

# 融合实体信息和时序特征的意图识别模型

郑思露,程春玲,毛毅

(南京邮电大学 计算机学院,江苏 南京 210023)

**摘要:**人机对话意图识别旨在通过人机之间简短的对话识别出用户意图,通过对话文本的分类进而实现意图的识别。针对人机对话中因篇幅短导致语境匮乏和因对话随意性导致意图模糊的问题,提出了一种融合实体信息和时序特征的人机对话意图识别模型。在文本表示阶段,通过捕捉对话中实体信息来增强文本语义表达,并利用双向注意力机制动态生成符合语境的文本表示;并利用双向GRU提取对话上下文的时序特征来获取上下文意图之间的关系;通过级联多层gMLP,利用其内部空间控制单元自适应融合实体信息和时序特征,从而提升意图识别的准确率。为验证所提模型在多种任务上的效果,在不同意图识别任务数据集CCKS2018和SMP2018上进行实验,分别取得了90.6%和93.7%的准确率,对比CLSTM、DBN、Attention-RNN等具有代表性的模型,均有3%以上性能的提升。

**关键词:**深度学习;意图识别;特征融合;实体信息;时序特征;gMLP

中图分类号:TP391

文献标识码:A

文章编号:1673-629X(2022)11-0171-06

doi:10.3969/j.issn.1673-629X.2022.11.025

## An Intention Recognition Model Combining Entity Information and Temporal Features

ZHENG Si-lu, CHENG Chun-ling, MAO Yi

(School of Computer Science, Nanjing University of Posts & Telecommunications, Nanjing 210023, China)

**Abstract:** The purpose of human-machine dialogue intention recognition is to identify the user's intention through the brief dialogue between human and machine, and then realize the intention recognition through the classification of the dialogue text. Aiming at the problems of lack of context due to short length and ambiguity of intention due to randomness of dialogue in man-machine dialogue, a man-machine dialogue intention recognition model integrating entity information and temporal features is proposed. In the stage of text representation, the text semantic expression is enhanced by capturing the entity information in the conversation, and the bidirectional attention mechanism is used to dynamically generate the text representation in accordance with the context. Bidirectional GRU is used to extract the temporal features of the dialogue context to obtain the relation between context intentions. In order to improve the accuracy of intention recognition, multi-layer gMLP is used to adaptively fuse entity information and temporal features with its internal spatial control unit. In order to verify the effectiveness of the proposed model on multiple tasks, experiments were carried out on CCKS2018 and SMP2018 data sets of different intention recognition tasks, and the accuracy was 90.6% and 93.7%, respectively. Compared with CLSTM, DBN, Attention-RNN and other representative models, the performance was improved by more than 3%.

**Key words:** deep learning; intention recognition; feature fusion; entity information; time series characteristics; gMLP

## 0 引言

近年来,随着语音识别和自然语言理解的快速发展,人机对话已经变得越来越流行。目前,国内外许多机构已经开发了自己的人机对话系统,如苹果的Siri语音助手、百度的小度智能助手等闲聊领域的系统,还有针对特定领域的系统,如智能点歌系统。其中用户意图识别是整个系统的核心,正确的意图不仅可以提

高系统的效率还可以提升用户的使用体验。

对话意图识别是指对人机的对话文本进行分类进而实现意图识别,不同的任务类型对应着不同的分类模型,如闲聊系统则属于多分类,即识别出相应领域再有针对性地进行聊天,而点歌系统则属于二分类,即判断用户是否有点歌的意图,为下游功能提供支撑。

目前,将一般的文本分类方法应用在人机对话意

收稿日期:2021-11-25

修回日期:2022-03-28

基金项目:国家自然科学基金青年项目(62002174)

作者简介:郑思露(1996-),男,硕士,研究方向为自然语言处理、意图识别;通信作者:程春玲(1972-),女,教授,CCF会员(15597M),研究方向为数据挖掘、大数据分析和数据管理。

