基于循环神经网络的汉语语言模型建模方法

王 龙[1,2]，杨俊安[1,2]，陈 雷[1,2]，林 伟[3]
(1. 中国人民解放军电子工程学院，安徽合肥 230037；2. 安徽省电子制约技术重点实验室，安徽合肥 230037；
3 安徽科大讯飞公司，安徽合肥 230037)

摘要：语言模型是语音识别系统的重要组成部分，目前的主流是 *n*-gram 模型。然而 *n*-gram 模型存在一些不足，对语句中长距信息描述差、数据稀疏是影响模型性能的两个重要因素。针对不足，研究者提出循环神经网络(Recurrent Neural Network, RNN)建模技术，在英语语言模型建模上取得了较好的效果。根据汉语特点将 RNN 建模方法应用于汉语语言建模，并结合两种模型的优点，提出了模型融合构建方法。实验结果表明：相比传统的 *n*-gram 语言模型，采用 RNN 训练的汉语语言模型困惑度(PerPLexity, PPL)有了下降，在对汉语电话信道的语音识别上，系统错误率也有下降，将两种语言模型融合后，系统识别错误率更低。

关键词：语音识别；循环神经网络；语言模型；模型融合
中图分类号：TP391 **文献标识码：**A **文章编号：**1000-3630(2015)-05-0431-06
DOI 编码：10.16300/j.cnki.1000-3630.2015.05.010

Recurrent neural network based Chinese language modeling method

WANG Long[1,2], YANG Jun-an[1,2], CHEN Lei[1,2], LIN Wei[3]
(1. *Electronic Engineering Institute of PLA*, *Hefei* 230037, *Anhui*, *China*; 2. *Key Laboratory of Electronic Restriction*, *Anhui Province*, *Hefei* 230037, *Anhui*, *China*; 3. *Anhui USTC iFlytek Corporation*, *Hefei* 230037, *Anhui*, *China*)

Abstract: Language model is an important part in the speech recognition system, the current mainstream technique is n-gram model. However, n-gram language model still has some shortcomings: the first is poorly to describe the long-distance information of a sentence, and the second is to arise the serious data sparse phenomenon; essentially they are the two important factors influencing the performances of the model. Aiming at these defects of n-gram language model, the researchers put forward a recurrent neural network (RNN) modeling technique, with which, the training for the English language model has achieved good results. According to the characteristics of the Chinese language, the RNN method is used for training the Chinese language model; also a model combination method to combine the advantages of the two models is proposed. The experimental results show that: the perplexity of RNN model has a certain decline, there is also a certain decline on the system recognition error rate，and after model combination, the recognition error rate reduces much more on the Chinese phone speech recognition, compared with the n-gram language model.
Key words: speech recognition; recurrent neural network; language model; model combination

0 引 言

语音识别(Speech Recognition)是指机器通过识别和理解，把人类的语音信号转变为相应文本或命令。由于语音信号的动态时变性、瞬时性和随机性，单靠声学模型的匹配与判断无法完成语音无误的识别和理解[1]，需要在此基础上结合语法、语义以及上下文内容等非声学的语言知识加以约束，进而提高系统的识别准确率。语言模型用于刻画自然语言中的内在规律，能够提供字或词之间的上下文和语义信息，因此成为语音识别系统的重要组成部分。

目前，基于回退(back-off)平滑算法的 *n*-gram 语言模型，在汉语语言模型建模领域占据主导地位。*n*-gram 建模技术具有很好的建模能力，实现也相对简单，当语料充足时，能够训练出性能很好的模型。但此建模技术仍有明显缺点。一是对语句中长距依存描述能力较弱，在训练时，模型阶数 *n* 通常取 2(Bi-gram)或 3(Tri-gram)；二是数据稀疏，由于训练语料中不可能覆盖所有的语言现象，此时就会造成“零概率”即数据稀疏。且模型阶数越大，数据稀疏越严重，需要另外结合数据平滑技术进行

收稿日期：2014-10-22；**修回日期：**2015-02-09
基金项目：国家自然科学基金(60872113)、安徽省自然科学基金(1208085MF94, 1308085QF99)资助项目。
作者简介：王龙(1989－)，男，硕士研究生，研究方向为声信号分析与识别技术。
通讯作者：王龙，E-mail: longwang0927@126.com

