

# 基于移动端的神经网络汉英翻译模型

成 洁

(陕西国际商贸学院 基础课部, 西安 712046)

**摘要:** 为了使中英文翻译更加智能以及更加合理, 采用句子级的释义对译文进行改写, 将其视为同一语言之间的翻译任务; 在没有大规模平行释义语料库的情况下, 利用机器翻译结果和源语言的参考翻译来近似平行释义语料库; 然后, 利用该模型训练一个从机器翻译结果到参考翻译的重复系统, 生成语义一致的句子级重复结果; 在此基础上, 将重述结果引入系统整合的翻译假设中; 最后, 在翻译和释义的基础上, 进行了面向移动应用的设计和开发, 实现了中英文机器翻译; 通过实验发现, 该方法相对于经典的基线系统提高了 1.02—1.71 BLEU 分数。

**关键词:** 机器学习; 机器翻译; RNN; 神经网络; 释义; 语料库

## Chinese—English Translation Model of Neural Network Based on Mobile Phone

Cheng Jie

(Shaanxi Institute of International Trade and Commerce, Xi'an 712046, China)

**Abstract:** In order to make Chinese and English translation more intelligent and more reasonable, this article uses sentence-level interpretation to rewrite the translation as a translation task between the same language. In the absence of a large-scale parallel paraphrasing corpus, the machine translation results and the reference translation of the source language are used to approximate the parallel paraphrasing corpus. Then, the model is used to train a repetitive system from machine translation results to reference translations to generate semantically consistent sentence-level repetition results. Based on this, the results of restatement are introduced into the translation assumption of system integration. Finally, on the basis of translation and interpretation, the mobile application-oriented design and development were carried out to realize Chinese and English machine translation. It was found through experiments that the proposed method improved 1.02—1.71 BLEU scores compared to the classic baseline system.

**Keywords:** machine learning; machine translation; RNN; neural network; paraphrase; corpus

## 0 引言

深度学习在自然语言处理的各个领域都得到了成功的应用, 基于强化学习的神经网络机器翻译的研究也得到了迅速的发展<sup>[1-4]</sup>。近年来, 基于神经网络的机器翻译与统计机器翻译相比已经取得了很大的成就, 因此研究者不断地、逐步地改变方法, 以利用神经网络来改进机器翻译<sup>[5-8]</sup>。尽管机器翻译有很多方法和技术, 但也存在一个问题。不同的机器翻译方法或技术不能完整、准确地描述同一源语句子的翻译信息, 而源语句子中的表达必然会产生错误<sup>[9-12]</sup>。复述可以纠正和补充机器翻译结果, 以尽可能地弥补上述问题的错误, 从而可以用源语言更全面地描述翻译结果。针对以上问题, 系统融合也提出了解决方案。对于不同的系统, 系统融合将不同机器翻译方法的输出结果融合在一

起, 以生成输出结果, 由于其改进, 该结果优于原始机器翻译。

本文运用目前的主要机器翻译方法来建立释义模型, 以完成提高汉英机器翻译的任务。然后, 它使用系统融合技术融合翻译, 以进一步提高翻译质量。然后, 在翻译和释义的基础上, 利用网络和移动设备设计和开发汉英翻译模型, 并在神经机器翻译和释义汉英移动翻译系统的基础上初步完成了开发工作。

## 1 神经机器翻译方法

### 1.1 基于递归神经网络的翻译模型

递归神经网络 (RNN) 是神经网络机器翻译中最为广泛的结构, 而其它方法在编码时只改变 RNN 的内部隐藏结构或固定向量计算<sup>[13-15]</sup>。与一般的前馈神经网络 FNN 不同, RNN 在其基础上引入了循环机制<sup>[16-18]</sup>。这样, RNN 可以处理前后不相关的输入, 但它不像以前的神经网络那样需要一次性完成输入特性。