图识别任务上已经取得了不错的效果,但是对话文本与常规文本在篇幅和内容上存在很大差异,因此要想取得更好的意图识别效果就必须针对对话文本存在的难点设计合理的分类模型。首先,在人机对话中,用户往往是通过少量的话语表达自己的意图,这就造成缺乏足够的篇幅从而导致语境匮乏,其中还会存在一些实体信息,如专有名词,常规模型无法识别出这些专有名词,从而把它和其他文本一样对待,这也会造成语境的匮乏;其次,用户在表达意图时比较随意,无法确定用户在对话开头、中间或是结尾表达了自己的确切意图,而且前后意图还存在相互的影响,这都会导致意图含义不明确。因此,针对上述在人机对话意图识别中语境匮乏和意图模糊的问题,该文提出一种融合实体信息和时序特征的模型。该模型能够增强对话文本的语义表示,同时能够捕获意图在时序上的表达,再通过特征融合提升意图识别的准确率。具体而言,首先,在文本表示模块,结合对话中的实体信息动态生成符合语境的词向量从而减少因对话篇幅短带来的文本语境匮乏;其次,在时序特征提取模块,利用双向长短期记忆网络提取对话中前后意图的联系从而减少因对话随意性带来的意图不清晰;最后,在特征融合模块,利用多层门控内部自相关机制自适应融合对话中的实体信息和时序特征,这样既考虑了对话篇幅短和随意性的问题,又对特征进行了不同程度的融合,从而提升意图识别的效果。

## 1 相关工作

### 1.1 文本表示

在进行意图识别前,首先要将对话文本转换成计算机可以处理的形式。一个能够蕴含丰富信息且噪音小的高质量文本表示可以极大地提升模型的性能。

早期主要采用 One-hot<sup>[1]</sup>、TF-IDF<sup>[2]</sup>等基于词频的模型对文本进行表示,但是随着词表的增加会产生维数灾难和语义缺失等问题。随着神经网络和表征学习的发展,Word2vec<sup>[3-4]</sup>和 Glove<sup>[5]</sup>利用词向量训练方法来得到文本表示可以解决维数灾难和缓解语义缺失的问题。但是词向量往往只包含了浅层、单一的语义信息,对于人机对话语境匮乏,词向量显然无法清晰地表达用户的意图。在 BERT<sup>[6]</sup>、ELMo<sup>[7]</sup>、RoBERT<sup>[8]</sup>以及 XLNet<sup>[9]</sup>等预训练模型提出后,利用大规模数据训练动态词向量的方法可以有效地解决一词多义等问题。但对于人机对话而言,由于数据量较少且篇幅较短,利用上述从文本数据量角度出发的模型显然无法丰富文本表示的语义信息。对于意图识别而言,如果能识别出对话中的实体信息,在多数情况下,如果将专有名词“快乐崇拜”识别为歌名,则可得出用户有要听

歌的意图。ERNIE<sup>[10]</sup>语言模型从实体信息出发,在结合了预训练语言模型特点的基础上,在文本表示阶段,对文本中的实体信息进行抽取并将实体信息加入预训练过程中,进而得到的文本表示既符合当前语境又包含了对话中的实体信息,可以有效地缓解对话篇幅短带来的语境匮乏的问题。

### 1.2 意图识别

对话文本意图识别属于一种特殊的文本分类。传统基于规则模板的模型需要人为事先针对不同领域的类别构造不同的匹配模板来进行意图的分类,基于规则的方法在初期单一领域尚可取得不错的效果,但也存在适应领域单一且人力成本较高的问题。随着神经网络的发展,其在表征学习上的优势越发明显,越来越多的研究开始将神经网络应用在意图识别问题上。