训练。

研究者致力于解决 *n*-gram 语言模型的固有缺陷问题，希望机器对语言的理解能够更加接近于人类的理解，进而提高实际语音识别系统的性能，故神经网络(Neural Network)逐渐在语音信号处理中得到应用[2,3]，并取得了较好的效果。用神经网络训练语言模型的思想最早由徐伟于 2000 年提出[4]，在实验中，采用神经网络训练出的二元语言模型取得了比 Bi-gram 结合平滑算法更好的模型性能。随后，Bengio 在网络中增加一个隐含层，构建了一个三层的前馈神经网络来训练模型[5]，此网络结构能够描述词与词之间更高元的依附关系，取得了比 *n*-gram 语言模型更好的效果。然而，模型训练时需要对训练语料进行特殊处理，还需要较好的参数选择，因此不容易实现。循环神经网络[6]的语言模型建模方法相比 Bengio 的网络结构，RNN 增加了一个可以反馈信息的存储层，能够存储当前词的所有历史信息，此时网络能够以当前词的整个上下文作为依据，预测下一个词的出现概率。另外，在模型训练的过程中，由于当前词的历史信息被映射到低维连续空间，语义相似的词被聚类，在语料中出现次数较少的词仍然能够得到很好地训练，解决了数据稀疏问题。

在汉语语言模型建模技术中，占主导地位的依然是基于统计规则的 *n*-gram 建模技术。本文根据汉语的特点首先将 RNN 建模方法应用到汉语语言模型建模，并在此基础上结合两个模型的优势，提出了一种模型融合构建方法。实验结果表明：基于 RNN 的汉语语言模型及融合方法构建的模型，在实际识别系统的识别率上都取得了较好的效果。

1 汉语语音识别系统

1.1 概 述

完整的语音识别系统基本原理框图如图 1 所示，先将语音信号数字化，其预处理包括预加重、加窗、分帧、端点检测等过程。再对其声学参数进行分析，提取出语音特征参数，形成特征矢量序列，包括短时平均幅度或能量、频谱、平均过零率，线性预测倒谱系数、Mel 倒谱系数等。在识别阶段，将待识别语音的特征矢量参数同训练得到的参考模板库中的模式进行相似性度量比较，将相似度较高的模式所属的类别作为识别中间候选结果输出。在后处理阶段，通过语言模型对初步识别候选结果进行判断和决策，可以有效地提高解码的效率和精

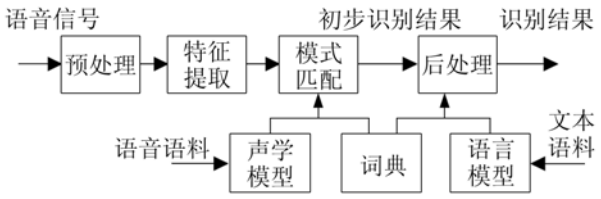

图 1 汉语语音识别系统基本框图

Fig.1 The basic block diagram of Chinese speech recognition system

度[7]，对于汉语存在大量的同音字、多音字，较高层次的语言知识的利用能够在声学识别的基础上减少模式匹配的模糊性，提高了系统识别的精确度。

上述工作完成后，得到的声学特征矢量，记为 o。自然语言可以被看作是一个随机序列，文本中的每个句子或词都是具有一定分布的随机变量。假设词(汉语中包含单字)是一个句子最小的结构单位，一个合理的包含 N 个词的语句 s 由词序列 $w=w_1, w_2, \cdots, w_N$ 组成。依贝叶斯准则，语音识别的过程就是根据式(1)，找出当前声学特征序列出现概率最大的词序列作为识别结果[8]。

$$\hat{w}=\arg\max_w p(w|o)=\arg\max_w \frac{p(o|w)p(w)}{p(o)}= \arg\max_w p(o|w)p(w) \tag{1}$$

$$p(w)=\prod_{i=1}^{N} p(w_i|w_1\cdots w_{i-1}) \tag{2}$$

其中，信号波形的先验概率 $p(o)$ 与词序列 w 选择无关，可不计算；$p(o|w)$ 代表在给定某个词序列的基础上，输出特征序列的可能性，在语音识别中用声学模型对其建模；而 $p(w)$ 表示了词序列 w 出现的可能性，在语音识别中用语言模型进行建模；$\hat{w}$ 表示概率值最大的句子，argmax 表示确定概率值最大句子时所对应的参数值，即组成句子的所有词序列。

