

基于优化“未定义”类话语检测的话语领域分类*

柯子烜, 黄沛杰, 曾真

(华南农业大学数学与信息学院, 广东 广州 510642)

摘要: “未定义”类话语在面向任务的对话语料之中广泛存在, 具有成分复杂, 与其余“已定义”类话语边界模糊的特点, 影响着话语领域分类的总体正确率。而且, “未定义”类话语一旦错分, 将使得用户对口语对话系统的功能有效性产生怀疑, 大大降低用户体验。本文提出一种基于优化“未定义”类话语检测的领域分类方案, 采用了两阶段法完成口语话语的领域分类任务。首先采用聚类方法, 将“已定义”类话语聚为几个大类, 简化众多的“已定义”类话语独立存在时和“未定义”类话语之间的边界。进而利用分类模型对聚类后的“已定义”类话语大类以及“未定义”类话语进行领域分类, 优化目标是“未定义”类话语的检测效率。最后, 将第一阶段分类为“已定义”类的话语, 在去除了绝大部分“未定义”类话语干扰的基础上进行再次分类。本文的分类模型采用了长短期记忆网络(long short-term memory, LSTM), 并利用无标签微博数据训练词向量用于话语特征表达。在 SMP 2017 意图领域分类比赛的多任务语料上的评测结果表明, 本文的方案在“未定义”类话语检测的 F1 值以及所有话语的领域分类总正确率上都有明显提升。

关键词: 话语领域分类; 未定义类; 两阶段法; 词向量; LSTM

中图分类号: TP391

文献标识码: A

Domain Classification Based on Undefined Utterances Detection Optimization

KE Zixuan, HUANG Peijie, ZENG Zhen

(College of Mathematic and Informatics, South China Agricultural University, Guangzhou 510642, China)

Abstract: Undefined utterances are omnipresent in task-oriented corpus. The complex component and the vague decision boundary between undefined and defined utterances have negative impact on the total accuracy of domain classification. Also, once the system misclassifies the undefined utterance, user may tend to have their doubt on the system's function, which may cause greatly decline of user experience. This paper tackles this problem by proposing an effective domain classification method based on optimizing the undefined utterance detection. We adopt a two-stage framework, i.e. firstly, a cluster algorithm is used to aggregate defined utterances into several super groups, which can simplify the decision boundary between the original defined utterance groups and the undefined one. Secondly, a domain classifier is used to classify the defined utterance super groups and undefined utterance group. The optimization object is the efficiency of undefined utterances detection. Finally, those utterances classified as defined ones will be classified again, which has removed the disturbance of most undefined utterances. We adopt the long short-term memory (LSTM) networks as the classifier and trained the word embedding from an unlabeled Weibo dataset for utterance feature representation. The evaluation in the multi-assignment corpus of SMP 2017 domain classification competition shows that the proposed method makes obviously improvement on the F1 score of undefined utterances detection and domain classification accuracy of all utterances.

Key words: domain classification; undefined tag; two-stage framework; word embedding; LSTM

*收稿日期:

定稿日期:

基金项目: 国家自然科学基金(71472068); 国家级大学生创新训练计划项目(201710564154)

1 引言

口语语言理解 (spoken language understanding, SLU) 是口语对话系统 (spoken dialogue system, SDS) 中的重要环节, 而话语分类 (utterance classification) 则是 SLU 的关键任务之一^[1]。话语分类可进一步细分为话语领域分类 (domain classification)、话语意图分类 (intent classification)、以及对话行为分类 (dialogue act classification)^[1-3]。其中, 领域分类的任务是把话语划分到定义好的不同领域标签^[3], 进而将话语正确的分进不同的 SLU 子系统。如用户提出“帮我写一封邮件”, 系统则应该将其划分到“邮件”领域之中, 使得系统能对该话语进行专门针对“邮件”领域的语言理解。

目前领域分类的标签多由人工设计, 尽管通过精巧的语料库设计, 往往能将领域覆盖得比较全面。但总有一些覆盖不到的话语, 不得不采用“未定义”的标签。而这些“未定义”标签在不同的语料之中可能表现为不同的类别。这些话语在面向任务语料之中往往表现为“闲聊”类、“超出领域”类或“其他”类, 与语料涉及的任务和领域无关。如第六届全国社交媒体处理大会 (SMP 2017) 中文人机对话技术评测中的用户意图领域分类语料 (SMP 2017 DC Dataset) 的“chat”领域与其他面向任务的领域如“health”、“cookbook”等明显不同, 故其可以为“未定义”类别, 而其他涉及任务的标签均为“已定义”类别。又如Tur等人^[1]使用的语料中的“其他”领域则为“未定义”类别, 其余25个领域为“已定义”类别。

