

基于 RNN-RBM 语言模型的语音识别研究

黎亚雄¹ 张坚强² 潘登³ 胡憺⁴

¹(湖北科技学院网络管理中心 湖北咸宁 437100)

²(弗吉尼亚理工大学信息技术中心 美国黑堡 VA24061)

³(湖北科技学院外国语学院 湖北咸宁 437100)

⁴(中南财经政法大学外国语学院 武汉 430073)

(yaxiong_li@live.cn)

A Study of Speech Recognition Based on RNN-RBM Language Model

Li Yaxiong¹, Zhang Jianqiang², Pan Deng³, and Hu Dan⁴

¹(Network Management Center, Hubei University of Science and Technology, Xianning, Hubei 437100)

²(Information Technology Department of Learning Sciences & Technologies Virginia Polytechnic and State University, Blacksburg, USA VA24061)

³(School of Foreign Languages, Hubei University of Science and Technology, Xianning, Hubei 437100)

⁴(School of Foreign Languages, Zhongnan University of Economics and Law, Wuhan 430073)

Abstract In the recent years, deep learning is emerging as a new way of multilayer neural networks and back propagation training. Its application in the field of language model, such as restricted Boltzmann machine language model, gets good results. This language model based on neural network can assess the probability of the next word appears according to the word sequence which is mapped to a continuous space. This language model can solve the problem of sparse data. Besides, some scholars are constructing language model making use of recurrent neural network mode in order to make full use of the preceding text to predict the next words. From these models we can sort out the restriction of long-distance dependency in language. This paper attempts to catch the long-distance information based on RNN-RBM. On the other hand, the dynamic adjunction of language model is analyzed and illustrated according to the language features. The experimental result manifests there are considerable improvement to the efficiency of expanding vocabulary continuing speech recognition using RNN-RBM language model.

Key words speech recognition; language model; neural network; recurrent neural network-restricted Boltzmann machine; relevance information

摘要 近年来深度学习兴起,其在语言模型领域有着不错的成效,如受限玻尔兹曼机(restricted Boltzmann machine, RBM)语言模型等.不同于 N -gram 语言模型,这些根植于神经网络的语言模型可以将词序列映射到连续空间来评估下一词出现的概率,以解决数据稀疏的问题.此外,也有学者使用递归神经网络来建构语言模型,期望由递归的方式充分利用所有上文信息来预测下一词,进而有效处理长距离语言约束.根据递归受限玻尔兹曼机神经网络(recurrent neural network-restricted Boltzmann machine, RNN-RBM)的基础来捕捉长距离信息;另外,也探讨了根据语言中语句的特性来动态地调整语言模型.实验结果显示,使用 RNN-RBM 语言模型对于大词汇连续语音识别的效能有相当程度的提升.

收稿日期:2014-03-11;修回日期:2014-07-07

关键词 语音识别;语言模型;神经网络;递归神经网络-受限玻尔兹曼机;关联信息

中图法分类号 TP183

语音是人类沟通的基本媒介,也是人们表达彼此情感的基本路径.在日常说话中人们通过言语和语调来了解对方的想法,听懂对方所传递的语音信息.人类所拥有的这种天赋引起了许多学者的研究兴趣,语音识别也就显示了它的研究地位.在语音识别的过程中,我们通过特征提取来处理语音信号,得到代表该信号的特征参数,然后转换成语音特征向量.另一方面使用训练语音语料和与之相对应的文本语料分别建构声学模型和语言模型,以此来表示语音与文字之间的对应关系以及语言中词汇的出现状况.最后根据词典、语言模型、声学模型和语音特征向量所提供的信息进行译码,从而得到最终的识别结果.本文我们所关注的是语音识别过程中的语言模型,希望通过语言模型来捕获语言中的规律.

经典的 N -gram 模型是较常见的语言模型之一,其易于实现的特性引发诸多研究者的研究与使用.但是 N -gram 模型有着 2 个几乎不可逾越的鸿沟:数据稀疏与缺乏长距离语言约束.自从 20 世纪 90 年代以来有多种不同类型的非线性语言模型被挖掘出来并期望解决这些问题.为了获得长距离信息,递归神经网络语言模型被发掘出来,但 Bengio 等学者^[1]指出,递归神经网络较难取得更长距离的信息,其理由是随着长度的增加,远距离信息通过概率相乘的结果趋向于零. BP 神经网络语言模型^[2]也是其中之一,它将长距离词序列的信息投影到连续空间,借以解决数据稀疏的问题,但对于长距离信息的取得仍不理想.本文延续先前对于神经网络语言模型的研究,尝试使用递归受限玻尔兹曼神经网络(recurrent neural network-restricted Boltzmann machine, RNN-RBM)语言模型来改进预测能力,希望在大词汇连续语音识别过程中有一定程度的改善.

