

图 7 直接免疫概率对传播节点及免疫节点数量的影响

Fig.7 The influence of direct immunity probability on the number of propagation node and immune node

由图 7 可知,在网络达到稳态之前, p_{er} 取值越大,则 $I(t)$ 的值会越小,而 $R(t)$ 的值会越大,这是因为 p_{er} 表示接收节点转变为免疫节点的概率, p_{er} 值的增大,意味着网络中会有更多处于接收状态的节点,因为与传播节点间的兴趣匹配不一致,未经传播状态而直接转变为免疫状态.此外, p_{er} 的取值越小, $I(t)$ 的值趋近于 0 的时间就会更加滞后,这是因为随着 p_{er} 取值减小,网络中接收节点转变成免疫节点的速度降低,这就将导致需要更长的时间才能使传播过程达到最终的稳定状态.

3.3.3 初始节点对舆情传播过程的影响

现有研究成果表明,社交网络中信息发布者“影响力”的大小,将会对信息接受者的转发行为产生巨大影响^[20].本文基于 PageRank 算法^[21],根据社交网络用户节点间的好友关系来评估用户节点的“影响力”.设置模型参数 $p_{ei} = 0.2, p_{er} = 0.1, p_{ir} = 0.3, p_{se} = 1$,并分别选取网络中入度(in-degree)最大、出度(out-degree)最大以及“影响力”最大的节点作为初始传播节点,得到如图 8 所示的网络传播节点和免疫节点的数量随初始节点不同的变化曲线.

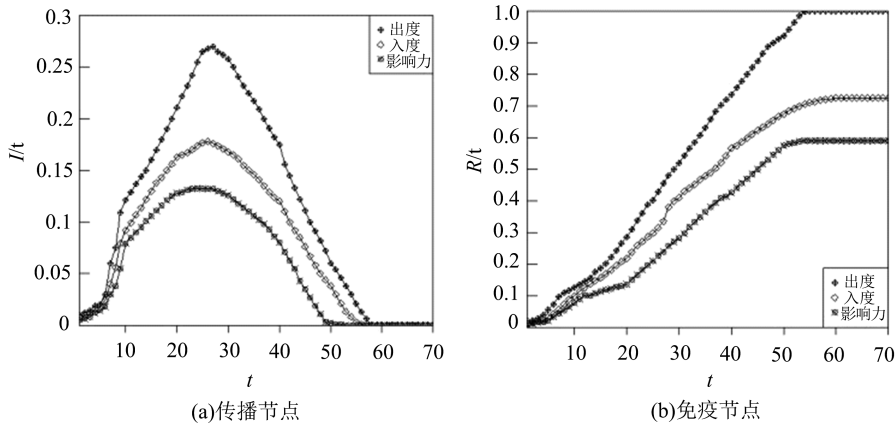


图 8 初始传播节点对传播节点及免疫节点数量的影响

Fig.8 The influence of initial propagation node on the number of propagation node and immune node

从图 8 的曲线可以看出,当网络中“影响力”最大的节点被选择为初始传播节点时,舆情信息在整个网络中的传播效果最好.

4 结论

实际的网络舆情传播过程,往往是通过在兴趣相似的网络用户之间进行信息分享而完成的,传播

主要面向兴趣匹配的传播主体.本文从传播学的角度,考虑网络中传播主体间兴趣差异对舆情传播的影响,建立与实际情况更为符合的网络舆情传播模型,重现网络舆情在整个网络上的传播过程链.

本文对从新浪微博中爬取到的数据集进行大数据分析,从中提取和分析社交网络中存在的多种显式或隐式关系.用户兴趣匹配关系是社交网络中广

泛存在的一种隐式关系,该关系直接影响网络舆情传播过程.基于此,本文构建了基于兴趣匹配的网络舆情传播模型,通过实验仿真分析了模型中各参数变化对传播过程的影响,从中发现网络舆情传播的规律以及影响舆情传播的主要因素,从而为社交网络中的舆情预测与危机管理提供理论支撑.针对现实互联网中,舆情信息传播往往是在多个社交网络中同时进行这一现象,应该研究舆情信息在多个社交网络中的传播规律,建立起多社交网络中的舆情传播模型.这些都是下一步的研究方向.

参考文献(References)

