第17章 问答系统算法

《人工智能算法》

清华大学出版社
2022年7月

提纲

- 引例
- 问答系统的基本思想
- 循环神经网络
- 长短期记忆网络
- 基于LSTM的问答系统构建
- 总结

引例 (1)

- **典型的问答系统场景**

q_1

客户 我当前使用的是多少的套餐？

客服 您当前使用的是七十八套餐，包含20GB流量和300分钟通话时间。

q_2

客户 我每个月使用多少流量？

客服 平均每个月使用25GB流量。

q_3

客户 目前最适合我的套餐是什么？

客服 九十八套餐，它包含30GB流量和500分钟通话时间。

引例 (2)

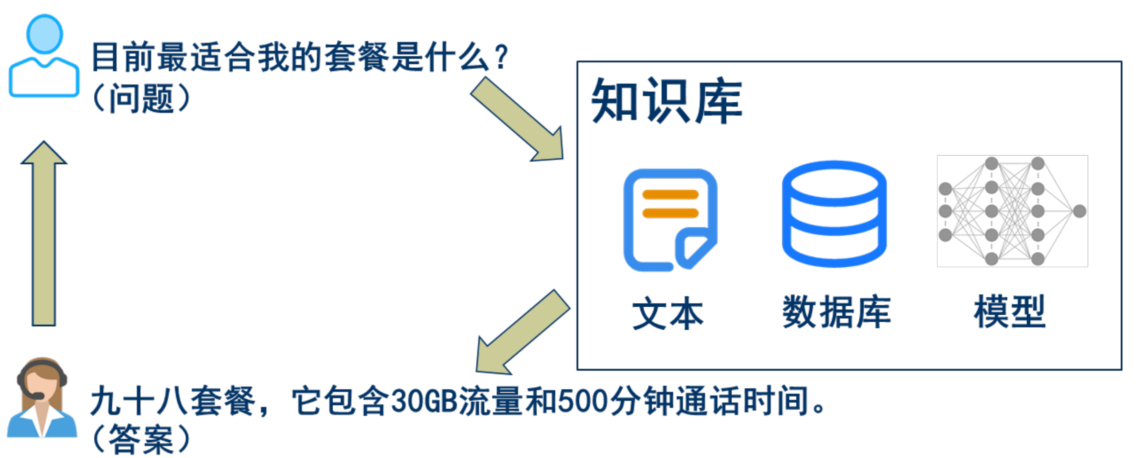
- q_1和q_2可直接查询知识库获得答案
- q_3需利用问答系统求解

（1）提取q_1和q_2的上下文语义信息。

（2）识别实体“七十八套餐”。

（3）链接知识库，确定与该实体对应的子集，生成候选答案集矩阵。

（4）计算候选答案中与问题相似度最高的实体。



提纲

- 引例
- 问答系统的基本思想
- 循环神经网络
- 长短期记忆网络
- 基于LSTM的问答系统构建
- 总结

问答系统的基本思想

- **问答系统**（**Question Answering System**）是信息检索系统的一种高级形式，它通过理解用户意图和问题语义而获取相关的知识，通过推理计算得到答案并返回给用户。
- 问答系统主要由问题分析、信息检索和答案抽取三部分组成。

问答系统
- 问题分析
  - 传统方法：词性分析、句法分析、语义分析等。
  - 新方法：深度神经网络模型，如RNN，LSTM等。
- 信息检索：表征答案候选集
- 答案抽取：计算问题与答案的最佳得分

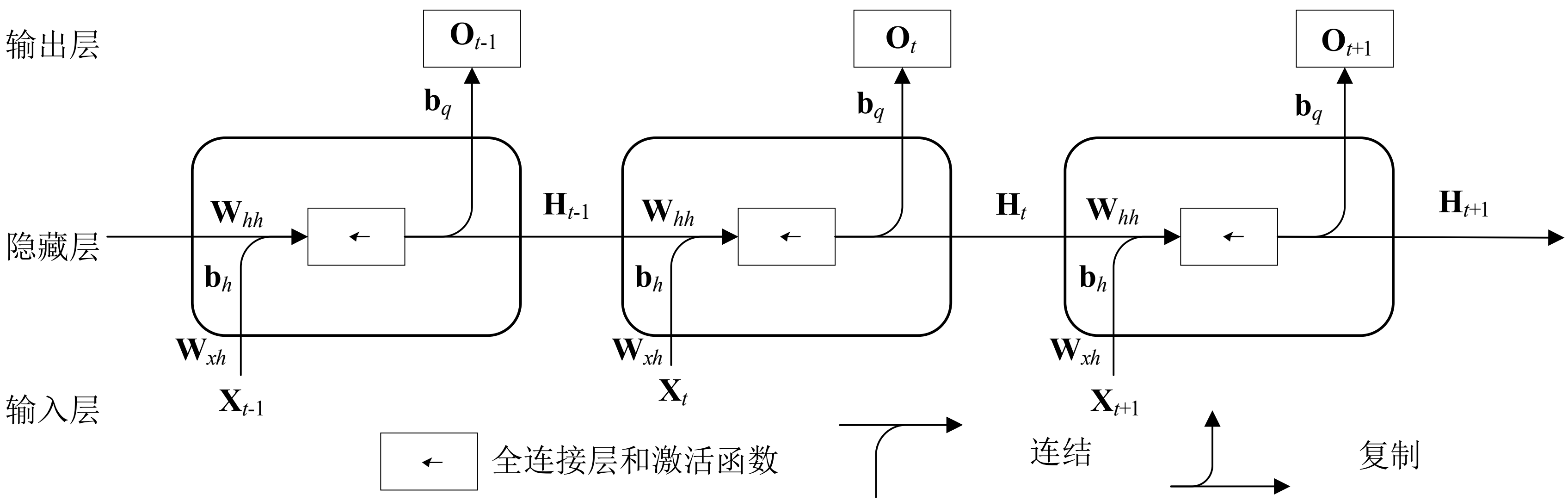
提纲

- 引例
- 问答系统的基本思想
- 循环神经网络
- 长短期记忆网络
- 基于LSTM的问答系统构建
- 总结

循环神经网络 (1)

