

# 基于网络行为分析的自助调节系统的设计与实现

作者姓名： 聂栋

校内指导教师： 于瑞云 博士

企业指导教师： 朱廷劭 教授

单位名称： 软件学院

专业名称： 软件工程

东 北 大 学  
2011 年 5 月

# **Design and Implementation of Self-adjustment System Network Based on Analysis for Network Behavior**

by Nie Dong

Supervisor: Doctor Yu RuiYun  
Associate Supervisor: Professor Zhu Tingshao

Northeastern University  
June 2011



# 基于网络内容分析的自助调节系统的设计与实现

## 摘 要

随着经济的发展，现代人的生活压力与日俱增，频发的心理健康问题正逐渐成为全社会的关注重点。在开展心理健康干预工作时，作为一个关键性前提，如何准确地、及时地鉴别出具有特定心理健康问题的目标人群将会直接影响到心理调节疏导的针对性与有效性。

互联网的高速发展，已逐渐深入到人们的日常生活，改变着人们之间的既有沟通方式。因此，基于人际互动过程的传统心理健康评估模式已经不能满足人们当前的需求，互联网的发展为人们解决心理健康监控提供了新的手段。

国内外学者研究发现，用户的网络使用行为与心理健康状态之间存在着一定的联系。基于此，我们提出一种基于网络用户行为分析的心理健康评估新模式。该模式分析用户当前的网络使用行为，利用心理健康行为相关性模型进行预测，计算用户心理健康各个维度的指标，实现对用户心理健康状况的实时感知。我们以网络使用行为作为预测变量，以不同种类的心理健康问题作为观测变量，建立一系列回归模型，并给出恰当的干预调节建议。

我们建立了一个 WebMind 系统为实验平台。利用训练获得的网络使用行为与心理健康的相关性模型，WebMind 系统实现了对用户的网络使用行为的跟踪记录、心理健康状态的实时预测以及干预调节建议的推荐。通过两周时间的用户实验表明，70.3%的干预建议对用户是有帮助的。证明我们的系统能够准确地预测用户心理健康状态并给出有效的干预建议。

**关键词：**心理健康预测，网络行为分析，线性回归，相关性分析，WebMind，心理健康调节

# **Design and Implementation of Self-Adjustment System Network Based on Analysis for Network Behavior**

## **Abstract**

With the development of economic and increasing pressure of modern life, frequent mental health problems are becoming a focus for the whole society. As a critical prerequisite to carrying out the work of mental health interventions, accurately and timely identifying mental health problems with a specific target population will directly affect the relevance of psychological adjustment to ease and effectiveness.

The rapid development of the Internet has gradually penetrated into people's daily lives, changing the existing means of communication between people. Therefore, the existing mental health assessment process model which based on the traditional interaction can no longer meet the needs of the people in current society. The latest research of the Internet will also address people to provide a new means of control to solve the mental health problems.

Scholars study found, there is some connection between the user's network usage behavior and mental health. Based on this, we propose a new model to predict mental health using the analysis of the user's network behavior. The model analysis the user behavior of the current network usage, and predict mental health problems through matching mental health behaviors using model. The model calculates mental health indicators in all dimensions, in order to achieve real-time user perception of mental health. We use the behavior of the network as a predictor variable, with different types of mental health problems as observed variables, the establishment of a series of regression models, and give a certain degree of adjustment proposals.

We established a WebMind system as the experimental platform. Using the behavior and mental health of models trained by the obtained Network data, the WebMind system achieve the behavior of the user's track record of network usage, mental health intervention in real-time forecasts and adjust the proposed recommendation. Through two weeks of the experiment, the result showed that 70.3% of the intervention proposed is helpful to the users. It proved that our system can accurately predict the state of mental health users and give recommendations of

effective interventions.

**Key words:** Predict Mental Health   Analysis Network Behavior   Linear regression  
Correlation Analysis   WebMind system



# 目录

DOCTOR YU RUIYUN .....	2
NORTHEASTERN UNIVERSITY .....	2
摘 要 .....	I
DESIGN AND IMPLEMENTATION OF SELF-ADJUSTMENT SYSTEM NETWORK BASED ON ANALYSIS FOR NETWORK BEHAVIOR .....	2
ABSTRACT .....	2
第 1 章 绪论 .....	7
1.1 课题背景、目的与意义 .....	7
1.1.1 目前社会上心理问题越来越严重 .....	7
1.1.2 互联网在中国迅速发展 .....	7
1.1.3 基于网络行为分析的自助式心理调节系统的意义 .....	7
1.2 国内外现状 .....	8
1.3 应用范围 .....	9
第 2 章 关键技术介绍 .....	11
2.1 关键性分析技术介绍 .....	11
2.1.1 分析过程 .....	11
2.1.2 分析结果 .....	12
2.1.3 心理健康网络行为关系模型 .....	14
2.2 关键性开发技术介绍 .....	19
2.2.1 .Net Framework 开发平台以及开发语言 C#介绍 .....	20
2.2.2 朴素贝叶斯文本分类理论介绍 .....	20
2.2.3 ZedGraph 控件介绍 .....	21
第 3 章 系统分析 .....	22
3.1 构架概述 .....	22
3.1.1 模块需求描述 .....	22
3.2 系统开发环境 .....	22
3.3 系统任务的可行性分析 .....	23
3.3.1 技术可行性 .....	23
3.3.2 系统安全性分析 .....	23
3.3.3 系统效率分析 .....	23
第 4 章 系统设计 .....	24
4.1 设计指导思想和原则 .....	24
4.1.1 指导思想 .....	24
4.1.2 软件设计原则 .....	24
4.2 系统结构设计 .....	24
4.3 系统流程设计 .....	25



4.4 系统模块设计 .....	26
4.4.1 浏览器模块详细设计: .....	26
4.4.2 绘图控件显示模块 .....	26
4.4.3 网络内容分析模块 .....	27
4.4.4 心理建议推荐模块 .....	29
4.5 系统数据库(历史网络行为库)设计 .....	29
4.5.1 网页浏览历史数据库设计 .....	30
4.5.2 浏览器操作数据库设计 .....	31
<b>第 5 章 系统实现 .....</b>	<b>33</b>
5.1 WebMind 系统框架 .....	33
5.2 浏览器模块 .....	33
5.2.1 浏览器基本功能 .....	33
5.2.2 记录浏览历史 .....	34
5.2.3 心理指标追踪 .....	34
5.2.4 心理健康建议 .....	35
5.3 历史网络行为库 .....	36
5.3.1 网页浏览历史 .....	36
5.3.2 浏览器操作历史 .....	36
5.4 用户网络行为分析 .....	36
5.4.1 访问时间频度统计 .....	36
5.4.2 URL 域名解析 .....	37
5.4.3 网页内容分析 .....	37
5.4.4 网页情感分析 .....	38
5.4.5 统计结果 .....	38
5.5 心理健康行为模型 .....	39
5.6 行为心理模型的自我更新 .....	40
5.7 实验结果及其分析 .....	42
5.7.1 实验步骤 .....	42
5.7.2 实验结果与分析 .....	42
<b>第 6 章 系统测试 .....</b>	<b>45</b>
6.1 测试方案及测试用例 .....	45
6.1.1 测试方案 .....	Error! Bookmark not defined.
<b>图 6.1 用户对心理建议推荐模型的评价 .....</b>	<b>47</b>
6.1.2 测试结果反馈 .....	47
<b>第 7 章 结论与展望 .....</b>	<b>48</b>
7.1 论文已完成的工作 .....	49
7.2 下一步研究工作 .....	49
<b>参考文献 .....</b>	<b>50</b>
<b>致 谢 .....</b>	<b>52</b>

# 第1章 绪论

## 1.1 课题背景、目的与意义

### 1.1.1 目前社会上心理问题越来越严重

现代社会竞争日益激烈，人们的心理早已不堪重负。失眠、抑郁、焦虑、紧张、压抑等一系列心理亚健康健康问题已严重影响人们的正常生活[1][2]。

据调查数据，中国存在 2600 万抑郁症患者，3000 万 17 岁以下的青少年有各种心理问题<sup>[3]</sup>，每年自杀成功者 28 万，自杀未遂的人群达到 250 万，各种神经症患者占人群的 3-5%，40-60%内科门诊的患者存在心理问题，尤其是白领群体。他们处于事业上升期、家庭组建初期，孩子早期教育关键阶段，因而在经济、生活方面自然承担着更多的压力，要付出更多的体力和精力，进而形成了 30—40%的白领人群存在严重的心理健康问题<sup>[4]</sup>。

我国人口基数庞大，决定了我国心理行业的发展前景十分广阔，但受传统文化的影响，我国民众思想比较保守，出现心理问题时，不愿求助心理专业人员的咨询服务；加之心理咨询服务市场良莠不齐、缺乏有效的监管机制，增加了一部分求助者的顾虑；另外心理咨询高昂的服务费，也让一部分人对心理咨询望而止步。我国心理行业发展滞后，心理咨询、心理调节途径缺乏，这就造成了在我国心理健康问题很严重，心理行业发展市场广阔，而相应的解决途径不足的矛盾。

### 1.1.2 互联网在中国迅速发展

2011 年 5 月，中国互联网络信息中心（CNNIC）发布了《第 27 次中国互联网络发展状况统计报告》。数据显示，截止 2011 年 3 月，我国网民规模已达 4.77 亿；手机网民达 2.8 亿。

互联网的高速发展，提供了一种获取信息、共享资源的革命性新途径，打破了信息流通和共享的时间限制与地域限制，实现了“时空压缩”，从而使人类真正进入了一个双向互动、平衡交换的理性状态。网络不仅仅从根本上改变着现有的生产结构、产业结构、劳动结构，而且也极大地影响着人们的生活方式、交往方式、工作方式、学习方式乃至思维方式，并导致人的价值观和伦理观的深刻变革，一种新型的社会——网络虚拟社会[5]应运而生。

### 1.1.3 基于网络行为分析的自助式心理调节系统的意义

近年来因心理健康问题引发的人身伤害事件不断发生，究其根本原因是心理

健康问题没有被及时发现，错过了最佳的处理时机。常用的心理健康监测机制主要靠定期心理普测、心理咨询室和心理咨询热线、亲人反映等途径来获得心理状况信息。这些途径都是被动途径，很难及早发现负面情绪和潜在的心理危机，错过了心理干预和调节的最佳时机。

但是人们的学习压力、经济压力、情感压力大，公共生活范围狭小，且极少主动寻求心理援助，这就进一步造成了人们的心理危机难以被心理辅导人员及时掌握。

心理危机难以及时进行监控，已成为现在心理辅导和心理调节的难点之一。因此，探索有效的心理健康的评估途径，成为当前心理学的一个重要课题。

基于此，我们面对的问题就是如何实现对心理健康的实时感知，也就是如何评价心理健康水平和心理健康状况。这需要建立一个评价标准，用于识别个体的心理是处于健康状态还是疾病状态，了解个体究竟患了何种心理疾病，出现何种心理困扰以及心理健康水平达到了什么程度。要制定这个评价标准，就要对心理健康的实质结构进行深入的科学研究，采用量化的方式对心理健康状况进行描述。有了评价标准，才能够对人的心理健康状况和心理问题进行比较、评估和分类，并根据不同的情况，采用危机干预、病患治疗、辅导咨询、困扰探究、社会教育、习性培养、自我发展等不同层次的途径解决问题，并对各种影响因素进行预防和控制，从根本上减少心理病患的产生，从而达到提高人类心理健康水平的目的。

作为发生在虚拟网络社会上的一种特殊的、有意识的社会活动，网络行为具有社会行为的一般特征和基本要素。通过研究网络用户行为的规律性，借以监控用户的心理健康状况变化，为心理危机的干预和心理调节提供良好的时机。

## 1.2 国内外现状

国外较早的展开用计算机技术解决心理方面问题的研究，到现在已经有一定规模和效用，而国内在利用计算机技术心里自助调节方面的研究目前还不是很多。

在国外，计算机辅助治疗、远程治疗、虚拟现实暴露疗法是计算机和网络心理治疗的三大特色[6]。它们在治疗焦虑、抑郁、恐怖症、外伤后应激障碍、倦怠和工作压力以及其它障碍方面具有显著效果。具体来看，国外主要着眼于行为认知疗法和接受与实现疗法的研究，而且以网站的形式呈现给用户，从而来帮助用户来治疗抑郁、焦虑等心理问题。认知行为疗法是一组通过改变思维或信念和行为的方法来改变不良认知，达到消除不良情绪和行为的短程心理治疗方法。具

有代表性的有埃利斯的合理情绪行为疗法(REBT)，贝克和雷米的认知疗法(CT)和梅肯鲍姆的认知行为矫正技术等。在美国心理学界，斯蒂文·海耶斯(Steven Hayes)开创了一种治疗心理和精神疾病的新疗法——“接受与实现疗法”

(acceptance and commitment therapy, 简称 ACT)，成为继行为疗法、认知疗法后，美国兴起的第三波心理疗法。这种新疗法主张拥抱痛苦，接受“幸福不是人生的常态”这一现实，然后再建立和实现自己的价值观。ACT 疗法在治疗抑郁症、上瘾症、癫痫病等精神类疾病方面都取得了不俗的成绩。目前，美国已有 1.2 万名心理专业人士接受了 ACT 疗法的培训，海斯的理论在 18 个国家都有追随者。

国内的利用计算机技术来支持心理学方面的研究起步较晚，从 2000 年开始才慢慢的出现了虚拟现实的心理治疗方式[7]。尽管近 20 多年来，心理学在我国迅速发展，全国很多高校相继成立了心理学系和心理学专业，但计算机技术在心理学方面的应用，与国外同行对比，还是显得落后许多。像北京大学心理系、中科院心理研究所等著名心理研究中心，在心理学研究方面都很少结合计算机技术。而用浏览器系统来跟踪用户的网络行为，并结合心理-行为模型分析出用户的心理状态，对于有心理问题的人们，给出相应心理康复服务，在国内应该是还没有人做过。

### 1.3 应用范围

通过广泛参考各种文献和借鉴国内外的研究成果[1][8]，实验室的老师、师兄们阐述了网络用户及其行为的定义与分类，由此提出了网络用户行为分析的体系框架。他们通过对中科院研究生院 400 多位学生的心理健康数据的分析，构建一个用户网络行为与心理健康的相关性模型。在他们的帮助下，我依据此模型建立起了 WebMind 系统。

