

一种面向初高中英语单选题的类人答题系统

陈志刚¹, 刘青文², 林伟², 汪洋², 陈小平¹

(1. 中国科学技术大学 计算机科学与技术系, 安徽 合肥 230026; 2. 科大讯飞股份有限公司研究院, 安徽 合肥 230000)

摘要:初高中英语考试中单选题的机器类人答题问题(sentence completion problem),其目标是从一个词或短语集合中选择最佳答案填入目标句子的适当位置使其成为一个合理通顺的句子.尽管针对该问题学术界已经提出了许多解决方案,但是都停留在语言模型和隐语义分析的算法框架下.为此以语言模型和隐语义分析两种方法为基线,分别从固定搭配抽取、动词时态分析两个方面抽取试题的词法、语法和语义特征,以初高中英语题库中的部分试题为训练集,构建了一个基于学习排序模型的类人答题系统.该系统的在初中和高中的单选题测试集上分别达到78%和76%的准确率,与基于语言模型和隐语义分析的基线模型相比均提高了4个百分点.

关键词:类人答题;句子完形;语言模型;信息抽取;时态分析

中图分类号:TP391 **文献标识码:**A **doi:**10.3969/j.issn.0253-2778.2016.03.008

引用格式: CHEN Zhigang, LIU Qingwen, LIN Wei, et al. An exam robot for sentence completion in high school English tests[J]. Journal of University of Science and Technology of China, 2016,46(3):231-237.

陈志刚, 刘青文, 林伟, 等. 一种面向初高中英语单选题的类人答题系统[J]. 中国科学技术大学学报, 2016,46(3):231-237.

An exam robot for sentence completion in high school English tests

CHEN Zhigang¹, LIU Qingwen², LIN Wei², WANG Yang², CHEN XiaoPing¹

(1. Department of Computer Science, University of Science and Technology of China, Hefei 230026, China;
2. Iflytek Research, Hefei 230026, China)

Abstract: Addressed in this paper is the problem of sentence completion in Chinese national college or high school entrance English examinations in which the most appropriate word or phrase from a given set should be chosen to complete a sentence. Although a variety of methods have been developed to solve this problem in the literature, these approaches mainly focused on language modeling (LM) and latent semantic analysis (LSA) to the best of our knowledge. An exam robot prototype was built by extending the language modeling and latent semantic analysis methods to verb tense analysis and long distance phrase extraction. Specifically speaking, the syntactic, lexical and semantic features are extracted separately using by means of LM and LSA as well as verb tense analysis and phrase extraction two methods developed by the authors. These features are then fed into a learning to rank model to build the exam robot. The proposed approach outperforms LM and LSA models by 4.0 percentage points, achieving 78% accuracy on the question sets for senior entrance exams and 76% accuracy on the question sets for college entrance exams.

Key words: exam robot; sentence completion; language model; phrase extraction; tense analysis

收稿日期:2015-09-12; **修回日期:**2015-12-29

作者简介:陈志刚,男,1977年生,博士/高级工程师.研究方向:语义数据挖掘. E-mail: zgchen@iflytek.com

通讯作者:刘青文,博士/工程师. E-mail: qwliu@iflytek.com

0 引言

移动互联网和大数据的背景下,以大数据驱动的深度学习等技术的发展,为人工智能提供了新的解决思路.自然语言理解是人工智能技术应用于实际领域的典型范例,虽然经过多年的努力,这一领域已经取得大量令人瞩目的成果,但是仍然面临着语义鸿沟、知识表示以及演绎推理等重大挑战.近年来,随着在线教育的兴起,大量教育领域的数据被收集起来作为自然语言理解任务的研究对象^[1],其中一项最具吸引力的研究就是类人答题系统.本文以中国高考英语单选题为切入点,对类人答题系统的关键技术进行了初步的探索,并构建出一个针对初高中英语单选题的类人答题原型系统,本文的主要贡献主要有如下几个方面.

(I)提出了一种英语句子固定搭配抽取算法,该方法在句法分析结果的基础上,从海量的语料中抽取短语搭配作为自动答题模型(学习排序模型)的特征,在语言模型和隐语义分析的基线模型(准确率分别为初中 74%,高中 72%)的基础上,提高了 3% 的答题准确率.

