

学校编码: 10384

分类号_____密级_____

学号: X2012230254

UDC _____

厦 门 大 学

工 程 硕 士 学 位 论 文

情感智能聊天系统的设计与实现

Design and Implementation of Emotion Intelligent Chat
System

黄 钊

指 导 教 师 : 王美红 助理教授

专 业 名 称 : 软 件 工 程

论文提交日期 : 2014 年 4 月

论文答辩日期 : 2014 年 5 月

学位授予日期 : 年 月

指 导 教 师 : _____

答辩委员会主席 : _____

2014 年 4 月

厦门大学学位论文原创性声明

本人呈交的学位论文是本人在导师指导下,独立完成的研究成果。本人在论文写作中参考其他个人或集体已经发表的研究成果,均在文中以适当方式明确标明,并符合法律规范和《厦门大学研究生学术活动规范(试行)》。

另外,该学位论文为()课题(组)的研究成果,获得()课题(组)经费或实验室的资助,在()实验室完成。(请在以上括号内填写课题或课题组负责人或实验室名称,未有此项声明内容的,可以不作特别声明。)

声明人(签名):

年 月 日

厦门大学学位论文著作权使用声明

本人同意厦门大学根据《中华人民共和国学位条例暂行实施办法》等规定保留和使用此学位论文，并向主管部门或其指定机构送交学位论文（包括纸质版和电子版），允许学位论文进入厦门大学图书馆及其数据库被查阅、借阅。本人同意厦门大学将学位论文加入全国博士、硕士学位论文共建单位数据库进行检索，将学位论文的标题和摘要汇编出版，采用影印、缩印或者其它方式合理复制学位论文。

本学位论文属于：

（ ） 1. 经厦门大学保密委员会审查核定的保密学位论文，
于 年 月 日解密，解密后适用上述授权。

（ ） 2. 不保密，适用上述授权。

（请在以上相应括号内打“√”或填上相应内容。保密学位论文应是已经厦门大学保密委员会审定过的学位论文，未经厦门大学保密委员会审定的学位论文均为公开学位论文。此声明栏不填写的，默认为公开学位论文，均适用上述授权。）

声明人（签名）：

年 月 日

摘要

随着人类科学的不断进步，人类社会已经进入了智能生活时代，把以前繁琐的工作交给计算机，让人们拥有更多时间去提高和优化流程。这几年移动互联网的快速发展，手中的手机已经不仅仅只用于通话，如今拥有一部智能手机便可轻松掌握生活的方方面面，甚至手机已经变成了一些人不可缺少的电子器官。智能移动时代已经到来，我们可以突破家用计算机的地域限制，将一个智能助手随身携带，将互联网握在手中，通过智能手机，我们生活中的出行、购物、工作和餐饮等方方面面可以更加方便管理。而技术实现不在成为主要门槛的今天，设计者们更加注重用户的使用体验，良好的人机交互是吸引用户的要点之一，而智能聊天系统也在被广泛的应用在各个领域。比如 Apple 设备上的 Siri 功能，京东商城的问答系统，我们可以看到智能聊天系统省去了大量人工过程，并提供良好快速的反馈效果，同时智能聊天系统的发展也是计算机拟人化道路上重要的一步。

本文尝试用全新的视角在计算机中模拟人类情感系统，试图设计出一个可以理解人类情感的分析系统，并且可以反馈做出实际且科学的判断来提供一些正确的建议，可以达到“情感伴侣”的级别。并且最终将用户的情感以数据的形式记录在计算机中，从另外一种形式将人的思维存储在计算机中。本文介绍智能聊天系统的关键技术，如分词算法，机器学习和自然语言库等的基础上，分析、设计并实现了一款符合情感聊的智能聊天系统。