- [1] SUDBURY A. The proportion of the population never hearing a rumour[J]. *Journal of Applied Probability*, 1985, 22(2): 443-446.
- [2] ZANETTE D H. Critical Behavior of Propagation on Small-world Networks[J]. *Physical Review E*, 2001, 64(1): 901-907.
- [3] ZANETTE D H. Dynamics of rumor propagation on small-world networks [J]. *Physical Review E*, 2002, 65(4): 1-9.
- [4] MORENO Y, ZANETTE D H. Dynamics of rumor spreading in complex networks[J]. *Physical Review E*, 2004, 69(2): 130-137.
- [5] LIU Z, LAI Y C, YE N. Propagation and immunization of infection on general networks with both homogeneous and heterogeneous components [J]. *Physical Review E*, 2003, 67(3):11-19.
- [6] ZHOU J, LIU Z, LI B. Influence of network structure on rumor propagation[J]. *Physics Letters A*, 2007, 368(6): 458-463.
- [7] KESTEN H, SIDORAVICIUS V. The spread of a rumor or infection in a moving population[J]. *Annals of Probability*, 2003, 33(6): 2402-2462.
- [8] 潘灶烽,汪小帆,李翔.可变聚类系数无标度网络上的谣言传播仿真研究[J]. *系统仿真学报*, 2006, 18(8): 2346-2348.
- [9] 刘常昱,胡晓峰,司光亚,等.基于小世界网络的舆论传播模型研究[J]. *系统仿真学报*, 2006, 18(12): 3608-3610.
- [10] 刘常昱,胡晓峰,罗批,等.基于不对称人际影响的舆论涌现模型研究[J]. *系统仿真学报*, 2008, 20(4): 990-992.
- [11] ZHOU T. A brief review of complex networks[J]. *Physics*, 2005, 34(1): 31-36.
- [12] FANG J. Theoretical research progress in complexity of complex dynamical networks [J]. *Progress in Natural Science*, 2007, 17(7): 841-857.
- [13] FU F. Social dilemmas in an online social network: The structure and evolution of cooperation [J]. *Physics Letters A*, 2007, 371(1): 58-64.
- [14] 张亚明,唐朝生,李伟钢.在线社交网络谣言传播兴趣衰减与社会强化机制研究[J]. *情报学报*, 2015, 34(8): 833-844.
- [15] LIU N, LU Y, TANG X, et al. Multi-document summarization algorithm based on significance topic of LDA[J]. *Journal of Frontiers of Computer Science and Technology*, 2015, 9(2): 242-248.
- [16] YAN M, ZHANG X, YANG D, et al. A component recommender for bug reports using discriminative probability latent semantic analysis [J]. *Information and Software Technology*, 2016, 73(3): 37-51.
- [17] YU J, ZHU T. Combining long-term and short-term user interest for personalized hash tag recommendation [J]. *Frontiers of Computer Science*, 2015, 9(4): 608-622.
- [18] 张健沛,姜延良.一种基于节点相似性的链接预测算法[J]. *中国科技论文*, 2013, 8(7): 659-662.
- [19] 孙翎,迟嘉昱.基于仓室模型的危机蔓延建模与演化分析[J]. *管理学报*, 2010, 7(12): 1878-1883.
- [20] 张旸,路荣,杨青.微博客中转发行为的预测研究[J]. *中文信息学报*, 2012, 26(4): 109-114.
- [21] 李志宏,庄云蓓.基于 PageRank 算法的双维度微博用户影响力实时度量模型[J]. *系统工程*, 2016, 7(2): 128-137.

应用于校园心理咨询的对话匹配度预测模型

谭嘉莉, 何钰, 吴燕晶, 孙广中

(中国科学技术大学计算机科学与技术学院, 安徽合肥 230027)

摘要: 聊天机器人在学术界及工业界均受到了广泛的关注. 目前在学术界, 关于端到端对话回复的研究成果众多. 其中, 采用数据驱动的对话回复研究方法占主要地位, 且多基于深度学习神经网络学习与理解自然语言. 已有的对话回复模型多针对开放领域. 在聊天机器人中比较成熟的应用目前也多为娱乐型聊天机器人. 专业领域内的聊天机器人(如心理咨询聊天机器人)目前还多基于规则及模板. 为了提高心理咨询类聊天机器人的智能性, 提出一种应用于校园心理咨询场景下的对话匹配度建模方法. 该方法基于心理咨询网站及贴吧语料, 提取单词及句子在心理咨询类别上的相关特征, 并将此特征应用于机器学习及深度学习网络中进行句对匹配度建模. 与传统的开放领域内句对匹配模型相比, 该模型利用了心理咨询的领域分析知识, 能够达到更好的匹配效果.

关键词: 心理咨询; 聊天机器人; 对话回复; 机器学习

中图分类号: TP391 **文献标识码:** A doi: 10.3969/j.issn.0253-2778.2018.09.008

引用格式: 谭嘉莉, 何钰, 吴燕晶, 等. 应用于校园心理咨询的对话匹配度预测模型[J]. 中国科学技术大学学报, 2018, 48(9):739-747.

英文作者, et al. Dialogue matching prediction model applied in campus psychological counseling[J]. Journal of University of Science and Technology of China, 2018, 48(9):739-747.

Dialogue matching prediction model applied in campus psychological counseling

TAN Jiali, HE Yu, WU Yanjing, SUN Guangzhong

(School of Computer Science and Technology, USTC, Hefei 230027, China)

Abstract: Chat-bots have received wide attention in both academia and industry. In academia, there have been many promising research results in the end-to-end dialogue response area. Among them, data-driven dialogue response methods predominate, which learn and understand natural language through deep neural networks. Existing dialogue response models are mainly designed for open domains. The current mature chat-bot applications are mostly used for entertainment. Methods used on professional chat-bots (like psychological counseling chat-bots) are mainly based on rule and template. To enhance the intelligence of the psychological counseling chat-bot, a new method of modeling dialogue matching pattern in the context of campus counseling is proposed. This method is based on the psychological counseling website and Tieba corpus, from which relevant characteristics of words and sentences in the category of psychological counseling types are extracted, and are applied to machine learning and deep learning networks to model the dialogue matching pattern. Compared with traditional dialogue matching models in open domain, the

收稿日期: 2018-03-09; 修回日期: 2018-05-18

作者简介: 谭嘉莉, 女, 1993年生, 硕士生. 研究方向: 机器学习、数据挖掘. E-mail: tanjiali@mail.ustc.edu.cn

通讯作者: 孙广中, 博士/副教授. E-mail: gzsun@ustc.edu.cn

proposed model achieved better matching results with the use of analyzed psychological counseling information.