WebMind 系统是一个基于 IE8 内核的、具备收集网络用户网络行为并对网络行为进行统计分析功能和推荐个性化心理治疗服务的自助调节系统。WebMind 系统应用范围很广泛，其硬件要求较小，只需要能够上网便可，而且 WebMind 系统是免费的，因此那些想要测试、调整自己心理的人们都可以上网下载 WebMind 发行版。

具体的，我认为 WebMind 能起到作用的人群应该是城市的诸多工作人员，尤其白领阶层，由于他们有一定的物质基础和文化基础，基本都会上网、都能上网，而他们工作压力很大，稍有不慎，很容易产生一些心理波动问题。近来，有不少过劳死的现象，其本质应该是疏于防范自己的心理问题和身体，所以，这些

人群应该对自己的心理和生理多加注意。心是一个人的灵魂，心理出现问题，其危害远远大于一般的身体健康问题，这促使心理咨询行业的前进。而心理咨询行业的高收费和小范围，让很多人望而却步，在这样的背景下，我相信，WebMind 系统就有其作为的一面。

## 第 2 章 关键技术介绍

### 2.1 关键性分析技术介绍

这里，要介绍的是心理健康网络行为模型[9][10]，整个 WebMind 系统的核心工作都是围绕着这个模型来开展的。

#### 2.1.1 分析过程

利用 SPSS 工具进行分析

1)在 SPSS 中录入数据（变量为 Y 和 X），并保存数据文件；在主菜单中选择[Analyze]=>[Regression]=>[Binary Logistic](如下图所示)。

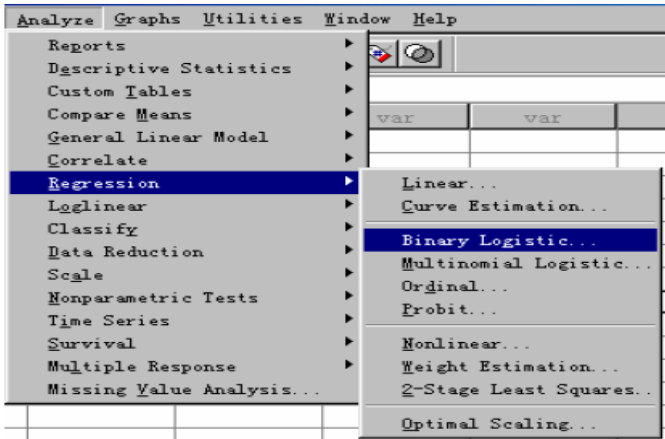


图 2.1 SPSS 分析过程：选择分析方法

2)在[Logistic Regression]对话框中，选择 Y 进入[Dependent]框作为因变量，选择 X 进入[Covariates]作为自变量。单击[Method]的下拉菜单，SPSS 提供了 7 种方法：

- ◆ [Enter]：所有自变量强制进入回归方程；
- ◆ [Forward: Conditional]：以假定参数为基础作似然比检验，向前逐步选择自变量；
- ◆ [Forward: LR]：以最大局部似然为基础作似然比检验，向前逐步选择自变量；
- ◆ [Forward: Wald]：作Wald 概率统计法，向前逐步选择自变量；
- ◆ [Backward: Conditional]：以假定参数为基础作似然比检验，向后逐步选择自变量；
- ◆ [Backward: LR]：以最大局部似然为基础作似然比检验，向后逐步选择自变量；
- ◆ [Backward: Wald]：作 Wald 概率统计法，向后逐步选择自变量。

在本试验中我们选择默认项[Enter]方法。

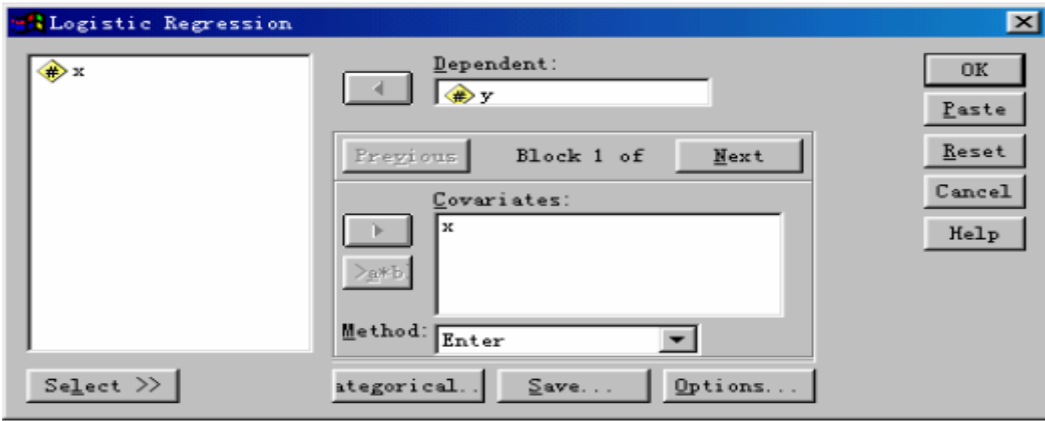


图 2.2 SPSS 分析过程：设置因变量和自变量

3)单击[Logistic Regression]对话框中的[Options]按钮，在显示的子对话框中选择 [Classification plots] 和 [Hosmer-Lemeshow goodness-of-fit] 等选项，并单击 [Continue]返回对话框。

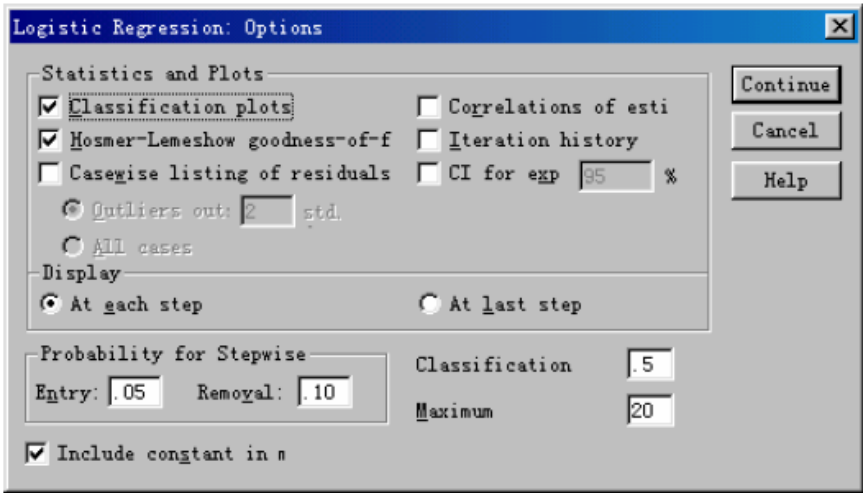


图 2.3 SPSS 分析过程：设置分析参数

4)单击主对话框中[OK]按钮，即可输出分析结果。

2.1.2 分析结果

以 PHI 问卷上各心理健康问题维度的高低组二分类（高分组的 27% 和 低分组 27%）变量为观测变量，一系列网络使用行为变量与人口统计学变量为预测变量，进行 Logistic 回归分析（ $\alpha = 0.05$ ）。结果，由于被试者在“抑郁”维度高低得分组中同时存在着大量彼此重叠的得分，使得高低得分组之间缺乏足够的区分度，因此共建立起 6 个回归模型。

表格 2.1 Logistic 回归分析的结果

Model	维度	$\chi^2$	df	p	总体正确率	击中率	正确拒绝率
1	躯体化	65.874	6	< .01	72.9%	64.6%	80.3%
2	焦虑	101.062	10	< .01	78.4%	73.3%	82.6%

3	病态人格	105.594	12	< .01	79.0%	72.8%	84.1%
4	疑心	94.420	10	< .01	77.9%	64.6%	88.6%
5	脱离现实	122.106	14	< .01	83.1%	80.0%	85.6%
6	兴奋状态	102.629	15	< .01	77.1%	72.4%	81.5%

其中， $\chi^2$  表示

$df$  表示自由度。在统计模型中，自由度指样本中可以自由变动的独立不相关的变量的个数，当有约束条件时，自由度减少。

自由度计算公式：自由度=样本个数-样本数据受约束条件的个数，即  $df = n - k$  ( $df$  自由度， $n$  样本个数， $k$  约束条件个数)。

**多重共线性诊断：**

由于目前 SPSS 在 Logistic 回归分析中尚不提供多重共线性诊断，替代方法之一是运用相同的观测变量与预测变量，拟合线性回归模型，并进行相应的多重共线性诊断。运用上述方法，以上各个模型预测变量的 Tolerance（容限度指数）和 VIF（方差扩大因子）的取值范围如下表所示：

表格 2.2 Logistic 回归分析的多重共线性诊断

Model	Tolerance 取值范围	VIF 取值范围
1	0.896-0.931	1.074-1.116
2	0.810-0.964	1.037-1.234
3	0.754-0.926	1.080-1.326
4	0.739-0.874	1.144-1.352
5	0.416-0.971	1.030-2.402
6	0.752-0.952	1.051-1.329

容限度和方差扩大因子是检验多重共线性的两个重要指标。容限度是由每个自变量 $X_j$ 作为因变量对其他自变量回归时得到的余差比例，当容限度很大时，说明 $X_j$ 包含的独立信息很多，可能成为重要解释变量；反之，容限度很小，说明 $X_j$ 对其他自变量的信息重复性越大，其对因变量  $Y$  的解释能力越小。容限度的大小是根据研究者的具体需要制定的，通常当容限度小于 0.1 时，便认为变量 $X_j$ 对其他变量之间的多重共线性超过了容许界限。

方差扩大因子（VIF）是容限度的倒数。它表示所对应的偏回归系数的方差由于多重共线性而扩大的倍数。当容限度为 0.1 时，VIF 为 10(倍)。通常当 VIF>10 时，便认为变量 $X_j$ 与其他变量之间存在多重共线性。



依据上述标准，上述 6 个 Logistic 回归模型均不存在严重的多重共线性问题。

2.1.3 心理健康网络行为关系模型

躯体化的行为模型：

表格 2.3 躯体化的行为模型

网络行为	取值	变量值	网络行为类别
参与话题小组	0、1	$T_{topic}$	PSI
发表日志	0、1	$T_{log}$	PSI
访问好友页面	0、1	$T_{friend}$	PSI
恐惧恐怖	1、2、3、4、5	$T_{terror}$	PAN
电子邮箱	1、2、3、4、5	$T_{email}$	PFN
网络购物拍卖	1、2、3、4、5	$T_{shopping}$	PFN

相应的计算公式： $Logit(躯体化) = 2.134 + 0.806 * T_{topic}$ （参与话题小组）+  
 $0.735 * T_{log}$ （发表日志）-  $1.229 * T_{friend}$ （访问好友页面）+  $0.447 * T_{terror}$ （恐惧恐怖）  
—  $0.789 * T_{email}$ （电子邮箱）+  $0.297 * T_{shopping}$ （网络购物拍卖）

焦虑的行为模型：

表格 2.4 焦虑的行为模型

网络行为	取值	变量值	网络行为类别
信息检索手段：其他	0、1	$T_{searchother}$	PSM
参与话题小组	0、1	$T_{topic}$	PSI
访问好友页面	0、1	$T_{friend}$	PSI

关注学术资料：> 3 小时	0、1	$T_{Academic}$	PTN
恐惧恐怖	1、2、3、4、5	$T_{terror}$	PAN
搜索引擎	1、2、3、4、5	$T_{se}$	PFN
短信服务	1、2、3、4、5	$T_{sms}$	PFN
无目的漫游	1、2、3、4、5	$T_{aimless}$	PFN

相应的计算公式： $Logit(焦虑) = 1.812 + 2.375 * T_{searchother}$ （信息检索手段：其它）+  $1.160 * T_{topic}$ （参与话题小组）-  $1.213 * T_{friend}$ （访问好友页面）-  $2.512 * T_{Academic}$ （关注学术资料：大于 3 小时）+  $0.443 * T_{terror}$ （恐惧恐怖）-  $0.876 * T_{se}$ （搜索引擎）-  $0.389 * T_{sms}$ （短信服务）+  $0.853 * T_{aimless}$ （无目的漫游）

病态人格行为模型：

表格 2.5 病态人格的行为模型

网络行为	取值	变量值	网络行为类别
网页平均逗留时间：30 秒—1 分钟	0、1	$T_{avg}$	ONB
网页平均逗留时间：> 5 分钟	0、1	$T_{avg}$	ONB
信息检索手段：其他	0、1	$T_{searchother}$	PSM
参与话题小组	0、1	$T_{topic}$	PSI
温馨	1、2、3、4、5	$T_{warm}$	PAN
软件下载	1、2、3、4、5	$T_{download}$	PFN
网络购物拍卖	1、2、3、4、5	$T_{shopping}$	PFN
同学录	1、2、3、4、5	$T_{classmates}$	PFN

无目的漫游

1、2、3、4、5

$T_{aimless}$

PFN

相应的计算公式： $Logit（病态人格）= 3.769 - 2.030 * T_{avg}$ （网页平均逗留时间：30 秒-1 分钟）-  $2.911 * T_{avg}$ （网页平均逗留时间 大于 5 分钟 ）+  $2.202 * T_{searchother}$ （信息检索手段：其它）+  $0.936 * T_{topic}$ （参与话题小组）-  $0.416 * T_{warm}$ （温馨）-  $0.964 * T_{download}$ （软件下载）+  $0.567 * T_{shopping}$ （网络购物拍卖）-  $0.332 * T_{classmates}$ （同学录）+  $0.690 * T_{aimless}$ （无目的漫游）

疑心行为模型：

表格 2.6 疑心的行为模型

网络行为	取值	变量值	网络行为类别
浏览不良信息网站次数（2 次）	0、1	$T_{unhealthy2}$	OBV
参与话题小组	0、1	$T_{topic}$	PSI
愤怒暴力	1、2、3、4、5	$T_{violence}$	PAN
同情	1、2、3、4、5	$T_{Sympathy}$	PAN
搜索引擎	1、2、3、4、5	$T_{se}$	PFN
同学录	1、2、3、4、5	$T_{classmates}$	PFN
无目的漫游	1、2、3、4、5	$T_{aimless}$	PFN

