

# 利用拼音特征的深度学习文本分类模型<sup>①</sup>

赵博轩<sup>②\*</sup> 房 宁<sup>③\*\*</sup> 赵群飞<sup>\*</sup> 张朋柱<sup>\*\*\*</sup>

( \* 上海交通大学自动化系 系统控制与信息处理教育部重点实验室 上海 200240)

( \*\* 上海交通大学人文学院 上海 200030)

( \*\*\* 上海交通大学安泰经济与管理学院 上海 200030)

**摘 要** 针对人-机器人语音交互中经过语音识别的文本指令,提出了一种利用汉语拼音中声韵母作为特征的深度学习文本分类模型。首先,以无人驾驶车语音导航控制为人机交互的应用背景,分析其文本指令结构并分别构建单一意图与复杂意图语料库;其次,在以字符作为文本分类特征的基础上,结合汉语拼音与英文单词的区别,提出了一种利用拼音声韵母字符作为中文文本分类的特征表示方法;然后,用门控递归单元(GRU)代替传统递归神经网络单元以解决其难以捕获长时间维度特征的不足,为提取信息的高阶特征、缩短特征序列长度并加快模型收敛速度,建立了一种结合卷积神经网络及 GRU 递归神经网络的深度学习文本分类模型。最后,为验证模型在处理长、短序列任务上的表现,在上述两个语料库上对提出的模型分别进行十折交叉测试,并与其他分类方法进行比较与分析,结果表明该模型显著地提高了分类准确率。

**关键词** 文本分类,意图理解,声韵母特征,门控递归单元(GRU)

## 0 引 言

人机交互是人和计算机设备之间相互影响的技术,其目标是使机器人帮助人高效、舒适、安全地完成任<sup>①</sup>。高效友好的人-机器人交互是任务型服务器机器人技术中的关键环节。人与机器人的实际交互方式已经多样化,主要包括语音交互、手势交互、表情交互等,其中语音交互是最直接、最便捷也是应用范围最广泛的交互方式。因此,对语音交互中文本指令理解方法的研究有着重要的意义。

在人-机器人语音交互中,用户的语音指令首先经语音文本转换(sound to text transform, STT)技术转成文本;其次,对文本指令的含义,也就是对用户的意图进行理解。

目前语音交互中文本指令的意图理解主要分为两种基本方法:基于规则的理解与基于统计语言学的理解,其本质是将对文本的理解问题转化为文本分类问题。前者指的是对语言的语法规则、词性和构词法等特性用计算机语言进行描述,从而对文本进行理解。而基于统计学习的理解方法则是通过对语料库进行特征提取后的统计学习,对文本类别进行判断。在大规模语料库的特定应用背景下,相比于基于规则的方法,基于统计学习的理解方法具有更理想的效果,也成为当今主流的文本理解方法。基于向量空间模型(vector space model, VSM)的文本指令模型可配合传统机器学习的分类方法进行文本分类,例如文献[2]中提出了一种利用向量空间模型与最大熵(maximum entropy, ME)模型结合的文本分类方法,文献[3]在传统向量空间模型

① 国家自然科学基金(91646205)资助项目。

② 男,1993年生,硕士生;研究方向:人-机器人交互,自然语言处理;E-mail: zbxuan@sjtu.edu.cn

③ 通讯作者,E-mail: nfang@sjtu.edu.cn

(收稿日期:2017-02-26)

础上引入词性向量空间模型,使特征信息更丰富,并引入深度去噪自编码器(stacked denosing autoencoder, SDAE)处理噪声词和无实义词的影响,且加入了同义词表增加文本先验知识并降低空间维度,但这些模型存在着对复杂问题表示能力不足,对复杂问题泛化处理能力不足等局限性<sup>[4]</sup>。

深度学习理论近些年开始被用于自然语言理解领域, Kim<sup>[5,6]</sup>、Kalchbrenner<sup>[7]</sup>等分别提出了不同结构的基于卷积神经网络的针对短句级语料库的文本分类模型,并在多个数据库上得到了较理想的结果。Dos Santos<sup>[8]</sup>、Xi<sup>[9]</sup>、Tang<sup>[10]</sup>等将卷积神经网络用于文本情感分析任务并取得了出色的效果。Lecun<sup>[11-13]</sup>将其处理图像的卷积神经网络应用于文本分类,而且提出字符可作为文本分类的一种新特征。Lai 等<sup>[14]</sup>提出了一种基于递归神经网络(recurrent neural network, RNN)的文本分类模型,并证明了深度学习在自然语言处理领域同样大有潜力。