事实上, 这样的话语绝非少数, 在 SMP 2017 DC Dataset 中, “未定义”类 (chat 类) 为 609 条, 占比达到 19.84%。研究表明, 对话系统中大约有 20% 的用户话语属于“未定义”类^[4-5], 其重要性以及误识别时的危害性可以从两个方面考究:

(1) 错把“已定义”类分到“未定义”类中

在实际应用场景中, 用户使用多或者单任务对话系统, 往往为了解决实际问题。如: 用户发出“发烧怎么处理?”的话语, 对话系统应该准确将其分到“健康”领域中。一旦错分到“未定义”类, 系统只能当作任务领域之外的话语来处理, 这会导致用户对系统的功能有效性产生怀疑。

(2) 错把“未定义”类分到“已定义”类中

有时候用户在对话过程中与系统闲聊。如用户发出“你多大了?”的话语, 系统应该将其分到“闲聊”话语之中。一旦错分到某个“已定义”类, 即某个领域内的话语, 则可能会做出“给某某打电话”这样出乎预料的反应, 大大降低用户体验。

本文提出一种基于优化“未定义”类的领域分类方案。整个方案可分为两阶段, 第一阶段首先区分“未定义”类与“已定义”类。由于“未定义”类语义开放性以及表达多样性的特点, 其与单个“已定义”类之间的边界可能较为复杂, 分类器进行区分时可能难度较大, 所以本文首先使用聚类方法, 将“已定义”类聚为几个大类, 再利用分类模型对聚类后的“已定义”类大类以及“未定义”类进行区分, 某种程度上解决了“未定义”类与“已定义”类中某些类别容易混淆, 边界不清的问题。随后, 被认为是“已定义”类的话语将进入第二个阶段, 将大类重新拆开, 分成符合要求的“已定义”类。在分类模型上, 本文采用长短期记忆网络 (long and short-term memory, LSTM), 并通过交叉验证的方式进行了聚类中心个数以及 LSTM 模型参数的选择, 相比于已有的研究, 本文的主要贡献包括:

(1) 提出优化“未定义”类话语检测的两阶段领域分类方案。首先完成“已定义”类与“未定义”类话语的分离, 对“已定义”类进行聚类的方法让邻近以及易于混淆的类别聚集成大类, 简化众多的“已定义”类话语独立存在时和“未定义”类话语之间的边界。进而将第一阶段分类为“已定义”类的话语, 在去除了绝大部分“未定义”类话语干扰的基础上进行再次分类。

(2) 在领域分类比赛数据集 SMP 2017 DC Dataset 上评测了本文提出的方法, 取得了优于目前研究进展方法的效果, 结果验证本文方法在“未定义”类话语检测的 F1 值以及

所有话语的领域分类总正确率上都有明显提升。

本文后续部分安排如下：下一节介绍相关工作。第 3 节介绍本文提出的方法。第 4 节给出测试结果及分析。最后，第 5 节总结了本文的工作并作了简要的展望。

2 相关工作

领域分类一直是学术界、工业界的研究热点。其任务是给每一段话语打上预先定义好的领域标签。由于口语对话具有长度短小的特点，领域分类通常会被看作是短文本分类。早期的领域分类多采用较为复杂的人工特征，如语法信息、韵律信息、词汇信息等^[6-8]。

深度学习流行以来，许多研究者开始用深度学习方法解决自然语言处理（**natural language processing, NLP**）任务，这使得许多任务得到了长足的发展，其中也包括了领域分类。一部分学者利用神经网络进行特征提取，在 **ATIS** 数据集上取得了较好的领域分类结果^[9-13]。代表性的有 **Sarikay**^[9]等人提出的深度置信网络（**deep belief network, DBN**）。另一部分学者使用卷积神经网络（**convolutional neural networks, CNN**）进行领域分类^[1,14]，如 **Tur**^[11]等人提出利用深度凸网络（**deep convex network, DCN**）进行领域分类，并取得优于传统提升方法（**boosting**）的结果。**Ravuri** 等人^[2]和 **Shi** 等人^[15]则利用 **LSTM** 得到语料中上下文序列信息，并取得优于传统方法的领域分类效果。也有一部分研究人员认为，槽标签（**slot tags**）和领域标签都能代表用户的语义信息，可以通过共享一部分信息来得到更好的结果，许多网络结构因此将此两者结合一起训练^[16-17]。

上述研究进展的方法都对领域分类提供了很好的借鉴。然而他们并没有特别关注“未定义”类，只是将其作为普通类别处理，更多从特征工程和模型优化方面考虑。实际上，与“已定义”类相比，“未定义”类的分类具有以下挑战：

（1）成分复杂。“未定义”类本身还可以分为若干个类别。如“骂人”、“身份信息”、“致歉”等。这样的特点决定了其成分复杂多样，难以直接与其他已经定义好的类进行区分。

