

文章编号: 2095-2163(2019)02-0262-07

中图分类号: TP399

文献标志码: A

面向问答系统的信息检索自动评价方法

张越, 杨沐昀, 郑德权, 赵铁军, 李生

(哈尔滨工业大学 计算机科学与技术学院, 哈尔滨 150001)

摘要: 对问答系统中的信息检索模块进行自动评价是开发问答系统中不可或缺的一环。采用传统的检索评价指标,就需要为每个问题标注正确的候选答案。为了避免这种代价,本文提出一种问答系统中的信息检索模块进行自动评价方法。该方法使用候选文档与问题本身以及问题参考答案间的信息,利用机器学习方法去拟合 *MAP*。在实验中,本文发现使用 *GBDT* 模型拟合 *MAP* 值最好,斯皮尔曼等级相关系数达到了 0.87。

关键词: 问答系统; 信息检索; 自动评价; *GBDT*

Derive *MAP*-like metrics from content overlap for IR in QA system

ZHANG Yue, YANG Muyun, ZHENG Dequan, ZHAO Tiejun, LI Sheng

(School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001, China)

[Abstract] During the development of a Question Answering System, implementing an automatic evaluation on the information retrieval module is an inevitable task. Using traditional information retrieval evaluation metrics requires labeling the correct candidate documents for each question. To avoid such cost, this paper proposes an automatic evaluation method for the information retrieval module in a Question Answering System. This method uses information between candidate documents and the question accompanied with the answer. This method utilizes machine learning technique to fit the *MAP* value. The experiment reviews that using *GBDT* model to fit the *MAP* value achieves the best performance and the corresponding Spearman rank correlation coefficient reaches 0.87.

[Key words] question answering system; information retrieval; automatic evaluation; *GBDT*

0 引言

问答系统是当前自然语言处理领域的研究热点之一。而且问答系统已被视作是搜索引擎的下一步发展方向。另据研究可知,百度、雅虎、谷歌等知名的互联网公司也都在不断地向这一领域持续增加研究投入。

问答系统的目标是让计算机使用自然语言来回答人的问题。总地来说,问答系统可以分为 3 个主要模块^[1-2],可对其阐释如下。

(1) 问题分析模块。分析问题得到句子词法、语法、语义成分信息,根据这些信息确定问题类别和潜在答案类型。

(2) 信息检索模块。在数据集中根据问题分析结果过滤出可能包含答案的数据,目的是缩小数据范围,减少下一阶段工作量。

(3) 答案抽取模块。从可能存在答案的数据中

抽取候选答案,对候选答案进行处理,最后组合候选答案为最终答案。

在问答系统的开发过程中,通过比较不同检索方法的优劣,从而达到不断地优化信息检索模块的目的,在此基础上致力于研发更好的问答系统。但是对问答系统中的信息检索模块进行评价仍面临着诸多不便。对此可概略分析如下。

信息检索模块的评价可以置于整个问答系统的评价中。即通过考察使用不同信息检索模块的系统回答用户提问的正确率,来评估信息检索模块的性能。但是,这种策略只有在后续的答案抽取模块设计完成后才能派上用场,而且只有高可靠的答案抽取模块,才能相对公平地反映检索模块的性能。显然,该策略在系统开发初期难以使用,而且会给模块化开发管理带来一定困难。此外,还可以对信息检索模块进行单独的评价,这样做的好处是无需等待整个系统开发流程结束就能开展信息检索模块的性

基金项目: 国家自然科学基金(61402134); 国家 863 计划项目(2015AA015405)。

作者简介: 张越(1994-),男,硕士研究生,主要研究方向:信息检索; 杨沐昀(1971-),男,博士,副教授,主要研究方向:信息检索、机器翻译; 郑德权(1968-),男,博士后,副教授,硕士生导师,主要研究方向:数据挖掘、信息抽取、应用人工智能; 赵铁军(1962-),男,博士,教授,博士生导师,主要研究方向:自然语言处理、人工智能应用、机器翻译; 李生(1943-),男,博士,教授,博士生导师,主要研究方向:自然语言处理、机器翻译、信息检索等。