- **循环神经网络（Recurrent Neural Network, RNN）**

  RNN可对时序信息进行建模，包括输入层、输出层和隐藏层。



循环神经网络 (2)

- **模型结构**

（1）输入层：输入RNN的序列，$\mathbf{X}_t$表示序列中时间步t的输入。

（2）隐藏层：读取当前时间步的输入$\mathbf{X}_t$和上一时间步的隐藏变量$\mathbf{H}_{t-1}$，通过激活函数得到当前时间步的隐藏变量$\mathbf{H}_t$：

$$\mathbf{H}_t = \varphi(\mathbf{X}_t\mathbf{W}_{xh} + \mathbf{H}_{t-1}\mathbf{W}_{hh} + \mathbf{b}_h)$$

（3）输出层：将$\mathbf{H}_t$经过权重参数和偏置参数计算得到输出$\mathbf{O}_t$：

$$\mathbf{O}_t = \mathbf{H}_t\mathbf{W}_{hq} + \mathbf{b}_q$$

循环神经网络 (3)

- **模型训练**

RNN在训练时使用反向传播算法，设RNN的激活函数为$\phi(x) = x$，

用交叉熵（Cross-Entropy）损失函数定义时间步t（$1 \le t \le T$）的损失：

$$L(\mathbf{O}_t, \mathbf{y}_t) = -\mathbf{y}_t \log \mathbf{O}_t$$

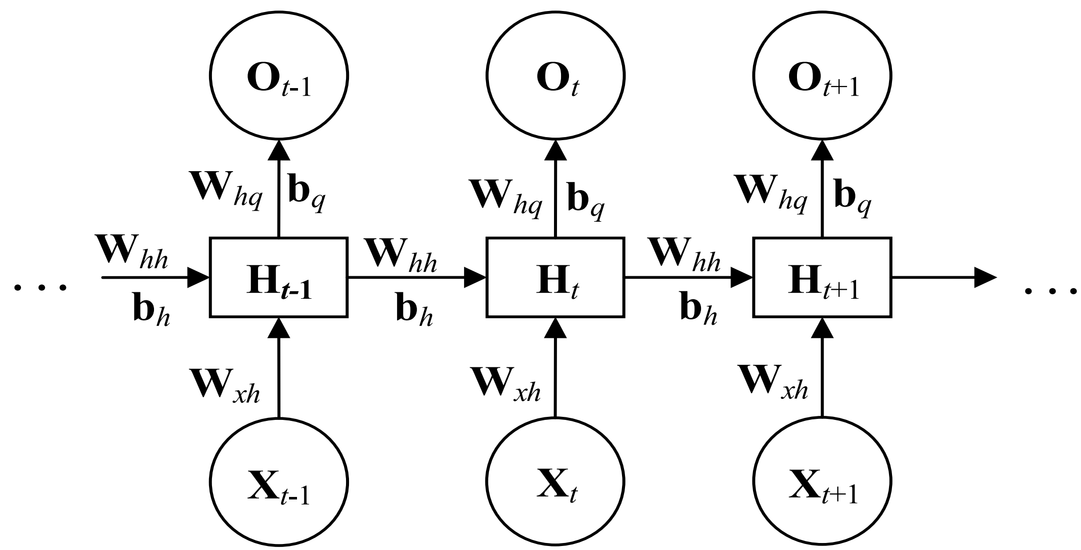
当时间步数为T时，模型的损失函数定义如下：

$$L = \frac{1}{T} \sum_{t=1}^{T} L(\mathbf{O}_t, \mathbf{y}_t)$$

循环神经网络 (4)

- **模型训练**

RNN在训练过程中的参数包括$\mathbf{W}_{xh}$、$\mathbf{W}_{hh}$、$\mathbf{W}_{hq}$、$\mathbf{b}_h$和$\mathbf{b}_q$， 其变量和参数在训练时的依赖关系：



循环神经网络 (5)

- **模型训练**

（1）计算各时间步输出层变量O_t的梯度：$\frac{\partial L}{\partial \mathbf{O}_t} = \frac{\partial L(\mathbf{O}_t, \mathbf{y}_t)}{T \cdot \partial \mathbf{O}_t}$

（2）计算隐藏变量$\mathbf{H}_t$的梯度

$$\frac{\partial L}{\partial \mathbf{H}_t} = \mathbf{W}_{hh}^{trans} \frac{\partial L}{\partial \mathbf{H}_{t+1}} + \mathbf{W}_{hq}^{trans} \frac{\partial L}{\partial \mathbf{O}_t} = \sum_{i=t}^{trans} (\mathbf{W}_{hh}^{trans})^{T-i} \mathbf{W}_{hq}^{trans} \frac{\partial L}{\partial \mathbf{O}_{T+t-i}}$$

（3）计算参数W_{hq}和b_q的梯度：$\frac{\partial L}{\partial \mathbf{W}_{hq}} = \sum_{t=1}^{T} \frac{\partial L}{\partial \mathbf{O}_t} \mathbf{H}_t^{trans}$，$\frac{\partial L}{\partial b_q} = \sum_{t=1}^{T} \frac{\partial L}{\partial \mathbf{O}_t}$