（2）与其他类边界难以确定。由于成分复杂，分类器进行分类时，如果不对“未定义”类进行专门处理，分类边界确定将变得尤为困难，容易混淆“未定义”类与“已定义”类。

Lane 等人^[18]提出的利用二层层叠泛化（**Stacked generalization**）进行“超出领域”（**out-of-domain, OOD**）话语检测的方案，某种意义上也相当于进行了“未定义”类话语检测。然而他们的研究中，训练集不含 **OOD** 话语，即 **OOD** 话语只在测试集，训练过程由领域内（**in-domain, ID**）话语轮流作为临时 **OOD**，这样的做法从某种程度上可以使得 **ID** 各类别之间的分类边界更加明确，但并没有解决“未定义”类话语与众多的“已定义”类话语边界复杂和模糊的问题。

针对以上的研究进展，本文提出一种基于优化“未定义”类话语检测的领域分类方案，经过实验评估，能够较好地解决领域分类中“未定义”类准确率、召回率以及 **F1** 值较低的问题，并有效提高领域分类的总正确率。

3 基于优化“未定义”类话语检测的领域分类

3.1 总体技术架构

图 1 是本文提出的方法的总体技术架构。

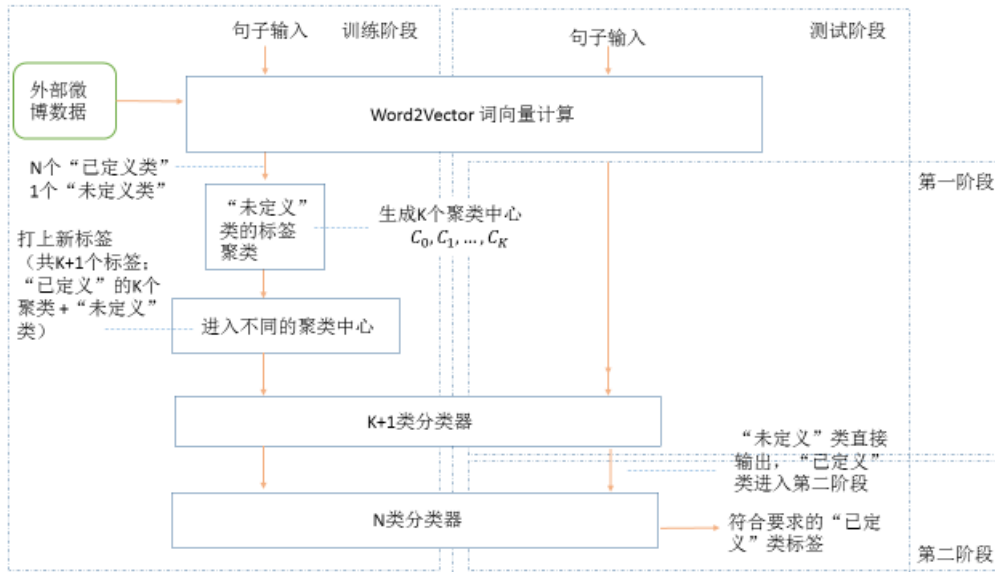


图 1 基于优化“未定义”类话语检测的领域分类

在这个架构中，主要分成两个阶段：

(1) 第一阶段首先进行“已定义”类与“未定义”类的区分：作为预处理，训练库利用外部微博数据以及 Word2Vector 生成词向量。同时，为了使得“已定义”类中相似标签之间更加易于“未定义”类相区分，我们选择合适的“已定义”类聚类中心数 K ，并由训练集生成聚类中心，为训练集打上聚类标签以及“未定义”类标签。测试开始时，首先利用 LSTM 模型将测试话语分到共 $K+1$ (K 个“已定义”类中心以及“未定义”类) 个标签中。

(2) 随后进入第二阶段，即进行“已定义”类内部的细分。第一阶段被分类为“未定义”类的话语将被保存，不进入第二个阶段的分类。第二个阶段只针对第一阶段中被分为 K 个“已定义”大类中某一个的话语进行分类。同样利用 LSTM 模型，将这部分话语分为满足条件的 N 个“已定义”类别。