1 RNN-RBM 语言模型及其应用

1.1 RNN-RBM 模型简介

RNN-RBM 模型^[3]由 Bengio 等学者提出以来,被用于生成复调音乐.它由递归神经网络(recurrent neural network, RNN)和一个单层的受限玻尔兹曼机(restricted Boltzmann machine, RBM)网络构

成, RBM 网络承担生成模型的功能(在文献[3]主要作为音乐生成的功能), RNN 网络承担判别模型的功能,其输出的值被作为提取的特征.

RNN-RBM 模型结构如图 1 所示^[3]:

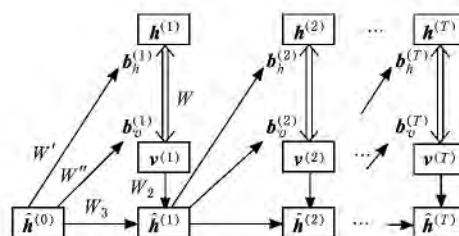


Fig. 1 RNN-RBM model.

图 1 RNN-RBM 模型

模型中 $h^{(T)}$, $v^{(T)}$ 和 $\hat{h}^{(T)}$ ($T=1, \dots, n$) 分别为 RBM 隐藏层、RBM 可见层和 RNN 网络中包含的神经元个数; $b_h^{(T)}$ ($T=1, \dots, n$) 为 RBM 隐藏层的偏置向量; $b_v^{(T)}$ ($T=1, \dots, n$) 为 RBM 可见层的偏置向量; $h^{(T)}$ ($T=1, \dots, n$) 为 RBM 隐藏层的状态向量; $v^{(T)}$ ($T=1, \dots, n$) 为 RBM 可见层的状态向量; $w = (w_{i,j}) \in \mathbb{R}^{h^{(T)} \times v^{(T)}}$ 为 RBM 隐藏层和可见层之间的权重矩阵; $w_2 = (w_{i,j}) \in \mathbb{R}^{h^{(T)} \times v^{(T)}}$ 为 RBM 可见层与 RNN 网络中相对应神经元之间的权重矩阵; w_3 为 RNN 网络中神经元之间的权重向量; $w' = (w_{i,j}) \in \mathbb{R}^{h^{(T-1)} \times h^{(T)}}$ 为 RBM 隐藏层与 RNN 网络神经元之间的权重矩阵; $w'' = (w_{i,j}) \in \mathbb{R}^{h^{(T-1)} \times v^{(T)}}$ 为 RBM 可见层与 RNN 网络神经元之间的权重矩阵.

对于上述参数的求解采用随机梯度下降法(SGD)^[4], RBM 部分的 Gibbs 采样采用 CD-k 算法^[5].

1.2 基于 RNN-RBM 的语言模型

1.2.1 神经网络语言模型介绍

神经网络语言模型^[6-7]的主要架构有输入层、隐藏层和输出层;有时还会在前两层之间加入一个投影层(projection layer),使长距离词序列投影至此连续空间.这样做的目的是在降低输入层维度的同时也使得词的表示更加柔和.与 N -gram 模型(产生数据稀疏问题)不同的是,输入的词序列中每个词都能贡献权重来估测下一个词出现的可能性.层与层之间的神经元靠着突触(synapse)来传递信息,各层以向量来表示,层之间的突触以矩阵表示.

输入层输入的是需要预测词的上文信息,其中每个词以 one-of- N 方式进行编码,如果词 d 是词典中的第 i 个词,在长度为 N 的向量中,词 d 只有第 i 维是 1 其余为 0. 例如:

“苹果”表示为 $[0\ 0\ 0\ 1\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ \dots]$;

“手机”表示为 $[0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 1\ 0\ 0\ \dots]$.