文献[11]利用无监督学习训练模型权重,然后利用反向传播做微调,在意图识别任务上取得了不错的效果。文献[12]利用卷积神经网络(Convolutional Neural Network, CNN)来提取用户的查询意图,相比于手动提取特征,不仅可以获得深层次的语义特征,还可以降低特征工程量,但 CNN 只能提取到对话的局部语义信息,无法提取到前后意图的联系。循环神经网络(Recurrent Neural Network, RNN)将文本看作一个序列,能够从序列中学习上下文的语义信息,文献[13]利用 RNN 抽取对话的上下文信息从而提高意图识别的准确度。一个简单的 RNN 存在梯度爆炸和梯度消失的问题,而且只考虑了前文对后文的影响,而人机对话存在用户说话随意性的问题,所以只考虑前文对后文影响显然是不全面的。长短期记忆网络 LSTM(Long Short-Term Memory)和 GRU(Gate Recurrent Unit)也是循环神经网络,它们的提出解决了梯度爆炸和梯度消失的问题。GRU 和 LSTM 都是时序神经网络,GRU 相比于 LSTM,GRU 更容易训练、参数量较少,能够很大程度上提高训练效率。文献[14]在 ATIS 和 Cortana 数据集上对 LSTM 和 GRU 进行了对比,结果表明在意图识别任务上 GRU 与 LSTM 的性能相当,但 GRU 的参数量更少并且模型相对简单、更容易训练,因此很多时候会倾向于使用 GRU。但 GRU 仅考虑上文对下文的影响,无法体现下文对上文的影响,而 BiGRU 则可同时考虑上下文的相互影响,因此该文采用 BiGRU 提取对话的时序特征。

将不同的神经网络进行级联相对于单一的神经网络往往有较大竞争力。文献[15]将注意力机制和循环神经网络进行简单级联,在意图识别任务上取得比单独 RNN 或 Attention 更好的效果。文献[16]则利用 BiLSTM 编码器和注意力机制从字级别、句子级别提取对于意图中的多级语义信息。文献[17]提出结合

CNN 与 LSTM 的 CLSTM 模型对用户对话中一定长度的上下文信息进行语义特征建模。Liu D<sup>[18]</sup> 等利用 BERT 作为预训练模型,并使用 BiLSTM 提取文本双向特征,构建了面向任务的人机对话的意图分类模型。针对不同的特征如果给予相同的关注程度,那么在最后进行分类的时候,会造成信息的冗余,常用注意力机制对特征给予不同的权值,从而提升特征的质量。最近计算机视觉领域火热的 gMLP<sup>[19]</sup> 利用多层基于门控的内部机制从空间和通道角度对图像特征进行增强或削弱能够达到和注意力机制同样甚至更好的效果,而 gMLP 在模型上更加简单,因此该论文将 gMLP 首次应用在对话文本意图识别上,对时序特征和实体信息进行融合。综上所述,针对人机对话文本的特点,选取 ERNIE 语言模型来获取符合语境且包含实体信息的文本表示,并利用 BiGRU 从时序上获取上下文意图之间的关系,再利用 gMLP 将时序特征和实体信息进行融合,进而提高意图识别效果。

## 2 融合实体信息和时序特征的意图识别模型

### 2.1 模型整体架构

在人机对话中存在对话短和随意性的特点,这会导致语境匮乏和意图不清晰的问题,而这两者对意图的影响较大,所以以这两个问题作为出发点进行研究。对于语境匮乏的问题,采用完全基于注意力机制的 Transformer,其能够依据不同的上下文动态生成词向量,同时在动态生成词向量时加入实体信息去缓解语境匮乏的问题。对话随意性会导致意图不清晰,考虑从对话的时序特征上对意图进行捕获,通过捕捉对话双向时序特征来获取对话中的具体意图。在解决上述两个问题后,考虑到两者是从不同层面对模型进行改进,该论文提出一种融合实体信息和时序特征的对话文本意图识别模型。模型的整体框架如图 1 所示。该模型在整体上可分为两个模块:文本表示模块和特征融合模块,两者分别对应图 1 中的文本表示和特征融合。

