

学界 | DiSAN: RNN/CNN 免费语言理解的定向自助注意网络

2017-11-17 机器海岸线

选自 arXiv

作者: Tao Shen, Tianyi Zhou, Guodong Long, Jing Jiang, Shirui Pan, Chengqi Zhang 等

机器海岸线编译

参与: 方建勇

DiSAN: Directional Self-Attention Network for RNN/CNN-free Language Understanding

Tao Shen[†]
Jing Jiang[†]

Tianyi Zhou[‡]
Shirui Pan[†]

Guodong Long[†]
Chengqi Zhang[†]

[†]Centre of Artificial Intelligence, University of Technology Sydney

[‡]Paul G. Allen School of Computer Science & Engineering, University of Washington

tao.shen@student.uts.edu.au, tianyizh@uw.edu

{guodong.long, jing.jiang, shirui.pan, chengqi.zhang}@uts.edu.au

论文链接: <https://arxiv.org/pdf/1709.04696>

摘要: 递归神经网络 (RNN) 和卷积神经网络 (CNN) 广泛应用于 NLP 任务, 以分别捕获长期和局部依赖。注意机制最近引起了极大的兴趣, 由于其高度可并行化的计算, 显著减少训练时间, 并在建模依赖的灵活性。我们提出了一种新颖的注意机制, 其中来自输入序列的元素之间的关注是有方向性和多维的, 即在特征方面。然后建立一个轻量级的神经网络“定向自我注意网络 (DiSAN)”, 在没有任何 RNN / CNN 结构的情况下完全基于预期的注意力来学习句子嵌入。DiSAN 只由一个定向自编码的时间序列编码, 然后是一个多维注意, 将该序列压缩成一个向量表示。尽管这种简单的形式, DiSAN 在预测质量和效率方面都优于复杂的 RNN / CNN 模型。它在所有句子编码方法中达到了最好的测试精度, 并且在斯坦福自然语言推理 (SNLI) 数据集上将最近的最佳结果提高了大约 1: 0%, 并且显示了斯坦福大学的最新测试精度情感树库 (SST), 涉及构成知识的句子 (SICK), TREC 问题类型分类和多种类型自然语言推理 (MultiNLI) 数据集。

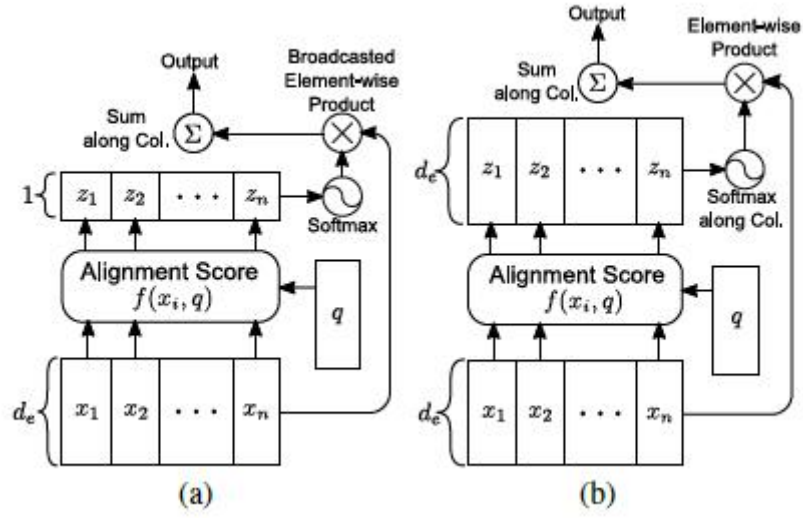


图 1: (a) 传统 (加性/乘性) 注意力和 (b) 多维注意力。 z_i 表示比对应得分 $f(x_i, q)$ ，其在 (a) 中的标量，而 (b) 中是矢量。