关键词：智能聊天；情感分析；情感反馈

Abstract

With the continuous progress of human science , human society has entered the era of intelligent life , the previously tedious job to a computer , so that people have more time to improve and optimize processes. In recent years the rapid development of mobile Internet , mobile phones already in the hands not just for the call, and now have a smart phone you can easily master all aspects of life , and even mobile phones have become indispensable in some electronic organs. Smart mobile era has arrived , we can break the geographical restrictions home computer , an intelligent assistant will carry the Internet in hand, through the smart phone , we live in travel, shopping, dining and other aspects of the work and can more easily manage . The technology does not become a major threshold today, designers pay more attention to the user experience , a good human-computer interaction is one of the points to attract users , but also intelligent chat system has been widely used in various fields. For example, Apple Siri functionality on the device , Jingdong Mall 's answering system , we can see a lot of intelligent chat system eliminates manual processes , and provide good quick feedback effect, while the development of intelligent chat system is also computer anthropomorphic important one step.

This dissertation attempts to use the new perspective of human emotions in a computer simulation system, trying to create an understandable human emotion analysis system, and can be made practical and scientific feedback to determine the right to provide some suggestions, you can reach "emotional partner" level. And finally the user's emotions as data recorded in the computer, from another form of the human mind stored in the computer. This dissertation will develop intelligent chat system, principles and related technologies to understand and analyze, for example, segmentation algorithms, machine learning , natural language libraries, and ultimately design a chat line with emotional intelligence chat system.

Keywords : Intelligent-chat; Sentiment Analysis; Emotional Feedback

目录

第 1 章 绪论	1
1.1 课题背景	1
1.2 智能聊天系统概述	1
1.3 智能聊天系统的发展状况	2
1.3.1 智能聊天系统的发展历史	2
1.3.2 国内外研究现状	3
1.3.3 智能聊天系统发展总结	4
1.4 本文研究内容与结构	6
第 2 章 相关技术介绍	7
2.1 人工智能	7
2.1.1 什么是人工智能	7
2.1.2 人工智能的哲学争论	8
2.2 机器学习	9
2.2.2 相关学科及它们对机器学习的影响	9
2.2.3 搜索匹配	10
2.3 服务端架构	11
2.3.1 Apache	12
2.3.2 PHP	13
2.3.3 开发框架 KOHANA	13
2.3.4 IDE	14
2.3.5 数据库	14
2.3.6 服务端接口测试工具	15
2.4 移动端	16
2.5 技术优化和扩展	18
2.6 本章小结	20
第 3 章 系统需求分析	22
3.1 现有类似系统分析	22

3.1.1 什么是 SimSimi.....	22
3.1.2 SimSimi 的基本原理.....	23
3.1.3 如何自制 SimSimi.....	28
3.1.4 从 SimSimi 身上学到什么.....	28
3.2 需求分析.....	28
第 4 章 系统设计.....	31
4.1 系统功能模块设计.....	31
4.2 分词子系统设计.....	33
4.2.1 中文分词算法.....	33
4.2.2 分词异常情况说明.....	34
4.3 节点管理子系统设计.....	35
4.4 数据库设计.....	36
4.5 简单语言系统难点分析.....	37
4.6 本章小结.....	38
第 5 章 系统实现.....	40
5.1 搭建系统服务端.....	40
5.2 分词的实现.....	42
5.3 交互界面实现.....	48
5.4 本章小结.....	51
第 6 章 总结和展望.....	52
6.1 总结.....	52
6.2 展望.....	52
参考文献.....	54
致谢.....	56

