Artificial Intelligence and Robotics Research 人工智能与机器人研究, 2022, 11(2), 114-121
Published Online May 2022 in Hans. http://www.hanspub.org/journal/airr
https://doi.org/10.12677/airr.2022.112013

Hans汉斯

基于机器阅读理解的生活情景常识预测

邓鉴格，刘宝锴，徐　涛*

西北民族大学，中国民族语言文字信息技术教育部重点实验室，甘肃　兰州

收稿日期：2022年4月7日；录用日期：2022年5月2日；发布日期：2022年5月9日

摘　要

机器学习研究的长期目标是产生适用于推理和自然语言的方法，建立智能对话系统。本实验通过回答日常生活的事件的问答问题来评估阅读理解，使用Facebook AI的BABI tasks中的四种类型数据完成模型训练，采用数字编码稀疏交叉熵损失函数对RNN模型、LSTM模型和BERT模型参数进行设置，采用多分类单标签的categorical_accuracy函数作为评价度量，预测样本数据集中的正确数量。实验结果表明，在RNN模型预测答案的准确率明显高于LSTM和BERT模型。

关键词

机器学习，RNN，LSTM，BERT，生活情境常识

General Knowledge Prediction of Life Situations Based on Machine Reading Comprehension

Jiange Deng, Baokai Liu, Tao Xu*

Key Laboratory of Chinese National Language and Character Information Technology, Ministry of Education, Northwest University for Nationalities, Lanzhou Gansu

Received: Apr. 7th, 2022; accepted: May 2nd, 2022; published: May 9th, 2022

Abstract

The long-term goal of machine learning research is to generate methods applicable to reasoning

*通讯作者。

文章引用: 邓鉴格, 刘宝锴, 徐涛. 基于机器阅读理解的生活情景常识预测[J]. 人工智能与机器人研究, 2022, 11(2): 114-121. DOI: 10.12677/airr.2022.112013

and natural language to build intelligent conversational systems. This experiment evaluates reading comprehension by answering question-and-answer questions about everyday events. Model training is completed using four types of data from Facebook AI's BABI tasks, and the RNN model, LSTM model, and BERT model parameters are set using a digitally encoded sparse cross-entropy loss function, and a multicategorical single-label categorical_ accuracy function is used as an evaluation metric to predict the number of corrections in the sample dataset. The experimental results show that the accuracy of predicting answers in the RNN model is significantly higher than that of the LSTM and BERT models.

Keywords

Machine Learning, RNN, LSTM, BERT, General Knowledge of Life Scenarios

1. 研究现状

随着大数据的来临，信息增长迅速，有效的信息查找离不开信息检索[1]，而对于信息检索的关键是机器阅读理解(machine reading comprehension, MRC) [2]。机器阅读理解[3]中，一般每个问题都对应了文本上下文的相关信息，机器阅读理解的目标是让计算机理解文本，并根据问题从文本中提取出正确的答案或者生成更加复杂的答案。

人工智能系统可以将常识知识[4]作为背景，常识(Commonsense)是社会对同一种事物普遍存在的日常共识。比如：太阳东升西落，树叶都会凋零等，这就是一些常识知识。生活情境常识(Common sense of life situation)是来源于日常生活中的语境所产生的事件，在我们现实生活中经常出现，与大众的现实生活有直接关联性，尤其是在我们熟知的生活情境中是能够看到。对于生活情景常识来说，回答并不是单纯的对内容的理解，还需要根据对话的对象、语境的理解以及基于记忆和知识去理解。

本文通过生活情景常识知识来评测模型的性能。首先使用 bAbI tasks 数据集[5]，选取部分数据分为 Time、Associations、Logic、Context 四种类型，划分训练集和测试集。其次对 N-grams 采用 trigram model，将单词向量转换为矢量嵌入，并构建一个二维的神经网络模型。三个模型对数据进行数字编码，采用语义分割，进行分类处理，将数据转成一维的格式进行实验。最后对 RNN 模型、LSTM 模型和 BERT 模型的损失部分进行改写，修改计算 loss 部分，定义损失函数 sparse_categorical_crossentropy 计算损失函数，在计算损失时使用了交叉熵原理，损失函数中的 from_logits 设置为 True 时，会将 y_pred 转化为概率，采用 softmax 函数，否则不进行转换，通常情况下用 True 结果更稳定，对 loss 进行处理，默认是求平均，计算公式如下所示，其中 n 为样本数，true 为 y_true，pred 为 y_pred。

$$\text{loss} = \frac{1}{n}\sum_i true_i \left(-\lim_e pred_i + le^{-7}\right) \tag{1}$$

