



# 智能系统学报

CAAI TRANSACTIONS ON INTELLIGENT SYSTEMS

## 智慧医疗促进分级诊疗——DUCG的初步实践

张勤

引用本文:

张勤. 智慧医疗促进分级诊疗——DUCG的初步实践[J]. 智能系统学报, 2020, 15(5): 1028–1032.

ZHANG Qin. Medical intelligent system to promote graded clinical diagnoses and treatments: application of dynamic uncertain causality graph[J]. *CAAI Transactions on Intelligent Systems*, 2020, 15(5): 1028–1032.

在线阅读 View online: <https://dx.doi.org/10.11992/tis.202101008>

## 您可能感兴趣的其他文章

### 非完备决策信息系统中的不确定性度量

Uncertainty measure in incomplete decision information system

智能系统学报. 2019, 14(6): 1100–1110 <https://dx.doi.org/10.11992/tis.201905052>

### 面对智能导诊的个性化推荐算法

A personalized recommendation algorithm for intelligent guidance

智能系统学报. 2018, 13(3): 352–358 <https://dx.doi.org/10.11992/tis.201711036>

### AI——人类社会发展的加速器

Artificial intelligence: an accelerator for the development of human society

智能系统学报. 2017, 12(5): 583–589 <https://dx.doi.org/10.11992/tis.201710016>

### 物联网中的智慧溯源服务系统Petri网建模与分析

Petri net modeling and analysis of an intelligent traceability service system based on the Internet of Things

智能系统学报. 2017, 12(4): 538–547 <https://dx.doi.org/10.11992/tis.201611031>

### 大数据情报分析发展机遇及其挑战

Opportunities and challenges of big data intelligence analysis

智能系统学报. 2016, 11(6): 719–727 <https://dx.doi.org/10.11992/tis.201610025>

### 基于用户移动轨迹的个性化健康建议推荐方法

Personalized recommendation algorithm of health advice based on the user's mobile trajectory

智能系统学报. 2016, 11(2): 264–271 <https://dx.doi.org/10.11992/tis.201511026>

微信公众平台



关注微信公众号，获取更多资讯信息

# 智慧医疗促进分级诊疗——DUCG 的初步实践

## Medical intelligent system to promote graded clinical diagnoses and treatments: application of dynamic uncertain causality graph

张勤<sup>1,2</sup>

(1. 清华大学核能与新能源技术研究院, 北京 100084; 2. 清华大学计算机系, 北京 100084)

我原来是搞核电的。核电站最重要的就是安全, 其中包含复杂工业系统的智能故障诊断问题。这个问题是 1988 年我在美国 UCLA 大学做访问学者的时候, 我的导师给我提出来的。那时美国三哩岛核事故刚刚发生不久。该事故提出了一个问题: 在有虚假信号、损失非常巨大、时间也非常紧迫、信号又动态急剧地变化的工况下, 人不可能正确地判断出故障到底在哪里, 就需要计算机进行辅助诊断。

有虚假信号存在的情况下, 如何做出正确的判断? 接受这个任务后, 我开始查阅各种人工智能的理论, 主要有两种: 一种是通过大数据机器学习, 另一种是传统的基于规则的专家系统。最后得出的结论是二者都不适合解决这个问题。核电站是高可靠系统, 出一次事故就会造成巨大损失, 而且每次出的都是新事故, 不具备重复性, 甚至连一个可重复利用的数据都没有, 所以没有可学习的数据, 如何进行人工智能的故障学习? 所以第一个模型——大数据机器学习模型就被否定了。第二个模型——基于规则的专家系统。可是在故障的情况下, 尤其是核电站, 很多知识都是不确定的, 规则的数量也非常巨大, 往往不知道从何入手, 写了后面的规则, 忘了前面的规则, 重叠、冲突、循环等在所难免, 也不直观。即知识的管理很困难; 另外, 它的不确定性传播是非常复杂的, 没有一个严谨的理论体系, 更不可能辨识出虚假信号这样的复杂工况, 以及负反馈这样的复杂的工况。怎么办? 我就只好自己创立了一套理论体系, 用于解决当时所提出来的这个问题, 经过一年的时间, 问题终于被解决。