(II)构建了英语动词时态预测模型,通过初高中教育专家的分析,我们总结了影响英语动词时态的词法和句法特征,并以这些特征为基础,将时态预测问题转化为一个多类分类问题.该分类模型覆盖 10 种常见英语时态,分类准确率达到 69%,使整体答题正确率提高了 1%.

(III)搭建了一种基于学习排序框架的类人答题原型系统,在该学习排序模型中,我们以语言模型和隐语义分析两种已有方法抽取的特征为基础,扩展了固定搭配特征和时态分析特征,使得答题的准确率在初中测试集上由 74% 提高到 78%,在高中测试集上由 72% 提高到 76%.

1 相关工作

句子完形任务由 Zweig 等^[1]首次提出,该文以福尔摩斯小说集为背景语料,利用语言模型筛选出 1040 个句子,对于每个句子去除其中的一个实词使其成为句子完形模型的测试对象,该数据集被命名为 MS_SCD.针对该任务,Zweig 等^[2]提出了语言模型和语义分析^[3]两种解决句子完形问题的方法.该任务与我们面临的中高考答题任务在形式上完全一

致,但是在内容上却存在较大差别,具体表现在:① MS_SCD 数据集中的题干只有一个句子,中高考试题可能存在多个句子,而且可能是以对话的形式出现,更为丰富的上下文虽然为答题提供了更多的可用信息,然而却带来了长距离语法依赖和语义依赖等难于解决的问题;② MS_SCD 数据集中的选项只含有一个单词,而试题选项可能包含多个单词,并且一个句子可能会存在多个缺失项;③ MS_SCD 数据集是经过 NGRAM 语言模型筛选的,这意味着集中的句子各选项对应的句子概率相差不大,NGRAM 语言模型在该任务上可能会失效,而我们则不存在该问题.Gubbins 等^[4]提出了一种融合句法分析结果的语言模型来解决句子完形问题,取得了比 RNN 语言模型^[5-6]更好的效果.

单词替换任务(lexical substitution task)是与句子完形任务十分相似的自然语言处理任务,它由 McCarthy 等^[7]提出,与句子完形任务的最大区别在于试题的选项用来替换题干中的某个特定的单词而不是用来填充某个缺失的单词从而补全句子,因此与句子完形任务相比该任务更简单,语言模型和隐语义分析以及本文提出的方法在该任务上同样适用.另外 Ritter 等^[8]提出的单词的句法参数合理性评估任务对本文中的答题任务也具有一定的参考价值,单词句法参数特征可以用来判断候选答案的合理性,本文中的句子搭配抽取和时态特征提取本质上也是在分析单词的句法参数的合理性.

2 自动答题系统整体架构

图 1 给出了自动答题系统的总体架构,逻辑上分为三层:

(I)数据资源层.该层主要负责语料的收集、格式化处理和存储.本文的第 3 节将详细介绍语料的规模、内容、格式及处理流程.

(II)特征提取层.该层负责输出上层模型训练所需要的特征,主要包括固定搭配抽取模块和动词时态特征抽取模块,特征抽取的细节将在第 4 节中作详细描述.

(III)模型层.该层包括五个子模型和一个融合模型,五个子模型分别从不同的维度给出各选项的特征,融合模型根据这些特征训练一个学习排序模型给出最终答题效果.各模型的细节将在第 4 节中详细描述.