图 1 描述了 RNN 的展开架构。 $x$  和  $o$  分别表示输入和输出;  $h$  表示隐藏层;  $t$  表示时间;  $U$  表示参数从输入层到隐藏层;  $V$  表示参数从隐藏层到输出层;  $W$  表示隐藏层之

收稿日期: 2020-03-03; 修回日期: 2020-03-27。

基金项目: 2019 陕西省教育厅专项科研项目 (19JK0090); 2019 陕西省社会科学基金项目 (2019M032)。

作者简介: 成 洁 (1986-), 女, 甘肃兰州人, 硕士研究生, 讲师, 主要从事机器翻译、英语教育方向的研究。

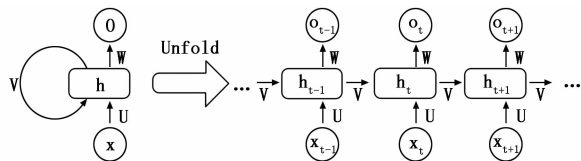


图 1 RNN 模型架构

间的递归参数。对于某一时刻的  $t$ ，其输出应考虑当前输入  $x_t$  和由某一状态  $S_{t-1}$  发送的值。

对于  $x_t$ ，它是在整个网络的时间  $t$  处输入的字向量。对于  $s_t$ ，它是时间  $t$  的状态，也是网络的存储单元。它包含时间  $t$  之前的所有状态，计算如下：

$$s_t = f(Ux_t + Ws_{t-1}) \tag{1}$$

在公式 (1) 中， $f$  是非线性激活函数。

对于  $o_t$  它是在时间  $t$  时的输出，通过公式 (2) 可以得到计算结果。

$$o_t = \text{softmax}(Vs_t) \tag{2}$$

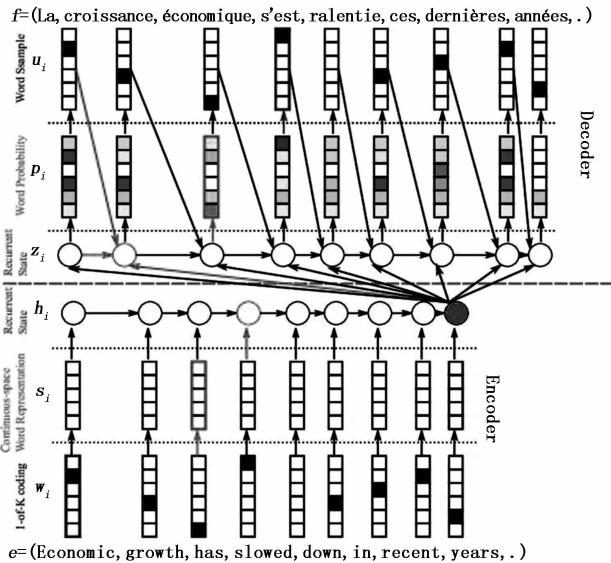


图 2 基于编码解码器的神经机器翻译系统

### 1.2 编解码模型

基于神经网络的机器翻译系统的典型模型是编码—解码。其基本思想是：将给定的源语言句子映射成连续的密集向量。解码是指它根据这个向量转换成目标语言句子。一个编码—解码：基于结构的神经机器翻译系统 (NMT) 如图 2 所示。编码和解码通常使用 RNN 实现。由于普通 RNN 具有梯度消失和梯度爆炸的特性<sup>[19-21]</sup>，因此通常采用长短期记忆网络 (LSTM)。

#### 1.2.1 LSTM 的原理

长短期记忆网络 (LSTM) 是递归神经网络种的一种特殊网络，这种网络可以对长期依赖关系进行学习和获取。Hochreiter 等人提出了该网络，并在接下来的数十年里，许多学者对该网络模型进行了完善和改进。LSTM 及相关变