相应的计算公式： $Logit（疑心）= 1.994 + 1.954 * T_{unhealthy2}$ （浏览不良信息网站次数：2 次）+  $1.255 * T_{topic}$ （参与话题小组）+  $0.879 * T_{violence}$ （愤怒暴力）-

$0.549 * T_{Sympathy}$ （同情）-  $0.596 * T_{se}$ （搜索引擎）-  $0.481 * T_{classmates}$ （同学录） +  
 $0.390 * T_{aimless}$ （无目的漫游）

脱离现实行为模型：

表格 2.7 脱离现实的行为模型

网络行为	取值	变量值	网络行为类别
社交网络日均使用时 间：0.5—1 小时	0、1	$T_{Social}$	ONB
社交网络日均使用时 间：1—2 小时	0、1	$T_{Social}$	ONB
访问好友页面	0、1	$T_{friend}$	PSI
主动网络行为所占时 间比例：1%—20%	0、1	$T_{Active1-20}$	ONB
主动网络行为所占时 间比例：41%—60%	0、1	$T_{Active41-60}$	ONB
主动网络行为所占时 间比例：61%—80%	0、1	$T_{Active61-80}$	ONB
主动网络行为所占时 间比例：81%—100%	0、1	$T_{Active81-100}$	ONB
愤怒暴力	1、2、3、4、5	$T_{violence}$	PAN
浏览新闻	1、2、3、4、5	$T_{news}$	PFN
搜索引擎	1、2、3、4、5	$T_{se}$	PFN
无目的漫游	1、2、3、4、5	$T_{aimless}$	PFN

相应的计算公式： $Logit$ （脱离现实）=  $4.181 + 1.422 * T_{Social}$ （社交网络日均  
使用时间：0.5—1 小时）+  $1.370 * T_{Social}$ （社交网络日均使用时间：1-2 小时）-  
 $1.409 * T_{friend}$ （访问好友页面）-  $4.452 * T_{Active1-20}$ （主动网络使用时间比例：1%-20%）  
-  $4.667 * T_{Active41-60}$ （主动网络行为所占时间比例：41%-60%）-  $4.627 * T_{Active61-80}$   
（主动网络行为所占时间比例：61%-80%）-  $6.761 * T_{Active81-100}$ （主动网络行为  
所占时间比例：81%-100%）+  $1.230 * T_{violence}$ （愤怒暴力）-  $0.517 * T_{news}$ （浏览新

闻) - 0.834\* $T_{se}$ （搜索引擎）+ 0.956\* $T_{aimless}$ （无目的漫游）

兴奋状态行为模型：

表格 2.8 兴奋状态的行为模型

网络行为	取值	变量值	网络行为类别
年龄	具体值	$T_{age}$	OBV
即时通讯软件好友数量：50—100 个	0、1	$T_{50-100}$	PSI
即时通讯软件好友数量：100—200 个	0、1	$T_{100-200}$	PSI
日均联络网友个数：5—10 个	0、1	$T_{contact5-10}$	PSI
新闻：2—3 小时	0、1	$T_{news2-3}$	PTN
电子邮箱	1、2、3、4、5	$T_{email}$	PFN
电子杂志	1、2、3、4、5	$T_{e-zine}$	PFN
多媒体娱乐	1、2、3、4、5	$T_{media}$	PFN
无目的漫游	1、2、3、4、5	$T_{aimless}$	PFN

相应的计算公式： $Logit（兴奋状态）= 12.741 - 0.484 * T_{age}（年龄） - 1.813 * T_{50-100}（即时通讯软件好友数量：50—100 个） - 1.919 * T_{100-200}（即时通讯软件好友数量：100-200 个） + 1.944 * T_{contact5-10}（日均联络网友人数：5-10 个） + 1.829 * T_{news2-3}（新闻：2-3 小时） - 0.590 * T_{email}（电子邮箱） + 0.646 * T_{e-zine}（电子杂志） - 0.392 * T_{media}（多媒体娱乐） + 0.452 * T_{aimless}（无目的漫游）$

抑郁的行为模型：

表格 2.9 抑郁的行为模型

网络行为	取值	变量值	网络行为类别
性别	0、1	$T_{sex}$	OBV
上网时段：（9:00 之前）	0、1	$T_{Internet9}$	ONB

浏览不良网页的次数 (3—6 次)	0、1	$T_{unhealthy3-6}$	OBV
周一到周五平均上网 时间：(2—6 小时)	0、1	$T_{MondayToFriday2-6}$	ONB
主动上网时间所占比 例 (0%)	0、1	$T_{Active0}$	ONB
同情	1、2、3、4、5	$T_{Sympathy}$	PAN
社交网络使用行为： 其他	0、1	$T_{social-other}$	PSI
周六周日平均上网时 间：(3—4 小时)	0、1	$T_{weekend3-4}$	ONB
使用搜索引擎	0、1	$T_{startbrowsing}$	OBV
周一到周五平均上网 时间：(10—12 小时)	0、1	$T_{MondayToFriday10-12}$	ONB

相应的计算公式： $Logit(抑郁) = 25.019 + 1.53 * T_{sex}$ （性别：男 1 女 0）+  
 $2.63 * T_{Internet9}$ （上网时段：9 点之前）+  $1.113 * T_{unhealthy3-6}$ （浏览不良网页次数：  
3-6 次）—  $0.368 * T_{MondayToFriday2-6}$ （周一到周五平均上网时间：2—6 小时）+  $1.4 * T_{Active0}$ （主动上网时间所占比例：0%）+  $0.195 * T_{Sympathy}$ （同情）+  $0.631 * T_{social-other}$ （社交网络使用行为：其他）—  $0.382 * T_{weekend3-4}$ （周六周日平均上  
网时间：3—4 小时）—  $0.188 * T_{startbrowsing}$ （搜索引擎使用）—  $0.594 * T_{MondayToFriday10-12}$ （周一到周五平均上网时间：10—12 小时）

2.2 关键性开发技术介绍

### 2.2.1 .Net Framework 开发平台以及开发语言 C#介绍

.NetFramework 是用于 Windows 的新托管代码编程模型。它强大功能与新技术结合起来，用于构建具有视觉上引人注目的用户体验的应用程序，实现跨技术边界的无缝通信，并且能支持各种业务流程<sup>[11]</sup>。

.NET 框架是以一种采用系统虚拟机运行的编程平台，以通用语言运行库（Common Language Runtime）为基础，支持多种语言（C#、VB、C++、Python 等）的开发。

C#(C Sharp)是微软(Microsoft)为.NetFramework 量身订做的程序语言，C#拥有 C/C++的强大功能以及 Visual Basic 简易使用的特性，是第一个组件导向(Component-oriented)的程序语言，和 C++与 Java 一样亦为对象导向(object-oriented)程序语言。

### 2.2.2 朴素贝叶斯文本分类理论介绍

假设已训练出模型类别集合  $S$ ，其中每个类别分别为  $s_1, s_2, s_3 \dots s_m$ ，而训练时提出  $s$  类别的特征，分别表示为  $f_1, f_2, \dots, f_n$  (注意:这里  $f_i$  只有符号意义)，现有待分类文本  $X$ ，其文本特征值分别为  $x_1, x_2, \dots, x_n$  (这里有数值意义)，要确认文本  $X$  属于  $S$  中的哪个类别，即求条件概率

$$\begin{aligned} P(s=s_i | f_1=x_1, f_2=x_2, \dots, f_n=x_n) \\ &= P(s=s_i, (f_1=x_1, f_2=x_2, \dots, f_n=x_n)) / P(f_1=x_1, f_2=x_2, \dots, f_n=x_n) \\ &= P((f_1=x_1, f_2=x_2, \dots, f_n=x_n) | s=s_i) * P(s=s_i) / P(f_1=x_1, f_2=x_2, \dots, f_n=x_n) \\ &\sim P((f_1=x_1, f_2=x_2, \dots, f_n=x_n) | s=s_i) * P(s=s_i) \end{aligned}$$

其中，我们要的是各个概率之间的相比较，取概率最大者，因分母相同(即  $P(f_1=x_1, f_2=x_2, \dots, f_n=x_n)$  相同)，当然这里 Navie Bayes 假设了各个特征之间相互独立，故分母可以省略掉。先验概率  $P(s=s_i)$  可以由专业统计部门给出(比如搜狗实验室)，故只需算  $P((f_1=x_1, f_2=x_2, \dots, f_n=x_n) | s=s_i) * P(s=s_i)$ ，而已经假设了各个特征值之间相互独立，因此

$$\begin{aligned} P(s=s_i | f_1=x_1, f_2=x_2, \dots, f_n=x_n) \\ &\sim P(f_1=x_1, f_2=x_2, \dots, f_n=x_n | s=s_i) \\ &= P(f_1=x_1 | s=s_i) * P(f_2=x_2 | s=s_i) \dots * P(f_n=x_n | s=s_i) * P(s=s_i) \quad (\#) \end{aligned}$$

故对于  $i=1, 2, \dots, m$ ，取其中使得式子'#'最大值的  $i$  值  $k$ ，则我们就可以认为文

本 X 属于文本 sk 这个类别。

### 2.2.3 ZedGraph 控件介绍

ZedGraph 是一个开源的 .Net 图表类库，全部代码都是用 C# 开发的。它可以利用任意的数据集创建 2D 的线性和柱形图表<sup>[12]</sup>。

ZedGraph 的类库具有很高的灵活性。几乎图表的每个层面都可以被用户修改。同时，为了保证类库的易用性，所有的图表属性都提供了缺省值。类库中包含的代码可以根据被划分的数据来选择适应的比例范围和步长、尺寸。ZedGraph 继承了 Framework 中的 UserControl 接口，所以允许用户在 VS 的 IDE 环境中进行拖放操作。增加了对其它语言的访问接口支持，如 C++，VB。



## 第 3 章 系统分析

### 3.1 构架概述

WebMind 系统是一个客户端应用程序，总体上，系统分为五个模块：浏览器模块、绘图控件模块、历史网络行为库模块、网络行为分析模块、心理建议推荐模块。这些模块之间相辅相成。浏览器模块是整个系统的基础，历史网络行为库和网络行为分析模块是整个系统的核心部分，分析得出的用户心理趋势值可以显示在绘图控件模块上，而心理建议模块又是在这个基础上相对应的给出心理调解建议。可以用图 3.1 清楚的描述这些模块之间的关系：

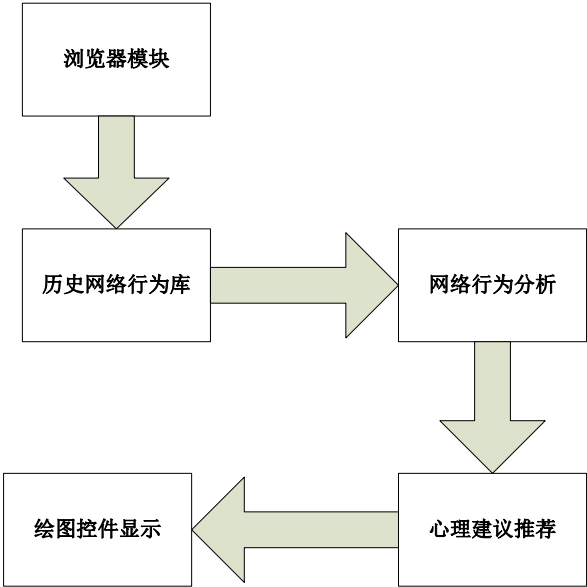


图 3.1 WebMind 系统模块分析

#### 3.1.1 模块需求描述

浏览器模块需要做到用户日常浏览器一样的功能需求。菜单栏、地址栏、状态栏，菜单栏具体可以参考 IE8 的菜单栏；地址栏的基本业务是允许用户输入地址，然后请求网页。同时有刷新、停止等动作。但最重要的是，需要在地址栏的右端显示用户心理状态；标签页需要按照 IE8 那种同窗口多标签页的形式；收藏夹、历史记录在浏览器左侧显示。

绘图控件显示模块的需求很单一，就是根据用户心理数据和行为数据来绘制用户曲线图。

### 3.2 系统开发环境

硬件配置：硬盘：320GB

内存：3GB

操作系统：Windows XP

开发环境：Visual studio 2008

环境配置：.Net Framework 3.5

### 3.3 系统任务的可行性分析

#### 3.3.1 技术可行性

(1)浏览器：基于 IE8 内核的 WebBrowser 控件，有完整的排版引擎和渲染引擎，因此可以直接调用 WebBrowser 控件，从而省去排版引擎和渲染引擎等深层次问题，而对于菜单项，完全可以通过本地设置文件的方式来显示持久化。

(2)中分分词：由于面向的用户，绝大部分是中文用户，而中文与英文有显著的差别——中文字之间没有空格。网络上有多种版本的中文分词程序，比如中科院提供的 ICTCLAS，但由于程序是由 C#编码，而 ICTCLAS 的编写太过复杂，因此选择了较为简单的双数组 trie 树算法，自己实现一个 c#版的中文分词程序。

(3).Net Framework 3.5:.Net 平台，在较新的操作系统(Vista、Win7)上，都自动安装有 .Net Framework 平台，而且其版本基本在 3.5 及其以上。

#### 3.3.2 系统安全性分析

随着经济的高速发展，社会节奏也变得越来越快，与此相对应的，是人们心理问题的凸显，而 WebMind 系统旨在通过用户的网络行为来跟踪分析他们的心理数据，以给出相应的心理调节建议。然而，要注意的是，近两年来，用户隐私问题在社会上多次引起轩然大波，而 WebMind 系统又是需要跟踪用户的一切网络行为的，因此，对于用户隐私的保护显得异常重要。在系统中，只是在客户端直接分析用户网络行为的数据，并不上传给服务器，如果用户愿意，可以将用户评价体系上传给服务器，这样，便一定程度上解决了系统的安全性问题。

#### 3.3.3 系统效率分析

心理所以及实验室的导师等的分析结果表明，市场上急需这样一套能跟踪反应人心理情况，同时还能给出相关的调节建议的系统。这样的系统，能大幅度增加心理治疗、咨询的人群范围，让很多过去因为费用问题或者距离问题的用户可以免费得到心理问题的预警反应和一定程度上的治疗。