本文在以字符作为特征的基础上,依据汉语拼音自身特点及其与英文单词在成分构成上的不同,提出了一种利用汉语拼音中的声母、韵母作为文本分类的新特征的特征选择方法。在深度网络分类模型中应用变长窗口的卷积神经网络对不同水平的特征进行有效表示,以得到不同抽象层次上的抽象数据表示;然后利用门控递归单元(gated recurrent units, GRU)对特征序列进行处理,得到隐层节点;最后,用 softmax 回归对文本指令进行预测分类。考虑到无人驾驶车将在未来智能交通中担任重要角色,而基于语音交互的控制方式是最直接、高效也是最常用的人与无人驾驶车交互的方式,本文以无人驾驶车作为人机语音交互的应用场景,分析了文本指令结构并分别构建了单一意图与复杂意图语料库。实验中用提出的语言模型在所构建的两个指令语料库上分别进行了训练与十折交叉验证,以验证模型在不同长度的语料任务上的表现,并与传统基于统计的文本分类方法进行了比较。

## 1 人机交互的指令意图理解流程

人机交互中的指令意图理解流程如图 1 所示,

主要包括语音识别、预处理、意图理解三个模块。语音识别模块采集用户发出的指令并将语音转化为文本;预处理模块对经语音识别得到的文本根据意图理解中的特征方法进行预处理,如分词、词性标注等,而本文所提出的特征方法只需将文本分词,再将其从汉字转化为拼音,目前汉字转拼音已经有成熟的模块如 pypinyin 等,此过程不需要手动转化;意图理解模块对预处理后的文本进行特征提取,并用预训练的文本分类模型将其分类,分类的准确率是高质量人机交互过程的关键,也是本文重点的研究内容。

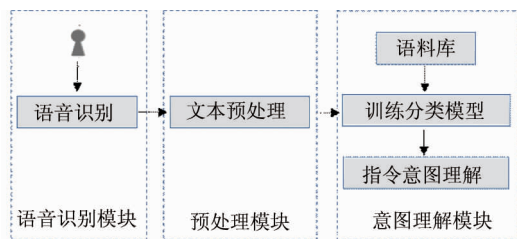


图1 人-机器人交互系统框架

## 2 文本特征表示方法

文本特征表示法中最简单直接的是独热(one-hot)编码表示法,其实质就是将词语编号,将词语映射到一个稀疏向量,此方法词与词之间是孤立的,无法得到词语间的关系信息,且当处理大语料库时将产生维度灾难。

向量空间模型<sup>[15]</sup>将文本转化为多维向量进行表示,每条文本都被映射为向量空间中的一个点 $(T_1, W_1; T_2, W_2; \dots, T_n, W_n)$ ,其中 $T_i$ 为词, $W_i$ 为词对应的权值,用以体现词在文本中的重要程度,此外还有加入了词性信息的词性向量空间模型(POS-VSM),词性的加入使文本表示信息更加丰富,但向量空间模型假设词与词之间是独立的,一个关键字唯一代表一个概念或语义单元,而实际情况是文档存在很多的一词多义和同义词现象,因此这种假设很难满足实际情况,且其在表示大规模语料库时也存在维度过高的问题。

此外词向量表示法<sup>[16]</sup>也广泛应用于文本分类问题中,词向量是从大量未标注的文本语料数据中

无监督地学习得到词语的低维度密集向量表示方法,有效解决了维度过大的问题,且可通过两个词向量的空间距离来判断两个词的关系。词向量法在处理大规模语料库上有着良好的效果,但在处理特定任务类型的语料库时,容易在训练词向量时产生信息丢失,造成分类效果下降。

Lecun<sup>[11,12]</sup>提出了一种基于字符的文本特征方法,其处理文本时不再以单词而以每个字符为单位,将每个字符映射到事先定义好的字符表上得到字符的稀疏表示,并将其应用在多个语料库分类任务上,得到出色的结果。