并对模型进行训练和测试，训练的过程中使用 RNN 模型、LSTM 模型和 BERT 模型对原文本数据设定了 max_epooch 的固定值 20 次，分别对模型进行对比实验，三种模型均采用 categorical_accuracy 对样本数据进行正确值预测。本实验的研究方法如图 1 所示。

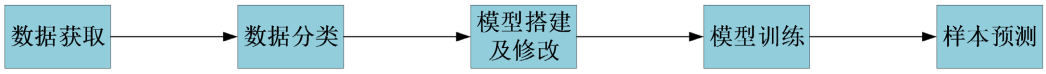

Figure 1. Research methods
图 1. 研究方法

2. 实验过程

2.1. 实验环境

本文中的实验使用 TeslaGPU_M60_16G，显存单卡_10 核，CPU_128G，内存进行训练，版本框架使用 Pytorch1.1.0，语言版本使用 Python 3.7。

2.2. 数据准备

本次实验的数据是由 Weston 等人提出，源于 Facebook AI 的 bAbI tasks 数据集，定义了二十个小的子类推理，涉及 20 种不同的理解能力。本实验选取 4 个典型的生活情境 task 的小样本进行模型训练，如表 1 所示。

所有的数据集的问题表示的形式为元组(question type, clause, support)。例如：

1 John is in the garden.

2 Where is John? garden 1

这个故事表示为一个子句，clause = (true, god, set_property, john, is_in, garden)，后跟一个问题，question = (evaluate, Clause, {1})。像“Is john in the garden?”这样的问题。而是表示为 question = (yes_no, Clause, {1})。

该 4 个 task 数据集训练样本数总数为 40000，测试样本 4000。类型分别为 Time、Associations、Logic、Context，分别去重数分别为 38，48，22 和 44。

Table 1. Sample test sets of different types
表 1. 测试集不同类型样例

类型	质量	排序	稿件
Time	1 This morning Mary moved to the kitchen. 2 This afternoon Mary moved to the cinema. 3 Yesterday Bill went to the bedroom. 4 Yesterday Mary journeyed to the school.	Where was Mary before the cinema?	Kitchen 1
Associations	1 Fred handed the football to Bill. 2 Jeff went back to the office.	Who received the football?	Bill 1
Logic	1 Greg is a rhino. 2 Julius is a rhino. 3 Lily is a swan. 4 Bernhard is a swan. 5 Julius is white. 6 Lily is yellow. 7 Brian is a rhino. 8 Bernhard is yellow. 9 Greg is yellow.	What color is Brian?	White 5
Context	1 Mary took the football there. 2 Sandra picked up the apple there. 3 Mary travelled to the hallway. 4 John journeyed to the kitchen.	Where is the football?	Hallway 3

2.3. 实验参数

本实验采用 RNN 模型、LSTM 模型和 BERT 模型来进行生活情境常识的训练，这三个模型实验所采用的共同参数如表 2 所示。本实验是在服务器上进行训练，配置好环境，分别搭建 RNN 模型、LSTM 模型和 BERT 模型，将四种类型的数据内容分别代入不同的模型中进行训练，并对其样本数进行评测正确值。

Table 2. The same parameters for all three models
表 2. 三个模型的相同参数

参数名称	参数含义	参数取值
dropout_rate	Dropout 概率	1e−1
max_epochs	最大迭代次数	20
batch_size	批量处理大小	64
learning_rate	学习率	1e−4
embedding	词向量维度	128

2.4. 模型选择

循环神经网络(recurrent neural network, RNN)是用于处理序列数据的神经网络，源自于 Saratha Sathasivam 于 1982 年提出的霍普菲尔德网络循环神经网络[6]。通过保存历史信息来帮助当前的决策，使用之前出现过的单词来加强对当前文字的理解，RNN 刻画了循环神经网络的隐藏层之间节点相互之间的链接，隐藏层的输入不仅包括了输入层的输出还包括了上一时刻隐藏层的输出。RNN 是一个非常强大的用于序列建模的神经网络，在深度学习[7]领域占据非常重要的地位。循环神经网络指的是网络的隐含层输出又作为其输入，结构如图 2 所示。

