DOI:10.3969/j.issn.1003-5060.2023.01.007

基于注意力机制的手语语序转换方法

张哲岩, 王青山
(合肥工业大学 数学学院,安徽 合肥 230601)

摘 要:文章考虑听障人士与健全人士在语法和语言结构上的差异,设计一种基于注意力机制的手语语序转换器,实现手语语序到书面表达的转换。语序转换器在编码阶段使用双向长短期记忆网络(long short-term memory,LSTM)提取手语语序特征,解码阶段使用一维卷积提取编码器隐藏状态的特征,并利用注意力机制避免了长距离的依赖问题,从而得到书面表达。实验结果表明,语序转换器准确率最高为92.64%。

关键词:注意力机制;语序转换;编解码模型;特征提取
中图分类号:TP183 **文献标志码**:A **文章编号**:1003-5060(2023)01-0042-06

A word order conversion method based attention mechanism for sign language

ZHANG Zheyan, WANG Qingshan
(School of Mathematics, Hefei University of Technology, Hefei 230601, China)

Abstract: Considering the differences in grammar and language structure between hearing-impaired and able-bodied people, this paper designs a sign language order converter based on attention mechanism to realize the conversion of sign language order to written expression. In the encoding stage, the word order converter uses bidirectional long short-term memory(LSTM) to extract the features of the sign language order, and in the decoding stage, it uses one-dimensional convolution to extract the features of the hidden state of the encoder. In addition, the attention mechanism is used to avoid the long-distance dependency problem, so as to obtain the written expression. The results show that the highest accuracy of the word order converter is 92.64%.

Key words: attention mechanism; word order conversion; encoder-decoder model; feature extraction

根据世界卫生组织最新的抽样调查[1],超过4.66亿人因听力受损而导致残疾。到2030年,预计将有近6.3亿人存在听力障碍问题,到2050年,这个数字可能会超过9亿。这些听障人士在日常生活中通常存在沟通障碍,导致其不能像健全人那样方便地学习、生活和就医等[2-4]。因此,准确理解手语对于听障人士与社会的正常交流是至关重要的。

文献[5]通过一系列的实验研究发现,听障学生在句子表征、局部连贯、整体连贯性上都不如健全学生。通过对听障人士手语和正常书面语比较[6],发现大部分听障人士的手语转化成书面语后虽然能够表达出关键词语,但句子语序混乱,健全人士无法准确理解,这种语序问题的原因是听障学生在进行书面语的表达时会遗失掉手语表达中的动作、身体姿态等一系列内容信息。对于手语作为第一语言的听障人士来说,学习第二语言书面语的难度较大[7]。因此,听障人士进行书面语表达时会使用便于他们自身的表达方式。促进听障人士与健全人士之间的交流已经成为全球关注的问题,目前已取得了一些成果[8-11]。但目前的工作都只考虑对手

收稿日期:2021-12-12;**修回日期**:2022-03-15
基金项目:国家自然科学基金资助项目(61571179)
作者简介:张哲岩(1997—),男,安徽蚌埠人,合肥工业大学硕士生;
王青山(1975—),男,安徽合肥人,博士,合肥工业大学教授,博士生导师,通信作者,E-mail:qswang@hfut.edu.cn.

语进行识别，没有考虑到听障人士和健全人士语序存在差异，导致很多听障人士的手语虽然被识别出来，但是健全人士难以准确地理解。因此本文提出一种基于注意力机制的手语语序转换方法，如图 1 所示，通过设计语序转换器来实现手语语序到书面表达的转换。如手语语序为“同学看我电影”，经过语序转换得到书面表达为“我和同学一起去看电影”。

