

## 循环神经网络速成

用来处理序列数据的神经网络，被称为循环神经网络（RNNs）。在传统的神经网络模型中，层与层之间是全连接的，每层的节点之间是无连接的。但是这种传统的神经网络对序列问题的处理是低效的。例如，在文本预测的问题中，要预测句子中的下一个单词是什么，一般需要考虑前面的单词，因为句子中的前后单词并不是独立的。处理序列数据的神经网络之所以称为循环神经网络，是因为一个序列当前的输出与前面的输出有关。具体的表现形式为：网络会对前面的信息进行记忆，并应用于当前输出的计算中，即隐藏层之间的节点不再是无连接的，而是有连接的，并且隐藏层的输入不仅包括输入层的输出，还包括上一时刻隐藏层的输出。理论上，循环神经网络能够对任何长度的序列数据进行处理。循环神经网络具有循环连接，随着时间的推移向网络增加反馈和记忆。这种记忆能力增强了循环神经网络对序列问题的网络学习和泛化输入能力。图 17-1 是一个简单的循环神经网络的拓扑结构。

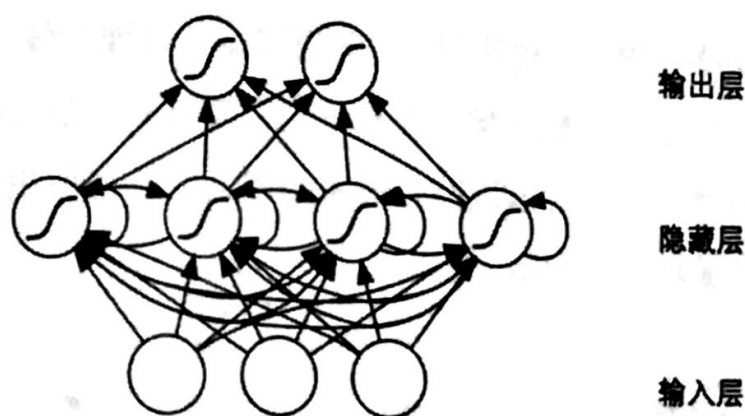


图 17-1

被称为长短期记忆网络 (LSTM) 的循环神经网络已被证明, 对处理从自然语言翻译到各种各样图像和视频自动字幕等序列问题具有非常良好的结果。目前, 基于 LSTM 的系统被广泛应用于自然语言翻译、控制机器人、图像分析、文档摘要、语音识别、图像识别、手写识别、聊天机器人、预测疾病、点击率、股票、合成音乐等领域。

## 17.1 处理序列问题的神经网络

有一些问题是非常典型的序列问题, 这些问题都涉及序列作为输入或输出。例如, 时间序列问题中的股票价格预测问题。这类问题的数据集可以通过定义窗口大小 (如 5), 来构建经典的前馈多层感知器网络, 并以固定窗口大小的输入进行短期预测的训练。

虽然使用窗口方法的多层感知器可以工作, 但是工作能力非常有限。输入窗口为问题增加了记忆功能, 但仅限于固定数量的神经元, 必须通过对问题的充分了解来选择合适的窗口大小。一个不合适的窗口将不会捕捉可能与预测有关的几分钟、几小时或几天等相关的变化趋势。从一个预测到下一个预测, 网络只知道提供的具体输入, 而不能很好地处理其中的相关性。考虑输入与输出的关系, 序列问题具有以下分类。

- 一对多: 序列输出, 用于图像字幕。
- 多对一: 序列输入, 用于情感分类。
- 多对多: 序列输入和输出, 用于机器翻译。
- 同步多对多: 同步序列输入和输出, 用于视频分类。

经典的前馈神经网络的输入与输出是一一对应的，如处理图像识别问题的神经网络。处理序列问题的神经网络是目前机器学习中的一个重要分类，最近的研究表明，深度学习对序列问题的处理具有非常好的适用性，最新的研究中一直使用一种专门针对序列问题设计的神经网络，即循环神经网络。

## 17.2 循环神经网络

循环神经网络是针对序列问题设计的一类特殊的神经网络。可以认为循环神经网络是在一个标准的多层感知器的架构上增加循环连接。例如，在给定的层中，每个神经元可以向其最近（侧向）的神经元传递信号，而不是只向前（下一层）传递信号；网络的输出可以作为下一次输入的输入向量反馈给网络等。

循环连接将状态或记忆能力添加到网络，并提高了网络从输入序列中学习和抽象的能力。循环神经网络的领域相当成熟，并具有非常流行的方法。为了使用循环神经网络有效地处理实际问题，需要解决两个主要问题。

- 如何训练具有反向传播的神经网络。
- 如何解决训练过程中梯度消失或梯度爆炸的问题。

用于训练前馈神经网络的主要技术是反向传播，使用计算错误的权重来更新网络权重。在循环神经网络中，因为神经元的重复连接或循环连接，反向传播存在不适用性，不能很好地完成网络权重的更新。这需要使用时间反向传播（BPTT）技术来解决循环神经网络的反向传播问题。

将循环神经网络的结构展开，可以比较清晰地描述时间反向传播。例如，具有自身连接的单个神经元（图 17-2 的左部分），展开网络结构，可以表示为具有相同权重值的多个神经元（图 17-2 的右部分）。这样可以将神经网络的循环图，变成与经典前馈神经网络同样的非循环图，并且可以使用反向传播。

