

文章编号: 2095-2163(2021)08-0185-04

中图分类号: TP391.1

文献标志码: A

基于度量学习的意图识别和槽填充方法

衣景龙^{1,2}, 赵铁军^{1,2}

(1 哈尔滨工业大学 计算机科学与技术学院, 哈尔滨 150001; 2 哈尔滨工业大学 机器智能与翻译实验室, 哈尔滨 150001)

摘要: 人机对话中小样本学习场景下的意图识别和槽填充, 是自然语言处理的一个重要课题。本文采用基于度量学习的方法, 通过计算 query set 中的样本与 support set 中样本的距离, 寻找距离最近的类别样本作为分类标签, 同时将两个任务联合进行训练, 用以提升模型的效果。从实验结果中可以得出, 本文提出的 Fine-tune 方法, 对意图识别和槽填充任务都有一定的帮助和提升。胶囊网络在意图识别中也起到了一定的效果, 可以帮助去除一部分无关信息, 但对槽填充任务的帮助不明显; 而任务自适应的投影网络, 可以更好地将不同类的向量分开, 提升了两个任务的性能。

关键词: 小样本学习; 意图识别; 槽填充

Intention detection and slot filling based on metric learning

YI Jinglong^{1,2}, ZHAO Tiejun^{1,2}

(1 School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001, China;

2 Machine Intelligence and Translation Lab, Harbin Institute of Technology, Harbin 150001, China)

[Abstract] Intent detection and slot filling in small sample learning scenarios in human-computer dialogue is an important topic in natural language processing. In this paper, a metric-based learning method was adopted. By calculating the distance between the samples in the Query set and the samples in the Support set, the closest category sample was found as the classification label. Meanwhile, the two tasks were trained jointly to improve the effect of the model. From the experimental results, it can be concluded that our fine-tune method has certain help and improvement for intention detection and slot filling task, and capsule network also plays a good effect in intention identification, which can help remove part of irrelevant information, but the help for slot filling task is not obvious. The task-adaptive projection network can better separate the vectors of different classes and improve the performance of the two tasks.

[Key words] few-shot learning; intention detection; slot filling

0 引言

近年来, 深度学习方法被广泛应用于计算机视觉以及人机对话等人工智能的多个领域^[1], 并极大地推动了这些领域的发展。但目前的深度学习方法都需要大量的标注训练数据, 而在工业界应用的过程中, 很难获得大规模标注好的数据。以任务型对话系统为例, 在系统研发时很难获得大量的用户对话语料, 若获取了大量的对话数据, 也面临着人工标注成本高昂的窘境。同时, 像任务型对话系统等类似的线上应用如百度的小度^[2], 用户的需求往往变化比较频繁, 会导致数据标注和模型训练反复进行。因此, 小样本场景下对自然语言理解技术的研究, 成为本课题的重中之重。在目前的人机对话系统研究中, 意图识别和槽填充任务的效果往往是影响人机对话系统性能的关键。小样本场景下的意图识别和槽填充任务面临着许多挑战: 由于样本量不足, 模型

往往难以学到足够的语言知识; 现有的基于原型向量的小样本下的分类方法, 往往会损失信息。另外, 还涉及到意图识别和槽识别任务如何结合等问题。

目前, 小样本场景中的自然语言理解方法已经取得了不错的进展, 文献[3]中提出了基于原型向量计算的非参学习的阈值调整方法用于小样本下的多意图分类, 取得了较好的实验效果。阿里巴巴团队提出基于 BERT 的预训练联合识别模型^[4], 用 BERT 动态表征词的多义性, 解决一词多义的问题, 在 ATIS 数据集上意图识别的准确率达到 97.5%, 槽填充的 F1 值达到了 96.1% 的结果。本文提出了加入预训练语言模型 Fine-tune 方法和胶囊网络动态计算原型向量的方法, 有效改善了小样本场景下的自然语言理解效果。

