

学校编码: 10384
学号: 31520091152835

分类号 _____ 密级 _____
UDC _____

廈門大學

硕士学位论文

基于灵敏度的中医人体亚健康检测研究
Research on Sub-health Diagnosis Based on Sensitivity
Theory

王翌

指导教师姓名: 李绍滋 教授

专业名称: 计算机技术

论文提交时间: 2012年 5月

论文答辩日期: 2012年 6月

学位授予日期: 2012年 月

答辩委员会主席: 陈国龙

评阅人: _____

2012年 5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(中医学健康检测)课题(组)的研究成果,获得(中医学健康检测)课题(组)经费或实验室的资助,在(实验)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): 王磊

2012年6月3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): 王翌

2012 年 6 月 3 日

厦门大学博硕士学位论文摘要库

摘要

当今社会人们的生活压力越来越大，这导致越来越多的人受到了亚健康 and 疾病的困扰，如今亚健康这一名词已经受到了世界性的广泛关注。然而，亚健康所描述的是健康与疾病之间的中间状态，其本身并没有明显的临床病症表现，因此很难被个体察觉，甚至在常规的检查中也很难被注意到，以西医为代表的现代医学往往对其束手无策，只能等到有实质性病变时才能进行治疗，因而研究人员正在试图将中医理论运用到亚健康的检测中。中医（TCM）指中国传统医学，是研究人体生理、病理，以及疾病的诊断和防治等的一门学科，在中医理论中，亚健康被定义为健康与疾病中的中间状态，能够被 13 个类别的中医证素所反映。中医使用四诊（望闻问切）诊断法对人体进行检测。四诊能在不给患者造成任何不适和伤害的情况下为医生提供了患者的大量外部特征，从而帮助医生对病人进行身体状况的判断。

但目前利用中医原理进行亚健康检测的研究仍然处于起步阶段，目前还存在诊断主观性，特征单一性，诊断数据良莠不齐等问题。针对这些问题，本文提出了一种基于中医理论的人体亚健康自动诊断方法，论文的主要工作和创新点如下：

(1)针对当前亚健康数据集相对缺乏的不足，建立了厦门大学亚健康数据库。数据库是亚健康检测基本而又重要的组成部分，数据库的好坏直接影响到分类器的“优劣”。本文建立的亚健康数据库具有分辨率较高、样本丰富、特征齐全的特点。为亚健康检测的研究提供了丰富的素材。

(2)针对亚健康特征获取单一的不足，提取了舌部、眼部、脸部等中医诊断中常用的人体部位的视觉信息，并结合心理和生理量表，提取了 77 维亚健康特征，从静态和动态、生理和心理的方面全面的反映了人体的健康状态。实验证明，高维的特征能够更好的描述人体的健康状态，有效防止过拟合，提高分类器的学习性能，进而优化实验结果。

(3)针对中医诊断中存在的主观性、多样性和不确定性的不足，本文提出了灵敏度理论，灵敏度可以客观的衡量特征及诊断数据的性能，从而作为选择特征和诊断数据的标准。本文进而在其之上提出基于灵敏度的特征选择方法和基于灵敏度的诊断数据融合方法，基于灵敏度的特征选择方法可以有效剔除无用特征，降低特征维度，提高分类效率。基于灵敏度的诊断数据融合方法可以综合不同专

家诊断结果，克服由单位专家诊断数据所造成的“偏科”现象。实验证明，使用基于灵敏度的特征选择方法和基于灵敏度的诊断数据融合方法可以有效的提高分类器分类效果。

总之，本文从特征提取以及特征选择与诊断数据融合这两个中医亚健康检测的核心问题入手，首先采用多部位的特征提取方式，提取丰富的亚健康特征，然后针对中医亚健康诊断的不确定性和主观性问题，提出了一种基于灵敏度的特征选择与诊断数据融合的方法，实验结果表明：多部位亚健康特征以及基于灵敏度的特征选择和诊断数据融合方法能有效的提高亚健康预测的准确率，相关性和剩余方差等性能，其结果有利于相关学科不断向前发展，并促进其应用价值的进一步实用化。