汉语拼音与英文单词有所不同,英文单词中的字符是独立的,而汉语拼音中则存在字符的“组合”,即多字符的声韵母,如‘ai’,‘zh’等其本身为一整体,如将它们像文献[11]中的方法那样拆分逐字符处理,将明显丢失有效信息,另外依据声韵母的字符映射方式可有效缩短序列长度。根据这一特点,我们提出了一种基于声韵母的拼音映射规则并规定了一个 57 维的字符映射表,包括 26 个英文字母,10 个数字,18 个由 2 个或以上字符组成的常见声韵母,以及 3 个其他符号,总字符表为

{abcdefghijklmnopqrstuvwxyz0123456789 ‘ai’ ‘ei’ ‘ao’ ‘ou’ ‘ui’ ‘ie’ ‘ue’ ‘an’ ‘en’ ‘in’ ‘un’ ‘ang’ ‘eng’ ‘ing’ ‘ong’ ‘zh’ ‘ch’ ‘sh’ \$ \_ \* }

不分大小写,在实验中同时与不加入多字符声韵母的字符表进行对比,即

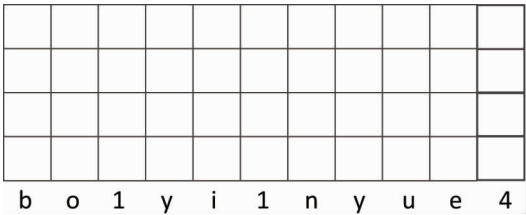
{abcdefghijklmnopqrstuvwxyz0123456789 \$ \_ \* }

在处理文本时,首先将文本转为拼音,拼音的声调在声母后用数字表示,再把拼音中的每个声韵母及数字根据字符表映射成稀疏向量,以“播音乐”短语为例,两种字符表的映射效果如图 2(a),图 2(b)所示。

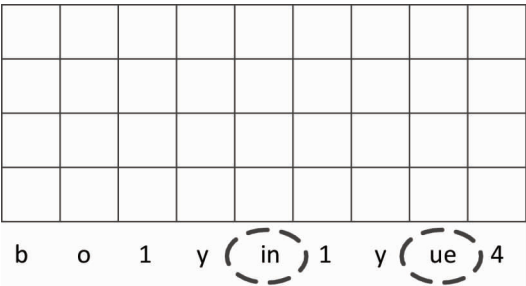
### 3 基于拼音特征的文本分类模型

深度神经网络在图像处理中已获得广泛应用,文本特征与图像特征的不同在于,图像中是利用如卷积神经网络提取其空间上的特征,而文本特征具有时间维即上下文信息丰富的特点,在对其处理时必须保留其前文信息。传统递归神经网络(RNN)

在处理长时间序列的过程中会出现梯度消失或梯度爆炸的现象<sup>[17]</sup>,改进的 GRU<sup>[18]</sup>网络单元可有效解决这一问题;同时,用卷积神经网络对文本高阶特征进行提取,在提高准确率的同时可以加快模型训练收敛速度。整个深度学习网络结构如图3所示,网



(a) 基于单字符映射



(b) 基于拼音中声韵母映射

图 2 拼音映射示例图

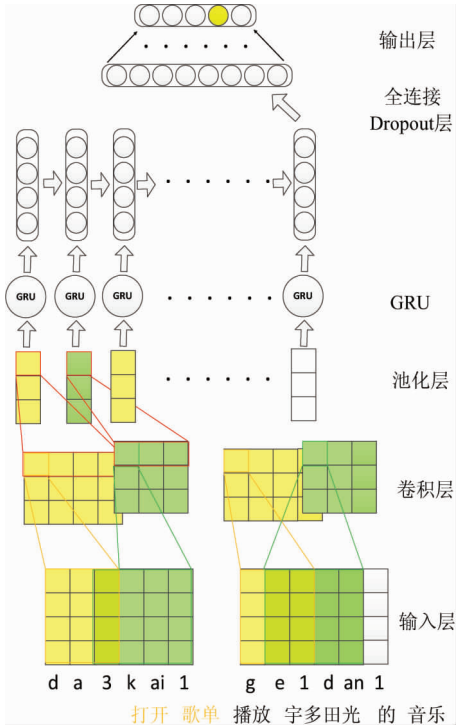


图 3 深度学习文本分类模型

络可分为输入层、CNN 层、池化层、门控递归单元 (GRU) 层、输出层。与通常对整句做卷积不同的是,在输入层中先对中文进行分词,再对每一个词的拼音进行映射得到的矩阵进行卷积操作,以得到每一个词语的卷积特征表示。对映射后长度小于 5 的词语对其进行补 0,使其长度变为 5,以避免在卷积中出现词语拼音长度小于卷积窗口大小的现象。