## 第 4 章 系统设计

### 4.1 设计指导思想和原则

#### 4.1.1 指导思想

- （1）坚持统一规划、科学管理、服务应用的指导方针；
- （2）充分利用现有工作基础和成果，进行改进、提升和整合，避免重复建设。

#### 4.1.2 软件设计原则

##### （1）标准化和规范化原则

制订适合于本系统的分类编码方案。为实现系统的集成，必须保证系统建设采用的软件平台、数据接口、开发技术符合公认的工业标准，符合国家和行业的有关标准、规范、规程；同时要保证在系统的分析、设计、实现、维护阶段中必须采取开放路线，遵循软件工程的标准、规范。

##### （2）安全性原则

首先要全面考虑各种例外情况，保证系统的安全可靠，其次是需要加强数据保密工作，即用户隐私控制，确保用户的隐私有保障。

##### （3）完备性原则

系统的数据结构与功能体系最大限度的满足业务需求，减少各类意外情况，提高效率，并在此基础上循环更新心理调节建议库，让系统的精确度更高。

### 4.2 系统结构设计

WebMind 系统是一个基于网络内容分析的自助调节系统，其主要包括浏览器的实现、网络行为跟踪与分析 and 自助推荐心理调节建议模块。下面分别给出系统的总体框架图和运行流程图。

具体的 WebMind 系统框架如图 4.1 所示：

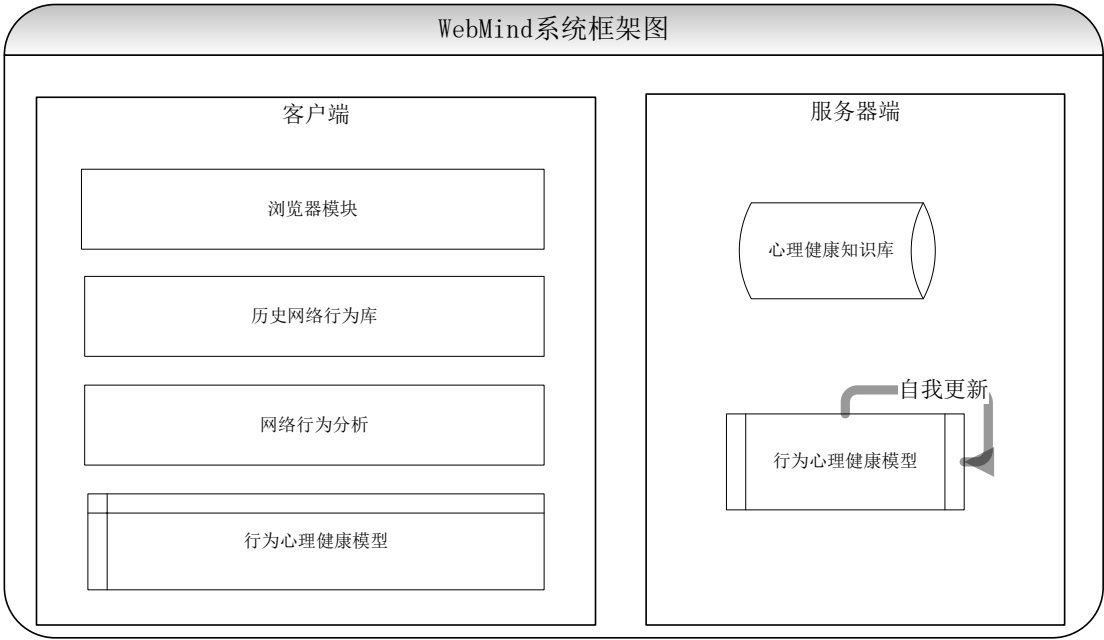


图 4.1 WebMind 系统总体框架图

4.3 系统流程设计

WebMind 系统总体运行流程图如图 4.2 所示：

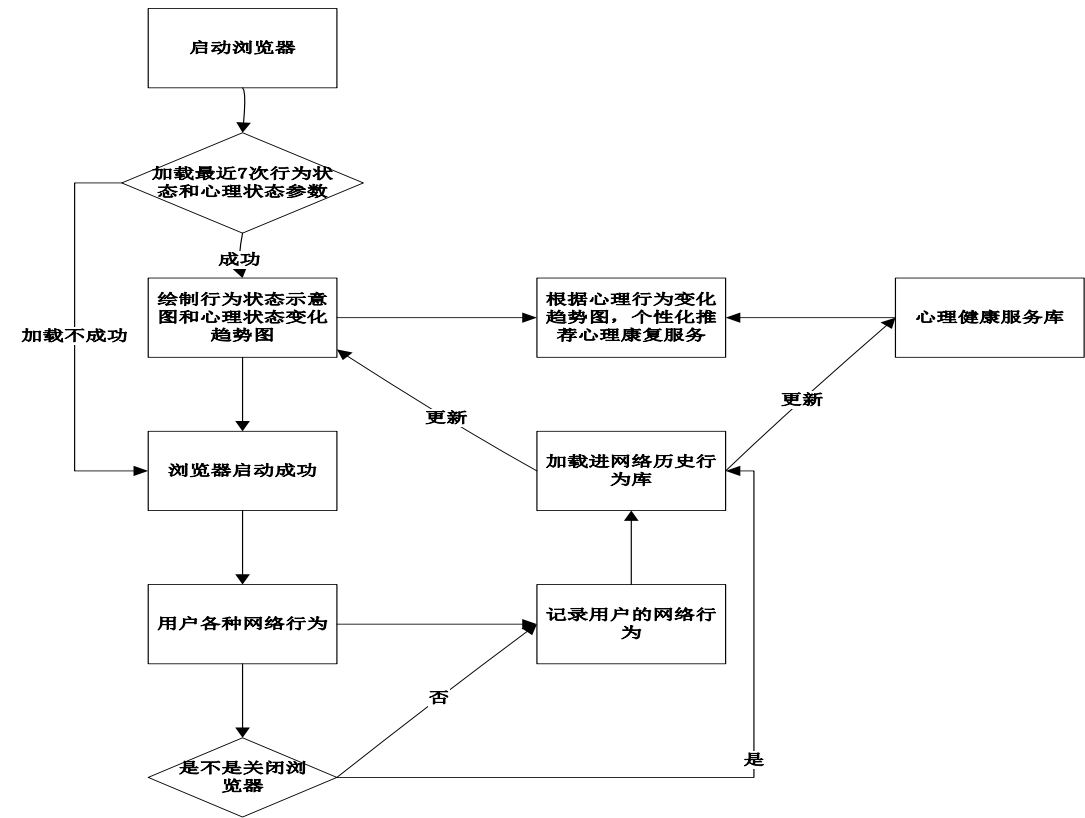


图 4.2 WebMind 系统运行流程图

4.4 系统模块设计

4.4.1 浏览器模块详细设计：

浏览器模块的基本构成是由各个窗体组成，自上往下分别是菜单栏、地址栏、标签栏、中间是网页显示的标签页控件、最下是状态栏。其具体位置关系大致如图 4.3 所述：

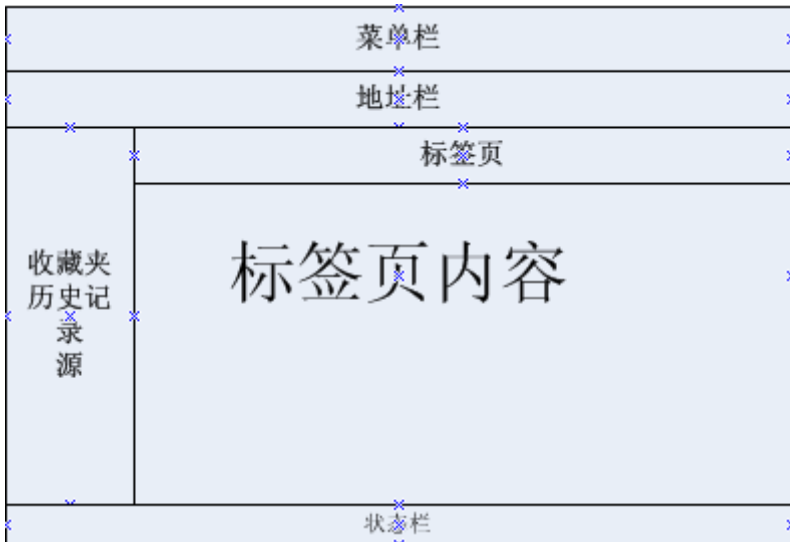


图 4.3 WebMind 浏览器的版块设计图

其中，菜单栏的选项有：文件、编辑、查看、收藏夹、工具、帮助，其具体排版的如图 4.4 所示：

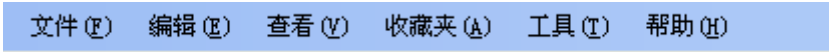


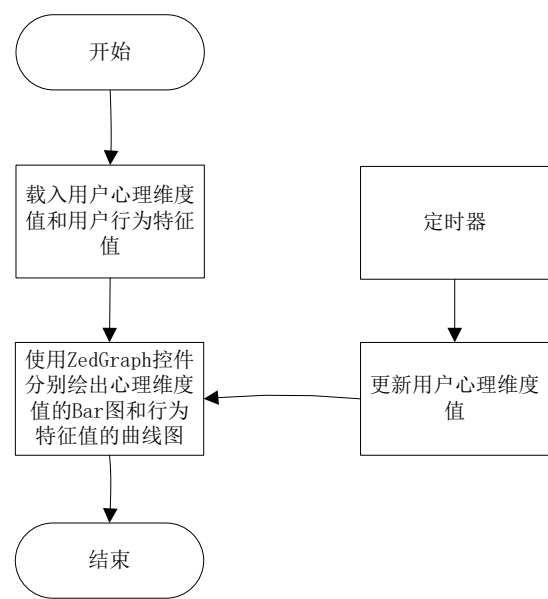
图 4.4 WebMind 浏览器的菜单栏设计

对于菜单栏，各个选项的下拉菜单基本上和用户日常所用的菜单项一样，收藏夹、历史记录的处理是自己记录在 xml 文件里，启动时加载那种模式，具体设计会在数据设计那个模块里陈述。

4.4.2 绘图控件显示模块

其实，在刚开始开题时，我并没有把绘图控件显示当做一个模块，只是在具体的实现过程中，发现绘图控件在地址栏右端的显示部分耗费了我很长很长时间，因而将绘图控件显示单独拿出当做一个模块来对待。

综合多种考虑，我最后选择了 c#程序里非常流行的 ZedGraph 控件当做绘图控件，其具体情况我已经在关键技术里介绍过了。图 4.5 具体介绍绘图模块的数据图形设计：



s

图 4.5 绘图控件显示设计的流程图

其中，用户心理维度值是使用 bar 图来绘制，而用户行为特征值是用曲线图来显示，二者的横坐标是用一次计算过程来统一的。

4.4.3 网络内容分析模块

网络行为分析的流程主要分为三个阶段，具体可参考图 4.6：

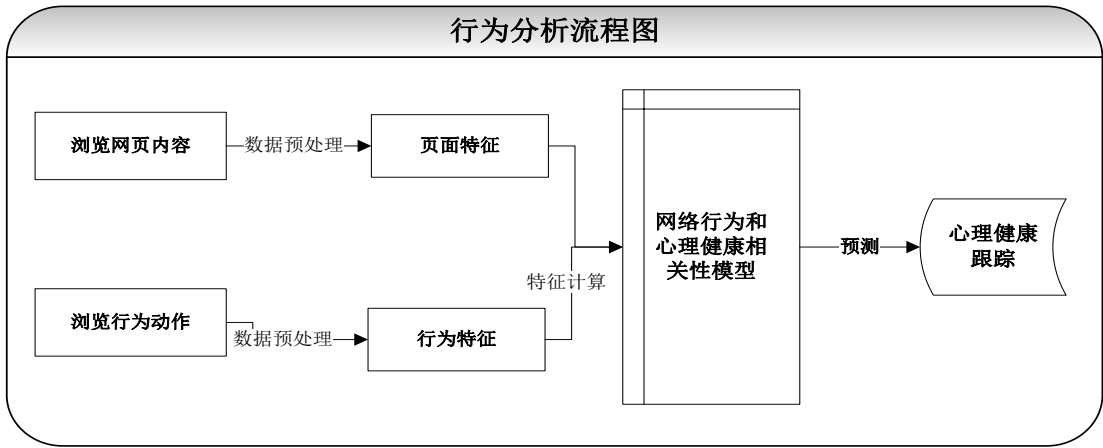


图 4.6 网络行为分析流程图

1、数据预处理

预处理过程定义为转化包括各种可利用数据源的使用记录、网络内容和结构信息为行为模式的数据在内的提取过程。预处理过程是整个行为分析过程的基础，主要包括使用记录预处理、内容信息预处理。

浏览行为动作预处理包括：识别用户、用户会话、用户浏览 URL 历史、用户的浏览操作行为等数据。浏览内容预处理包括：转化文本、图像、脚本、多媒体文件成为所需要的形式。当然，由于中文不同于英文，中文文字之间没有空格，因此，为了内容分类和情感分类，在预处理阶段需要进行中文分词，而本系统利用“双数组 trie 树算法”[13][14]来进行中文分词，下面给出该算法的流程图：