1.2 *n*-gram 语言模型

统计规则的 *n*-gram 语言模型于 1980 年提出[9]，是应用广泛的语言模型，它采用 Markov 假设，即认为每一个预测变量出现的可能性只与长度为 $n-1$ 的上下文有关。若将词 w_i 的历史信息表示为 $h_i=w_1w_2\cdots w_{i-1}$，则根据条件概率公式及 Markov 假设，词 w_i 和句子 s 出现的概率分别为：

$$p(w_i|w_1, w_2, \cdots, w_{i-1})=p(w_i|h_i)=p(w_i|w_{i-n+1}^{i-1}) \tag{3}$$

$$p(s)=p(w_1, w_2, \cdots, w_N)=\prod_{i=1}^{N} p(w_i|h_i) \tag{4}$$

对于 *n*-gram 语言模型而言，通常需要对大规模的语料进行训练，语料中出现频率越高的词往往能够训练得越好，而对低频词训练的效果不理想。另外，阶数 n 越大，模型的约束力越强，然而随着

n 的增大，模型规模成指数级增长，训练时计算复杂度增大，对存储空间也提出了更高的要求。因此合适的 n 值是语言模型精确度与复杂度之间的折衷。在实际识别系统中，一般选择 n=3 来构造 Tri-gram 语言模型，即训练数据的句子中每个词出现的概率只与其前两个词有关，可表示为

$$p(w_i|h_i)\approx p(w_i|w_{i-2},w_{i-1}) \tag{5}$$

本文针对汉语语音识别系统中的语言模型进行改进，在基线系统其他模块不变的情况下，用 RNN 语言模型对初步识别候选结果重打分，进行识别后处理，完成整个系统的识别过程。

2　RNN 语言模型

参考文献[10]指出：循环神经网络又称为 Elman 网络，其结构如图 2 所示，由三个网络层构成，分别是输入层、隐含层、输出层，存储层作为输入的一部分，保存了上一时刻隐含层的状态。文本语料经过此 RNN 结构训练后，当前词 w_i 出现的概率表示为[11]

$$p(w_i|h_i)=p_{rnn}(w_i|w_{i-1},h_{i-1})=p_{rnn}(w_i|h_i) \tag{6}$$

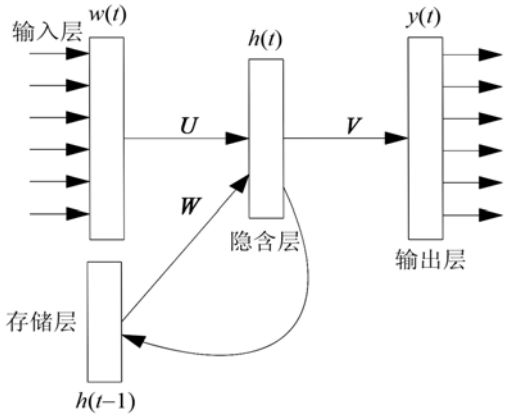

图 2　循环神经网络框架

Fig.2　The structure of recurrent neural network

参考文献[10]还指出：假设在 t 时刻网络输入词样本为 $w(t)$，即当前词的向量、维数由语料中词样本数 $|V|$ 决定；隐含层的状态 $h(t)$ 由输入当前词向量 $w(t)$ 和上一时刻隐含层的状态即历史信息 $h(t-1)$ 共同决定，通过隐含层到输入层的连接，将 $t-1$ 时刻的隐含层状态作为 t 时刻输入的一部分；输出层 $y(t)$ 表示当前历史下后接词的概率分布信息，输出层节点与输入层节点数相同也是 $|V|$。各个层之间计算关系用下列式子表示：

隐含层的输入：　$x(t)=w(t)+h(t-1)$

隐含层状态：　$h_j(t)=f(\sum_i x_i(t)u_{ji})$

输出层状态：　$y_k(t)=g(\sum_j h_j(t)v_{kj})$

其中，$f(z)=\dfrac{1}{1+e^{-z}}$ 为 sigmoid 激活函数，$g(z_m)=e^{-z_m}\Big/\sum_k e^{-z_k}$ 为 Softmax 函数。