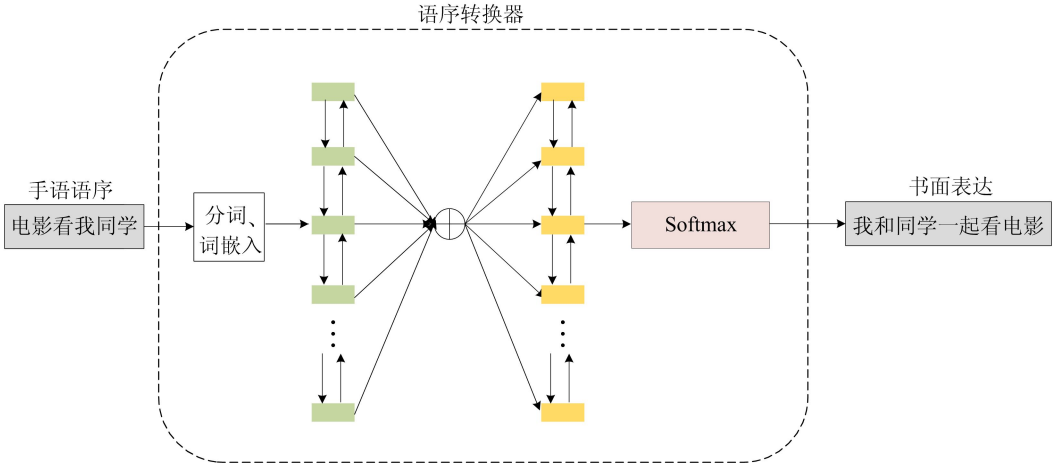

图 1　语序转换实例

目前，循环神经网络（recurrent neural network，RNN）模型是用于处理手语时序的主要方法。文献[12]使用带有注意力机制的双层 RNN 构建编解码器模型，使用类似神经机器翻译的框架，端到端地将连续手语视频翻译成口语化语句；文献[13]指出，随着输入序列长度的增加，基于 RNN 的编解码器性能迅速恶化，在反向传播时，过长的序列导致梯度计算异常，容易发生梯度消失或爆炸；为了解决上述问题，文献[14]提出使用注意力机制向解码器传递附加信息，通过构建上下文向量来减少信息损失，实现句子的翻译对齐。

本文根据手语语序特点，在编码阶段使用双向长短期记忆网络（long short-term memory，LSTM）分别从序列两端出发，更完整地提取手语语序的信息，解码时使用一维卷积提取编码器隐藏状态的特征，并利用注意力机制避免长距离的依赖问题，实现手语语序到书面表达的转换。

1　语序转换器

本节提出的语序转换器是一个基于注意力的编解码器模型，实现手语语序到书面表达语序的转换。编码阶段使用双向 LSTM 将分词后的句子以固定大小的向量形式投射到潜在空间中，在解码阶段加入注意力机制，确定关注输入序列的具体部分，减少将输入的所有信息编码成固定长度向量的负担，让解码器选择性地检索输入序列的隐藏状态。工作原理如图 2 所示。

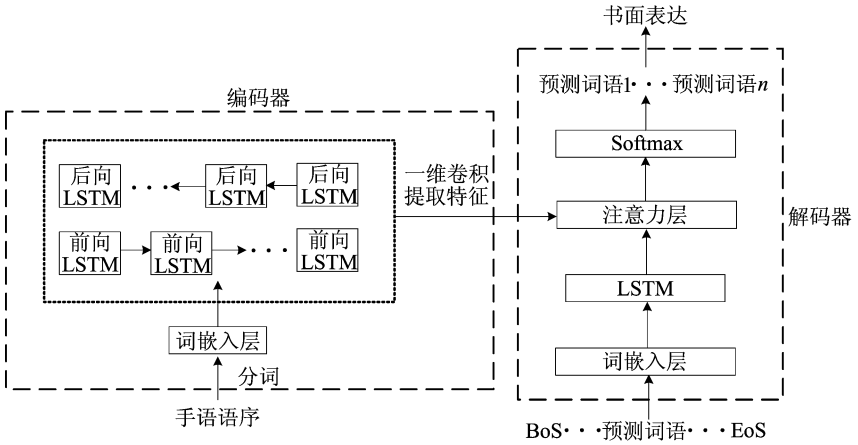

图 2　语序转换器工作原理

1.1 编码阶段

在编码阶段,由于连续手语信号中的词语边界难以确定,因此对手语识别器得到的手语语序利用隐马尔可夫模型(hidden Markov model, HMM)进行分词。在分词时,计算每个位置上词语的出现概率:

$$P(\text{Observed}[t],\text{Status}[t-1]) = p(\text{Status}[t-1]) \times p(\text{Observed}[t] \mid \text{Status}[t-1]) \tag{1}$$