这个理论发展到现在已经 30 多年, 我称之为动态不确定因果图 (dynamic uncertain causality graph, DUCG)。这个理论首先应用于我们的核电站、化工系统、航天系统的状态监测、故障预报、诊断、发展预测、决策支持和概率安全评价 (prob-

abilistic safety assessment, PSA)。我们进行了上百项的故障诊断, 到目前为止, 正确率是 100%, 但是还没有大规模应用。后来, 我们想到核电站是大型复杂系统, 其实人也是大型复杂系统, 既然 DUCG 理论可以用于解决复杂的核电站的故障诊断问题, 为什么不可以用于给人看病呢? 因此近些年来我们用 DUCG 理论来解决看病问题。

中国的看病难、看病贵是非常现实的问题, 主要原因在于优质的医生和医疗资源不够。

据国家卫健委 2018 年统计数据, 我国的医疗机构有 100 万个, 其中 95% 的医疗机构是基层, 也就是县级以下的。我们的三甲医院只占 0.14%。

最近 WHO 发布了一个数据, 每年由于误诊而死亡的人数大约是 260 万, 没有死但耽误了病情的就更多了。每个病人都希望能够向高水平医生求医, 以免被误诊漏诊。在这种情况下, 各国政府都面临着医改问题, 我国也像其他的国家一样, 多年以前就开始实施“分级诊疗”医疗体制改革, 总结成四句话就是“基层首诊, 双向转诊, 急慢分治, 上下联动”。这里关键的就是“基层首诊”, 基层如果不能准确诊断, 那么分级诊疗就失去了基础, 就不可能实现。

我们国家提出的目标是“90% 大病不出县”。主要的问题在于区县、乡镇、村和社区这些基层机构, 高水平医生是很难留住的。

国家派医学临床专家到西藏、青海、新疆、内蒙这些边远地区巡诊, 但专家的时间是有限的, 当专家赴外地巡诊, 有些专程来大医院就医的患者就挂不上号了; 远程医疗也会占用专家有限的时间。培养一名专家的时间是非常长的, 8 年的医学院学习, 再加上 20 年的临床实践不断的积累, 才有可能成为知名的专家。而且专家个人的能力是有限的, 他们分身乏术。

所以基层缺少好医生, 这个问题始终解决不了。于是我们国家又推出另外一项重大的计划,

就是全科医生培养。全科医生培养有两种系统,一个是“5+3”,即5年的本科全科学习,再加上3年的全科医生规范化培训,国家补贴,希望8年学习完成之后能够注册为全科医生,然后到基层去服务至少5年,后来减为2年(3年的规范化培训也算作是服务)。但全科学习难度大,每年的招生名额招不满、质量不佳。毕业之后愿意注册为全科医生的就更少,到乡下去,到穷困的地方去,到缺医少药的地方去的人就更少,所以这个问题其实是无解的。于是我们就想到用人工智能技术来解决这个问题。

我认为人工智能系统解决临床诊断问题应当具备四个要素。第一,必须要在任何场景都能够应用,不能只在三甲医院的训练集上是正确的,然后一到基层去就错误百出。第二,必须要准确,它的水平要比基层医生高得多,最好是90%以上。第三,既然是辅助诊断,就必须要有可解释性。一个没有可解释性的系统将无人敢用。包括核电站故障诊断系统,如果不能解释为什么诊断出的是这个结果,那么操作人员就不容易相信系统的诊断结果。而看病是人命关天的事,要让医生签字来承担医疗责任,他必须要了解人工智能系统给出的诊断结论是怎么得出来的。因为现在还不是以人工智能代替医生看病,而只能是辅助诊断。人工智能系统尚不能承担看病的法律责任。第四,智能看病不是所有的数据都已经收集好了,然后让智能系统来判断到底是什么病,而是要引导基层医生一步一步地去收集应当收集的证据,不该做的检查不要做,该做的检查不要漏掉,在准确收集证据的基础上看病,这样才能节省成本,并且准确地诊断出疾病。但是每位病人都有个体差异,每个人的临床检查路径都应当是不同的。所以人工智能系统应当能够根据每位病人的情况动态地生成一个“个体优化”的临床检查和诊断路径。只有满足这4个要求的人工智能系统,才是一个好的临床辅助诊断系统。

