**专利技术交底书**

提出部门：上海交通大学

全部发明人：曹健、朱能军

第一发明人身份证号码：320422197209242817

联系人：曹健

联系电话：13651963843

电子邮箱：cao-jian@sjtu.edu.cn

1. **发明名称：**

基于多元证据融合的治疗方案推荐方法

1. **所属技术领域：**

本发明主要涉及医疗决策领域和机器学习相关技术。同时，基于Dempster-Shafer证据理论，本发明从医生对历史病例治疗方案的支持度，案例之间的相似度（距离），医疗指南的建议以及机器学习推荐方案等多个角度提出针对最终决策方案的多元证据主体以及证据可靠性的表达。然后，利用Dempster合成规则将这些证据进行融合。最后，根据融合后的证据支持以及Pignistic概率为目标病例确定最为合理的治疗方案，并提供证据支持。

1. **技术背景**

不同于其它运用机器学习方法进行决策辅助的领域，医疗决策领域对于机器推荐治疗方案的可解释性和相应的证据支持拥有更高的要求。因此，这对于单纯的利用历史病例数据建立的机器学习方法来说，是非常具有挑战性的。因为他们往往只追求指标上的最优（如最大化推荐结果的准确度），而不关心这些推荐结果是否可以解释。另外，有些机器学习方法，比如神经网络，本身就无法给出推荐结果的解释。所以，直接把这些结果推荐给医生或者患者，是很难被接受的。不过，可以把这些机器学习方法的推荐结果作为一个独立的证据，同时设定该证据的可靠性（或不确定性），以此支持最终的决策。另外，为了降低选择不恰当的治疗方案的风险，许多权威的医学指导方针，如国际上的NCCN（the National Comprehensive Cancer Network）医疗指南，已被各种医疗机构发布和使用。因此，这些通过经验总结出来的规则也可以作为一个独立的证据支持最终的决策。

另外，基于K近邻（KNN）的治疗方案推荐方法因为它结果的可解释性，已经被广泛运用在医疗决策领域。然而，这些方法仅仅从病例之间的距离（或相似度）的角度，去刻画和总结最终的推荐方案，并没有考虑单个病例作为独立证据的可靠性。本发明认为，即使是已经确定了治疗方案的历史病例，在确定治疗方案的过程中也会存在一定的分歧度，尤其是通过基于少数服从多数准则讨论过后确定的结果，从而带入其作为证据的不可靠性。因此，把这部分信息融合到决策当中，会进一步促进推荐结果的合理性，从而降低决策风险。

Dempster-Shafer证据理论为多元证据的合成提供了技术支持，具有直接表达“不确定”和“不知道”的能力。这些信息表示在mass函数中，并在证据合成过程中保留了这些信息。Dempster-Shafer证据合成过程主要涉及两个核心步骤：基本置信分配（Basic Belief Assignment,BBA）和Dempster证据合成。假设X为可供决策的治疗方案集合，将它的所有子集的集合表示为，这个被称为识别框架，其中的元素被称为事件。基本置信分配的作用是确定一个mass映射函数，同时，空集的mass函数的值为0，即，其他事件mass值的和为1，即。通过不同的证据主体，可以构造出不同的mass函数，如，和，Dempster合成规则就是要融合这些函数，如下：





其中，。

在得到合成后的mass函数后，就可以根据融合后的置信度分配确定最终的治疗方案。

1. **发明目的**

针对上述现有技术中的缺陷，本发明要解决的技术问题体现在以下几点：

1）从医生对历史病例治疗方案的支持度，案例之间的相似度，医疗指南的建议以及机器学习推荐方案等多个角度提出针对最终决策方案的多元证据主体以及证据可靠性的表达

2）利用Dempster合成规则将这些证据进行融合

3）根据融合后的证据支持以及Pignistic概率为目标病例确定最为合理的治疗方案

1. **技术方案**

本发明中的方法包括以下步骤：

步骤1：以目标病例的K个近邻为证据主体，分别为每个近邻构造mass函数（基本置信度指派），确定证据可靠度，其主要包括以下子步骤：

步骤1.1：根据患者信息确定病例的特征向量V，以及对应的治疗方案（y既可以是由单个治疗方案构成的单元素集合，也可以是由多个治疗方案构成的集合，前者一般是由基于数据的机器学习方法提供的推荐方案，后者一般符合基于医疗指南的推荐结果）；另外，根据医生讨论投票记录和支持该治疗方案的医生比例，确定该治疗方案的先验支持度，其中表示应该参与投票的所有医生人数，表示支持该治疗方案的医生人数，表示缺席投票的人数，表示该次投票中所有候选治疗方案的个数。

