

文章编号: 1003-0077(2014)04-0111-06

# 交互式问答系统中待消解项的识别方法研究

张超,孔芳,周国栋

(苏州大学 计算机科学与技术学院自然语言处理实验室, 江苏 苏州 215006)

**摘要:**交互式问答系统能够与用户进行对话式交互进而处理用户提出的一系列问题。交互式问答技术是近年来问答技术的一个热门方向。该文首次深入研究交互式问答中待消解项的识别方法。根据语料统计了交互式问答中待消解项的分布情况并进行相关实验,运用前人研究的启发式规则与平面特征相结合的方法在交互式问答中测试识别待消解项的性能。结合交互式问答的特点提出了专有名词的两个基于交互式问答特点的特征,并在 TREC QA 问题集语料中进行相关实验。实验结果表明,代词、有定名词用已有的方法识别效果较好,在加入本文提出的新特征后,在专有名词上也取得了较好的效果。

**关键词:**交互式问答;待消解项识别;指代消解

**中图分类号:** TP391      **文献标识码:** A

## Detecting Anaphoricity for Coreference Resolution in Interactive Question Answering

ZHANG Chao, KONG Fang, ZHOU Guodong

(NLP Lab, School of Computer Science and Technology, Soochow University, Suzhou, Jiangsu 215006, China)

**Abstract:** Interactive Question Answering (IQA), a hot research topic in the area of QA, can interact with users to process a series of questions from users just like talking to them. This paper systematically explores anaphoricity determination for coreference resolution in IQA. The statistic of the corpus shows the distribution of anaphoricity and the experiment in the TREC QA questions set which uses the rule-based and flat feature-based method shows the performance of anaphoricity determination for coreference resolution in IQA. On the basis of the characteristic of IQA, two flat features about proper noun are proposed. Experimental results show that the proper method and the proposed feature is effective.

**Key words:** interactive question answering; Anaphoricity Determination; coreference resolution

## 1 引言

问答系统(Question Answering, QA)作为一种能够对用户以自然语言提出的问题作出自动回答的技术受到越来越多的关注。问答系统的研究一直以来都是自然语言处理领域的重要内容<sup>[1]</sup>。但传统问答技术只局限于对单一问题回答的方法研究,这种一问一答的形式所能提供的相关信息具有一定的局限性,在很多情况下并不能满足用户对同一实体或事件多方面复杂信息的需求。交互式问答系统允许用户提出一系列与主题相关的问题,以对话的形式逐个回答用户关于同一实体或事件多方面的问題<sup>[1]</sup>。目前,交互式问答技术已经作为问答系统的

一项重要技术受到广泛关注:文本信息检索会议(TREC)在 2004 年的 QA 任务中开始以系列问题的方式给出问题评测集<sup>[2]</sup>,每个系列提供一个对象(Target),系列问题中的每个问题都是与该对象相关(例 1: Q1:What film introduced Jar Jar Binks? Q2:What actor is used as his voice? Q3:To what alien race does he belong?);同时,交互式问答技术在工业界的应用也受到青睐,例如,苹果公司的 Siri 语音助手、许多不同领域的自动咨询机器人等。

然而,目前许多运用交互式问答技术的工业界应用(如 Siri)都存在一个缺陷:对于交互式问答中的上下文信息运用得不够全面和充分。而这同样是交互式问答技术的重点和难点。以对话的形式来回答用户的提问必定会出现其他自然语言文本中常见