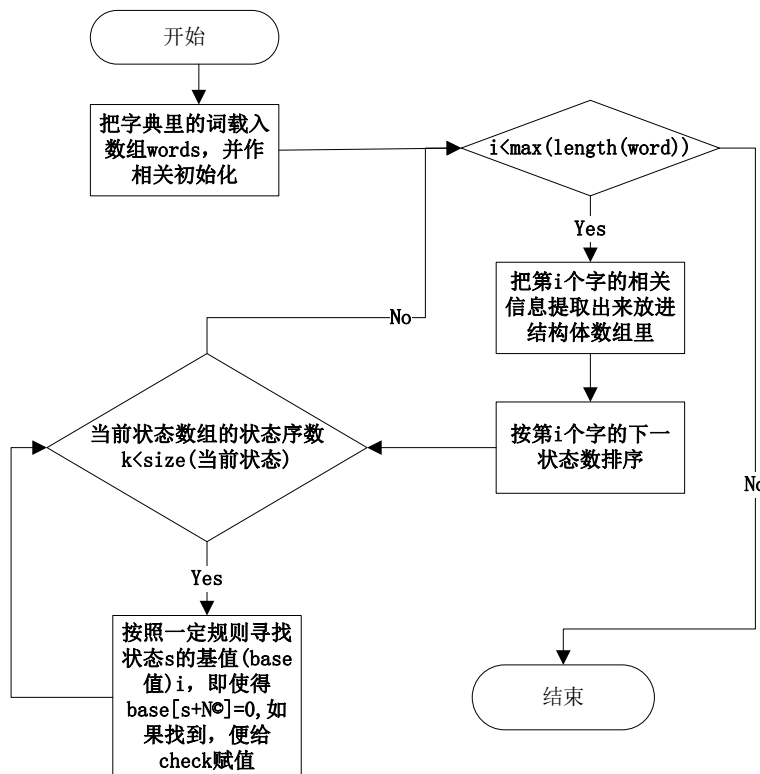


图 4.7 “双数组 trie 树”的算法流程图

通过上面的流程，便可以得出双数组，从而能实现中文分词。

## 2、数据特征提取和统计计算

### (1) 统计分析

统计分析是抽取用户访问网站知识的最常用方法。通过分析会话文件，依据浏览过的页面、浏览时间、导航路径的长度，我们能获取不同种类的描述统计信息（频度、均值、中值等）。有一些访问工具能获取一个周期性的统计报表，包括站点的最频繁的访问页面、平均浏览时间、平均浏览路径长度。还可以发现一些常见的错误，例如没有授权实体点和非法的 URL。

### (2) 内容分类

分类是将一个数据项映射到几个预先给定类别中的某一个的方法，其中朴素贝叶斯分类[15][16]是一种较容易实现而且效果较好的分类方法。在 WebMind 浏览器获取到一个网页时，使用朴素贝叶斯分类器对该网页的内容进行文本分类，其类别为：汽车、财经、IT、健康、体育、旅游、教育、招聘、文化、军事。

(3) 情感分类

在朴素贝叶斯分类器的基础上，增加情感字典[17]来加强分析用户浏览网页的情感分类，其中，情感暂设以下分类： 愤怒、搞笑、难过、无聊、新奇。

3、心理维度值的计算

网络行为分析的目标就是得出用户心理状况值。根据跟踪的用户网络行为以及多种分析结果，通过网络行为和心理健康状况的相关性模型进行心理状况预测，从而得出用户当前阶段心理状况。最后，通过曲线图的方式跟踪心理健康的变化趋势。

4.4.4 心理建议推荐模块

心理建议推荐模块是在网络行为分析模块的基础上，根据心理学规律，针对网络行为分析得出的各类心理问题，分别构造心理调节建议库。同时，也是根据网络行为分析所反映出的用户的心理问题，个性化的推荐用户的心理调节建议。目前，在系统中，已经尝试使用三种推荐模型：完全随机推荐模型，行为得分推荐模型，行为变化推荐模型。其中，完全随机推荐模型是用来做对照组的；而行为得分推荐模型，是依据用户心理问题严重度最大的那个维度来推荐相应的心理调节建议；行为变化推荐模型的原理是认为连续几次之间波动最大的那个心理维度便是用户心理最严重的问题之所在，因此，可以根据这个来推荐用户对应的心理调解建议。图 4.8 分别给出三个推荐模型的设计示意图：

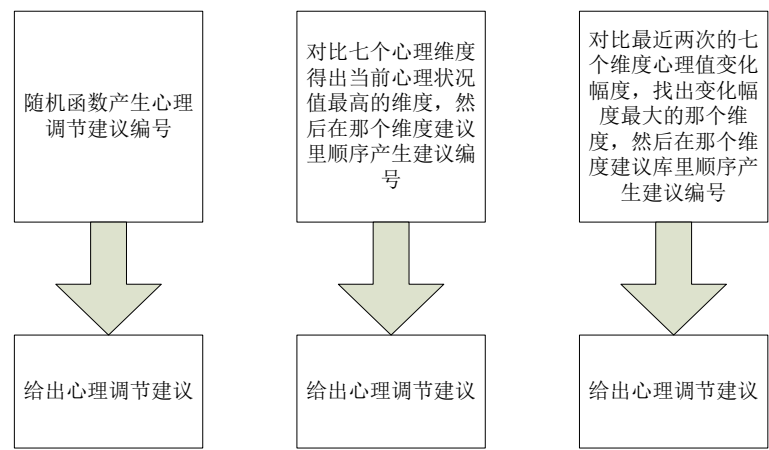


图 4.8 三种心理推荐模型的设计原理

其实，在三种推荐模型中，完全随机推荐模型是对照组，方便对比出另外两种推荐模型哪种模型的推荐效果更好，以进一步去优化系统的推荐效率。

4.5 系统数据库(历史网络行为库)设计

WebMind 在本质上是一款有跟踪分析功能的浏览器，因此，系统数据库的设计相对来说就显得很重要。在系统数据库的设计过程中，分为网页浏览历史数



数据库和浏览器操作历史数据库。其中，网页浏览历史数据库就是网络行为分析的主要对象，包括网页内容、url 解析、页面停留时间等多个条目；而浏览器操作历史数据库就是设置用户的个性化设置，比如隐藏、显示某个栏、收藏夹、浏览历史记录、浏览器选项设置等。羡慕就分别对两类数据库进行设计。

4.5.1 网页浏览历史数据库设计

网页浏览方面需要记录 url 分析、url 映射内容、行为特征分析、行为特征分析历史、内容分析、内容分析历史、情感分析、情感分析历史、心理维度值、三个推荐模型建议库、用户对心理推荐模型的评价等表项。这些表项均是以 xml 文件的形式在用户的终端存在。网页浏览操作所需要记录的内容之间的内在关联可用图 4.9 表示：

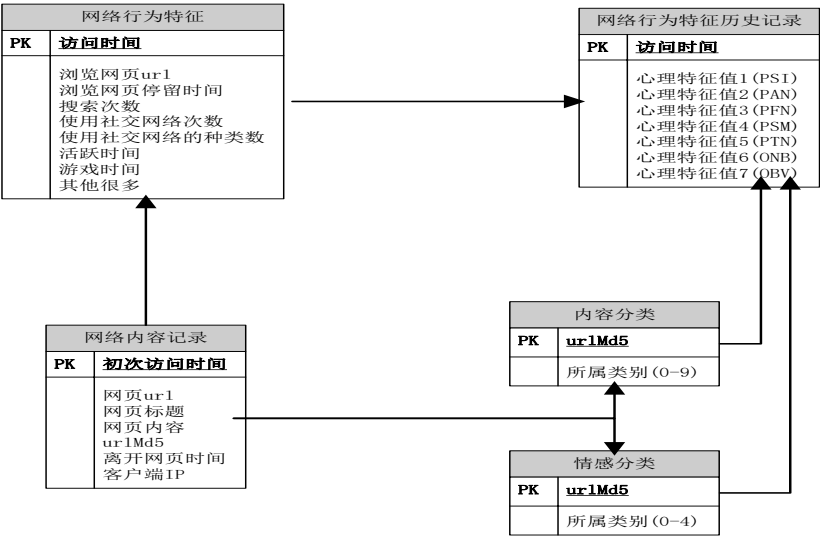


图 4.9 WebMind 系统网页浏览数据库设计

根据心理调节建议库的三种模型的格式设计如图 4.10 所示：



图 4.10 WebMind 系统推荐模型数据库设计

设计用户对心理调节建议的三种模型的评价数据库可以按照图 4.11 所示的方式设计：

用户评价	
PK	建议ID
	评价时间 模型类别 评价分数

图 4.11 WebMind 系统用户评价数据库设计

通过上面几种方式，可以在用户客户端设计出完整的用户浏览历史记录，并能据此和一些相关技术知识得出用户当前阶段心理状况趋势数值，从而为调节用户心理做好准备。心理建议推荐模型的提出，使得可以直接在用户体验的基础上实验哪种用户推荐模型是最有效果的，因此，用户评价数据库的作用就很明了了。

4.5.2 浏览器操作数据库设计

这部分是根据用户的日常浏览器操作体验来设计的，设计的基本思想就是与用户的其他浏览器体验一致，为此，我设计了收藏夹、浏览历史记录、上次会话内容，还有 IE 选项设置。按照传统的浏览器习惯，收藏夹都是收录用户的添加的一些重要或者感兴趣的网页等内容，浏览历史记录是收录用户浏览过的网页的 url，上次会话内容是为了用户因为没有正常关闭浏览器而引起的浏览网页的丢失而设计的，这样一来便可以还原上次会话的内容，浏览器选项我设计的比较简陋，因为可以调用微软提供的 API，但我还是自己设计了一个简陋的浏览器选项设置。

浏览器选项设置，比如默认主页，比如工具条是否隐藏等，其相关设置可以用图 4.12 所示的数据库模型图来设计。

浏览器选项设置	
PK	更新时间
	工具条显示 主页

图 4.12 浏览器选项设置的数据格式设计

浏览器上次会话的 url 记录，在浏览器非正常关闭时会自动打开上次会话的网页，其 url 存储数据库表设计如图 4.13 所示。

上次会话	
PK	更新时间
	url(不定数)

图 4.13 浏览器上次会话的数据格式设计

对于浏览器历史记录的处理，我在客户端设置了一个 xml 文件，记录用户的网页浏览记录，图 4.14 显示了其设计内容。

浏览历史记录	
PK	<u>url</u>
	访问时间

图 4.14 浏览器浏览历史数据格式设计

浏览器收藏夹，收藏了用户感兴趣或者觉得重要的网页，因此需要在客户端设计相应的存储 url 的数据库，可参考图 4.15。。

收藏夹	
PK	<u>url</u>
	添加时间

图 4.15 浏览器收藏夹数据格式数据

这些用户偏好的数据库都是以客户端 xml 文件的方式存在于用户的主机上的，而且这些内容的设计方式与 IE 浏览器的设置方式基本一样，目的是让用户习惯这款浏览器。不过，由于设计过程中时间限制和考虑不足，用户偏好的设置应该还有很多可以改进的地方。

## 第 5 章 系统实现

### 5.1 WebMind 系统框架

WebMind 系统分为两部分：客户端和服务端。

客户端系统以 WebMind 浏览器为载体，进行网络行为的收集、分析和预测，同时还进行心理健康状况的跟踪。

服务端系统主要负责心理健康知识的更新和行为心理健康模型的自我更新。

其中，客户端系统分为四个模块：浏览器模块、历史网络行为库、网络行为分析和行为心理健康模型。服务器端系统分为：心理健康知识库和行为心理健康模型更新。

### 5.2 浏览器模块

WebMind 系统是 WebMind 系统的主体显示部分，是一个可以跟踪用户心理健康变化的浏览器。它不仅具有通用浏览器的常用功能：解析 HTML 网页、标签页显示、收藏夹管理、下载管理等，还可以通过分析用户的网络访问行为，跟踪用户的心理健康变化，并在浏览器上对心理健康变化给出合理性的建议，是一个懂得人心理的智能化浏览器。

WebMind 浏览器的截图如下：



图 5.2 WebMind 浏览器总体图

如图 5.2 所示，WebMind 具备通用浏览器的功能，以标签页的形式进行网页显示。对比常见的浏览器，WebMind 浏览器的右上角有个小窗口，这个小窗口通过曲线图跟踪显示用户心理健康的变化情况，同时可以及时给予心理的干预调节建议，调节心情，使人们更加健康的上网。

#### 5.2.1 浏览器基本功能

当前 WebMind 浏览器具备的基本功能有：

- （1）网页基本内容的解析和显示：WebMind 浏览器采用 IE8.0 浏览器内核对网页进行解析和显示
- （2）收藏夹：添加收藏夹、从收藏夹浏览网页
- （3）历史记录：存储用户浏览过的网页 url
- （4）其他一些 IE 的基本功能:比如菜单栏上的一些功能选项
- （5）以绘图的方式显示用户当前心理状态和网络行为特征值
- （6）推荐心理调剂建议

### 5.2.2 记录浏览历史

WebMind 系统主要记录用户的浏览历史的域名信息、URL 和相应的时间信息，以 XML 文件的形式保存在客户端。

记录浏览历史记录的内容有：

- （1）网页浏览的信息，具体包括：页面开始访问时间、页面访问结束的时间、用户、IP、网页 URL、网站类别、标题
- （2）抓取网页的具体内容：DOM 树的 Title 和 Body 标记中，去除网页标识内容，仅仅抓取文本内容

具体的记录结果以 XML 文件的形式保存在本地机器，其中，XML 文件的数据格式的定义详见图 5.3 和图 5.4

```
<UrlRecord UrlMD5="http://www.sina.com.cn/">
  <DomainName>www.sina.com.cn</DomainName>
  <Title>新浪首页</Title>
  <FirstTime>2011-05-20 11:14:33</FirstTime>
  <LastTime>2011-05-2011:14:49</LastTime>
  <ClientIP>210.77.29.188</ClientIP>
  <StaySecond>17</StaySecond>
</UrlRecord>
```

图 5.3 WebMind 系统记录的浏览历史记录一

```
<UrlRecord UrlMD5="">
  <Count>0</Count>
  <ContentCategory>0</ContentCategory>
  <EmotionCategory>0</EmotionCategory>
</UrlRecord>
```

图 5.4 WebMind 系统记录的浏览历史记录二

### 5.2.3 心理指标追踪

在浏览器的右上角有个小窗口（见图 5.2），通过曲线图来显示该用户心理健康的变化情况。通过曲线图，可以非常直观的把用户的心理变化情况呈现出来。用户可以点击小窗口，从而打开心理状态跟踪的详细窗口，如下图所示：

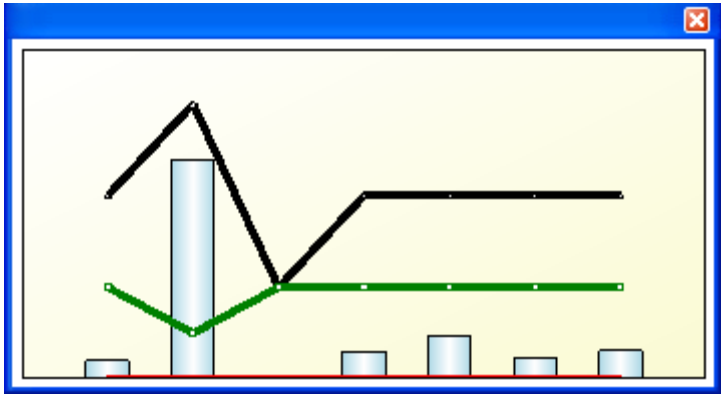


图 5.5 心理状况跟踪图

如图 5.5 所示，我们把心理健康状况分为六个指标进行分别跟踪，相隔一段时间，WebMind 系统就会针对当前的网络行为重新计算一个当前的心理健康指标。

曲线图的横坐标是评估心理健康指标的时间，纵坐标是计算的评估值。其中，评估值为正数，表明用户当前的心理指标处于模型的高分组；评估值为负数，表明用户当前的心理指标处于模型的低分组。

