

# 基于汽车领域的情感问答系统设计与实现

栾家阳 张文波 姚天昉

上海交通大学计算机科学与工程系 上海 200240

kalpop@sina.com, hellowenniu@gmail.com, yao-tf@cs.sjtu.edu.cn

**摘要:** 传统的问答系统通常只针对比较简单的、基于事实的问句。随着网络上大量的评论信息的涌现, 针对这类信息的观点和情感问题的答案往往不是唯一的。因此, 专门针对情感问题的情感问答系统已经成为一个研究热点。本文设计和实现了一个基于汽车领域的情感问答系统。首先介绍问答系统相关知识以及本文问答系统的构造, 然后重点介绍语料库的生成, 情感候选答案检索和分类过程中主题匹配, 答案的条件过滤以及排序的实现, 最后对实验结果进行分析并提出将来的工作。

**关键词:** 情感问答系统, 检索与分类, 答案语料库, 主题匹配, 条件过滤

## Design and Implementation of Opinion QA System Based on Automobile Domain

Jiayang Luan, Wenbo Zhang, Tianfang Yao

Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai 200240

kalpop@sina.com, hellowenniu@gmail.com, yao-tf@cs.sjtu.edu.cn

**Abstract:** Normally, traditional Question/Answer systems aim to answer the questions that are easy or based on the facts. With greatly increasing of comments on the Internet, the answers to opinion and sentiment question are not exclusive anymore. Consequently, research on opinion Q/A system is becoming more and more popular. An opinion question answering system based on automobile domain is designed and implemented in this paper. At first, we introduce related work about QA system and the architecture of our system; and then, put great emphasis on how to build answers corpus and how to get the final results by means of topic matching, condition filtration and sorting; at last, carry out experiments, analysis the final results and propose our future work.

**Keywords:** Opinion question answering system, search and classification, answers corpus topic matching, condition filtration

### 1. 引言

从 TREC1999 TREC Question Answering Track[1]的推出到 ACL2000 专题讨论会“Open Domain Question Answering”[2]的开设, 自动问答越来越成为最受关注的主题之一。目前比较成熟的问答系统有 START[3], AnswerBus[4], AskJeeves[5], ASKMSR[6]等。

问答系统是从问题分析到对相关信息进行精确检索的一个过程。作为一种新的信息抽取技术, 自动问答系统允许用户以自然语言形式提问, 而且返回给用户的结果不再是一段相关文本, 而是一个句子或者短语等确定性答案[7]。问答系统的工作流程一般包括问题分析、信息检索和答案抽取三部分。

## 2. 系统组成

本文设计的汽车领域情感问答系统分为四个模块：情感问题分析模块，答案语料库生成模块，答案检索模块和答案抽取模块。

当用户输入问题时，系统会将其传递给问题分析模块，后者对问题进行分析，得到问题的类型、主题词、限定条件、所包含的情感极性等信息，并将分析结果传递给答案检索模块。答案检索模块从答案语料库中找到相关的语料后，从上一步得到初步答案的答案抽取模块，对答案进行筛选和排序，并将最终答案传递给用户。