1 Fine-tune 方法

以 BERT 为代表的预训练模型^[5]是在海量数据

作者简介: 衣景龙(1997-), 男, 硕士研究生, 主要研究方向: 任务型对话; 赵铁军(1962-), 男, 博士, 教授, 博士生导师, 主要研究方向: 自然语言生成、机器翻译、人机对话。

收稿日期: 2021-06-08

上训练得到的,虽然含有大量的先验语言知识,可以在小样本场景下提高小样本学习的性能,但当应用于不同的数据集时,往往需要微调,来使表征向量更为贴合新数据集的语义。

由于 BERT 的预训练过程与本文的模型训练过程不一致, BERT 中 **CLS** 向量虽然被视为句子的表示向量,但其在训练过程中,却是用来对句子对之前是否是顺序关系做判断,这与本文在训练过程中对句子进行分类过程并不一致。当面临多任务时,预训练过程与训练过程不一致导致的误差会被放大。对此,本研究对模型进行了微调,在 BERT 后,链接两个全连接输出层,一个接在 **CLS** 向量之后,用以进行意图分类,一个接在 **token** 向量之后用以进行槽填充。具体实现过程如下:

(1) 在海量数据集 (wiki) 上训练出 BERT 模型 (也可直接调用已经训练好的)。

(2) 在改进的 BERT 后加一个全连接层,其参数随机初始化生成,直接调用 **CLS** 向量和各个词向量进行意图和槽的分类,进行 Fine tune。

(3) 丢弃第二步中的全连接层,将现在的 BERT 作为新的模型的编码器,后面接入小样本的意图识别和槽填充模型。

整体模型框架如图 1 所示。

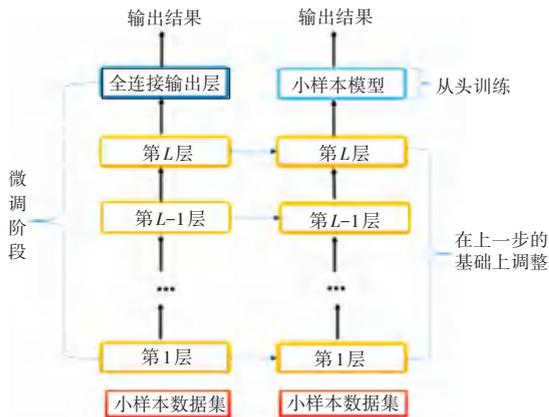


图 1 预训练过程图

Fig. 1 Pre-training process

在本文中,源模型采用哈尔滨工业大学社会信息与检索团队预训练的中文语言模型,将该模型进行改进,在其后直接加一个全连接层进行分类,依次来弥合小样本数据与源预训练数据的分布误差后,舍弃掉全连接层,在其后加入小样本训练模型,用于原型向量的计算和匹配分类。因为在 Fine tune 方法中可能会存在“灾难性遗忘”问题,因此在训练时可以固定 BERT 编码器的参数,只训练小样本模型。

当选择 Fine tune 编码器时,还采用了学习率预热方法 warm up,即在训练开始时使用一个较小的学习率,训练了一定的 epoch 或者 step 之后,再修改为原来设置的学习率来进行学习。这是因为最开始的时候,若学习率较大,模型可能不稳定,通过 warm up 的方式,可以在一定程度上避免模型震荡。

2 基于度量学习方法的意图识别和槽填充

通过 BERT 编码器,提取出句子和字的表示向量,在 Support Set 中将不同的意图标签和槽标签所对应的向量通过取平均等方式作为原型向量 c_i ,同时利用 BERT 编码器获取 Query Set 中语句和字的表示向量与原型向量计算相似度,通过最近邻的思想来确定每个语句的意图标签和每个字的槽标签。

最初求原型向量的做法是对每个意图或槽所对应的向量直接取平均^[6],但在每个句子中可能存在一些与意图无关的干扰信息,如果直接加和或者取平均的方式很容易导致这些与意图或者槽无关的干扰信息累加,影响最终的效果。因此,本课题采用了胶囊网络的方式,去除从样本表示计算类别过程中的无关信息^[7]。其模型架构如图 2 所示。