5.2.4 心理健康建议

通过评估的心理健康指标，结合心理学干预调节建议库，给用户推荐合适的干预调节建议。调节建议由两种方式进行显示：自动推荐和用户主动查看。

- （1）自动推荐：系统每隔一定时间，就会在显示器右下角弹出一个意见窗口，显示干预调节建议。
- （2）用户主动查看：用户可以通过 WebMind 浏览器主动查看心理指标跟踪的详细情况，同时系统会给出相应的干预调节建议，建议出现方式如图 5.6 所示。

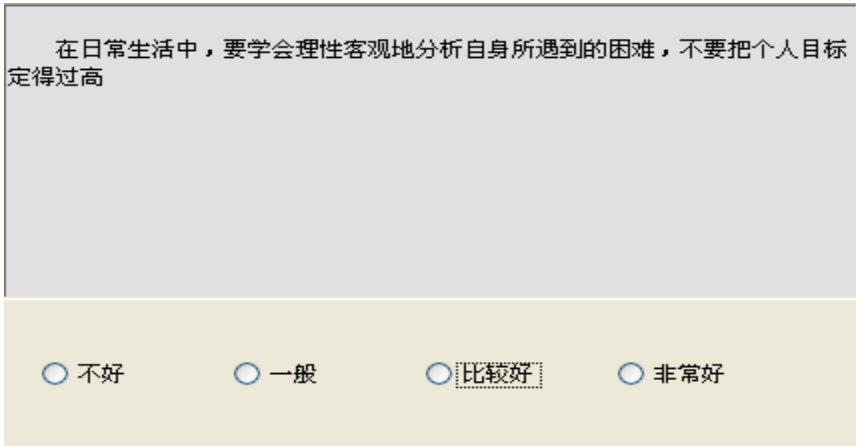


图 5.6 心理健康干预建议

5.3 历史网络行为库

历史网络行为库是通过 WebMind 记录的用户网络行为集合。依据网页的语言类别，记录在本地硬盘中。具体可参考图 5.7。

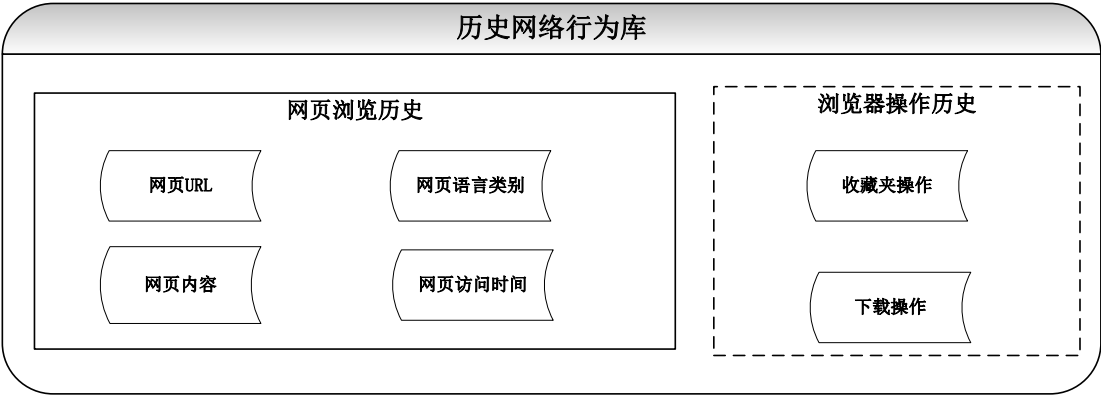


图 1 WebMind 系统的历史网络行为库

5.3.1 网页浏览历史

网络浏览历史的记录格式：

- ✧ 页面打开时间
- ✧ 页面关闭时间
- ✧ 标签页的激活时间
- ✧ 用户 IP
- ✧ 网站 URL
- ✧ 网站标题

区分网页语言类别作用：

在进行文本内容分类的时候，需要依据网页内容的类别进行选择采用中文的文本分类或者是英文的文本分类。

5.3.2 浏览器操作历史

在浏览器上的操作历史，现在只能记录两种操作：把网页添加到收藏夹，用户的下载历史。

5.4 用户网络行为分析

用户网络行为分析主要通过以下四种途径进行分析：

5.4.1 访问时间频度统计

统计用户访问 URL 的时间和次数信息。统计的网络行为有：周六、周日上网时间；周一到周五的上网时间；、上网的时间段（上网最频繁）；每个网页的平均停留时间；主动网络行为和被动网络行为 所占百分比；最常使用

的开始网络浏览的方式；使用搜索引擎时，您一般会选择哪种检索手段；每周主动浏览不良网站的次数。

### 5.4.2 URL 域名解析

通过建立域名库，解析出用户访问的网页所属的类别。主要区分类别有：电子邮箱；浏览新闻；搜索引擎；软件下载；网络购物、拍卖；短信服务；电子杂志；多媒体娱乐；同学录、校友录；无目的漫游。

### 5.4.3 网页内容分析

利用朴素贝叶斯分类器，以搜狗实验室的语料数据为训练集进行模型训练，对用户访问页面的内容进行分类。主要分出的类别有：计算机类；网络游戏类；健康；新闻；休闲娱乐；学术研究资料；科学动态与趋势；购物；社会；体育。

文本分类我们采用的朴素贝叶斯分类算法实现的分类器。下面主要介绍下在系统里文本分类实现过程。

在浏览器的网页下载事件发生后，便把该 document 的内容进行处理。当然还有别的行为分析，比如记录在一个网页上停留时间等等等。在这里我讲核心的东西——文本分类。下面我简单介绍下这个过程。

首先，中文网页，当然得中文分词，我用的是双数组 trie 树写的一个分词程序，在系统设计的 4.3.3 的数据预处理里我已经详细介绍过。

得到分词后的中文内容后，用 naive bayes 原理写了个文本分类器，其详细理论可见 2.2.2。至于训练文本，我是用搜狗语言实验室的训练数据，分为内容分类和情感分类(情感分类接下来会做特殊处理，以使得效果更好)。

当得出训练模型后，内容分类和情感分类大小分别为 2.37M 和 1.01M，由于读取模型耗费时间较长，我尝试了序列化存储对象，后来我改变了训练模型在文件里的存储格式，使得缩小到 2.07M 和 837K，然后我在分类器里读取训练模型时将 SortedDictionary 改成了 Dictionary，一下子时间缩短了不少，同时，我改了些别的地方，最后使得读取模型时间缩减到 0.18s 和 0.083s，这样是可以接受的。

集成后，每次启动一个新网页，我都要将其文本分类，开始时我很傻的采用了每次启动网页，每次都读取一次模型，因为我觉得模型比较大，存在内存里太浪费空间了。后来老师问我每次读模型什么时候释放的模型，我终于想通了，直接用了个单态模式(singleton pattern)，当第一次浏览网页时读取模型，接下来数次便不用读取分类模型了，用户关闭浏览器时自动释放模型。这样明显加快了浏览器的请求链接的速度。

另外一方面，在分类后，其分类结果(以一个整数值代表分类类型)会以字典的方式存储客户端的 xml 文件里。



5.4.4 网页情感分析

我们采用语料库和词典相结合的方法作为情感文本的分类[18][19]。我们的情感语料库是从新浪社会新闻排行榜抓取的从 2007 年 4 月 23 日到 2009 年 11 月 14 日的中文新闻文本。分为八类：感动、同情、难过、无聊、搞笑、温馨、新奇、愤怒。新浪社会新闻可以根据读者对看完新闻后的感受选择这八个情感中的哪一个，每天各个类别的新闻根据读者投票得分最高的前八条新闻列入心情排行榜。

情感词典的建立采用自动获取加人工校正的方式，并去除重叠词语和非情感词汇，共得到 1302 个情感词。

主要分出的类别：喜悦、开心；愤怒、暴力；哀伤、难过；恐惧、恐怖；新奇、奇特；感动；温馨；同情。

5.4.5 统计结果

依据 4.3.1 中定义的网络行为分析对收集到的网络行为进行分类。具体分为七类：PSI、PAN、PFN、PSM、PTN、ONB 和 OBV。在系统里实现结果可参见图 5.8。

```
<Record Time="-8589066809367432058">
  <SOM>-0.449</SOM>
  <DEP>25.604</DEP>
  <ANX>-0.807</ANX>
  <PSD>2.318</PSD>
  <HYP>0.3109999999999999</HYP>
  <UNR>0.529</UNR>|
  <HMA>-3.208</HMA>
</Record>
```

图 5.8 WebMind 系统网络行为分析的结果

以 PSI 为例，来讲一下网络行为分析的结果。针对每一种网络行为（例如：每天联络的网络好友数），通过设定一些统计阈值，把网络行为离散为一系列统计变量：ContactFriends2(2 个以下)、ContactFriends2\_5(2-5 个)、ContactFriends5\_10(5-10 个)、ContactFriends10\_20(10-20 个)和 ContactFriends20(20 个以上)。通过分析得出用户的网络使用习惯，把对应统计量置为 1，其余统计量置为 0。详细见下图 5.9。

```
<EmotionTracking>
  <Record Time="2011-04-11 09:22:51">
    <PSI>
      <Behavior_A>1</Behavior_A>
      <Behavior_B>1</Behavior_B>
      <Behavior_C>0</Behavior_C>
      <Behavior_D>0</Behavior_D>
      <Behavior_E>0</Behavior_E>
      <Behavior_F>0</Behavior_F>
      <Behavior_G>0</Behavior_G>
      <Behavior_H>1</Behavior_H>
      <Behavior_I>0</Behavior_I>
      <IMActiveTimeA>0</IMActiveTimeA>
      <IMActiveTimeB>0</IMActiveTimeB>
      <IMActiveTimeC>0</IMActiveTimeC>
      <IMActiveTimeD>0</IMActiveTimeD>
      <IMActiveTimeE>1</IMActiveTimeE>
      <IMFriends10>0</IMFriends10>
      <IMFriends10_50>0</IMFriends10_50>
      <IMFriends50_100>0</IMFriends50_100>
      <IMFriends100_200>1</IMFriends100_200>
      <IMFriends200>0</IMFriends200>
      <ContactFriends2>0</ContactFriends2>
      <ContactFriends2_5>1</ContactFriends2_5>
      <ContactFriends5_10>0</ContactFriends5_10>
      <ContactFriends10_20>0</ContactFriends10_20>
      <ContactFriends20>0</ContactFriends20>
    </PSI>
  </Record Time="2011-04-11 09:22:51">
</EmotionTracking>
```

图 5.9 WebMind 系统网络行为分析的结果（PSI 详细结果）

5.5 心理健康行为模型

心理健康行为模型选取的是 PHI 方式来作为用户心理量化指标，其共有七个维度。根据这些心理健康行为的 phi 值，可以形成心理健康建议知识库。它们之间的关系可以用图 5.10 表示。

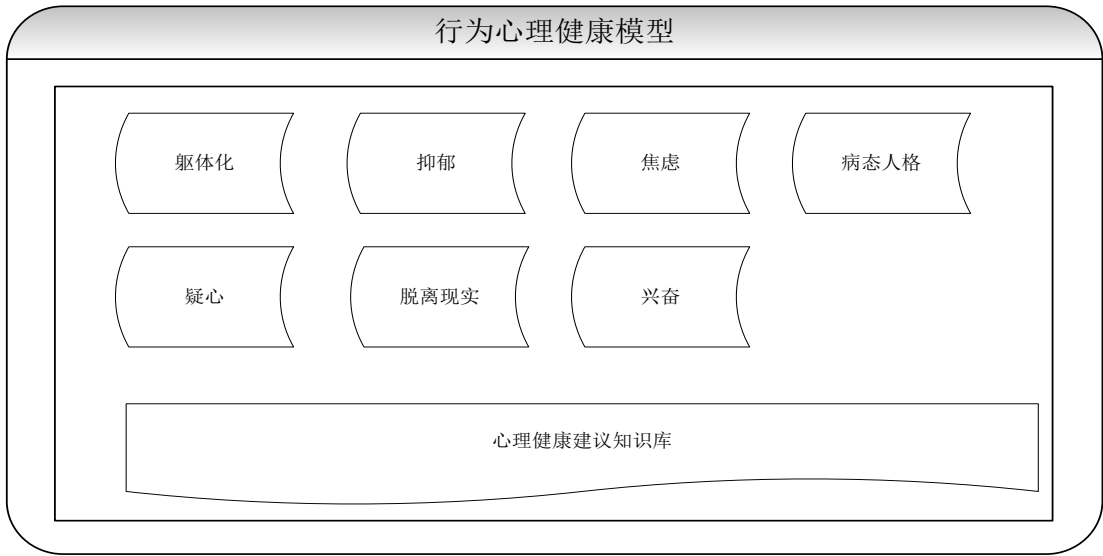


图 5.10 WebMind 系统心理健康行为模型

以 PHI 问卷上各心理健康问题维度的高低组二分类（高分组的 27% 和低分组 27%）变量为观测变量，一系列网络使用行为变量与人口统计学变量为预测变量，进行 Logistic 回归分析（ $\alpha = 0.05$ ）。分析结果表明，被试者在“抑郁”维度高低得分组中同时存在着大量彼此重叠的得分，使得高低得分组之间缺乏足够的区分度，因此对于“抑郁”维度建立一个线性回归模型，而对其他 6 个维度建立 6 个 Logistic 线性回归模型，一共七个模型。下图所示的就是其中一个模型：躯体化维度评估的模型。

```

3      /// <summary>
      /// 躯体化预测结果
      /// </summary>
-      /// <returns></returns>
3      public static double Calc_SOM_Prediction_Value()
      {
          //PSI_Behavior_B : 参与话题小组
          //PSI_Behavior_F : 发表日志
          //PSI_Behavior_H : 访问好友页面
          //PAN_Fear : 恐惧/恐怖
          //PFN_Email : 电子邮箱
          //PFN_OnlineShopping : 网络购物拍卖
          double dSOMValue =
              2.134 + 0.806 * PSI_Behavior_B + 0.735 * PSI_Behavior_F
              - 1.229 * PSI_Behavior_H + 0.447 * PAN_Fear - 0.789 * PFN_Email
              + 0.297 * PFN_OnlineShopping;
          return dSOMValue;
      }
-