现在我们已经完成知识库建造和第三方测试验证的知识库覆盖36个主诉症状。DUCG的知识库都是以主诉为切入点。例如腹痛,病人并不知道自己该挂哪个科室,很多科室都有可能,因为腹痛病的范围很广,患者只知腹痛,这就是他的主诉。我们以每一个主诉为切入点,把所有相关科室的专家组织起来建造DUCG知识库,来看腹痛的病。这样的知识库我们建造并且完成了其内部测试,又完成了第三方测试,包括关节痛、呼

吸困难、呕血、鼻塞、鼻出血、咳嗽咳痰、发热伴皮疹、腹痛、便血等。

临床学告诉我们,在基层做临床看病大概有30多个主诉,把它分得细一点,可能会有四十多个主诉。我们已经完成了36个主诉,基本覆盖全科。经过第三方验证,正确率95%以上,其中少见病正确率也在80%以上,现已在山东胶州和重庆忠县上线应用。

我们是怎么做出来的呢?

首先我们有一个知识库编辑平台,放在云上,任何能上网的地方都能使用这个平台。与我们合作的专家主要集中在北京,包括协和医院、北大医院、宣武医院、朝阳医院、北京医院、安贞医院、阜外医院等,我们跟这些医院的主任级别的专家进行深度的合作,来建造我们的知识库。

我们的第三方验证主要有10家医院:四川遂宁市中心医院、内蒙呼和浩特市人民医院、解放军306医院、清华长庚医院、还有6家重庆的三甲医院。下面通过实例来介绍DUCG系统。

通过HIS系统点选“DUCG”进入DUCG智能临床辅助诊断系统,把病人姓名、年龄自动转入DUCG系统,还有病案号,以便查询。但不涉及隐私。页面左边有一个“主诉类”,选择并打开“主诉类”,选择要看的病——关节痛(这是个真实的病人),点击进入“关节痛”后,选择到底是哪些风险因素、症状和体征?点选之后,相应变量的颜色就会改变。每一个变量都有解释和说明。不确定的证据怎么办?选择不确定的隶属度,颜色变成紫色。

我们的算法可以进行精确的推理,在有不确定证据的情况下,我们的算法本身是精确的。为节省时间,我们直接把真实的病例证据调出来,但去掉了所有的化验和其他的检查结果,只保留风险因素,如:年龄、性别、家族史、症状、体征等,是社区医生通过简单的问话、观察和测量血压、心跳或者摸一摸他的骨关节大小就可以得到的信息。我们只把这些信息中的一部分,也就是建库专家认为必要的信息先输入进来,大多数是绿色的,表示阴性或正常,只有少部分是其他颜色(异常的)。有了这些必选项后,就可以避免漏诊误诊。

回答完必选问题之后就可以进行诊断了。点一下“诊断”按钮,不到一秒钟,结论就出来了:这个病人患系统性红斑狼疮的概率大概41%,当然还有其他的病,比如说类风湿关节炎等,参考概率只有3.5%,其他疾病的概率就更低了。

也就是说, 仅仅根据村医可知的必选信息, 已经基本上判断出这位病人患系统性红斑狼疮。为什么是这个病? 要有可解释性。点击“图解”可打开图形解释。从中可以看到患者是一位女性、青年, 青年女性的发病率是要增高的。下游的绿色节点是预期异常, 但实际上还没有异常, 其他颜色都是已经异常, 可由系统性红斑狼疮来解释。

为什么不是别的病? 排第二的类风湿关节炎的图解有3个孤立症状无法解释, 导致它的概率是比较低的, 它只能解释4个异常症状, 另外3个解释不了。其他概率更低的疾病的解释类似。

