

# 一种结合话语伪标签注意力的人机对话意图分类方法

丁健德, 黄沛杰\*, 许嘉宝, 彭佑铭

华南农业大学, 数学与信息学院, 广东广州, 510642

bighead@stu.scau.edu.cn, pjhuang@scau.edu.cn, xujiabao@stu.scau.edu.cn,  
ympengcoder@stu.scau.edu.cn

## 摘要

在人机对话中, 系统需要通过意图分类判断用户意图, 再触发相应的业务类型。由于多轮人机对话具有口语化、长文本和特征稀疏等特点, 现有的文本分类方法在人机对话意图分类上还存在较大困难。本文在层次注意力网络(hierarchical attention networks, HAN)基础上, 提出了一种结合话语伪标签注意力的层次注意力网络模型PLA-HAN (HAN with utterance pseudo label attention)。PLA-HAN通过优选伪标签集、构建单句话语意图识别模型以及设计话语伪标签注意力机制, 识别单句话语意图伪标签, 并计算话语伪标签注意力。进而将单句话语伪标签注意力嵌入到HAN的层级结构中, 与HAN中的句子级别注意力相融合。融合了单句话语意图信息的句子级注意力使模型整体性能得到进一步的提升。我们在中国中文信息学会主办的“客服领域用户意图分类评测比赛”的评测语料上进行实验, 实验结果证明PLA-HAN模型取得了优于HAN等对比方法的意图分类性能。

**关键词:** 意图分类; 伪标签注意力; 层次注意力网络; 人机对话

## A Human-machine Dialogue Intent Classification Method using Utterance Pseudo Label Attention

Jiande Ding, Peijie Huang\*, Jiabao Xu, Youming Peng

College of Mathematics and Informatics, South China Agricultural University, China  
bighead@stu.scau.edu.cn, pjhuang@scau.edu.cn, xujiabao@stu.scau.edu.cn,  
ympengcoder@stu.scau.edu.cn

## Abstract

In human-machine dialogue system, it needs to judge the user's intent through the intent classification, and then triggers the corresponding business type. Due to the characteristics of colloquialization, longer texts and sparse features of multi-turn dialogues, the existing classification methods still have great difficulties in the classification of human-machine dialogue intent. Based on hierarchical attention networks (HAN), we propose PLA-HAN model that combines the utterance pseudo-label attention. Through selecting utterance intent set, constructing utterance intent detection model and designing an utterance pseudo-label attention mechanism, PLA-HAN recognizes the pseudo-label of utterance intent and then computes utterance pseudo-label attention. Furthermore, the utterance pseudo-label attention is embedded into

\*通讯作者

©2020 中国计算语言学大会

根据《Creative Commons Attribution 4.0 International License》许可出版

the hierarchical structure of HAN and is integrated with its sentence-level attention. Sentence-level attention that incorporates utterance intent information further improves the overall performance of the model. We conducted experiments on the shared task dataset of “Customer Intent Classification Evaluation Competition for Customer Service Domain” sponsored by the Chinese Information Processing Society of China. Experiment results show that the proposed model achieved better performance than HAN on dialogue intent classification.

**Keywords:** Intent classification , Pseudo label attention , Hierarchical attention network ,

## 1 引言

近年来，人机对话由于其潜在的发展潜力和诱人的商业价值而收到越来越多的关注 (俞凯等, 2015; Chen et al., 2017)。意图分类是人机对话中的基础任务。在人机对话领域中，通常的做法是，先将用户语音通过语音识别技术转换成文本，然后再通过意图分类算法将其识别到相应类型下的具体意图。这样可以识别出用户输入到对话系统执行某个动作之间的一个映射关系，达到识别和理解用户要表达的意图的目的。在人机对话系统中，系统通过意图分类判断用户意图，再触发相应的业务类型。在面向智能客服的意图分类任务中，业务类型通常包括查询类、咨询类、办理类和投诉类等(在本文的意图分类数据上，查询类和咨询类合并成一个类型),每个业务类型下还有多种不同的用户意图。表1给出了一个例子，表示客服和用户的语音对话记录，以及对应的意图类别。

序号	对话段
1	客服: 您好请讲
2	用户: 我问一下就是那个我我在那个中国移动这个官方网页上看到的是六十元可以充值呢十二g那个流量哦坐车电话
3	客服: 现在准确的但是前提那个得是一百三十八以上的套餐才能办理而且我这没有套餐的能办理什么嗯这个只能办理正常包月流量一号
4	用户: 包月咋办的
5	客服: 十元一百兆二十元三百兆三十元五百兆四十元七百兆五十元一个g的这种
6	用户: 哦那就不能换
7	客服: 对那种的办理不了那种前提您套餐资费也比较高的您得先改成一百三十八以上才可以办理的
8	用户: 我这卡要变成那个国通卡最低消费多钱的
...	.....
21	客服: 请问您还有其他问题需要咨询吗
22	用户: 六
23	客服: 哎好了祝您生活愉快再见好再见哎
业务类型: 咨询(含查询)	
用户意图: 营销活动信息	
类别合并: 咨询(含查询)-营销活动信息	