Contents

Chapter 1 Introduction	1
1.1 Background Topic	1
1.2 Overview of an Intelligent Chat System	1
1.3 Development of Intelligent Chat System	2
1.3.1 Historical Development of Intelligent Chat System	2
1.3.2 the Research Status	3
1.3.3 Summary	3
1.4 Contents and Structure	4
Chapter 2 Overview of the Related Technologies	7
2.1 Artificial Intelligence	7
2.1.1 Artificial Intelligence	7
2.1.2 Artificial Intelligence Philosophical Debate	8
2.2 Machine Learning	9
2.2.2 Related Disciplines and Their Impact on Machine Learning	9
2.2.3 Search Match	10
2.3 Server Architecture.....	11
2.3.1 Apache	12
2.3.2 PHP	13
2.3.3 Development Framework KOHANA	13
2.3.4 IDE	14
2.3.5 Database	14
2.3.6 Server Interface Testing Tool	15
2.4 Mobile Terminal	16
2.5 Technology to optimize and expand	18
2.6 Summary	20
Chapter 3 System Requirements Analysis	22
3.1 Analysis SimSimi System	22
3.1.1 SimSimi	22
3.1.2 the Basic Principle of SimSimi	23
3.1.3 How to Homemade SimSimi	28
3.1.4 SimSimi Analysis Summary	28
3.2 Requirements Analysis	28
3.3 Summary	29

Chapter 4 System Design	31
4.1 The Design of System Function Module	31
4.2 Segmentation Subsystem Design	33
4.2.1 Chinese word segmentation algorithm	33
4.2.2 Description Word Anomalies	34
4.3 Node Managerfeatures Subsystem Design.....	35
4.4 Database design.....	36
4.5 Simple Language System Initially Conceived	37
4.6 Summary	38
Chapter 5 System Implementation	40
5.1 Build System Server	40
5.2 Achieve Pairs of Words	42
5.3 Interface to Achieve	48
5.4 Summary	51
Chapter 6 Conclusions and Prospects.....	52
6.1 Conclusions	52
6.2 Prospectd	52
References	54
Acknowledgements	56

第1章 绪论

1.1 课题背景

随着网络时代的到来,其高速数据传输的蜂窝移动通信技术使人们利用手机去获取复杂的网络信息成为可能,而越来越多的年轻人和商务人士都乐意去使用这一便捷的信息获取方式。智能手机的发展也赢得了手机行业的辉煌十年,从当年黑白屏幕超大个头只能打电话的“专用产品”到现在想要囊括人们生活方方面面的“贴身管家”,手机已经成为人们工作生活中必不可少的一部分,而且也在悄悄的改变人们的生活方式。

迄今,中国的上网人群已使它成为全球互联网人口第二,仅次于美国。随着电子书刊和各大门户网站的用户群体扩大,互联网信息更新的及时性互联网成为我们获取信息的主要渠道。智能聊天系统给我们提供了便捷的信息入口和交互平台,我们可以快捷了解周边的天气交通和饮食情况,当然我们每天在网上浏览各种信息,高速的互联网管道将全世界相互连接,这也确定了未来电子化工作和生活的必要性和重要性,或许在未来,我们不用走出家门便可以完成工作,我们根据计算机安排的方式生活,让计算机来处理繁琐无谓的工作,让人有时间去做更伟大的事情。

不可否认,智能手机给移动互联网甚至人们的生活方式的变化和改善都是毋庸置疑的,也是标志着人类进入全新科技时代的开始,我们可以不断改变生活方式来迎合世界的变化。同时科技也是正在迎合我们的生活习惯。在未来,我们可以过上全智能化的生活,通过智能聊天系统,计算机便可以帮助我们分析解决大部分日常工作,甚至我们可以和计算机做朋友。

1.2 智能聊天系统概述

信息技术的发展,大量即时交易平台以聊天的形式出现在人们的面前。例如京东的自助式客服查询系统,Apple 公司提供的 Siri 拟人问答系统等不仅减少了人力资源的消耗也在便捷效率上都大大提高的服务质量。这些智能聊天系统的