### 2.2 文本表示

从提升文本表示所蕴含的信息角度出发,通过捕捉对话中的实体信息并动态生成符合当前语境的词向量来丰富文本表示的信息。例如一个句子的输入包含  $n$  个词语,  $T = \{T_1, T_2, \dots, T_n\}$ , 首先在进行词向量训练前,对  $T$  中部分词进行 mask,对比 BERT 的随机 mask 词语,该论文采用的 ERNIE 通过 mask 句子中的实体,这样模型在训练完成后,也就学习到了这些实体的信息。ERNIE 在经过 mask 之后得到  $E$ ,将  $E$  在训练过程中采用一个多层的 Transformer 作为动态词向量的编码器,Transformer 通过多头注意力机制可以依据

词在文本中的上下文信息,动态生成符合语境的特征表示。经过文本表示模块之后文本表示为  $x = \{x_1, x_2, \dots, x_n\}$ ,其中词向量的维度为  $k$ 。实体动态词向量生成过程如图 2 所示。

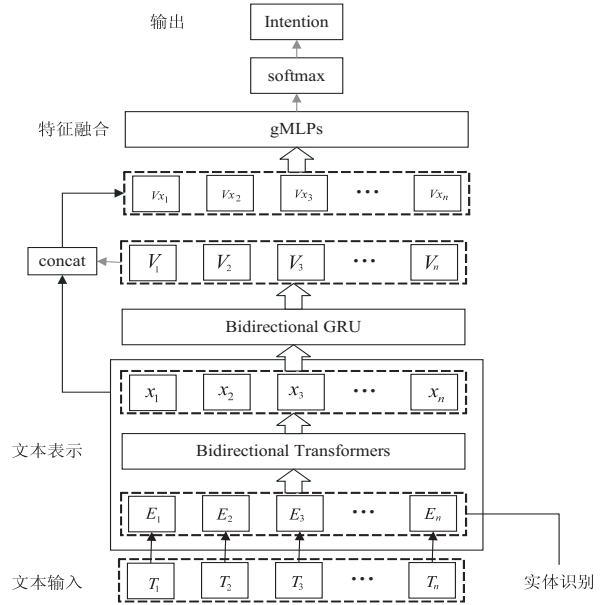


图 1 融合实体信息和时序特征的对话文本意图识别模型架构

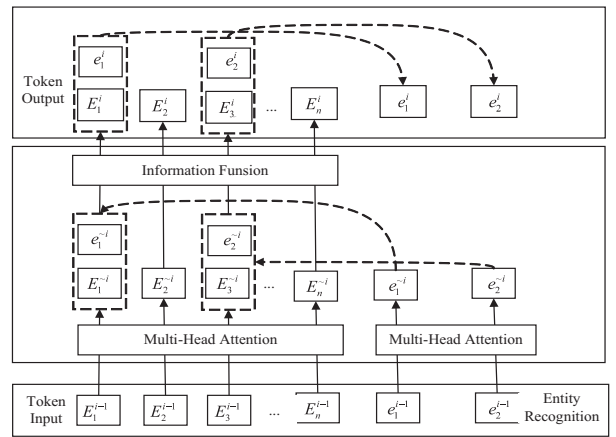


图 2 基于实体信息的词嵌入

其中,  $T^{i-1}$  表示在训练过程中  $i$  时刻词向量的状态,  $e^{i-1}$  表示在时刻  $i$  输入训练过程的实体信息。核心公式如公式(1):

$$Z = \text{softmax}\left(\frac{Q \times K^T}{\sqrt{d_k}}\right) \times V \quad (1)$$

其中,  $Z$  表示注意力的值,  $Q, K, V$  分别表示查询向量 (Query Vector) 矩阵、键向量 (Key Vector) 矩阵和值向量 (Value Vector) 矩阵。  $\sqrt{d_k}$  对注意力进行缩放,可以分散注意力而不至于过于集中某个 token。

### 2.3 时序特征提取

从对话文本的时序角度出发,通过提取对话前后意图之间的联系来捕获用户的意图信息。考虑前后之间的关系,如序列前文内容在意图的表达上对后文是



否有影响,影响有多大,相同的也需要考虑后文内容对前文意图的影响。但是如果不加距离地考虑前文哪一段对后文产生了影响,而是考虑前文所有内容,无疑是会造成信息冗余从而导致精度下降。所以采用 BiGRU 对  $x$  进行双向长短距离的特征提取,提取得到对话时序上的特征。其中单向 GRU 时序特征提取器如图 3 所示。