Table 1: 一个意图分类例子

人机对话意图分类属于文本分类任务。已经有很多文本分类的研究致力于提高分类器性能，早期的典型代表是SVM和最大熵等 (Haffner et al., 2003; Phan et al., 2008)。随后，深度学习在自然语言处理(natural language processing, NLP)中受到关注，主流的应用包括深度信念网络(deep belief networks, DBN) (Sarikaya et al., 2011)、CNN (Xu and Sarikaya, 2013; Kim, 2014)和RNN (Xu and Sarikaya, 2014)等，尤其是RNN中最常用的LSTM (Cheng et al., 2016; Ravuri and Stolcke, 2016; Vu et al., 2016; 柯子等, 2018)。近年来，注意力机制被引入到了NLP中，实验证明其善于在文本分类任务中抽取文本的含义，例如话语意图检测 (Liu and

Lane, 2016)、话语领域分类 (Kim et al., 2018)、问答情感分类 (安明慧等, 2019)和针对语篇长文本的文档分类 (Yang et al., 2016)等。尽管基于深度学习的分类模型, 尤其是层次注意力网络(hierarchical attention network, HAN) (Yang et al., 2016), 将长文本的分类性能提高到一个新的水平, 但对于同时具有口语化、长文本和特征稀疏的多轮人机对话段文本, 其意图分类仍然存在较大挑战。如图1的真实客服对话段例子可以看到, 人类对话文本以寒暄和多轮询问式对话为主, 文本通常只包含极少实际有意义的词语, 导致内容具有特征稀疏性问题, 增加了抽取有效特征的难度。此外, 由于用户不同的口音和语义表达偏好, 以及容易出现的不规则特征词和未登录词, 也加剧了语音识别的错误比例, 进一步给对话段的意图分类带来困难。尤其是长的对话段(如本文实验里的数据集, 对话段的平均话语数达到20句左右)挑战更大。

针对上述挑战, 经过分析和借鉴人类理解对话段的经验, 我们发现每个单句话语所表达的意图也是一种十分重要的信息, 这些信息对于整体对话段的意图理解有着举足轻重的作用。对于人机对话意图分类任务而言, 标注数据仅为整个对话段的意图标签, 参加到对话段意图识别的单句话语的意图标签只能靠额外构建的单句话语意图识别模型预测得到, 有别于真实标注的标签, 本文称之为伪标签(pseudo label)。本文提出了一种结合单句话语伪标签注意力的层次注意力网络模型PLA-HAN (HAN with utterance pseudo label attention)。PLA-HAN通过构建单句话语意图识别模型和设计话语伪标签注意力机制, 识别单句话语意图伪标签, 并计算话语伪标签注意力。进而将单句话语伪标签注意力嵌入到HAN的层级结构中, 与HAN中的句子级别注意力相融合。此外, 我们还对伪标签集进行了优选, 选用了和人机对话任务关联度高的单句话语意图伪标签。融合了单句话语意图信息的句子级注意力使模型整体性能得到进一步的提升。我们在中国中文信息学会主办的“客服领域用户意图分类评测比赛”的评测语料上进行实验, 实验结果证明PLA-HAN模型取得了优于HAN的意图分类性能。

本文的其余部分组织如下: 在第2节, 我们介绍了相关的基础工作, 包括层次注意力网络HAN以及单句话语意图识别模型。第3节介绍本文提出的方法, 包括我们设计的伪标签注意力机制, 以及结合伪标签注意力的人机对话意图分类模型。第4节是实验和分析。第5节简要总结了本文的工作。

## 2 基础工作

### 2.1 层次注意力机制

更符合长文本层级结构的是HAN(Yang et al., 2016)所采用的层次注意力机制, 它带有两个层级的注意力机制, 分别是词级别和句子级别, 能够更好地表示长文本中重要信息的位置。