（4）计算参数$\mathbf{W}_{xh}$、$\mathbf{W}_{hh}$ 和$\mathbf{b}_h$的梯度：

损失L通过$\mathbf{H}_1, \mathbf{H}_2, \cdots, \mathbf{H}_T$依赖于参数$\mathbf{W}_{xh}$、$\mathbf{W}_{hh}$和$\mathbf{b}_h$，根据链式求导法则，有：

$$\frac{\partial L}{\partial \mathbf{W}_{xh}} = \sum_{t=1}^{T} \frac{\partial L}{\partial \mathbf{H}_t} \mathbf{X}_t^{trans}, \quad \frac{\partial L}{\partial \mathbf{W}_{hh}} = \sum_{t=1}^{T} \frac{\partial L}{\partial \mathbf{H}_t} \mathbf{H}_{t-1}^{trans}, \quad \frac{\partial L}{\partial \mathbf{b}_h} = \sum_{t=1}^{T} \frac{\partial L}{\partial \mathbf{H}_t}$$

（5）权重更新：使用梯度下降法，以目标的负梯度方向对参数进行更新：

$$\mathbf{W}_{hq} \leftarrow \mathbf{W}_{hq} - \eta\, \partial L / \partial \mathbf{W}_{hq}, \ \mathbf{W}_{xh} \leftarrow \mathbf{W}_{xh} - \eta\, \partial L / \partial \mathbf{W}_{xh}, \mathbf{W}_{hh} \leftarrow \mathbf{W}_{hh} - \eta\, \partial L / \partial \mathbf{W}_{hh}$$

$$\mathbf{b}_h \leftarrow \mathbf{b}_h - \eta\, \partial L / \partial \mathbf{b}_h, \ \mathbf{b}_q \leftarrow \mathbf{b}_q - \eta\, \partial L / \partial \mathbf{b}_q$$

循环神经网络 (6)

- ## 训练算法

随机初始化网络的权重参数$\mathbf{W}_{xh}$、$\mathbf{W}_{hh}$和$\mathbf{W}_{hq}$，使其均服从正态分布，初始化i=1

While $i \le n_{RNN}$ Do

For each $(\mathbf{X}_t, \mathbf{y}_t) \in D$ Do

由$\mathbf{H}_t = \varphi(\mathbf{X}_t\mathbf{W}_{xh} + \mathbf{H}_{t-1}\mathbf{W}_{hh} + \mathbf{b}_h)$计算时间步$t$的隐藏变量$\mathbf{H}_t$

由$\mathbf{O}_t = \mathbf{H}_t\mathbf{W}_{hq} + \mathbf{b}_q$计算时间步$t$的输出$\mathbf{O}_t$

由$L(\mathbf{O}_t, \mathbf{y}_t) = -\mathbf{y}_t\log\mathbf{O}_t$计算时间步$t$的损失函数$l(\mathbf{O}_t, \mathbf{y}_t)$

由$L = \frac{1}{T}\sum_{t=1}^{T} L(\mathbf{O}_t, \mathbf{y}_t)$计算损失函数$Loss$

由$\frac{\partial L}{\partial \mathbf{O}_t} = \frac{\partial L(\mathbf{O}_t, \mathbf{y}_t)}{T \cdot \partial \mathbf{O}_t}$计算时间步$t$变量$\mathbf{O}_t$的梯度$\partial Loss/\partial \mathbf{O}_t$

由$\frac{\partial L}{\partial \mathbf{H}_t} = \mathbf{W}_{hh}^{trans}\frac{\partial L}{\partial \mathbf{H}_{t+1}} + \mathbf{W}_{hq}^{trans}\frac{\partial L}{\partial \mathbf{O}_t} = \sum_{i=t}^{trans}(\mathbf{W}_{hh}^{trans})^{T-i}\,\mathbf{W}_{hq}^{trans}\frac{\partial L}{\partial \mathbf{O}_{T+t-i}}$ 计算时间步t变量$\mathbf{H}_t$的梯度$\partial L/\partial \mathbf{H}_t$

End For

由$\frac{\partial L}{\partial \mathbf{W}_{hq}} = \sum_{t=1}^{T}\frac{\partial L}{\partial \mathbf{O}_t}\mathbf{H}_t^{trans}$ 计算第i次迭代参数$\mathbf{W}_{hq}$的梯度

由$\frac{\partial L}{\partial \mathbf{b}_q} = \sum_{t=1}^{T}\frac{\partial L}{\partial \mathbf{O}_t}$计算第$i$次迭代参数$\mathbf{b}_q$的梯度

由$\frac{\partial L}{\partial \mathbf{W}_{xh}} = \sum_{t=1}^{T}\frac{\partial L}{\partial \mathbf{H}_t}\mathbf{X}_t^{trans}$ 计算第i次迭代参数$\mathbf{W}_{xh}$的梯度

由$\frac{\partial L}{\partial \mathbf{W}_{hh}} = \sum_{t=1}^{T}\frac{\partial L}{\partial \mathbf{H}_t}\mathbf{H}_{t-1}^{trans}$ 计算第i次迭代参数$\mathbf{W}_{hh}$的梯度

由$\frac{\partial L}{\partial \mathbf{b}_h} = \sum_{t=1}^{T}\frac{\partial L}{\partial \mathbf{H}_t}$ 计算第i次迭代参数$\mathbf{b}_h$的梯度

由梯度$\partial L/\partial \mathbf{W}_{xh}$、$\partial L/\partial \mathbf{W}_{hh}$、$\partial L/\partial \mathbf{W}_{hq}$、$\partial L/\partial \mathbf{b}_h$和$\partial L/\partial \mathbf{b}_q$，更新参数$\mathbf{W}_{hq}$、$\mathbf{W}_{xh}$、$\mathbf{W}_{hh}$、$\mathbf{b}_h$和$\mathbf{b}_q$

End While

Return $\mathbf{W}_{xh}$, $\mathbf{W}_{hh}$,$\mathbf{W}_{hq}$,$\mathbf{b}_h$和$\mathbf{b}_q$