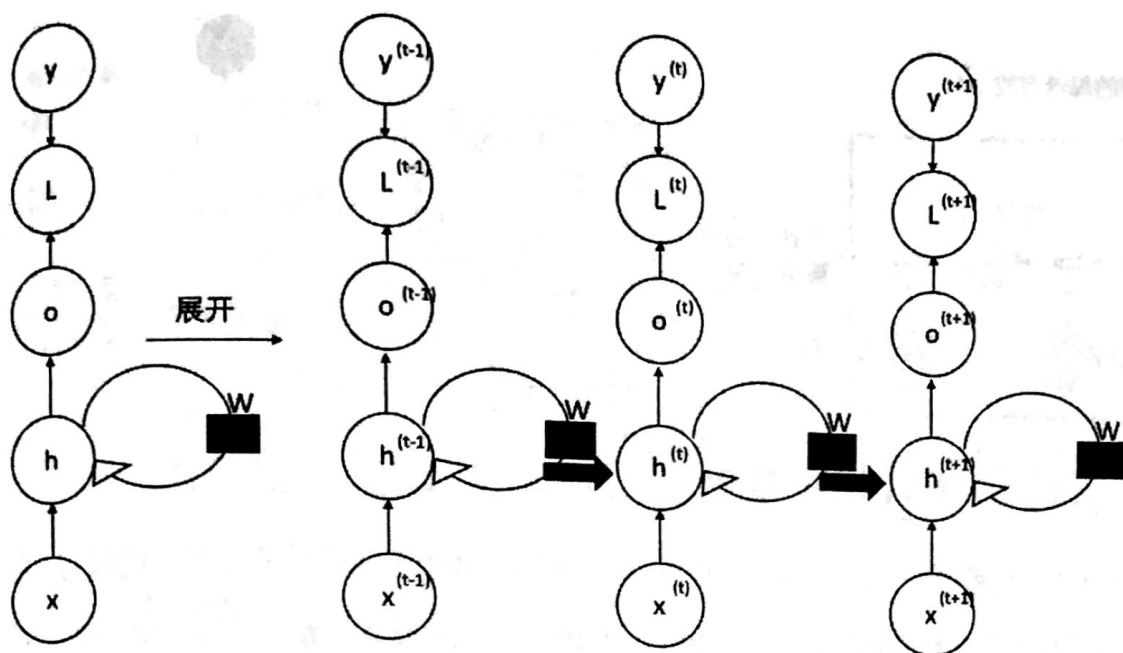


图 17-2

当反向传播用于非常深的神经网络和循环神经网络时，为了更新权重而需要计算的梯度可能变得不稳定，它们可能变成非常大的数值（梯度爆炸）或非常小的数值（梯度消失）。使用这些数据来更新网络中的权重，使训练变得不稳定，并且使神经网络生成的模型不可靠。

在多层感知器神经网络中，通过使用激活函数使这个问题得到缓解，甚至使用无监督的预训练层来缓解这个问题；在循环神经网络架构中，使用长短期记忆网络可以缓解这个问题。

### 17.3 长短期记忆网络

长短期记忆网络是一种常见的循环神经网络，使用时间反向传播训练，并解决了梯度消失的问题。因此，可以利用长短期记忆网络来创建大型的循环神经网络，用于解决机器学习中比较复杂的序列问题，并且长短期记忆网络对序列问题的处理具有很高的效率。在长短期记忆网络中使用存储单元代替常规的神经元，每个存储单元由输入门、输出门和自有状态构成（见图 17-3）。

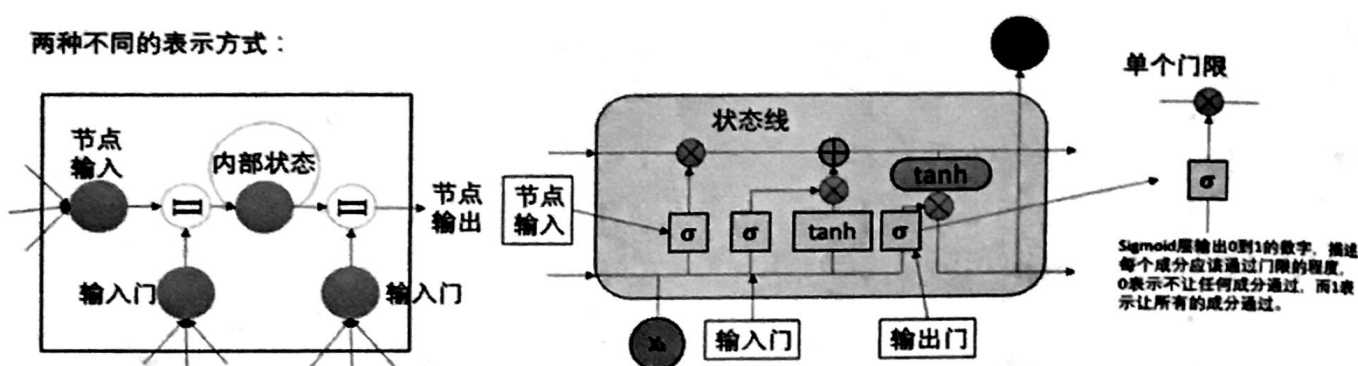


图 17-3

存储单元是比神经元更聪明的组件，并且对最近序列具有记忆功能。存储单元包含自有状态和输出门，根据输入序列来操作每一个存储单元，存储单元内的每个门使用 **sigmoid** 激活函数来控制它们是否被触发，有条件地进行状态改变，以及控制通过存储单元中的信息流动。存储单元内有以下三种类型的门。

- 遗忘门：有条件地决定哪些信息从单元中抛弃。
- 输入门：有条件地决定在单元中存储哪些信息。
- 输出门：有条件地决定哪些信息需要输出，并输出信息。

每个存储单元就像一个迷你状态机，其中单元的门具有在训练过程中学习的权重。理解了一层的长短期记忆网络如何实现复杂的学习和记忆，就不难想象如何通过多个这样的层进行更复杂的抽象学习。