**注意力机制。**近年来, 在文本分类问题上, 基于注意力机制的模型在效果和效率上都展现出了一定的优越性。我们考虑非层级结构的软注意力机制以及层级结构的分层注意力机制。

注意力机制最早在机器翻译中应用, 现在已经成为神经网络相关中一个十分具有影响力的概念, 注意力机制相当于模仿人类将大量视觉信息压缩成描述性语言的非凡能力(Xu et al., 2015), 它将编码器的输出映射为注意力权重, 将权重与编码器输出进行加权, 如式(1)-(3)所示。

$$u_i = \text{score}(h_i) \quad (1)$$

$$\alpha_i = \frac{\exp(u_i)}{\sum_i^L (u_i)} \quad (2)$$

$$h_* = \sum_i^L \alpha_i h_i \quad (3)$$

其中, 本文采用前馈神经网络为对齐函数(score)、在语音识别(Graves et al., 2013)和口语语言理解(Xu and Sarikaya, 2014)领域有着非常成功应用的双向LSTM作为编码器。通过对齐函数将编码器的输出对齐, 得到注意力分数, 进行归一化后得到注意力权重 $\alpha$ , 最终将编码器输出与注意力权重进行加权。

**词级别。**对于一个句子的词 $w^*$ , 在经过embedding之后, 获得词向量 $w_{emb}^*$ , 编码器对词向量进行编码, 获得词级别的表达 $h_{word}^*$ , 通过单层MLP对 $h_{word}^*$ 进行对齐计算, 获

得 $u_{word}^*$ ，使用softmax函数得到字级别的注意力权重 $\alpha_{word}^*$ ，在这之后，通过字级别注意力权重 $\alpha_{word}^*$ 与 $h_{word}^*$ 加权求和得到句子向量 $s^*$ 。如式(4)-(8)所示。

$$w_{emb}^* = embedding(w^*) \quad (4)$$

$$h_{word}^* = Encoder(w_{emb}^*) \quad (5)$$

$$u_{word}^* = \tanh(W_w h_{word}^* + b_w) \quad (6)$$

$$\alpha_{word}^* = \frac{\exp(u_{word}^* u_w)}{\sum_t \exp(u_{word}^* u_w)} \quad (7)$$

$$s^* = \sum_t (\alpha_{word}^* word_{emb}^*) \quad (8)$$

**句子级别。**与词级别的做法相似，编码器对句子向量 $s^*$ 进行编码获得句子向量的表达 $h_{sentence}^*$ ，同样通过单层MLP对句子向量进行对齐计算，获得 $u_{sentence}^*$ ，使用softmax函数得到句子级别的注意力权重 $\alpha_{sentence}^*$ ，最终通过句子级别注意力权重 $\alpha_{sentence}^*$ 与 $h_{sentence}^*$ 加权求和作为文档向量 $v$ 。如式(9)-(12)所示。

$$h_{sentence}^* = Encoder(s^*) \quad (9)$$

$$u_{sentence}^* = \tanh(W_s h_{sentence}^* + b_s) \quad (10)$$

$$\alpha_{sentence}^* = \frac{\exp(u_{sentence}^* u_s)}{\sum_t \exp(u_{sentence}^* u_s)} \quad (11)$$

$$s^* = \sum_t (\alpha_{sentence}^* sentence_{emb}^*) \quad (12)$$

## 2.2 单句话语意图识别模型

对于单句话语的意图识别，我们采用了基于BERT的双向LSTM模型，BiLSTM结构在语音识别(Graves et al., 2013)和口语语言理解(Xu and Sarikaya, 2014)领域有着非常成功的应用。在BERT层，我们采用预先训练好的中文BERT模型对固定长度为L的话语序列 $w^* = \{w_1, w_2, \dots, w_L\}$ 进行编码，每个字 $w_i$ 会被编码为字向量 $e_i$ ；这样 $w^*$ 就被编码成了 $emb_u = \{e_1, e_2, \dots, e_L\}$ 。如式(13)所示。

$$emb_u = BERT(w^*) \quad (13)$$

在BiLSTM网络层，我们将 $emb_u = \{e_1, e_2, \dots, e_L\}$ 输入到双向的长短时记忆网络，然后分别从网络中得到正向的输出 $h^{fw} = \{h_1^{fw}, h_2^{fw}, \dots, h_L^{fw}\}$ 和反向的输出 $h^{bw} = \{h_1^{bw}, h_2^{bw}, \dots, h_L^{bw}\}$ 。如式(14)所示。

$$h^{fw}, h^{bw} = BiLSTM(emb_u) \quad (14)$$

我们将正向和反向的结果拼接起来得到双向LSTM的输出 $h$ 。如式(15)所示。

$$h = [h^{fw}, h^{bw}] \quad (15)$$

对于输入序列中的每个元素，每个LSTM结构计算以下函数，其中 $h_t$ 是t时刻的隐藏层， $c_t$ 是t时刻的记忆单元， $x_t$ 是t时刻的输入， $h_{t-1}$ 是t-1时刻的隐藏层， $i_t$ 、 $f_t$ 、 $\tilde{c}_t$ 和 $o_t$ 分别