Key words: psychological counseling; chat-bot; conversation reply; machine learning

0 引言

大学阶段可以看作大学生从校园迈向社会的过渡阶段,也是其生理、心理发展的重要转折阶段,在此阶段学生生理上、心理上都将发生很大的变化。学业压力、社会竞争、生活方式的转变,可能使学生出现比以往更多的心理应激反应,产生心理失衡及不适应感,引发各种心理问题^[1]。学校应给予学生足够的关注,及时地发现问题,给予一定的引导与疏通,帮助学生不断地调整认知、正确面对,使他们能够保持心理上的健康。

在心理学领域,罗杰斯的理论和工作成为心理咨询和心理治疗的一个结合点。在此前,心理治疗都由医生来做,此后,非医疗人员也可以参加心理咨询的培训,对表现为轻度心理问题及心理紊乱的人进行辅导。这为聊天机器人在心理咨询领域的应用提供了一种可能性。在校园环境下,考虑到学生可获得的心理咨询帮助并不够,使用聊天机器人及时为同学们答疑解惑有利于他们保持心理健康。

目前,在对话回复领域中,聊天机器人在学术界、工业界受到了广泛的关注。工业界,微软小冰、苹果的 Siri、百度小度等聊天机器人的出现推动了聊天机器人的产品化发展。目前,在线客服、娱乐、教育、个人助理、智能问答等领域已经有了一些较成熟的应用,如京东 JIMI、微软小冰、讯飞开心熊宝、百度度秘、IBM Watson 等^[2]。学术界,对话回复的研究在不断深入。开放领域内对话回复方法得到了不断发展,但在一些特殊领域内,如心理咨询领域,尚未出现智能性较好的聊天机器人。

心理咨询场景下的对话具有一定的开放性。具体表现在咨询语句可以有无穷多种,对于同一个咨询句,它的答句是不固定的。使用基于规则及模板匹配的对话回复方法进行心理咨询领域内对话时,对话能力将会受到很大的限制。与完全开放领域内的对话回复相比,心理咨询领域内的对话回复对回复句的质量要求更高。给出的回复句不仅要求语句通顺,有一定的意义,也要与咨询句保持一定的一致性。同时,在进行心理咨询领域内对话回复时,需要根据来访者要咨询的类别,给出合适的回复。特别

地,当来访者所要咨询的类别为“异常心理”,如“人格障碍”、“自杀”等时,不应继续进行常规对话,而应当帮助引导用户寻求专业治疗人员的帮助。

开放领域内对话回复方式由于没有对句子的咨询类别进行分析,不能很好地针对来访者的咨询意图进行回复,同时也无法判断对来访者的干预是否在聊天机器人能力范围之内,因而不适合直接应用到心理咨询领域内。

基于数据的对话回复方式包括生成型对话回复及选择型对话回复,考虑到生成型对话回复本身容易产生无意义的句子,甚至是语法不正确的句子,不适合应用于心理咨询领域内的对话回复。本文基于选择型对话回复方式,提出了一种可应用于校园心理咨询的对话匹配度预测模型(consult type aware sentence matching model,CTASM)。该模型结合校园环境常见的若干种咨询类别,定义出单词及句子在咨询类别下的度量方式,并通过机器学习模型识别出句子所属的咨询类别,以此咨询类别信息指导后续句对匹配度的学习。与完全开放领域内的对话回复模型相比,CTASM 模型可以根据来访者的咨询语句判断出所属的咨询类别,一方面可以规避高危对话,另一方面也可以根据咨询类别更好地找出句子中的关键词。同时,模型中定义了单词及句子在各咨询类别下的相似度及相关度度量方式,训练的过程中将会对问句及答句在咨询类别上的相似性进行学习,从而使模型在判断句对匹配与否时,可以将问答句在咨询类别上的一致性考虑进去,而不仅仅是问答句在语义上的相似性。

本文的主要工作和贡献点在于:①提出了一种获取心理咨询相关语料的方法。②定义了单词及句子在咨询类别下的度量方式。③提出了一种新的寻找句子关键词的方法。④提出了一个应用于校园心理咨询场景下的对话匹配度预测模型。

实验结果表明,在心理咨询领域内进行选择型对话回复时,CTASM 模型效果要好于完全开放领域内对话回复方法。

1 相关工作

对话系统分为任务导向的对话系统以及非任务

导向的对话系统.现有的任务导向型对话系统多通过手动设定的特征及规则来完成对话状态跟踪及策略制定.设计一个可用的对话系统将耗费巨大的人力及时间,且设计出的系统缺乏灵活性,不能应用到其他领域内^[3].

