

基于融合模型的用户会话领域识别系统

王正魁, 胡嘉南, 桑海岩, 邱伟伟, 张世西, 贾志强

wangzhengkuiml@sina.com, {jianan.hu, haiyan.sang,

gavin.qiu, shixi.zhang, steve.jia}@cloudminds.com

达闼科技(北京)有限公司, 北京 100102

摘要: 本文介绍了达闼科技公司的自然语言处理实验室自主研发的会话领域识别系统, 该系统将语言规则与深度学习技术的优势有机地结合起来, 实现了对系统预制领域和用户自定义领域的良好支持。针对产品响应速度和并发量的需求, 该系统放弃使用循环神经网络等低效率模型, 而是采用先进、高效的 Transformer 和卷积神经网络模型。针对 SMP 2018-ECDT 竞赛要求进行该系统进行适当定制后, 荣获竞赛第 1 名。正文中详细介绍会话领域识别系统所采用的多模型融合的技术方案。

关键词: 人机对话, 领域识别, 语言规则, 卷积神经网络, Transformer 模型

1 引言

人机对话一方面为被服务者提供了更及时可在企业提高效率、降低成本方面发挥重要作用, 其中关键一步即是在会话过程中准确地识别出用户所关心的垂直领域。相对于其他文本分类任务, 具有待选类别多、语句相对口语化和实时性强等难点。

SMP 2018-ECDT 竞赛的数据集包含 31 个会话领域, 包括聊天类和 30 个垂直领域。一方面, 如果为 30 个垂直领域分别构造规则模板, 则工作量较大, 成本高; 另一方面, 由于语料资源有限, 完全依赖于深度学习模型难以获得良好的识别效果。因此, 本文针对 SMP 2018-ECDT 任务一所提供的语料, 提出了一种可将语言规则与机器学习模型有机结合的方案, 并在竞赛中取得了良好的效果, 在开发数据集上 F1 的宏平均值为 95%。

通过对本次竞赛的数据进行评测, 发现以下难点:

- 1) 语料资源较为有限;
- 2) 样本在不同领域中的分布严重不平衡;
- 3) 某些领域之间存在较高的相似度, 例如 tvchannel(电视台)和 ep(节目单), app 和 website。在缺乏上文的情况下, 无法判断“打开京东”是 app 领域或 website 领域。

针对上述三个难点, 本文分别采取了以下解决方法:

- 1) 利用大规模语料进行预训练, 提高模型泛化能力;
- 2) 对 chat 等样本较多的领域进行欠采样, 对 datetime 等样本较少的领域进行数据增强。但在增强过程中, 应注意不能影响样本的分布规律;
- 3) 采用标签光滑、dropout^[1]和复杂度正则化等技术降低模型过拟合风险, 从而弱化噪声的影响。

2 系统架构介绍

首先是进行实体扩展工作, 可靠的实体识别结果对于领域识别具有重要作用。例如, music 类的“来一首路灯下的小姑娘”和 poetry 类的“来一首静夜思”, 单纯依靠机器学习模型难以对其进行区分。一种有效的解决方案是先得到某些类的实体词典, 将其加入分词词典中, 如果分词得到了歌名则识别为 music 类。完成了多个实体词典, 并发现该方法显著提高了对应领域的召回率。

如图 1 所示，本系统分为 2 个阶段。第一阶段采用语言规则进行识别，如果识别的准确率高于阈值时，直接返回对应的识别领域；如果语言规则识别的准确率低于阈值时，则进入第二阶段，利用融合模型进行识别。