时间复杂度：

$O(n_{RNN}Td^2h)$

提纲

- 引例
- 问答系统的基本思想
- 循环神经网络
- 长短期记忆网络
- 基于LSTM的问答系统构建
- 总结

长短期记忆网络 (1)

- **长短期记忆网络（Long Short-Term Memory, LSTM）**

✓ 当需处理长文档的提取问题与答案的表征时，传统的RNN具有“梯度爆炸”或“梯度消失”的局限性。

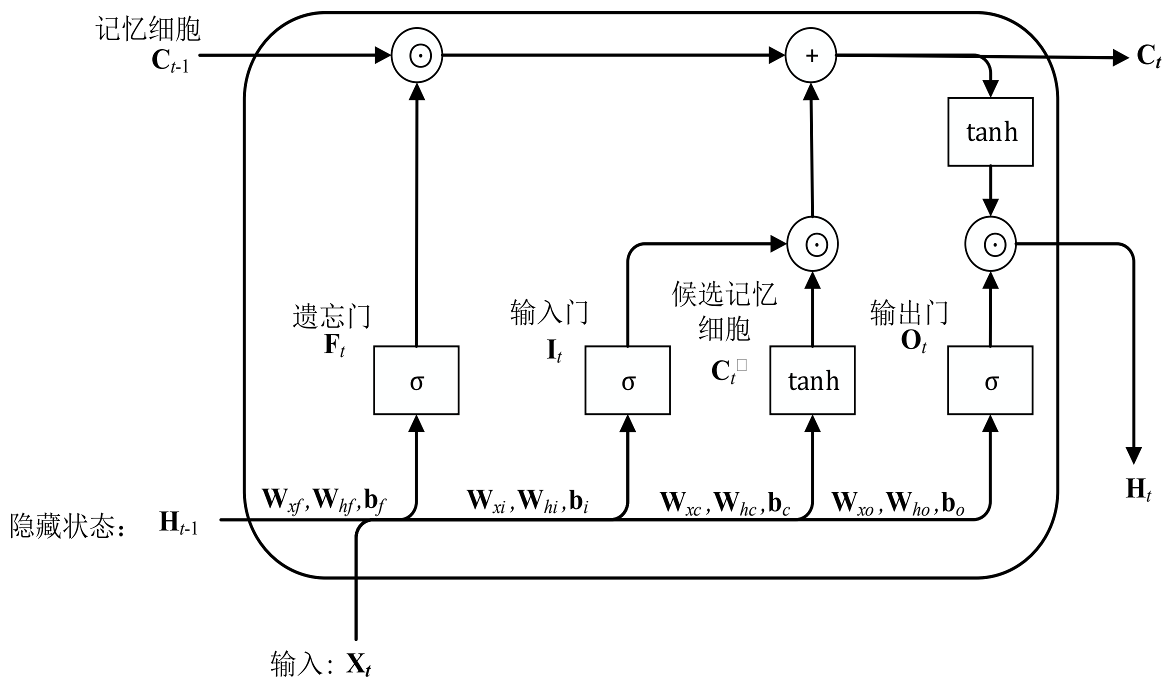
✓ LSTM运用门控机制，可很好地处理长序列之间的依赖关系。

传统的RNN存在什么问题

长短期记忆网络 (2)

- **模型结构**

LSTM引入遗忘门（Forget Gate）、输入门（Input Gate）、记忆细胞和输出门（Output Gate），记录额外信息，避免RNN梯度消失和梯度爆炸问题。



长短期记忆网络 (3)

- **模型结构**

（1）遗忘门：LSTM根据遗忘门决定丢弃的信息。

读取时间步t–1的输出$\mathbf{H}_{t-1}$和时间步t的输入$\mathbf{X}_t$，通过Sigmoid函数（记为σ）来计算遗忘门$\mathbf{F}_t$：

$$\mathbf{F}_t = \sigma(\mathbf{X}_t \mathbf{W}_{xf} + \mathbf{H}_{t-1} \mathbf{W}_{hf} + \mathbf{b}_f)$$

（2）输入门：通过σ决定值的更新。

$$\mathbf{I}_t = \sigma(\mathbf{X}_t \mathbf{W}_{xi} + \mathbf{H}_{t-1} \mathbf{W}_{hi} + \mathbf{b}_i)$$

长短期记忆网络 (4)

- **模型结构**

（3）记忆细胞： 由tanh函数创建新的候选值$\mathbf{C}_t{}'$，并将$\mathbf{C}_{t-1}$更新为$\mathbf{C}_t$：

$$\mathbf{C}_t{}' = \tanh(\mathbf{X}_t\mathbf{W}_{xc} + \mathbf{H}_{t-1}\mathbf{W}_{hc} + \mathbf{b}_c)$$

$$\mathbf{C}_t = \mathbf{F}_t \odot \mathbf{C}_{t-1} + \mathbf{I}_t \odot \mathbf{C}_t{}'$$

（4）输出门： 通过σ函数得到初始输出$\mathbf{O}_t$，通过tanh函数计算时间步t的隐藏变量$\mathbf{H}_t$：

$$\mathbf{O}_t = \sigma(\mathbf{X}_t\mathbf{W}_{xo} + \mathbf{H}_{t-1}\mathbf{W}_{ho} + \mathbf{b}_o)$$

$$\mathbf{H}_t = \mathbf{O}_t \odot \tanh(\mathbf{C}_t)$$

长短期记忆网络 (5)

- **模型训练**

（1）计算t时刻的误差项δ_t，时间步t的隐藏变量为$\mathbf{H}_t$：

$$\delta_t = \frac{\partial L}{\mathbf{H}_t}$$

为便于描述梯度计算方法，定义$net_{f,t}$、$net_{i,t}$、$net_{c,t}$和$net_{o,t}$：

$$net_{f,t} = X_t W_{xf} + H_{t-1} W_{hf} + b_f$$

$$net_{i,t} = X_t W_{xi} + H_{t-1} W_{hi} + b_i$$

$$net_{c,t} = X_t W_{xc} + H_{t-1} W_{hc} + b_c$$

$$net_{o,t} = X_t W_{xo} + H_{t-1} W_{ho} + b_o$$

长短期记忆网络 (6)