种网络的被设计出来的目的是为了 避免长期依赖及 相关的问题。对于 LSTM 来说，记住长期信息是其本身的自发默认行为，是不需要通过大量学习就能够实现的。

所有普通的循环神经网络种，都有重复模块，这些模块以链式存在。在标准的 RNN 中，重复模块非常简单，例如只有一个 tanh 层，如图 3 所示。

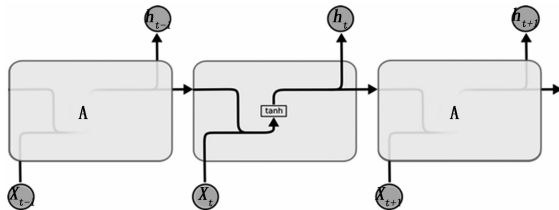


图 3 标准 RNN 的重复模块包含一个单层网络

基于以上背景可知，在 LSTM 及相关变种网络中，同样存在这样的链状结构，但是与普通 RNN 中的区别在于，LSTM 中的重复模块的结构是不一样的，具体如图 4 所示。在 LSTM 中，原来的单层网络被一个四层网络所代替，并且他们之间的交互方式也十分特别。

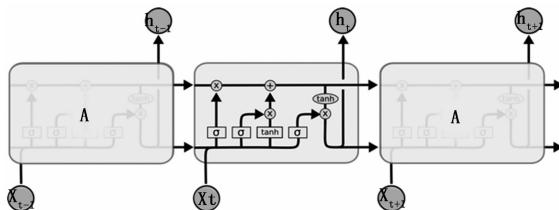


图 4 LSTM 网络的重复模块包含一个四层的交互网络

cell 状态是 LSTM 及相关变种网络的关键，图 5 表示了一个横穿整个 cell 顶端的水平线。

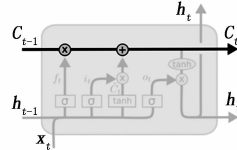


图 5 水平线横穿 cell 示意图

在长短期记忆网络中，cell 状态类似于传送带。为了保证承载的信息在传送过程中保持不变，以及方便传输，cell 状态是直接穿过链，并且只在少数地方进行一些较小的线性交互。

在 LSTM 中，为了能够改变 cell 状态，如对信息的添加或删除，所以引入一种叫做“门”的结构来对其进行操作。在这里，一个 sigmoid 神经网络层和一个位乘操作组成了门，如图 6 所示。门在 LSTM 中的作用是控制信息，有选择的让信息通过。

sigmoid 层输出数值是 0 或者 1，这两个值代表对应的信息是否允许通过。当数值为 0 时，当前不允许信息通过，

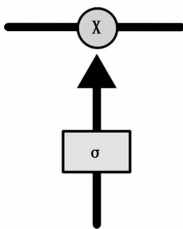


图 6 门结构示意图

而当数值为 1 时，则表示当前所有信息都能够通过。一个 LSTM 中门有 3 个，从而实现对 cell 状态的保护和控制。下面将具体介绍这 3 个门。

第一个门：首先在 LSTM 中需要决定哪些信息要从 cell 状态中被抛弃。这个选择是由一个被称为“遗忘门”的 sigmoid 层控制。这个门的输入是  $h_{t-1}$  和  $x_t$ ，输出为 0 或者 1。1 代表保留这个信息，0 代表这个信息需要被抛弃。

第二个门：其次 cell 状态中将要存储哪些新信息也需要被判定。这步可以细分为两步。首先，更新哪些信息是由一个叫做“输入门”的 sigmoid 层决定的。然后，一个新的候选向量  $\tilde{C}_t$  会由 tanh 层生成，这个值可能被加入到 cell 状态中。最后，这两个值将结合起来用于更新 cell 状态。

第三个门：接着，在 LSTM 中，是需要对最终输出做出决定。当前的 cell 状态会影响输出结果，但是为了使结果更加准确，将会对结果进行过滤。首先，建立一个输出门，这个门是基于 sigmoid 神经网络层，这个门的作用是用来决定哪一部分的 cell 状态是可以输出。然后，将 cell 状态通过 tanh 函数后，乘以输出门。最终就输出了只允许被输出的部分。