的指代现象和成分省略现象,这就需要运用上下文信息来进行指代消解和省略恢复以便于更准确更全面地理解用户的提问,例如,例 1 中 Q2 代词 his 的消解问题。

指代是自然语言文本中常见的语言现象,它是指语篇中的指代词回指之前的某个语言单位。指代的作用在于其能够使文章更加简洁明了,并能够帮助保持文章的连贯性。指代消解是指为篇章中的指代词寻找所指代的语言单位的过程。在交互式问答中,指代消解就是为用户问题中的指代词寻找其具体指代的内容,以便于更好地理解用户的问题来获取更准确的答案。通常来讲,指代消解具体分为两个步骤:(1)待消解项的识别;(2)对待消解项的消解。作为指代消解任务的第一步,待消解项识别的主要工作是识别哪些词或短语是具有指代现象的指代词。近期的许多指代消解方面的相关研究表明:待消解项的识别工作能较好地提高指代消解的性能<sup>[3]</sup>。虽然待消解项识别已经存在相关研究<sup>[4]</sup>,但是由于交互式问答技术的对话文本与传统普通文本有着很大的差异,交互式问答系统中问句形式比较多并且口语化表达比较频繁,这会对待消解项识别工作造成一定的影响,因此交互式问答系统中待消解项的识别工作不同于传统普通文本中的待消解项识别工作。

本文将在接下来的第 2 节中介绍交互式问答以及待消解项识别的相关研究;在第 3 节介绍本文提出的关于交互式问答系统的框架;第 4 节详细阐述交互式问答系统中的待消解项识别方法并在第 5 节给出了实验结果和分析;最后一节是本文的内容总结及工作展望。

## 2 相关工作

自 TREC2004 的 QA 任务中加入系列问题任务以来,部分研究者便意识到交互式问答技术的重要性,并相继开展了交互式问答技术的相关研究。Chai<sup>[5]</sup>等人最早指出了用户与问答系统对话的形式所获取的信息比一问一答形式获取的信息更加全面

准确。用户往往是想询问关于一个特定主题的信息或者是如何解决一个特定的任务,仅仅一个问题往往满足不了用户的需求。交互式问答技术能够以对话的形式解答用户的一系列相关问题,Carbonell<sup>[6]</sup>和 Nils<sup>[7]</sup>等都指出了对话领域中指代现象出现频繁,是计算机理解人机对话的一大障碍。Dongsheng Wang<sup>[8]</sup>提出使用本体与模板的方法来利用交互式问答中的上下文信息,Joyce Y Chai<sup>[9]</sup>根据交互式问答的特点提出了对话篇章模型来使用上下文信息,但两者都并未真正在指代词上进行消解工作。

早期的指代消解并未考虑待消解项的识别任务,代表性的是 Soon 等人<sup>[4]</sup>提出的机器学习方法,将指代消解问题化为二元分类问题,并未进行待消解项识别。随着指代消解研究的日益深入,待消解项识别越来越受到重视。Lappin 等人<sup>[10]</sup>最早在指代消解平台中加入了用于识别“it”是否是待消解项的识别模块;Bean 和 Riloff<sup>[11]</sup>提出了将基于统计的待消解项识别特征运用到指代消解平台中。

随着大规模语料的出现,基于机器学习方法的待消解项识别方法逐渐成为热点。Evans 等<sup>[12]</sup>从语法、句法等角度分析了成为待消解项的“it”的特点,归纳出判定“it”为待消解项的特征集;Ng 等人<sup>[4]</sup>选用包含了语法、句法、语义、位置、构词等多方面信息的特征,给出了一种基于机器学习方法的待消解项识别方法;Müller 等人<sup>[13]</sup>探索了对话领域中非待消解项“it”的判定方法。Zhou 等人<sup>[14]</sup>给出 17 个关于代词、有定名词、指示性代词、专有名词的特征,利用标记传播算法,在机器学习的基础上对待消解项识别进行了全局优化。但上述研究都未在交互式问答中作性能测试。

同以往研究不同的是,本文首次在交互式问答任务方面提出待消解项识别问题,并针对该任务给出相应的识别方法和有效特征。

## 3 交互式问答系统框架

根据交互式问答中存在的指代和省略现象,本文提出了改进的交互式问答系统框架,如图 1 所示。

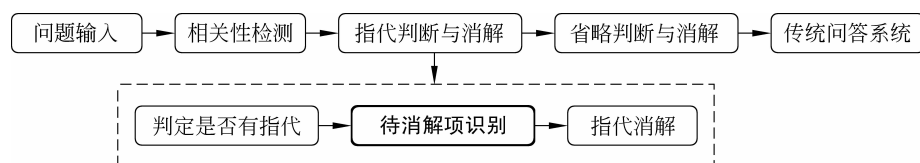


图 1 考虑指代和省略消解的交互式问答系统框架图

同传统的交互式问答系统不一样的是,该框架将用户提出的问题进行指代消解和省略恢复工作,将原问题恢复到传统问答技术能够回答的状态。本文充分利用已有的问答技术的研究成果,并结合文本篇章结构处理方法,所提出的系统框架能够更好地处理实际应用中的交互式问答问题。