产生都是人们为了满足人们快速增长的物质文化的信息化需求,极度需要自动化工具实时的帮助人们在日常生活学习中快速得到所需的信息。智能聊天系统就是在这样的环境下产生的。聊天智能机器人简单来说就是一类像真人陪同用户一起聊天,帮助用户查询信息,为用户提供大部分整合网络服务的具备高度的人工智能信息互动平台。从技术原理上看,它属于基于语料库的专家信息系统。它的聊天功能主要指在聊天室或论坛上根据人们提出的不同问题,通过分析聊天语句并做出自动相应的应答的一种问答信息系统;能够模拟人和人对话,具有人性的特质,可以模仿人与人之间对话的智能语言系统;是利用人工智能技术开发的,目的在于和人交流对话的计算机执行程序。它可以不受时间、空间等因素限制,不知疲倦地与人进行聊天,不仅供人娱乐消遣,也可以在一定意思上帮助进行语言的学习的计算机执行程序。总所周知,日常生活中语言具有歧义性、模糊性和不确定性,对计算机来说,很难像现实中那样对语句进行感性的理解,因此智能聊天系统对接收到的语句进行分析和判断并做出信息反馈,回答不同的提问并进一步提出问题。具有自我学习能力的智能聊天机器人不仅能回答提问,还能使用数据库录入新的对话内容,从中解析出关键词汇进行记忆,自动充实自身的语料库,使聊天的过程更加真实^[1]。

1.3 智能聊天系统的发展状况

1.3.1 智能聊天系统的发展历史

一切自 Joseph Weizenbaum 于 60 年代在麻省理工学院 (MIT) 开发 Eliza 开始,它被认为是第一个聊天机器人。Eliza 最有名的程序是 DOCTOR script,它模仿在最初的精神病面谈中非定向精神治疗师的答复。具有讽刺意味的是,即使产生于人们身上的情感反应使 Weizenbaum 相信系统有极大价值,但他从未把它看作是智能系统。

6 年后,即 1972 年,精神病医生 Kenneth Colby 在斯坦福大学 (Stanford University) 创造了 PARRY。它的基本原理与 Eliza 一样,但是 Parry 模拟的是妄想型精神分裂症患者,而不是心理治疗师。即使 Colby 做出了更多的努力,其结果是完全相似的。

1978, Michael Mauldin 是试图将推理引入聊天机器人的第一人。他创造的 PET 聊天机器人能设想新信息, 并由于如下对话而闻名。

Subject: 我喜欢我的朋友。(later)

Subject: 我喜欢食物。

PET: 我听说食物是你的朋友。

随后, Mauldin 创造了虚拟人物(即现在的 Conversive)和两个有名的聊天机器人: Sylvie (1994) 和 Julia (1997)。除了加入动画和合成语音, 由 Maulding 和 Peter Plantec (另一个创造者)提供的重要方面是有能力探索收集一个虚拟世界的信息(如: 网站), 并在对话中使用收集的信息。

聊天机器人历史中另一个重要人物是 Richard Wallace, A.L.I.C.E (人工语言互联网计算机实体)的创始人。Wallace 采取不同的方法, 使 Alice 三次获得罗纳奖(2000 年, 2001 年, 2004 年)。他回归到一种纯模式匹配的方法, 为具体说明启发式的会话规则, 他创造了一个名为 AIML (人工智能标记语言)的 XML Schema 语言。这种方法的优势在于能容易地在一个 AIML 文档中创建和共享知识, 以及把多个 AIML 文档加载到一起组成一个“更聪明的”机器人。

所有付诸于基于模式匹配的聊天机器人的工作都是对时间和金钱的巨大浪费(我们将看到, 浪费将达数亿美元)。你不相信的话就试用一下 Eliza, 并将其与 2008 年的罗纳奖获得者 Elbot 相比较, 然后再来判定这 50 年的努力是否值得。我们探究的领域正确吗^[2]?