收稿日期: 2018-06-19

能评估。

TREC 在组织历届信息检索评测时,开发了一套对信息检索系统进行评价的方法,比如最常见的 $MAP^{[3]}$ 指标。待评测的信息检索系统根据查询返回一个对文档的排序列表,人工评价者将逐一审查列表中的文档,根据文档与查询间的相关程度给文档赋予最终的二元相关度等级,即相关和不相关两档。之后使用 MAP 的定义公式求得该列表的 MAP 值。对于搜索引擎来说,这种答案设置可以用户的点击来代替,从而方便地完成搜索引擎的评价指标计算^[4]。当然,对于某些存在多种检索模块的情形,例如在 TREC、NTCIR 等评测中,还可以使用文献^[5]提出的利用不同检索系统的结果彼此互评的方法。

综上所述这些方法并不完全适合对问答系统中的信息检索模块进行评价。遵循 TREC 的做法,使用 MAP 、 $NDCG$ 等指标进行评价却要对返回的每个文档均需经历一番人工审查,耗时耗力。文献^[4]中的方法也不具可行性,因为问答系统在还未上线前缺乏点击日志。而文献^[5]则需要获得参加评测的其它检索系统的结果,无法仅就单个系统进行评价。

为了实现问答系统中信息检索模块的自动评价,且能够达到类似于 MAP 的效果,同时能够避免之前方法的种种缺陷,本文提出针对一种问答系统中的信息检索模块进行自动评价方法。该方法使用候选文档与问题本身以及问题参考答案间的信息,利用机器学习方法去拟合 MAP 。在实验中,研究发现使用 GBDT 模型拟合 MAP 值最好,斯皮尔曼等级相关系数达到了 0.87。

论文中,首先讨论用于拟合 MAP 值的多种机器学习模型和使用的特征。然后,则设计论述了实验准备、实验内容,并运行得到了实验结果。最后,就是总结全文,同时给出未来研究展望。

1 基于机器学习的信息检索评价指标建模

本文选择使用机器学习方法来建模拟合过程,而非人工构建拟合公式。

该方法的目标是在机器学习大框架下,寻找一个好的自动评价函数。也就是 $y = f(x)$ 。其中, x 是排序列表中的所有文档以及排序列表对应问题答案, y 是预测的 MAP 。研究中,期望 y 与真实的 MAP 尽量一致。

在机器学习建模过程中,将需要对纯文本信息

x 进行特征抽取,抽取出一个列表中所有文档的相关性特征,记做 $features$ 。而且,模型也不可能完美地预测 MAP 值,为了寻找最好的模型,需要定义损失函数 $loss$, 并最小化损失函数 $loss$ 。综上,本文将寻找一个好的机器学习模型 g_{best} 满足如下数学定义:

$$g_{best} = \min_g \text{loss}(g(features), MAP), \quad (1)$$

围绕着如何寻找 g_{best} 这一主题,本文拟从回归和排序 2 个角度构建研究体系,并深入剖析模型特征,相应内容可探讨分析如下。

1.1 回归策略

寻找能够最小化预测值和真实值间损失函数的 f , 这种任务定义形式天然地属于回归问题。

为了解决这个回归问题,本文首先尝试了普通最小平方回归,又称 OLSR, 来拟合 MAP 值,即求解以下优化问题:

$$\min_w \sum_{i=1}^N \|(w^t x_i - y_i)\|^2, \quad (2)$$

其中, w 是回归参数, x 是第 i 个排序列表 R_i 中所有文档的相关性指标, y_i 是第 i 个排序列表 R_i 对应的 MAP 值。

接下来,本文又考虑了岭回归和 Lasso^[6] 2 种方法。岭回归和 Lasso 对 OLSR 加以改进,分别对回归参数进行了 L_1 和 L_2 正则化。其中,岭回归将对应以下优化问题:

$$\min_w \sum_{i=1}^N \|(w^t x_i - y_i)\|^2 + C \|w\|^2, \quad (3)$$

而 Lasso 则对应以下优化问题:

$$\min_w \sum_{i=1}^N \|(w^t x_i - y_i)\|^2 + C |w|, \quad (4)$$

其中, C 是正则化系数。

在此基础上,本文还着重探究了 SVM^[7]、决策树、随机森林^[8]、GBDT^[9] 等 4 种更复杂的非线性模型。

在 GBDT 中,本文使用平方误差函数作为损失函数。研究推得数学公式如下:

$$\min_w \sum_{i=1}^N (y_i - f(x_i))^2, \quad (5)$$

$$f(x_i) = w_0 + \sum_{m=1}^M w_m \phi_m(x_i), \quad (6)$$

其中, w_m 是权重, ϕ_m 是弱决策树模型。

在弱决策树模型中,本文采用了最小平方误差准则来选择待分裂特征和最佳分裂点,对应以下优化问题:

$$\min S = \sum_{c \in \text{leaf}(T)} \sum_{i \in c} (y_i - m_c)^2, \\ m_c = \frac{1}{N_c} \sum_{i \in c} y_i. \quad (7)$$

其中, i 表示节点 c 区域内的一个样本; N_c 表示节点 c 区域内样本数量; m_c 表示对于节点 c 区域内样本的预测值。

1.2 排序策略

针对何种自动评价函数才可称作一个好的自动评价函数这个问题, 本文将从数值逼近和排序结果逼近 2 个角度进行释义解读, 具体研究如下。

对于一个自动评价方法, 若其对一个信息检索系统的评价结果在数值上越接近真实的 MAP 值, 这个自动评价方法的性能也就越好。

但数值上的偏离并不意味着这种自动评价方法没有价值。当研究者进行自动评价时, 其目的常常是获得不同系统间性能的高低关系而并非仅是着眼于某些特定的性能指标。如果使用该自动评价方法对不同系统的排名与人工评价方法对不同系统的排名已然相近, 也就是在排序效果上相近, 那该方法即可判定为是优秀的算法。

回归方法就是在数值上逼近人工评价结果, 从上述的分析可知在最终排序上逼近人工评价结果也是一种可行性选择。与回归函数相似的是, 排序函数的输入是一个列表中所有文档的相关性特征。排序输出则是对排序列表的 MAP 值的排序等价的预测值。

学习排序函数的过程实际上是一个 Learning to Rank 的过程。Learning to rank 算法是机器学习研究的研究核心与关键, 有着丰富的研究成果。Learning to Rank 分为 3 类, 分别是: Pointwise、Pairwise 和 Listwise。本文选用了属于 Pairwise Learning to Rank 算法的 Rank SVM^[10] 作为学习算法。

1.3 模型特征

在问答系统中, 问题本身和问题答案是人工评价者判别文档相关性的依据。一个好的文档既要覆盖问题的答案, 又要将问题的答案按照问题本身的要求组织起来, 2 个部分缺一不可。

本文使用 12 个特征来衡量文档与问题本身和问题答案间的相关度。这 12 种特征来自 4 种粒度和 3 种度量的两两组合。为了消除不同分词粒度的影响并获取不同粒度的语义表示, 本文在 4 种粒度下分别计算 3 种度量。这 4 种粒度包括有: 一元、二

元、三元和词。3 种度量可表述为: 精确率、召回率和置信度。也就是说, 单个文档将对应 12 个相关性指标。文中, 关于 3 种度量的研究计算可详见如下。

(1) 精确率和召回率。从集合论的角度, 文档与问题答案可以视作 2 个集合。在集合论中, 如果要衡量 2 个集合间相似程度, 通常做法是取 2 个集合的交集。为了将交集的大小规整到 0~1 之间, 可以除以任意一个集合的大小。在机器学习的定义中, 此番推导的结果实际上是在计算文档与问题答案间的精确率和召回率。于是, 本文将选取文档与问题答案间的精确率和召回率作为预测文本与问题答案间匹配程度的特征。