### 1.3 基于神经网络的释义模型

在此对机器翻译中释义技术的改进主要来自两个方面：

- 1) 扩展语料库信息。释义技术弥补了训练集、开发集和测试集中句子信息的缺陷。
- 2) 改变句子表达。复述技术对机器翻译的输入句子进行复述，使翻译系统能够更好地理解或复述输出的翻译文本，使翻译句子更符合目的语的语言习惯。

通过不同的释义语料库和不同的神经网络建立不同的释义模型，并通过第二种方法对机器翻译结果进行复述，以提高机器翻译的结果质量。释义模型的生成需要几个步骤来完成：

- 1) 得到神经网络翻译模型；
- 2) 得到释义语料库；
- 3) 训练并得到释义模型。

## 2 基于释义的机器翻译融合模型

### 2.1 RNN 编解码器的改进分析

在改进方法的基础上，引入了译码器中的对齐模型，使该模型能够同时学习对齐和平移。与传统的 RNN 编解码

器最大的区别是改进后的模型不将整个输入序列变换成固定长度的中间矢量表达式，而是通过应用双边递归神经网络将输入序列变换成具有相同输入长度的隐单元表达式。在此基础上，解码器可以在解码过程中选择一个或多个隐藏表达式进行对齐和翻译。该模型在处理长句子时具有较好的处理能力。

#### 2.1.1 编码器：双向 RNN 表达式输入序列

对于给定的输入序列  $x = (x_1, \dots, x_T)$ ，传统的递归神经网络模型将根据该序列从第一个序列到最后一个序列的顺序依次输入该模型，最终将整个序列转化为一个中间向量表达式。为了使每个字的隐藏层都能在这个序列中携带其前、后信息，该模型采用双向 RNN 模型来代替传统的 RNN 模型。

双向递归神经网络模型由前向和后向递归神经网络组成。前向递归神经网络  $f_{\rightarrow}$  将根据输入序列  $\vec{x}_1, \dots, \vec{x}_T$  的前向序列读取，并生成改进序列的前向隐藏层状态向量；相反，逆循环网络  $f_{\leftarrow}$  将输入序列按逆顺序读取，生成逆隐藏状态向量  $\overleftarrow{h}_1, \dots, \overleftarrow{h}_T$ 。

这样，隐藏的分层状态表达式就包含了整个输入序列的信息。RNN 模型的隐层表达趋向于更大程度上的表达和该词附近的信息。因此，隐藏的分层向量  $h_j$  对于每个词的信息主要集中在该词附近的上下文信息上，每个词都可以以给定词的上下文信息为中心来表达。在此基础上，解码器引入对齐模型，对输出序列进行更合理、更准确的解码

#### 2.1.2 解码器：训练对齐模型

在该模型中，重新定义了以下方程：

$$p(y_i | y_1, y_2, \dots, y_{i-1}, x) = g(y_{i-1}, s_i, c) \quad (3)$$

其中： $s_i$  是时间  $i$  时的隐藏层状态，并且隐藏层状态的计算可以通过式 (4) 得到：

$$s_i = f(s_{i-1}, y_{i-1}, c_i) \quad (4)$$

其中： $c_i$  可以由输入序列的隐藏状态向量表示。这种隐层状态向量的实现依赖于上下文中双向递归神经网络模型表达式的输入序列。这种隐层矢量计算是通过每个输入序列的隐层状态进行矢量加权来实现的。