非任务型对话系统主要应用于开放领域中的对话回复,对应的回复方式通常分为两种,一种为生成型对话回复,它通过编解码的方式逐字生成回复句,优点是生成语料中从未出现过的句子,但这种方式很容易生成较短的、与问句不相关,甚至语法错误的句子^[4],因此很多学者从不同方面尝试提高对话回复的质量.文献[5]通过最大化互信息提高生成式对话回复的回复相关性与多样性.也可以增加额外的信息来指导句子生成,这些信息包括关键词^[6]、主题^[7]、情感^[8]等.还有学者从对话系统的角度,提高系统的会话质量,如文献[9]赋予聊天机器人人格,使其在会话中可以保持一致性,文献[10]通过强化学习,提升对话交互的轮次,文献[11]让机器人在会话的某些时机主动引导对话.文献[12-13]将句子的情感向量嵌入到句子的语义向量中,生成的答复句在情感上有较好的表现.另一种是选择型对话回复模型,这种模型通过对句对的匹配度进行学习,从若干个备选答复中选择与问题匹配度最高的句子作为答复句.与生成型对话回复相比,选择型对话回复模型筛选出的句子在信息量及流畅性方面有较大的优势,回复句质量有一定的保证.文献[14]中利用循环神经网络对句子进行编码,得到问句及答句的单一向量表示,再通过对问答句相似度进行学习,计算问句及答句的匹配度,从而筛选出最佳的答复句.该方法比较简单,仅使用了句子编码的最终结果,是一个比较基础的对话匹配模型.文献[15]结合了局部信息及内在层次,提出了一个基于DNN的短文本对话回复模型.文献[16]通过使用深度卷积神经网络来学习问答句的向量表示,由单词之间的相似度矩阵,学习问答句之间的匹配模式,进一步提高了模型效果.文献[17]在利用相似度建模时,使用了句子向量生成过程中产生的隐变量,同时综合单词及句子两个层面的相似度矩阵,进行匹配度建模.文献[18]通过主题模型获取句子的主题,在建模过程中,不仅学习问答句在语义上的关联性,也会学习到问答句在主题上的关联性.文献[19]提出了一个多轮对话场景下的选择型对话回复模型.

有关对话系统在心理咨询领域的研究也有很

多.文献[20]研究了聊天机器人在精神疾病干预方面的适用性,并特别研究了聊天机器人在饮酒习惯评估上的有效性.研究表明,聊天机器人在治疗中可以起到积极的作用,但由于系统主要通过关键词及模板匹配完成,因此功能上十分受限.文献[21]通过收集用户手机上的行为数据,使用SVM进行多分类,判断用户的大五人格.文献[22]利用社交平台上的多模态数据预测用户的人格,并将此方法应用到聊天机器人DiPsy中,使其成为为私人电子心理医生.文献[23]则结合Carl Gustav Jung的理论,更加精细地对人格进行了分析.上述研究的目均准确地分析用户人格,而较少描述如何将模型应用到聊天系统中,实现一个可进行心理干预的聊天机器人.其中文献[22]中的聊天系统DiPsy在进行对话回复时使用的是开放领域内对话回复方法,仅当对用户人格的判断有一定把握时,会结合模型判断出的人格及预先设定的策略进行回复,因此目前为止还不是一个成熟的应用.

本文将开放领域内对话回复模型与心理咨询领域相结合,提出了一个用于校园心理咨询的句对匹配度预测模型CTASM.模型延续了单轮对话中问答句匹配所使用的二维匹配范式.在相似度衡量时,增加了从词共现角度计算得到的句对中单词之间的相关度矩阵.出发点在于,相匹配的句子,可能是句子中含有相似度较高的部分,也可能是含有相似度不高但相关度较高的部分.

在衡量相似性时,不仅从单词的Word2Vec分布式向量表示这个角度进行相似度衡量,还从单词在咨询类别下的重要性这个角度进行单词相似度衡量.使模型可以学习到问句及答句在咨询类别上的匹配模式.同时,考虑到直接将计算得到的相似度矩阵送入CNN时,默认矩阵中的各元素是没有区别的,而实际上应当着重考虑问句中的关键词与答句中各单词的相似度,而给非关键词较少关注,因此本文模型利用问句所属的咨询类别,得到问句中的关键词,并通过对关键词加权获得加权后的相似度及相关度矩阵,作为CNN的输入,以获取更好的匹配效果.

2 模型介绍

问句 $U = [u_0, u_1, \dots, u_m]$ 包括 m 个单词,待匹配的答句 $R = [r_0, r_1, \dots, r_n]$ 包括 n 个单词. $Label_{U,R} \in \{0, 1\}$ 表示问句 U 与答句 R 之间的匹配

度标签, Label 值为 0 时表示两个句子不匹配, 值为 1 时表示两个句子是匹配的.

2.1 模型概要

模型框架如图 1 所示. 其中 U, R 分别代表预处理之后的问句及答句, 主要包括两个模块.

咨询类别判断模块: 基于校心理咨询网站及贴吧的语料数据, 从中提取出单词及句子在各类别下的重要性特征, 学习得到一个可以判断句子所属咨询类别的模型图. $consultType$ 为句子 Sentence 对应的目标咨询类别. $consultVec$ 为句子在目标咨询类别下的重要性向量.

句对匹配度判断模块: 基于贴吧的对话语料数据, 及上一个模块中的咨询类别判断模型, 从单词与句子的层面计算出若干相似度与相关度矩阵, 进行句对匹配度建模. 图中 $word_wv, word_tf, CoApp_matrix$ 分别代表单词的两种向量表示方式及词共现矩阵, 均由语料计算得到. Sim_* 为计算得到的问答句之间的相似度及相关度矩阵, $weighted_*$ 为加权后的相似度及相关度矩阵.

下面详细介绍咨询类别判断模型、句对匹配度判断模型以及两者之间的关系.