其中:Status[$t-1$]为第 $t-1(1\leqslant t\leqslant n)$个位置的状态值;$n$ 为句子长度;Observed[t]为第 t 个位置上词语的观察值。选择概率值最大的词语放在第 t 个位置。接着,利用词嵌入[15]将单个词语投影到连续空间中,得到更密集的表示形式。在这种形式中,具有相似含义的词语更接近。对于句子中第 i 个词语 x_i $(i=1,2,\cdots,n)$,使用一个全连接层,将词语从独热向量转化到更密集空间上的线性投影 $\boldsymbol{x}_i^{\text{emb}}$,即

$$\boldsymbol{x}_i^{\text{emb}} = \text{WordEmbedding}(x_i) \tag{2}$$

所得到的 $\boldsymbol{x}_i^{\text{emb}}$ 为 x_i 对应的词向量。

因为语言习惯的不同,听障人士的语序中往往会包含主谓宾语残缺、语序错误、搭配不当、用词颠倒等问题,位置靠后的语句可能也会包含关键信息[16-17],所以生成词语需要根据上下文来做出判断。对于当前时刻,使用双向 LSTM[18]作为编码器,用 2 个 LSTM 分别训练前向和后向序列,使得输出序列包含每一个输入词语完整的上下文信息。对于 i 时刻的隐藏状态 $\boldsymbol{h}_i^*$,由给定序列的前向隐藏状态$\overrightarrow{\boldsymbol{h}_i^*}$ 和后向隐藏状态$\overleftarrow{\boldsymbol{h}_i^*}$ 连接而得到。

$$\begin{gathered}\boldsymbol{X} = [\boldsymbol{x}_1^{\text{emb}},\boldsymbol{x}_2^{\text{emb}},\cdots,\boldsymbol{x}_n^{\text{emb}}],\\ \boldsymbol{h}_i^* = [\overrightarrow{\boldsymbol{h}_i^*},\overleftarrow{\boldsymbol{h}_i^*}],\\ \overrightarrow{\boldsymbol{h}_i} = \text{LSTM}_{\text{ENC}}(\boldsymbol{x}_i^{\text{emb}},\overrightarrow{\boldsymbol{h}_{i-1}^*}),\\ \overleftarrow{\boldsymbol{h}_i} = \text{LSTM}_{\text{ENC}}(\boldsymbol{x}_i^{\text{emb}},\overleftarrow{\boldsymbol{h}_{i-1}^*})\end{gathered} \tag{3}$$

从(3)式可以看出,每个时刻的隐藏状态 $\boldsymbol{h}_i^*$ 均与前面的输入序列以及前一时刻的隐藏状态 $\boldsymbol{h}_{i-1}^*$ 有关。若句子过长,$\boldsymbol{h}$ 的维度有限,则可能无法表示序列的所有特征。在解码的初始阶段使用一维卷积提取编码器输出向量的特征,如图 3 所示。

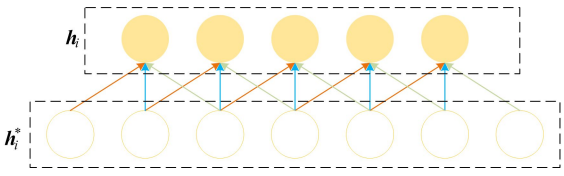

图 3 一维卷积提取特征

图 3 中,$\boldsymbol{h}_i$ 为提取特征后的隐藏状态。

1.2 解码阶段

在解码阶段,本文设计的解码器在文献[19]模型的基础上加入注意力机制,它克服了使用固定长度的向量表示整个序列信息所造成的长距离依赖和梯度消失问题,如图 2 右侧所示。解码器加入基于注意力机制可以确定关注输入序列的具体部分,从而减少将输入的所有信息编码成固定长度向量的负担,可以让解码器选择性地检索输入序列的隐藏状态。