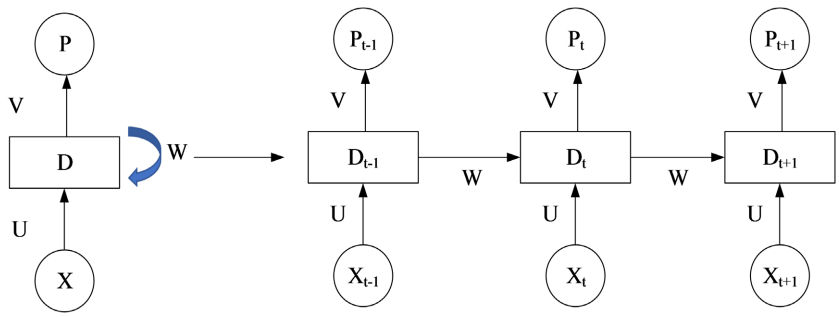

Figure 2. RNN architecture diagram
图 2. RNN 结构图

其中的 U、W、V 作为其输入层到隐含层、隐含层到隐含层和隐含层到输出去层的参数，在使用循环神经网络的时候，设置一个有限的循环次数，将其展开后相当于堆叠多个共享隐含层参数的前馈神经网络。

$$h_t = tanh\left(W^{xh}x_t + b^{xh} + W^{hh}h_{(t-1)} + b^{hh}\right) \tag{2}$$

$$y = softmax\left(W^{hy}h_n + b^{xy}\right) \tag{3}$$

Sepp Hochreiter等人于1997年提出长短时记忆网络(Long Short-Term Memory Neural Network, LSTM) [8]，LSTM是在RNN的基础上进行改进，包含遗忘门、输入门和输出门的网络结构。LSTM是隐藏单元加入了复杂的门控机制的循环网络结构，能够使数据保持长期的依赖性，在实际应用中得到了广泛的应用。图3为LSTM的结构图。

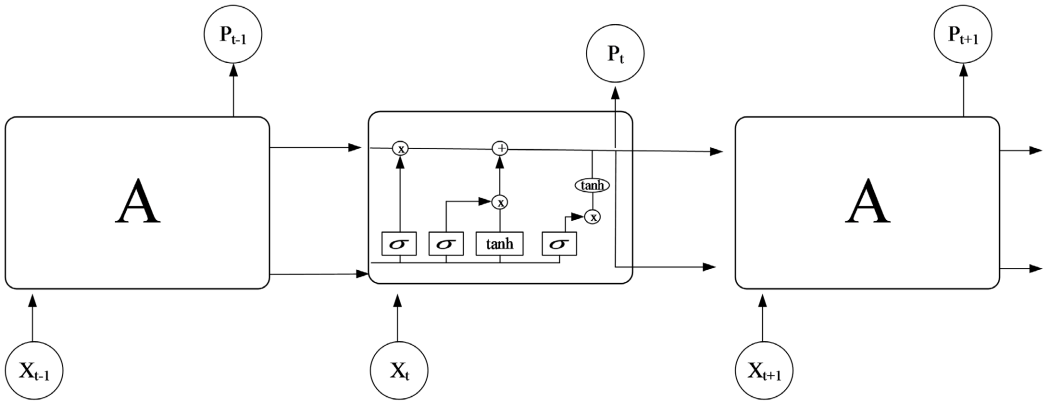

Figure 3. LSTM architecture diagram
图3. LSTM结构图

遗忘门(forget gate)和输入门(input gate)是使循环神经网络更有效的保存长期信息的关键。遗忘门根据当前时刻的输入x_t以及上一时刻的输出h_(t−1)选择性的遗忘上个时刻的状态。输入门根据x_t和h_(t−1)决定将哪些记忆输入到上个时刻的单元状态上并更新单元状态C_t。输出门则用来过滤单元状态信息，并产生此时刻的输出。

$$i_t = sigmoid\left(W_i h_{t-1} + V_i x_t + b\right) \tag{4}$$

$$f_t = sigmoid\left(W_f h_{t-1} + V_f x_t + b\right) \tag{5}$$

$$o_t = sigmoid\left(W_o h_{t-1} + V_o x_t + b\right) \tag{6}$$

$$\hat{C}_t = sigmoid\left(W_f h_{t-1} + V_i x_t + b\right) \tag{7}$$

BERT (Bidirectional Encoder Representations from Transformers)是一个语言表示模型[9]，是利用了Transformer的encoder部分，以CLS为开始部分，SEP为其结束部分所进行。BERT模型是采用了两个预训练任务所进行，一个是双向语言模型，另一个是判断下一段的文本。而BERT在深度学习中是作无监督学习。

BERT预训练模型是结合了GPT (Generative Pre-Training)模型[10]以及EMLO模型的相关优势，引入了Transformer的编码模型，先掌握上游的任务并运用到下游任务中。如图4是BERT模型。

