

scau_SIGSDS 中文人机对话技术用户意图领域分类任务评测报告

杜泽峰, 梁泳诗, 何宇虹, 陈俊安, 梁超杰, 黄培松, 丁健德, 朱建恺, 黄沛杰*

(华南农业大学数学与信息学院, 广东 广州 510642)

摘要: 本文介绍华南农业大学口语对话系统研究室 (scau_SIGSDS) 开发的话语领域分类系统。我们的系统结合了基于训练语料的多分类器以及基于知识的领域识别。分类模型采用了长短期记忆网络 (LSTM), 领域知识包括了单关键词和组合关键词, 关键词来源包括训练语料以及人工构建的知识表。我们参加了 2018 中文人机对话技术评测 (ECDT) 中的用户意图领域分类任务, 在训练数据的交叉验证中取得了 95.07% 的正确率, 但在最终测试中由于测试集和训练集差异较大, 正确率有较大幅度的下降。

关键词: 话语领域分类; 人机对话; LSTM; 封闭式; 开放式

1 引言

近年来, 人机对话技术, 也称为口语对话系统 (spoken dialogue system, SDS) 受到了学术界和产业界的广泛关注。学术上, 人机对话是人机交互最自然的方式之一, 其发展影响及推动着语音识别与合成、口语语言理解、对话管理以及自然语言生成等研究的进展; 产业上, 众多产业界巨头相继推出了人机对话技术相关产品, 如个人事务助理、娱乐型聊天机器人等。以上极大地推动了人机对话技术在学术界和产业界的发展。在第八届全国社交媒体处理大会 (SMP 2018) 上, 哈尔滨工业大学和科大讯飞股份有限公司组织了 SMP 2018 中文人机对话技术评测 (ECDT), 为人机对话技术相关的研发人员提供了一个良好的沟通平台。

口语语言理解 (spoken language understanding, SLU) 是 SDS 中的重要环节, 而话语领域分类 (domain classification) 则是 SLU 的关键任务之一^[1]。话语领域分类的任务是把话语划分到定义好的不同领域标签^[2], 进而将话语正确地分进不同的 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“email”领域之中, 对该话语进行专门针对“邮件”领域的语言理解。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^[3-4], 分类模型采用传统的统计学习模型, 如随机森林、隐马尔科夫、条件随机场等。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, 许多任务得到了长足的发展, 也包括了领域分类^[2, 5-6]。代表性的模型包括了深度置信网络 (deep belief network, DBN)、卷积神经网络 (convolutional neural networks, CNN) 和长短期记忆网络 (long and short-term memory, LSTM) 等。

本文介绍我们参加 ECDT 评测中用户意图领域分类任务的系统, 采用了 LSTM 分类模型进行领域识别, 并采用基于数据的提取算法和外部领域知识表构建方法, 进一步优化领域分类效果。

2 scau_SIGSDS 话语领域分类系统

2.1 总体技术架构

图 1 是本文提出的方法的总体技术架构。在这个架构中, 主要有两个阶段:

(1) 基于 LSTM 的 31 领域分类器。我们选择了 LSTM 结合隐层的 31 分类的多分类器, 采用训练集经过交叉验证训练得到用于预测的模型。

(2) 基于知识的领域识别。初步开发了三种不同的基于知识的领域识别, 分别是基于领域知识表的领域识别、基于领域关键词的领域识别、以及基于组合词搭配的领域识别。

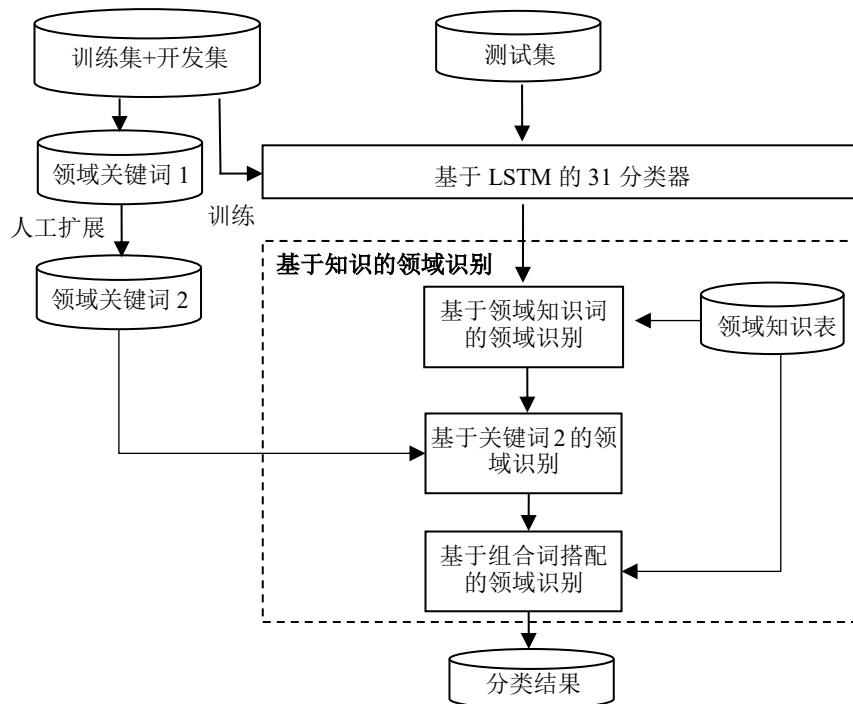


图 1 总体技术架构

2.2 基于 LSTM 的领域分类

相比于CNN，循环神经网络(Recurrent Neural Networks, RNN)有利于学习到句子中单词间的长距离依赖关系，但存在梯度消失/发散问题。目前常用的是RNN的一些变体，如LSTM、GRU(Gated Recurrent Unit)等，他们通过门控机制很大程度上缓解了RNN的梯度消失问题，并防止梯度发散。经典的LSTM模型整体结构如图2所示，LSTM记忆单元细节参阅文献[7]。