$$c_i = \sum_{j=1}^{T_x} a_{ij} h_j \quad (5)$$

每个向量  $h_j$  的权重  $a_{ij}$  可以通过式 (6) 计算得到：

$$a_{ij} = \frac{\exp(e_{ij})}{\sum_{k=1}^{T_x} \exp(e_{ik})} \quad (6)$$

其中： $e_{ij}$  是一个对准模型，用于估计对准概率，具体可以通过式 (7) 计算得出：

$$e_{ij} = a(s_{i-1}, h_j) \quad (7)$$

基于隐层状态向量  $s_{i-1}$  和  $j_m$  双向递归神经网络，获得了这种对准概率。

## 2.2 对准模型

通过上下文中可以看出, 在  $T_x \times T_y$  的每一个时间段中, 双语句子都将与输入模型保持一致。在模型训练和解码过程中, 需要进行大量的运算。因此, 为了降低计算复杂度, 所以将使用感知器来完成这样的功能。具体如式 (8) 所示:

$$a(s_{i-1}, h_j) = \mathbf{v}_a^T \tan(\mathbf{w}_a s_{i-1} + \mathbf{U}_a h_j) \quad (8)$$

其中:  $\mathbf{w}_a$ ,  $\mathbf{U}_a$  和  $\mathbf{v}_a^T$  是模型的权重矩阵。由于  $\mathbf{U}_a h_j$  的计算不依赖于  $i$ , 因此它的值可以预先计算, 以减少模型训练中的计算量。

## 2.3 释义生成

在复述任务中, 把复述语言的输入和输出作为不同的语义表达。与翻译任务不同, 释义任务的输入和输出属于同一种语言。基于这一思想, 所以利用该模型对同一语言的复述系统进行训练, 并以机器翻译结果为输入, 生成句子规模的复述结果, 使每个句子语义一致。由于在同一语言中很难获得大规模的平行释义语料库, 因此采用机器翻译文本和参考翻译文本的平行语料库来近似释义语料库。在系统融合任务中, 更多希望得到的是机器翻译结果的复述句, 因此与机器翻译文本和参考文本相比, 并行复述语料库在一定程度上具有更大的优势。同时, 在一定程度上提高了机器翻译质量的效果。

当得到释义语料库时, 释义模型的生成相对简单。释义模型的生成过程与翻译模型的训练过程相似。在已有 RNN 网络和 CNN 网络的基础上, 利用释义语料库分别训练模型, 从机器翻译文本和参考翻译文本中获得释义模型。模型实现后, 将机器翻译结果输入到“复述系统”中进行翻译, 从而为输入的句子生成理想的句子级复述结果。

## 3 基于移动系统的模型应用与测试

### 3.1 系统设计与翻译过程

本文的软件设计框架采用 MVC (model view controller) 软件构建模式<sup>[22]</sup>。MVC 三层架构通过业务逻辑、数据和界面显示的分离, 促进系统组织代码, 减少耦合, 提高可重用性。系统生命周期成本低, 可维护性高。内容提供者主要用于管理共享数据, 实现不同应用程序之间的数据共享。

通过内容提供者提供的完整性机制, 可以实现一个程序对另一个程序的数据访问。如果内容提供者允许, 它甚至可以更正数据。目前, Android 提供的标准跨程序数据共享正在使用内容提供者。移动翻译软件系统的整体流程图如图 7 所示, 在该系统中, 移动终端视图层用来告诉用户输入和收集源语言语句。控制器用于将句子传输到服务器。

当系统接收到需要翻译的源句后, 系统将该句子输入到模型中, 通过控制器, 将源句传输到远端服务器, 然后

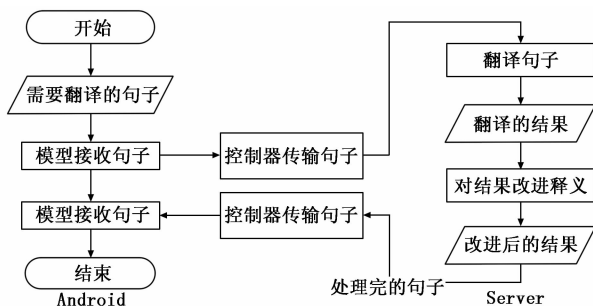


图 7 系统工作流程

服务器通过利用前文所提的基于递归神经网络的翻译模型对句子进行翻译, 然后根据基于释义的机器翻译融合模型对释义进行优化, 增加译后句子的可读性。然后远端服务器将处理后的句子再次通过控制器传输回移动端, 最后在移动端展示出结果。