- **模型训练**

（2）计算梯度$\partial L/\partial \mathbf{net}_{f,t}$、$\partial L/\partial \mathbf{net}_{i,t}$、$\partial L/\partial \mathbf{net}_{c,t}$和$\partial L/\partial \mathbf{net}_{o,t}$：

$$\frac{\partial L}{\partial \mathbf{net}_{f,t}} = \delta_t^{trans} \odot \mathbf{O}_t \odot (\mathbf{1} - \tanh(\mathbf{C}_t)^2) \odot \mathbf{C}_{t-1} \odot \mathbf{F}_t(\mathbf{1} - \mathbf{F}_t)$$

$$\frac{\partial L}{\partial \mathbf{net}_{i,t}} = \delta_t^{trans} \odot \mathbf{O}_t \odot (\mathbf{1} - \tanh(\mathbf{C}_t)^2) \odot \mathbf{C}_t' \odot \mathbf{I}_t(\mathbf{1} - \mathbf{I}_t)$$

$$\frac{\partial L}{\partial \mathbf{net}_{c,t}} = \delta_t^{trans} \odot \mathbf{O}_t \odot (\mathbf{1} - \tanh(\mathbf{C}_t)^2) \odot \mathbf{I}_t(\mathbf{1} - (\mathbf{C}_t')^2)$$

$$\frac{\partial L}{\partial \mathbf{net}_{o,t}} = \delta_t^{trans} \odot \tanh(\mathbf{C}_t) \odot \mathbf{O}_t(\mathbf{1} - \mathbf{O}_t)$$

长短期记忆网络 (7)

- **模型训练**

（3）将各个时间步的梯度进行累加，得到$\mathbf{W}_{hf}$、$\mathbf{W}_{hi}$、$\mathbf{W}_{hc}$和$\mathbf{W}_{ho}$的梯度：

$$\frac{\partial L}{\partial \mathbf{W}_{hf}}=\sum_{t=1}^{T}\frac{\partial L}{\partial \mathbf{net}_{f,t}}\mathbf{H}_{t-1}^{trans},\ \frac{\partial L}{\partial \mathbf{W}_{hi}}=\sum_{t=1}^{T}\frac{\partial L}{\partial \mathbf{net}_{i,t}}\mathbf{H}_{t-1}^{trans}$$

$$\frac{\partial L}{\partial \mathbf{W}_{hc}}=\sum_{t=1}^{T}\frac{\partial L}{\partial \mathbf{net}_{c,t}}\mathbf{H}_{t-1}^{trans},\ \frac{\partial L}{\partial \mathbf{W}_{ho}}=\sum_{t=1}^{T}\frac{\partial L}{\partial \mathbf{net}_{o,t}}\mathbf{H}_{t-1}^{trans}$$

（4）将各时刻的梯度进行累加，得到$\mathbf{b}_f$、$\mathbf{b}_i$、$\mathbf{b}_c$和$\mathbf{b}_o$的梯度：

$$\frac{\partial L}{\partial \mathbf{b}_{f}}=\sum_{t=1}^{T}\frac{\partial L}{\partial \mathbf{net}_{f,t}},\ \frac{\partial L}{\partial \mathbf{b}_{i}}=\sum_{t=1}^{T}\frac{\partial L}{\partial \mathbf{net}_{i,t}}$$

$$\frac{\partial L}{\partial \mathbf{b}_{c}}=\sum_{t=1}^{T}\frac{\partial L}{\partial \mathbf{net}_{c,t}},\ \frac{\partial L}{\partial \mathbf{b}_{o}}=\sum_{t=1}^{T}\frac{\partial L}{\partial \mathbf{net}_{o,t}}$$

（5）计算$\mathbf{W}_{xf}$、$\mathbf{W}_{xi}$、$\mathbf{W}_{xc}$和$\mathbf{W}_{xo}$的梯度：

$$\frac{\partial L}{\partial \mathbf{W}_{xf}}=\sum_{t=1}^{T}\frac{\partial L}{\partial \mathbf{net}_{f,t}}\mathbf{X}_{t}^{trans},\ \frac{\partial L}{\partial \mathbf{W}_{xi}}=\sum_{t=1}^{T}\frac{\partial L}{\partial \mathbf{net}_{i,t}}\mathbf{X}_{t}^{trans}$$

$$\frac{\partial L}{\partial \mathbf{W}_{xc}}=\sum_{t=1}^{T}\frac{\partial L}{\partial \mathbf{net}_{c,t}}\mathbf{X}_{t}^{trans},\ \frac{\partial L}{\partial \mathbf{W}_{xo}}=\sum_{t=1}^{T}\frac{\partial L}{\partial \mathbf{net}_{o,t}}\mathbf{X}_{t}^{trans}$$

长短期记忆网络 (8)