### 3.1 卷积层特征提取

模型中卷积层接受字符映射得到的向量作为输入,使  $\mathbf{x}^i \in \mathbb{R}^{57 \times l}$  表示指令语句中第  $i$  个词经字符映射后得到的矩阵,假设指令句中有  $n$  个词语,其矩阵表示为

$$\mathbf{x}^{1:n} = \mathbf{x}^1 \oplus \mathbf{x}^2 \oplus \cdots \oplus \mathbf{x}^n$$

令  $\mathbf{x}_{j:j+h-1}^i$  表示词中第  $j$  到  $j+h-1$  个字符映射构成的矩阵,用一个大小为  $\mathbf{w} \in \mathbb{R}^{57 \times h}$  的卷积窗口依次对每个单词做卷积得到新特征表示,例如特征  $c_j^i$  由词中  $\mathbf{x}_{j:j+h-1}^i$  得到

$$c_j^i = f(\mathbf{w} \cdot \mathbf{x}_{j:j+h-1}^i + b) \quad (1)$$

假设词  $\mathbf{x}^i$  中字符经映射后长度为  $l$ ,词  $\mathbf{x}^i$  得到的特征序列为:

$$\mathbf{c}^i = [c_1^i, c_2^i, \dots, c_{l-h+1}^i]$$

将得到的特征序列取最大池化,即:

$$m^i = \max\{c\}$$

得到对应词  $\mathbf{x}^i$  的拼音表示的卷积特征,整条指令的特征表示为

$$\mathbf{m} = [m^1, m^2, \dots, m^{l-h+1}]$$

Kim 在文献[5]中提出在卷积层中取变长度卷积窗口以提取不同抽象层次上的卷积特征,为测试其在本文所研究的应用背景中的影响,我们定义卷积窗口大小为  $h$ ,且定义: $h = \{2, 3\}$  表示取边长卷积窗口长度为 2 和 3。实验中在多个数据库上测试定窗口长度与不同组合的变窗口长度的结果进行对比。

### 3.2 递归神经网络

递归神经网络 (RNN) 是一种用来处理时间序列信息的网络模型,RNN 接受在当前时刻  $t$  的节点输出  $\mathbf{h}_t$  不仅接受当前时刻输入  $\mathbf{x}_t$ ,还受上一时刻节点输出  $\mathbf{h}_{t-1}$  的影响:

$$\mathbf{h}_t = f(\mathbf{W}\mathbf{x}_t + \mathbf{U}\mathbf{h}_{t-1} + \mathbf{b}) \quad (2)$$

$\mathbf{W} \in \mathbb{R}^{m \times n}$ ,  $\mathbf{x}_t \in \mathbb{R}^n$ ,  $\mathbf{U} \in \mathbb{R}^{m \times n}$ ,  $\mathbf{b} \in \mathbb{R}^m$ ,  $m$  为 RNN 输出维数。传统的 RNN 增加了时序信息,但对长时间序列处理的过程中会出现梯度消失或梯度爆炸的现象<sup>[15]</sup>。

GRU 是由 Cho<sup>[18]</sup>等于 2014 年提出的具有特殊结构的 RNN 单元,并将其用于短句级的文本建模取得良好效果<sup>[19]</sup>。与普通 RNN 单元不同的是,GRU 单元引入门的概念,且当前时刻的输出不仅受当前输入与上一时刻输出的影响,还受候补输出  $\tilde{\mathbf{h}}_t$  的影响,  $\mathbf{h}_t$  中第  $j$  个元素  $h_t^j$  具体计算式为

$$h_t^j = (1 - z_t^j) h_{t-1}^j + z_t^j \tilde{h}_t^j \quad (3)$$

$$z_t^j = \sigma(\mathbf{W}_z \mathbf{x}_t + \mathbf{U}_z \mathbf{h}_{t-1})^j$$

$$\tilde{h}_t^j = \tanh(\mathbf{W}_x \mathbf{x}_t + \mathbf{U}(\mathbf{r}_t \odot \mathbf{h}_{t-1}))^j$$

$$r_t^j = \sigma(\mathbf{W}_r \mathbf{x}_t + \mathbf{U}_r \mathbf{h}_{t-1})^j$$

其中  $\odot$  是逐元素乘法,GRU 通过重置门(reset gate)与更新门(update gate)将重要特征保留,在处理文本序列时使其重要的上下文关系不会丢失。