3.2 LSTM 分类模型

本文使用的 LSTM 模型整体结构如图 2 所示。

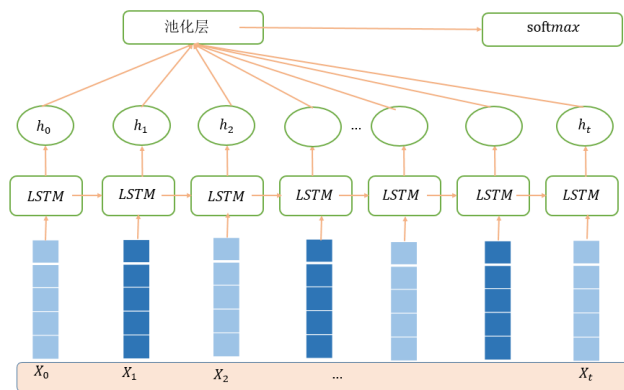


图 2 LSTM 模型

对于给定的 d 维词向量以及给定的 t 个时间序列构成的 X 输入，我们有：

$$h_t = \sigma(W^{hh}h_{t-1} + W^{hd}X_t) \quad (1)$$

$$y_t = \text{softmax}(W^{(s)}h_t) \quad (2)$$

其中， W 为权重矩阵， σ 为sigmoid， y 为类别标签。在池化层，我们利用所有 h 生成一个向量。一般选择机制有：平均池化、最大池化以及最后池化。平均池化选择所有 h 向量的平均值，最大池化选择最大的 h 向量，最后池化则选择最后一个 h 向量（如： h_t ）^[19]。

如果将每个时间序列中的LSTM单元（Cell）展开来看，那么其结构如图3^[17]所示。

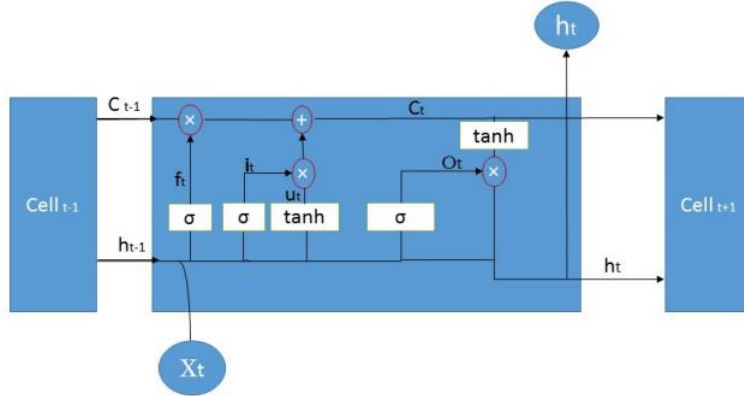


图 3 LSTM 模型单元结构及其对应的参数

图3中：

$$i_t = \sigma(W^{(i)}X_t + U^{(i)}h_{t-1} + b^{(i)}) \quad (3)$$

$$f_t = \sigma(W^{(f)}X_t + U^{(f)}h_{t-1} + b^{(f)}) \quad (4)$$

$$o_t = \sigma(W^{(o)}X_t + U^{(o)}h_{t-1} + b^{(o)}) \quad (5)$$

$$u_t = \tanh(W^{(u)}X_t + U^{(u)}h_{t-1} + b^{(u)}) \quad (6)$$

$$c_t = i_t \odot u_t + f_t \odot c_{t-1} \quad (7)$$

$$h_t = o_t \odot \tanh(c_t) \quad (8)$$

其中，输入为 d 维向量， i_t 为输入门， f_t 为遗忘门， o_t 为输出门， c_t 为记忆单元， h_t 为隐藏状态， \odot 为矩阵点乘。

3.3 聚类

聚类算法能将无标签的数据根据聚类中心数 K 分成 K 组，将含有某中联系的类别分到一个聚类标签之中。不同的聚类方法有不同的几何距离计算方法。本文采用经典的K-means算法对“已定义”类进行聚类，该方法已经在大量实际应用领域取得了较好的成果。

其效果示意图如图4所示。

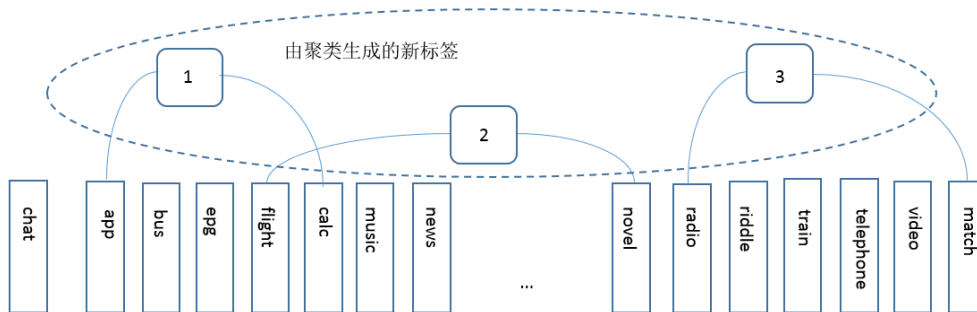


图 4 K-means 聚类效果示意图

如图 4 所示，本文实验将“已定义”类由 K-means 聚合为 K 个不同的类别，据此为“已

定义”类的话语打上新的标签。

其基本原理是不断更新聚类中心，使得聚类内平方误差最小：

$$\sum_{i=0}^n \min_{\mu_j \in C} (\|x_j - \mu_i\|^2) \quad (9)$$

大量的理论和实证研究都证明了 K-means 聚类方法能较好的将相似的样本聚集到某一特定聚类中心下。同时聚类中心个数 K 是影响聚类效果的重要因素^[20-21]。