假设在某种粒度下排序列表中第 i 个文档可表示为 Doc_i , 排序列表对应答案可表示为 Ans 。那么该粒度下 Doc_i 的精确率和召回率的数学定义表述可见如下:

$$\text{精确率} = \frac{|Doc_i \cap Ans|}{|Doc_i|}, \quad (8)$$

$$\text{召回率} = \frac{|Doc_i \cap Ans|}{|Ans|}. \quad (9)$$

其中, $|Doc_i|$ 、 $|Ans|$ 和 $|Doc_i \cap Ans|$ 分别为文档 Doc_i 的长度、答案 Ans 的长度和 Doc_i 与 Ans 的交集的长度。

(2) 置信度。因为信息检索系统的查询来自于问题本身, 就使得系统赋予文档的置信度评分代表着问题本身与文档间的相关性。本文使用语言模型作为系统的检索模型来为问题打分。语言模型的基本思想是从候选文档中估计出一个语言模型, 而后计算该语言模型下生成查询的概率, 最后利用此概率来为文档设定排序。在实际求解时, 采用概率的对数形式以简化计算。

假设在某种粒度下排序列表 R 对应的查询可表示为 $Q = \{q_1, q_2, \dots, q_n\}$, 其中 q_j 是 Q 的第 j 个查询词。排序列表中, 第 i 个文档可表示为 Doc_i 。用 tf_{q_j} 表示查询词 q_j 在文档 Doc_i 中的出现次数, cf_{q_j} 表示查询词 q_j 在整个文档集中的出现次数, $|Doc_i|$ 表示该粒度下文档 Doc_i 中词的总数, $|C|$ 表示该粒度下文档集中词的总数。参照文献[11-12], 研究推得 Doc_i 在该粒度下置信度的数学运算可表示如下:

$$\text{置信度} = \log \prod_{j=1}^n P(tf_{q_j} | Doc_i), \\ P(tf_{q_j} | Doc_i) = \frac{tf_{q_j} + \mu \frac{cf_{q_j}}{|C|}}{|Doc_i| + \mu}. \quad (10)$$

基于前述分析可知,本文使用的回归函数和排序函数的输入都是一个列表中所有文档的相关性特征。所以本文首先计算每个文档与问题本身和问题答案间的相关性指标,再将每个文档的相关性指标按照其在排序列表中的位置依次拼接在一起构建成完整的输入数据。

在计算诸如 *MAP* 和 *NDCG* 之类的人工评价指标时,排名靠前的文档对指标有着更大的影响。故而即使仅采用了由排名靠前的文档的相关性特征组成的部分输入数据,模型也将获得不错的性能。同时,每个模型自身的复杂度和泛化能力都不相同,这就导致了每个模型最适合处理的输入长度也将各有不同。

研究至此可知,在后续的仿真实验中,每个模型使用多少文档成为该模型可以调节的参数之一,记为 N 。称某个模型使用 N 个文档,意味着该模型使用由排序列表前 N 个文档的相关性特征组成的部分输入数据。

2 实验

2.1 数据来源

本文研究使用的检索系统是面向高考历史简答题的问答系统的一部分。检索系统面对的检索范围是 9 本人教版高中历史教科书,包括 3 本必修教材和 6 本选修教材。在检索和自动评价过程中使用的问答对来自 2013~2015 年北京高考历史简答题真题,共 15 题。

录入的教材按照每个自然段对应文档集中的一个文档的规律进行切分,将自然段所属的课文标题作为文档标题。平均段落长度为 127 词。实验中,研究文档的一个实例见表 1。

表 1 文档例子

Tab. 1 An example of the document

标题	内容
从禅让到王位世袭	约公元前 2070 年,禹建立中国历史上第一个王朝——夏。中国出现早期国家政治制度。禹年老时沿用禅让的惯例,选举继位人。但是禹死后,他的儿子启夺得王位,并传位给自己的后代。这样,政治权力由“传贤”变成“传子”,王位在一家一姓中传承,“家天下”的局面逐渐形成。原始社会后期的禅让制被王位世袭制所取代。商朝建立后,王位有时父子相传,有时兄终弟及。