关键词：灵敏度；亚健康；中医；神经网络

Abstract

In modern society, cause of the stronger and stronger pressure people suffer, more and more people are facing the problem of sub-health and sickness, the word 'sub-health' draws a worldwide attention in medical already. However, since sub-health is a condition between health and illness, it is hard to be checked all by oneself. In addition, it is difficult to diagnose by routine physical exam. Based on this situation, researchers have put their attentions to sub-health diagnosis on other methods like the Traditional Chinese medicine (TCM). TCM is the prevention, diagnosis, and treatment of human diseases. In TCM theory, sub-health is an intermediate state between health and disease. The Sub-health can be represented by TCM syndromes, which is categorized into 13 classes. TCM doctors use four means called "Sizhen" for diagnosis. "Sizhen" can provide the TCM doctors good exterior information of the patients, while don't cause them any discomfort.

However, the research of the sub-health diagnosis based on TCM theory is still at its early stage, Some problems like the subjectivity of the diagnosis, the inabundance of the features, and the quality of the diagnosis data varying greatly are still exist. To solve these problems, we propose a TCM based sub-health diagnosis system, the major works and contributions are summarized as follow:

(1)To solve the lacking of the sub-health data problem, we build the Xiamen University sub-health database. The database is a basic and important part of the system. The quality of the database will affect the classifier's result directly. The database we build in this paper has high resolution ratio, abundant samples and plenty of features, providing the sub-health diagnosis system plentiful materials.

(2)To solve the inabundance of the features problem, we extract the features that always used in TCM diagnosis which includes the visual information of the tongue, the eyes and the face. We also extract the features from the physical and psychological scale, we totally extract 77 features to describe the health of the body from not only static and dynamic ways, but also physical and psychological ways. The experiment results show that, the high dimension features can describe people's health state

greatly, and avoid over fitting while training the classifier and improve the performance of the classifier, thus gets better results in the end.

(3) To solve the subjectivity, the diversity and the uncertainty of the TCM diagnosis, we propose the sensitivity theory. The sensitivity can measure the quality of the features and diagnosis data objectively, as the standard of feature selection and diagnosis data fusion. Moreover, we propose a sensitivity based feature selection method and the diagnosis data fusion method. The sensitivity based feature selection method can delete the useless feature, reduce the feature dimension and improve the efficiency of the classification. The sensitivity based diagnosis data fusion method can fuse the diagnosis data from different experts to void the incomprehensive abilities of them. The experiment results show that the sensitivity based feature selection method and the sensitivity based diagnosis data fusion method can improve the performance of the classifier effectively.

In sum, we start the problem from the sub-health feature extraction; the feature selection and the diagnosis data fusion which are two main problems in TCM based sub-health diagnosis. Firstly, we extract abundant features from different parts of human body. Secondly, we propose the sensitivity based feature selection method and the sensitivity based diagnosis data fusion method to solve the uncertainty and subjectivity of the TCM based sub-health diagnosis. The experiment results show that the feature extraction from different parts of body, the sensitivity based feature selection method and the sensitivity based diagnosis data fusion method can improve the system's inference performance on the accuracy, correlation and residual variance. The result of this paper can help to improve the technology of the sub-health diagnosis in a way.

Key Words: sensitivity; sub-health; TCM; neural networks

目录

摘要.....	I
ABSTRACT.....	III
目录.....	V
CONTENTS.....	VII
第一章绪论	1
1.1 研究背景	1
1.2 国内外研究概况	3
1.3 本文的主要研究工作及组织结构	4
1.3.1 本文的主要研究工作及创新	4
1.3.2 本文的组织结构	6
第二章灵敏度理论	7
2.1 灵敏度的定义	7
2.2 基于灵敏度的特征选择	9
2.3 基于灵敏度的诊断数据融合	10
2.4 本章小结	11
第三章亚健康数据库	13
3.1 亚健康数据库内容选择依据	13
3.2 亚健康数据库采集说明	15
3.3 本章小结	20
第四章数据预处理与特征提取	21
4.1 面部图像	21
4.1.1 haar-like 特征和积分图	21