本文主要研究图 1 虚线框中加黑部分,即待消解项识别方法。下一部分将详细阐述待消解项识别的具体方法。

4 待消解项识别方法

本文以孔芳等人<sup>[15]</sup>基于 ACE 语料库提出的规则与机器学习方法相结合的待消解项识别方法为基础,在交互式问答对话中进行待消解项识别方法的研究工作,探究在传统语言文本上的待消解项识别方法在交互式问答对话文本上的适应性,找出两者的不同之处,并根据交互式问答的特性优化孔芳等人提出的方法。

待消解项的识别任务与一些其他的 NLP 研究任务的不同之处在于某些待消解项的识别是与上下文无关的,这时候就需要更多地运用语言学的知识去建立相关的规则进行判断。由于语言现象纷繁复杂,全部列举出不同的语言现象并制定规则非常困难,而且某些语言现象很难用简单的规则来准确表达,同时,部分待消解项识别任务需要运用到上下文信息,因此本文运用规则方法与机器学习方法相结合的方法。

Zhou 等<sup>[14]</sup>对新闻语料统计后发现,能够成为指代词的语言单位主要有以下几类:代词、专有名

词短语、有定名词短语、无定名词短语、指示性名词短语和未知类别名词短语。本文在交互式问答对话语料中统计后发现在交互式问答中的待消解项主要是代词、专有名词以及有定名词三类,而无定名词短语、指示性名词短语和未知类别名词短语所占的比重极少,因此本文忽略这几类,仅考虑代词、专有名词以及有定名词这三种类别的规则与特征方法研究。

4.1 启发式规则方法

基于规则的待消解项识别方法是运用启发式规则来判定候选词是否是待消解项的方法。本节余下部分将分别介绍针对代词、专有名词短语和有定名词短语制定的不同的启发式规则。

4.1.1 针对代词的消解项识别规则

代词是最常见的指代词,但并非所有代词都有指代内容,最常见的情况是代词“it”。中性代词“it”常常用在固定的短语结构和句式中,并未有指代任何语言单位,例如,“It is so hot today.”、“It is Bill that makes me happy.”。孔芳等<sup>[15]</sup>参考 Lappin 和 Leass<sup>[10]</sup>的研究成果总结了 3 种“it”系表结构:关于气象的“it”,关于时间的“it”以及被动结构中的“it”,归纳了 9 种不同的“it”是否是待消解项的判断规则。

但上述的规则集都是在陈述句中的固定结构,而交互式问答文本与新闻文本有很大不同,问答对话中多是问句,并有很少一部分的祈使句。为此根据交互式问答文本的特点,在前人的研究基础上,本文总结了在问句中的“it”为非待消解项的特征集。

表 1 交互式问答中非待消解项“it”的表现形式

类别	表现形式	说 明
天气	<be> it ADV * PCP	PCP 表示 12 个常用的气候名词,例如,raining。
	<be> it ADJ	ADJ 是 44 个关于气候的形容词,例如,cloudy。
	<be> it N	N 是 5 个季节名词:spring、summer、autumn、fall、winter。
时间	<be> it ADV * NUM	NUM 代表数字,例如,one 等。
	<be> it ADV * PREP * time	PREP 代表介词,time 是英文代词,例如,It's about time。
	<be> it ADV * (early late)	表示有 early 或 late 出现,例如,It's early。
被动结构	<be> it modadj that	Modadj 是形态形容词,例如,easy、usefull 等。
	<be> it modadj [for N] to V	[for N]可选,这一形式也是被动结构的主要表现形式。
	N Vcog it modadj [for N] to V	Vcog 是一些情感动词的形式。
	<be> it modadj PRON	PRON 表示代词,例如,It seems he will. . .