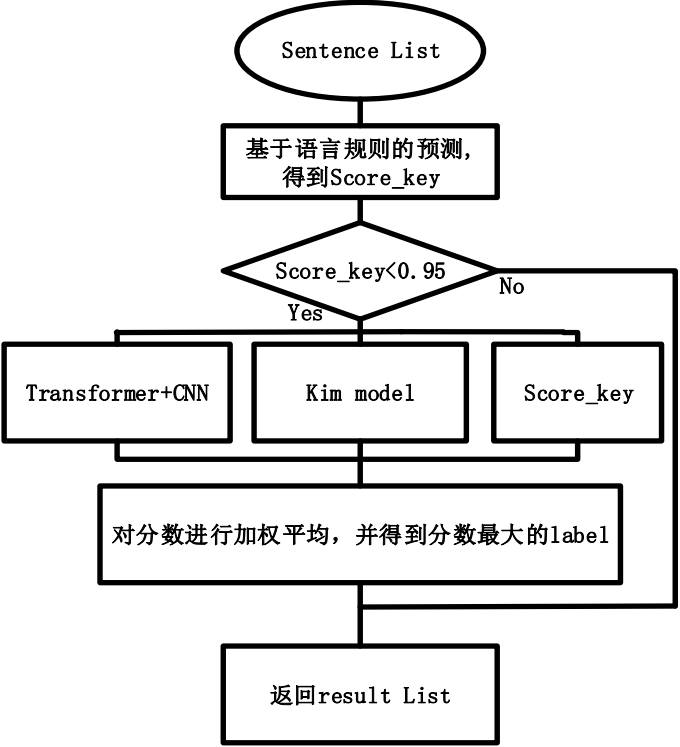


图 1 会话领域识别系统技术架构图

3 语言规则模型

通过分析语料，发现某些垂直领域具有明显的关键词特征，例如句子中包含“股票”即认为是 stock 领域的。然而，该方法简单快速，但可能会识别错误，例如将“股票用英语怎么说？”也划分到 stock 领域。此种方法未能充分利用句子中的信息，便武断地给出识别结果，无法获得具有较高的识别精度。因此，达闼科技公司生产环境中的领域识别系统采用的是更为可靠稳定的模板匹配方案，但是公司产品与本次竞赛对领域的划分方法存在一定差异，需要重新编写模板。由于时间仓促而未能将该模块迁移至本次参赛的系统中。

受限于语料的规模，本次竞赛中机器学习模型的识别能力较为有限。因此，不适合为语言规则模型设定一个较高的阈值，本次竞赛中将该阈值设定为 95%。开发数据集的评估结果表明，该语言规则模型在 5 个垂直领域上的准确率超过了 95%。

第二阶段中采用了三个模型，并将其预测的伪概率分布与语言规则预测的准确率进行加权融合，并给出得分最高的类别。下面详细介绍两个机器学习模型。为降低领域识别问题的难度，机器学习模型忽略了语言规则可以较为准确预测的 5 个垂直领域，给出句子在 26 个类别上的伪概率分布。

4 Kim 模型

Yoon Kim 在 2014 年提出了应用于文本分类问题的卷积神经网络模型^[2]，推广了卷积神经网络在 NLP 领域中的应用。Kim 模型将卷积神经网络在图像领域中捕捉局部相关性的优势迁移至文本分类任务中，先用不同的卷积窗口提取多种特征，然后分别进行最大池化，对特征进行拼接后，再用 2 层全连接网络进一步提取特征，最后由 softmax 层预测伪概率分布，如图 2 所示。

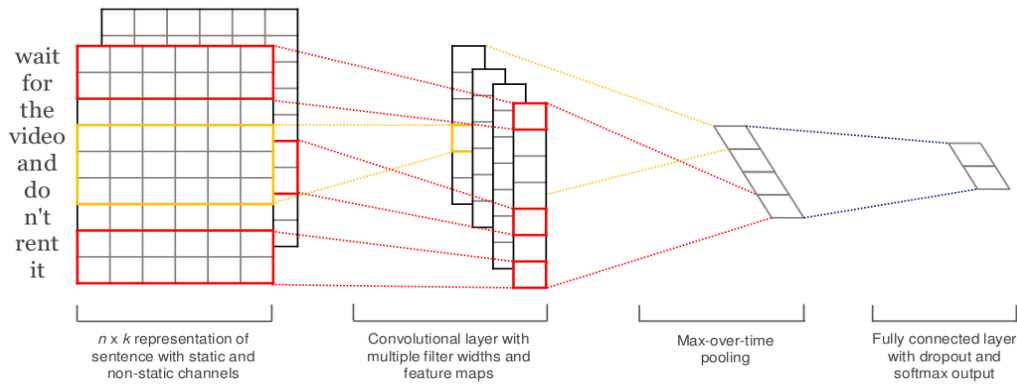


图 2 原始的 Kim 模型示意图

如图 2 所示，原始的 Kim 模型采用一层卷积神经网络对输入进行卷积，由于卷积窗口较小，难以考虑较长的位置序列信息。本文采用正弦和余弦函数对其输入张量添加位置编码。由于语料不足，所得词典无法覆盖常见会话用词，所以将原有模型的词向量部分换成字向量。为进一步提高模型的识别能力，利用达闼科技有限公司所研发的 SmartVoive 平台上累计的数据对字向量进行预训练。本文改进后的 Kim 模型在开发数据集上 F1 的微平均值和宏平均值均为 92.5%。