Softmax 函数保证了当前词下后接词的概率分布是合理的，即对于任意一个词 m 的 $y_m(t)>0$，且 $\sum_k y_k(t)=1$。在模型参数的初始化设置上，隐含层初始状态 $h(0)$ 一般设为零，或随机初始化为很小的值。输入词向量 $w(t)$ 形式如(0 0 0 1 0 0……)，(0 0 1 0 0 0……)，其中一维为 1，其他维置为 0。隐含层节点数通常取 100 到 1000，根据具体训练数据的大小进行调节。$\boldsymbol{U}$、$\boldsymbol{W}$、$\boldsymbol{V}$ 为各层之间权值矩阵，随机初始化为较小的值，在模型训练的过程中，通过标准的反向传播算法(Back Propagation, BP)结合随机梯度下降法学习更新[10,12]：

$$\boldsymbol{e}(t)=\text{desired}(t)-y(t) \tag{7}$$

$$\Theta(t)=\Theta(t-1)+\alpha\nabla_{SGD} \tag{8}$$

其中：$\boldsymbol{e}(t)$ 表示每处理一个样本时的输出端的误差向量，desired(t) 表示特定上下文中期望输出的概率值。Θ 包括网络中 $\boldsymbol{U}$、$\boldsymbol{W}$、$\boldsymbol{V}$ 三个权值矩阵变量，∇_{SGD} 表示相应的梯度下降值，经过不断的迭代学习直到模型收敛。

3　汉语的特点

相比于英语语音，汉语语音识别更加复杂。汉语普通话中有 6000 多个常用字，大约有 60 个音素，407 个无调音节，1332 个有调音节，每个音节由声母、韵母和声调组成，每个汉字代表一个音节，音节和音节之间的连音现象不明显[13]，给声学模型的匹配计算带来难度。同时，汉语中还存在大量的同音、多音字现象，必须通过上下文语境等高层次的非声学知识加以约束才能完成识别。

在语言方面，英语语句注重结构，而汉语语句注重语义，同一个词在不同的语境下有不同的含义，词之间的长距离依附关系比较紧密。采用 RNN 训练语言模型时，考虑了更多高层次的语义信息，更能反映出汉语词与词之间的约束关系。因此，RNN 建模技术将更适合于汉语语言模型的训练。

另外，在英文语料句子中词与词之间都有明显的空格间隔，语料稍加处理可直接进行模型训练。而汉语语料则不同，汉语句子中字与字之间没有明显的界限，还需要根据分词模型对训练语料进行分词，将句子分隔成子词单元。经过一系列处理得到纯净的文本语料后才能进行模型训练。

汉语训练语料处理流程图如图 3 所示，文本语料要进行清洗，将粗语料中的字母、标点符号等噪声信息删除，去除冗余信息；语料中会存在大量数字，还要完成语数字正规化，将语料中的阿拉伯数字转换成汉字；这样语料中只存在汉字信息，分词根据分词模型将词与词用空格分开，将句子划分成词(或字)单元；词典过滤删除英文边界符和语料中的非词典词句子。如此得到能够用来训练的语料，然后进行模型训练。

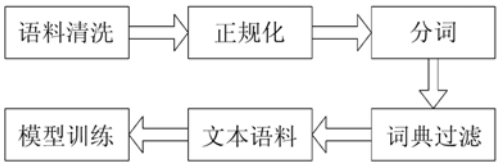

图 3 汉语语料的处理流程

Fig.3 The processing flow of Chinese corpus

4 RNN 模型与 *n*-gram 模型融合建模

语料中频数越高的词汇 *n*-gram 建模技术能够训练得越好，而对频数较低的词汇则相反。采用 RNN 对语料进行训练时，尽管语料中有些词的频数比较低，但依然能够训练得很好，可以对 *n*-gram 模型进行有效的补充。为了充分发挥两种建模技术的优势，得到更好的识别效果，本文研究了一种基于 RNN 模型与 *n*-gram 模型的融合建模方法。

如图 4 所示，对于同一条语音，从解码器生成的词图(lattice)中，可以获得 *n*-best 列表，再利用训练好的 RNN 语言模型对 *n*-best 列表进行重打分。然后将 *n*-best 列表的 *n*-gram 模型得分信息与 RNN 模型的重打分信息进行插值融合，计算出每一个候选单元新的语言模型得分。