是输入门、遗忘门、记忆单元门和输出门， $\sigma$ 是sigmoid函数， $\odot$ 是哈达玛积。如式(16)-(21)所示。

$$i_t = \sigma(W_{ii}x_t + b_{ii} + W_{hi}h_{t-1} + b_{hi}) \quad (16)$$

$$f_t = \sigma(W_{if}x_t + b_{if} + W_{hf}h_{t-1} + b_{hf}) \quad (17)$$

$$\tilde{c}_t = \tanh(W_{ig}x_t + b_{ig} + W_{hg}h_{t-1} + b_{hg}) \quad (18)$$

$$o_t = \sigma(W_{io}x_t + b_{io} + W_{ho}h_{t-1} + b_{ho}) \quad (19)$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t \quad (20)$$

$$h_t = o_t \tanh(c_t) \quad (21)$$

接着将 $h$ 累加得到话语序列的句子向量 $s$ ，最后接上全连接-softmax层得到各个意图类别的概率输出。如式(22)所示。

$$y = \text{softmax}(Wh + b) \quad (22)$$

### 3 本文的方法

在本节中，我们将介绍如何将意图伪标签模型学到的伪标签注意力权重和层级注意力权重相融合，从而构成层级结构中嵌入单句话语意图伪标签注意力的机制。

#### 3.1 伪标签注意力

在面向长文本的任务型意图分类场景中，能够提取到长文中的关键单句话语将会对整体的分类效果带来提升。除了在模型内部采用注意力机制来实现，也可以通过外部给模型带来额外的信息以更好地提取到关键单句话语。在此，我们通过单句话语的意图伪标签来实现将外部信息注入模型结构中。长文本的每个单句话语意图伪标签的加入，可以视为构建了一个内部的“对话意图-伪意图标签”的分布，反映总体对话所表达的意图倾向。我们通过一个注意力模型学习到了每个伪标签所对应的注意力，这种注意力作为一种额外的信息，反映了其对应的单句话语对于总体对话意图的重要程度，利用该信息可以让模型更好地选择出重要的单句话语。

通过相关联的单句话语意图任务，我们训练了一个针对单句话语的意图识别模型(详见2.2节)，记为 $model_{intent}$ ，通过此模型，给主任务中的每个子句标注上意图伪标签 $intent^{pseudo}$ ，使用one-hot对意图伪标签进行独热编码，获得意图伪标签的文本表示 $intent_{emb}^{pseudo*}$ ，经过编码器得到 $intent_h^{pseudo*}$ ，在此处采用软注意力来获取注意力权重 $\beta_{pseudo}$ (详见2.1节的注意力机制部分)，记为伪标签意图注意力，通过伪标签意图注意力可以反映出“对话意图-伪意图标签”的分布，从而表达单句话语意图的重要程度。如式(23)-(25)所示。