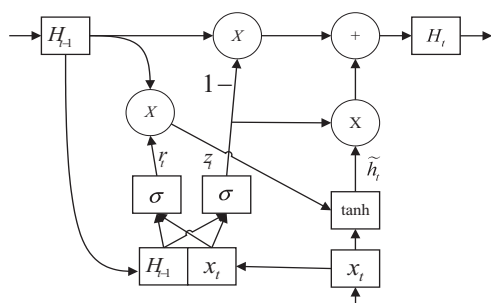


图 3 单向长短距离特征提取器

其中,  $x_t$  表示  $t$  时刻输入词,  $H_{t-1}$  表示  $t-1$  时刻输出的隐藏状态,  $r_t$  表示重置门,  $z_t$  表示更新门。更新门用于控制前一时刻的状态信息被带入到当前状态中的程度,更新门的值越大说明前一时刻的状态信息带入越多,如果为 1,则表明前一时刻信息完全可以覆盖当前状态,如果为 0 则说明前一刻的状态信息没有任何作用。重置门  $r_t$  控制前一状态有多少信息被写入到当前的候选集  $\tilde{h}_t$  上,重置门越小,前一状态的信息被写入的越少。具体计算过程如公式(2)~公式(5):

$$r_t = \sigma(W_r \cdot [H_{t-1}, x_t]) \quad (2)$$

$$z_t = \sigma(W_z \cdot [H_{t-1}, x_t]) \quad (3)$$

$$\tilde{h}_t = \sigma(W_{\tilde{h}} \cdot [r_t * H_{t-1}, x_t]) \quad (4)$$

$$H_t = (1 - z_t) * H_{t-1} + z_t * \tilde{h}_t \quad (5)$$

其中,  $W_r$ 、 $W_z$ 、 $W_{\tilde{h}}$  为参数矩阵,  $\sigma$  为 sigmoid 激活函数,  $\cdot$  表示矩阵乘法。双向时序特征是将文本的前向特征  $\vec{H}$  和后向特征  $\overleftarrow{H}$  进行拼接,从而得到包含文本双向序列特征的  $V$ 。计算公式如公式(6)所示。

$$V = [\vec{H}, \overleftarrow{H}] \quad (6)$$

## 2.4 特征融合

在获得含有实体信息的动态词向量和文本的序列特征的信息后需要对这些特征进行融合。尽管序列特征信息中包含有实体信息,但是经过双向长短距离的特征提取后实体信息会有所衰减,常用的解决方法是将含有实体信息的文本表示与序列特征进行简单地拼接。拼接过程如公式(7)所示。

$$Vx = [V, x] \quad (7)$$

通过对不同层特征进行拼接后,得到的  $Vx$  所包含的语义信息既有原始的实体信息又有文本序列上的信息,如果不对  $Vx$  进行有效地融合则会导致计算量过大

同时也会导致信息的冗余,对特征给予相同的关注程度会影响到那些可以体现意图的特征不明显。故在得到多层的特征之后,采用门控机制对特征沿着通道的维度推断出权重,然后与原特征图相乘来对特征进行自适应调整,这样就能将多层的特征进行有效地增强或削弱。其中特征选择器内部结构如图 4 所示。

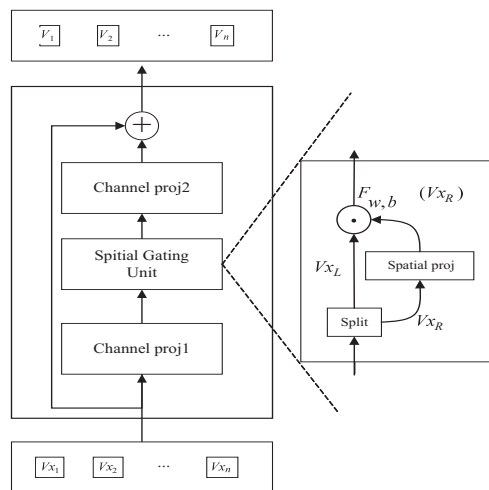


图 4 特征融合器内部结构

其中 Channel proj1 和 Channel proj2 都是线性映射,前者将输入  $n * d$  的输入向量  $VT$  映射成  $n * d_{fnn}$  的向量,后者将  $n * d_{fnn}$  映射成  $n * d$ ; split 是将  $n * d_{fnn}$  一分为二为  $V_{x_L}$  和  $V_{x_R}$ 。