4.1.2 彩色模型.....	23
4.1.3 面部区域划分及特征提取.....	26
4.2 眼部区域.....	28
4.2.1 光流场.....	28
4.2.2 眼部区域分割及特征提取.....	29
4.3 舌部区域.....	30
4.4 本章小结.....	32
第五章基于灵敏度的中医人体亚健康在线自动诊断系统.....	33
5.1 BP 神经网络.....	33
5.2 实验流程.....	34
5.3 中医人体亚健康在线自动诊断系统的界面与说明.....	39
5.4 本章小结.....	44
第六章总结与展望.....	45
6.1 现有工作总结.....	45
6.2 未来研究展望.....	46
6.3 本章小结.....	48
参考文献.....	49
致谢.....	54
附录攻读硕士学位期间发表的论文.....	56

Contents

Abstract in Chinese	I
Abstract in English	III
Contents in Chinese	V
Contents in English	VII
Chapter 1 Introduction	1
1.1 Research Background.....	1
1.2 Survey of Home and Abroad	3
1.3 Major Works and Framework.....	4
1.3.1 Major Works and Innovation	4
1.3.2 Framework	6
Chapter 2 Sensitivity Theory	7
2.1 The Definition of Sensitivity	7
2.2 Sensitivity-based Feature Selection.....	9
2.3 Sensitivity-based Diagnosis Data Fusion	10
2.4 Conclusion of This Chapter.....	11
Chapter 3 The Sub-health Database	13
3.1 The Reason for the Database Collecting	13
3.2 Introduction of the Collection of the Database	15
3.3 Conclusion of This Chapter.....	20
Chapter 4 Data Pretreatment and Feature extraction	21
4.1 Facial Part.....	21
4.1.1 Haar-like Feature and Integral Image	21

4.1.2 Color Moduls	23
4.1.3 Region Segmentation and Feature Extraction of the Facal Part	26
4.2 Eyes' Part.....	28
4.2.1 Optimal Flow Field	28
4.2.2Region Segmentation and Feature Extraction of the Eyes' Part	29
4.3 Region Segmentation and Feature Extraction of the Tongue's Part.....	30
4.4 Conclusion of This Chapter.....	32
Chapter 5 The Sensitivity-based Sub-health Diagnosis System	33
5.1BP Neural Networks.....	33
5.2The Measurement Processes	34
5.3The Introduction of the System	39
5.4 Conclusion of This Chapter.....	45
Chapter 6 Conclusion and Prospect.....	46
6.1 Conclusion of This Paper	46
6.2 Prospect	47
6.3 Conclusion of This Chapter.....	49
References.....	50
Acknowledgements	55
Appendix Publications.....	57

第一章 绪论

1.1 研究背景

WHO（世界卫生组织）将机体无器质性病变，但是有部分功能改变的非健康状态定义为“第三状态”，在我国，我们则将这种状态称为“亚健康”状态。亚健康所代表的是一种临界状态，处于亚健康状态的人，虽然生理或心理上并没有明确的疾病，但却会出现精神活力和适应能力的下降，这种状态如果无法得到及时的纠正，则非常容易引起后继的心身疾病。在当今社会中，人们的生活压力越来越大，这导致越来越多的人受到了亚健康和疾病的困扰，而随着人们对健康的关注日益加深，如今亚健康这一名词已经受到了世界性的广泛关注，在上个世纪末甚至被冠以了“世纪疾病”的称号。据统计，按照 WHO 于 1978 年所颁布的健康标准来评测，全球人口中仅有 5% 符合健康标准，20% 处于疾病状态，而其余高达 75% 的人口都面临着亚健康的困扰[1]，在全部人口当中，中青年人口、城市人口、脑力劳动者和白领阶层由于生活无规律和缺乏适量运动等原因，都已成为了亚健康的高发人群。然而，亚健康所描述的是健康与疾病之间的中间状态，其本身并没有明显的临床病症表现，因此难以被个体察觉，甚至在常规的医疗检查中也很难被注意到以西医为代表的现代医学往往对其束手无策，只能等到有实质性病变时才能进行治疗[2]。