$$intent_{emb}^{pseudo*} = \text{OneHot}(intent^{pseudo}) \quad (23)$$

$$intent_h^{pseudo*} = \text{Encoder}(intent_{emb}^{pseudo*}) \quad (24)$$

$$\beta_{pseudo} = \text{Attention}(intent_h^{pseudo*}) \quad (25)$$

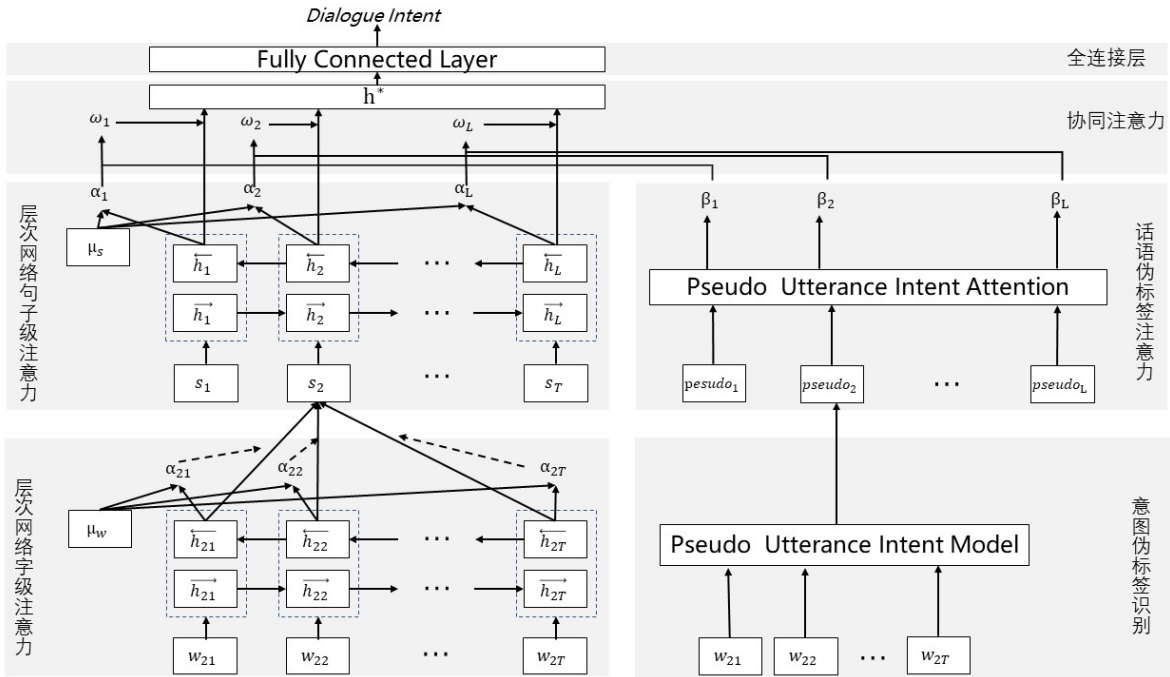


Figure 1: PLA-HAN模型的架构

### 3.2 结合伪标签注意力的人机对话意图分类

本文的模型是在HAN模型基础上融入伪标签注意力构成的话语伪标签注意力层次注意力网络PLA-HAN，模型采用BERT编码 (Devlin et al., 2019)，由字级层、句子级层、协同注意力层和全连接层组成，如图1所示。

**BERT编码。**我们采用中文BERT预训练模型对字进行编码。固定长度为L的序列输入，通过BERT编码后将会被编码为(L, 768)的向量。

**PLA-HAN字级层。**包括了HAN模型中的字级注意力网络(图1的左下部分)和单句话语的意图伪标签识别模型(图1的右下部分)。

(1) HAN的字级注意力：在HAN中，对于经过BERT编码的篇章结构作为输入计算得到词级别的注意力 $\alpha_{word}$ (详见2.1节)。

(2) 意图伪标签识别：通过单句话语意图识别模型，给对话段中的每个话语打上伪意图标签，此时得到对话段的“对话意图-伪标签意图”分布。

**PLA-HAN句子级层。**包括了HAN模型中的句子级注意力网络(图1的左中部分)和单句话语的意图伪标签注意力机制(图1的右中部分)。

(1) HAN的句子级注意力：在HAN中，在句子向量的基础上进一步计算句子级别的注意力 $\alpha_{sentence}$ (详见2.1节)。

(2) 话语伪标签注意力：同时，我们采用3.1节中介绍的方法，通过额外构建的注意力模型，将“对话意图-伪标签意图”分布转化为伪标签注意力 $\beta_{pseudo}$ 。

**协同注意力层。**我们采用加法将分层注意力机制的句子级注意力和伪标签注意力相融合，记为 $\omega_i$ 。如式(26)所示。

$$\omega_i = \alpha_{sentence} + \beta_{pseudo} \quad (26)$$

最后我们利用注意力权重对分层注意力结构中的句子向量 $s^*$ 进行加权缩放，并将其累加得到的 $h^*$ 作为注意力层的输出。如式(27)所示。

$$h^* = \sum_t (\omega_i * s_i) \quad (27)$$

**全连接层。** 我们利用全连接层将模型的输出映射为相应意图类别数量，接着使用softmax输出各个意图类别的概率，最终采用概率最高的意图类别作为输出。公式如式(28)-(29)所示。

$$y^* = \text{softmax}(W * h^* + b) \tag{28}$$

$$\text{predict} = \text{argmax}(y^*) \tag{29}$$

### 3.3 伪标签的挑选

对于单句话语义图模型所带来的意图伪标签，并不是所有的伪标签都能对人机对话段的意图分类产生正面效果。为此，本文的方案中，通过全部伪标签的初步试验，计算出每个伪标签在对话段中的平均覆盖率：

$$pseudo_{cover} = \text{意图伪标签在对话段中的出现次数} / \text{对话段的对话轮次} \tag{30}$$

然后，筛选出覆盖率高的伪标签：

$$pseudo_{select} = \text{topk}(pseudo_{cover}) \tag{31}$$

在覆盖率分析的基础上，我们进一步从语义角度挑选出高关联度的伪标签嵌入到模型中，去掉了部分关联度不高的标签。在本文的实验中，用以训练单句话语义图识别模型的数据集共有48种单句话语义图类别。经过伪标签挑选，我们的最终方案中，选取了25种与人机对话意图分类任务高关联度的单句话语义图，同时将其它的单句话语义图标签标记为“其它”。不同伪标签集的实验效果见4.5小节。