### 3.3 网络输出层

GRU 层最终得到隐层输出  $\mathbf{h}_t$ ,通过全连接层及 softmax 回归获得类别输出,其中加入 dropout,其作用是在训练中以概率  $1 - p$  随机使隐层节点变为 0,以抑制过拟合现象,输出表达式为

$$\mathbf{y} = \mathbf{W}(\mathbf{h}_t \odot \mathbf{r}) + \mathbf{b} \quad (4)$$

$\odot$  是逐元素乘法,  $\mathbf{r} \in \mathbb{R}^m$  且满足概率为  $p$  的伯努利分布,即

$$P(r_i = 1) = p$$

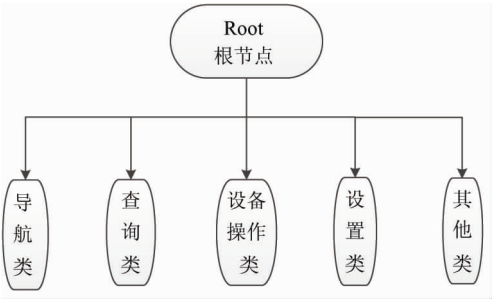
$$P(r_i = 0) = 1 - p$$

在训练时,梯度下降只会影响未被置 0 的节点。

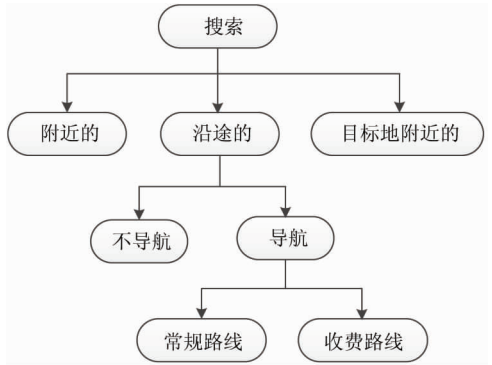
## 4 数据库与实验

实验中以无人驾驶车的语音导航为人机交互应用背景,分析其应用环境中的文本指令结构并分别构建简单意图和复杂意图语料库,简单意图为单一意图指令集,复杂意图是由数个单一意图组成的复合意图,实验中分别在两个语料库上进行交叉测试以验证模型在不同长度指令库上的表现。单意图语料库指令结构见图 4(a),以复意图指令库的搜索子

类为例,其指令结构如图 4(b) 所示,两个语料库的具体数据信息如表 1 所示。



(a) 单意图语料库指令结构



(b) 复意图搜索子类指令结构

图 4 两个语料库的意图指令结构

表 1 数据库统计信息

| 数据库  | 类别数 | 指令数  | 平均句长 |
|------|-----|------|------|
| 简单意图 | 53  | 8220 | 12.5 |
| 复杂意图 | 36  | 4816 | 25.7 |

4.1 模型间分类结果比较

将数据库随机分为 10 份,每次取 9 份进行训练,1 份测试,进行十折交叉验证。

首先,在实验中本文方法与文献[2,3]中的方法进行对比,为保证所有实验先验知识一致,本节实验都先对语料库进行同义词替换处理。模型中卷积层取卷积窗口大小为  $h = \{3, 4\}$ ,为保证学习到充分的特征,与文献[5]中相同每种窗口大小取 100 个卷积核,GRU 层取节点数为 128,Dropout 层取  $p = 0.5$ ,CNN 激活方式设为 softplus 函数,与传统 sigmoid 等激活方式相比,softplus 函数在训练过程中可有效加速模型收敛<sup>[20]</sup>,即

$$f(x) = \ln(1 + e^x) \tag{5}$$

代价函数取交叉熵函数,即

$$Cost = - \frac{1}{n} \sum_x [y \ln a + (1 - y) \ln (1 - a)]$$

其中  $y$  是期望输出,  $a$  是神经元实际输出。

其次,为测试 CNN 层特征提取对实验结果的影响,如果在模型中去掉 CNN 层及池化层,直接用 GRU 处理拼音经映射得到的向量序列,即如图 5 所示的网络结构,我们将这个模型称为 CHAR-GRU 模型,将图 3 中加入 CNN 及池化层的模型称为 CHAR-CNN-GRU 模型。