如果我们用 $C(d_i)$ 来表示 d_i 所对应的向量,输入层就是将 $C(d_{i-n+1}), \dots, C(d_2), C(d_1)$ 这 $n-1$ 个向量首尾相连,形成一个 $(n-1) \times N'$ (N' 为词典维

数)的向量矩阵. 隐藏层如同普通的神经网络,通过权重和偏置直接计算. 再利用 Sigmoid 作为激活函数. 输出层中的每个节点表示下一词的未归一化 log 概率,利用 Softmax 激活函数将输出值归一化成概率. 最后,输出层的结果可视为一个 N 维的向量,其第 K 维的意义是在上文词序列发生的情况下,目前预测的词 d_i 发生的概率在数学上表示为 $p(d_i=k|(d_{i-n+1}, \dots, d_2, d_1))$. 其网络机构和过程如图 2 所示:

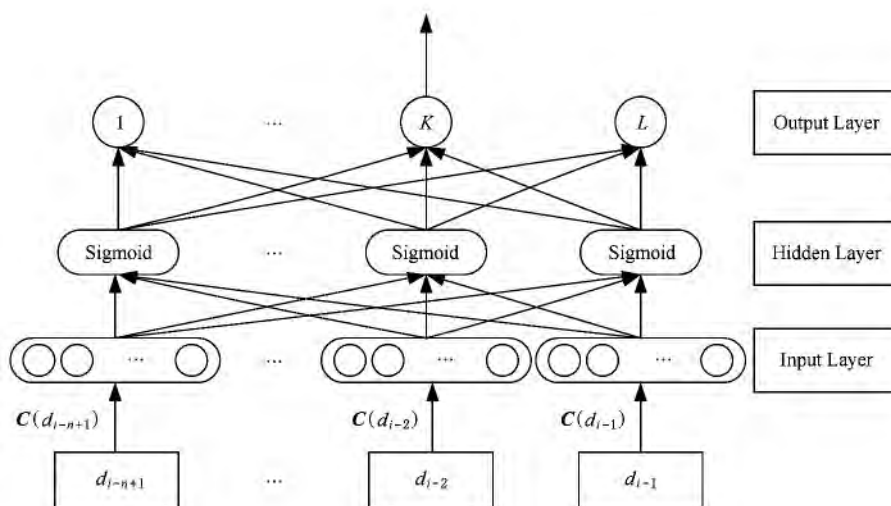


Fig. 2 Language model based on neural network.

图 2 神经网络语言模型

1.2.2 RNN-RBM 语言模型

与神经网络语言模型所不同的是, RNN-RBM 语言模型减去了投影层,除了增加一个前次递归隐藏层之外,还用 RNN-RBM 模型替代神经网络语言模型中输入层和隐藏层. 训练过程中的差别是,输入

层是一次以一个词来表示并训练,词的表示方法则与传统神经网络语言模型相同. 其网络的结构是把输入层加大,将上一时间点的隐藏层预先储存起来,若以时间方式来阶层展开,将会更清楚地看出其递归的概念,如图 3 所示:

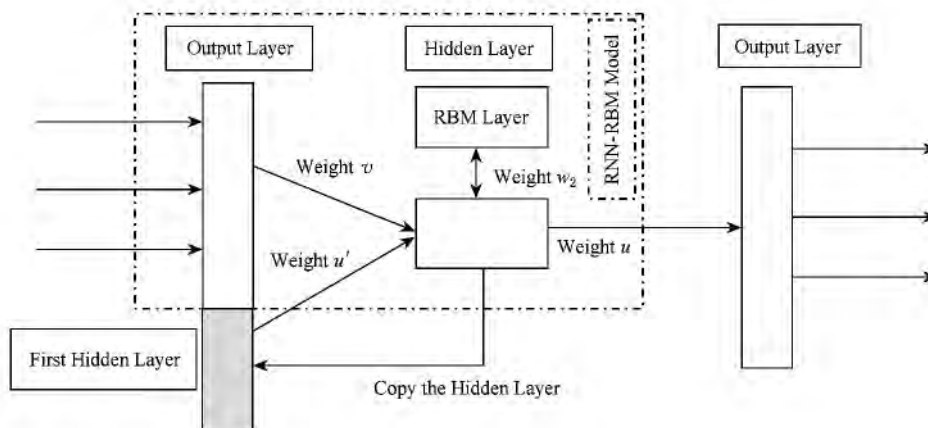


Fig. 3 RNN-RBM language model.