4.1.2 针对专有名词短语的消解项识别规则

专有名词是指表示人、地方、组织机构等特有事物的名词。专有名词待消解项的识别需要结合语义、语法、句法、词法等多方面信息,仅仅使用规则信息很难包含这些信息,因此专有名词的待消解项识别主要通过特征方法。规则方法主要根据 WordNet 判断专有名词是否是人名(Person)、地名(Location)、机构名(Organization)、日期(Date)、时间(Time)、百分数(Percentage)、货币(Monetary value)中的一类,若不在其中则为非待消解项。

4.1.3 针对有定名词短语的消解项识别规则

有定名词是指带有定冠词 the 的名词短语,例如: the organization。有定名词在指代词中占有很大比重。Poesio 等人<sup>[16]</sup>将其分为 5 类: 语义功能的有定名词(例如, the second place)、专有型有定名词(例如, the FBI)、同位语型有定名词(例如, Bush, the president of USA)、特定型有定名词(例如, the place where...)、广义有定名词(例如, the earth)。对于有定名词在交互式问答中作为待消解项的特征,可总结出以下确定为非待消解项的启发式规则:

1. the+形容词最高级/序数词/专有名词+none 的有顶名词。
2. the+广义名词,如 the earth、the sun 等。
3. the + none+that 的形式。
4. 有定名词+of 的形式。
5. 有定名词+关系从句。

满足上述规则的有定名词将判定为非待消解项。

4.2 机器学习方法

基于规则的待消解项识别方法的优点是能够利用语言学知识,对规律性强的类别有较好的识别能力。但是基于规则方法不能够完全穷举,且无法运用词法、句法、语法等信息。随着大规模语料的出现,基于特征的方法在待消解识别研究中起着重要作用。本文结合交互式问答的特点与前人的研究提出了针对代词、专有名词短语和有定名词短语的基于特征的待消解项识别方法,经过实验后选择了适应交互式问答特点的平面特征。

特征方法主要描述候选词的两方面信息: 当前对象自身信息与当前对象所处的上下文信息。其中描述三类候选词的上下文信息有一些是共同的特

征,主要描述对象的语义特征,如表 2 所示。

表 2 描述候选词上下文信息的特征集

特征	说 明
Str_Match	在当前对象之前是否有全串匹配的对象。
Arg0	当前对象是否承担某一动词驱动的 A0 角色。
Args	当前对象是否承担任何语义角色。

同时,根据三类候选词的不同特点,选取并制定了各自不同的描述三类候选词上文信息的特征以及描述它们各自自身信息的特征,如表 3~5 所示。

根据特征方法的要求,特征集分为描述当前对象自身信息和描述上下文信息的两类特征。其中描述当前对象自身信息的特征主要为词性、大小写、短语构成等信息;描述上下文信息的特征主要为描述匹配、语义以及一些句法信息。表 4 中加黑部分的两个特征是本文根据交互式问答的特点提出的两个新特征。

表 3 代词待消解项分类器特征集

特征集	特 征	说 明
描述当前对象自身信息的特征	Possessive	是否是所有格名词。
	Embedded	是否是名词前的修饰语。
	Number	是否是单数。
	MaleorFemale	是否是男性。
	FirstPersonPRON	是否第一人称代词。
	SecondPersonPRON	是否第二人称代词。
	ThirdPersonPRON	是否第三人称代词。
描述当前对象上下文信息的特征	WordSense	在对象之前是否出现语义类相同的对象。
	Appositive	与对象具有同位语关系的对象存在。
	NestOut	是否嵌套在某一名词短语中。
	Arg0MainVerb	是否在承担 A0 角色时由主驱动词驱动。
	FirstNP	是否是语句中第一个名词短语。
	DistanceForBack	距后一个 that、if、whether 等词的距离。
	DistanceForFront	距前一个 that、if、whether 等词的距离。