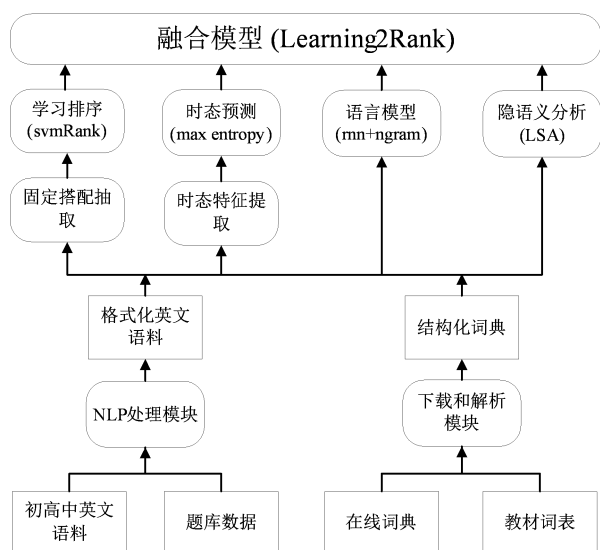


图 1 自动答题系统架构

Fig. 1 Automatic answering system

3 数据集介绍

3.1 初高中英语教学领域语料

因为答题针对中高考英语试题,所以我们收集了初高中教育领域的相关英文文本语料,共 7.6 GB,语料的统计数据如表 1 所示。

表 1 英文语料明细

Tab. 1 Details of English corpus

语料类型	规模 (MB)	句子数量
教材	472	9 372 680
英语学习网站	1 600	21 210 788
新闻播报	1 700	27 786 936
词典例句	50	4 351 138
初中题库	1.3	17 962
高中题库	4	43 748

为了满足上层特征抽取的需求,我们将表 1 中的文本语料进行了流程化处理,处理的流程如图 2 所示。

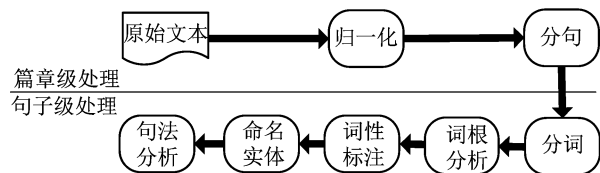


图 2 语料处理流程

Fig. 2 Corpus processing procedure

整个流程分为两个子模块:篇章级处理包含归一化和分句操作两个步骤,其中归一化操作指去除特殊字符、汉字以及归一化单词间隔,具体的处理流程如下:首先根据字符编码去除语料中杂质,仅保留

英文字符、中文字符、数字以及标点符号;然后将汉字替换为一个统一的字符串;最后对多种不同编码的空格符做归一化;分句采用基于统计的句子边缘检测模型. 具体而言,我们对英文句号是否为句子的边缘标注了较大规模的训练集,并采用神经网络训练一个句子边缘检测模型,准确率达到 95%. 句子级处理流程包含五个步骤,即分词、词性标注、词根分析、命名试题识别、句法分析^[9-10]. 处理后,我们一共得到 600M 个条目,以序列文件的格式存储在 HDFS 上。

3.2 初高中英语单选题库

答题目标为初高中英语单项选择题,试题的一般形式为

You were _____ to take your shoes off when you entered the lab, but you didn't.

- A. invited B. supposed
- C. selected D. Approved

我们一共收集了 25 660 道初中英语试题、62 498 道高中英语试题,并按照 14 : 3 : 3 的比例分别分为训练集、测试集和开发集。

不同的试题所考察的知识点不尽相同,所采用的答题算法也不尽相同. 为了准确分析自动答题模型在不同考点上的效果,我们按照试题的易错原因总结了 10 种类型,并为测试集中的每道试题人工标注了易错原因(一共标注了 4 313 道试题),图 3 展示了测试集中各类型试题的分布情况。

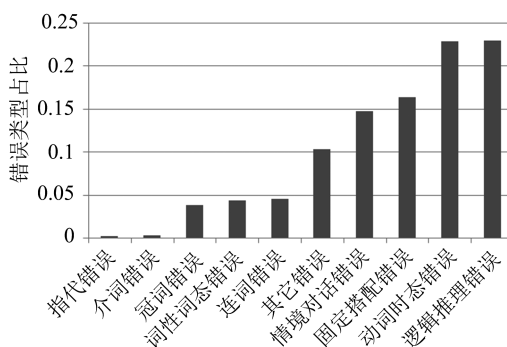


图 3 试题类型标注统计

Fig. 3 The test type annotation statistics

4 自动答题核心算法

本节将详细对答题系统中的各核心模块作较详细的阐述. 首先介绍已有工作中的两种答题思路,在我们的系统中,利用这两种方法所抽取的特征所能达到的答题效果,将作为系统的对比基线;然后分别