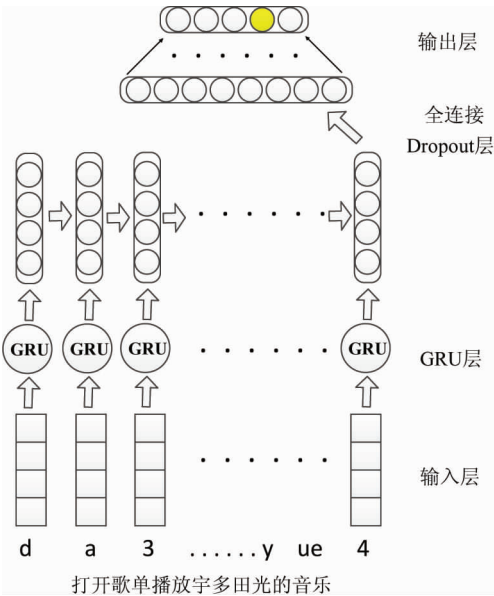


图 5 CHAR-GRU 模型

训练中设 batch-size 为 32,训练轮数为 200,取十次交叉验证的结果均值,在两个数据库上的结果见表 2。

从结果中可以看出本文所提出的网络模型与文献[2,3]中方法均有较大提升,且模型在长、短两个任务语料库上均有很高的分类精度。并且可以看出词向量在处理本文所构建的两个指令语料库时效果不理想,主要原因是利用大规模语料库训练所得到的词向量模型并不适合处理词汇量较少且方向应用性较强的语料库,而通过自身语料库训练得到的词向量模型因为语料规模较小,在训练时有效信息丢失严重。

表 2 多种分类模型在两个数据库上结果对比

| 模型           | 特征选取                            | 简单意图   | 复杂意图   |
|--------------|---------------------------------|--------|--------|
| CHAR-CNN-GRU | 基于声韵母的拼音特征(57 维)                | 95.21% | 94.40% |
|              | 无声韵母的拼音特征(39 维) <sup>[11]</sup> | 93.12% | 92.52% |
| CHAR-GRU     | 基于声韵母的拼音特征(57 维)                | 94.48% | 94.08% |
|              | 无声韵母的拼音特征(39 维) <sup>[11]</sup> | 93.47% | 91.71% |
| ME           | POS TF-IDF <sup>[2]</sup>       | 90.62% | 88.42% |
|              | 词向量                             | 71.05% | 72.51% |
| SDAE + SVM   | POS TF-IDF <sup>[3]</sup>       | 93.40% | 92.89% |
|              | 词向量                             | 74.69% | 72.72% |

其次,可以看出,基于声韵母成分构成的 57 维拼音特征比文献[11]中提出的不依据声韵母的 36 维普通拼音特征在两个任务上均有更优的表现,可以说明本文所提出的文本特征表示方法在处理中文语料库时有优秀的效果。

并且从表 2 中可以看出,CHAR-GRU 模型也可取得较高的分类结果,但两个模型在训练时的收敛速度有所不同,取两个模型在简单意图的数据库上训练中前150轮的代价函数值,训练趋势如图6

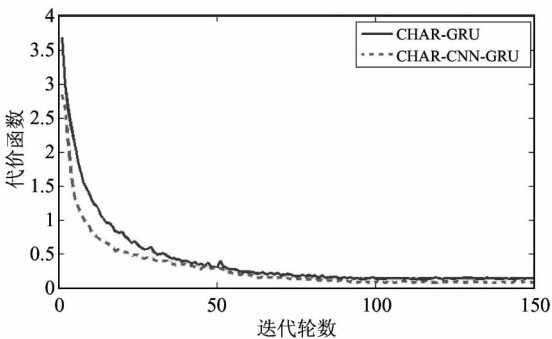


图 6 两种模型代价函数下降趋势

所示。可以看出模型在训练 80 轮左右后代价函数曲线稳定在很小的区间内,且 CHAR-CNN-GRU 模型在训练时有更快的收敛速度,且最终的代价函数值低于 CHAR-GRU 模型,CNN 层的加入可有效提取文本的高阶特征,提高模型分类精度且加快模型训练的收敛速度。