1.3.2 国内外研究现状

随着智能手机的普及, 给智能聊天系统提供了更广阔的发展平台, 而且智能聊天的出现, 也符合目前移动互联网和智能生活的需求。近几年国内外不断涌现出许多智能聊天系统。

Siri 是苹果公司在其产品 iPhone 上应用的一项语音控制功能。Siri 让一台 iOS 设备变为智能机器人, Siri 用户可以通过移动设备了解美食、读取讯息信、接收天气服务、设置闹钟等。Siri 能够支持解析口语输入, 并且和 iOS 系统自身多项功能紧密的联系起来, 还能不断调整和学习新的语调和语法, 提供口语形式的交互体验。Siri 源于 2007 年美国国防部高级研究规划局所的 CALO 计划: 可是使

军方简化处理一些繁复杂务，并具有自我学习以及认知能力的智能助理，最后发展来的平民版 Siri 智能语音助理。苹果在 2010 年以 2 亿美金收购 Siri，最初是以文字聊天服务为主，随后通过与全球最大的语音识别厂商 Nuance 合作，Siri 实现了语音识别功能。使用者可以通过声控、文字输入的方式，来搜寻餐厅、电影院等生活信息，同时也可以直接收看各项相关评论，甚至是直接订位、订票；另外其适地性服务的能力也相当强悍，能够依据用户默认的居家地址或是所在位置来判断、过滤搜寻的结果。不过其最大的特色，则是人机的互动方面，不仅有十分生动的对话接口，其针对用户询问所给予的回答，也不至于答非所问，有时候更是让人有种心有灵犀的惊喜，例如使用者如果在说出、输入的内容包括了“喝了点”、“家”这些字，Siri 则会判断为喝醉酒、要回家，并自动建议是否要帮忙叫出租车。Siri 的后台技术也是相当的先进，其中包括以 Google 为代表的网页搜索技术；以 Wolfram Alpha 为代表的知识搜索技术；以 Wikipedia 为代表的知识库技术；以 Yelp 为代表的问答以及推荐技术。

目前国内关于智能聊天系统的开发和相关技术发展也日趋成熟，也已经出现了不少智能聊天机器人，比如赢思软件的小 i，爱博的小 A，小强，和爱情玩偶等等。这些机器人也已经日益成为网民上网的好伙伴。赢思软件推出的小 i 还有很多丰富的功能，比如 msn 群，让办公室白领能够更加轻松的交流。

小 i 是由国内赢思公司在 2001 年研究开发的，也是现今全球用户量最大的中文聊天机器人，从最早的 MSN 上出现，后拓展到 QQ、短信、网页等交互平台上，拥有超过 1 亿的用户。2012 年 2 月，小 i 机器人整合了自然语音识别功能，并陆续发布了智能手机、智能电视以及电话上的语音交互版本。其中关键技术有自然语言处理，语义分析和理解，知识构建和学习体系，语音技术，整合通信技术，云计算基础技术等。可见一款全面智能的中文智能对话系统是需要靠不断的积累和反复科研才能逐渐成形的。

还有之前微软亚洲研究院推出的读心机器人，可以通过用户询问问题回答出。

1.3.3 智能聊天系统发展总结

观看以往的智能聊天系统多是由于商务系统，人们更关注它即时性回答问题

的能力,所以开发者和设计者更偏向将精力集中在搜索匹配问题的算法和集成语料库上,为匹配上最佳答案而设计,将学习能力居于次位,这样智能聊天系统更偏向的是机器而不是人,相当于一个网络搜索引擎,负责网络信息的查询和处理,只是将回答进行人类语言的润色,更符合人际交流的习惯。没有长时记忆体的功能,只能对一句话做出快速的反应,没有连贯的逻辑能力和思维能力。而本质上也根本不具备思考和联想的能力。

系统设计基于情感的管理系统,侧重点却与之不同。系统设计的重点并非“机智聪明”,“精通万物”,目标需求是能模拟人类之间的情感对话,具有长时记忆体的功能,用带有“温度”的交互,用户可以和它分享快乐,发泄不快,而且它不仅仅是个最佳的聆听者,它也会欢笑,鼓励或者给予对方肯定,用完全模拟人类情感的方式去实现这个管理系统,所以这套系统是不断进行数据积累,是通过不断的分析和了解用户的情感状态而随之改变的,所以每一个用户都会有一个专属于自己的“Robot Friend”,它只属于也只了解这个用户的感情状态,它知道用户的喜好,它了解用户的习惯,它不需要精通天文地理,它只需“关心”用户的一个模拟人类情感的系统。