阐述了本文提出的固定搭配抽取和时态预测算法,这两种方法所提取的特征将与基线方法所提取的特征一起作为学习排序答题模型的特征。

4.1 基于语言模型和隐语义分析的自动答题模型

目前学术界关于这类问题的解决方法主要有两个,第一个是基于语言模型的方法^[2,6],第二个是基于隐语义分析的方法^[2]。基于语言模型的答题模型的基本思想是利用海量语料训练语言模型,将候选答案代入句子中,通过语言模型计算句子的概率值,选择概率最大的作为正确答案;基于隐语义分析的答题模型的基本思想是:通过计算选项中的单词与题干中单词在语义隐语义空间的语义距离(通过对句子-单词矩阵做奇异值分解来得到词的语义向量),来确定其与题干的相关程度,最终选择语义距离最近的选项作为正确答案。

在收集的英文语料上,我们分别训练了 NGRAM 语言模型和 RNN 语言模型,并在 RNN 模型中融入了词性序列的信息。同时我们在大规模的语料上进行隐语义分析(SVD),得到初高中词表中所有单词的语义向量。

4.2 基于固定搭配抽取的自动答题模型

从图 3 中我们可以看到,固定搭配类试题在初高中试题集中占有较大的比例,语言模型能够解决大部分短距离搭配类问题,但是解决长距离搭配类问题的能力有限(即使是目前较先进的 RNN 语言模型的长距表达能力也比较有限),因此我们提出了一种基于句法分析结果的固定搭配抽取算法来获取单词之间的搭配关系。

图 4 展示了固定搭配抽取算法的基本思想,我们通过句法分析产生句子的语法树,并在此基础上抽取短语级别的搭配;进一步,我们通过在句法树的基础上对句子做依存分析,在依存分析的结果上得到词级别的固定搭配。图中的圆角矩形中的内容是词性标注的结果,词性标注规范参考斯坦福词性标注器^[10],弧线上为依存分析的结果,依存关系的规范参考斯坦福依存关系分析器^[10]。

本文采用该算法在海量的教育领域文本上抽取单词的固定搭配模式,并将其作为上层学习排序模型的基本特征。

4.3 时态预测模型

从图 3 中,我们看到动词时态是初高中英语单选题考察的重点,而上文中提到的语言模型、隐语义分析和固定搭配抽取的方法均无法解决时态类问题,因此我们针对这类问题提出了一个时态预测

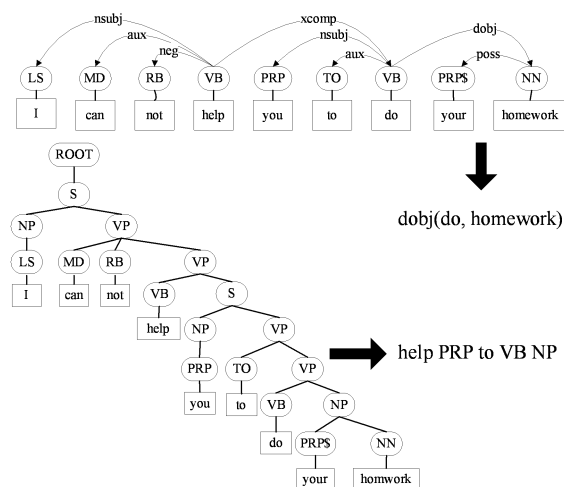


图 4 固定搭配特征抽取

Fig. 4 Fixed collocation feature extraction

模型。

经过分析,影响动词时态的因素主要分为两个方面:第一方面是动词所处位置的上下文的词序列和词性序列;第二方面是句子前后文的时态以及时间状语等一系列语法特征,我们将影响动词的语法特征总结如下。

(I) 时间状语. 句子中的时间状语通常用来表达动作发生的时间,对时态的判定具有决定性的作用,如时间状语是“tomorrow”,则动词时态一般为将来时。

(II) 修饰副词. 有一部分副词能够反应动词的时态,如“often”后常接动词一般现在时。

(III) 介词短语. 介词短语能够表明动作的发生时间、地点和目的等,对时态的判定也具有一定的参考价值,如“in the next three years”。

(IV) 连词从句. 由连词连接的两个从句之间的时态具有明显的约束关系,如“and”或“or”连接的两个从句时态要保持一致。

(V) 状语从句. 状语从句和副词的功能相当,从句和主句的时态具有约束关系,如由“when”引导的状语从句,如果主句是一般过去时,则从句需要用一般过去时或者过去进行时。

(VI) 前后文句子时态. 前后文句子的时态对当前句子的时态有影响。

根据上述分析,我们将分三步来抽取动词时态相关的特征:第一步,按照图 2 的处理流程得到词性标注和句法分析结果;第二步,结合词性标注的结果,根据一系列固定规则识别动词的时态(时态识别的准确率能够达到 95%),生成训练和测试数据;第

三步,抽取动词前后文词性序列以及上述的若干种语法特征.下面我们用一个例子对第三步进行详细的说明.