Spatial proj 为了捕捉词与词之间复杂空间的交互信息,计算过程如公式(8)和公式(9):

$$F_{w,b}(V_{x_R}) = W \cdot V_{x_R} + b \quad (8)$$

$$s(Vx) = V_{x_L} \odot F_{w,b}(V_{x_R}) \quad (9)$$

其中,  $b$  是偏置项,  $s(Vx)$  捕捉了词与词之间的关系信息,通过  $s(Vx)$  多层特征可以进行自适应调整以此来对特征进行选择。

## 3 实验及其结果分析

### 3.1 实验设置与内容

为验证所提模型在人机对话意图识别任务上的有效性,在不同任务的意图识别数据集 CCKS2018 和 SMP2018 进行实验;其中 CCKS2018 是识别用户是否具有点歌意图的数据集, SMP2018 则是判定用户闲聊的领域。将每个数据集随机打乱,平均分成 10 份,按照训练集:验证集:测试集 = 8 : 1 : 1 的比例进行数据划分。数据集的划分如表 1 所示。

表 1 数据划分

Data sets	CCKS2018	SMP2018
Train set	9 600	2 455
Test set	1 200	307
Dev set	1 200	307

以下参数均由实验所得,其中 batch\_size 为 32; Dropout 为 0.1;优化器为 BertAdam,学习率为 0.001,设置 warmup 为 0.05;句子的最大长度为 30,策略是取长补短;gMLP 层数为 6。选用准确度 precision 和综合评价指标 f1 作为模型的评价指标。

实验开发平台如下:操作系统为 Windows10,CPU 为 R7-2700,GPU 为 RTX2060,开发工具为 PyCharm,深度学习框架为 Pytorch-1.8.0。

### 3.2 模型整体性评估

为验证所提模型在意图识别准确率上的提升,对比近年来在对话意图识别任务上具有代表性的模型方法,实验结果如表 2 所示。

表 2 代表性意图识别模型性能比较

Dataset	CCKS2018		SMP2018	
	precision	f1	precision	f1
CLSTM	0.871	0.869	0.896	0.883
DBN	0.676	0.615	0.745	0.729
Attention-RNN	0.853	0.851	0.875	0.863
OurModel	0.906	0.905	0.937	0.928

在准确率和 f1 上对比其他三个模型,说明无论是在单意图识别任务上还是多意图识别上,所提模型在两个数据集上均取得最好的表现,说明面对不同的意图识别任务,融合实体信息和时序特征的方法相比于已有的模型具有更好的效果。

### 3.3 融合实体信息性能分析

为验证实体信息对语境丰富的影响,设计在文本表示阶段不捕获实体信息的 BERT-CBMLP 模型,同时设计在文本表示阶段加入主题词的 Th-CBMLP 模型。实验结果如表 3 所示。

表 3 融合实体信息性能比较

Dataset	CCKS2018		SMP2018	
	precision	f1	precision	f1
BERT-CBMLP	0.880	0.879	0.859	0.831
Th-CBMLP	0.892	0.889	0.874	0.856
OurModel	0.906	0.905	0.937	0.928

从表 3 可以得出以下结论,在文本表示阶段没有加入其他信息的 BERT-CBMLP 比加入主题信息的 Th-CBMLP 和捕获实体信息的 OurModel 效果要低,表明在文本表示阶段主题信息和实体信息可以增强语境的表达。主题信息来自对话文本中关键字,属于内部信息,而实体信息则属于外部信息,相比内部信息,外部信息能够提供更多的语义信息。