针对高考真题,对每道题通过人工在历史教材中标出了相关文档,再根据题目人工拟定合适的查

询。因此每道题由问题、答案、查询和相关文档 4 部分组成。研究中也随即给出了高考题的另一实例可参见表 2。

表 2 高考题例子

Tab. 2 An example of college entrance examination

问题	答案	查询	相关文档
15.世界博览会通过展示科学技术成果,预示世界发展方向。1962 年世博会在美国西雅图举行。1957 年,苏联发射了人类历史上第一颗人造地球卫星。1961 年,苏联宇航员加加林成为第一个遨游太空的人,美国随即提出阿波罗登月计划。1962 年,在西雅图举办了主题为“太空时代的人类生活”的博览会。	政治背景:冷战时期美苏争霸;太空军备竞赛。我国航天成就:人造卫星;长征系列火箭;神舟系列载人飞船;探月工程;宇宙空间站。	美国 苏联 太空时代 政治背景 航天成就 中国	1953 年 3 月,斯大林逝世,美苏双方的关系开始出现微妙变化。美国总统艾森豪威尔主张通过对话谈判,逐步“软化”苏联。1961 年入主白宫的约翰·肯尼迪采取的对苏策略是“一手抓箭,一手抓橄榄枝”的“和平战略”。苏联最高领导人赫鲁晓夫提出一套争取同美国平起平坐,实现苏美共同主宰世界的战略目标。美苏争霸局面逐渐形成。
(2) 阅读材料,结合所学,指出人类进入“太空时代”得到国际政治背景;(4 分)列举 20 世纪 60 年代以来我国取得的航天成就。(6 分)			

2.2 数据准备

由于模型参数的数量较多,训练机器学习模型需要大量的样本。而本文仅有来自先前实验的十余个排序列表。为了解决样本匮乏问题,本文需要生成大量的排序列表,如此才能充分训练机器学习模型。

经过了对各种研究方案的斟酌考量后,本文选择了直接通过随机交换已有排序列表中的文档来生成训练排序列表。然而直接对整个排序列表使用随机洗牌算法是效率低下的。相关文档在原排序列表中的位置通常比较靠前,这一性质会使原排序列表评价指标值通常维持在一个较高的水平。随机洗牌使得相关文档在新排序列表中变成了随机分布,这极大降低了新排序列表的评价指标值。针对这一问

题,研究选择每次只在部分排序列表上使用随机洗牌算法,这就决定了相关文档不会偏离其原位置太远,从而保证能够获得比较高的评价指标。同时,每次选择的部分间互有重叠,使得每个文档都有较小可能连续多次被交换,不会出现诸如前10的文档永远在前10这种违背随机的现象。综合以上探讨过程,可设计推得算法1的研发代码详见如下。

算法1 样本生成算法

```
Function SHUFFLE(rank, start, end)
random shufflerank[start...end]
return rank
Function GEN(rank, size, shift)
l ← length(rank)
s ← 0
while s < l do
SHUFFLE(rank, s, s + size)
s ← s + overlap
end while
rank ← SHUFFLE(rank, s, l)
map ← EVAL(rank)
return rank, map
Function PREPARE(rank, num, size, shift)
ranks ← list()
for i = 1, 2, ..., num do
genrank ← GEN(rank)
APPEND(ranks, genrank)
end for
return ranks
```

本文研究的15道高考题中只有8道题在教材中找到了相关文档,因此带有人工标注的相关文档的8道题参与了样本生成过程。本文对每道高考题生成100个排序列表,并通过人工标注得到了生成的排序列表的MAP值,研究中总计得到了800个 < 排序列表, MAP > 对。