4.2 卷积窗口大小对模型影响

上一节实验中模型的卷积窗口大小均为  $h = \{3,4\}$ ,为比较模型中取不同卷积窗口大小对准确率的影响,现取窗口大小为不同组合,其余参数不变,在两个语料库上的测试结果如表 3 所示。首先,可以看出,与固定长度窗口相比,组合长度窗口可以在一定幅度上提高分类准确率;其次,在不同数据库上的最优组合情形并不是固定的,但通常情况应不大于 5,因为每个词语映射后长度有限,如窗口过大则卷积不能进行有效的特征提取。在不同的语料库任务中,尝试多种卷积窗口大小组合可以在一定程度上提高分类准确率,Zhang<sup>[21]</sup>等在其文章中在此方面进行了讨论。

表 3 不同窗口大小卷积组合测试结果

| 窗口组合<br>数据库 | $h = \{2\}$ | $h = \{3\}$ | $h = \{4\}$ | $h = \{5\}$ | $h = \{2,3\}$ | $h = \{3,4\}$ | $h = \{4,5\}$ | $h = \{3,4,5\}$ |
|-------------|-------------|-------------|-------------|-------------|---------------|---------------|---------------|-----------------|
| 简单意图        | 94.13%      | 94.43%      | 94.01%      | 93.11%      | 95.33%        | 95.21%        | 94.19%        | 94.32%          |
| 复杂意图        | 92.81%      | 93.14%      | 93.95%      | 91.20%      | 93.77%        | 94.40%        | 92.51%        | 93.87%          |

4.3 GRU 隐层单元数对模型影响

本节通过改变 GRU 隐层单元数来讨论其对模型效率的影响。以简单意图数据库测试,其余参数设置与 4.1 节中实验相同,仅改变 GRU 节点数  $k$ ,

取  $k = 32,64,128,256$ ,测试准确率与每轮平均训练时间如表 4 所示。



表 4 GRU 节点数影响

|      | $k = 32$ | $k = 64$ | $k = 128$ | $k = 256$ |
|------|----------|----------|-----------|-----------|
| 准确率  | 77.49%   | 88.53%   | 95.21%    | 95.29%    |
| 时间/轮 | 15.8s    | 33.7s    | 65.3s     | 133.9s    |

可以看出 GRU 的隐层单元个数对模型效率有明显的影响,如 GRU 隐层单元数过小,则分类精度会下降,如过大则会影响训练速度,根据任务语料库的规模、语句的长短选择适当的 GRU 节点数  $k$ ,平衡好训练时间与训练精度的关系,对构建一个高效模型非常重要。

5 结 论

本文以字符为特征,总结了英文单词与汉语拼音的区别,结合汉语拼音的特点,提出了一种针对中文语料库分类的基于拼音声韵母字符的新特征,大大减少了特征维度且不需要对文本做过多处理只需将文本转化为拼音,并利用递归神经网络单元与 GRU 在处理时间序列样本上的优势,提出了一种结合 CNN 与 GRU 的深度学习文本分类模型,对人机对话应用场景下的不同长度的指令语料库进行文本意图分类,并与多种分类模型对比。实验表明,该模型在人与机器人对话过程中意图识别准确率有明显提高。

参考文献

[ 1 ] 俞凯,陈露,陈博等. 任务型人机对话系统中的认知技术——概念、进展及其未来. 计算机学报, 2015, 38 (12):2333-2348

[ 2 ] 王文,赵群飞,朱特浩. 人-服务机器人交互中自然语言理解研究. 微型电脑应用, 2015, (3): 45-49

[ 3 ] 李瀚清,房宁,赵群飞等. 利用 SDAE 深度学习的指令意图理解方法. 上海交通大学学报, 2016, 50 (7): 1102-1107

[ 4 ] 奚雪峰,周国栋. 面向自然语言处理的深度学习研究. 自动化学报, 2016, 42 (10):1445-1465

[ 5 ] Kim Y. Convolutional neural networks for sentence classification. In: Proceeding of the 2014 Conference on Empirical Methods in Natural Language Processing, Doha, Qatar, 2014. 1-6

[ 6 ] Kim Y, Jernite Y, Sontag D. Rush Alexander M. Character-aware neural language models. arXiv preprint arXiv:1508.06615

[ 7 ] Kalchbrenner N, Grefenstette E, Blunsom P. A convolutional neural network for modelling sentences. In: Proceeding of the Association for Computational Linguistics, Baltimore, USA, 2014. 655-665