图 3 RNN-RBM 语言模型示意图

但 Bengio 等学者^[8]发现,在递归神经网络中利用梯度下降法(gradient descent method)求解,对于学习长距离的信息十分困难.要获得长距离信息必须要具有任意时间序列分析并且拥有抵抗其他信息干扰的能力.因为随着时间序列的变化,距离较远的信息会被每一次递归所输入的信息干扰,因此我们还需要对该模型作进一步改进.

1.3 改进 RNN-RBM 语言模型

1.3.1 将词的关联信息融入 RNN-RBM 语言模型

神经网络语言模型能有效解决数据稀疏问题,但是在处理长距离词语信息时仍稍嫌不足;因此,RNN-RBM 语言模型希望能取得更多长距离词语信息.多种研究也表明递归神经网络语言模型在处理长距离词语信息时的确能比一般神经网络语言模型带来更好的成效.这也是本文使用采取类似递归神经网络语言模型来探讨的原因.递归神经网络语言模型中的回馈方式是使用时序性倒传递算法;然而,该方法同样被证明这类模型仍然无法有效获得长距离文本信息.因此,我们探讨通过加入词的关联信息(word relevance information)来帮助预测下一个词出现的可能性.

关联信息以向量的方式来表示,向量的大小同上文中的输入层一样.因此本文将输入层扩增为两倍,前半段为原来训练数据的关联向量,后半段为与之对应训练数据的关联向量.我们将关联向量定义为两种:一种是句子间的关联,称为语句关联信息,另一种是词与词之间的关联,称为词关联信息.语句关联信息的产生,是将需要检索的句子放进训练语料中进行一次检索,检索完成后便可知对所有训练语句的关联系数,根据关联系数可以决定要使用多少关联语句来当作关联向量.现假设存在句子 S_i , S_i 中依序包含了词 d_1, d_2, d_3, d_4, d_5 (其量分别为 10000, 00100, 01000, 00010, 00001) 则可以定义 $R_1(d_1, d_3, d_5)$ 为对应 S_i 的语句关联信息.在训练时仍是以词为单位进行训练,赋予每个词的关联信息皆为此句的语句关联信息.因此 $R_{d_1}, R_{d_2}, R_{d_3}$ 的向量相同为 11001. 词 d_1, d_2, d_3 所对应向量分别为 1000011001, 0010011001, 0100011001.

词的关联信息则为了避免使每个词对应到相同的关联信息,其产生的方式是从训练语料中收集每一个词左右相邻文段中相隔一定距离内其他词出现的频率,结果会得知数个关联的词,每一个词的关联信息长度也皆不同.

本文将词的关联信息分为 3 种,分别是词频、正则化以及自定义(1 或 0)来进行讨论,目的是为了观察关联信息带来的影响.词频表示我们使用实际的次数来增加词与词之间的关联性.正则化表示时所有关联信息的总和为 1,贡献大即次数越多其值越高;反之贡献小则其值较低.自定义表示是将有次数出现的维度以 1 来表示,没有出现为 0.如果文本语料中每个词所对应的关联向量是固定的,我们将上文信息重新组合得到新的关联信息.因为每个词的上文信息大部分皆不同,所以组合出来的关联向量也皆不同.根据上文信息的远近,分别使用不同权重来结合,越远的上文信息则随着时间越来越小,我们用式(1)来表示:

$$\begin{cases} R_i = R_0, & \text{if } i=0; \\ R_i = (1-\alpha)R_{i-1} + \alpha R_i, & \text{if } 0 < i \leq L; \end{cases} \quad (1)$$

其中, R_0 表示词语 i 的原始关联信息; R_i 为新获得的关联信息; L 为在该语句所含词的数目; α 则为可调控之参数.本文我们取 α 值为 0.5.假设句子 S_i 中含有 4 个词如图 4 所示,图 4 中 d_1, \dots, d_4 为词的关联信息.可看出其关联信息为所有上文中词的权重之和,权重部份则取决于词的距离.

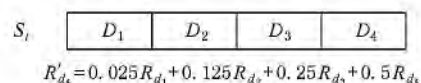


Fig. 4 Examples of dynamic word relevance information.

图 4 动态词关联信息实例

1.3.2 将语句的关联信息融入 RNN-RBM 语言模型

本节讨论利用动态语言模型调整的方式来增加语言模型的预测能力,对于不同测试语句使用不同的 RNN-RBM 语言模型或是结合不同族群的 RNN-RBM 网络语言模型.不同于神经网络语言模型所提出的方法,所有测试语句都使用由相同训练语料训练出的语言模型,我们希望针对不同的测试语句,以线性组合的方式结合不同训练语料来进行训练,期望找到适合测试语句本身的 RNN-RBM 语言模型.