2.5. 实验结果与分析

通过RNN模型对数据集tasks1-4进行训练得出训练损失值。如图5所示，训练集的损失值不断的在减少，在第7个时期左右，数据集都开始趋于平稳，而task2还有明显下降的趋势。

通过LSTM模型对数据集tasks1-4进行训练得出训练损失值，如图6所示，训练集的损失值不断的在减少，除task2以外，其余三个数据集都有小范围的波动。在10个时期，Context数据集增加达到17.5%，然后下降趋于平滑。

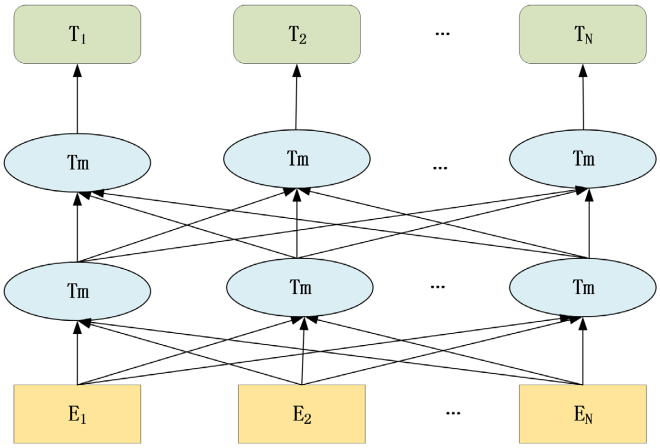

Figure 4. BERT architecture diagram
图 4. BERT 结构图

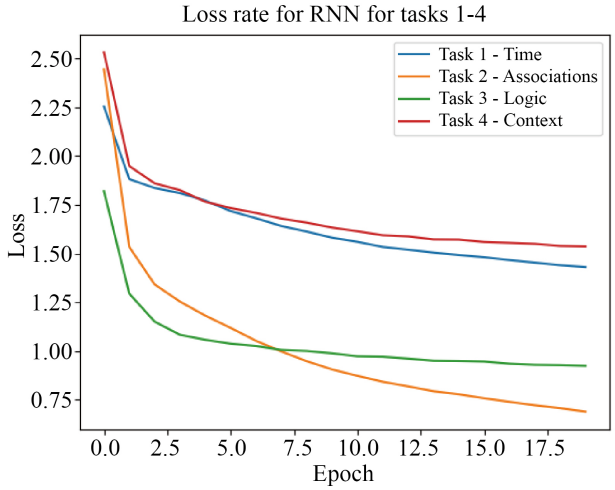

Figure 5. RNN model task1-4 loss rate
图 5. RNN 模型 task1-4 损失率

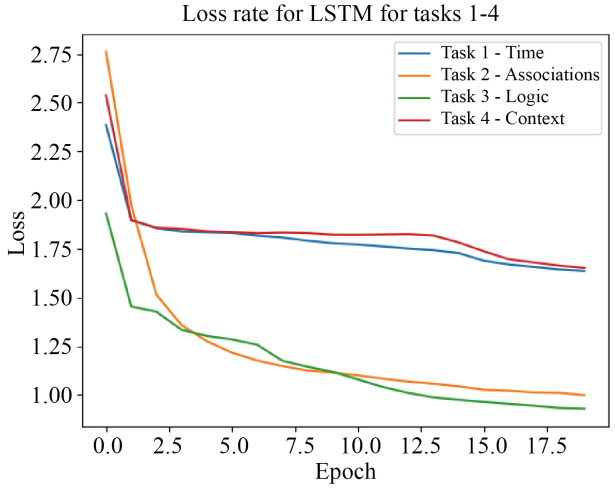

Figure 6. LSTM model task1-4 loss rate
图 6. LSTM 模型 task1-4 损失率

通过 BERT 模型对数据集 tasks1-4 进行训练得出训练损失值，如图 7 所示，整体情况 task1、task2 和 task4 均在第一个时期之后趋于平缓，而 task3 在第九个时期之后趋近于平缓，损失值为 1.38。