由于不同人体自身体质、年龄、适应能力、文化背景等个体特征的不同，处于亚健康状态中的人的临床表现往往也不尽相同。比较常见的临床表现有容易疲劳、精力不济、抵抗力下降、易患感冒、头痛虚汗、食欲不振、失眠焦虑、萎靡易怒、人际关系不协调以及家庭关系不和睦等[11]。

亚健康虽然只是一种未病的临界状态而非真正的疾病，但其实已经是向身体发出了警告，倘若这种状态不能得到及时的纠正，则很容易影响到患者的正常工作生活，甚至引起其他身心疾病，包括：心理障碍、胃肠道疾病、高血压、冠心病、癌症、肺部疾病、糖尿病等重大疾病[23]。

中医（TCM）是指中国传统医学，是一门研究人体生理、病理，以及疾病的诊断和防治等的学科，至今已经有 2500 多年的历史。它承载着中国古代人民同疾病作斗争的经验和理论知识，是在古代朴素的唯物论和自发的辩证法思想指

导下,通过长期医疗实践逐步形成并发展成的医学理论体系。[20]在研究方法上,以整体观、相似观为主导思想,以脏腑经络的生理、病理为基础,以辨证论治为诊疗依据,具有朴素的系统论、控制论,分形论和信息论内容。

养生素来是中医所倡导的理念,“不治已病治未病”的防病养生谋略早在《黄帝内经》中就被正式提出,它的出现标志着中医将养生防病作为其主导思想,也是我国卫生界所遵守的“预防为主”战略所参照的最早思想,它包括未病需先防、已病需防变、已变需防渐等多个方面的内容,这就要求人们不但要治病,而且要防病,不但要防病,而且也要注意阻挡病变发生的趋势、并在病变未产生之前就想好能够采用的救急方法,这样才能全面的掌握面对疾病的主动权,也就能达到“治病十全”的“上工之术”[28]。朱震亨在《格致余论》中说:“与其求疗于有病之后,不若摄养于无疾之先;盖疾成而后药者,徒劳而已、是故已病而不治,所以为医家之怯;未病而先治,所以明摄生之理。如是则思患而预防之者,何患之有哉?此圣人不治已病治未病之意也。”[40]

实际上,亚健康就相当于中医当中的“未病”这一概念,“未病”一词最早出自《黄帝内经》中的《素问·四气调神论》篇。文中有“是故圣人不治已病治未病,不治已乱治未乱,此之谓也。夫病已成而后药之,乱已成而后治之,譬犹渴而穿井,斗而铸锥,不亦晚乎!”[36]。这段话就是从正、反两个方面强调了治未病的重要性,已成为预防医学的座右铭。

除此之外,《黄帝内经》中出现“治未病”一词的还有另外2篇。《素问·刺热篇》内记载:“病虽未发,见赤色者刺之,名曰治未病”[44]。这里的“未发”,类似于唐代孙思邈所说的“欲病”,实际上就是指已经有先兆小疾存在,也就是疾病时期中症状较少且又较轻的阶段。在这种情况下,及时发现,早期诊断治疗无疑起着决定性的作用。《灵枢经·逆顺》篇中也记载到:“上工刺其未生者也;其次,刺其未盛者也,……上工治未病,不治已病,此之谓也”。即最好的医生在病灶还没有发作之前,就可以开始着手于预防和治疗了[41]。两篇均强调在疾病发作之先便把握时机,予以治疗,从而达到“治未病”的目的。

中医理论认为,人体内在的健康状态可以由人体的外在生理表征表现出来。因此,中医医生通常使用“四诊”即:“望”、“闻”、“问”、“切”来进行诊断。四诊具有朴素性和直观性的特点,可以在医生感官所及的范围内,直接的获取患者

信息，随后医生即刻进行分析综合，及时作出判断。四诊的基本原理是建立在整体观念和恒动观念的基础上的，是阴阳五行、藏象经络、病因病机等基础理论的具体运用。四诊的理论基础是物质世界的统一性和普遍联系[45]。在现代中医诊断中，四诊中的“望”、“问”两诊使用的最为频繁，因为这两种方法相对更为准确，能够为医生提供丰富的病人外部信息，整个诊断过程也不会为患者带来任何的不适感。这种非侵入性的诊断方式非常符合二十一世纪大家所提倡的无痛苦、无伤害的医学理念[47]。