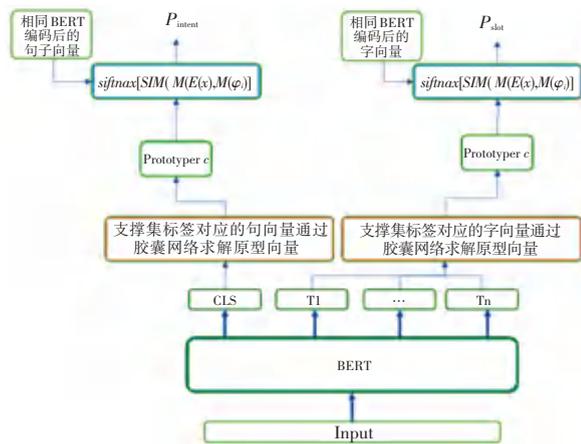


图 2 模型架构图

Fig. 2 Model architecture diagram

将支撑集的样本通过 BERT 编码表示为向量之后,将这些表示向量视为输入的胶囊,输出的胶囊视为类别的语义特征表示,从输入胶囊到输出胶囊的过程中采取动态路由算法。

首先是对所有的句子(字)表示向量接一个前馈神经网络,这等同于做了一次矩阵转化,其意义在于从样本表示的语义空间转换到类比表示的语义空间。

$$\hat{e}_{ij} = W e_{ij}, \quad (1)$$

式中, e_{ij} 表示类别 i 中第 j 个样本的句子(词) 向量表示, \hat{e}_{ij} 是转换后的表示。然后采用动态路由算法来过滤无关信息, 提取类别的表示特征。在每次路由迭代的过程中, 将其权重因子输入到 *softmax* 函数, 确保其加和为 1。

$$d_i = \text{softmax}(b_i), \quad (2)$$

其中, b_i 为连接系数, 初始化为 0, 等同于在最开始的时候采用平均化的操作。每一类的类别向量 \hat{c}_i , 是其转后候选类向量的 \hat{e}_{ij} 加权之和。

$$\hat{c}_i = \sum_j d_i \cdot \hat{e}_{ij}, \quad (3)$$

为了保证每个类的向量模长不超过 1, 使用一个非线性的 *squash* 函数来对其进行处理。

$$c_i = \frac{\|\hat{c}_i\|^2}{1 + \|\hat{c}_i\|^2} \cdot \frac{\hat{c}_i}{\|\hat{c}_i\|}, \quad (4)$$

每次迭代结束后, 通过路由协议来调整不同胶囊之间的连接权重。如果产生类别表示的向量 c_i 与该样本的候选向量 \hat{e}_{ij} 在向量空间中较为相近, 则增大该连接权重, 否则减小该连接权重。

$$b_{ij} = b_{ij} + c_i \cdot \hat{e}_{ij}. \quad (5)$$

在本文模型中, 经过实验, 迭代次数设置为 3 次。

3 实验

3.1 实验数据

本文的实验数据为小样本数据集, 数据来源主要有两个途径: 一是 SMP2020 中文人机对话技术评测 (ECDT) TASK1; 此外, 少部分数据来源于 SNIP 等公开数据集上进行翻译, 设计脚本, 抽取小样本所需要的样本组织形式。数据规模为:

(1) 训练集: 查询城市、APP 问答、询问价格、翻译、查询天气、航班预订等在内的 45 个领域。

(2) 开发集: 单词查询、垃圾分类、笑话、假期、温度查询 5 个领域。

(3) 测试集: 时间规划、讲故事、虚拟查询、星座问答、戏剧问答等在内的 9 个领域。共计: 6 694 个句子, 143 个意图及 205 个槽位。

格式如下:

```
{
  "text": "请帮我打开 uc",
  "domain": "app",
  "intent": "LAUNCH",
  "slots": {
```

```
"name": "uc"
```