### 3.2 系统展示

该软件的主要功能是将英语翻译成汉语。当软件打开时, 图 8 中的初始界面将显示在手机上。可以看到在打开软件后, 有一个友好的界面建议。

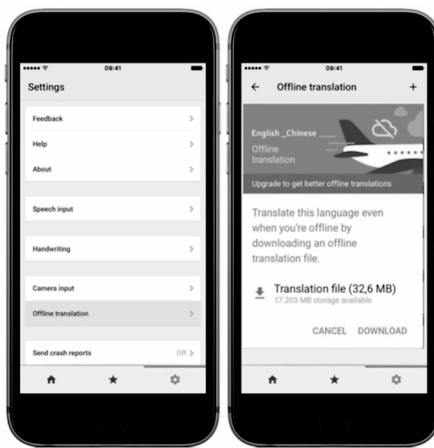


图 8 软件初始界面

输入要翻译的句子时, 单击相应的按钮翻译该单词或句子。点击翻译后, 输入的英文将被传送到服务器进行翻译。翻译和释义后, 翻译结果将显示在图 9 的界面中, 供用户检查。

### 3.3 性能评估

为了更好地评估所提出的翻译模型的性能, 本实验基于 NIST 进行英文翻译任务, 使用的语料测试集为 NIST05, NIST06, NIST08。开发集才用的是 NIST02<sup>[8]</sup>。

接下来通过 BLEU 算法评价性能<sup>[23-24]</sup>。BLEU 算法的计算公式为:

$$BLEU = \frac{\text{出现在 } reference \text{ 中的 } candidate \text{ 的单词的个数}}{\text{candidate 中单词的总数}} \quad (9)$$



图 9 翻译结果示意图

表 1 列出了在基线系统中的方法和转换的结果。

表 1 翻译结果的对比

	NIST05	NIST06	NIST08	ALL
Baseline	35.98	33.88	30.07	35.12
Proposed	37.33	35.12	31.02	36.83

表 1 给列出了相应的实验结果，ALL 表示将所有的测试集放在一起，从表中结果可以发现，对于 NIST08 测试集，基线系统的 BLEU 值为 30.07，是所有结果中最低的。相应的，提出的模型，在 NIST08 测试集上的 BLEU 值有所提高，为 31.02，但同样为所有测试集中的最低的，但结果提高了 1.02。在 3 个测试集中，NIST05 的结果最好，分别为 35.98 和 37.33。从表中可知，使用释义纠正短语概率的方法在一定程度上得到了改进。对于所有的测试集结果，基线系统的 BLEU 值为 35.12，提出的模型的 BLEU 值为 36.83，提高了 1.71BLEU。结果表明，修正后的短语翻译概率在一定程度上解决了训练语料数据的稀疏性，修正后的短语翻译概率更加准确合理。然而，通过短语向量提取的短语复述过程中存在一定的噪声，因此机器翻译中的振幅改善并不大，不需要进一步改进。

#### 4 结束语

本文提出了一种面向移动终端的释义机翻译系统融合。该方法利用 RNN 编解码模型生成语义一致性的句子级释义。我们把释义作为同一种语言之间的翻译任务。在没有大规模并行释义语料库的情况下，利用机器翻译结果和源语言的参考翻译文本来逼近并行释义语料库。利用该模型可以训练一个从机器翻译结果到参考翻译文本的释义系统，并生成语义连贯的句子级释义结果。然后，将释义结果引入系统融合的翻译假设候选集。从实验结果来看，该方法的改进比基于词尺度的短语尺度释义的系统融合结果要大，因此我们的机器翻译系统融合具有更大的潜力。

#### 参考文献:

[1] 刘 洋. 神经机器翻译前沿进展 [J]. 计算机研究与发展,

2017, 54 (6): 1144-1149.