```

图 5.11 心理健康行为模型-躯体化评估

## 5.6 行为心理模型的自我更新

WebMind 系统初始的心理健康行为模型是基于中科院研究生院 400 多位学生建立的。因为训练数据人群特点，不可避免存在着误差。因此，在 WebMind 系统设计之初，就考虑了采用模型自我更新的途径来解决存在的误差。其具体实现可参见图 5.12。

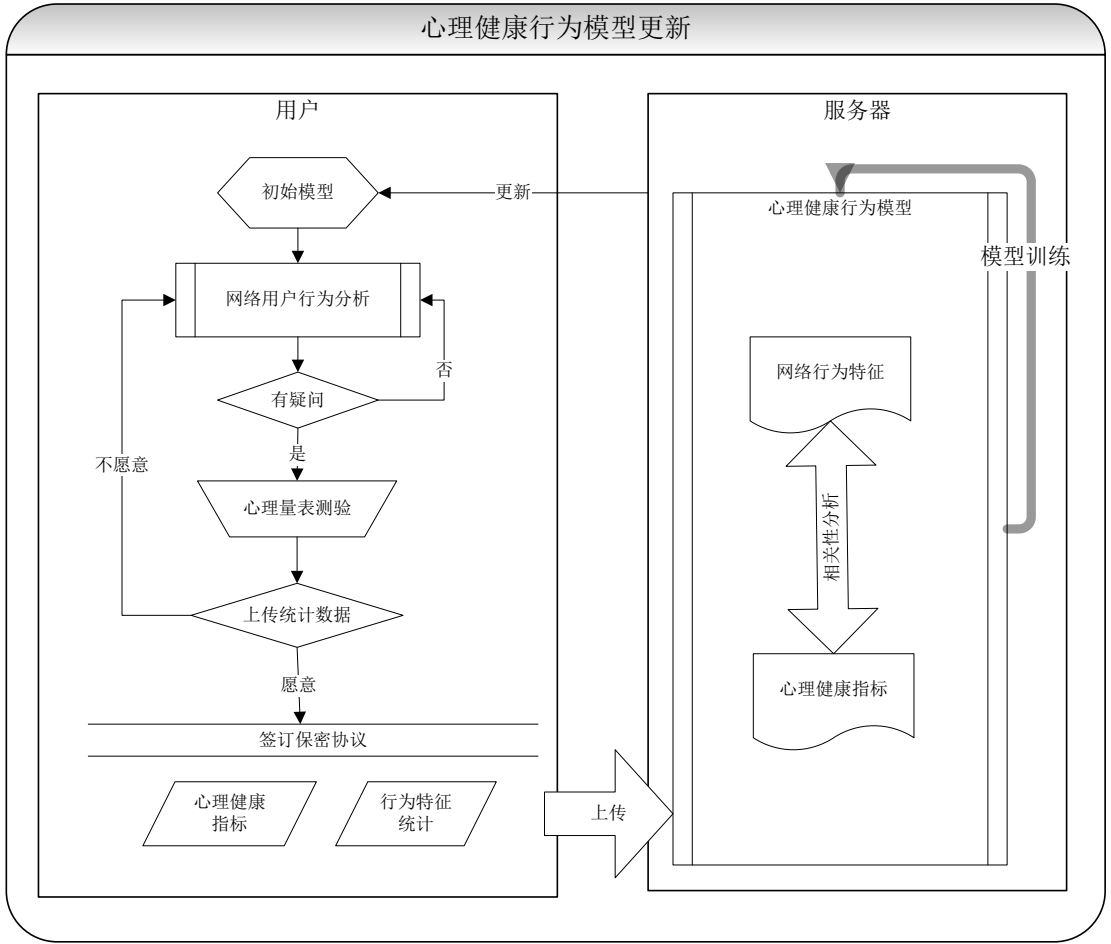


图 5.12 WebMind 心理健康行为模型更新

更新策略

处于用户隐私的考虑，系统并不主动收集用户的网络行为，所有的网络收集和分析工作均在本地客户端进行，模型更新需要收集的网络数据，需要用户依据意愿主动上传。模型的再次训练和更新工作在服务器端进行，计算出最新模型后，再由服务器端提供给用户下载使用。

具体的更新策略：

- ◆ 客户端：对收集的用户网络行为数据进行统计分析，生成标准 XML 格式的统计结果文档。如果用户对测试结果存有异议，可以选择进行标准的心理量表评估，给出权威的心理评估结果。依据意愿，如果用户愿意贡献网络数据，用户需要把存储统计结果的 XML 文档和所做心理量表测试的结果的 XML 文档一起上传到服务器的指定位置。
- ◆ 服务器端：每隔一定周期，服务器对收集到的网络行为特征数据和心理测试结果整理，重新建立线性回归模型，从而计算出最新的心理健康网络行为相关性模型。
- ◆ 生成模型后，由服务器端提供给用户下载使用。

## 5.7 实验结果及其分析

### 5.7.1 实验步骤

- 第一步：每个参与实验的用户填写“网络使用行为问卷”(IUBCL)，此问卷用以计算生成初始的心理健康行为模型。
- 第二步：每个用户安装使用 WebMind 系统两个星期，每天使用 WebMind 系统至少一个小时。收集用户的网络行为、相应的心理预测结果。
- 第三步：用户在使用 WebMind 系统过程中，系统会定时依据心理预测结果给出干预调节建议。参与实验的用户针对调节建议是否对自己有用，进行评分反馈。
- 第四步：试验结束后，参与实验的用户需要参与心理量表的测评。填写自评抑郁量表（SDS）和焦虑自评量表（SAS），量表测评结果用以和网络行为结果做比照分析。
- 第五步：参与实验的用户需要提交记录网络行为统计信息的 XML 文件、用户使用反馈报告和相应的两个心理量表测试结果。

### 5.7.2 实验结果与分析

- 通过两个星期的实验，我们总共收集 68 个有效网络用户的数据，有效用户使用反馈 672 条。实验结果的分析主要从两个方向来进行：心理健康状况评估效果分析和 WebMind 系统用户使用反馈分析。
- 心理健康状况评估效果分析：

通过 WebMind 系统网络行为分析的用户心理健康状况和用户通过 SDS 和 SAS 量表计算出的心理健康状况进行对比。主要分析二者对于心理健康指标评估的区域分布是否一致。

以心理健康维度焦虑指数为例进行比对分析。

第一步：通过 WebMind 系统计算出参与实验 68 位用户的焦虑维度的得分。Logistic 线性回归得分把实验用户分为正负两组（高分组 23 个，低分组 45 个），其中，正数表明网络用户处于该维度的高分组，负数表明网络用户处于该维度的低分组。

第二步：把高低分组的网络用户，按照 SAS 焦虑自评结果的原始分再进行分组。具体分组详见表格 5.1

表格 5.1 用户焦虑指数对比

行为分析分组	SAS 焦虑自评测评结果（原始分）	命中率
	35 以下    35-38 分    38 分之上	

高分组（23 个）	6	7	10	73.9%
低分组（45 个）	33	7	5	73.3%

第三步：分析实验结果。SAS 焦虑自评量表的平均分应为 38 分，但由于我们建立测试模型的人群中焦虑指数偏低，低于全国的常模。因此，在试验统计中采用 35 分和 38 分为两个统计基准，分为三组，可以得出 WebMind 系统通过网络行为评估心理维度焦虑的命中率在 73.3%-73.9% 。

以心理健康维度抑郁指数为例进行比对分析。

第一步：计算 68 个用户的 SDS 抑郁自评得分，依据得分分布情况分为三组：低分组 20 个记录，原始分从 20 到 31；中分组 25 个记录，原始分从 32 到 36；高分组 23 个记录，原始分从 37 到 54。

第二步：依据 SDS 分组情况，对 WebMind 系统网络分析的数据也进行对应分组，计算对应的命中记录个数。详细结果见表格 13

表格 5.2 用户抑郁指数的对比

SDS 分组	命中	成功率	总体命中率
高分组（23 个记录）	19	82.6%	75%
中分组（25 个记录）	17	68%	
低分组（20 个记录）	15	75%	

依据表格 5.2，WebMind 系统通过网络行为分析，在设定的三个区域中命中 51 个数据。除去命中失败的 17 个数据点，得到抑郁指数的区域分布曲线图如 5.13 所示，可以看出 WebMind 系统在心理健康维度抑郁指数预测方面与 SDS 抑郁自测量表存在一定的区域一致性。

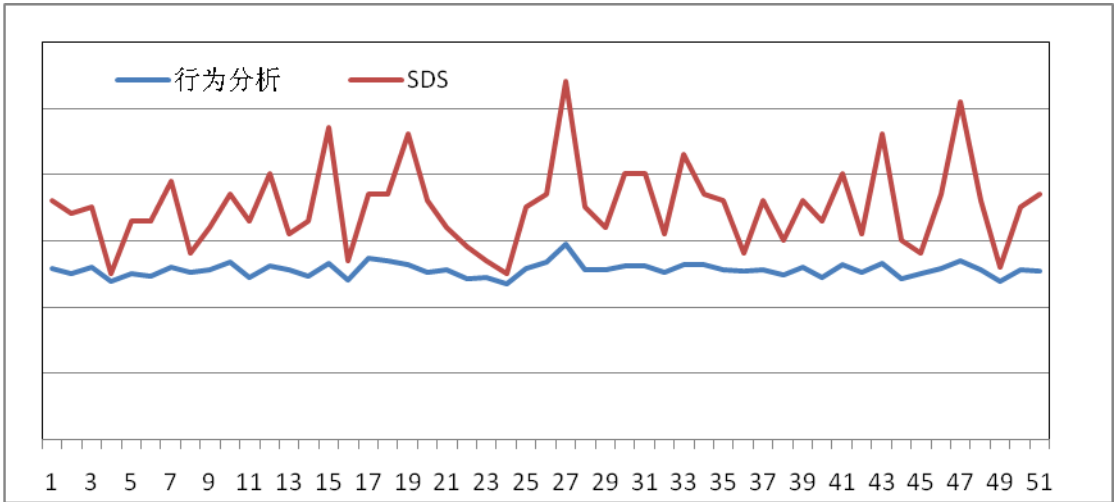


图 5.13 抑郁曲线图

**WebMind 系统用户使用反馈分析：**WebMind 系统采用三种方式对用户进行心理调节建议的推荐：随机推荐、依据心理变化幅度大小推荐和依据心理指标数值高低进行推荐。

WebMind 系统每次随机给出一种推荐方式，我们通过两个星期的测试，共收

到 672 条有效的反馈建议。对应于三种推荐模式，具体的情况见表格 5.3

表格 5.3 用户体检反馈结果

推荐模型	用户体验反馈结果				
	不好	一般	比较好	好	好评率
随机模型	68	94	26	27	24.6%
维度分值	37	35	87	79	69.7%

通过表格 14 可以看出，通过心理变化幅度和心理健康维度分值进行的干预调节相对于随机推荐的干预调节，更加符合网络用户的需求，更能得到用户的认可。三种推荐模式的对比见图 5.14，其中，纵坐标是该评价得分在对应模型中所占的百分比。

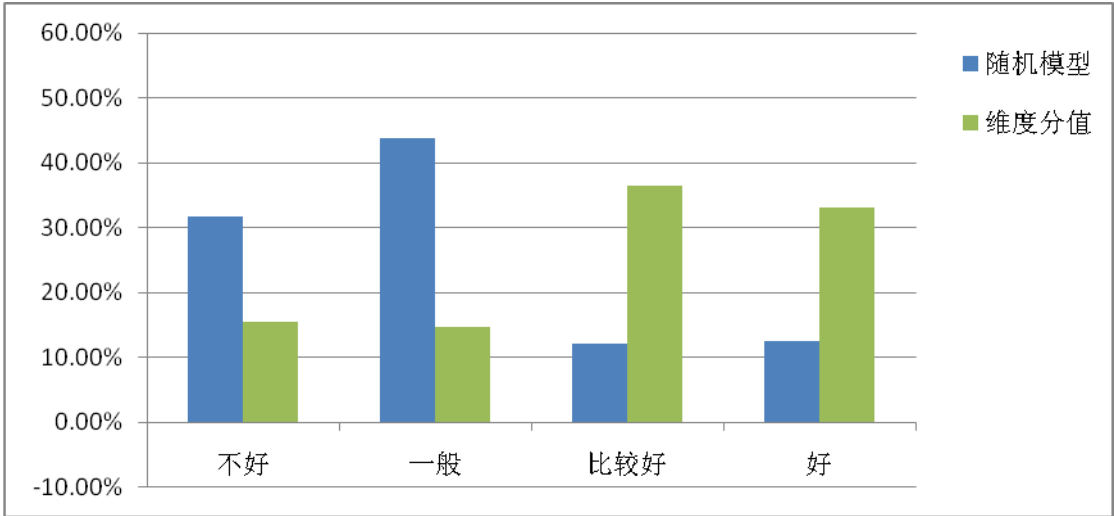


图 5.14 用户反馈结果示意图

综上，通过分析心理健康状况评估效果和 WebMind 系统用户使用反馈，我们可以得出 WebMind 系统能够及时和较为有效的跟踪心理健康状况，并能够给出用户满意的干预调节建议，取得了满意的理论和应用价值。

# 第 6 章 系统测试

软件测试是一项十分复杂多样化的工作。它在软件的整个生存期中占据重要的位置，对软件可靠性保证具有极其重要的意义。下面将结合本系统软件开发对系统进行了如下几个方面的测试。

## 6.1 测试方案

软件测试是为了发现错误而执行程序的过程。软件测试在软件生存期中横跨两个阶段：通常在编写完每一个类或功能模块之后就对其做必要的测试，即单元测试。编码和单元测试属于软件生存期中的同一个阶段。在结束这个阶段后对软件系统还要进行各种综合测试，这是软件生存期的另一个独立阶段，即测试阶段。