1.2 国内外研究概况

随着亚健康受到越来越多人的重视，使用计算机技术来进行亚健康方面的研究也变得越来越流行，但使用图像技术或其他多媒体技术去解决亚健康问题的相关文章依然不多。其主要原因在于，要结合多媒体技术解决亚健康问题需要研究人员同时掌握计算机技术和医学知识，而具备这样能力的研究人员为数尚不多。目前绝大多数亚健康检测方面的成就都依赖于传感器技术和血液检测。这两种方式所依靠的传感器技术和西医理论都相对较为成熟[5][22][25][46]，而基于传统中医的技术方面由于一直存在有主观性的问题，发展较为缓慢。文献[1]阐述了利用中医来解决亚健康诊断问题的可能性和优势。文献[2]对亚健康和中医进行了量化研究。文献[12][32][49][50][55]对面部轮廓提取和区域划分进行了一系列的而研究。文献[3]利用患者的眼部视觉特征建立了亚健康自测系统。文献[4]运用多特征 CBIR 方法进行舌体检测，文献[21]研究了对舌部视频中舌头的自动跟踪方法。文献[5]总结了中医脉象传感器的研究成果。文献[6]构造了一种基于中医的脉搏获取装置。文献[7]利用中医理论建立了一个专家系统，然而，对于症状的评测仍然由受试者自己评定，因此，中医诊断中的主观性依然没有得到很好的解决。文献[8]使用 SVM 分类器对健康和亚健康人群的腕部脉象特征进行了分类。然而，这些方法都忽视了中医检测中的不确定性和主观性问题，从而影响了实际应用。文献[24]建立了一种内置于服装中的基于 ECG 的亚健康智能诊断系统。文献[12]提出了一种基于双椭圆可变形轮廓理论的舌部自动分割方法。文献[13]利用舌部信息完成身体状况分析检测工作。文献[27]建立了一个基于模糊集理论[15]和 BP 神经网络的亚健康诊断系统。文献[28]研究了亚健康人群的普遍特

征及普遍症状表现。文献[9][10]结合多媒体信息和中医医案来进行亚健康检测，在一定程度上考虑了不确定性和主观性的问题。在文献[9]中，郭锋提出了一种是用面部、眼部视频以及生理心理量表进行亚健康检测的方法，使用云技术[14][16]进行推导，可以推导出受试者的亚健康程度及其对应的13种中医证素。然而，如何选择以及选择多少特征及诊断数据这一问题却没有得到解决。随后，在文献[10]中，郭锋提出了“基于一致性的规则挖掘方法”[29][34][35][37]来解决特征选择和数据选择的问题。这种方法在一定程度上解决特征选择和数据选择的问题，然而，一致性对于数据融合存在一定的限制，同时这种方法容易导致过拟合现象，对诊断效果有较大的影响。

1.3 本文的主要研究工作及组织结构

1.3.1 本文的主要研究工作及创新

本文围绕中医亚健康自动诊断性能的改进以及“灵敏度”理论在中医亚健康诊断中的应用开展研究，做了以下的研究工作：

- 1 对若干种机器学习分类器进行了全面的归纳和总结，分析了这些分类器的优缺点和适用范围，为本论文分类器的选择提供了理论的支持。
- 2 提出了“灵敏度”理论，基于“灵敏度”理论的特征选择方法和诊断数据融合方法。

由于中医医生水平的良莠不齐以及中医诊断中不可避免的主观性，中医诊断的结果往往充满多样性和不确定性。为了克服中医诊断中存在的主观性、多样性和不确定性，本文从三个方面进行创新。

- (1) 大量的特征和诊断数据优劣无法得到客观的评价，鱼龙混杂，有用的数据和无用的数据无法得到有效的区分。针对这个问题，本文提出了“灵敏度”理论，灵敏度可以有效的反映特征改变量与类别值改变量这两者之间的变化关联程度，为评测特征及诊断数据的优劣程度提供了形象、客观的评价标准。
- (2) 过高维数的特征不但会影响整体系统的运行速度，其中无用的特征对系统性能不但没有益处甚至还会降低系统的诊

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