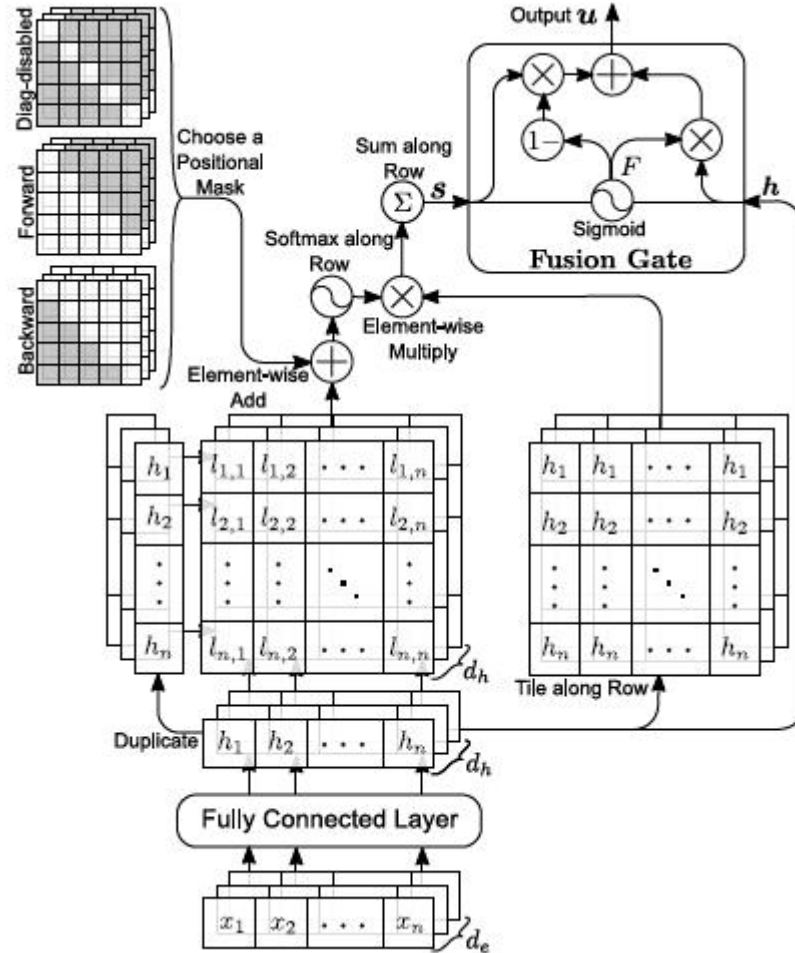


图 2: 定向自我注意机制, 这里我们使用 $l_{i,j}$ to denote $f(h_i, h_j)$ in Eq. (15).