[ 8 ] Dos Santos N, Gatti M. Deep convolutional neural networks for sentiment analysis of short texts. In: Proceeding of the 25th International Conference on Computational Linguistics, Dublin, Ireland, 2014. 69-78

[ 9 ] Xi O Y, Zhou P, Li C H, et al. Sentiment analysis using convolutional neural network. In: Proceeding of the IEEE International Conference on Computer and Information Technology, Liverpool, UK, 2015. 2359-2364

[10] Tang D Y, Qin B, Liu T. Document modeling with gated-recurrent neural network for Sentiment Classification. In: Proceeding of the 2015 Conference on Empirical Methods in Natural Language Processing, Lisbon, Portugal, 2015. 1422-1432

[11] Zhang X, Zhao J B, Lecun Y. Character-level convolutional networks for text classification. In: Proceeding of the Neural Information Processing Systems, Montreal, Canada, 2015. 647-657

[12] Zhang X, Lecun Y. Text understanding from scratch. arXiv preprint atarXiv: 1502.01710

[13] Conneau A, Schwenk H, Lecun Y. Very deep convolutional networks for natural language processing. arXiv preprint arXiv: 1606.0178

[14] Lai S W, Xu L H, Liu K, et al. Recurrent convolutional neural networks for text classification. In: Proceeding of the Conference on Artificial Intelligence, Austin, USA, 2015. 2267-2273

[15] Mikolov T, Chen K, Corrado G, et al. Efficient estimation of word representations in vector space. arXiv preprint atarXiv: 1301. 3781

[16] Levy O, Goldberg Y. Dependency-based word embeddings. In: Proceeding of the Association for Computation Linguistics, Dublin, Ireland, 2014, 2:302-308

[17] Bengio Y, Simard P, Frasconi P. Learning long-term dependencies with gradient descent is difficult. *IEEE transactions on Neural Networks*, 1994, 5 (2):157-166

[18] Chung J, Gulcehre K, Cho K, et al. Gated feedback re-

- current neural networks. *Computer Science*, 2015. 2067-2075
- [19] Chung J, Gulcehre K, Cho K, et al. Empirical Evaluation of gated recurrent neural networks on sequence modeling. 2015. arXiv preprint at arXiv:1412.3555
- [20] Zheng H, Yang Z L, Liu W J, et al. Improving deep neural networks using softplus units. In: Proceeding of the International Joint Conference on Neural Networks, Killarney, Ireland, 2015. 1-4
- [21] Zhang Y, Byron C W. A sensitivity analysis of (and practitioners' guide to) convolutional neural networks for sentence classification. arXiv preprint arXiv:1510.03820

## A deep learning model for text classification using phonetic features

Zhao Boxuan<sup>\*</sup>, Fang Ning<sup>\*\*</sup>, Zhao Qunfei<sup>\*</sup>, Zhang Pengzhu<sup>\*\*\*</sup>

(<sup>\*</sup> Laboratory of System Control and Information Processing, Department of Automation,  
Shanghai Jiaotong University, Shanghai 200240)

(<sup>\*\*</sup> School of Humanities, Shanghai Jiaotong University, Shanghai 200030)

(<sup>\*\*\*</sup> Antai College of Economics & Management, Shanghai Jiaotong University, Shanghai 200030)

### Abstract

A deep learning model using the features of the consonant and vowel in Chinese Pinyin was proposed for the intention texts speech recognized in human-robot voice interaction. Firstly, by taking unmanned vehicle voice navigation as the application scenarios of human-robot interaction, the intention text structure was analyzed, and a single intention corpus and a complex intention corpus were built respectively; Secondly, based on the character-level features in text classification, a feature representation method using consonant and vowel in Pinyin for Chinese text classification was proposed with considering the differences between Chinese Pinyin and English words. Thirdly, traditional recurrent neural network (RNN) units were replaced by gated recurrent units (GRU) for the problem of difficulties in capturing long-term dependencies. To extract high-level features, shorten the length of feature sequences and increase the convergence rate of the model, a deep learning model combining the convolutional neural network (CNN) with the GRU-RNN was established. Finally, to evaluate the performance of the model on short and long sequence tasks, 10-fold cross validations were implemented on corpuses for two tasks respectively, and then the comparisons and analysis were carried out against other classification methods. The result shows that the proposed model can significantly improve the accuracy of classification for the intention texts.

**Key words:** text classification, intention understanding, features of consonant and vowel, gated recurrent units (GRU)