4 实验

4.1 数据集

本文采用SMP 2017意图领域分类比赛提供的公开数据集SMP 2017 DC Dataset。数据共计31个类别，每个类别数据单独为一个文件，具体包括聊天类（chat）和垂类（30个垂直领域），其数据集中类别分布、举例及统计，如表1所示。

表 1 SMP 2017 DC Dataset 语料

标签	举例	话语数	标签	举例	话语数
“已定义”类					
cookbook	人造鸡蛋的配方。	358	weather	天气南京	88
video	帮我找一找龙门飞甲	242	music	我要听还是会这首歌	88
epg	今天湖南卫视的电子节目单	143	telephone	打电话给之山	84
poetry	搜索兰陵王。	136	message	给庄小雷发短信	84
stock	五八零零二二的市盈率是多少	95	flight	明天回广州的航班	83
train	贵溪至泉州的火车票	94	translation	中国制用英文怎么说	82
tvchannel	CCTV8高清电视剧	94	news	答复，我要听新闻。	78
health	颈椎病用什么药物治疗？	74	website	打开三g门户网站	72
app	请帮我打开uc	71	riddle	来个字谜我猜一下。	45
contacts	把哥哥的电话发给殷龙	40	schedule	十分钟后叫我起床	38
radio	调幅am六百二十八	32	lottery	我要查彩票的开奖信息。	32
calc	4乘6乘七再除以二	32	email	我想要转发这条邮件	32
cinemas	电影院有什么电影？	32	match	热火NBA季前赛的那个赛程。	32
bus	从中山到西安的汽车。	32	novel	我要看灵异推理小说	32
datetime	明天上午9点离现在还有多久	24	map	合肥百货大楼在哪里	91
“未定义”类					
chat	你睡午觉吗	609			

值得注意的是，这份数据集属于多任务语料库，除“chat”类外，其余每个类代表一个领域。根据之前的定义可知，这份数据中，“chat”类属于“未定义”类，其余30个类别均属于“已定义”类。

由于SMP 2017 DC Dataset的测试集还没有给出，我们利用目前已经发布的验证集作为测试集，利用训练集进行K折交叉验证（本文中K=5）模型参数选择。

SMP 2017 DC Dataset的总体情况如表2所示。

表 2 训练验证以及测试语料的情况

训练（交叉验证）话语数	测试集话语数
2299	770

具体地，整个SMP 2017 DC数据集语料中包括训练集2299条和测试集770条，共有3069句话语，其中“未定义”类话语有609句，占了19.84%。训练集中含“未定义”455句，占19.79%，测试集中“未定义”类有154句，占20%，表明了优化“未定义”类话语检测的研究和应用的价值。

4.2 实验设置

本文采用词向量的特征表达，训练词向量的外部数据库采用的是中国中文信息学会社交媒体专委会提供的 SMP 2015 微博数据集(SMP 2015 Weibo DataSet)。该数据集超过 500G，目前我们采用了其中的一个子集（1000 万条微博，519,734 词汇，约 1.5G）。词向量采用 Python Gensim 主题模型包中的 word2vec 进行训练。

实验方案为：

(1) 聚类中心个数选择：验证使用不同聚类中心进行聚类时，通过交叉验证“未定义”类检测的 F1 值，并从中选择“未定义”类 F1 值最高的方案

(2) 研究进展方法领域分类性能对比：对比了本文提出的方法与研究进展方法的领域分类结果。

本文的方法，先对“已定义”类聚类，优化“未定义”类话语检测，进而对“已定义”类再次分类的二阶段领域分类方案，记为 2Stage(Clustering+LSTM)，分类模型采用的是 LSTM。对比的研究进展的方法如下：

(1) LSTM：在研究进展中，LSTM 模型被采用在不同语料库上利用上下文特征进行领域分类^[3,14-15]。本文实验利用其 LSTM 模型直接对语料中的 31 个类别进行 31 分类。

(2) OOD Detection：Lane 等人^[18]提出的利用置信度层叠泛化模型（Stacked Generalization）进行OOD检测的方法。第一层输出ID分类器的置信度，第二层对第一层的置信度进行学习，并检测出OOD类。

4.3 实验结果与分析

4.3.1 聚类中心个数选择

聚类中心个数是进行聚类时的重要参数，本文验证 8 个不同的“已定义”类聚类中心个数方案，即 K=1 至 8，如图 5 所示。K=1 即为将所有“已定义”类话语作为一个大类，“未定义”类话语作为一类。