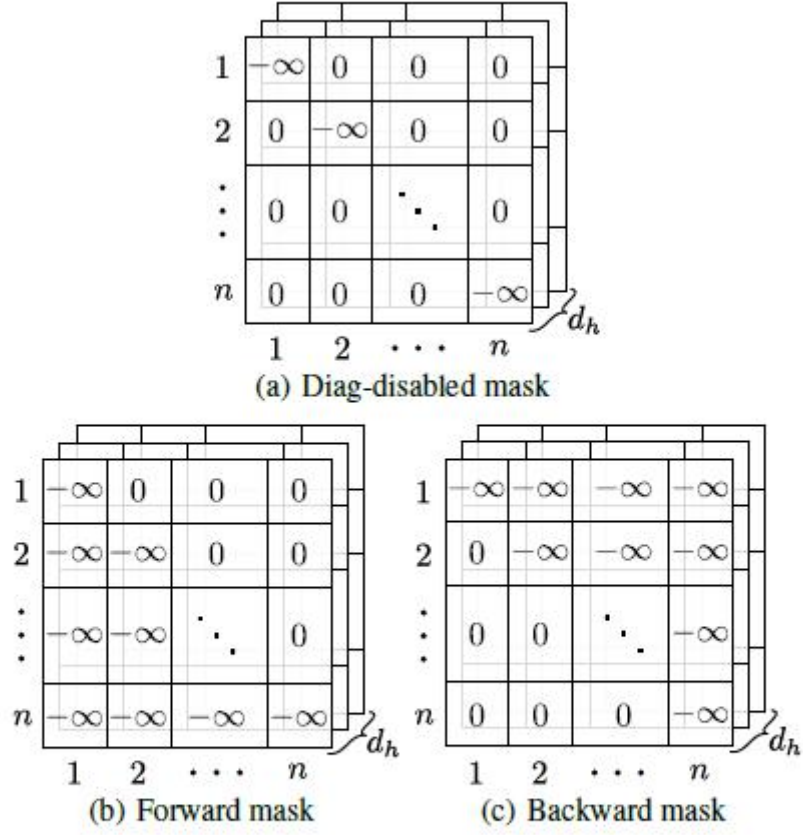


Figure 3: Three positional masks: (a) is the diag-disabled mask M^{diag} ; (b) and (c) are forward mask M^{fw} and backward mask M^{bw} .

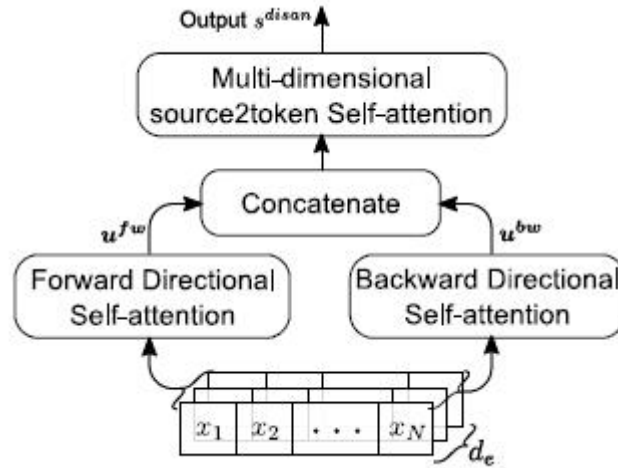


图 4: 定向自我关注网络 (DiSAN)。

Model Name	$ \theta $	T(s)/epoch	Train Accu(%)	Test Accu(%)
Unlexicalized features (Bowman et al. 2015)			49.4	50.4
+ Unigram and bigram features (Bowman et al. 2015)			99.7	78.2
100D LSTM encoders (Bowman et al. 2015)	0.2m		84.8	77.6
300D LSTM encoders (Bowman et al. 2016)	3.0m		83.9	80.6
1024D GRU encoders (Vendrov et al. 2016)	15m		98.8	81.4
300D Tree-based CNN encoders (Mou et al. 2016)	3.5m		83.3	82.1
300D SPINN-PI encoders (Bowman et al. 2016)	3.7m		89.2	83.2
600D Bi-LSTM encoders (Liu et al. 2016)	2.0m		86.4	83.3
300D NTI-SLSTM-LSTM encoders (Munkhdalai and Yu 2017b)	4.0m		82.5	83.4
600D Bi-LSTM encoders+intra-attention (Liu et al. 2016)	2.8m		84.5	84.2
300D NSE encoders (Munkhdalai and Yu 2017a)	3.0m		86.2	84.6
Word Embedding with additive attention	0.45m	216	82.39	79.81
Word Embedding with multi-dimensional attention	0.54m	261	86.22	83.12
600D Bi-LSTM encoders with multi-dimensional attention	2.88m	2080	90.39	84.53
Two self-attention with multi-dimensional attention	2.35m	592	90.18	84.66
Directional self-attention network (DiSAN)	2.35m	587	91.08	85.57

表 1: SNLI 不同方法的实验结果。 $|\theta|$: 参数个数 (不包括字嵌入部分)。T (s) /epoch: 每个历元的平均时间 (秒)。 训练 Accu (%) 和测试 Accu (%) : 训练和测试集的准确性。

Model	Test Accu
MV-RNN (Socher et al. 2013)	44.4
RNTN (Socher et al. 2013)	45.7
Bi-LSTM (Li et al. 2015)	49.8
Tree-LSTM (Tai, Socher, and Manning 2015)	51.0
CNN-non-static (Kim 2014)	48.0
CNN-Tensor (Lei, Barzilay, and Jaakkola 2015)	51.2
NCSL (Teng, Vo, and Zhang 2016)	51.1
LR-Bi-LSTM (Qian, Huang, and Zhu 2017)	50.6
DiSAN	51.72