表 4 专有名词待消解项识别分类器特征集		
特征集	特 征	说 明
描述当前对象自身信息的特征	Uppercase	是否为英文大写字母。
	First_Word_U	是否首字母大写。
	PERSONProperNP	是否表示人名。
	LOCProperNP	是否表示地名。
	ORGProperNP	是否表示机构名。
描述当前对象上下文信息的特征	NameAliasMatch	在对象之前是否出现与之别名匹配的对象。
	Head_Match	在对象之前存在与之别名匹配的对象。
	Appositive	与对象具有同位语关系的对象存在。
	NestIn	是否存在名词短语嵌套在当前对象中。
	NestOut	是否嵌套在某一名词短语中。
	First_NP_Paper	是否是第一句中的专有名词。
	NEROut	前后是否有命名实体。
	NNOut	前后是否有名词。

表 5 有定名词待消解项识别分类器特征集		
特征集	特 征	说 明
描述当前对象自身信息的特征	Head_Uppercase	中心词是否是大写单词。
	The_N	“the”+一个普通名词。
	The_2N	“the”+两个普通名词。
	The_ADJ_N	“the”+形容词+普通名词。
	The_NUM_N	“the”+数量词/序数词+普通名词。
	The_NE	“the”+命名实体。
	The_SING_N	“the”+单数名词。
描述当前对象上下文信息的特征	Head_Match	存在与之有相同中心词的对象。
	NameAliasMatch	出现与之别名匹配的对象。
	First_NP_Paper	是否是对话第一句中的名词短语。
	First_NP_Sent	是否是语句中第一个的名词短语。
	DistanceForBack	距后一个 that 的距离。

5 实验结果与分析

5.1 实验设置

本文采用 TREC2004 至 TREC2007 的 QA 评

测任务的 286 个问题集 1 962 个问题作为实验语料<sup>[2]</sup>。通过对语料统计后发现共有 771 处指代,其中代词在语料中所占的比例最大,占 53.48%;有定名词也有一定比例,占 31.97%;专有名词占 13.43%;而其他情况只占 1.12%。因此,本文仅仅考虑代词、有定名词和专有名词这三类。

在基于机器学习方法中,本文采用 SVM-Light 工具中径向基核函数(RBF)来进行训练与测试各类别的待消解项识别分类器的性能,选取 60%的语料进行训练,20%的语料作为开发集,其余 20%作为测试语料。

在指代消解任务中加入待消解项识别任务的目的是尽可能地保留需要消解的消解元素并且能够过滤掉干扰信息。多位研究者(例如,Bean 等人<sup>[11]</sup>;Müller 等人<sup>[13]</sup>)认为在待消解项识别任务中召回率比准确率重要,因为如果待消解项被误认为非待消解项,则会直接降低后续指代消解的性能。本文采用 Zhou 等人<sup>[14]</sup>提出的两个准确率来评估待消解项识别器的性能:Acc<sup>+</sup>和 Acc<sup>-</sup>。

$$Acc^{+} = \frac{\text{正确识别的待消解项个数}}{\text{应识别的待消解项个数}}$$

$$Acc^{-} = \frac{\text{正确识别的非待消解项个数}}{\text{非待消解项总个数}}$$

Acc<sup>+</sup>表示的正例准确率,这一准确率越高说明丢失的待消解项越少,指代消解在这一环节损失的召回率越低;Acc<sup>-</sup>表示的负例的准确率,这一准确率越高说明过滤的干扰信息越多,引入的噪声越少。

5.2 实验结果分析

表 6 给出了基于本文改进的规则的各类别待消解项识别分类器的性能。

表 6 基于规则的各类别待消解项识别分类器性能			
	代词/%	专有名词/%	有定名词/%
Acc <sup>+</sup>	100	99.07	100
Acc <sup>-</sup>	100	20.55	26.04

从表中可以看出:

1) 代词的正负准确率均达到 100%,通过观察语料发现交互式问答中的代词大部分都有所指。在新闻语料中占有非待消解项很大比重的“it”在交互式问答中出现很少非待消解项的情况,只有少数几例非待消解项并能被本文上述规则全部识别。这与新闻语料中代词的情况有很大不同,分析其原因在于,“it”在做系表结构时多为表达观点或陈述事实,

而交互式问答中多为问句,少有表达观点的陈述句。

2) 规则方法在专有名词和有定名词上都有很高的正例准确率。然而,规则方法在负例准确率上面表现非常差,所获得的正确率非常低。该实验结果表明规则方法能识别出大部分待消解项,但同时会引入较多的噪声,对非消解项的过滤能力极低。此种情况与在新闻语料上出现的情况十分类似。