首先,由编码器的隐藏状态计算上下文向量。考虑到编码器的隐藏状态 $\boldsymbol{h}_i(i=1,2,\cdots,n)$包含整个输入序列的信息,特别是输入序列的第 i 个词语附近的部分,可根据 $\boldsymbol{h}_i$ 来计算上下文向量 $\boldsymbol{c}_i$:

$$\boldsymbol{c}_i = \sum_{j=1}^{n} o_{ij}\boldsymbol{h}_j \tag{4}$$

其中,o_{ij} 为注意力权重,表示 i 时刻编码器输入与解码器输出的相关性。o_{ij} 的计算公式为:

$$o_{ij} = \frac{\exp(\text{score}(\boldsymbol{g}_i,\boldsymbol{h}_j))}{\sum_{k=1}^{n}\exp(\text{score}(\boldsymbol{g}_i,\boldsymbol{h}_k))} \tag{5}$$

其中,$\boldsymbol{g}_i$ 为解码器中 LSTM 层输出的隐藏状态。评分函数 score 取决于所使用的注意力机制,本文主要考虑 2 种评分函数,具体如下:

$$\text{score}(\boldsymbol{g}_i,\boldsymbol{h}_j) = \boldsymbol{g}_i\boldsymbol{W}\boldsymbol{h}_j \tag{6}$$

$$\text{score}(\boldsymbol{g}_i,\boldsymbol{h}_j) = \boldsymbol{V}^{\text{T}}\tanh(W[\boldsymbol{g}_i,\boldsymbol{h}_j]) \tag{7}$$

其中:(6)式为文献[20]提出的乘法函数;(7)式为文献[14]提出的基于连接的函数;$\boldsymbol{W}$ 和 $\boldsymbol{V}$ 为待学习的参数。

然后,将上下文向量 $\boldsymbol{c}_i$ 与隐藏状态 $\boldsymbol{g}_i$ 进行连接,得到注意力向量 $\boldsymbol{a}_i$:

$$\boldsymbol{a}_i = \tanh(W_c[\boldsymbol{c}_i,\boldsymbol{g}_i]) \tag{8}$$

最后,使用前一个时刻的隐藏状态 $\boldsymbol{g}_{i-1}$ 和词向量 $\boldsymbol{y}_{i-1}^{\text{emb}}$作为输入,通过解码生成序列中下一个时刻的词语 $\boldsymbol{y}_i$ 并更新其隐藏状态 $\boldsymbol{g}_i$:

$$(\boldsymbol{y}_i,\boldsymbol{g}_i) = \text{LSTM}_{\text{DEC}}(\boldsymbol{y}_{i-1}^{\text{emb}},\boldsymbol{g}_{i-1},\boldsymbol{a}_{i-1}) \tag{9}$$

通过对每个词语的交叉熵来计算损失,这些损失通过网络向后传播到 LSTM 和词嵌入层,从而更新网络参数。解码器逐词生成句子:

$$p(\boldsymbol{y} \mid \boldsymbol{x}) = \prod_{i=1}^{n} p(\boldsymbol{y}_i \mid \boldsymbol{y}_{i-1},\boldsymbol{g}_i) \tag{10}$$

其中:$\boldsymbol{y}_0$ 为句子的开始标记<BoS>(begin of sentence);$\boldsymbol{g}_0$ 为<BoS>的隐藏状态。预测过程持续到表示句子结束的标记<EoS>(end of sentence)为止。

2 实　验

本文从机器翻译大赛中所用的专业数据集 WMT 中挑选 5 000 条中文句子,共 5 730 个词,来训练语序转换器,训练集和测试集的划分为 9∶1。用 BLEU-1、BLEU-2、BLEU-3、BLEU-4 的分值以及准确率评估语序转换器的性能,在不同层次上对翻译结果进行打分。实验从递归网络、注意力机制的选择、批次大小、波束宽度 4 个方面验证不同模块的选择对语序转换性能的影响,后续也给出语序转换的实验结果。