首先将与语音语料相对应的文本语料以单字向量(unigram word vector)表示,用 K-means 进行聚类.然后用聚类后的结果来训练语句所对应的 RNN-RBM 语言模型.假设所有训练语句可分为 S 类,针对 S 类中每个类分别算出各自的特征权重向量再求平均值,即平均向量,此部分可由每一群中各训练语句的单字向量进行相加并取平均求得.

当有语句需要进行测试时,可以利用此语句的单字特征向量和已经计算好的各类特征权重向量来求取其相似度,同时选取欲使用的 RNN-RBM 语言模型或结合不同的族 RNN-RBM 语言模型. 由于我们无法事先得知语句(语音)正确转写的语句(文本). 因此,语句的单字特征向量是从多条识别结果中最佳的来表示.

在计算相似度时,使用广义 Dice 系数来计算:

$$\cos(U_k, v_s) = \frac{2 \cdot u_k \cdot v_s}{u_k^2 + v_s^2}, \quad (2)$$

其中, U_k 表示第 k 句测试语句; u_k 表示测试语句中第 k 句多条识别结果中最佳的单连词向量; v_s 表示用第 S 类中训练语句的单字特征向量. 本文主要采用 2 种选取方法:

方法 1. 最大权重相似度. 只选取和测试语句最相似的训练语句群:

$$RNN-RBM_{U_k} = \arg \max_s \text{sim}(U_k, v_s); \quad (3)$$

$$p(U_k) \approx RNN-RBM_{U_k}(U_k). \quad (4)$$

方法 2. 线性组合相似度. 组合系数由计算测试语句与 S 群中训练语句间的相似度来获得, 如果语句和某一类相似度较大, 则表示该类比较符合该测试语句的特性. 语句与各训练语句群的综合系数为

$$\xi_{s,k} = \frac{\text{sim}(u_k, v_s)}{\sum_{s'=1}^s \text{sim}(u_k, v_{s'})}, \quad (5)$$

$$p(U_k) = \sum_{s=1}^S \xi_{s,k} \times p_{RNN-RBM_s}(U_k), \quad (6)$$

式(6)中, $p(U_k)$ 表示测试语句经过线性结合后 RNN-RBM 语言模型概率; $p_{RNN-RBM_s}(U_k)$ 表示第 S 类 RNN-RBM 语言模型的概率. 我们根据分数结合 RNN-RBM 语言模型所预测的概率, 来得到最终的结果.

2 实验结果与分析

2.1 实验语料

本文使用的实验语料来自 2011~2013 年湖北省咸宁市普通话等级考试中朗读部分^①. 我们在考试成绩为一级甲等、一级乙等、二级甲等和二级乙等水平中选取 46.7 s 的语音语料及与之相对的文本标记语料 1489 个句子. 然后将这些语料随机分为训练语料集、验证语料集和测试语料集, 结果如表 1 所示:

Table 1 Number of Sentences and Time Statistics on Experimental Corpus

表 1 实验语料句数与时间长度统计

Corpus	Number of Sentences	Time/s
Train Set	962	33
Validation Set	266	7
Test Set	260	6

2.2 RNN-RBM 语言模型基础实验结果

从文本语料中挑选 100 句 ($M=100$) 最佳候选词序列, 经过训练好的 RNN-RBM 语言模型得到各句词序列的分数, 并将其结果加入声学模型的信息以此获得语音识别的总分数, 再选出排名最高的词序列与文本语料中与之对齐的句子, 计算出相似度从而得到字正确率, 最后将本文所提出的关联信息融入 RNN-RBM 语言模型. 目的是希望再次通过 RNN-RBM 语言模型来重新排序, 并找出字正确率最高的词序列.

在对语言模型进行训练及语音识别过程中, 句子和句子之间是彼此独立的, 即上一个语句与目前正在训练的句子是不相关联的. 实验结果如表 2 至表 5 所示:

Table 2 Experimental Results of RNN-RBM Language Model

表 2 RNN-RBM 语言模型的基础实验结果^②

Experimental Approaches	Language Complexity of Train Set	Language Complexity of Test Set	Word Correct Rate of Train Set	Word Correct Rate of Test Set	Absolute Increase Rate	Relative Increase Rate
3-gram	461.82	472.67	0.8523	0.8455	—	—
RNN-RBM	610.02	631.40	0.8306	0.8314	-0.141	-0.847
Oracle(RNN-RBM)	—	—	0.9433	0.9354	—	—

① 语音语料来源: 湖北科技学院资格鉴定所, 为 2011 年至 2013 年湖北省咸宁市普通话等级考试中朗读部分的录音, 与之相对应的文本语料来自: 湖北省普通话水平测试专用教材(中国和平音像出版社)中朗读部分 1~50 号文章.