### 3.4 时序特征提取性能分析

为探究对话中时序特征对意图表达的影响,设计 ERNIE-gMLP 来探究加入时序特征和不加入时序特

征对意图识别的影响;设计加入 GRU 提取单方向时序信息的 ERNIE-CGRU-gMLP 来探究双向时序特征在提取意图上的优势;设计利用 Transformer 加入词位置信息的 ERNIE-CTrm-gMLP 来对比从不同角度考虑对话前后意图的联系。实验结果如表 4 所示。

表 4 时序特征性能分析

Dataset	CCKS2018		SMP2018	
	precision	f1	precision	f1
ERNIE-gMLP	0.887	0.877	0.892	0.880
ERNIE-CGRU-gMLP	0.891	0.891	0.921	0.911
ERNIE-CTrm-gMLP	0.895	0.893	0.914	0.908
OurModel	0.906	0.905	0.937	0.928

由表 4 可以得出,对比不提取时序特征的 ERNIE-gMLP,其他三个提取时序或位置信息的模型在单意图和多意图数据集上均取得了较大的性能提升,说明对于对话意图识别而言,应该考虑对话前后在表达意图上的关系。对比提取单向时序特征和加入词的位置信息,两者在不同数据集上的结果各有高低但相差较小,说明对于对话意图识别而言,通过加入词的位置信息可以达到仅提取单向时序特征同样的效果。由于用户在表达意图是比较随意,比如,用户在开始表明有多个意图的可能性,最后通过指代来说明具体意图,这时候单向的时序则无法准确识别出具体的意图,所以利用 BiGRU 从双向提取对话的时序信息可以有效地提取到具体的意图。

### 3.5 特征融合性能分析

为验证对特征进行不同程度融合可以提升模型性能,对比将实体信息和时序特征进行 concat 的 ERNIE-CBiGRU、进行 add 的 ERNIE-ABiGRU 以及不进行特征融合的 ERNIE-BiGRU。实验结果如表 5 所示。

表 5 不同特征融合方法对比

Dataset	CCKS2018		SMP2018	
	precision	f1	precision	f1
ERNIE-BiGRU	0.878	0.874	0.917	0.901
ERNIE-CBiGRU	0.891	0.891	0.930	0.920
ERNIE-ABiGRU	0.898	0.888	0.923	0.916
OurModel	0.906	0.905	0.937	0.928

由表 5 可以看出,不进行特征融合的在两个数据集上 precision 和 f1 都是最小的,对特征从不同角度进行融合可以提高意图识别的效果,所以对话意图识别上可以从不同角度对实体信息和时序特征进行融合,从而有效地提高意图识别的效果。但是对于不同的方法而言,add 方法是将特征图相加,保持原有的通道数不变,concat 则是通道数的增加,在两个数据集上进行 concat 的效果要略好于 add 方法,特征图的相加无疑

会增加后面分类的信息量,也会有较多信息的损失,而 concat 则是把信息全部带到下一分类层中。但进行 concat 也会带来信息的冗余,所以 OurModel 在 concat 之后利用多层 gMLP 对实体信息和时序信息进行自适应的融合,通过 gMLP 的内部内部空间控制单元对 concat 之后的重要特征进行增强,对于一些噪音特征进行弱化,从而有效地提高意图识别的效果。

#### 4 结束语

该文提出了一种特征融合的方法,具体来说,首先利用对话中的实体信息在语义表达上和 BiGRU 在时序特征提取上的优势,再利用多层 gMLP 的门控机制自适应调整实体信息和时序特征的权值。实验结果表明,所提模型对不同任务人机对话领域均具有较好的性能表现。该模型方法可以应用在任务型对话中,比如,预定酒店、机票等。大多数模型都假定用户的语句只有一个意图分类,但是很多情况并不是这样,真实环境下用户往往存在多个意图。目前多意图的研究还比较少,这也是未来一个重要的研究方向。

