

RNN进展

1982年，美国加州理工学院物理学家John Hopfield发明了一种单层反馈神经网络 Hopfield network，用来解决组合优化问题。这是最早的RNN的雏形。86年，Michael I. Jordan 定义了recurrent的概念，提出 Jordan network。1990年，美国认知科学家Jeffrey L. Elman 对 Jordan network进行了简化，并采用BP算法进行训练，便有了如今最简单的包含单个自连接节点的RNN 模型。但此时RNN由于梯度消失 (gradient vanishing)及梯度爆炸 (gradient exploding)的问题，训练非常困难，应用非常受限。直到1997年，人工智能研究所的主任 Jurgen Schmidhuber 提出长短期记忆 (LSTM)，LSTM使用门控单元及记忆机制大大缓解了早期RNN训练的问题。同样在1997年，Mike Schuster 提出双向RNN模型 (Bidirectional RNN)。这两种模型大大改进了早期RNN结构，拓宽了RNN的应用范围，为后续序列建模的发展奠定了基础。此时RNN虽然在一些序列建模任务上取得了不错的效果，但由于计算资源消耗大，后续几年一直没有太大的进展。

2010年，Tomas Mikolov对Bengio提出的feedforward Neural network language model (NNLM) 进行了改进，提出了基于RNN的语言模型 (RNN LM)，并将其用在语音识别任务中，大幅提升了识别精度。再此基础上Tomas Mikolov于2013年提出了大名鼎鼎的word2vec，与NNLM及RNNLM不同，word2vec的目标不再专注于建模语言模型，而是如何利用语言模型学习每个单词的语义化向量 (distributed representation)，其中distributed representation概念最早要来源于

Hinton 1986年的工作。Word2vec引发了深度学习在自然语言处理领域的浪潮，除此之外还启发了knowledge representation，network representation等新的领域。

另一方面，2014年，Bengio团队与google几乎同时提出了seq2seq架构，将RNN用于机器翻译。没过多久，Bengio团队又提出attention机制，对seq2seq架构进行改进。自此机器翻译全面进入到神经机器翻译（NMT）的时代，NMT不仅过程简单，而且效果要远超统计机器翻译的效果。目前主流的机器翻译系统几乎都采用了神经机器翻译的技术。除此之外，attention机制也被广泛用于基于深度学习的各种任务中。

近两年，相关领域仍有一些突破性进展，2017年，facebook人工智能实验室提出基于卷积神经网络的seq2seq架构，将rnn替换为带有门控单元的cnn，提升效果的同时大幅加快了模型训练速度，此后不久，google提出transformer架构，使用self-attention代替原有的RNN及CNN，更进一步降低了模型复杂度。在词表示学习方面，Allen人工智能研究所2018年提出上下文相关的表示学习方法ELMo，利用双向LSTM语言模型对不同语境下的单词学习不同的向量表示，在6个nlp任务上取得了提升。OpenAI团队在此基础上提出预训练模型GPT，把LSTM替换为transformer来训练语言模型，在应用到具体任务时，与之前学习词向量当作特征的方式不同，GPT直接在预训练得到的语言模型最后一层接上softmax作为任务输出层，然后再对模型进行微调，在多项任务上GPT取得了更好的效果。不久之后，Google提出BERT模型，将GPT中的单向语言模型拓展为双向语言模型（Masked Language Model），并在预训练中引入了sentence prediction任务。BERT模型在11个任务中取得了最好的效果，是深度学习在nlp领域又一个里程碑式的工作。