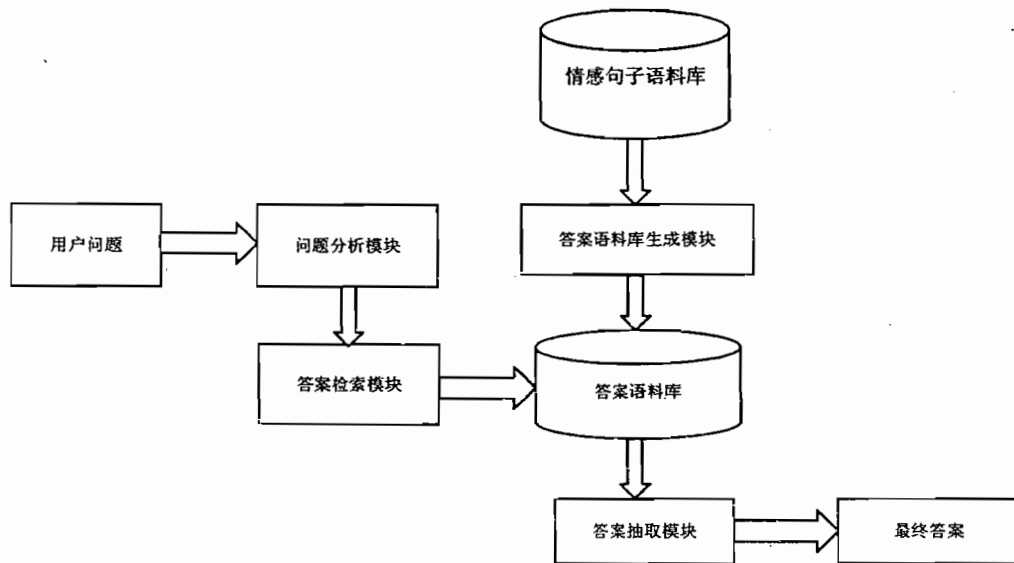


图 1 系统组成

## 3. 答案语料库的生成

本系统是基于固定语料的，高质量的答案语料库是系统快速找到答案的基础。通过分析语料的标注结果，按每句语料中所描述的主题结合汽车领域本体进行树状方式存储。标注语料符合《汉语意见型主观性文本标注规范》[8]，每句语料存于一个 xml 中。

### 3.1 语料库中的答案模式

由于语料库规模较大，所以定义了一个答案模式来区分语料之间的相同点和不同点，作为系统进行答案的分类与检索的依据之一。即对所有情感语料库中的情感答案的集合定义一个关系 Key(ID, 领域, 情感持有人, 情感表达方式, 情感类型, 情感极性)，具体属性如下：

- (1)ID: 每个情感答案有一个唯一的 ID
- (2)领域: 考虑到日后的可扩展性，每个候选答案都应给出所属的领域，如车、笔记本电脑、手机等。
- (3)情感持有人: 情感持有人一般包括大多数人、我、大家等。
- (4)情感表达方式: 指情感持有人表述情感所用到的词法模式和句法模式等，这是排序的重要依据之一。
- (5)情感类型: 指情感持有人表达情感的针对性的强弱程度，是排序的重要依据之一。
- (6)情感极性: 整句句子的情感的褒贬性和极性的强弱程度，其值为语料中所有情感元素的极性值之和。

### 3.2 语料存储结构

语料的存储结构是根据汽车领域本体建立的树状存储结构。树的节点对应汽车领域本体(ontology[9])的某个概念元素, 部件或子属性对应其子节点, 例如“车身”与“车门”、“安全性”与“避震”等。每个节点中存储的信息为含有对该节点所代表的本体或主题表达情感倾向的语料。如, 假设语料库中有以下4个文件, 对应4句话:

a.xml: “我喜欢桑塔纳, 真的非常不错哦。” (综合评价)

b.xml: “桑塔纳的性价比还行。” (性价比)

c.xml: “奥迪车质量很好。” (质量)

d.xml: “这车的质量有待提高。” (质量)

这四句话在语料库中的存储位置如图2所示:

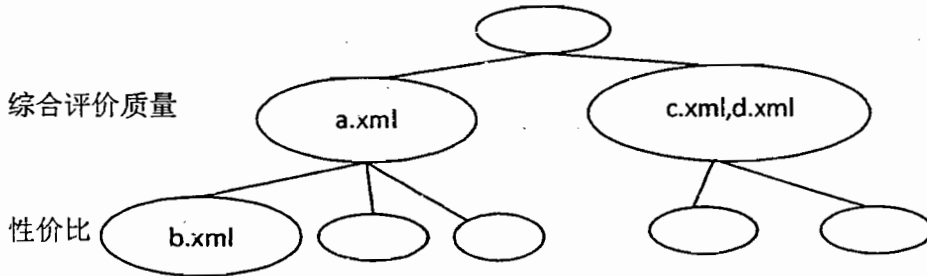


图2 答案语料库存储结构

### 3.3 答案语料的主题的识别

本文总结主题词识别方法主要有以下四种:

(1)字符串匹配。如“桑塔纳的性价比很高。”中的主题“性价比”。

(2)一词多义的处理。如“价格”和“价钱”、“价位”所表达的意思一样, 但这些近义词难以用单纯的字符串匹配来识别。本文利用了知网(HowNet)[10]的词语知识库来识别主题中的近义词。如“声誉”的属性(attribute|performance|implement)与“名声”的概念解释是完全一样, 所以“名声”被识别为这句语料所描述的主题之一。