最后生成的排序列表的MAP分布如图1所示。已有排序列表的MAP分布如图2所示。

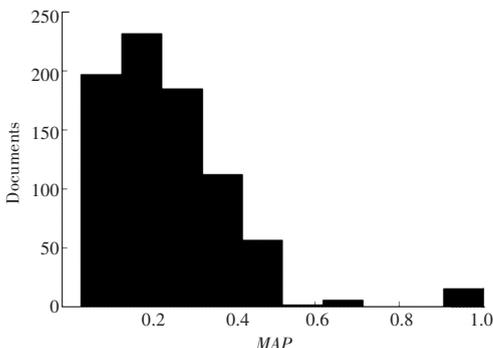


图1 原始MAP分布

Fig. 1 Original MAP distribution

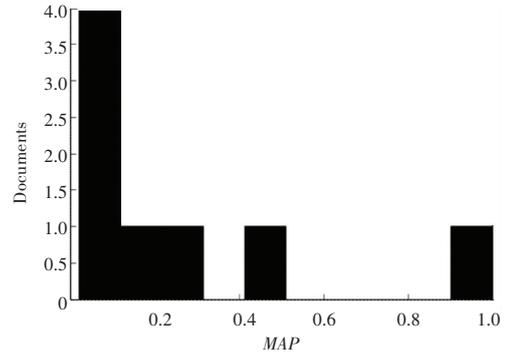


图2 生成样本MAP分布

Fig. 2 MAP distribution of generated sample

本实验中的每个排序都包含100个文档,而每个文档包含12个相关性指标,因此一个样本包含排序列表的1200个相关性指标和排序列表对应的MAP值。

2.3 模型评价

本文采用斯皮尔曼等级相关系数来衡量模型对于评价指标的预测能力。斯皮尔曼等级相关系数是衡量2个变量的依赖性的一项统计指标。相比于皮尔逊相关系数,斯皮尔曼等级相关系数,只关心变量间的大小关系而不是变量的具体大小,对于排序错误更加敏感,对于离群点更加不敏感,因此也较为符合进行自动评价以改进系统这一出发点。其计算公式如下:

$$\rho = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2 (y_i - \bar{y})^2}} \quad (11)$$

其中, x_i, y_i 是原始变量 X_i, Y_i 对应的等级。原始数据依据其在总体数据中平均的降序位置,被分配了一个相应的等级。 \bar{x}, \bar{y} 是所有原始变量等级的平均值。

2.4 特征选择

因为每个样本包含1200个相关性指标,而这里只有800个样本,容易出现过拟合的情况。为应对这一局面弊端,本文对回归模型和排序模型分别进行了特征选择。本文采用的是前向选择的特征选择算法。特征子集从空集开始,每次选择一个使得模型最优的新相关性指标加入特征子集,使得斯皮尔曼等级相关系数最优,是一种简单的贪心算法。为了避免挑选最优相关性指标时受到噪声的干扰,本文使用四重交叉校验,并将4次结果取平均。最终本文在回归模型和排序模型中都保留了7项相关性特征,细节详情见表3。单独使用每一个特征得到性能可见表4。在特征选择过程中,每增加一个

特征,性能增加的幅度如图 3 所示。

表 3 单独特征性能

Tab. 3 Individual feature performance

特征	回归模型	排序模型
精确率(1-gram)	0.35	0.25
召回率(1-gram)	0.11	0.02
精确率(2-gram)	0.23	0.17
召回率(2-gram)	0.12	0.06
精确率(3-gram)	0.45	0.30
召回率(3-gram)	0.52	0.46
精确率(word)	0.03	-0.02
召回率(word)	0.34	0.23
置信度(1-gram)	0.47	0.37
置信度(2-gram)	0.57	0.40
置信度(3-gram)	0.63	0.54
置信度(word)	0.42	0.25

表 4 特征选择结果

Tab. 4 Feature selection result

回归模型	排序模型
置信度(1-gram)	置信度(1-gram)
置信度(2-gram)	置信度(2-gram)
置信度(3-gram)	置信度(3-gram)
置信度(word)	精确率(1-gram)
精确率(1-gram)	召回率(2-gram)
召回率(2-gram)	召回率(3-gram)
召回率(word)	召回率(word)

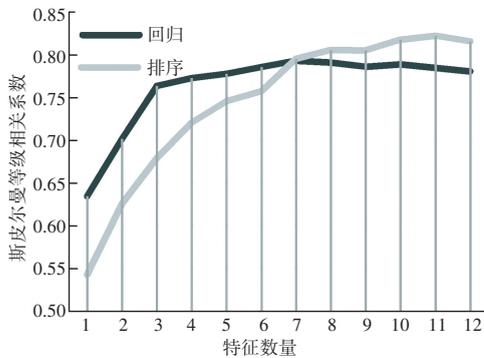


图 3 特征数量与性能关系

Fig. 3 The relationship of feature quantity and performance

2.5 实验结果

本文使用四重交叉测验来评估每个模型的性能。对于每个模型,本文首先利用开发集寻找使得模型性能达到最优的文档数目,然后在此基础上调节每个模型各自的参数使得模型性能再次达到最优。最终得到的实验结果可参见表 5。