## 4 实验

### 4.1 数据集

本文的人机对话实验数据来自于中国中文信息学会主办的“客服领域用户意图分类评测比赛”<sup>1</sup>，属于客服领域对话文本，视为多轮话的长文本。同时这项比赛还有一个子任务是单句用户话语的自然语言理解(包括意图识别和槽填充)。本文将其中的话语意图识别任务视为本文人机对话任务的辅助任务，用于产生人机对话段中每一句单句话语的意图伪标签。

**人机对话数据集** 人机对话数据集为2万条真实客服对话段标注数据，此数据集中共有35种人机对话意图类别，表2给出了业务类型与用户意图的种类。我们按照8: 2的比例分别划分训练集和测试集。训练时，再从训练集中划分出20%作为验证集。

**单句话语义图识别数据集** 单句话语义图的数据集为2万条真实的单句话语义图数据，此数据集中共有48种单句意图类别，如表3所示。经过伪标签优选，选取了与对话段意图分类任务相关度高的25种单句意图类别，其它类别的伪标签标记为“其它”。按照8: 2切分训练集和验证集，不设置测试集。

业务类型	对话意图
咨询(含查询)	业务订购信息查询、业务规定、业务订购信息查询、业务资费、产品/业务功能、使用方式、办理方式、号码状态、宽带覆盖范围、工单处理结果、工单处理结果、服务渠道信息、用户资料、电商货品信息、营销活动信息、账户信息
办理	下载/设置、停复机、取消、变更、开通、打印/邮寄、移机/装机/拆机、缴费、补换卡、重置/修改/补发、销户/重开
投诉(含抱怨)	不知情定制问题、业务使用问题、业务办理问题、业务规定不满、信息安全问题、服务问题、网络问题、营销问题、费用问题

Table 2: 业务类型种类与35种对话意图类别

<sup>1</sup><http://www.cips-cl.org/static/CCL2018/call-evaluation.html>

单句话语意图标签
查询、查询套餐余量、查询充值缴费记录、查询流量、查询本机号码、查询本机业务、查询余额、查询短信、查询积分、查询语音、查询月初扣费、查询账单、查询宽带、查询手机、咨询宽带、具实帮助、具实返回、具实转人工、具实退出、具实重听、具实业务列表、具实转ivr、办理套餐、办理手机充值、取消流量、预约宽带、修改宽带、开通流量、重置、重置手机、重置宽带、GPRS、拒识、修改、确认、手机、修改手机、短信、宽带、流量、empty、集外说法批评,抱怨,脏话、集外说法机器人、集外说法集外业务、咨询、集外说法结束、集外说法短拒识、集外说法友好问候、集外说法感谢

Table 3: 48种单句话语意图类别

#### 4.2 实验设置

人机对话意图分类实验 batch\_size设置为16, epoch设置为20, Dropout设置为0.1, 采用Adam优化器, 学习率为1e-5, 非层级模型设置最大句子长度为512, 层级模型设置最大句子数量为25, 最大句长为25。

话语伪标签预测实验 batch\_size设置为32, epoch设置为100, Dropout设置为0.1, 采用Adam优化器, 学习率为1e-3, 非层级模型设置最大句子长度为30。

#### 4.3 对比方法

本文提出的PLA-HAN模型将与以下代表性的基线方案进行比较, 为了公平比较, 全部模型都采用了BERT编码:

- BERT FineTune: 该方法将BERT fine-tuning (Devlin et al., 2019)应用到分类任务, 在BERT分类层增加了一个新的输出层。
- BERT BiLSTM: 该方法是文本分类的经典基线 (Vu et al., 2016), 适合于序列问题。
- BERT SoftAtt: 该方法采用了Liu等(Liu and Lane, 2016)在ATIS数据集的话语意图识别的BiLSTM模型中采用的软注意力。
- BERT HAN: Yang等(Yang et al., 2016)提出的更加适合篇章结构文本的层次注意力模型结构。

#### 4.4 整体性能

我们提出的PLA-HAN模型与几种对比方法进行比较, 包括BERT FineTune、BiLSTM、带软注意力的BiLSTM以及层次注意力模型HAN。实验结果如表4所示。

模型类别	模型	意图分类正确率(%)			
		总体	咨询(含查询)	办理	投诉
长文本结构	BERT FineTune	53.11	49.50	66.72	48.14
	BERT BiLSTM	55.75	52.39	69.83	49.35
	BERT SoftAtt	55.87	51.98	71.03	49.91
层次结构	BERT HAN	56.31	52.55	71.08	50.40
	<b>BERT PLA-HAN</b>	<b>56.94</b>	<b>53.33</b>	<b>71.26</b>	<b>51.35</b>