其中, text 是要查询的文本; domain 是领域; intent 是文本意图; slots 是文本中存在的词槽。本课程抽取了 3-shot 和 5-shot 的数据集进行研究。

3.2 实验结果

对于小样本意图分类任务, 采用准确率 (Accuracy) 作为评价指标。对于小样本语义槽填充任务, 采用 *F1 - score* 作为评价指标, 当预测槽位的一个槽和其值组合与标准答案的一个槽和槽值组合完全一致, 视为一个正确预测。用句子准确率作为二者联合训练的评价指标。只要当该句子的意图识别和槽填充全部正确时, 该样本才算一个正确样本。

实验结果见表 1~表 3。

表 1 在 3-shot 数据集上联合学习实验结果

Tab. 1 Joint learning experiment results on 3-shot dataset

Model	Intent Acc.	Slot F1	Sentence Acc
BiLSTM	30.30	22.21	11.40
BERT+Proto	67.22	40.73	21.64
Fine-tune+BERT+Proto	67.82	62.73	38.97
Fine-tune+BERT+ Proto+Capsule	68.08	53.29	31.39

由表 1 可知, 当意图识别和槽填充任务联合训练时, Fine-tune 的方法对槽识别任务效果提升明显。而胶囊网络的引入, 虽然提高了意图识别的效果, 但也造成了槽填充任务的效果下降。针对此, 本文单独以胶囊网络对槽填充和意图识别的影响做了消融实验。

表 2 意图识别实验结果

Tab. 2 Intent detection experimental results

Model	3-shot Intent Acc	5-shot Intent Acc
BERT+ Proto	71.44	75.03
BERT+ Proto +capsule	73.95	78.93

表 3 槽填充实验结果

Tab. 3 Slot filling experimental results

Model	3-shot Slot F1	5-shot Slot F1
BERT+ Proto	35.79	38.53
BERT+ Proto +capsule	33.87	34.86

由表 2 和表 3 可以看出, 胶囊网络对意图识别实验有明显的提升作用, 但是在槽填充任务上并没有获得提升效果。其原因可能是对于细粒度的词级分类而言, 并没有太多的无关信息, 而胶囊网络的引入反而使最终的原型向量与对应的类别中特征平缓的单词更为相像, 反而引入了误差和偏置; 而对意图

识别任务而言,其粒度是句子级,可能由于表述习惯不同,或者无关单词较多等原因,含有较多的与意图类别无关的干扰信息,此时引入胶囊网络,可以起到一个很好的消除干扰信息的作用。

4 结束语

本文对人机对话中的小样本学习场景下的意图识别和槽填充展开了研究,采用了基于度量学习(Metric based)的方法^[8]。度量学习方法通过计算 query set 中的样本与 support set 中样本的距离,寻找距离最近的类别样本作为分类标签,同时将两个任务联合进行训练,用以提升模型的效果。从实验结果中可以得出,本文的 Fine-tune 方法对意图识别和槽填充任务都有一定的帮助和提升,胶囊网络在意图识别中也起到了不错的效果,可以帮助去除一部分无关信息,但对槽填充任务的帮助不明显。因此下一步可以继续研究在词级别的原型向量计算上是否有更好的改进方法,能更准确的表征槽的类别向量,同时在 Metric based 的小样本学习方法中可以看出,主要的改进方向有两个:一是对样本语义表