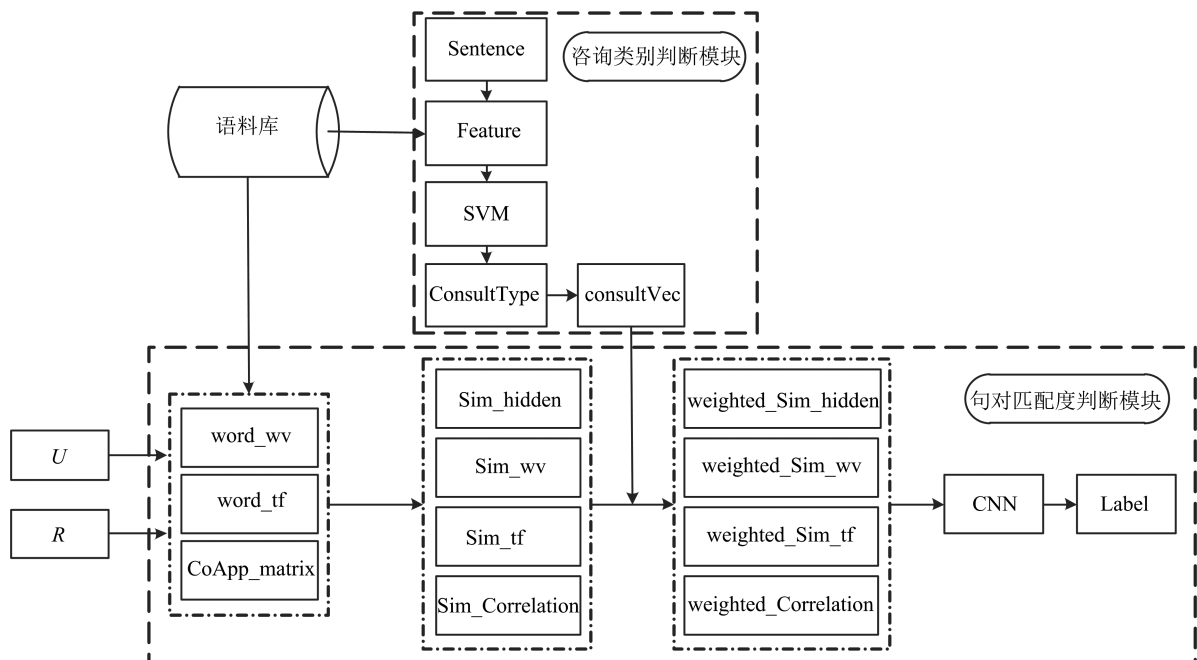


图 1 模型框架示意图

Fig.1 Model frame diagram

2.2 咨询类别判断模型

设咨询类别判断模型的语料为 X_{CT} , $X_{CT} = \{(S_1, consultTypes_{S_1}), (S_2, consultTypes_{S_2}), \dots\}$, 其中 S_i 表示语料中第 i 个句子, 即样本 i , 由若干个单词组成. 句子 S_i 对应的咨询类别标签为 $consultTypes_{S_i}$. 在校园场景下常见的咨询类别包括学习、恋爱、人际、择业、适应、异常心理这几个类别. 咨询类别判断模型能够根据用户发来的咨询语句判断出其所要咨询的类别. 其中在识别为异常心理时, 不宜继续进行常规对话回复, 而应当引导用户寻求其他渠道的帮助. 另外增加一个“无关”类别, 表示该句子不属于心理咨询相关的语句.

对于句子 S , 其真实咨询类别为 $consultTypes_S$, 我们要学习到一个咨询类别判断模型 f_{CT} , 使对于任意句子 S , 它对应的咨询类别

$consultTypes_S$ 可以通过 $f_{CT}(S)$ 进行预测.

该咨询类别判断模型有以下几个作用: ①判断咨询语句的目标咨询类别, 帮助理解来访者的咨询意图. ②帮助聊天系统判断对话是否超出能力范围. ③利用该模型判断咨询语句是否为心理咨询相关的句子, 在实验部分将帮助筛选出心理咨询相关的语料. ④作为前导模型, 帮助从句对中提取心理咨询相关的特征, 为后续的句对匹配模型提供必要的信息. 如提供一种咨询类别上单词及句子的分布式向量表示, 从咨询类别这个角度计算出单词之间的相似度. 同时, 结合模型判断得到句子的目标咨询类别及单词在目标咨询类别下的重要性值, 有助于更好地找到句子中的关键词.

设有 $numConsultType$ 个咨询类别, 经过预处理之后语料中共有 $numWords$ 个单词, 其中单词

w_i 在咨询类别 T_j 上出现次数为 $TF_{w_i}^{T_j}$. 下面给出单词及句子在咨询类别下度量方式的定义.

定义 2.1 单词在咨询类别下的度量 $tw_{w_i}^{T_j}$ 、 tw_{w_i}

$tw_{w_i}^{T_j} \in R$ 为单词 w_i 在咨询类别 T_j 上的重要性值, 计算公式为

$$tw_{w_i}^{T_j} = \frac{TF_{w_i}^{T_j}}{\sum_k TF_{w_i}^{T_k}} \quad (1)$$

即单词在该类别下出现的次数与在语料中总出现次数的比值. 它可以表示单词在指示咨询类别上的能力. 对于不同类别 T_j , 单词 w_i 计算得到 $tw_{w_i}^{T_j}$ 值不同. 在同一类别下的重要性值相近的单词, 指示类别 T_j 的能力也是相似的, 因此从这一角度来说单词在咨询类别下的重要性值也可以被视为单词的一种浅层语义抽象.