2.1 递归单元

本文通过将双向 LSTM、LSTM 和文献[21]中单一 RNN 进行比较,验证对于递归单元的选择。语序转换器使用文献[20]中提出的注意力机制,批次大小为 128,波速宽度设置为 1,实验结果见表 1 所列。

从表 1 可以看出,使用双向 LSTM 时语序转换器的表现最佳,使用 RNN 效果最差。原因是 LSTM 存在门结构,反向传播有较多的路径选择,而且只用了对应元素相乘和相加使其更为稳定,因此 LSTM 发生梯度消失的概率远远小于 RNN。

双向 LSTM 则能更全面地获取到输入序列的信息。因此,在后面的实验中递归单元均选择使用双向 LSTM。

表 1　递归单元的选择

递归单元	BLEU-1 分值	BLEU-2 分值	BLEU-3 分值	BLEU-4 分值	准确率/%
RNN	40.25	19.82	19.34	6.29	82.76
LSTM	41.54	20.90	20.66	16.40	86.30
双向 LSTM	43.51	30.29	20.95	18.58	89.00

2.2 注意力机制

本文讨论不同的注意力机制对语序转换器的影响,使用文献[14]和文献[20]提出的注意力机制进行比较,还使用了一个不含注意力机制的模型作为空白对照。语序转换器使用双向 LSTM、批次大小为 128,波速宽度设置为 1,实验结果见表 2 所列。

从表 2 可以看出,加入文献[20]提出的注意力机制后的语序转换器性能最好,这是由于在生成词语时使用了解码器的隐藏状态,可以提取更多信息。因此,在后面的实验中均使用文献[20]提出的注意力机制。

表 2　注意力机制的选择

注意力机制	BLEU-1 分值	BLEU-2 分值	BLEU-3 分值	BLEU-4 分值	准确率/%
无注意力机制	40.55	27.19	20.28	16.29	83.04
文献[14]	42.39	29.71	22.43	17.99	85.43
文献[20]	43.51	30.29	20.95	18.58	89.00

2.3 批次大小

本节讨论不同的批次大小对语序转换性能的影响,实验中均使用双向 LSTM、波速宽度设置为 1 以及文献[20]提出的注意力机制。虽然更大批次的训练能够使得梯度更加平滑,但是在收敛速度上有一定程度的降低,且容易陷入局部极小值。小批量训练能够使得网络更有效地表示数据使其更接近目标分布,实验结果见表 3 所列。

表 3　批次大小的选择

批次大小	BLEU-1 分值	BLEU-2 分值	BLEU-3 分值	BLEU-4 分值	准确率/%
128	43.11	29.89	22.55	17.18	89.00
64	42.29	29.96	22.76	17.64	90.10
32	44.07	30.84	23.48	18.07	91.14
1	43.80	31.23	24.01	18.56	92.05

从表 3 可以看出,批次大小为 1 时模型性能最好,但整体训练的复杂度更高,这是梯度差异过大造成的。在后续实验中,将批次大小设为 32 进行随机梯度下降训练。

2.4 波速宽度

使用波束搜索(beam search)作为解码预测词语方法,相对贪心搜索扩大了搜索空间,但远不及穷举搜索指数级的搜索空间,是两者的一个折中方案。然而更大的波束宽度并不能带来更好的性能,反而会增加解码的时间和内存。本节通过实验验证不同波束宽度对语序转换器性能的影响。语序转换器使用双向 LSTM 以及文献[20]中提出的注意力机制,批次大小为 32,使用实验准确率和 BLEU-4 作为评估指标,结果见表 4 所列。从表 4 可以看出,波束宽度为 2 时在验证集上得到了最好的性能。

表 4 波速宽度的选择

波速宽度	BLEU-4 分值	准确率/%
4	16.95	89.36
3	17.68	91.10
2	18.07	92.64
1	17.43	91.14

2.5 语序转换结果

本节展示语序转换器的具体结果,使用基于 Theano 深度学习框架实现的 RNNSearch 模型作为基准系统,实验结果见表 5 所列。

表 5 语序转换结果

手语语序	基准系统	语序转换器	书面表达
商品,问题,退换	商品有退换问题	商品如果有问题可以来退换	商品如果有问题,可以退换
同学,上课,黑板	上课看黑板	上课的时候同学们看黑板	上课时,同学们请注意看黑板
房间,空调,热水	房间有空调热水	房间内有空调为你们提供热水	房间内有空调,24 h 提供热水