#### 参考文献:

- [1] JOSEPH T, RATINOV L, BENGIO Y. Word representations: a simple and general method for semi-supervised learning[C]//Proceedings of the 48th annual meeting of the association for computational linguistics. Uppsala: ACL, 2010: 384-394.
- [2] YIH W T, GOODMAN J, CARVALHO V R. Finding advertising keywords on web pages[C]//International conference on world wide web. Edinburgh: DBLP, 2006: 213-222.
- [3] MIKOLOV T, CORRADO G, KAI C, et al. Efficient estimation of word representations in vector space[C]//International conference on learning representations. New York: Curran Associates, 2013.
- [4] MIKOLOV T, SUTSKEVER I, CHEN K, et al. Distributed representations of words and phrases and their compositionality[C]//Conference and workshop on neural information processing systems. MA: MIT Press, 2013: 3111-3119.
- [5] PENNINGTON J, SOCHER R, MANNING C. Glove: global vectors for word representation[C]//Empirical methods in natural language processing. PA: ACL, 2014: 1532-1543.
- [6] DEVLIN J, CHANG M W, LEE K, et al. BERT: pre-training of deep bidirectional transformers for language understanding[J]. arXiv:1810.04805, 2018.
- [7] PETERS M, NEUMANN M, IYYER M, et al. Deep contextualized word representations[C]//Proceedings of the 2018 conference of the North American chapter of the association for computational linguistics; human language technologies. [s. l.]: ACL, 2018: 2227-2237.
- [8] LIU Y, OTT M, GOYAL N, et al. RoBERTa: a robustly optimized BERT pretraining approach[J]. arXiv: 1907. 11692, 2019.
- [9] YANG Z, DAI Z, YANG Y, et al. Xlnet: generalized autoregressive pretraining for language understanding[C]//33rd conference on neural information processing systems (NeurIPS 2019). Vancouver: [s. n.], 2019: 5754-5764.
- [10] SUN Y, WANG S, LI Y, et al. ERNIE: enhanced representation through knowledge integration[J]. arXiv: 1904. 09223, 2019.
- [11] SARIKAYA R, HINTON G E, RAMABHADRAN B. Deep belief nets for natural language call-routing[C]//Proceedings of the IEEE international conference on acoustics, speech, and signal processing. Prague: IEEE, 2011: 5680-5683.
- [12] KIM Y. Convolutional neural networks for sentence classification[C]//Empirical methods in natural language processing. PA: ACL, 2014: 1746-1751.
- [13] BHARGAVA A, CELIKYILMAZ A, HAKKANITUR D, et al. Easy contextual intent prediction and slot detection[C]//Acoustics, speech and signal processing (ICASSP). NJ: IEEE, 2013: 8337-8341.
- [14] RAVURI S, STOLCKE A. A comparative study of recurrent neural network models for lexical domain classification[C]//Proceedings of the 2016 IEEE international conference on acoustics, speech and signal processing (ICASSP). NJ: IEEE, 2016: 6075-6079.
- [15] YU W, SHEN Y, JIN H. A bi-model based RNN semantic frame parsing model for intent detection and slot filling[C]//Annual conference of the North American chapter of the association for computational linguistics; human language technologies (NAACL-HLT). [s. l.]: [s. n.], 2018: 309-314.
- [16] 徐 扬, 王建成, 刘启元, 等. 基于上下文信息的口语意图检测方法[J]. 计算机科学, 2020, 47(1): 205-211.
- [17] ROJASBARAHONA L, SU P, ULTES S, et al. Exploiting sentence and context representations in deep neural models for spoken language understanding[C]//Proceedings of COLING 2016, the 26th international conference on computational linguistics; technical papers. NY: ACM, 2016: 258-267.
- [18] LIU D, ZHAO Z, GAN L, et al. Intention detection based on bert-bilstm in taskoriented dialogue system[C]// 2019 16th international computer conference on wavelet active media technology and information processing. [s. l.]: IEEE, 2019.
- [19] LIU H, DAI Z, SO D R, et al. Pay attention to MLPs[J]. arXiv:2105.08050, 2021.