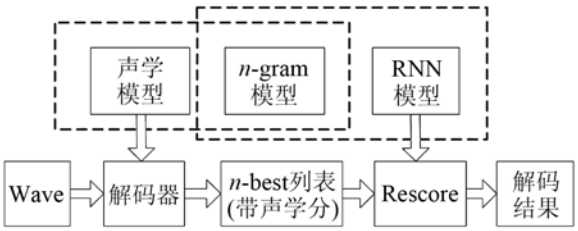

图 4 基于 RNN 与 n-gram 语言模型融合

Fig.4 The combination of RNN based model and n-gram language model

在模型的融合算法中，线性插值融合是目前较常用的方法[14]，根据上下文 h_i 预测当前词 w_i 的概率：

$$p_x(w_i|h_i)=\sum_j^L \lambda_j p_j(w_i|h_i) \tag{9}$$

$$p(w_i|h_i)=\lambda_1 p_{rnn}(w_i|h_i)+\lambda_2 p_{ng}(w_i|h_i) \tag{10}$$

式中：L 为插值模型的个数；各模型的插值权重 λ_j 非负，且总和为 1，即 $\sum_{j=1}^{L}\lambda_j=1$。模型融合后对每一个 *n*-best 列表句子重新打出对数似然得分：

$$\lg L(s)=\sum_{i=1}^{n} asc_i+lms\sum_{i=1}^{n}\lg p_x(w_i|h_i)+n\cdot wp \tag{11}$$

其中：n 是句子中词的个数；wp 是词的惩罚分；asc_i 为词 w_i 声学模型得分；lms 为模型规模；$p_x(w_i|h_i)$ 代表每个词的 *n*-gram 与 RNN 模型的融合得分，由式(10)计算得出。根据式(11)将语言与声学模型得分以及惩罚分信息结合起来，计算出每一个列表的总体得分并对比，从中选出得分最高的一个作为此 *n*-best 列表的最终识别结果，将所有的识别结果与对应的语音标注数据进行比对后计算出系统的识别率。

5 语言模型评价方法及实验分析

5.1 语言模型评价标准

评价语言模型性能优劣是根据信息论知识。通过计算语言模型在测试文本上困惑度的大小对语言模型的性能进行衡量。困惑度是指用语言模型预测某文本集中每个词的出现概率时，这些概率的几何平均值的倒数。假设测试文本中有 M 个词，则困惑度为

$$Perplexity=\frac{1}{\sqrt[M]{2^{\sum_{i=1...M}\log_2 P(w_i|w_{i-n+1},...,w_{i-1})}}} \tag{12}$$

通常情况下，该值越小，表明语言模型对语言约束力强，所训模型的性能好[15]。除了用困惑度衡量语言模型外，最直观的想法是将模型应用到系统中，通过测试系统误字率(Word Error Rate, WER)进行衡量。一般模型训练得好，则系统的识别率就高。在本文实验中将两种评价标准结合起来对语言模型进行测试分析。

5.2 实验设计

实验 1 用于验证 RNN 语言模型在汉语语音识别中的效果。实验中分别用 RNN 和 *n*-gram 建模在同一数据集上进行模型训练，并且通过改变 RNN 隐含层节点数训练出不同的 RNN 语言模型，研究不同参数下 RNN 模型困惑度和系统识别率的变化，并与 *n*-gram 模型的性能进行比较。实验 2 用于验证提出的模型融合算法的有效性，将实验 1 中训练出来的 *n*-gram 模型与 RNN 模型分别进行线性插值融合，并测试识别率的变化。为了加速 RNN

模型的训练，本文中 RNN 训练是在 GPU(NVIDIA GTX 650)服务器上结合 CUDA Toolkit 5.5 进行的，相比在 CPU 上训练速度相对提升 2~3 倍，缩短了 RNN 模型的训练时间。

5.3　实验数据

训练数据来源于科大讯飞公司提供的汉语电话语音转写任务标注数据，共 16 M，包含 550 千个句子，4342 千个词。模型困惑度测试本文共 9332 个句子，包含 23 千个词，语音测试集为对电话语音解码结果 3433 句 100-best 列表，大小为 87 kB。RNN 训练模型时隐含层节点数共设六组参数。