$tw_{w_i} \in R^{1 \times \text{numConsultType}}$ 为单词在所有咨询类别下的重要性向量(后文简称为单词的 tw 编码表示), 由单词在所有咨询类别上的重要性值组成, 即 $tw_{w_i} = [tw_{w_i}^{T_1}, tw_{w_i}^{T_2}, \dots, tw_{w_i}^{T_{\text{numConsultType}}}]$, 表示单词 w_i 指示不同咨询类别的能力. 向量长度为咨询类别的个数公式表示.

定义 2.2 句子在咨询类别下的度量 tw_S 、 consultVec_S

对于句子 $S = [w_1, w_2, \dots, w_{\text{numWords}_S}]$, 它在所有咨询类别下的重要性向量 $tw_S \in R^{1 \times \text{numWords}_S}$ 为句子中各单词在各咨询类别下的重要性向量的平均值, 计算公式为

$$tw_S = \frac{1}{\text{numWords}_S} \sum_{i=1}^{\text{numWords}_S} tw_{w_i} \quad (2)$$

向量长度为咨询类别的个数 numConsultType .

$\text{consultVec}_S \in R^{1 \times \text{numWords}_S}$ 为指定咨询类别下, 句子中各个单词在该目标咨询类别 T^* 下的重要性值所组成的向量, 即 $\text{consultVec}_S = [tw_{w_1}^{T^*}, tw_{w_2}^{T^*}, \dots, tw_{w_{\text{numWords}_S}}^{T^*}]$, 向量长度为句子中单词的个数 numWords_S . consultVec 向量中值越大的元素对应的单词在句子中就越重要, 因此当已知句子所属的目标咨询类别时, 通过 consultVec_S 可以指示出句子中的关键词. 这种方式计算得到的关键词, 与根据单词对上下文的 tf-idf 值来寻找关键词的方法不同. 它的计算仅与语料相关, 效果不受句子长短的影响, 同时在选择的过程中又能参考句子的咨询类别信息, 过滤掉在语料中比较重要但在具体句子中不需要关注的关键词.

在咨询类别判断模型中, 对于语料中的句子

S , 使用句子对应的重要性向量 tw_S 作为 SVM 或 LR 模型的输入特征, 可学习得到一个简单的咨询类别判断模型 f_{CT} , 并可以通过模型 f_{CT} 来预测句子的咨询类别. 在句对匹配模型建模过程中, 需要使用模型 f_{CT} 计算出问句对应的 consultVec 向量对相似度及相关度矩阵进行加权. 对于新句子 S_{new} , 它的目标咨询类别 $T_{\text{new}} = f_{CT}(S_{\text{new}})$, 根据定义 2.2, 句子在目标咨询类别下的度量 $\text{consultVec}_{S_{\text{new}}} = [tw_{w_1}^{T_{\text{new}}}, tw_{w_2}^{T_{\text{new}}}, \dots, tw_{w_{\text{numWords}_{S_{\text{new}}}}}^{T_{\text{new}}}]$.

2.3 句对匹配模型

设语料为 $X_L = \{((U_0, R_0), \text{Label}_0), ((U_1, R_1), \text{Label}_1), \dots\}$, 其中 (U_i, R_i) 代表一个句对, $\text{Label}_i \in \{0, 1\}$ 为对应的匹配度标签, $i = 0, 1, \dots$. 我们要学习一个句对匹配模型 f_L , 使得对于语料中的任意句对 (U_i, R_i) , 它对应的匹配标签 Label_i 可以通过 $f_L(U_i, R_i)$ 进行预测.

本文延续文献[19]中的思路, 通过单词层面及句子层面的相似度及相关度对句对的匹配度进行建模. 另外在前期增加一个咨询类别判断模型, 使用机器学习的方法获得句子所属的目标咨询类别, 由此指示出句子中的关键词. 通过对关键词加权, 使得模型在学习的过程中可以区分问句中单词的重要性, 更多地关注句子中关键词之间的相似度及相关度, 从而提高句对匹配模型的效果.

句对之间相似度的衡量方式有多种, 首先可通过 word2vec 将单词进行分布式向量表示. 这种表示方式具有良好的特性, 即相似单词的词向量具有较高的相似性, 同时还能表示出单词之间的类比性质. 另外, 我们通过计算单词在各咨询类别上的重要性, 对单词及句子按照定义 2.1 和 2.2 所述, 进行另一种分布式向量表示. 它可以将语义上不是那么相似的单词, 从咨询类别这个角度相似化.