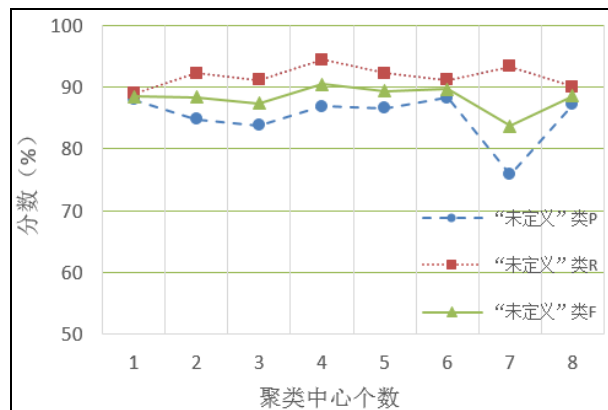


图 5 “已定义”类聚类中心个数验证结果

在本文的实验中，当 $K=4$ 时，相应的“未定义”类的 F1 值最高，据此，本文选择 $K=4$ 作为“已定义”聚类中心数。

4.3.3 研究进展方法领域分类性能对比

本文的方法与研究进展方法的领域分类结果如表 3 所示。根据图 4 的对比，本文方案的“已定义”聚类中心个数选择为 4。

表 3 本文方法与研究进展方法的领域分类对比

方案	总体分类正确率(%)	“未定义”类话语检测		
		准确率(%)	召回率(%)	F1 值 (%)
LSTM	77.2	88.6	85.7	87.1
OOD Detection	56.2	3.12	0.65	1.01
2Stage(Clustering+LSTM)	83.8	97.2	90.9	94.0

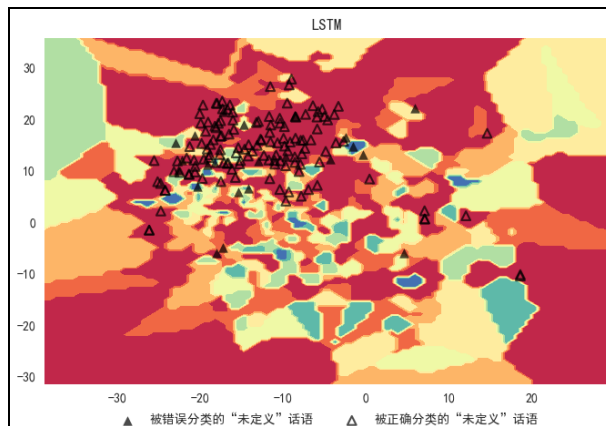
从表 3 可以看到，本文提出的方法比 Lane 等人^[18]的 OOD Detection 方法和 LSTM 模型方法分别提高了 27.6% 和 6.6% 的领域分类总体正确率。在“未定义”类话语检测的 F1 值方面，本文提出的方法也比 LSTM 模型提高了 6.9%。其中，OOD Detection 方法的“未定义”类 F1 值几乎为零，其原因主要在于按该方法，在第一层输入之前去掉了训练集中所有的“未定义”类，使得信息大大丢失，因此在第二层进行 OOD 检测时失准。

4.4 案例分析

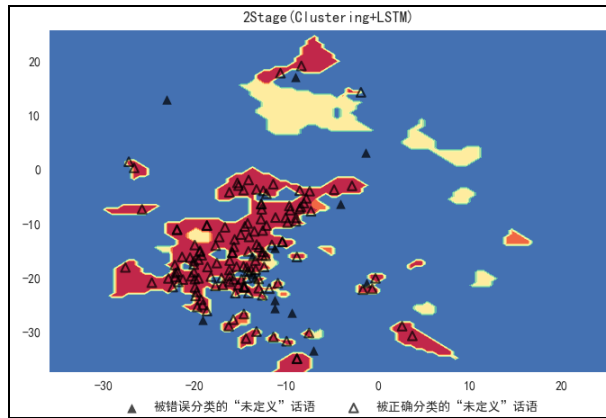
上述实验表明，本文提出的基于优化“未定义”类话语检测的领域分类方案能有效提高总的分类正确率以及“未定义”类话语检测的 F1 值。

4.4.1 “未定义”类分类结果

从整体上看，利用 T-SNE^[22-23]以及 KNN^[23]模拟出如图 6 所示边界（三角形代表“未定义”类，空心代表正确分类，实心代表错误分类。圆形代表“已定义”类，同样是空心代表正确分类，实心代表错误分类。深红色区域表示分类器认为是“未定义”类的区域）。可以看到，相比于本文的 2Stage(Clustering+LSTM)方法，单纯 LSTM 分类时，有较多的“未定义”类错分到“已定义”之中，造成了“未定义”类的召回率较低。相反的，应用聚类方法后，红色区域能够包揽更多的“未定义”类话语，使得召回率大大提升。