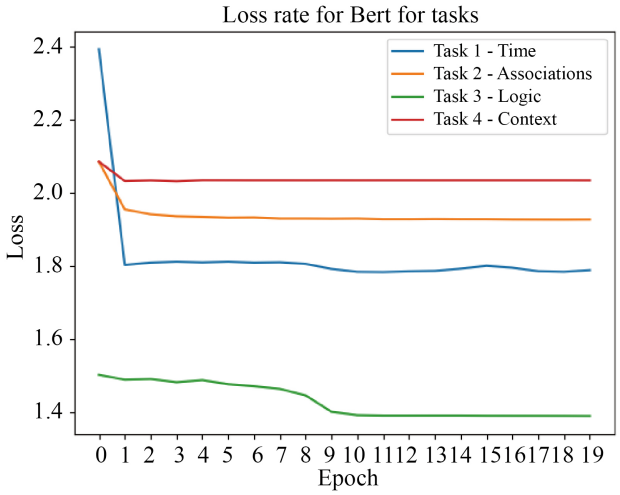

Figure 7. BERT model task1-4 loss rate
图 7. BERT 模型 task1-4 损失率

本实验采用 Keras.metrics 中的 categorical_accuracy 作为评价指标。categorical_accuracy 针对的是 y_true 为 onehot 标签，y_pred 为向量的情况，对多分类问题，计算再所有预测值上的平均正确率。本实验在样本数上进行正确样本数预测，得出其四类数据正确预测结果。如下表 3 所示。

Table 3. Evaluation results of different experimental models
表 3. 不同实验模型评测结果

数据类型	RNN	LSTM	BERT
Time	0.46	0.37	0.08
Associations	0.74	0.51	0.17
Logic	0.54	0.53	0.25
Context	0.40	0.34	0.16

通过对比 RNN 模型、LSTM 模型和 BERT 模型对四种生活情景常识数据进行样本预测的实验结果来看，RNN 模型对于小样本的准确率要高于 LSTM 模型和 BERT 模型，且损失值都低于 LSTM 模型和 BERT 模型，因此 RNN 模型对于数据预测正确值的效果好，BERT 在样本数据少的情况下，呈现出来效果不佳。

3. 结论

在本文中，针对生活情境常识知识任务，选取了 babi task 中四种类型的小样本数据集进行训练和测试，对循环神经网络 RNN、长短时记忆网络 LSTM 和语言表示模型 BERT 进行模型概述，然后设置固定的参数指标进行实际训练，采用 sparse_categorical_crossentropy 交叉熵得出模型在四个 tasks 的损失率，并进行分析，最后样本数据选择 categorical_accuracy 进行结果的预测。分析对比三者实验的结果，验证了 RNN 模型对于小样本的生活情境常识准确性好。

基金项目

国家科技重大专项(2017YFB1002103)、甘肃省青年科技基金(21JR1RA21)、中央高校基本科研业务费专项资金项目(31920210017)、国家档案局科技项目(2021-X-56)与甘肃省档案科技项目(GS-2020-X-07G)。

参考文献

[1] 张海涛, 张枭慧, 魏萍, 刘雅姝. 网络用户信息检索行为研究进展[J]. 情报科学, 2020, 38(5): 169-176. https://doi.org/10.13833/j.issn.1007-7634.2020.05.024

[2] 顾迎捷, 桂小林, 李德福, 沈毅, 廖东. 基于神经网络的机器阅读理解综述[J]. 软件学报, 2020, 31(7): 2095-2126. https://doi.org/10.13328/j.cnki.jos.006048

[3] Liu, S., Zhang, X., Zhang, S., *et al*. (2019) Neural Machine Reading Comprehension: Methods and Trends. *Applied Sciences*, **9**, 3698. https://doi.org/10.3390/app9183698

[4] 李闪闪, 曹存根. 事件前提和后果常识知识分析方法研究[J]. 计算机科学, 2013(4): 185-192.

[5] Weston, J., Bordes, A., Chopra, S., *et al*. (2015) Towards AI-Complete Question Answering: A Set of Prerequisite Toy Tasks. https://arxiv.org/abs/1502.05698

[6] 杨丽, 吴雨茜, 王俊丽, 刘义理. 循环神经网络研究综述[J]. 计算机应用, 2018, 38(S2): 1-6.

[7] 刘建伟, 刘媛, 罗雄麟. 深度学习研究进展[J]. 计算机应用研究, 2014, 31(7): 11.

[8] Sundermeyer, M., Schlüter, R. and Ney, H. (2012) LSTM Neural Networks for Language Modeling. Interspeech. https://doi.org/10.21437/Interspeech.2012-65

[9] Devlin, J., Chang, M.W., Lee, K., *et al*. (2018) BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding. *Proceedings of the* 2019 *Conference of the North American Chapter of the Association for Computational Linguistics*: *Human Language Technologies*, Minneapolis, 4171-4186.

[10] Radford, A., Narasimhan, K., Salimans, T., *et al*. (2018) Improving Language Understanding by Generative Pre-Training. https://www.semanticscholar.org/paper/Improving-Language-Understanding-by-Generative-Radford-Narasimhan/cd18800a0fe0b668a1cc19f2ec95b5003d0a5035