(3)缺省、隐含主题的处理。如“这车好贵啊!”。主题为车的价格, 但价格与价格的形容词并未出现。而知网中描述某个极性词的概念时, 会给出该词通常所描述的领域和方向。“贵”的概念及义原为“aValue/price/expensive/undesired”。根据这个概念, 系统很容易就能识别出这句语料中有“价格”这一隐含主题。

(4)复杂主题的处理。如“奥迪的发动机性能很好”, 既有“发动机”又有“性能”, 那么这句话该如何存储呢? 这句话即有发动机的评价, 又可用于评价性能。因此在单个主题含有两个或多个汽车领域本体的情况时, 将这句语料存储在相应的多个位置上。

主题识别的算法伪代码如下:

算法:Find\_Topic

输入: S, S 为被标注出了各意见型元素的句子

输出: Topic[n] (表示识别结果的一个数组)

begin

if there is no topic in S,

for all the sentiment s in S

Topic\_Identify(s);

else

for all the annotated topics t in S

Topic\_Identify(t);

End

```

Function Topic_Identify(t);
begin
if t in the set of ontology
    store S in Node t;
else:
    String meaning := MeaningInHowNet(t);
    for all topics in ontology
        if MeaningInHowNet(ontology_topic)=meaning
            store S in Node ontology_topic;
end

```

其中：Topic\_Identify()是主题识别的函数，该函数先进行普通的字符串匹配，如果没有成功，调用函数 MeaningInHowNet() 查询答案语料中的主题在《知网》中的概念，查询汽车领域中的本体有无概念与答案语料的主体概念一致，如果有的话则匹配成功。

## 4. 答案的检索与分类

答案的检索与分类，即从答案语料库中找出与问句分析结果匹配的答案，对其进行条件过滤及排序，是本文研究的重点内容。本节将重点介绍其工作流程和方法。

### 4.1 问题分析

针对不同的情感问题，分析问题后所输出的结果除情感问题的原句外，还包含了以下五个要素：

- (1) 问题类型。问题的类型有五种，分别为 who、what、how、why 和 YesOrNo。
- (2) 情感持有人。通常指问题中出现的人称代词。如“你认为宝马车好吗？”中的“你”。
- (3) target，即问题中涉及的主题词和领域本体。
- (4) attitude，即问题中的情感元素及其倾向性。
- (5) 限制条件。如“你认为宝马车好吗？”，问的是人们对宝马车的看法，而不是其他车型，所以答案抽取模块会根据这个限制条件对候选答案进行条件过滤。

### 4.2 答案检索的工作流程

答案的检索过程主要有三个步骤：主题匹配、答案的条件过滤以及排序。如图 3

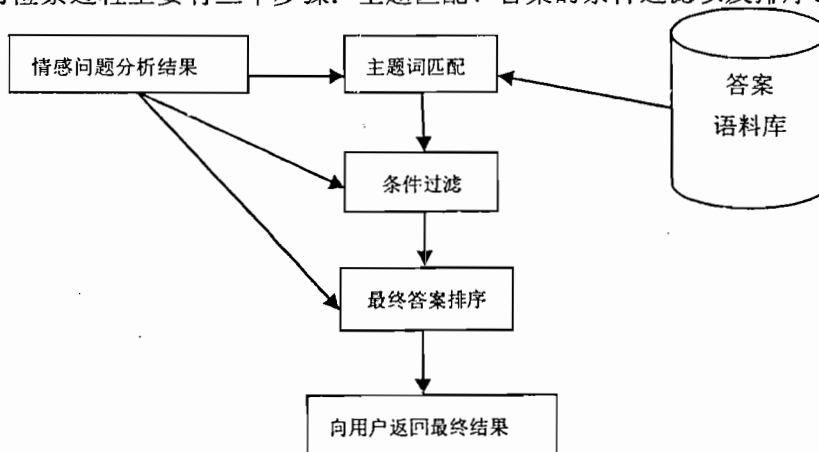


图 3 答案检索的工作流程

### 4.3 主题词匹配

主题匹配是根据问题中出现的主题词，在答案语料库中检索对这一主题进行评价或者发表情感的语料。判定答案语料库中的语料 B 是否与某情感主题 a 匹配的标准有两条：1. 语料 B 中是否含有主题 a 或其近义词 a'、a''等；2. 语料 B 中是否含有对主题 a 做出评价或者表达情感的情感元素。主题词匹配算法有下列四种：