表 5 不同模型性能

Tab. 5 The performance of different models

模型	相关系数	N
GBDT	0.87	100
随机森林	0.83	5
SVM 回归	0.83	100
Lasso	0.82	100
决策树	0.78	70
Ridge	0.72	100
OLSR	0.53	100
Rank SVM	0.59	100

由表 5 可知,GBDT 方法都取得了颇高的相关系数,达到了 0.87 左右,由此表明这种方法能够有效实现自动评价,同时可以看到非线性回归方法中除了决策树外,其余的所有方法都取得 0.8 以上的相关系数,彼此间的差异并不明显。值得一提的是,作为线性回归方法的一员,Lasso 的相关系数也达到了 0.82。相比于其同族的 OLSR 和岭回归有较明显的优势。

Rank SVM 作为排序模型,其性能与 OLSR 相近,且明显低于大部分回归模型,该算法存在的劣势,由此可见一斑。

在训练时间上,Rank SVM 作为 pairwise 的 Learning to rank 方法相比其它方法却更显劣势。而 GBDT、随机森林、决策树、SVM 回归,因为模型复杂度的原因,在训练时间上相比 Lasso、岭回归、OLSR 也略逊一筹,但都属于可以接受的范围。

综合以上分析即可得出,GBDT 这一基于提升树的方法已经能够有效解决面向问答系统的自动评价问题。

3 结束语

本文探讨了如何对问答系统中信息检索系统进行自动评价的 3 种方法。分别是分类、回归和排序。

本文重点研究了 12 种刻画问题本身和问题的答案两者与文档间相关性的指标,并通过特征选择获得了比较重要的 7 种。

本文使用 Lasso、岭回归、OLSR、SVM 回归、GBDT、随机森林、决策树、Rank SVM 来解决回归方法的学习问题,使用 Rank SVM 来解决排序方法的学习问题。经过实验发现,GBDT 模型结合了较快的训练速度和最高性能 2 个优点,本文推荐使用该模型进行自动评价。

应该指出,目前的实验数据集规模还仅为有限,理想状态是在完全真实而非人工生成的数据集上验证本文的结论。此外,本文的方法还需要提供问题的参考答案或者答案要点,如何在无参考答案的情况下进行检索模块的性能估计,仍是一个有待探讨的技术问题。

参考文献

- [1] 毛先领, 李晓明. 问答系统研究综述[J]. 计算机科学与探索 2012, 6(3): 193-207.
- [2] DANG H T, LIN J, KELLY D. Overview of the TREC 2006 question answering track [C]//15th Text REtrieval Conference. Gaithersburg, Maryland; NIST, 2006:99-116.
- [3] National Institute of Standards and Technology. TREC - 2004 common evaluation measures [EB/OL]. <http://trec.nist.gov/pubs/trec14/appendices/CE.MEASURES05.pdf>.
- [4] THORSTEN J. Evaluating retrieval performance using clickthrough data [M]//FRANKE J, NAKHAEIZADEH G, RENZ I. Text Mining. Berlin/Heidelberg: Physica/Springer Verlag, 2003: 79-96.
- [5] SOBOROFF I, NICHOLAS C, CAHAN P. Ranking retrieval systems without relevance judgments[C]//Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval. Louisiana, USA; ACM,

2001:66-73.