(a) LSTM 方案



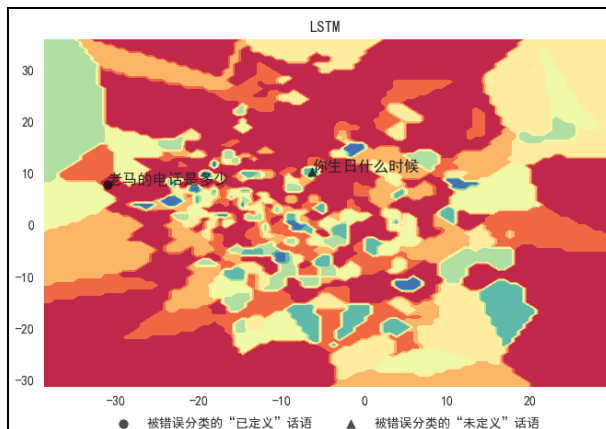
(b) 本文方案

图6 “未定义”类分类结果

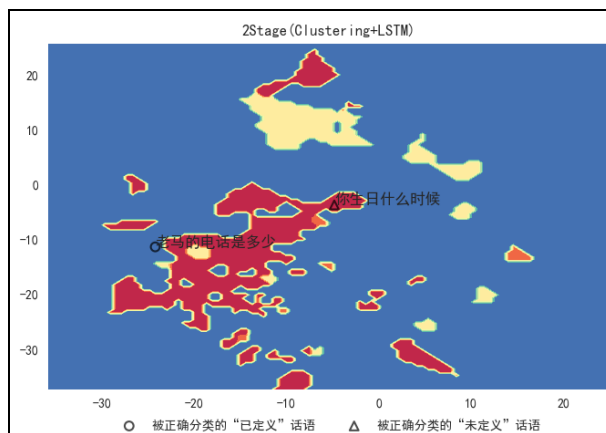
值得注意的是，本文使用 KNN 算法近似算出边界，实际边界必定在近似边界附近，但不保证一定处在近似边界位置^[24]。同时，由于 T-SNE 降维算法对于展示效果的自适应，图6中(a)和(b)两子图的点投影后位置发生改变，但仍可以通过话语序号追踪到点的情况。

4.4.2 典型话语追踪

下面结合两个典型的从错分中纠正过来的话语例子探讨本文方法的效果，如图7所示。



(a) LSTM 方案



(b) 本文方案

图7 典型话语的追踪

(1) 错分为“未定义”类的纠正

在未对“未定义”类进行专门处理时，分类器容易误把“已定义”类认为是“未定义”类。如：话语“老马的电话是多少”，“多少”这个词在“未定义”类中出现多次，而在“已定义”类的“contacts”类的训练集中的没出现，极容易被误分到“未定义”类中。使用本文方案后，该句在第一阶段被分到了某个大类，进而可以在第二阶段排除“未定义”类的干扰进行“已定义”类的细分，大大提高其正确率。如图 7 圆形所示，(a)中直接被分到“未定义”类中，而(b)则分到“已定义”类中，并且正确分到“contacts”类。

(2) 错分为“已定义”类的纠正

同样，在未对“未定义”类进行专门处理时，一些“未定义”类的话语也容易被误分为“已定义”类。如：话语“你生日是什么时候”，出现“什么”的关键词，“什么”在“epg”类中出现多次，如“今天电影频道演什么”、“现在热门的电视剧是什么”等，含“什么”的话语占比达 47%，所以分类器容易将其直接判定为“epg”类。但使用本文方案后，“什么”并不在某一个大类中占多数，如果在第一阶段先区分“已定义”与“未定义”，则该句在第一阶段能被划分到“未定义”类。如图 7 三角形所示，(a)中被分到了“epg”之中，(b)中该话语重新被“未定义”类有效召回。

5 结束语

本文利用聚类算法和 LSTM 分类模型，提出了基于优化“未定义”类话语检测的领域分类方案。在 SMP 2017 意图领域分类比赛数据集的测试表明，本文的方法取得了优于研究进展中的领域分类方法的效果。在有限训练数据集的条件下，相比于基于置信度的 OOD 检测和 LSTM 等模型，在“未定义”类话语检测的 F1 值以及所有话语的领域分类总正确率上都有明显提升。未来计划通过分析存在的分类错误样例，并结合进一步扩大的训练库和测试集，深入探索领域分类决策边界优化机制。