#### 4.3.1 字符串匹配

字符串匹配是最简单的主题匹配算法。当发现问题中的主题词恰好是汽车领域本体中的某一个概念时，在语料库中找出相应的节点，将存储在该节点中的所有语料取出，进行答案的条件过滤。

#### 4.3.2 近义词匹配

对概念或主题的一个或多个近义同义词需要进行近义词匹配。进行近义词匹配时，对主题 a，答案检索模块查询词语 a 在知网中的概念 meaning\_a。若语料库中的主题 b 在知网中的概念与 a 相同，则主题 b 中的所有答案语料与该问题匹配成功。

#### 4.3.3 缺省主题匹配

如果用户提出的问题中找不到无法找到显式主题。如“宝马车贵吗？”，询问的是宝马车的价格，但无法从问题中找到“价格”这个主题。系统会从问题中的情感词分析问题中的“attitude”属性，查询问题中的情感词 a 在知网中的概念，就可以找出这个情感词通常所描述的对象或主题 A。“贵”在知网中的概念为“aValue/price/expensive/undesired”，可知它通常用于描述价格，故与“价格”成功匹配。

#### 4.3.4 子主题的匹配

假设某情感问题 Q 询问的是对车的属性 A 的评价，A 即是这个情感问题的 attitude，同时 A 也是汽车领域本体的一个概念，并有一个子属性 B。那么，语料库中对属性 B 发表评价或是情感的语料也是可以用来回答情感问题 Q。

### 4.4 条件过滤

由主题匹配部分检索出来的语料还不能作为最终结果返回给用户，因为每个情感问题都蕴含某些对答案的限制条件，因此需要根据情感问题中的限制条件，对检索出的语料进行条件过滤。汽车领域的情感问题中主要含有的三种限制条件及处理方法如下：

#### (1) 汽车品牌的限制。

如“宝马车的性能怎么样？”，需要根据“宝马”这一汽车品牌的限制条件对检索到的语料进行条件过滤，去除那些对其他品牌、车型发表情感的语料。

#### (2) 主题修饰语部分的限制。

情感问题中，在主题词前后出现的定语、状语、补语都可能含有对该问题的限制条件。如“你认为近年来宝马车的价格调整合理吗？”，“调整”一词对此问题给出了明确的限制，即用户对价格的变化趋势的看法。对于此类问题目前所用的方法主要是搜集常见修饰词，如“趋势”、“变化”等，并搜集在回答含有这类限制条件的情感问题中，出现频率较高的关键词，在答案检索时进行匹配。

#### (3) 问题类型与情感极性的限制。

有一些情感问题本身就对答案的情感极性做出了限制。如在回答“谁觉得宝马车好？”时，所检索和条件过滤出的答案都应包含对宝马车“好”的性能评价。

### 4.5 最终答案排序

当返回给用户多个答案时，系统有必要对这些候选答案按一定的规则进行排序。排序评判依据为关系模式 key 中的情感持有人，情感表达方式，情感类型，极性强度，各部分所占比重如表 2 所示。

情感持有人只有在对 who 类型问题时的候选答案进行排序时，占的比重较大，因为只有回答 who 问题时语料中的情感持有人存在与否才比较重要。

不论在回答何种类型的情感问题时，适当的情感表述方式和情感类型都是必须的，所

问题类型	情感持有人(%)	情感表述方式(%)	情感类型(%)	极性强度(%)
Who	30	20	20	30
What	10	30	30	30
How	20	40	40	--
Why	10	30	30	30
Yes/No	20	40	40	--

表 2 候选答案与问题的匹配度判定

以它们的比重在面对各个问题类型时都大体相同。

在回答 How 和 Yes/No 类型的问题时，前者问题本身并不一定出现关键的情感词，后者的情感词只用于限制答案语料中相关情感元素的极性，所以答案语料的极性强度对排序结果没有影响，相反在对其余三种问题的候选答案的排序结果会有较大影响。