表 7 给出了基于机器学习方法的各类待消解识别分类器性能。表中原特征是针对于新闻语料制定的特征集,新特征是加入了本文提出的两个新特征之后的特征集。

表 7 基于机器学习的各类别待消解项识别分类器性能

类别	专有名词(%)		有定名词(%)
	原特征	新特征	
Acc <sup>+</sup>	67.43	82.33	97.34
Acc <sup>-</sup>	62.23	85.38	70.98

从表 8 可以看出:

1) 原特征在专有名词上的正负例准确率都较低,说明原特征在交互式问答中的性能不是很好。交互式问答是以对话的形式出现,其中的表达不如新闻文本那么正式,对于一些结构名与人名则是能省则省,往往只用单独一个命名实体来表示。本文在总结了交互式问答的这个特点后在原特征基础上加入了两个在专有名词上的新特征:NEROUT 和 NNOUT(如表 4 中加黑部分所示)。加入这两个特征后正负例准确率都有很大幅度的提高,证明新特征十分有效。

2) 原有特征在有定名词上的正负例准确率都比较高。正例准确率能达到 98% 以上,说明识别出的待消解项比较全面,而且负例样例也有较好的效果,能过滤掉大部分非待消解项。本文又在原有特征上尝试加入了其他几类特征,但最后的性能并未有太大改变,由此说明在新闻语料上的有定名词特征集同样适用于交互式问答中。

6 小结与展望

本文将现有的关于普通文本的识别方法运用到交互式问答中,观察方法的适应性。利用启发式规则与基于特征方法相结合的方法构造二元分类器来识别待消解项。针对交互式问答任务的特点,本文在机器学习方法上面提出了新的特征。实验结果表

明已有的识别方法在交互式问答中取得了一定的效果,在加入本文提出的新特征后,负例准确率有明显的提升。虽然本文在交互式问答中待消解项识别的实验的效果取得了较好的效果,但由于语料的限制只研究了代词、有定名词、专有名词三种类别的识别研究。待消解项识别研究完成后将要进行下一步工作:指代消解,最后对于待消解项识别工作对在交互式问答中的指代消解的影响将是下一步研究的一项重要内容。

参考文献

[1] Nick Webb. Introduction of Interactive Question Answering Workshop[C]//Proceedings of the Interactive Question Answering Workshop at HLT-NAACL 2006, 2006.

[2] Ellen M Voorhees. Overview of the TREC 2004 Question Answering Track. <http://trec.nist.gov/>, 2004.

[3] Ng V, Cardie C. Identify anaphoric and non-anaphoric noun phrases to improve coreference resolution[C]//Proceedings of the 19th Int Conf on Computational Linguistics (COLING), 2002:976-984.

[4] Soon W M, Ng H T, Lim D. A machine learning approach of coreference resolution of noun phrase[J]. Computational Linguistics, 2001, 27(4): 521-544.

[5] Chai J, Jin R. Discourse structure for context question answering [C]//Proceedings of HLT-NAACL 2004 Workshop on Pragmatics of Question Answering, 2004:23-30.

[6] Carbonell J G. Discourse pragmatics and ellipsis resolution in task-oriented natural language interfaces [C]//Proceedings of 21st Annual Meeting on Association for Computational Linguistics, 1983:164-168.

[7] Nils D, Jonsson A. Empirical studies of discourse representations for natural language interfaces[C]//Proceedings of 4th Conference of the European Chapter of the ACL, 1989:291-298.

[8] Dongsheng Wang. Answering contextual questions based on ontologies and question templates[J]. Frontiers of Computer Science in China, 2011, 5(4): 405-418.

[9] Joyce Y Chai, Rong Jing. Discourse structure for context question answering [C]//Proceedings of the Workshop on Pragmatics of Question Answering at HLT-NAACL, 2004:23-30.

[10] Lappin S, Herbert J L. An algorithm for pronominal anaphora resolution [J]. Computational Linguistic, 1994, 20(4):535-561.