参考文献

- [1] Tur G, Deng L, Hakkani-Tür D, et al. Towards deeper understanding: Deep convex networks for semantic utterance classification[C]// Proceedings of the 37th IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2012), 2012:5045-5048.
- [2] Stolcke, A., K. Ries, N. Coccaro, et al. Dialogue act modeling for automatic tagging and recognition of conversational speech [J]. Computational Linguistics, 2000, 26(3), 339-373.
- [3] Ravuri S, Stolcke A. A comparative study of recurrent neural network models for lexical domain classification C// Proceedings of the 41th IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2016), 2016: 6075-6079.
- [4] Gorin A, Riccardi G, Wright J. How may I help you?[J]. Speech Communication, 1997, 23(1-2): 113-127.
- [5] Durston P, Farrell M, Attwater D., et al.. OASIS natural language call steering trial[C]// Proceedings of 7th European Conference on Speech Communication and Technology (Eurospeech 2011), 2001: 1323-1326.
- [6] Haffner P, Tur G, Wright J H. Optimizing SVMs for complex call classification[C]// Proceedings of the 28th IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2003), 2003: I-632-I-635.
- [7] Yang X, Loukina A, Evanini K. Machine learning approaches to improving pronunciation error detection on an imbalanced corpus[C] // Proceedings of the 4th IEEE Workshop on Spoken Language Technology (SLT 2014), 2014:300-305.
- [8] Chelba C, Mahajan M, Acero A. Speech utterance classification[C]// Proceedings of the 28th IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2003), 2003:I-280-I-283.

- [9] Sarikaya R, Hinton G E, Ramabhadran B. Deep belief nets for natural language call-routing[C]// Proceedings of the 36th IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2011), 2011:5680-5683.
- [10] Sarikaya R, Hinton G E, Deoras A. Application of Deep Belief Networks for natural language understanding[J]. IEEE/ACM Transactions on Audio Speech & Language Processing, 2014, 22(4):778-784.
- [11] Wang Z, Yang Y, Chang S, et al. Learning a deep ∞ encoder for hashing[C]// Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence (IJCAI 2016), 2016: 2174-2180.
- [12] Chen Y N, Hakkani-Tur D, Tur G, et al. Knowledge as a teacher: Knowledge-guided structural attention networks[J]. arXiv preprint arXiv:1609.03286, 2016.
- [13] Chen Y N, Hakkani-Tur D, Tur G, et al. Syntax or semantics? knowledge-guided joint semantic frame parsing[C]. Spoken Language Technology Workshop. IEEE, 2017:348-355.
- [14] Xu P, Sarikaya R. Contextual domain classification in spoken language understanding systems using recurrent neural network[C]// Proceedings of the 39th IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2014), 2014: 136-140.
- [15] Shi Y, Yao K, Chen H, et al. Contextual spoken language understanding using recurrent neural networks[C]// Proceedings of the 40th IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2015), 2015: 5271-5275.
- [16] Bhargava A, Celikyilmaz A, Hakkani-Tur D, et al. Easy contextual intent prediction and slot detection[C]// Proceedings of the 38th IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2013), 2013: 8337-8341.
- [17] Xu P, Sarikaya R. Convolutional neural network based triangular CRF for joint intent detection and slot filling[C]// Proceedings of the 2013 IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU 2013), 2013: 78-83.
- [18] Lane I.R., Kawahara T., Matsui T., et al. Out-of-Domain Utterance Detection based on Confidence Measures from Multiple Topic Classification [J]. IEEE Transactions on Audio, Speech, and Language Processing, 2007, 15(1): 150-161.
- [19] Khanpour H., Guntakandla N., Nielsen R. Dialogue Act Classification in Domain-Independent Conversations Using a Deep Recurrent Neural Network[C]// Proceedings of the 26th International Conference on Computational Linguistics (COLING 2016), 2016: 2012–2021.
- [20] Rodriguez A., Laio A. Machine learning. Clustering by fast search and find of density peaks[J]. Science, 2014, 344 (6191) :1492
- [21] Collobert R., Weston J., Bottou L., et al. Natural language processing (almost) from scratch[J]. Journal of Machine Learning Research, 2011, 12:2493–2537.
- [22] Maaten L. Accelerating t-SNE using tree-based algorithms [J]. Journal of Machine Learning Research, 2014, 15(1): 3221-3245
- [23] Maaten L. Learning a parametric embedding by preserving local structure[J]. Journal of Machine Learning Research, 2009, 5:384-391.
- [24] M. A. Migut, M. Worring, C. J. Veenman. Visualizing multi-dimensional decision boundaries in 2D[J]. Data Mining and Knowledge Discovery, 2015, 29(1): 273–295.

作者简介：



柯子焜（1995—），本科，主要研究领域为自然语言处理。

Email: iscauzixuanke@gmail.com



黄沛杰（1980—），通讯作者，博士，副教授，主要研究领域为人工智能、自然语言处理、口语对话系统。

Email: pjhuang@scau.edu.cn



曾真（1996—），本科，主要研究领域为自然语言处理。

Email: zz0807@126.com