表 2: 斯坦福情树 (SST) 细粒度情感分类的测试准确性。

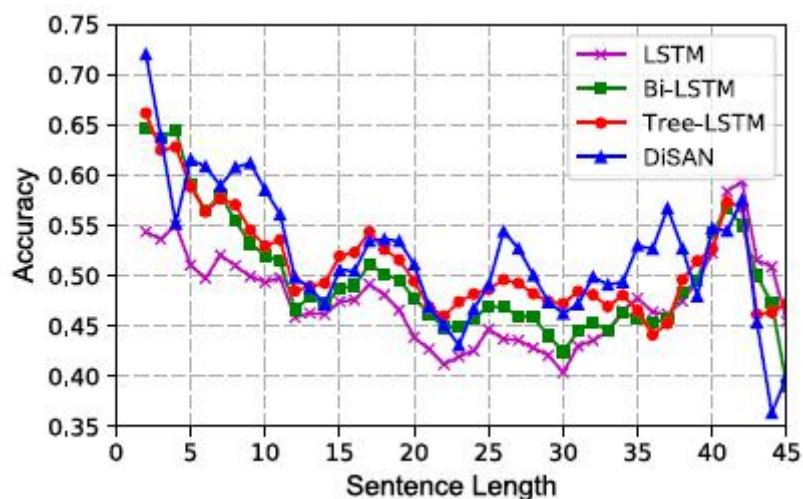


图 5: 对不同长度句子的细粒度情感分类准确率。LSTM, Bi-LSTM 和 Tree-LSTM 的结果来自 Tai, Socher 和 Manning (2015), DiSAN 的结果是五个以上的随机试验的平均值。

Pred. \ Label	Very Neg.	Neg.	Neu.	Pos.	Very Pos.
Very Neg.	106	153	1	18	1
Neg.	72	447	20	91	3
Neu.	13	185	45	140	6
Pos.	1	65	15	351	78
Very Pos.	0	15	3	178	203

表 3: 斯坦福情绪树库 (SST) 测试集上 DiSAN 预测的混淆矩阵。Pred., Neg., Neu. 和 Pos. 是预测, 否定, 中立和正面的缩写。

Method	Pearson's r	Spearman's ρ	MSE
Meaning Factory (Bjerva et al. 2014)	0.8268	0.7721	0.3224
ECNU (Zhao, Zhu, and Lan 2014)	0.8414	-	-
DT-RNN (Socher et al. 2014)	0.7923	0.7319	0.3822
SDT-RNN (Socher et al. 2014)	0.7900	0.7304	0.3848
LSTM (Tai, Socher, and Manning 2015)	0.8528	0.7911	0.2831
Bidirectional LSTM (Tai, Socher, and Manning 2015)	0.8567	0.7966	0.2736
Skip-Thought (Kiros et al. 2015)	0.8655	0.7995	0.2561
Constituency Tree-LSTM (Tai, Socher, and Manning 2015)	0.8582	0.7966	0.2734
Dependency Tree-LSTM (Tai, Socher, and Manning 2015)	0.8676	0.8083	0.2532
DiSAN	0.8704	0.8164	0.2861

表 4: 不同方法对 SICK 句子相关性的实验结果。

Method	Test Accu
cBoW (Zhao, Lu, and Poupart 2015)	87.3
RNN (Zhao, Lu, and Poupart 2015)	90.2
CNN (Kim 2014)	93.6
AdaSent (Zhao, Lu, and Poupart 2015)	92.4
Skip-Thought (Kiros et al. 2015)	92.2
DiSAN	94.4

表 5: TREC 问题类型分类数据集不同方法的实验结果

Method	Matched	Mismatched
cBoW	0.65200	0.64759
Bi-LSTM	0.67507	0.67248
DiSAN	0.70977	0.71402