② 表 2 中, Oracle(RNN-RBM) 是文本语料库中挑选的最佳候选词序列(语句). 其语言复杂度(Language Complexity)和提升率(Increase Rate)不用计算. 3-gram 识别的提升率(Increase Rate)不是本文所关注的对象.

表 2 中语言复杂度我们仿照 Mikolov 等学者^[9]计算而得,从语言复杂度的角度来看,RNN-RBM 语言模型的语言复杂度为 610.02.从 3-gram 的实验结果可以看出,虽然语言复杂度确实低,但其效果明显比 RNN-RBM 语言模型要好得多,Oracle (RNN-RBM)的实验结果字正确率可到达 94.33%,这意味着我们有很大的改进空间.

继续使用语句关联信息来帮助 RNN-RBM 语

言模型的改进.在训练模型时,语句关联信息挑选最相关的训练语句.但如果挑选过多的关联信息反而有可能会降低识别率,因此我们只挑选关联最紧密的部分.因为 RNN-RBM 语言模型是以词为单位进行训练,所以每个词需要对应到一个关联向量,词的关联向量为该词的关联向量.而关联向量的表示方式分别使用了句子中词出现的频率、正则化和自定义值.得到的结果如表 3 所示:

Table 3 Experimental Results on Model with Sentence Relevance Information

表 3 语句关联信息融入模型中的实验结果

Sentence Relevance Information	Language Complexity of Train Set	Language Complexity of Test Set	Word Correct Rate of Train Set	Word Correct Rate of Test Set	Absolute Increase Rate	Relative Increase Rate
Fundamental Recognition Rate of RNN-RBM Model	610.02	631.40	0.8306	0.8314	—	—
Word Frequency	601.22	603.19	0.8217	0.8306	-0.0008	-0.0047
Regularization	608.46	631.27	0.8311	0.8318	0.0004	0.0028

从表 3 我们可以看出,语言复杂度部分以正则化的实验结果较好.其余 2 种识别结果则比 RNN 基础识别率低.探究其原因应该是每一语句内,我们分析为词的关联向量皆为此句的关联向量,因此关联向量的重复率与句中词的频率有很大关系,成为识别率降低的原因之一.还有一个原因是原始文本所输入的信息有可能被关联信号所干扰.因此我们尝试将关联信息切得更细致,使用词的关联信息来进行改进.

词关联信息的生成是在训练语料中词出现的地方找出其相邻的词作为关联信息,本文则选取左右距离为 3.

表 4 为词的关联信息融入 RNN-RBM 语言模型的实验结果,由于部分词的关联信息太多,包含关联词和非关联的词,因此,我们试着去调整词关联信息的使用程度,其中词关联信息的长度是根据训练集语料最好的结果来设定.在语言复杂度方面,词频、正则化和自定义方面也都有进步;而字正确率方面,词频、正则化以及自定义的识别结果均有提升.从这 3 种识别率中我们看到自定义的效果最佳,因为此方法对于每个关联词的关联度较公平,皆设定为 1.而词频、正则化,因为每个关联词之间的歧异度较高,尤其是词频的差距很大,导致有些关联词的贡献被埋没.

Table 4 Experimental Results on Model with Word Relevance Information

表 4 词的关联信息融入模型中的实验结果

Word Relevance Information	Language Complexity of Train Set	Language Complexity of Test Set	Word Correct Rate of Train Set	Word Correct Rate of Test Set	Absolute Increase Rate	Relative Increase Rate
Fundamental Recognition Rate of RNN-RBM Model	610.02	631.40	0.8306	0.8314	—	—
Word Frequency	609.05	629.93	0.8388	0.8333	0.0019	0.0114
Regularization	609.13	629.75	0.8383	0.8331	0.0017	0.0102

这部分实验结果也显示了当自定义值设为 1 且调整词关联信息的长度较好;于是我们进一步观察使用这种表示法在不同词关联信息长度上的结果.虽然从实验结果得知使用词的关联信息的确能够提