对于句对 (U, R) , 设问句长度为 m , 答句长度为 n , $WV_U = [\text{wordVec}_{U,1}, \dots, \text{wordVec}_{U,m}]$ 及 $WV_R = [\text{wordVec}_{R,1}, \dots, \text{wordVec}_{R,n}]$ 分别代表问句 U 及答句 R 对应的 word2vec 表示序列. $TW_U = [tw_{U,1}, \dots, tw_{U,m}]$ 及 $TW_R = [tw_{R,1}, \dots, tw_{R,n}]$ 为问句 U 及答句 R 对应的 tw 表示序列. $H_U = [h_{U,1}, \dots, h_{U,m}]$ 及 $H_R = [h_{R,1}, \dots, h_{R,n}]$ 为问句 U 及答句 R 对应的隐状态序列. 本文由相似度这个角度出发, 计算出问答句在隐状态序列上的相似度矩阵 $\text{Sim_hidden} \in R^{m \times n}$ 、 word2vec 编码表示单词层面上的相似度矩阵 $\text{Sim_wv} \in R^{m \times n}$ 、 tw 编码表示下单词层面上的相似度矩阵 $\text{Sim_tf} \in R^{m \times n}$, 计算公

式如下:

$$\text{Sim_hidden}_{i,j} = h_{U,i} W h_{R,j}^T \quad (3)$$

$$\text{Sim_wv}_{i,j} = \text{wordVec}_{U,i} \cdot \text{wordVec}_{R,j}^T \quad (4)$$

$$\text{Sim_tf}_{i,j} = t w_{U,i} \cdot t w_{R,j}^T \quad (5)$$

在判断句对之间是否匹配的时候,除了通过相似性这个角度进行判别之外,还可以通过相关性这个角度进行判别.当问句与答句相似度较低但具有较高的相关度时,也应当认为是匹配的.本文在建模过程中引入句对之间的相关性矩阵,从词共现这个角度对句对中单词之间的相关性进行衡量.当两个单词同时在某个语境里出现时,就认为这两个单词具有某种关联,具体来说,当两个单词同时出现在语料中的某个句对中时,记为一次共现.单词之间共现的频率越高则相关度就越高.

设有单词 A 及单词 B , 则 A 、 B 之间的相关度值 $\text{CoApp}(A, B) = \frac{n(A, B)}{n(A)}$, 其中 $n(A, B)$ 表示语料中单词 B 在单词 A 之后出现的次数, $n(A)$ 表示语料中单词 A 出现的总次数.由相关度值的公式可知, $\text{CoApp}(A, B) \leq 1$. 问答句之间的相关度矩阵 Sim_Correlation 计算公式为

$$\text{Sim_Correlation}_{i,j} = \text{CoApp}(u_i, r_j) \quad (6)$$

考虑到直接将上述计算得到的相似性及相关性矩阵作为后续 CNN 的输入特征时,默认矩阵中各元素的权重是一样的,而在实际中应该更关注问句中的关键词与答句中单词之间的相似性及相关性,较少关注问句中非关键词与答句中单词之间的相似性,因此找出句子中的关键词,并通过关键词对相似度及相关度矩阵进行加权十分必要.

在单轮对话的句对匹配中,由于句子的长度较短,单词出现的频率普遍较低,不适合通过 tf-idf 方式计算出句子中的关键词.我们按照定义 2.2 计算出问句对应的 consultVec 向量,由该向量描述问句中所有单词的重要性值,其中向量值较大的元素对应的单词被定义为关键词.这种方式计算出的关键词与句子长度无关,而仅与使用的语料及句子所属咨询类别相关,因此在缺少上下文或上下文长度较短的情况下也可以找到句子中的关键词.

在计算出问句对应的 consultVec 向量之后,通过公式

$$\text{weighted_Sim_wv} = (\alpha * \text{consultVec} + 1) \cdot 1 \otimes \text{Sim_wv} \quad (7)$$

$$\text{weighted_Sim_tf} = (\beta * \text{consultVec} + 1) \cdot 1 \otimes \text{Sim_tf} \quad (8)$$

$\text{weighted_Correlation} =$

$$(\lambda * \text{consultVec} + 1) \cdot 1 \otimes \text{Sim_Correlation} \quad (9)$$

对原始的相似度矩阵及相关度矩阵进行加权.其中参数 $\alpha, \beta, \lambda \in R$, 用于控制重要性向量的作用,使模型可以从语料中自行学习关键词在句对匹配中的作用.

3 实验

为了验证 CTASM 模型在心理咨询对话匹配度建模上的有效性,实验通过 TensorFlow 搭建 CTASM 模型及若干对比模型,在约 10 万条心理咨询相关的对话数据上进行模型训练,比较各模型在不同评价指标下的表现.考虑到句对匹配与否本身为二分类问题,在进行模型评价时,选取常见的二分类评价指标如准确率、精准率、召回率、 $F1$ 值作为模型的评价指标.同时考虑到该模型在进行对话回复时,需要从若干备选回复中选取最佳回复句,选取文献[14]中使用的评价指标 $R_n@k$ 作为实验中模型的评价指标之一.

实验硬件环境为: Windows7, Intel(R) Core(TM) i5-4460 CPU @ 3.20GHz, 8GB 内存, 64 位操作系统.软件环境为: TensorFlow 1.4.0.下面介绍实验中数据集的获取,对比实验的选取,模型参数的设置及最终的结果.

3.1 数据及预处理

由于心理咨询的特殊性,心理咨询语料的获取非常困难.目前尚无相关的公开语料,心理咨询网站上的语料又比较少.可获取的较大的语料库多为开放领域内对话语料,可以在开放领域内对话语料中进行筛选,以获取足够量的心理咨询对话语料.直接通过人工打标签筛选出心理咨询相关语料不太现实,因此我们通过机器学习的方法,训练出一个可以自动对语料中的句子打标签的模型,通过该模型快速地从开放领域对话语料中筛选出大量的心理咨询相关语料.模型构建时首先在较小的语料上建立一个简单的模型,判断句子是否与心理咨询相关.再将其应用于一个较大的语料库里筛选出问句与心理咨询相关的句对,并标注其所属咨询类别.