给定如下两个英文句子,要求抽取动词“got”时态特征.首先我们识别出“got”对应的时态标签为一般过去时,然后在句法分析结果的基础上抽取如表2所示的语法特征.

Did you visit your parents yesterday?

Yes, when I got there, they were working in the fields.

表2 时态特征
Tab.2 Tense characteristics

特征名	特征值
词性特征	pos1:UH pos2:, pos3:CC pos4: WBR pos5:PRP pos6:RB pos7:, pos8:PRP pos9:VBD pos10:VBG
前文时间状语	pre_tmod:yesterday
后文时间状语	post_tmod:null
前文时态	pre_tense:[SPA](一般过去时)
后文时态	post_tense:null
时间状语	tmod:null
修饰副词	adv_mod:when
介词短语	prep_mod:null
连词从句	conj_porc:null conj_mark:null conj _tense:null
状语从句	advcl_porc:clause advcl_mark:when advcl_tense:[PAPR](过去进行时)

在表2中,状语从句相关的特征中 advcl_proc 表示目标动词是位于主句还是从句中,如果位于主句中则 advcl_proc 的值为 primer, 否则为 clause, advcl_mark 表示状语从句的引导词, adv_tense 表示除了目标子句之外另一个句子的时态.

由于时间状语、副词、从句引导词的种类繁多,且有相当大一部分与时态没有强相关性,因此我们在训练集中统计了各特征词与时态之间的相关性,并根据相关性的强弱做特征筛选以缓解特征稀疏的问题.得到上述特征后,采用最大熵模型来预测动词的时态.

4.4 基于学习排序的融合模型

为了提高整体答题正确率,我们在语言模型和隐语义分析所产生的特征的基础上,扩展了固定搭配和时态分析的特征,以一部分试题为训练集,训练学习排序模型(pairwise svmRank),并在标注了试题类型的测试集上测试答题效果.

下面我们简要叙述一下该模型的构建过程:对于任意一个选项,我们把该选项回代到题干中得到一个完整的句子,然后通过下面四个步骤得到该选项所对应的特征:

(I)通过语言模型解码得到句子的概率.

(II)通过隐语义分析模型,计算选项中的单词与题干中单词的平均语义距离.

(III)通过固定搭配抽取模型得到句中所有可能的搭配,并在预先抽取的固定搭配库中查找该搭配出现的频率,最终以抽取出来的固定搭配数量、长度以及频率作为分类器的特征.

(IV)如果选项全为动词,则通过时态测模型预测得到每个选项所对应的时态出现的概率,并以该概率作为分类器的特征;否则该维特征值统一置为一个相同的值(这样做的目的是为了让非时态类考点的试题,不受该维特征的影响).

特征抽取完成后,我们以试题的正确选项对应的句子作为正例样本,错误选项对应的句子作为错误样本训练学习排序模型.

5 实验结果与分析

5.1 语言模型

表3展示了基于语言模型的自动答题模型的正确率,其中“RNN+NGRAM”表示RNN语言模型的解码概率与NGRAM语言模型的解码概率的线性差值结果,“RNN_POS”表示在RNN语言模型中加入词性单元.从表3可以看到,通过线性插值的方式融合两种语言模型的概率能够显著提高答题正确率,并且在RNN语言模型中加入词性信息也能在一定程度上提升答题效果.由于RNN语言模型能够建模词之间或者词性之间的长距离依赖关系,因此两种语言模型的融合能够有效提高答题效果.

由于试题集合是领域受限的(初高中英语教学),所以在训练语言模型时的语料选择也会对答题效果造成较大的影响.我们尝试使用全部语料和只使用部分语料(题库数据、课文以及学习网站语料)分别训练语言模型.在初中测试集上,后者较好,而在高中测试集上前者较好,这是因为初中阶段的词汇较少且句式也较高中要匮乏许多,与新闻语料的匹配程度差.为了准确分析各部分语料对答题任务的贡献,我们分别在初高中的开发集上测试了不同语料的契合程度,如表4所示.从表4可以看到,高中英语试题和新闻语料的匹配程度显著大于初中试题与新闻语料的匹配程度.