- **模型训练**

（6）使用梯度下降法，给定学习率$\eta(0<\eta<1)$，以目标的负梯度方向调整参数，更新权重矩阵和偏置：

$$\mathbf{W}_{hf} \leftarrow \mathbf{W}_{hf} - \eta \frac{\partial L}{\partial \mathbf{W}_{hf}} \qquad \mathbf{W}_{xc} \leftarrow \mathbf{W}_{xc} - \eta \frac{\partial L}{\partial \mathbf{W}_{xc}}$$

$$\mathbf{W}_{hi} \leftarrow \mathbf{W}_{hi} - \eta \frac{\partial L}{\partial \mathbf{W}_{hi}} \qquad \mathbf{W}_{xo} \leftarrow \mathbf{W}_{xo} - \eta \frac{\partial L}{\partial \mathbf{W}_{xo}}$$

$$\mathbf{W}_{hc} \leftarrow \mathbf{W}_{hc} - \eta \frac{\partial L}{\partial \mathbf{W}_{hc}} \qquad \mathbf{b}_{f} \leftarrow \mathbf{b}_{f} - \eta \frac{\partial L}{\partial \mathbf{b}_{f}}$$

$$\mathbf{W}_{ho} \leftarrow \mathbf{W}_{ho} - \eta \frac{\partial L}{\partial \mathbf{W}_{ho}} \qquad \mathbf{b}_{i} \leftarrow \mathbf{b}_{i} - \eta \frac{\partial L}{\partial \mathbf{b}_{i}}$$

$$\mathbf{W}_{xf} \leftarrow \mathbf{W}_{xf} - \eta \frac{\partial L}{\partial \mathbf{W}_{xf}} \qquad \mathbf{b}_{c} \leftarrow \mathbf{b}_{c} - \eta \frac{\partial L}{\partial \mathbf{b}_{c}}$$

$$\mathbf{W}_{xi} \leftarrow \mathbf{W}_{xi} - \eta \frac{\partial L}{\partial \mathbf{W}_{xi}} \qquad \mathbf{b}_{o} \leftarrow \mathbf{b}_{o} - \eta \frac{\partial L}{\partial \mathbf{b}_{o}}$$

长短期记忆网络 (9)

◆ LSTM训练算法

随机初始化网络的权重参数W_{hf}、W_{hi}、W_{hc}、W_{ho}、W_{xf}、W_{xi}、W_{xc}、W_{xo}、b_f、b_i、b_c和b_o，使其均服从正态分布，初始化i=1

While $i \leq n_{LSTM}$ Do

For each $(X_t, y_t) \in D$ Do

由$F_t = \sigma(X_t W_{xf} + H_{t-1} W_{hf} + b_f)$计算时间步$t$的遗忘门$F_t$

由$I_t = \sigma(X_t W_{xi} + H_{t-1} W_{hi} + b_i)$ 计算时间步t的输入门I_t

由$C_t' = \tanh(X_t W_{xc} + H_{t-1} W_{hc} + b_c)$计算时间步$t$的候选记忆细胞$C_t'$

由$C_t = F_t \odot C_{t-1} + I_t \odot C_t'$计算时间步$t$的记忆细胞$C_t$

由$O_t = \sigma(X_t W_{xo} + H_{t-1} W_{ho} + b_o)$计算时间步$t$的输出$O_t$

由$H_t = O_t \odot \tanh(C_t)$计算时间步$t$的隐藏变量$H_t$

计算时间步t的损失函数$L(O_t, y_t)$与损失函数L

由$\delta_t = \frac{\partial L}{H_t}$计算时间步$t$的隐藏变量误差项$\delta_t$

计算时间步t的梯度$\partial L/\partial net_{f,t}$、$\partial L/\partial net_{i,t}$ 、 $\partial L/\partial net_{c,t}$ 和 $\partial L/\partial net_{o,t}$

End For

计算第i次迭代时权重矩阵W_{hf}、W_{hi}、W_{hc}、W_{ho}、W_{xf}、W_{xi}、W_{xc}、W_{xo}和偏置b_f、b_i、b_c、b_o的梯度

更新权重矩阵W_{hf}、W_{hi}、W_{hc}、W_{ho}、W_{xf}、W_{xi}、W_{xc} 和W_{xo}

更新偏置b_f、b_i、b_c和b_o

End While

Return $W_{hf}, W_{hi}, W_{hc}, W_{ho}, W_{xf}, W_{xi}, W_{xc}, W_{xo}, b_f, b_i, b_c, b_o$

时间复杂度：
$O(n_{LSTM} T d^2 h^5)$

长短期记忆网络 (10)

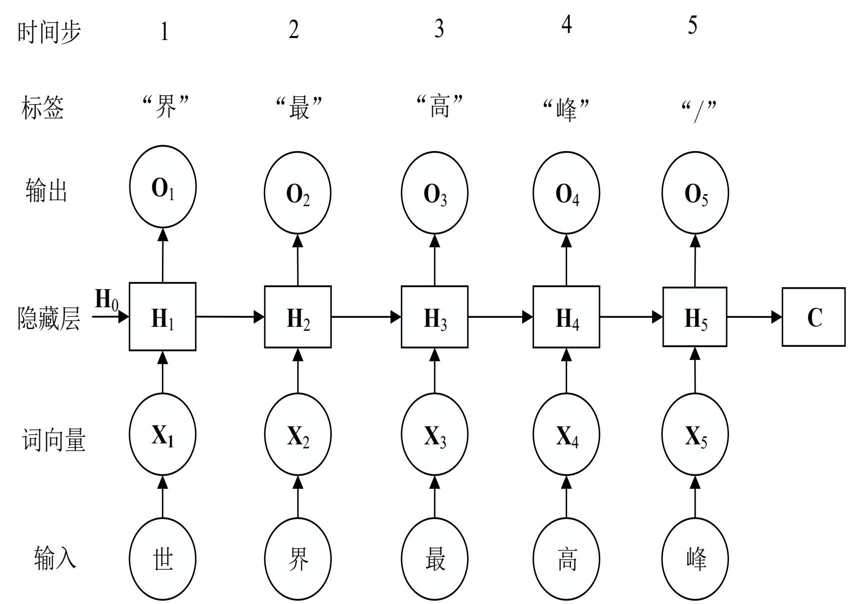
- 例：问答对q=“世界最高峰”、a=“珠穆朗玛峰”，LSTM提取文本序列特征的过程如下：

（1）转为词向量，记$\mathbf{q}' = (\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_5)$，$\mathbf{a}' = (\mathbf{Y}_1, \mathbf{Y}_2, \cdots, \mathbf{Y}_5)$。