现在我们看到“检测完备度”只有45%, 这是因为我们还没做任何化验拍片或者病理检查, 只是在诊室里问诊, 所以可信度还不够高, 信息还不够多, 系统性红斑狼疮的概率也只有41%, 于是要做进一步检查。该做什么检查呢? 为指导基层医生做检查, 点选“推荐检测”按钮。

现在“推荐检测”结果出来了, 排第一位的是一个“问诊”: 关节痛是自限性的还是非自限性的? 这个病人是“非自限性”的。再次诊断, 结果有所改变, 系统性红斑狼疮概率升到45%, 类风湿关节炎概率下降到1%, 其他疾病也有所调整, 但变化不大。

再次点选“推荐检测”, 可以看到推荐的第一个选项是“自身免疫性抗体”检测, 共有5个指标, 虽然抗体的检测有很多, 但对这个患者只需要做这5个指标的检测。点选之后, 检查单已经发给化验科了。现在我们进到化验科的界面, 作为检查医生, 对“自身免疫性抗体”的这5个检查项目进行检查。经过化验之后, 将结果点选进去, 其中有1个是正常的, 3个是异常的, 还有一个未知, 医院没做这项检查。勾选保存, 这个检查结果就传到门诊医生那里。

加入“自身免疫性抗体”的检测结果后, 进行新的诊断结果是系统性红斑狼疮的概率为80%, 为什么是这个结果? 在“图解”中我们看到了一个六边形的“金标准”, 即双链DNA抗体阳性, 是系统性红斑狼疮的“金标准”, 所以确定80%得了这个病。其他疾病的可能性很小, 不到1%, 可以基本确诊了。

我们将原始的病历调出(四川遂宁真实病人)医院给病人做了24项指标的检查, 最后得到的诊断结果与DUCG系统得出的诊断结果比对, 结论完全一样, 而DUCG仅通过5项指标检查就得出与病人做24项指标检查完全一样的诊断结果。可见DUCG可显著缩短门诊的临床检查路

径, 减少检查项, 节省费用和时间。

诊断结果出来后, 如何治疗? 点击“方案”后, 推荐的治疗方案就展现出来(由中华医学会每两年更新一次系统性红斑狼疮指南中的治疗方案), 让医生从中选择他认为可以采纳的治疗方案。

有了处方, 还需要写病历。点击“生成病历”, 就生成了符合卫健委要求的一个规范化的自然语言病历, 这个自然语言病历是从DUCG的结构化病历生成的。所以系统中实际上保存了两份病历。如果临床医生认为病历写得不合适, 还可以进行修改, 处方自动加进去, 再将病历加密存储, 跟随病人一辈子。实际上, 任何连网的地方, 只要输入身份证号和密码, 就可以调取他过去所有的病历。

英文版DUCG现在也已经初步做了, 还可以做中医知识库。湖南中医药大学已经建造了两个知识库, 即病机十九条和“六经辩证”。

下面简单介绍知识库是如何建造的。我们针对每个主诉症状, 例如腹痛, 构建十几到上百个疾病的模块, 模块中的每条红线打开都是个矩阵, 都有参数, 每个变量都有描述。然后通过计算机自动合成为一张比较复杂的DUCG图, 也就是我们用于诊断的知识库。

上线应用之前, 我们对每个知识库都要进行第三方验证。验证的方法是将第三方医院的病历库中符合该知识库主诉的病历全部找出来, 再按照库中疾病分类, 每类随机抽取10份质量可靠的病历进行验证, 即将病历中的病情信息输入DUCG系统, DUCG不要求的不输入。然后看DUCG的诊断结论与病历记载的诊断结论是否一致, 一致的记为正确, 否则记为错误。结果所有知识库的诊断正确率都在95%以上, 其中少见病的正确率80%以上。

这种分类验证的方法是李兰娟院士提出来的, 更加科学严谨。因为每个库中少数常见疾病的病历数占了病历总数的绝大部分, 例如90%以上。如果不分类验证, 90%以上会是常见病, 如果这些常见病能正确诊断, 正确率就是90%以上, 即使对少见病的诊断都是错的。这样的智能系统是不能实际使用的, 因为基层医生对常见病通常能够正确诊断, 但对少见病诊断错误率很高。如果不能正确诊断少见病, 对基层医生就没有帮助。另外, DUCG所做的不限于某种疾病的筛查, 而是对库中疾病进行鉴别诊断, 不仅告诉你是不是某种疾病, 而且告诉你如果不是某种疾病, 那么是什么疾病。这样才能有效实现“基层首

