Chinese people have high quality of life."

lattice based CNN layer

卷积核尺寸为n 的lattice CNN层对词w在word lattice G=<V,E>G=<V,E>下的输出特征向量是:

$$
\begin{aligned}
F_w = g\{f(\boldsymbol{W}_c(\boldsymbol{v}_{\boldsymbol{w}_1} : \dots : \boldsymbol{v}_{\boldsymbol{w}_n}) + \boldsymbol{b}_c^T) | \\
\forall i, w_i \in V, (w_i, w_{i+1}) \in E, w_{\lceil \frac{n+1}{2} \rceil} = w\}
\end{aligned}
$$

门池化的公式表示如下所示:

$$
\begin{aligned}
\alpha_1, \dots, \alpha_t &= \mathtt{softmax}\{\boldsymbol{v}_g^T \boldsymbol{v}_1 + b_g, \dots, \boldsymbol{v}_g^T \boldsymbol{v}_t + b_g\} \\
\mathtt{gated\text{-}pooling}\{\boldsymbol{v}_1, \dots, \boldsymbol{v}_t\} &= \sum_{i=1}^{n} \alpha_i \times \boldsymbol{v}_i
\end{aligned}
$$

数据集选取

DBQA：是一个基于文档的问题回答数据集。 在测试集中有8.8 k 的问题和182k 的问句对用于训练，6k 的问题和123k 的问句对用于测试。

KBQA：是一种基于知识的关系抽取数据集。 在训练集中有14.3 k 问题，其中问题谓词对为273k，问题谓词对为156k，问题谓词对为9.4 k。

03 Lattice CNNs —— 实验结果分析

实验结果

	DBQA			KBRE	
	MAP	MRR	P@1	P@1	MRR
MatchZoo					
Arc1	.4006	.4011	22.39%	32.18%	.5144
Arc2	.4780	.4785	30.47%	76.07%	.8518
CDSSM	.5344	.5349	36.45%	68.90%	.7974
MP	.7715	.7723	65.61%	86.21%	.9137
MV-LSTM	**.8154**	**.8162**	**71.71%**	**86.87%**	**.9271**
State-of-the-Art DBQA					
(Fu et al. 2016)	.8586	.8592	79.06%	—	—
(Xie 2017)*	**.8763**	**.8768**	—	—	—
Single Granularity CNNs					
CNN-jieba	.8281	.8289	75.10%	86.85%	.9152
CNN-PKU	.8339	.8343	76.00%	89.87%	.9370
CNN-CTB	.8341	.8347	76.04%	88.92%	.9302
CNN-char	**.8803**	**.8809**	**82.09%**	**93.06%**	**.9570**
Word Combine CNNs					
jieba+PKU	.8486	.8490	77.62%	90.57%	.9417
PKU+CTB	.8435	.8440	77.09%	90.48%	.9410
CTB+jieba	**.8499**	**.8504**	**78.06%**	90.29%	.9399
PKU+CTB+jieba	.8494	.8498	78.04%	**91.16%**	**.9450**
Word+Char CNNs					
Word+Char	.8566	.8570	78.94%	91.64%	.9489
Char+Word	.8728	.8735	80.76%	92.78%	.9561
Char+Lattice	**.8810**	**.8815**	**81.97%**	**93.12%**	**.9582**
DGCs					
DGC-ave	**.8868**	**.8873**	**83.02%**	**93.49%**	**.9602**
DGC-max	.8811	.8818	82.01%	92.79%	.9553
DGC-gated	.8790	.8795	81.69%	92.88%	.9562
LCNs					
LCN-ave	.8864	.8869	83.14%	**93.60%**	**.9609**
LCN-max	.8870	.8875	83.06%	93.54%	.9604
LCN-gated	**.8895**	**.8902**	**83.24%**	93.32%	.9592

敬请各位大佬批评指正