升识别率,但我们仍希望突破目前的瓶颈.我们发现虽然词的关联信息解决了语句关联信息的问题,但是仍有类似语句关联信息的缺点存在,语料中每个词所对应的词关联信息一样,造成在训练中重复使

用同样的词关联信息. 我们希望在训练 RNN-RBM 语言模型时得到独特的词关联信息. 由语言学知识可知相同的词在不同的句子代表不同的意思, 所以我们希望用这种特点来获得上文或句子中的信息, 预测下一词时能够更加符合句子中的意思. 因此, 我们希望利用动态词关联信息来进一步改善语言模型. 其计算方式如图 4 所示.

表 5 显示了动态调整词的关联信息效果比一般的词关联信息较差一点. 寻找其原因, 应该是在词关联信息中常包含关联和非关联的信息, 因此我们难以准确知道越近距离的词关联信息有较相关的信

息, 造成使用的关联信息无法正确代表与该词的关联信息.

本文提出的在 RNN-RBM 语言模型中融入关联信息的确有助于提升识别率, 但是其效果不是很明显, 我们分析其原因有 3 种: 一是关联信息可能会对输入层所要传递的信息造成干扰, 使得输入层所要传递的信息减弱, 甚至有可能将关联信息当成主要传递的信息; 二是关联信息也可能只是将其表示方式作了延伸, 而关联信息的表示可能有更佳的表现方法; 三则是难以准确地决定关联信息, 导致效果无法彰显.

Table 5 Experimental Results on Model with Dynamic Word Relevance Information

表 5 动态调整词的关联信息并融入模型中的实验结果

Dynamic Relevance Information	Language Complexity of Train Set	Language Complexity of Test Set	Word Correct Rate of Train Set	Word Correct Rate of Test Set	Absolute Increase Rate	Relative Increase Rate
Fundamental Recognition Rate of RNN-RBM Model	610.02	631.40	0.8306	0.8314	—	—
Word Frequency	607.72	629.35	0.8323	0.8323	0.0009	0.0054
Regularization	609.42	631.19	0.8322	0.8331	0.0017	0.0112

2.3 语句的关联信息融入 RNN-RBM 的实验结果

本节的目的是讨论将语句的关联信息融入 RNN-RBM 语言模型中的实验结果. 表 6 将训练集语料分为 2 类用于测试集语料的结果. 表 7 将测试集语料的最佳设定用在测试语料的识别结果; 表 8 与表 9 分别为将训练集语料分为 4 类的验证集语料

与测试语料的识别结果.

表 7 将训练语料分成 2 类后, 分别使用最大权重和线性组合相似度测量的识别结果. 比较表 6 我们可以看出, 利用最大权重法有较好的结果, 与基础递归式类神经网络语言模型相比, 提升率有所提升.

Table 6 Recognition Results on Validation Corpus with a Two-category Training Corpus

表 6 语料分 2 类时验证语料字的识别结果

RNN-RBM Weight	Maximum Weight-Similarity	Linear Combination-Similarity
0	0.8329	0.8329
0.2	0.8487	0.8488
0.4	0.8481	0.8492
0.6	0.8470	0.8457
0.8	0.8405	0.8397
1	0.8156	0.8162

Table 7 Recognition Results on Test Corpus with a Two-category Training Corpus

表 7 语料分 2 类时测试语料字的识别结果

Calculation Method	Word Correct Rate	Absolute Increase Rate	Relative Increase Rate
Maximum Weight-Similarity	0.8441	0.0023	0.0155
Linear Combination-Similarity	0.8424	0.0006	0.0048

Table 8 Recognition Results on Validation Corpus with a Four-category Training Corpus
表 8 语料分 4 类时验证语料字的识别结果

RNN-RBM Weight	Maximum Weight-Similarity	Linear Combination-Similarity
0	0.832 9	0.832 9
0.2	0.847 3	0.847 3
0.4	0.848 6	0.848 6
0.6	0.846 1	0.845 0
0.8	0.841 5	0.841 3
1	0.817 5	0.823 9