表 6: MultiNLI 不同方法预测精度的实验结果。

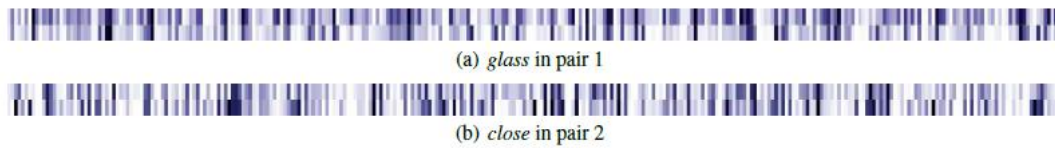


图 6: 在不同的句子语境中, 对同一词的两个注意概率比较。

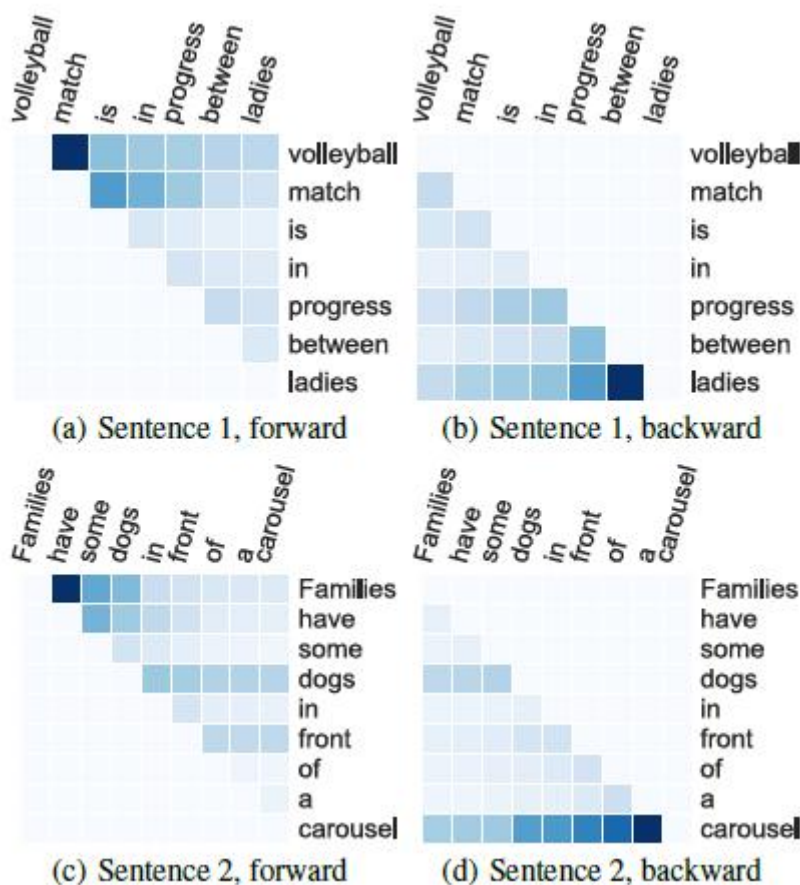


图 7: 前向/后向自我注意块的注意概率。

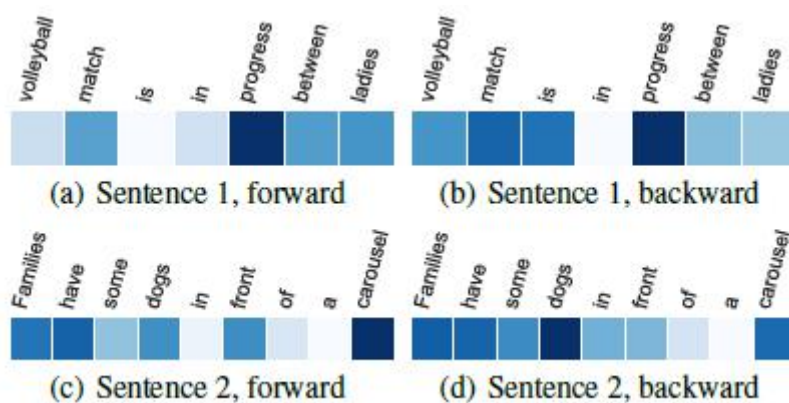


图 8: 融合门 F 在前进/后退方向自我注意块。