考虑到校园场景,我们选取中国科学技术大学心理咨询网站及中国科学技术大学、四川音乐学院、武汉大学、郑州大学、河南大学等若干个活跃校园贴吧上的对话数据作为原始语料.对原始语料进行预处理操作,如去除停用词及无意义的符号,去掉长度

过短及过长的句子之后,人工标注其中约 1 300 条句子作为咨询类别判断模型的语料,在此之上学习得到可以进行咨询类别判断的模型 f_{CT} .对于语料中的句对,若其问句所属的咨询类别标签不是“无关”类,则保留该句对.

最终提取到约 10 万条与心理咨询相关的对话语料.这些句对的匹配度标签均设置为 1,表示句对之间是匹配的.接着对语料中的每个问句,随机从语料中选择一个句子作为其答句,匹配标签为 0,即句对之间是不匹配的.以此方式可以得到约 20 万条训练句对,其中标签为 0 的句对数与标签为 1 的句对数一致.

3.2 Baseline

实验中选取文献[14]模型(Ubuntu-LSTM)作为基础模型,该模型计算句子的向量表示,通过计算句子相似度进行匹配度判断,是一个比较基础且在多个论文中用作对比试验的模型.选取文献[19]中的模型(SMN)作为多角度进行相似度建模的对比模型,它是多角度对句对匹配度进行建模的模型中目前实验效果最好的.为了验证 CTASM 模型中关键词加权策略及从咨询类别角度进行单词相似度及相关度计算策略的有效性,增加一个中间模型 weighted-SMN,它在 SMN 模型的基础上使用了关键词加权策略,通过 SMN 模型与 weighted-SMN 模型的对关键词加权策略的有效性进行验证.同时 CTASM 模型与 weighted-SMN 模型的对关键词加权策略的有效性进行验证.

对比实验中用到的模型如下:

Ubuntu-LSTM:文献[14]中的模型,通过 LSTM 模型对问答句进行编码,对句子向量进行相似度计算,学习句对匹配模式,是句对匹配问题中常用的对比模型.

SMN:文献[19]中的模型,通过 LSTM 模型对问答句进行编码,利用编码过程中的隐状态序列,计算问答句在句子层面上的相似性矩阵,同时利用单词的 word2vec 编码计算问答句在单词层面上的相似性矩阵,从两个层面对相似性进行度量,学习句对的匹配模式,是目前多角度进行句对匹配度学习的模型中表现最佳的.

Weighted-SMN:在 SMN 模型的基础上增加关键词加权机制,通过关键词对句对间的相似度矩阵进行加权.

CTASM:本文提出的模型,在 weighted-SMN

模型的基础上,增加从咨询类别的角度出发计算得到的问答句的相似度及相关度矩阵,对句对的匹配模式进行学习.

3.3 参数设置

在咨询类别判断模型中,训练集、验证集、测试集的划分比例为 5 : 1 : 1,通过网格搜索及交叉验证选择模型的最佳参数.

在句对匹配模型中,随机从语料中选择约 1 000 条句对作为验证集,1 000 条句对作为测试集,剩余句对作为训练集.保证训练集、测试集、验证集之间没有重叠.考虑到语料数量及语料中句子的长度,在实验中我们设定词向量的维度为 100 维,LSTM 网络中隐层单元的维度为 50,卷积层及池化层窗口大小及特征图的个数通过网格搜索选定.卷积及池化的窗口从{(2,2),(3,3),(5,5)}中选取,特征图的个数从{4,8,10}中选取,最终选取的窗口大小为(3,3),特征图个数为 8.通过反向传播优化目标函数,Adam 算法对参数进行更新.训练时设置批大小(batch size)为 150,句子的最大长度设置为 15.句子长度超过 15 的做截断处理,不足 15 的用 0 填充.训练中使用早停(early stop)机制避免模型的过拟合.在模型训练过程中,若连续多次在验证集上的训练效果没有提升,则认为发生“过拟合”现象,停止训练并选择整个训练过程中在验证集上表现最佳的模型作为输出模型.

3.4 实验结果

对于咨询类别判断模型,它需要更准确地提取句子所属的咨询类别信息,同时作为句对匹配模型的一部分,它的分类效果将影响句对匹配的准确性.咨询类别判断是一个多分类问题,在实现时将多标签问题转化为多个二分类问题进行求解,对每个咨询类别建立一个二分类模型.对于单个二分类模型,评价时采用常见的评价指标准确率(accuracy)、精确率(precision)、召回率(recall)及 F1 值(F1-score),结果如表 1 所示.

表 1 咨询类别判断模型效果

Tab.1 The effect of consultation category judgment model

类别	准确率	精确率	召回率	F1 值
学习	0.94	0.70	0.93	0.81
恋爱	0.97	0.85	0.93	0.89
人际	0.92	0.75	0.98	0.85
择业	0.98	0.71	0.97	0.82
适应	0.93	0.76	0.98	0.86
异常	0.96	0.80	0.95	0.87
无关	0.91	0.81	0.86	0.84