5.3.1　实验 1

n-gram 语言模型训练结合了平滑性能较好的 Kneser-Ney 回退平滑算法，模型阶数取 3 即 3-gram，由 SRILM 工具箱构建。模型中 *n*-gram 数量为：30274+6568660+13830707=20429641，模型大小为 463599 kB。然后用同样的训练数据训练 RNN 语言模型，分别测试模型 PPL 值以及识别系统 WER 的变化。实验结果如表 1 所示。

由表 1 可见，同样的训练数据分别训练 RNN 和 3-gram 语言模型，在困惑度上 RNN 语言模型相对降低 7%左右；在语音识别的误字率上，RNN 语言模型相对下降 5%左右，证明了 RNN 语言模型在汉语语音识别中的有效性。一般情况下，生成模型的困惑度越低，系统的识别率就越高。

另外，从表 1 还可以看出：(1) 随着隐含层节点数的增加，RNN 语言模型的困惑度以及系统误字率呈逐渐下降的趋势，说明网络的学习能力随着节点数的增加而增强；(2) 当隐含层的节点数增加到一定程度后，生成模型的困惑度反而会升高，系统识别率也随之下降，说明隐含层节点数越多，网络结构越复杂，通过学习可以使训练样本的误差减少到足够小，然而过分地追求在训练样本上的学习会产生过度训练。在训练样本数有限的情况下，当学习进行到一定阶段后，如果学习样本集的平均训练相对误差继续减小，而测试样本集的平均测试相对误差(泛化误差)反而增大，导致网络的泛化能力降低，影响所训练模型的性能。因此，对于不同规模的训练语料，采用 RNN 训练时参数需要进行调整，才能达到较好的实验效果。

5.3.2　实验 2

在实验 1 识别结果的基础上，采用线性插值方法，将两套模型对 100-best 列表中的每个词的语言模型得分进行插值重新打分，再根据式(11)结合声学模型得分等，计算出每一句话的概率，从中选出得分最高的作为该条语音的识别结果，如表 2 所示。实验中插值系数为 0.6，即 λ_1=0.6 时取得了较好的识别效果。

由表 2 可见，经过线性插值后模型在识别率上，相对于 3-gram 模型误字率下降 8%左右，相比于 RNN 模型误字率也下降了 3%左右，模型融合后的识别率相对于 *n*-gram 模型的识别率提升较为明显，这是因为 RNN 模型对低频词的训练效果较好，能够有效地解决数据稀疏问题。同时可以看出模型融合后的识别率比任何一个单独模型的识别率高，说明两种模型具有互补作用，证明了模型融合方法的

表 1　RNN 语言模型与 3-gram 语言模型性能对比

Table 1　The performance comparison between RNN and 3-gram language models

隐含层节点数	*n*-gram(KN3)		RNN		PPL 下降率/%	WER 下降率/%
	PPL	WER/%	PPL	WER/%		
100	166.3	42.09	158.3	40.2	4.81	4.49
200	166.3	42.09	157.6	40.1	5.23	4.72
300	166.3	42.09	156.6	40.08	5.83	4.77
400	166.3	42.09	155.4	39.96	6.55	5.06
500	166.3	42.09	154.5	39.87	7.1	5.27
600	166.3	42.09	155.2	39.93	6.67	5.13

表 2　两种模型融合后识别性能对比

Table 2　The comparison of recognition performance after combination of two models

隐含层节点数	RNN+*n*-gram	*n*-gram	RNN	与 *n*-gram 相比下降率/%	与 RNN 相比下降率/%
	WER/%	WER/%	WER/%		
100	39.23	42.09	40.2	6.79	2.41
200	39.14	42.09	40.1	7	2.39
300	38.6	42.09	40.08	8.29	3.69
400	38.62	42.09	39.96	8.24	3.35
500	38.56	42.09	39.87	8.38	3.28
600	38.59	42.09	39.93	8.31	3.35

有效性。在实际的识别系统中，可以先训练一个大语料的 *n*-gram 语言模型，并以此模型作为通用模型，然后与用少量语料训练的 RNN 模型在语音识别系统的后处理模块进行插值融合，采用这种方法对解码结果进行后处理，提高系统的识别率。在本实验中，由于模型训练语料有限，系统整体识别率不是很高，但是仍然可以看出 RNN 在汉语语言模型建模方面的优越性，以及模型融合构建方法的有效性。