步骤1.2：根据病例的特征向量确定病例之间的距离d(i,j),其中，i表示目标病例，j表示历史病例。

步骤1.3：根据距离d(i,j)，确定目标病例i的K近邻集合N(i)

步骤1.4：每个近邻作为一个独立证据，根据其推荐的治疗方案y以及根据步骤1.1中计算得到的，连同其同目标病例的距离d（i,j）确定基本置信度指派如下：

其中表示所有证据主体涉及到的治疗方案的集合，指示了该证据的不确定性（不可靠性）。

步骤2：确定基于医疗指南的证据表达（基本置信度指派）和证据可靠度表达，其主要包括以下子步骤：

步骤2.1：根据经验或医疗指南历史表现设定医疗指南作为证据主体的不确定性，

步骤2.2：根据医疗指南确定目标病例的推荐方案

步骤2.3：确定该证据的mass函数如下：

步骤3：确定其它机器学习模型（如神经网络）的证据表达（基本置信度指派）和证据可靠度表达，其主要包括以下子步骤：

步骤3.1：根据历史病例数据和对应的治疗方案，学习出机器学习模型

步骤3.2：根据模型在测试集中的推荐准确accuracy度，设定将该模型结果作为证据主体的不确定性

步骤3.3：根据模型给出的推荐方案y以及步骤3.2中确定的模型推荐准确度，确定该证据的mass函数如下：

步骤4：融合以上从三种不同角度提出的证据，如下：



步骤5：根据融合后的证据做出最终的治疗方案推荐，主要步骤是计算每个候选治疗方案y的Pignistic概率（BetP）如下：



最后，根据BetP(y)的大小，确定最终的推荐方案。

1. **附**

表1 各治疗方案的基本置信度分配表示例

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 方案集合 |  |  |  |  |  |
| {Tcb} | 0 | 0.8 | 0 | 0 | 0.53 |
| {Pcb} | 0.7 | 0 | 0 | 0 | 0.31 |
| {No need} | 0 | 0 | 0 | 0.6 | 0.02 |
| Need={Tcb,Pcb} | 0 | 0 | 0.9 | 0 | 0.12 |
| {No need,  Need,Tcb,Pcb} | 0.3 | 0.2 | 0.1 | 0.4 | 0.02 |

1. **本专利申请的有益效果**

由于采用了上述技术方案，本发明的有益效果是：

能够充分收集和利用多元证据主体，结合Dempster-Shafer理论，为医疗决策提供证据更为充分的推荐治疗方案，并且充分考虑了各个证据主体的不确定性（证据可靠性）。另外，最终的推荐方案具有很好的解释性，它能把不同结构的证据有机结合起来，并且以一种统一的方式表达出来，更容易被医生和患者接受。并发明通过证据融合，可以自适应得把不同的证据的特性挖掘出来，例如基于医疗指南的推荐结果一般为多种治疗方案的集合，而基于历史数据的推荐结果往往会得到更细的治疗方案，以及将机器学习方法的推荐准确度、医生投票的支持度都纳入到证据的考量等。

1. **可以实现本发明的方式与途径**（结合交底书第五项的技术方案来详细描述实施举例实现的过程）

下面结合具体实施例对本发明进行详细说明。以下实施例将有助于本领域的技术人员进一步理解本发明，但不以任何形式限制本发明。应当指出的是，对本领域的普通技术人员来说，在不脱离本发明构思的前提下，还可以做出若干变形和改进。这些都属于本发明的保护范围。

假设需要为目标病例i推荐辅助化疗方案，现在通过KNN算法得到的2个历史近邻j1和j2，且d(i,j1)=d(i,j2)=0。计划应该参与对j1和j2治疗方案投票医生均为10人，实际参与j1投票的有10位医生，且有7位投票支持最终治疗方案为Tcb,实际参与j2投票的有8位医生，7位投了Pcb方案，另一位投了其他方案。另外，根据NCCN医疗指南，认为目标患者需要化疗，但没有给出更细化的化疗方案。最后，通过神经网络算法，预测得出目标病例不需要化疗，且该算法在历史测试中验证得出准确度为0.6。所有这些方法中，总共涉及到的推荐方案有：Tcb,Pcb,Need（需要），No need（不需要），如表1所示，且Tcb和Pcb均属于细化了的化疗方案，所以Need={Tcb,Pcb}。根据这些信息，我们的基于多元证据融合的治疗方案推荐方法具体步骤如下：