在系统实现的初期,的知识、语料库还十分匮乏,当我们与其进行对话,大量的未知名词需要进行归类 and 定义,这时候,我们需要提供一套分析算法,对当前问题的处理和反馈。接下对算法的基础模型进行描述。

举例一:我喜欢吃西瓜。

通过分词系统我们可以得到:

我/喜欢/吃/西瓜。

首先获得人物[我],其属于人物主库(Host)中,

其次获得情感动词[喜欢]和[吃],分别是情感库(Emotion)和行为库(Action)。

还有对象[西瓜],放入事物库中(Object)中。

然后将其关联起来。

举例二:今天我吃了个西瓜。

带有日期时间的,我们要将其加入行为库(Event)中。并记录时间。

事件库包含的字段有:时间,地点,人物,事件,对象,情感等。

1.4 本文研究内容与结构

本文将对智能聊天系统的发展历史,涉及的相关学科和理论进行深入的调查和研究,针对人工智能、机器学习和系统实现上设计的各方面技术进行理解和实践,通过分析现有成熟的智能聊天系统并结合自身系统特点制定出系统需求,最终完成系统设计。

本文主要分为绪论、相关技术介绍、系统需求分析、系统设计和系统实现 5 个组成部分。

第一章绪论。介绍智能聊天系统的概念和发展历史。

第二章相关技术介绍。介绍智能聊天系统涉及的相关技术原理和实现。

第三章系统需求分析。介绍通过分析现有的智能聊天系统,并结合自身系统特点进行需求分析。

第四章系统设计。介绍在需求的基础上定义和划分系统模块,并设计出实际可用的系统。

第五章系统实现。介绍通过系统设计,结合现有成熟的技术实现系统的功能模块。

第2章 相关技术介绍

本章将对智能聊天系统涉及到的相关技术和原理进行调查和分析,全面地对智能聊天系统的概念定义、发展历史、系统结构和实现技术进行深度的剖析,最终对智能聊天系统有全面的认识 and 正确的理解,对系统需求、设计打下坚实的基础。

2.1 什么是人工智能

现今时代人工智能对于我们来说已经不再像从前那样蒙着一层神秘的面纱,它涉及到生活的方方面面。

2.1.1 介绍人工智能

人工智能的一个目前比较流行的定义,也是较早该领域的定义是由美国麻省理工学院的约翰·麦卡锡在1956年的达特茅斯会议上提出,人工智能是要让机器行为看起来就像人所体现出的智能行为一样。但是这定义似乎忽略强人工智能的可能性。另一个定义指人造机器所表现出的智能行为。

总而言之,对人工智能的定义目前可划分为四个类别。

- ① 像人一样的思考;
- ② 像人一样的行动;
- ③ 理性地思考;
- ④ 理性地行动。

这里“行动”在广义地解释应该是采取动作或者决定行动的决策,而不是身体的动作。强人工智能的观点认为是有可能会制造出能够推理和解决实际问题的智能类机器,并且这类型机器能被视为是有知觉,有自我意识。强人工智能可以分为两类:

- ① 类人人工智能,是指像人类的思维一样思考和推理的机器;
- ② 非类人人工智能,是指机器产生与人完全不等的意识和知觉,使用与人

Degree papers are in the "[Xiamen University Electronic Theses and Dissertations Database](#)". Full texts are available in the following ways:

1. If your library is a CALIS member libraries, please log on <http://etd.calis.edu.cn/> and submit requests online, or consult the interlibrary loan department in your library.
2. For users of non-CALIS member libraries, please mail to etd@xmu.edu.cn for delivery details.

厦门大学博硕士论文摘要库