Table 4: 不同模型性能对比

从表4的结果可以看到:

(1) 在以整篇长文本直接作为输入的模型中, 通过加入编码器BiLSTM和注意力机制的应用, 可以在BERT的基础上进一步提升对话段意图分类的性能。

(2) 相比于整篇长文本结构的输入, 层次结构模型取得了更好的分类性能。这一方面得益于不受BERT输入长度的限制, 另一方面也由于分层次的注意力能更好地建模字到句子再到对话段的语义结构。

(3) 融合了单句话语伪标签注意力的PLA-HAN取得了最好的性能, 优于HAN模型。



## 4.5 进一步分析

通过比较以上结果可以看出PLA-HAN模型取得了良好的性能，我们也想进一步探究模型能有所提升的原因。我们首先分析了不同的伪标签集对模型性能的影响。然后，我们给出了一个不同长度对话段情况下的PLA-HAN模型与基础的HAN模型的性能对比的定量分析。

**不同伪标签集对模型性能的影响。**为了研究不同伪标签集在PLA-HAN模型中效果，我们对采用三种不同的伪标签集的BERT PLA-HAN的实验结果进行观察。PLA-HAN(All)代表不做选择采用了全部的48种伪标签，PLA-HAN(Fit)代表只选用了和人机对话意图相关的32种伪标签，PLA-HAN(Select)代表进一步优选的25种伪标签。结果如表5所示。

模型	意图分类正确率(%)			
	总体	咨询(含查询)	办理	投诉
BERT PLA-HAN(All)	56.06	52.48	70.41	49.95
BERT PLA-HAN(Fit)	56.59	53.31	70.94	50.72
BERT PLA-HAN(Select)	<b>56.94</b>	<b>53.33</b>	<b>71.26</b>	<b>51.33</b>

Table 5: 不同伪标签集的PLA-HAN模型性能对比

结果表明，我们的伪标签集选择策略是必要的，不同的伪标签集对模型整体性能存在明显的影响。详细的分析如下：

- PLA-HAN(All): 在此方案中，我们采用了所有48种伪标签。其中部分伪标签实际上和人工对话意图分类任务中的35种意图的相关性并不强。从实验结果可以看到，分类性能都不够好，甚至都略微低于HAN模型。
- PLA-HAN(Fit): 在此方案中，我们选取与人工对话任务的意图有较高覆盖率的伪标签嵌入到模型中，将基本不相关的标签标记为“其它”。实验中通过验证集选取了32个伪标签。从实验结果可以看到，分类性能相比于PLA-HAN(All)有了明显提高，相同时也优于HAN模型。
- PLA-HAN(Select): 在此方案中，我们在PLA-HAN(Fit)的基础上，进一步从语义角度挑选出高关联度的伪标签嵌入到模型中，去掉了部分关联度不高的标签，一共选取了25种伪标签。从实验结果看，PLA-HAN(Select)取得了最好的性能。

**不同长度对话段的模型性能对比。**我们进一步对比不同对话段长度情况下的PLA-HAN模型与基础的HAN模型的性能。我们按照对话段的长度分为长(600字以上)、中(301-600字)和短(300字以下)三类进行观察，结果如图2所示。

从图2可以看到：

- (1) 长的对话段的意图分类存在较大的挑战，正确率明显低于短的对话段。
- (2) 我们的PLA-HAN，在伪标签注意力的帮助下，在不同长度的对话段性能均优于HAN。尤其是长的对话段(超过600字)，意图分类正确率提升较为显著，达到1.56%。

## 5 结束语

针对现有文本分类方法在人机对话意图分类上存在的挑战，本文提出了一种结合话语伪标签注意力的层次注意力网络模型PLA-HAN。PLA-HAN通过优选伪标签集，设计和计算单句话语意图伪标签注意力，并将其嵌入到HAN的层级结构中，与HAN中的句子级别注意力相融合，提升了人机对话意图分类性能。我们在中国中文信息学会主办的“客服领域用户意图分类评测比赛”的评测语料上进行实验，实验结果证明PLA-HAN模型取得了优于HAN等研究进展文本分类方法的意图分类正确率。