5 Transformer+CNN 模型

谷歌大脑在 2017 年提出 Transformer 模型^[3]，并在文本翻译任务中取得了突破。该模型不但在评价指标方面获得了 state-of-the-art 的水平，而且训练速度和推理速度也具有明显优势，自注意力机制具有易于并行计算和能捕获长距离依赖关系的优势。本文将编码模块中的 Transformer 模型（如图 3 所示）迁移至文本分类任务中。该模型中采用了多头注意力机制（见图 4 和图 5），允许模型在不同的表示子空间里学习到相关的信息。

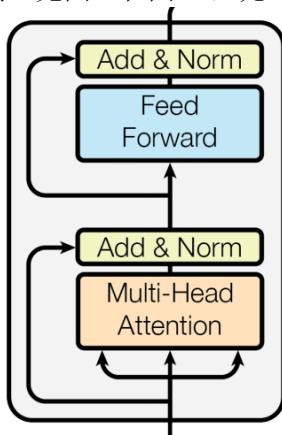


图 3 Transformer 模型

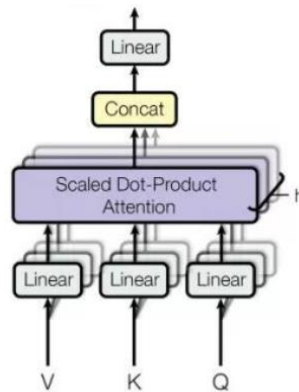


图 4 多头注意力机制

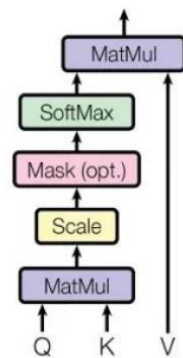


图 5 缩放点积计算注意力

但由于该模型在逐层叠加的过程中，与句子长度相关的维度保持不变，故多层的 Transformer 模型对于长句子训练效率相对较低。因此，本文只采用了一层 Transformer 模型，并在其上加上卷积网络和全连接层进一步提取特征，然后用 softmax 层预测其伪概率分布。因此，本文提出的 Transformer+CNN 模型结构如图 6 所示。



图 6 本文提出的 Transformer+CNN 模型示意图

通过分析语料发现，有小部分数据在划分领域时存在一定的模糊性，因此在计算损失函数时采用了标签平滑来缓解过拟合。为进一步避免过拟合问题，由于某些领域语料明显不足，在 Transformer 模型中和池化层后加入 dropout 机制，在全连接层采用了二范数的正则化。本文提出的 Transformer+CNN 模型在开发数据集上 F1 的微平均值为 92.9%，宏平均值为 91.4%。

6 分析与总结

比较本文改进的 Kim 模型和 Transformer+CNN 模型，发现它们在不同的评价指标下各有所长，具体到各个领域也是各有优势，而语言规则模型则在样本少的领域中效果较好。因此，本文利用信息融合的原理，对多个模型的打分进行加权平均，进一步提高了系统的识别能力。经测试，本文提出的两个神经网络模型不仅具有与 LSTM、GRU 等模型相当、甚至更高的识别能力，而且其推理速度分别为 17ms 和 55ms 比后者高出多倍，尤为适合产业应用。

由于时间有限，本次参赛的领域识别系统仍有待于进一步优化，例如采用交叉验证和数据增强可以有效缓解语料不足的问题。为适应社会的发展进步，系统采用的实体库也需要不断更新和扩充。值得庆幸的是，达闼科技公司研发的 SmartVoice 平台已经在众多场景下投入应用，在服务客户的同时也在向客户学习，不断累积资源，越来越智能。

人机对话是人工智能领域中的重要技术，也是一个融合了经验常识、个人情感和社会环境的复杂问题，仅依靠机器学习的统计方法和函数拟合的思想难以解决。达闼科技有限公司提出了将语言规则、知识图谱与深度学习技术相结合的解决方案，并将其成功应用于自主研发的 SmartVoice 平台。该平台易定制扩展、使用方便、响应迅速，并且采用了多媒体融合的交流技术，已经与中国移动、中国银行、郑州医院、金地集团、龙泉寺机器僧等客户进行业务的深度对接。达闼科技将与多方企业共同努力，在销售、服务和信息等领域实现 AI 化，用技术让生活更智能。

参考文献

- [1] Srivastava N, Hinton G E, Krizhevsky A, et al. Dropout: a simple way to prevent neural networks from overfitting[J]. Journal of machine learning research, 2014, 15(1): 1929-1958.
- [2] Kim Y. Convolutional neural networks for sentence classification[J]. arXiv preprint arXiv:1408.5882, 2014.
- [3] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin. Attention is all you need[C]. In Advances in Neural Information Processing Systems, pages 6000–6010, 2017.