线性组合相似度略差于 RNN-RBM 基础识别率,原因应该是数据较偏向某一类,因此线性组合相似度则较差,被偏向的类就能有较好的效果。

表 9 将训练集语料分 4 类来进行实验,从该表来看最大权重相似度是较好的。但和 2 类的实验相比,虽然与 RNN-RBM 基础识别率相比有所进步,

但其结果仍然不理想。分析其原因应该是训练语料不足和我们所用的实验硬件环境受到限制有关。由于分类数目增多,每类中的训练语料则随之减少(因实验硬件受到限制训练的时间太长)。因此无法训练出学习能力较好的 RNN-RBM 网络语言模型,导致字正确率的下降。

Table 9 Recognition Results on Test Corpus with a Four-category Training Corpus
表 9 语料分 4 类时测试语料字的识别结果

Calculation Method	Word Correct Rate	Absolute Increase Rate	Relative Increase Rate
Maximum Weight-Similarity	0.843 3	0.001 5	0.010 5
Linear Combination-Similarity	0.843 1	0.001 3	0.009 2

3 结论与展望

参 考 文 献

语言模型,希望通过动态调整语言模型的方法来使其拥有更好的识别能力。随着分类的增加(如分成 4 类),RNN-RBM 语言模型的识别能力有一定程度上的增加,可惜的是由于训练语料的不足和实验硬件环境的限制,我们没有训练出学习能力理想的 RNN-RBM 语言模型。

随着时代的变迁,语言也不断进化,许多以前没有的词语也不断涌现,因此用不同平滑化的方法来处理未登录词问题也是相当重要的议题。另外,也可以与现行的语言模型相结合,如 LDA(latent dirichlet allocation)语言模型或 DTW(dynamic time warping)语言模型等。使语言模型更具有一般性和适应性能力。神经网络语言模型在一定程度上类似,差别在于一种是监督式的,另一种是非监督式的。倘若在类似神经网络语言模型中利用 LDA 语言模型的思想对其进行改良,识别效果应该会有更好的提升,期望在未来能将这两种语言模型相结合,并进一步获得更好的识别结果。

[1] Bengio Y, Ducharme R, Vincent P, et al. A neural probabilistic language model [J]. Journal of Machine Learning Research, 2003, 3(2): 1137-1155

[2] Mikolov T, Kopecky J, Burget L, et al. Neural network based language models for highly inflective languages [C] // Proc of the 34th IEEE Int Conf on Acoustics, Speech and Signal. Piscataway, NJ: IEEE, 2009: 4725-4728

[3] Boulanger-Lewandowski N, Bengio Y, Vincent P. Modeling temporal dependencies in high-dimensional sequences: Application to polyphonic music generation and transcription [C] //Proc of the 29th Int Conf on Machine Learning. New York: ACM, 2012: 590-598

[4] Bottou L. Stochastic gradient learning in neural networks [C] //Proc of Neuro-Nimes 91, Nimes: EC2, 1991: 687-699

[5] Hinton G E, Osindero S, Teh Y W. A fast learning algorithm for deep belief nets [J]. Neural Computation, 2006, 18(7): 1527-1554

[6] Bengio Y, Frasconi P, Simard P. The problem of learning long-term dependencies in recurrent networks [C] //Proc of IEEE Int Conf on Neural Networks. Piscataway, NJ: IEEE, 1993: 1183-1188

- [7] Xu W, Rundicky A. Can artificial neural networks learn language models? [C] //Proc of the 6th Int Conf on Spoken Language Processing. Beijing, China; ISCA, 2000
- [8] Bengio Y, Simard P, Frasconi P. Learning long-term dependencies with gradient descent is difficult [C] //Proc of IEEE Trans on Neural Networks. Piscataway, NJ: IEEE, 1994: 157-166
- [9] Mikolov T, Kombrink S, Deoras A, et al. RNNLM- Recurrent neural network language modeling toolkit [C] // Proc of IEEE Workshop on Automatic Speech Recognition and Understanding. Piscataway, NJ: IEEE, 2011: 5528-5531



Li Yaxiong, born in 1979. Engineer in the Network Management Center at the Hubei University of Science and Technology. His main research interests include computational linguistics and machine learning.



Zhang Jianqiang, born in 1980. PhD candidate at the Virginia Polytechnic and State University. His main research interests include human-computer interaction, machine learning and computational linguistics.



Pan Deng, born in 1977. Associate professor in the School of Foreign Languages at the Hubei University of Science and Technology. His main research interests include instructional technology, corpus linguistics.



Hu Dan, born in 1971. Professor in the School of Foreign Languages at the Zhongnan University of Economics and Law. His main research interests include language resources and translation technology.