• Sentiment classification 情感分析 • 机器翻译 视屏检测 • 人物对象检测 符号 • 实体识别问题 x^t • 时间序列 • T_x = 9, T_y= 9 • NLP natural language processing Representing words One-Hot vector 问题 • 输入数据和输出数据长度不一致 • 不能共享特征 ◦ 维度太大 (one-hot) **RNN** • 双向RNN 处理视屏 • 反向传播在RNN中 ● 逻辑回归,交叉熵 穿越时间反向传播 **Examples of RNN architectures** • 多对多结构 ● 多对一结构 • 一对一结构 • 一对多结构 音乐生成 • 多对多, 机器翻译, encoder to deconder Summary of RNN types $x^{<1>} x^{<2>}$ One to one Many to one One to many $\chi^{< T_{\chi}>}$ χ <1> χ <2> Many to many Many to many of many-to-many Andrew Ng 语言模型 使用RNN 构架语言模型 语音识别 概率 语音识别系统 如何构架语言模型 。 训练数据, 大量的英语文本 。 句子结尾 UNK token 不在常用的单词列表 ● 基于单词的语言模型(character-level language model) • 计算成本比较高 • 序列生成(Sequence generation) 梯度消失 vanishing gradients with RNNS ● 梯度爆炸 • exploding gradients, 梯度修剪 clipping GRU 门控制单元 ● 解决梯度消失 • memory cell 记忆细胞 • Tanh 激活函数 Full GRU $\widetilde{c}^{<t>} = \tanh(W_c[\lceil \times c^{<t-1>}, x^{<t>}] + b_c)$ $\Gamma_u = \sigma(W_u[c^{<t-1>}, x^{<t>}] + b_u)$ [= & (W, [c(+-1) x (+>)] + pr) $c^{<t>} = \Gamma_u * \tilde{c}^{<t>} + (1 - \Gamma_u) + c^{<t-1>}$ The cat, which ate already, was full. 然后我们为什么有了了 And why do we have gamma r? LSTM 长短时记 GRU and LSTM GRU LSTM $\underline{\tilde{c}^{< t>}} = \tanh(W_c[\underline{\Gamma_r} * \underline{c^{< t-1>}}, x^{< t>}] + b_c)$ $\Gamma_{u} = \sigma(W_{u}[c^{< t-1>}, x^{< t>}] + b_{u}) \qquad \text{where} \qquad \Gamma_{u} = \sigma(W_{u}[c^{< t-1>}, x^{< t>}] + b_{u})$ $\Gamma_{r} = \sigma(W_{r}[c^{< t-1>}, x^{< t>}] + b_{r}) \qquad \text{where} \qquad \Gamma_{g} = \sigma(W_{g}[c^{< t-1>}, x^{< t>}] + b_{g})$ $C^{< t>} = \Gamma_{u} \times \tilde{c}^{< t>} + (1 - \Gamma_{u}) \times c^{< t-1>} \text{where} \qquad \Gamma_{g} = \sigma(W_{g}[c^{< t-1>}, x^{< t>}] + b_{g})$ $C^{< t>} = \Gamma_{u} \times \tilde{c}^{< t>} + \Gamma_{g} \times C^{< t-1>}$ $C^{< t>} = \Gamma_{u} \times C^{< t>} + \Gamma_{g} \times C^{< t-1>}$ $C^{< t>} = \Gamma_{u} \times C^{< t>} + \Gamma_{g} \times C^{< t-1>}$ alto = [* clt> [Hochreiter & Schmidhuber 1997. Long short-term memory] Andrew Ng LSTM in pictures $\tilde{c}^{<t>} = \tanh(W_c[a^{<t-1>}, x^{<t>}] + b_c)$ $a^{< t>}$ $\Gamma_u = \sigma(W_u[\ a^{< t-1>}, x^{< t>}] + b_u)$ $\Gamma_f = \sigma(W_f[\ a^{< t-1>}, x^{< t>}] + b_f)$ tanh $\Gamma_o = \sigma(W_o[a^{< t-1>}, x^{< t>}] + b_o)$ $c^{< t>} = \Gamma_u * \tilde{c}^{< t>} + \Gamma_f * c^{< t-1>}$ output gate $a^{< t>} = \Gamma_o * \tanh c^{< t>}$ χ <t> LSTM in pictures $\tilde{c}^{< t >} = \tanh(W_c[a^{< t-1>}, x^{< t>}] + b_c)$ $\Gamma_u = \sigma(W_u[a^{< t-1>}, x^{< t>}] + b_u)$ $\Gamma_f = \sigma(W_f[\underline{a}^{< t-1>}, x^{< t>}] + b_f)$ $\Gamma_o = \sigma(W_o[a^{< t-1>}, x^{< t>}] + b_o)$ $c^{< t>} = \Gamma_u * \tilde{c}^{< t>} + \Gamma_f * c^{< t-1>}$ $a^{< t>} = \Gamma_o * \tanh c^{< t>}$ $x^{< t>}$ c<2> a<3> a<1> ewellot eas op medt bled fæyt bans Andrew Ng • peephole connection 偷窥孔洞 • 什么时候使用GRU 或是 LSTM 没有统一标准。 双向RNN bi-directional RNN BRNN 。 需要完整的语句 **Deep RNN** Deep RN example **NLP and Word Embeddings** ● 词嵌入 Word representation ○ one-hot representation, 每个词独立,内积为0 特征化 Featurized representation: word embedding 。 单词的泛化 Visualizing word embedding, t-SNE algorithm ● mapp to 相似的特征, 高维度到低维度的可视化 叫做 Word embedding **Use Word embedding** Name entity recognition example Transfer learning and word embeddings relation to face encoding fixed encoding 词嵌入的属性 • Analogies 类比推理 • Analogies using word vectors • Similarity function • 余弦相似度 Cosine similarity **Embedding matrix** embedding layer • Neural language model 获取多个300维度的向量,然后作为输入 Other context/target pairs Skim-grams model • 上下文 Context Target Problems with soft-max classification 分级的soft-max 分类器 negative sampling 负采样 Word2vec skip-grams word bag Defining a new learning problem • 10000 个二分类器 Glove word vectors global vectors for word representation stop words • 情感分类 Sentiment classification The problem of bias in word embeddings 减少词中的偏见问题 Addressing bias in word embeddings

深度学习 - 序列模型

• 语音识别

• 音乐生成

● DNA序列分析

2. Improving Deep Neural Networks: Hyperparameter tuning, Regularization and Optimization

Specialization outline

Identify bias direction

○ Neutralize: for every word that is not definitional, project to get rid of bias. 性别中立

。 奇异值分解

Equalize pairs.

• RNN GRU LSTM

• Image captioning 图片描述

• CNN -> RNN

• Greedy search algorithm

• Beam search algorithm

• 启发式搜索, 误差分析

• 计算错误比例

• Bleu score (optional)

Computing attention

Speech recognition

problems

end to end

总结

• Encoder, Decoder

Picking the most likely sentence

◦ Language model 语言模型

• Refinements to beam search 改进

Evaluating machine translation

Attention model intuition 注意力模型

• Bleu score on bigrams

The problem of long sequences

• 300 hours, 3000, 10,000 hours

• CTC cost for speech recogintion

• Attention model for speech recognition

• Audio Data 关键字 Trigger word detection

beam number id 10

• Length normalization 长度归一化, 概率值太小了,数值下溢出

○ Machine translation 翻译模型

• PCA

Basic model

Seq2Seq

4. Convolutional Neural Networks5. Sequence Models

1. Neural Networks and Deep Learning

3. Structuring Machine Learning Projects

dels