[2] 罗华珍, 潘正芹, 易永忠. 人工智能翻译的发展现状与前景分析 [J]. 电子世界, 2017 (21): 21-23.

[3] 刁 洪. 国内翻译技术研究综述 [J]. 北京第二外国语学院学报, 2017, 39 (6): 69-81, 125.

[4] 邵 华, 杨晓峰. 人工+智能语言翻译的现状与发展 [J]. 电子世界, 2018 (16): 28-29.

[5] 杨寅冬, 姚 洁. 神经网络机器翻译之语境式学习 [J]. 太原师范学院学报 (自然科学版), 2019, 18 (3): 26-29.

[6] 高明虎, 于志强. 神经机器翻译综述 [J]. 云南民族大学学报 (自然科学版), 2019, 28 (1): 72-76.

[7] 李亚超, 熊德意, 张 民. 神经机器翻译综述 [J]. 计算机学报, 2018, 41 (12): 2734-2755.

[8] 刘宇鹏, 马春光, 张亚楠. 深度递归的层次化机器翻译模型 [J]. 计算机学报, 2017, 40 (4): 861-871.

[9] 张志锐. 面向神经机器翻译的数据增强方法及应用 [D]. 合肥: 中国科学技术大学, 2019.

[10] 邵 华. 概述机器翻译及其语用分析缺失 [J]. 海外英语, 2019 (3): 146-148.

[11] 韩 冬, 李军辉, 熊德意, 等. 基于子字单元的神经机器翻译未登录词翻译分析 [J]. 中文信息学报, 2018, 32 (4): 74-79, 119.

[12] 冯志伟. 机器翻译与人工智能的平行发展 [J]. 外国语 (上海外国语大学学报), 2018, 41 (6): 35-48.

[13] 李婧萱. 基于深度神经网络的统计机器翻译模型研究 [D]. 哈尔滨: 哈尔滨工业大学, 2016.

[14] 张学强. 基于深度学习的机器翻译技术研究 [D]. 沈阳: 沈阳航空航天大学, 2018.

[15] 李 威. 面向 RNN 语言模型基于主题与词特征的语义强化研究 [D]. 厦门: 华侨大学, 2019.

[16] 杨 丽, 吴雨茜, 王俊丽, 等. 循环神经网络研究综述 [J]. 计算机应用, 2018, 38 (s2): 1-6, 26.

[17] 白瑞芳. 基于 RNN 编码器的交互式机器翻译平台控制技术 [J]. 计算机测量与控制, 2019, 27 (7): 89-92.

[18] 全志斌. 改进的递归网络及其在汉语语言处理中的应用 [D]. 南京: 东南大学, 2018.

[19] 张 尧. 激活函数导向的 RNN 算法优化 [D]. 杭州: 浙江大学, 2017.

[20] 张晶晶, 黄 浩, 胡 英, 等. 口语理解中改进循环神经网络的应用 [J]. 计算机工程与应用, 2019, 55 (18): 155-160.

[21] 刘礼文, 俞 弦. 循环神经网络 (RNN) 及应用研究 [J]. 科技视界, 2019 (32): 54-55.

[22] 陶 南. 基于手机移动端的助教机器人的设计与实现 [J]. 信息与电脑 (理论版), 2019, 31 (24): 76-78.

[23] 王 星. 融合结构信息的神经机器翻译模型研究 [D]. 苏州: 苏州大学, 2018.

[24] 杨 南. 基于神经网络学习的统计机器翻译研究 [D]. 合肥: 中国科学技术大学, 2014.