在 WebMind 系统中，主要是对浏览器模块、网络历史行为分析模块、心理建议推荐模块进行了测试。主要采用了黑盒测试，对功能和性能分别进行了测试。测试了各功能点的精确性、稳定性、强壮性，以及界面美观性，以保证该系统的正常的运行。

为了测试 WebMind 系统的推荐性能，以更新心理调节建议库，优化推荐模型，实验室安排了 130 名研究生院的学生(选我导师课的学生)进行了为期两周的测试。

## 6.2 测试内容

下面就按模块来叙述测试的内容，并给出相应的测试用例表。

### 6.2.1 浏览器模块

浏览器模块的测试基本就是测试浏览器的常用功能，由于浏览器模块测试内容过多，考虑到篇幅影响，只选重点测试用例，其具体测试用例内容如下面所示。

#### (1) 网页请求

表 6.1 网页请求功能测试用例

1	网页浏览 功能	输入想要访问网页的地址， 并按‘Enter’键	WebMind 系统标签页内容显示相应 请求网页内容	正常
---	------------	----------------------------	-------------------------------	----



2	多标签页功能	双击多标签页按钮	新建一个标签页，并将焦点放到新建标签页上	正常
3	网页的滚动条	滚动鼠标滚轮	网页相应的向上或者向下滚动	正常

结论：与网页有关的操作都能够照常运行，并遵守用户友好的原则，符合预期结果。

（2）测试菜单栏的文件选项功能

表 6.2 浏览器模块测试菜单栏的文件选项测试用例表

序号	测试内容	测试方法	预期结果	测试结果
1	文件->属性	点击文件，显示下拉菜单里的“属性”功能	弹出一个当前网页的属性对话框	正常
2	文件->新建标签页	点击文件下拉菜单的“新建标签页”功能	WebMind 浏览器打开一个新的标签页，并将焦点放在新标签页里	正常
3	文件->打印预览	点击文件下拉菜单里的“打印”功能	显示当前网页的打印预览模式	正常
4	文件->关闭标签页	点击文件，显示下拉菜单里的“关闭标签页”	关闭当前标签页，转移焦点到相邻标签页上，如果没有标签页了，就显示没有标签页状态	正常

结论：文件选项里的属性、新建标签页、打印预览、关闭标签页等操作都能够得到 WebMind 系统正确响应并处理，与用户的使用习惯一致，符合预期结果。

（3）测试菜单收藏夹选项测试用例。

表 6.3 浏览器模块菜单栏收藏夹选项测试用例表

序号	测试内容	测试方法	预期结果	测试结果
1	收藏夹->添加到收藏	点击收藏夹，显示下拉菜单里的“添加到收藏夹”功能	弹出“添加到收藏夹”对话框，让用户将当前网页添加到收藏夹	正常
2	收藏夹->整	点击收藏夹下拉菜单的“整	弹出“整理收藏夹”对话框，供用户	正常

	理收藏夹	理收藏夹”功能	整理收藏夹（添加或者删除）。	
--	------	---------	----------------	--

结论：收藏夹的所有操作都能够得到正确处理，符合预期结果。

6.2.2 浏览器模块

WebMind 系统测试包括性能测试和效果测试。性能测试的对象是浏览器的基本功能，比如，网页浏览功能，多标签页功能，收藏夹功能，历史记录功能等常用浏览器拥有的功能；效果测试的对象是针对系统提供的用户心理趋势图和用户心理调节建议，在实现系统时，用一个 xml 文件来实现用户心理调节建议。

```
<Record Time="2011-05-24 22:46:16">
  <Category>1</Category>
  <Suggestion>21</Suggestion>
  <Scoring>2</Scoring>
</Record>
<Record Time="2011-05-24 22:46:08">
  <Category>1</Category>
  <Suggestion>21</Suggestion>
  <Scoring>3</Scoring>
</Record>
```

图 6.1 用户对心理建议推荐模型的评价

6.3 测试结果反馈

经过 130 名研究生院的学生两个星期的测试，分析得出了测试反馈结果。  
性能测试结果参考图 6.2。

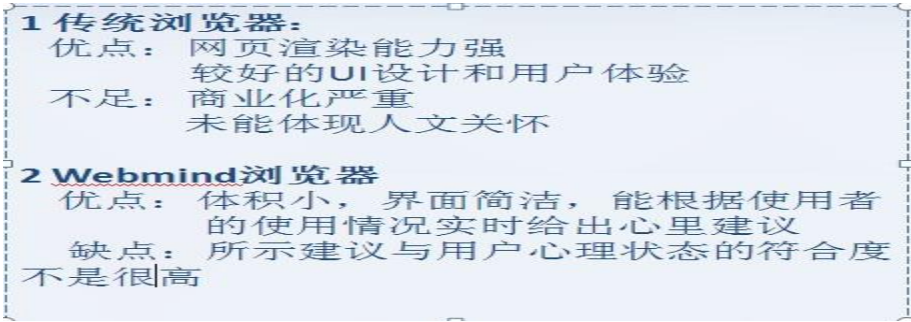


图 6.2 用户反馈性能测试结果

效果测试结果，可参考图 6.3。

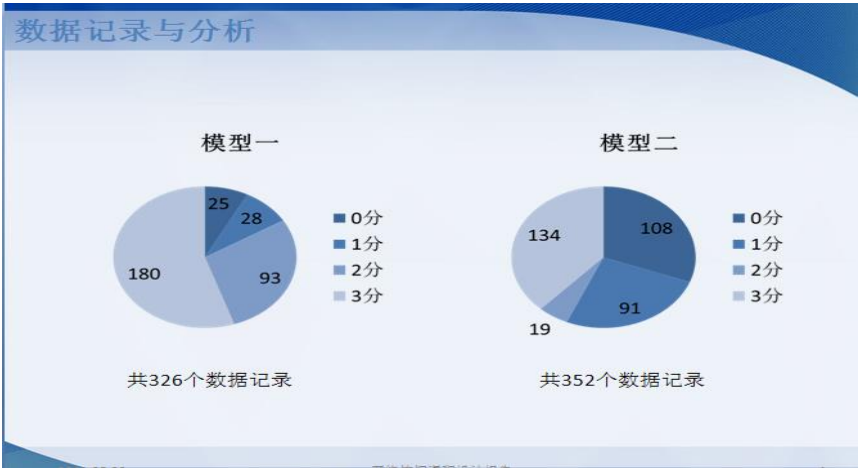


图 6.3 用户反馈效果测试结果

其中，模型一是心理维度值变化推荐模型，模型二完全随机推荐模型。

由表项可以看出，在误差可控的情况下，1 类模型的得分显著优于 2 类模型。

尽管 2 类模型得分比较均衡，但是其优差得分极差不大，并且出现大幅度得分差的情况，而 1 类优差极差明显。且优得分大幅度领先。由此可以分析出 1 类模型更加合理。

尽管结果表明心理维度值变化推荐模型更好，但好评率还是偏低，不到 60%，所以需要进一步去更新心理调节建议库，同时去优化心理调节推荐模型。

## 第 7 章 结论与展望

随着互联网的发展，通过网络作为媒介来治疗人们的心理问题将是必然的趋势。其中，通过分析网络行为来预测用户心理健康问题是目前新兴的研究方向，由于它的客观性和可操作性，这种研究方向将是心理学研究的热点。而建立用户网络行为和心理健康状况的相关性模型则是实现的关键环节，相关性模型的优劣，将会直接导致预测结果的好坏。可是，建立相关性建模非常复杂，需要长久时间的数据积累和大量实验的验证分析。因此，本文只是为通过行为分析预测心

理健康提供了一个可行性的论证，本文建立相关性模型，需要随着用户数据的积累不断进行更新。最后，实现了一个对绝大多数人群都适应的用户网络行为和心理健康相关性模型。

## 7.1 论文已完成的工作

本论文在深入调研国内外关于心理学和行为分析研究现状的基础上，探讨了用户网络行为和心理健康状况存在一定的相关性，特别是网页内容的情感分析，更能够体现网络用户的心理健康，为以后网络行为分析指明了一个新的方向。

在实验基础上，建立了一个网络行为和心理健康的相关性模型，并依据此模型提出了一种新型的心理健康测验手段——基于网络用户行为分析的心理健康预测。同时，我也设计完成了一个通过网络行为分析来预测心理健康，并给出调节建议的实验平台——WebMind 系统。通过分析 WebMind 系统的实验数据，发现我们提出新的心理健康测验手段能够及时高效的跟踪心理健康症状，给出用户满意的调节建议，取得了满意的理论和使用反馈。

同时，本论文提出的通过网络行为来预测心理健康的思路，搭建了心理学和信息科学之间的桥梁，为两个行业的发展都提供了新的契机。本论文的尝试，也为相关领域的发展提供了实证性研究。

## 7.2 下一步研究工作

本文对用户网络行为预测心理健康的各个环节都进行了详细的论述，也提出了一些新的思路、观点和方法。但是由于作者本身的精力有限以及时间的关系，研究中不可避免的存在着一些不完善的地方，今后对这些内容还有待进一步的研究。基于目前的研究现状，下一步研究的内容如下：

- 1) 初始模型的局限性：由于大规模收集网络使用行为的难度，初始的网络行为模型数据是通过问卷形式收集的。以后的模型需要依据客户端收集的网络行为数据逐渐更新。
- 2) 网络行为收集的局限性：网络行为分为多种，每类网络行为对于网络用户的心理健康状况的影响是不一样的，我们需要逐步改进对于网络行为特征提取，找出最能反映心理健康状况的网络行为。
- 3) 分析的结果是 Logistic 线性回归模型，只是利用测试人群的各项心理健康指标，建立一个常模，现有的模型只是能够给出网络用户在当前常模的水平（高分组、低分组）。这个常模受限于测试数据，有很大的局限性。

- 4) 对于相关性模型，需要实现自我学习和完善，达到个性化的程度。

## 参考文献

1. H. Meltzer, R. Gatward, R. Goodman, T. Ford. Mental health of children and adolescents in great Britain, International Review of Psychiatry 15(1-2) (2003) 185-187.
2. 邱培媛, 杨洋, 吴芳等. 国内外流动人口心理健康研究进展及启示, 中国心理卫生杂志 24 (2010) 64-68.
3. 罗鸣春, 中国青少年心理健康服务需求现状研究, Ph.D. thesis, 西南大学,

(2010).

4. 中国白领健康调查研究报告.

<http://wenku.baidu.com/view/98553222192e45361066f576.html>.2010.8.

3. 固犇.信息过载问题及其研究,中国图书馆学报（双月刊）,2000,5:43-46

4. D. Manasian. Digital dilemmas: a survey of the internet society, Economist Jan.25 (2003) 1-26.

5. 宋大伟. 网络心理咨询的局限与应用对策研究. 河南师范大学学报(哲学社会科学版), 2008, 35(5): 221-223

6. 朱晓红, 蒋波. 国外计算机和网络心理治疗评介[J],中国临床心理学杂志, 2010,18(1): 130-133

7. 李涛. 虚拟现实技术在心理治疗中的应用[J], 心理科学, 2005,28(3): 654-655

8. 崔丽霞, 雷雳, 蔺雯雯等. 网络心理咨询的疗效与展望, 心理科学进展, 2007, 15(2): 350-357

9. 刘鹏. 网络用户行为分析的若干问题研究[D].北京: 北京邮电大学, 2010.

10.赵瑞刚, 黄家林. 校园网常见网络行为特征分析的研究, 网络安全技术与应用, 2010,10:34-36

11. .NET Framework. <http://baike.baidu.com/view/18370.htm>, 2010,10.

12. ZedGraph 控件介绍.

<http://hi.baidu.com/christole/blog/item/184f7e55288cb40b3b29355d.html>. 2011.4.

13. 王思力,张华平,王斌. 双数组 Trie 树算法优化及其应用研究[J], 中文信息学报, 2006,20(5): 24-30

14. 赵欢, 朱红权. 基于双数组 Trie 树中文分词研究[J], 湖南大学学报自然科学版, 2009, 36(5): 77-80

15. 李祥, 周波. 一种基于朴素贝叶斯分类的性能预测方法[j],计算机应用与软件, 2011,28(1): 231-290

16. Domingos, Pedro & Michael Pazzani (1997). On the optimality of the simple Bayesian classifier under zero-one loss, Machine Learning, 29:103–137

17. 杨鼎, 阳爱民. 一种基于情感词典和朴素贝叶斯的中文文本情感分类方法. 计算机应用研究, 2010, 27(10): 3737-3743.

18. 李军。中文评论的褒贬义分类试验研究 [D]。北京: 清华大学, 2008.
19. 王素格, 魏英杰。停用词表对中文文本情感分类的影响. 情报学报, 2008, 27( 2) : 175-179.

## 致 谢

衷心地感谢我的毕业设计导师于瑞云老师和我的研究生导师朱廷劭博士。他们在繁忙的工作中抽出宝贵的时间给我深切的关怀和悉心的指导,使我得以顺利地完成了本科毕业设计阶段的学习和研究工作。两位导师渊博的学识、敏锐的洞察力、严谨的治学态度、高尚的师德和慈善的心怀,让我如沐春风,使我受益匪浅。在两位导师身上,我不仅仅学到了扎实、宽广的专业知识,也学到了做人的道理,这将是受益终生的财富。在此,我要向两位导师致以最衷心的感谢和深深地祝福。

在读本科期间,我的周围是一群风华正茂、意气风发的有志青年,与之朝

夕相处，受益颇丰。毕设阶段，师兄、师姐等 WSI 实验室的全体同学都给予了我极大的帮助，祝愿他们事业有成，一生平安。

衷心感在本科期间教授过我课程的每一位老师。在我本科在读期间，他们给了我很多方便和帮助。让我最为感激的是东北大学软件学院，她拥有良好的学习氛围和求真务实的学术作风，使我终身受益，伴我成长，催我奋进。

深深感谢含辛茹苦养育我、为我的成长付出无数的心血、不断地鼓励和支持我上进的父母亲。今后我将竭尽所能，加倍补偿这份一辈子也还不清的深情。

最后要感谢所有帮助过我的人们，我不能在这里尽数你们的名字，但我会在我心中永记你们的恩惠。

聂栋

2011 年 5 月 于北京