（2）将$\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_5$输入LSTM，得到词典$E$中词的概率分布。

（3）计算背景向量$\mathbf{C} = \mathbf{H}_5$，包含序列“世，界，最，高，峰”的信息。

（4）重复（2）～（3）步，得到$\mathbf{C}'$，即“珠，穆，朗，玛，峰”的信息。



提纲

- 引例
- 问答系统的基本思想
- 循环神经网络
- 长短期记忆网络
- 基于LSTM的问答系统构建
- 总结

基于LSTM的问答系统构建 (1)

- **基于LSTM构建问答系统的总体思路**

（1）分析问题：链接知识库生成候选答案集。

（2）利用LSTM表征问题和答案。

（3）训练模型，使得问题与正确答案的表征向量相似度得分最高。

如何利用LSTM构建问答系统

基于LSTM的问答系统构建 (2)

- **基于LSTM的问答系统构建算法**

- 输入:

  $D_{qa}=\{(q_i,a_i)\}_{i=1}^{N}$：训练数据集，$q$：用户提出的问题，$G$：知识图谱

- 输出:

  a：问答系统的答案

- 步骤:

  1. 由$\arg\max_{\theta_1,\theta_2} \frac{1}{N}\sum_{i=1}^{N} log\left(\sum_{y\epsilon V(G)} P_{\theta_1}(y|q_i)P_{\theta_2}(a_i|y,q_i)\right)$学习参数$\theta_1$和$\theta_2$
  2. 由$P_{\theta_1}(y|q)=Softmax\left(W_y' M_q\right)=\frac{exp(W_y' M_q)}{\sum_{y'\epsilon V(G)} exp(W_{y'}' M_q)}$计算$G$中实体是$q$对应主题实体的概率，选取概率最大的实体$y_{tp}$为主题实体
  3. 对y_{tp}邻域2跳以内的所有实体$A_y=\{a_1,a_2,\cdots,a_m\}$进行拓扑排序，得到一个有序节点的子图，进而构造推理子图G_y
  4. 由$\left(G_{y\to a}\right)=\frac{1}{\#Parent(a)}\sum_{a_j\in Parant(a),(a_j,r,a)\ or\ (a,r,a_j)\in G_y} \tau\left(V\times\left[g\left(G_{y\to a}\right),\vec{e_r}\right]\right)$计算$y_{tp}$的推理子图$G_y$中任一实体的向量表示，记为$g\left(G_{y\to y_{tp}}\right)$
  5. $a\leftarrow argmax_a\left(P_{\theta_2}(a|y,q)\right)$ //计算G_y中的每个实体是答案的概率，选取概率最大时的实体作为答案
  6. Return a

基于LSTM的问答系统构建 (3)

- **基于LSTM构建问答系统的主要步骤**

（1）通过训练集$D_{qa}=\{(q_i,a_i)\}_{i=1}^{N}$，学习参数$\theta_1$和$\theta_2$：

$$\arg\max_{\theta_1,\theta_2}\frac{1}{N}\sum_{i=1}^{N}\log\left(\sum_{y\epsilon V(G)}P_{\theta_1}(y|q_i)P_{\theta_2}(a_i|y,q_i)\right)$$

（2）在已有知识图谱中寻找问题对应的主题实体：

① 利用LSTM学习问题q的表征，将q映射到d维实值向量空间$\mathbb{R}^d$，记为$\mathbf{M}_q\in\mathbb{R}^d$。

② 利用TransE算法将知识图谱G中的每个实体映射到低维向量空间，记为$\mathbf{M}_G$。

③ 使用以下方法计算G中实体作为问题q主题实体的概率：

$$P_{\theta_1}(y|q)=\mathrm{Softmax}\left(\mathbf{W}_y'\mathbf{M}_q\right)=\frac{\exp\left(\mathbf{W}_y'\mathbf{M}_q\right)}{\sum_{y'\epsilon V(G)}\exp\left(\mathbf{W}_{y'}'\mathbf{M}_q\right)}$$

④ 当$P_{\theta_1}(y|q)$概率最大时，则对应实体为G中对应主题实体，即$\mathrm{argmax}_y\left(P_{\theta_1}(y|q)\right)$。

基于LSTM的问答系统构建 (4)

- **基于LSTM构建问答系统的主要步骤**

（3）利用主题实体构造推理子图：

设y_{tp}为主题实体，对G中y_{tp}邻域内的所有实体$A_y = \{a_1, a_2, \cdots, a_m\}$进行拓扑排序，得到$G_y$，即为针对$y_{tp}$的推理子图。

（4）计算从主题实体对应的推理子图向量表示：

对于推理子图中的任一实体a，基于a的父节点向量表示来计算a的推理子图的向量表示：

$$\left(G_{y\to a}\right) = \frac{1}{\#Parent(a)} \sum_{a_j \in Parant(a),(a_j,r,a)\ or\ (a,r,a_j)\in G_y} \tau\left(\mathbf{V} \times \left[\mathbf{g}\left(G_{y\to a}\right), \overrightarrow{\mathbf{e}_r}\right]\right)$$

唯一递归返回条件：

$$y = a\text{时，}\ \mathbf{g}\left(G_{y\to a}\right) = \vec{\mathbf{0}}$$

基于LSTM的问答系统构建 (5)

- **基于LSTM构建问答系统的主要步骤**

（5）计算推理子图中每个实体作为答案的概率：

$$P_{\theta_2}(a|y,q) = \text{Softmax}\left(\mathbf{M}_q{}'\mathbf{g}(G_{y\to a})\right) = \frac{\exp\left(\mathbf{M}_q{}'\mathbf{g}(G_{y\to a})\right)}{\sum_{a'\in V(G_y)}\exp\left(\mathbf{M}_q{}'\mathbf{g}(G_{y\to a'})\right)}$$