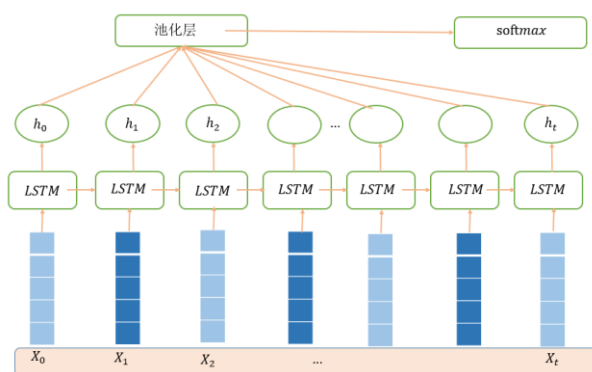


图 2 LSTM 模型

2.3 基于知识的领域识别

2.3.1 基于领域知识词的领域识别

这属于单关键词领域识别，关键词来自于领域知识表。通过分析领域话语特点，我们对除了 chat 领域的 30 个领域，结合外部信息构建了 30 领域知识表。当句子中含有的领域知识词没有在 chat 语料中出现且只在一个领域知识表中出现时，直接识别到对应的领域。对于有些领域，由于领域知识词较为开放，而训练语料又相对不足，不出现在 chat 语料中且只在一个领域知识表中出现的领域知识词不代表就具有领域识别性。因此，对于一些领域，如 music、stock 等，我们进一步要求领域知识词就是整个句子时才进行领域识别。

2.3.2 基于领域关键词的领域识别

这同样属于单关键词领域识别，但关键词来源不同。我们采用了基于数据的领域关键词提取算法，通过对训练集和开发集进行统计，抽取足够置信度和支持率的 2 和 3 字“词”构成

的领域关键词表。本系统采用的是置信度=0.95，支持率阈值根据领域类别样本数量分了 0.10、0.15 和 0.18 三个等级。并根据包含关系精简了领域关键词列表。例如，“七乐”和“七乐彩”都在抽取的到的另一关键词集时，保留了“七乐”。这做虽然预期具有更好的识别能力，但也在一定程度上提高了误判的风险。根据开放式评测的特点，我们进一步结合人工知识扩展了领域关键词表。

2.3.3 基于组合词搭配领域识别

这属于组合关键词的领域识别。通过分析领域话语特点，我们对 9 个领域设计了一套基于组合关键词的领域识别，包括 cinemas、stock、epg、radio、datetime、schedule、music、calc 和 contracts。设计的一套组合关键词采用正则表达式和领域知识表抓取句子中的特征词语，并归结为特征表达。例如，根据领域知识“北京”会识别为城市名，当句子同时出现“几点”时，形成 datetime 领域的“城市名”+“几点”的关键词组合，句子识别到 datetime 领域。

3 实验结果

3.1 数据集及任务介绍

SMP 2018意图领域分类任务的数据集共计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)。按比赛组织方提供的数据，训练集由2017年的训练、开发和测试集构成，数据集总体情况如表1所示。

表 1 训练和测试语料的情况

训练集	测试集
3750	4530

评价指标采用了正确率。

3.2 实验结果与分析

表 2 给出了我们的方案在 2017 测试结果(训练时采用的是 2017 的训练集和开发集进行交叉验证)，以及 2018 训练集(如前所述，由 2017 的训练集、开发集和测试集组成)的交叉验证结果。其中“+ 基于知识的领域识别”代表在基于 LSTM 的 31 分类器的基础上，进一步增加了 3 种基于知识的领域识别方法，包括了单和组合关键词的方法。

表 2 实验结果

方法	领域分类正确率(%)	
	2017测试集	2018训练集交叉验证
基于LSTM的31分类器	93.56	93.04
+ 基于知识的领域识别	95.05	95.07

从结果可以看到，我们的方案在2017的测试集以及2018的训练集交叉验证中都取得较好的分类正确率。但在2018的测试中，我们从测试日志（正式测试结果在本报告撰写时尚未公布）看到，由于语料相比于训练集有了很大的变化，正确率有了一定程度的下降，尤其是数据驱动的方法（基于训练语料训练的31分类器，以及基于训练语料提取的关键词）面临着很大的挑战。

4 结束语

本文介绍了我们参加 2018 中文人机对话技术评测（ECDT）中的用户意图领域分类任务的参赛系统的技术方案和线下实验情况。本文的方案在 2017 测试集以及 2018 训练集上都取得较好结果。在 2018 的测试中，由于语料相比于训练集有了很大的变化，正确率有了较大程度的下降，需要在未来工作中进一步完善本文的方案，尤其是知识表，以及知识和统计有机结合的方案。

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] 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.
- [3] 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.
- [4] 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.
- [5] 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.
- [6] Ravuri S, Stoicke A. A comparative study of neural network models for lexical intent classification[C]// Proceedings of the 2015 IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU 2015), 2015: 368-374.
- [7] Hochreiter S. and Schmidhuber J. Long Short-Term Memory [J]. Neural Computation , 1997, 9(8):1735–1780.