6 结束语

本文将 RNN 语言模型建模应用到汉语语言模型上，通过与传统的 *n*-gram 模型对比，生成模型的困惑度降低 7%左右，在对实际电话信道语音识别上误字率降低了 5%，验证了 RNN 建模方法在汉语语言处理中的有效性。另外，本文提出 RNN 语言模型与 *n*-gram 语言模型的融合方法，使语音识别系统的性能得到进一步提升，识别效果优于任一单个模型，证明了融合算法的优越性。虽然 RNN 语言模型性能较高，但由于其相对较高的计算复杂度，导致训练效率很低，本文中 RNN 汉语语言模型的训练是在 GPU 上进行的，相对提升了训练效率。如何进一步提升模型训练效率，是下一步研究的重点。

参 考 文 献

[1] 倪崇嘉, 刘文举, 徐波. 汉语大词汇量连续语音识别系统研究进展[J]. 中文信息学报, 2009, **23**(1): 114-117.
NI Chongjia, LIU Wenju, XU Bo. Research on large vocabulary continuous speech recognition system for mandarin Chinese[J]. Journal of Chinese Information Processing, 2009, **23**(1): 114-117.

[2] 杨云升, 温晓杨, 吕敏. 一种基于 BP 神经网络的语音相空间客观干扰效果评估模型[J]. 声学技术, 2009, **28**(4): 507-511.
YANG Yunsheng, WEN Xiaoyang, LÜ Min. A BP artificial neural network model for evaluating jammed effect in speech phase-space[J]. Technical Acoustics, 2009, **28**(4): 507-511.

[3] 陈存宝, 赵力. 嵌入时延神经网络的高斯混合模型说话人辨认[J]. 声学技术, 2010, **29**(3): 292-296.
CHEN Cunbao, ZHAO Li. Speaker identification based on GMM with embedded TDNN[J].Technical Acoustics, 2010, **29**(3): 292-296.

[4] XU Wei, AlexRudnicky. Can artificial neural networks learn language models?[C]// Proceedings of International Conference on Spoken Language Processing. 2000.

[5] Bengio Yoshua. A neural probabilistic language model[J]. Journal of Machine Learning Research, 2003, **10**(3): 1137-1155.

[6] Tom´aˇs Mikolov. Statistical language models based on neural networks[D]. Brno University of Technology, Czech Republic, 2012.

[7] 甘海波. 语音识别系统中声学层模型的研究[D]. 哈尔滨: 哈尔滨工业大学, 2008.
GAN Haibo. The research about the acoustic model in speech recognition system[D]. Harbin: Harbin Institute of Technology, 2008.

[8] 张强. 大词汇量连续语音识别系统的统计语言模型应用研究[D]. 成都: 西南交通大学, 2009.
ZHANG Qiang. Application research on statistical language model of large vocabulary continuous speech recognition system[D]. Chengdu: Southwest Jiaotong University. 2009.

[9] 邢永康, 马少平. 统计语言模型综述[J]. 计算机科学, 2003, **30**(9): 22-29.
XING Yongkang, MA Shaoping. A survey on statistical language models[J]. Computer Science, 2003, **30**(9): 22-29.

[10] Mikolov T, Karafi´at M, Burget L, et al. Recurrent neural network based language model[C]// Proceedings of Interspeech, 2010: 1045-1048.

[11] Kombrink S, Mikolov T, Karafi´at M, et al. Recurrent neural network based language modeling in meeting recognition[C]// Proceedings of Interspeech, 2011: 2877-2880.

[12] Mikolov T, Kombrink S, Burget L, et al. Extensions of recurrent neural network language model[C]// Proceedings of ICASSP, 2011: 5528-5531.

[13] 吴斌. 语音识别中的后处理技术研究[D]. 北京: 北京邮电大学, 2008.
WU Bin. Post-processing technique for speech recognition[D]. Beijing: Beijing University of Posts and Telecommunication, 2008.

[14] Mikolov T, Deoras A, Kombrink S, et al. Empirical evaluation and combination of advanced language modeling techniques[C]// Proceedings of Interspeech, 2011: 605-608.

[15] 张仰森, 曹大元, 俞士汶. 语言模型复杂度度量与汉语熵计算[J]. 小型微型计算机系统, 2006, **27**(10): 1931-1934.
ZHANG Yangsen, CAO Dayuan, YU Shiwen. Perplexity measure of language model and the entropy of Chinese[J]. Mini-micro Systems, 2006, **27**(10): 1931-1934.