当$P_{\theta_2}(a|y,q)$达到最大值时，对应实体a即为问题q的答案$\text{argmax}_a P_{\theta_2}(a|y,q)$。

基于LSTM的问答系统构建 (5)

- **基于LSTM的问答系统构建示例**

设长序列复杂问题q为“在遥远的东方有个国家叫做中国，美丽富饶、山川壮丽，其最高的山峰叫什么?”。执行过程：

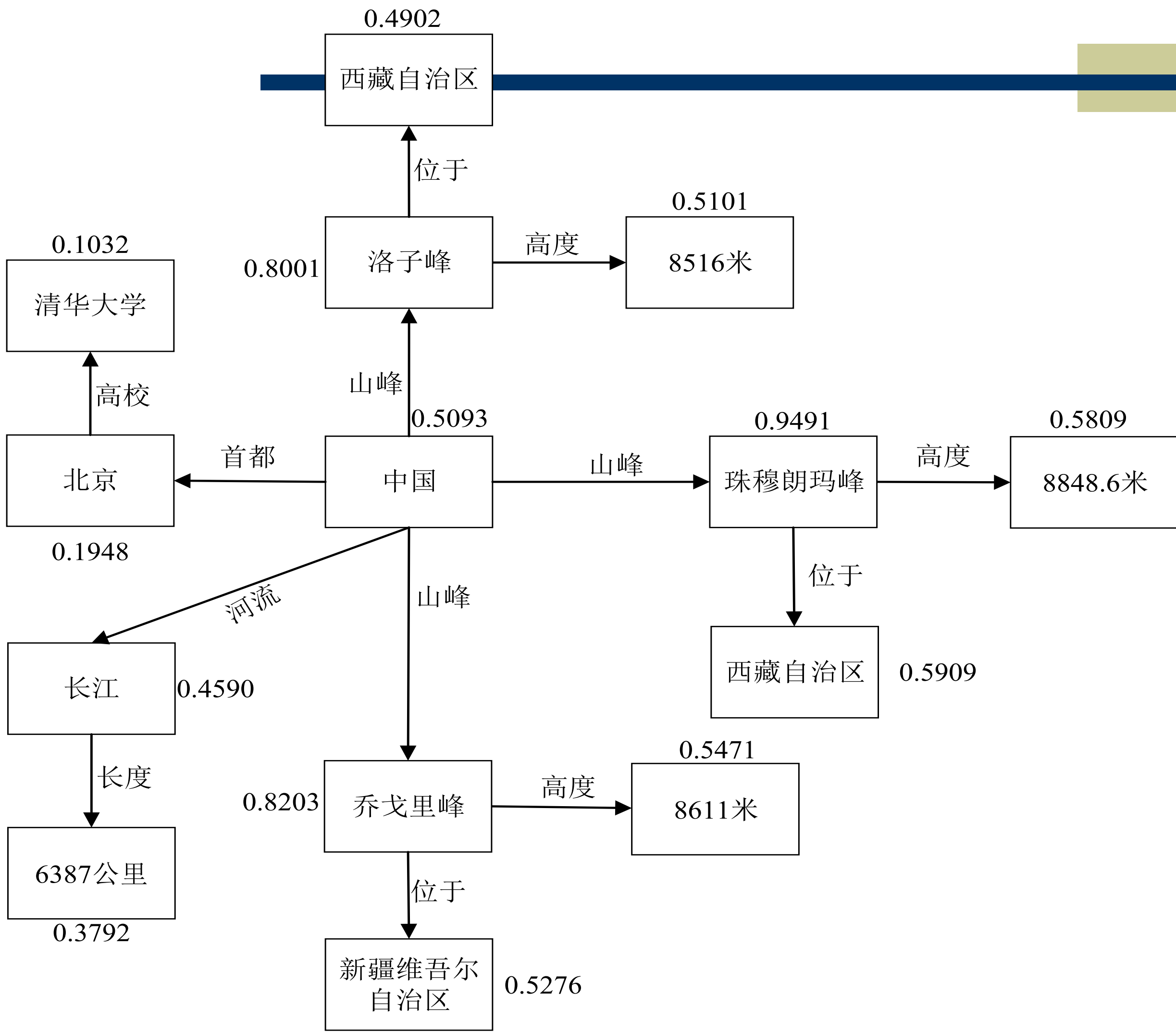
（1）使用LSTM对问题编码，得到问题的编码矩阵。

（2）计算G中每个实体为q的主题实体的概率。

若y表示“中国”时，$P_{\theta_1}(y|q)=0.9273$，且此时概率最大，则问题q的主题实体为“中国”。

（3）从主题实体“中国”开始，在G中选取邻域2跳以内的所有实体进行拓扑排序，得到如图所示的推理子图（不包含图中实体旁的数字）。

基于LSTM的问答系统构建 (6)



基于LSTM的问答系统构建 (7)

（4）由实体“中国”对应的推理子图中任一实体的向量表示，计算推理子图中的实体为最终答案的概率。

（5）计算推理子图中每个实体是问题q所对应答案的概率，最大值对应实体为问题的最终答案。

实体“珠穆朗玛峰”对应的$P_{\theta_2}(a|y,q)$最大，则问题最终答案为“珠穆朗玛峰”。

提纲

- 引例
- 问答系统的基本思想
- 循环神经网络
- 长短期记忆网络
- 基于LSTM的问答系统构建
- 总结

总结

- 问答系统任务的主要思想
- 问答系统任务的特殊性

  - 与检索系统性质的区别

  - 主要的输入数据源
- 面向问答系统的深度学习算法的特点

  - 深度学习在问答系统中起到的关键性作用
- 基于LSTM的问答系统构建

  - 基于LSTM构建问答系统的算法及主要步骤

  - 基于LSTM算法构建问答系统的优势

结语

谢谢！