- [6] TIBSHIRANI R J. Regression shrinkage and selection via the lasso [J]. Journal of the Royal Statistical Society. Series B (Methodological), 1996, 73(1): 273-282.
- [7] SMOLA A J, SCHÖLKOPF B. A tutorial on support vector regression[J]. Statistics and computing, 2004, 14(3): 199-222.
- [8] HO T K. The random subspace method for constructing decision forests [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1998, 20(8): 832-844.
- [9] FRIEDMAN J H. Greedy function approximation: A gradient boosting machine [J]. The Annals of statistics, 2001, 29(5): 1189-1232.
- [10] JOACHIMS T. Optimizing search engines using click through data [C]// Proceedings of the eighth ACM SIGKDD international conference on knowledge discovery and data mining. Alberta, Canada; ACM, 2002:133-142.
- [11] STROHMAN T, METZLER D, TURTLE H, et al. Indri: A language-model based search engine for complex queries [C]// Proceedings of the International Conference on Intelligence Analysis. Guangzhou, China; IEEE, 2005: 1-2.
- [12] ZHAI Chengxiang, LAFFERTY J. A study of smoothing methods for language models applied to Ad Hoc information retrieval [C]// Proceedings of the 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '01). New Orleans, Louisiana, USA; ACM, 2001: 334-342.

(上接第 261 页)

4 结束语

智能交通协调控制系统的设计研究已经成为时下学界的热点研究课题之一。本文则是重点围绕当前国内部分城市不时发生的交通拥堵问题,基于多种成熟研发工具的强大特点优势,并引入了最优路径规划算法设计与数据挖掘及分析技术,探讨给出了一个智能交通协调控制系统。该系统推出后,即能将从出发点到目的地规划所得的最优结果路径快捷、准确、且清晰地显示在用户界面上,这就不仅可以有效节省日后人们驾车出行的时间及生活成本,而且有助于进一步调整优化未来的交通出行环境。经过系统仿真测试,该系统运行效果良好,具有一定的实用性和推广价值。

参考文献

- [1] 王伟耀. 人工智能技术在智慧交通领域中的应用[J]. 电子技术与软件工程, 2018(3): 251.
- [2] 周济, 陈锋, 韦达利, 等. 城市交通区域协调的模糊控制研究[J]. 电子技术, 2010, 37(11): 22-24, 21.
- [3] 李军. 城市智能交通中的动态路径规划研究[D]. 杭州: 杭州电子科技大学, 2016.
- [4] KALMAN D. A singularly valuable decomposition: The SVD of a

matrix [J]. The College Mathematics Journal, 1996, 27(1): 2-23.

- [5] 王海梅. 基于 GIS 的最优路径算法研究与实现 [D]. 南京: 南京理工大学, 2008.
- [6] 滕国文, 丛飏, 李闯, 等. 算法设计方法与优化 [M]. 北京: 清华大学出版社, 2013.
- [7] 林清岩. 智能交通中车辆最优路径规划策略研究 [D]. 长春: 吉林大学, 2013.
- [8] 伍伟. 智能化交通路径优化算法的设计与实现 [D]. 南昌: 南昌大学, 2016.
- [9] 王峰, 游志胜, 曼丽春, 等. Dijkstra 及基于 Dijkstra 的前 N 条最短路径算法在智能交通系统中的应用 [J]. 计算机应用研究, 2006(9): 203-205, 208.
- [10] 温毓铭, 滕国文, 杨建强. 基于物联网的智能交通道路导航系统 [J]. 计算机时代, 2018(9): 11-13.
- [11] 王珊, 萨师焯. 数据库系统概论 [M]. 4 版. 北京: 高等教育出版社, 2006.
- [12] 光俊华, 闫献国, 郭宏. 基于网络化制造 ASP 平台的虚拟组合夹具站的研究 [J]. 机械工程与自动化, 2008(6): 60-62, 65.
- [13] 那光宇. 车辆生产管理系统的设计与实现 [D]. 长春: 长春工业大学, 2018.
- [14] 阴增辉. 基于 J2EE 的 OA 系统的设计与实现 [D]. 天津: 天津大学, 2016.
- [15] 侯亚君. 基于 Java 语言信息系统平台的开发与设计 [J]. 信息安全与技术, 2013, 4(12): 66-67.
- [16] 杨剑, 郭建新, 杨菊英, 等. 数据结构 [M]. 北京: 清华大学出版社, 2011.
- [17] 马文阁. 基于模糊控制的单交叉口信号控制方法与算法研究 [D]. 大连: 大连海事大学, 2007.