从表 5 可以看出,第 1 句使用基准系统出现了语序颠倒的情况,应该是“有问题之后再进行退换”;第 2 句缺少主语“同学”;而在第 3 句中则出现了“有空调热水”这一缺少谓语的情况。而在使用语序转换器时能有效地避免这些问题,但和完整的书面表达仍有差距,这可能是由于输入的手语语序与书面表达之间的信息缺失所导致的。

3 结 论

本文研究了手语语序到书面表达转换的问题,提出了一种基于注意力机制的手语语序转换方法。该方法将手语语序分词放入基于注意力机制的语序转换器中,从而实现了从手语语序到书面表达的转换。最后,本文通过实验验证了该方法的性能,结果表明,语序转换器的准确率最高能达到 92.64%。

[参 考 文 献]

[1] BACALA T. Hearingchat for world hearing day 2018: recap [EB/OL]. [2022-12-14]. https://journals.lww.com/thehearingjournal/blog/OnlineFirst/pages/post.aspx? PostID=26.

[2] HUMPHRIES T, KUSHALNAGAR P, MATHUR G, et al. Avoiding linguistic neglect of deaf children[J]. Social Service Review, 2016, 90(4): 589-619.

[3] WILCOX P. My mother made me deaf: discourse and identity in a deaf[J]. Journal of Anthropological Research, 2019, 75(3): 425-426.

[4] DAVIS G. Made to hear: cochlear implants and raising deaf children[J]. American Journal of Sociology, 2017, 32(3): 436-437.

[5] 任媛媛. 聋人学生汉语书面语语法研究综述[J]. 中国特殊教育, 2011(3): 16-19.

[6] 吴铃. 九年制聋校毕业生书面语言能力发展研究[J]. 中国特殊教育, 2006(8): 29-34.

[7] 贺微. 浅谈如何提高聋生的书面语能力[J]. 课程教育研究, 2020(19): 89-90.

[8] MING J C, OMAR Z, JAWARD M H. A review of hand gesture and sign language recognition techniques[J]. International Journal of Machine Learning and Cybernetics, 2019, 10(1): 131-153.

[9] ANIBLE B. Iconicity in American sign language English translation recognition[J]. Language and Cognition, 2020, 12(1): 138-163.

[10] SUNEETHA M, PRASAD M, KISHORE P. Multi-view motion modelled deep attention networks (M2DA-Net) for video based sign language recognition[J]. Journal of Visual Communication and Image Representation, 2021, 78: 103-161.

[11] SHARMA S, KUMAR K. ASL-3DCNN: American sign language recognition technique using 3-D convolutional neural networks[J]. Multimedia Tools and Applications, 2021, 80(17): 26319-26331.

[12] KYUNGHYUN C, BART M, DZMITRY B, et al. On the properties of neural machine translation: encoder-decoder approaches[C]//Proceedings of Computation and Language. [S. l.: s. n.], 2014: 103-111.

(下转第 59 页)

4 结　论

关键场景中的 AI 应用对 DLA 的可靠性要求很高。基于 RR、CR、DR 等常规阵列的冗余方法虽然能简单地修复一些故障，但是容易受到故障分配不均的影响，即使冗余资源充足往往也无法修复所有故障。针对这一问题，本文提出了重计算结构(RCA)，与传统即时的故障修复策略不同，它有一组冗余 PE 组成的重计算单元(RCU)，能够从阵列边缘获得流经故障 PE 的数据，利用冗余 PE 重新进行故障 PE 的计算，并在存储单元中替换错误的计算结果。当 2D 计算阵列中故障 PE 的数量不超过 RCU 大小时，RCA 可以完全修复 2D 计算阵列。即使 PE 错误率进一步增加，RCA 可以通过增加 RCU 的冗余规模，进一步提高容错能力，也可以选择修复出最大的计算阵列区域，减少性能损失。实验表明，RCA 具有较高的可扩展性和较小的芯片面积开销，并在故障修复能力上大大优于之前的冗余方法。

[参　考　文　献]

[1] TAKANAMI I, HORITA T. A built-in circuit for self-repairing mesh-connected processor arrays by direct spare replacement[C]//2012 IEEE 18th Pacific Rim International Symposium on Dependable Computing. [S. l.]: IEEE, 2012: 96-104.