## 5. 实验结果与分析

实验时我们选取了 15 个不同的问题，5 种类型各三个，问题由简单到复杂，分别是：

QA1：有人认为奥迪是一款好车吗？

QA2：是否有人认为奥迪价格的确贵了？

QA3：谁认为奥迪比帕萨特好？

QB1：什么车贵？

QB2：什么车很贵？

QB3：什么车健步如飞？

QC1：捷达车怎么样？

QC2：捷达车零部件怎么样？

QC3：捷达和其它车相比怎么样？

QD1：为什么说这车好？

问题	准确率	召回率
QA1	100%	78%
QA2	88%	70%
QA3	66%	50%
QB1	80%	80%
QB2	80%	80%
QB3	---	0%
QC1	100%	60%
QC2	100%	63%
QC3	---	0%
QD1	88%	90%
QD2	80%	83%
QD3	80%	70%
QE1	93%	72%
QE2	---	---
QE3	50%	50%

表 3 系统针对以上 15 个问题的实验结果

- QD2: 为什么说大家都喜欢买桑塔纳?  
 QD3: 为什么有人认为桑塔纳没有其它车好?  
 QE1: 桑塔纳便宜吗?  
 QE2: 桑塔纳性能好吗?  
 QE3: 宝来和其它车相比, 有优势吗?

最后得到的实验结果如表 3 所示。

从实验结果可以看出, 对于简单问题, 系统的准确率和召回率都有不错的表现。随着问题由简单到复杂, 系统的性能有所下滑, 这也符合问答系统性能变化的一般规律。但实验中有些项存在重大缺陷, 例如语料库中无法找出任何答案来回答 QC3 和 QE2, 回答复杂问题的性能不理想。原因总结为以下两点:

- (1) 语料库规模有限。这是语料库问答系统无法回避的一个瓶颈。有些问题在语料库中无法找出答案, 这就需要根据用户需要随时对语料库进行改善。
- (2) 对于复杂问题的处理和对于复杂答案的识别能力不强。系统往往无法很好处理那些带有询问两者比较、某属性变化趋势方面的问题, 这也是汉语语言处理长久以来的难点。

## 6. 结论与展望

本文实现了汽车领域的情感问答系统。通过构建本地候选答案语料库, 在得到情感问题及其分析结果后, 对答案语料进行主题匹配与条件过滤, 检索出初步的候选答案, 再对候选答案进行排序, 并将最终的候选答案提交给用户。其中重点介绍了候选答案的检索与分类的流程及算法, 通过实验论证, 得到了比较理想的结果。如何将问答系统从汽车领域扩展到多领域, 如何将系统从本地语料库扩展至网络问答系统是将来研究重点和突破方向。

## 参考文献

- [1] E. Voorhees. Overview of the TREC-9 Question Answering Track[C]. In: Proceedings of the 9th Text Retrieval Conference (TREC9), NIST, Gaithersburg, MD, 2000, 71-80.
- [2] Ulf Hermjakob. Parsing and Question Classification for Question Answering[R]. Proceeding of the workshop on Open-Domain Question Answering at ACL-2001
- [3] Lin J, Fernandes A, Katz B, Marton G, et al. 2002. Extracting Answers from the Web Using Knowledge Annotation and Knowledge Mining Techniques[C]. In proceedings of the Eleventh Text REtrieval Conference (TREC2002).
- [4] Zhiping Zheng. 2002. AnswerBus question answering system[C]. In Proceeding of 2002 Human Language Technology Conference (HLT 2002).
- [5] Ask Jeeves, <http://www.ask.com>
- [6] hwick A B. A Maximum Entropy Approach to Named Entity Recognition : [ Ph D Thesis ] [D]. Department of Computer Science , New York University , 19991
- [7] 郑实福, 刘挺, 秦兵, 李生. 自动问答系统综述[J]. 中文信息学报. 2002, 16(6): 46-52
- [8] 刘全升, 姚天昉, 黄高辉, 刘军, 宋鸿彦. 汉语意见型主观性文本类型体系的研究[R]《中文信息学报》. 2008 22(6)
- [9] Gruber T R. Towards Principles for the Design of ontologies Used for Knowledge Sharing[J]. International Journal of Human Computer Studies, 1995, 43: 907-928.
- [10] Fuchun Peng, Fangfang Feng, Andrew McCallum. Chinese segmentation and new word detection using conditional random fields(to appear)[C]. Proc of COLING, 2004.