表 3 语言模型的测试效果

Tab. 3 Test effects of a language model

算法	初中		高中	
	测试集	开发集	测试集	开发集
—				
NGRAM	0.71	0.70	0.70	0.69
RNN	0.70	0.70	0.63	0.61
RNN+NGRAM	0.73	0.74	0.72	0.70
RNN_POS+NGRAM	0.74	0.75	0.72	0.70

表 4 语料的匹配程度

Tab. 4 Corpus matching degree

语料集	初中试题匹配度(%)	高中试题匹配度(%)
新闻	1.7	7.7
课文	4.9	4.9
学习网站	28.1	32.0
题库	64.8	55.0

5.2 固定搭配的抽取

从 600M 个句子中,我们一共抽取出了 10M 个搭配,给定一个句子,我们以从中抽取出的搭配个数、每个搭配的长度和频率作为特征,以试题的正确选项对应的句子为正样本,错误答案对应的句子为负样本,训练一个学习排序模型.由于固定搭配的数量巨大,为了特征的缓解稀疏性,我们根据搭配的频率对这些特征进行了裁剪.训练数据中,每道试题有三个错误选项和一个正确选项,因此我们的正负样本比例为 1:3,由于正例偏少,我们在训练时尝试了不同的样本权重,在初中测试集上实验结果如表 5 所示.

表 5 基于固定搭配的答题模型效果

Tab. 5 Effect of answer model based on sentence collocation

搭配频率阈值	正样本权重		
	1.0	2.0	3.0
10	0.55	0.59	0.59
100	0.57	0.60	0.59
1 000	0.55	0.58	0.57
10 000	0.47	0.49	0.48

从表 5 可以看到,正样本的权重设置为 2.0,且只取频率高于 100 的固定搭配作特征时,答题效果最好,但是远低于 5.1 节中的语言模型的答题正确率,这是因为这种方法的目标是解决那些选项与题干中的词能构成长距离搭配关系的问题,而对于短距离搭配和其他类型的试题其效果则不如语言模型.

5.3 时态预测模型

我们在 12 045 个句子上测试了动词时态预测

的效果,整体的正确率为 69%.表 6 展示了各种时态的精度和召回率以及 F 值.

表 6 时态预测算法的效果

Tab. 6 The effect of tense prediction algorithm

时态名	标注名	精度	召回率	F 值
一般现在时	SPR	0.77	0.78	0.78
一般过去时	SPA	0.47	0.57	0.52
一般将来时	SF	0.50	0.26	0.34
现在进行时	PRPR	0.35	0.25	0.29
过去进行时	PAPR	0.61	0.33	0.43
现在完成时	PD	0.53	0.23	0.32
非谓不定式	INF	0.60	0.43	0.50
非谓过去分词	NFN	0.49	0.61	0.54
非谓现在分词	NFV	0.76	0.82	0.79

从表 6 可以看到,一般现在时与非谓语动词的现在分词的预测效果最好,其他时态的预测效果都相对较差,主要是因为一般现在时在训练样本中的比例最大,而对于非谓语动词的现在分词,一般句子中都存在强的句法特征.

5.4 融合模型

以语言模型的效果为基线,我们在融合模型中逐步添加隐语义分析、固定搭配以及时态分析的特征,其效果变化如表 7 所示:

表 7 融合模型的效果

Tab. 7 The effects of fusion model

特征	高中正确率		初中正确率	
	绝对值	增量	绝对值	增量
基线(LM)	0.721	0.0	0.739	0.0
+隐语义	0.723	+0.002	0.737	-0.002
+固定搭配	0.752	+0.029	0.770	+0.033
+时态特征	0.762	+0.010	0.781	+0.011

从表 7 可以看到,加入隐语义分析特征对答题效果的提升并不明显,在初中测试集上效果反而有所下降,这是因为初高中试题中考察语义辨析类的试题所占比例不高,且该方法在许多情况下会与语言模型的结果相冲突.由于中高考试题中考察长距离搭配的试题比例较高,语言模型擅长解决短距离依赖问题而对长距离搭配类问题失效;在海量领域语料上抽取的固定搭配能够显式地覆盖绝大部分初高中英语中的常用搭配,因此加入固定搭配特征后,我们看到在初中测试集上答题效果提高了 3.3%,在高中测试集上效果提高了 2.9%.时态特征的加入也在一定程度上提高了答题的正确率.