[2] TAKANAMI I, FUKUSHI M. A built-in circuit for self-repairing mesh-connected processor arrays with spares on diagonal[C]//2017 IEEE 22nd Pacific Rim International Symposium on Dependable Computing (PRDC). [S. l.]: IEEE, 2017: 110-117.

[3] XU D, CHU C, WANG Q, et al. A hybrid computing architecture for fault-tolerant deep learning accelerators[C]//2020 IEEE 38th International Conference on Computer Design (ICCD). [S. l.]: IEEE, 2020: 478-485.

[4] XU D, WANG Q, LIU C, et al. HyCA: a hybrid computing architecture for fault tolerant deep learning[J]. IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, 2022, 41(10): 3400-3413.

[5] ZHANG J J, BASU K, GARG S. Fault-tolerant systolic array based accelerators for deep neural network execution [J]. IEEE Design & Test, 2019, 36(5): 44-53.

[6] LI L, XU D, XING K, et al. Squeezing the last MHz for cnn acceleration on FPGAs[C]//2019 IEEE International Test Conference in Asia (ITC-Asia). [S. l.]: IEEE, 2019: 151-156.

[7] MEYER F J, PRADHAN D K. Modeling defect spatial distribution[J]. IEEE Transactions on Computers, 1989, 38(4): 538-546.

[8] ZHANG J J, GU T, BASU K, et al. Analyzing and mitigating the impact of permanent faults on a systolic array based neural network accelerator[C]//2018 IEEE 36th VLSI Test Symposium (VTS). [S. l.]: IEEE, 2018: 1-6.

(责任编辑　张　锴)

(上接第 46 页)

[13] CHO K, MERRIENBOER B V, GULCEHRE C, et al. Learning phrase representations using RNN encoder-decoder for statistical machine translation [EB/OL]. [2022-12-14]. https://arxiv.org/abs/1406.1078.

[14] DZMITRY B, KYUNGHYUN C, YOSHUA B. Neural machine translation by jointly learning to align and translate[C]//Proceedings of Conference on Learning Representations. [S. l. : s. n.], 2015: 1409-1423.

[15] TOMAS M, KAI C, GREG C, et al. Efficient estimation of word representations in vector space [EB/OL]. [2022-12-14]. https://arxiv.org/pdf/1301.3781.pdf.

[16] 郭学慧，王志强. 聋校听觉障碍学生书面语语法偏误研究综述[J]. 绥化学院学报，2020，40(1)：35-38.

[17] 张帆. 国内近年来聋人学生汉语书面语句法研究述评[J]. 长春大学学报(社会科学版)，2015，25(6)：133-136.

[18] HOCHREITER S, SCHMIDHUBER J. Long short-term memory[J]. Neural Computation, 1997, 9(8): 1735-1780.

[19] SUTSKEVER I, VINYALS O, LE O V. Sequence to sequence learning with neural networks[C]//Proceedings of Conference and Workshop on Neural Information Processing Systems. [S. l. : s. n.], 2014: 3104-3112.

[20] LUONG M, HIEU P, CHRISTOPHER D M. Effective approaches to attention-based neural machine translation [C]//Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing. [S. l. : s. n.], 2015: 1412-1421.

[21] KALCHBRENNER N, BLUNSOM P. Recurrent continuous translation models[C]//Proceedings of 2013 Conference on Empirical Methods in Natural Language Processing. [S. l. : s. n.], 2013: 1700-1709.

(责任编辑　李　凯)