## 致谢

本文受到国家自然科学基金(项目编号:71472068)的资助。

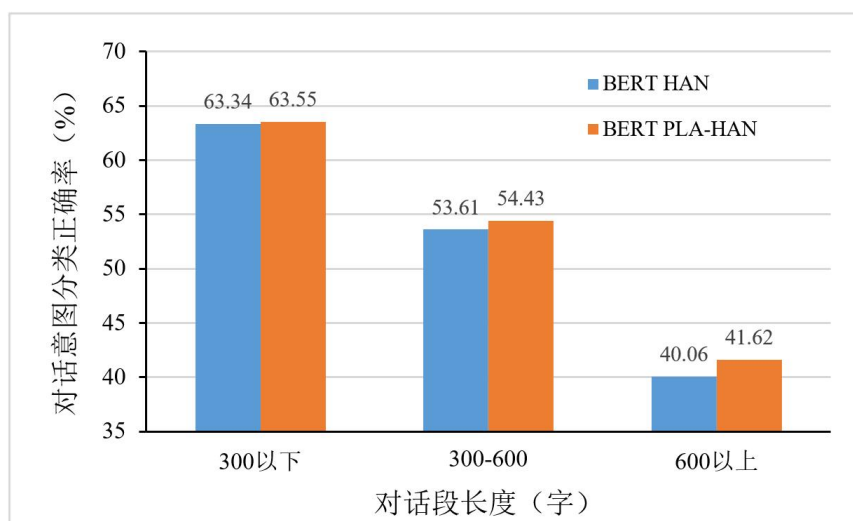


Figure 2: 不同长度对话段的模型性能对比

## 参考文献

- 安明慧, 沈忱林, 李寿山, 等. 2019. 基于联合学习的问答情感分类方法. 中文信息学报, 33(10):119-126.
- 柯子, 黄沛杰, 曾真. 2018. 基于优化“未定义”类话语检测的话语领域分类. 中文信息学报, 32(4):105-113.
- 俞凯, 陈露, 陈博, 等. 2015. 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, 38(12):2333-2348.
- Chen, Hongshen and Liu, Xiaorui and Yin, Dawei and Tang, Jiliang. 2017. *A survey on dialogue systems: Recent advances and new frontiers*, volume 19. ACM New York, NY, USA.
- J. P. Cheng, L. Dong, and M. Lapata. 2016. Long short-term memory-networks for machine reading. *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing (EMNLP 2016)*, pp. 551-561.
- J. Devlin, M. Chang, K. Lee, et al. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *Proceedings of the 17th Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT 2019)*, pp. 4171-4186.
- A. Graves, N. Jaitly and A. Mohamed. 2013. Hybrid speech recognition with deep bidirectional LSTM. *Proceedings of the 2013 IEEE workshop on automatic speech recognition and understanding (ASRU 2013)*, 273-278.
- P. Haffner, G. Tur, and J. H. Wright. 2003. Optimizing SVMs for complex call classification. *Proceedings of the 28th International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2003)*, pp. 632-635.
- B. Liu and T. Lane. 2016. Attention-based recurrent neural network models for joint intent detection and slot filling. *Proceedings of the 17th Annual Conference of the International Speech Communication Association (INTERSPEECH 2016)*, pp. 685-689.
- Y. Kim. 2014. Convolutional Neural Networks for Sentence Classification. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP 2014)*, pp. 1746-1751.
- Y. Kim, D. Kim, A. Kumar. 2018. Efficient large-scale neural domain classification with personalized attention. *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (ACL 2018)*, pp. 2214-2224.
- X. H. Phan, L. M. Nguyen, S. Horiguchi. 2008. Learning to classify short and sparse text & web with hidden topics from largescale data collections. *Proceedings of the 17th International Conference on World Wide Web (WWW 2008)*, pp. 91-100.

- S. Ravuri and A. Stolcke. 2016. A comparative study of recurrent neural network models for lexical domain classification. *Proceedings of the 41th IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2016)*, pp. 6075-6079.
- R. Sarikaya, G. E. Hinton, and B. Ramabhadran. 2011. Deep belief nets for natural language call-routing. *Proceedings of the 36th IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2011)*, pp. 5680-5683.
- N. T. Vu, P. Gupta, H. Adel, et al. 2016. Bi-directional recurrent neural network with ranking loss for spoken language understanding. *Proceedings of the 41th IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2016)*, pp. 6060-6064.
- P. Y. Xu and R. Sarikaya. 2013. Convolutional neural network based triangular CRF for joint intent detection and slot filling. *Proceedings of the 2013 IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU 2013)*, pp. 78-83.
- P. Y. Xu and R. Sarikaya. 2014. Contextual domain classification in spoken language understanding systems using recurrent neural network. *Proceedings of the 39th International Conference on Acoustics, Speech and Signal Processing (ICASSP 2014)*, pp. 136-140.
- K. Xu, J. Ba and R. Kiros, et al. 2015. Show, attend and tell: Neural image caption generation with visual attention. *Proceedings of the 31st International Conference on Machine Learning (ICML 2015)*, pp. 2048-2057.
- Z. C. Yang, D. Y. Yang, C. Dyer, et al. 2016. Hierarchical attention networks for document classification. *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics (NAACL 2016)*, pp. 1480-1489.