诊”,病人才愿意接受“基层首诊”,而非现在这样:到大医院首诊,到基层拿药,与“分级诊疗”的医改方案完全相反。

DUCG在科学方法论上是对现代临床诊断学的颠覆。因为现代临床诊断是指医生凭自己的经验,从各种检查结果来推断到底是何种疾病,这要靠经验的积累,所以一定要20多年才能成为一个专家。而DUCG只需要从因到果构建知识库,然后能够反向计算,由果到因诊断出疾病,乃至可以动态地生成病人个体优化的临床检查路径。DUCG的个体优化临床检查路径考虑了疾病的危险度、检测的代价和检测效。这样就使得临床诊断学从一门经验科学变成了一门精确计算的严谨的科学,我们的每一个公式都是经得起证实或证伪的,是透明的,它的物理意义都是清晰可辨的。

目前我们的系统通过第三方验证后,已经在山东胶州和重庆忠县的县、乡镇和村(社区)三级医疗机构实际应用,并与当地的HIS系统双向对接,不干扰当地诊疗的正常流程。目前胶州已积累9000多个病例,异议率3.3%,其中只有6例是DUCG的问题,即我们的主诉知识库中没有包含两个应该包含的疾病。例如腹痛中没有包含妇科的盆腔炎。现在将盆腔炎加入腹痛后,诊断就正确了。也就是说,DUCG在基层实际使用的正确率近乎100%。忠县开展的晚些,也积累了6000多个病例,异议率更低。

目前有一种误解,似乎人工智能就是大数据机器学习。但这是不对的,DUCG就不用大数据,而是大知识。我们建库不用一个病历数据,只跟专家合作。当然,专家后面通常是一个团

队,包含主诉症状的各种跨科室疾病的知识。病历数据只是用于第三方验证,而不是构建知识库。此外,现在知识库被误解为各种教科书、病历、论文、指南、规范等文献的集成,其实那只是可检索的数据库,而非知识库。知识库应当与推理算法相融合,无论其表达方式是什么。也就是说,知识库是可以活用的知识,而不是静态的碎片化的知识。习近平总书记在党的十九大报告中提出“推动互联网、大数据、人工智能和实体经济深度融合”。大数据和人工智能是并列关系,不是等同关系。大数据机器学习是对数据的拟合(黑箱);DUCG是对知识的图形表达(白箱),其维护和升级是通过专家验证新知识后在知识库编辑器中修改模块再合成而得,两者有本质区别。基于知识的DUCG更加稳定可靠并具有可解释性。张钹院士认为“AI奇迹难再现,深度学习潜力近天花板”;谭铁牛院士在2018年院士大会上指出,目前大数据机器学习人工智能存在六大瓶颈:1)数据瓶颈;2)泛化瓶颈;3)能耗瓶颈;4)语义鸿沟瓶颈;5)可解释性瓶颈;6)可靠性瓶颈。而DUCG无这些瓶颈。当然大数据也有优势,如医学的图像识别,声音识别。医学影像识别和声音识别结果可作为DUCG的输入,彼此是合作关系,相得益彰。

#### 作者简介:



张勤,教授,博士生导师。全国政协常委,国际核能院院士,国家核电重大专项战略咨询专家组组长,中国人工智能学会会士、不确定性人工智能专委会主任,中国知识产权研究会副理事长兼学术顾问委员会主任。

中文引用格式:张勤.智慧医疗促进分级诊疗——DUCG的初步实践[J].智能系统学报,2020,15(5):1028–1032.

英文引用格式:ZHANG Qin. Medical intelligent system to promote graded clinical diagnoses and treatments: application of dynamic uncertain causality graph[J]. CAAI transactions on intelligent systems, 2020, 15(5): 1028–1032.