步骤1：为近邻j1和j2构造mass函数（即，基本置信度指派），确定证据可靠度，其主要包括以下子步骤：

步骤1.1：根据患者信息确定病例的特征向量V，以及对应的治疗方案（y既可以是由单个治疗方案构成的单元素集合，也可以是由多个治疗方案构成的集合，前者一般是由基于数据的机器学习方法提供的推荐方案，后者一般符合基于医疗指南的推荐结果）；另外，根据医生讨论投票记录和支持该治疗方案的医生比例，确定该治疗方案的先验支持度，其中表示应该参与投票的所有医生人数，表示支持该治疗方案的医生人数，表示缺席投票的人数，表示该次投票中所有候选治疗方案的个数。基于前面的假设，j1对应的，j2对应的。

步骤1.2：根据病例的特征向量确定病例之间的距离d(i,j),其中，i表示目标病例，j表示历史病例

步骤1.3：根据距离d(i,j)，确定目标病例i的K近邻集合N(i),基于前的假设N(i)={j1,j2}

步骤1.4：每个近邻作为一个独立证据，根据其推荐的治疗方案y以及根据步骤1.1中计算得到的，连同其同目标病例的距离d(i,j)确定基本置信度指派如下：

其中表示所有证据主体涉及到的治疗方案的集合，指示了该证据的不确定性（不可靠性）。基于前面的假设，以j1为证据主体，其配置信度分配如下：

其中={Tcb,Pcb,No need,Need}，且Need={Tcb,Pcb}，j2相关计算方法同j1，结果如表1所示。

步骤2：确定基于医疗指南的证据表达（基本置信度指派）和证据可靠度表达，其主要包括以下子步骤：

步骤2.1：根据经验或医疗指南历史表现设定医疗指南作为证据主体的不确定性，假设为0.1.

步骤2.2：根据医疗指南确定目标病例的推荐方案，即y=Need（需要化疗方案）。

步骤2.3：确定该证据的mass函数，得到表1中分配，如下：

步骤3：确定其它机器学习模型（如神经网络）的证据表达（基本置信度指派）和证据可靠度表达，其主要包括以下子步骤：

步骤3.1：根据历史病例数据和对应的治疗方案，学习出机器学习模型

步骤3.2：根据模型在测试集中的推荐准确accuracy度，设定将该模型结果作为证据主体的不确定性

步骤3.3：根据模型给出目标病例的推荐方案y={No need}以及步骤3.2中确定的模型推荐准确度，确定该证据的mass函数，得到表1中分配，如下：

步骤4：融合以上从三种不同角度提出的证据，如下：



基于以上的案例，其计算结果如表1所示。

步骤5：根据融合后的证据做出最终的治疗方案推荐，主要步骤是计算每个候选治疗方案y的Pignistic概率（BetP）如下：



例如，



最后，根据BetP(y)的大小，确定最终的推荐方案。

1. **相关专利、公知技术检索结果及其与本发明/实用新型的异同点**

（检索到的专利和公知技术文献摘要以及这些专利和公知技术中存在的不足。这些文件可以是专利文件，也可以是期刊、杂志、手册和书籍等非专利文件。前者要写明国别、公开号和公开日期，后者应写明文件的详细出处。发明人应客观地指出这些专利和公知技术文献中的技术存在的问题和缺陷，这些问题和缺陷仅限于本申请的技术方案所解决的问题和缺点，可能的话，应说明产生这些问题和缺点的原因以及解决这些问题的难度。）

**关键词：** (填写专利与公知技术检索需要的关键词)

**关键词：** (填写专利与公知技术检索需要的关键词)

**相关检索结果1：**

申请（专利）号: xxxxxxxxxxxx 名称:xxxxxxxxxxxxxxxxx

摘要: xxxxxxxxxxxxxx

技术要点比较：

1．XXXXXXXXX：xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx

2．XXXXXXXXXXXXXXXXXXXXXXXXX：xxxxxxxxxxxxxxxxxxxx

**相关检索结果2：**

申请（专利）号: xxxxxx 名称: xxxxxxxxx

摘要: xxxxxxxxxxxxxx

技术要点比较：

xxxxxxxxxxxxxxxxxxxxxxxxx

**相关检索结果3：**

现有公知技术名称，出处，

技术要点比较