下面我们对一些典型的例子作简要分析:

例 5.1 Martin Luther King Jr. was _____ the Noble Peace Prize of 1964 for advocating nonviolence policy in the movement for civil rights.

- A. awarded B. offered
C. rewarded D. received

对于该题,基于语言模型的方法选择答案为“received”,这是因为“received the noble peace prize”在训练预料中出现的频率较高,而基于固定搭配的方法能够选择正确答案 A,因为我们抽出了“PRP BE awarded < * > prize”这样的搭配.

例 5.2 Did you visit your parents yesterday?

Yes, when I _____ there, they were working in the fields.

- A. got B. get C. getting D. have got

对于该题,基于语言模型的方法选择的答案为“get”,这是因为“get there”在语料中的出现频率高于“got there”,而加入时态特征后,答题模型能选出正确答案 got,因为答题模型能够通过前一句的时态、时间状语 yesterday 以及状语从句的时态特征判断出正确时态应该为一般过去时.

例 5.3 What's one fourth and a half, do you know?

Yes, It is _____.

- A. two sixths B. three fourths
C. one three D. three sixths.

上题考查的是演绎推理及语义理解,语言模型以及本文中提出的方法均无法解决.

6 结论

本文针对初高中英语单选题的自动答题任务,构建了一个类人答题原型系统,该系统通过融合语言模型、隐语义分析、固定搭配抽取以及动词时态分析等特征,在初中单选题测试集上的正确率达到 0.78,在高中测试集上的正确率达到 0.76.

经过我们对错误进行分析后,发现时态类错误和逻辑推理类错误分别占错误总数的 18%和 30%,后续的研究工作可以具体针对这两类问题展开.动词的时态预测问题不仅对于答题任务本身很重要,而且在英文的语法自动纠错、智能作文评分等领域也有重要的价值.解决逻辑推理类问题是人工智能技术面临的重大难题,后续可以从语义深度理解、知

识库构建和演绎推理等方面寻求解决方案.

参考文献(References)

- [1] ZWEIG G, BURGESS C J C. The Microsoft research sentence completion challenge[R]. Technical Report MSR-TR-2011-129, Microsoft, 2011.
- [2] ZWEIG G, PLATT J C, MEEK C, et al. Computational approaches to sentence completion [C]// Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics. Jeju Island, Korea: Association for Computational Linguistics, 2012, 1: 601-610.
- [3] LANDAUER T, DUMAIS S T. Latent semantic analysis[J]. Annual Review of Information Science and Technology, 2004, 38(1): 188-230.
- [4] GUBBINS J, VLACHOS A. Dependency language models for sentence completion[C]// Proceedings of the Conference on Empirical Methods in Natural Language Processing. Seattle, USA: ACL Press, 2013: 1405-1410.
- [5] MIKOLOV T, KARAFIÁT M, BURGET L. Recurrent neural network based language model[C]// 11th Annual Conference of the International Speech Communication Association. Makuhari, Japan: ISCA Press, 2010: 1045-1048.
- [6] MIKOLOV T. Statistical language models based on neural networks [R]. Presentation at Google, Mountain View, 2nd April 2012.
- [7] MCCARTHY D, NAVIGLI R. The English lexical substitution task [J]. Language Resources and Evaluation, 2009, 43(2): 139-159.
- [8] RITTER A, Mausam, Etzioni O. "A latent dirichlet allocation method for selectional preferences [C]// Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics. Stroudsburg, USA: ACL Press, 2010: 424-434.
- [9] DE MARNEFFE M C, MACCARTNEY B, MANNING C D. Generating typed dependency parses from phrase structure parses[C]// Proceedings of the Language Resources Evaluation Conference. Stanford, USA: IEEE Press, 2006, 6: 449-454.
- [10] MANNING C D, SURDEANU M, BAUER J, et al. The Stanford CoreNLP natural language processing toolkit[C]// Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics. Baltimore, USA: ACL Press, 2014: 55-60.