征向量的计算;二是对原型向量的计算,这对以后的研究也是一个好的启发。

参考文献

- [1] LÓPEZ-CÓZAR R, CALLEJAS Z, GRIOL D, et al. Review of spoken dialogue systems[J]. *Loquens*, 2014, 1(2): 12-26.
- [2] 陈东升. 百度 AI 黑科技小度在家智能音箱[J]. *计算机与网络*, 2018, 44(22): 26-28.
- [3] HOU Y, LAI Y, WU Y, et al. Few-shot Learning for Multi-label Intent Detection[J]. 2020.
- [4] CHEN Q, ZHUO Z, WANG W. BERT for Joint Intent Classification and Slot Filling[J]. 2019.
- [5] DEVLIN J, CHANG M W, LEE K, et al. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding[J]. 2018.
- [6] LOZANO M, SOTOCA J M, JS SÁNCHEZ, et al. Experimental study on prototype optimization algorithms for prototype-based classification in vector spaces[J]. *Pattern Recognition*, 2006, 39(10): 1827-1838.
- [7] XIA C, ZHANG C, YAN X, et al. Zero-shot User Intent Detection via Capsule Neural Networks[J]. 2018.
- [8] KULIS B. Metric Learning: A Survey[J]. *Foundations and Trends © in Machine Learning*, 2012, 5(4).

(上接第 184 页)

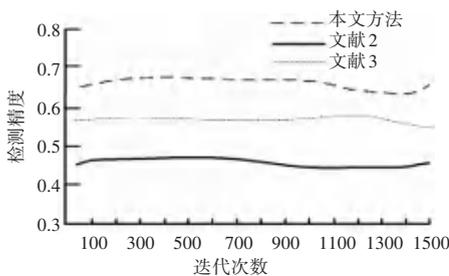


图 1 D-S 理论的模糊图像智能识别精度对比测试

Fig. 1 Comparative test of fuzzy image intelligent recognition accuracy based on D-S theory

分析图 1 得知,本文方法进行模糊图像智能识别精度较高,较其他方法性能更为优越。

4 结束语

当前,多媒体技术的快速发展,使得图像识别技术的应用需求不断提升,因此合理的图像识别技术能够为人们的日常生活和工作带来诸多的便利。本文通过上述论述,基于 D-S 推理理论实现了对模糊图像的智能识别方法设计,并通过实验进一步验证了该方法在实际应用中的性能和优势。将本文提出的模糊图像识别方法应用于实际,不仅能够实现

对图像的高精度识别,同时还能够进一步提高图像资源的利用。

参考文献

- [1] 钟国韵,常艳荣,汪宇玲,等. 旋转局部梯度模式特征及其在模糊图像识别中的应用[J]. *计算机应用与软件*, 2020, 37(10): 180-187, 255.
- [2] 陈晓文,刘光帅,刘望华,等. 结合 LoG 边缘检测和增强局部相位量化的模糊图像识别[J]. *计算机科学*, 2020, 47(12): 197-204.
- [3] 袁颖,李论,杨英仓. 基于 Capsule Net 架构的模糊指纹图像识别[J]. *软件*, 2020, 41(8): 48-51.
- [4] 杨旭,李君,宋美艳,等. 基于证据理论的电力视角图像边缘检测[J]. *电力学报*, 2020, 35(4): 71-79.
- [5] 刘争,祝利,解梦奇. 基于邻域粗糙集和 D-S 证据理论的雷达辐射源识别[J]. *舰船电子对抗*, 2020, 43(2): 48-54.
- [6] 金铃子,曹越操,元元浩,等. 基于声发射与 D-S 证据理论的截齿磨损状态识别[J]. *煤炭科学技术*, 2020, 48(5): 125-133.
- [7] 阮敏. 基于模糊特征抽取的图像识别算法研究[D]. 扬州:扬州大学, 2019.
- [8] 段宾,符祥,江毅,等. 结合 GAN 的轻量级模糊车牌识别算法[J]. *中国图象图形学报*, 2020(9): 1813-1824.
- [9] 郑仁思. 基于视觉传达的多维彩色图像模糊区域特征识别方法[J]. *自动化与仪器仪表*, 2020(2): 27-30.
- [10] 胡北辰,蔡瑞瑞. 多重纹理图像局部边缘模糊特征清晰化识别方法[J]. *太原学院学报(自然科学版)*, 2019, 37(4): 37-42.