**说 明 书 摘 要**

本发明提供了一种适用于概念漂移的医疗方案的推荐系统及方法，该系统包括：用户界面模块，与推荐模块相连，用于读取一开始的训练样本；推荐模块，与工作流系统相连，根据历史病例的数据，自动智能地为当前的病例计算出一个最合适的医疗方案；外部数据库，与推荐模块相连，用来读取一开始的训练样本；外部数据库用于存放每一个病例的详细信息，例如姓名身高体重等个人信息，以及具体的各项医疗指标信息。本发明能够考虑到了概念漂移现象对结果产生的影响，兼顾了样本的采集时间等因素，能检测到概念漂移的发生，并且修正了过时的样本以适应新的规律，使得此后的预测更加准确。

**摘 要 附 图**



**权 利 要 求 书**

1、一种适用于概念漂移的医疗方案的推荐系统，其特征在于，包括：

用户界面模块，与推荐模块相连，用于读取一开始的训练样本；

推荐模块，与工作流系统相连，根据历史病例的数据，自动智能地为当前的病例计算出一个最合适的医疗方案；

外部数据库，与推荐模块相连，用来读取一开始的训练样本；外部数据库用于存放每一个病例的详细信息，例如姓名身高体重等个人信息，以及具体的各项医疗指标信息；

工作流数据库，与工作流系统相连，工作流数据库用来存储和读取每一个病例的状态；

工作流系统，工作流系统用于控制每一个病例的被整个系统处理的基本流程。

2、根据权利要求1所述的适用于概念漂移的医疗方案的推荐系统，其特征在于，所述推荐模块包括方案推荐模块、方案制定模块，其中：方案推荐模块是推荐方法给待预测的样本推荐一个方案 的过程；方案制定模块是确定实际采用的方案的过程。

3、一种用于概念漂移的医疗方案的推荐方法，其特征在于，其包括以下模块：

步骤一，开始；

步骤二，输入待推荐方案的病例信息x；

步骤三，读取历史病例数据Da，从历史病历样本集合中读取历史病历数据Da；

步骤四，同时计算对概念敏感的探测器；

步骤五，输出自适应分类器的推荐结果Yac，自适应分类器的结果将作为此病例的推荐医疗方案输出；

步骤六，确定时间采用的方案y，待医生根据推荐结果进行讨论分析，并将病人个体的特征作为参考，最终得出实际采用的诊疗方案之后；

步骤七，将病例信息x和步骤六的方案y加入历史并样本集合，将此例病人的信息加入历史数据样本集合中，以供将来使用；

步骤八，判断对比自适应分类器的结果、CSD分类器的结果和实际采用的方案是否相同，是则转步骤九，否则转步骤十；

步骤九，更新每个历史样本Dt的冲突列表Ct，自适应分类器预测错误，而CSD分类器预测正确，则需要对历史病例样本集合中的相关数据进行更新，使旧的样本适应新的规律，从而更准确地预测当前规律下的医疗方案；

步骤十，结束。

4、根据权利要求1所述的用于概念漂移的医疗方案的推荐方法，其特征在于，所述步骤二使用病例信息以及各项临床指标等属性作为特征建立模型，使用模型逐条计算新病例与现有的所有病例记录的相似度，并根据相似度排序，根据相似度最高的几条结果，综合产生推荐方案。

5、根据权利要求1所述的用于概念漂移的医疗方案的推荐方法，其特征在于，所述步骤三中历史病例数据Da等于训练样本集合，每一条历史病例数据，就是一个训练样本。

**说 明 书**

适用于概念漂移的医疗方案的推荐系统及方法

**技术领域**

本发明涉及一种推荐系统及方法，具体地，涉及一种适用于概念漂移的医疗方案的推荐系统及方法。

**背景技术**

在医疗决策的过程中，使用kNN（k-NearestNeighbor，邻近算法）等基于案例的推荐算法可以有效地利用以往的相似病例来提供医疗方案的推荐。但是实际情况下，对于同一特征病人所采用的治疗方案不断地发生着变化，新的医学研究进展、新药物的研发等都会引起治疗方案以不规律的方式变化着。这种目标变量的统计特性随着时间的推移不规律变化的现象，我们通常称作概念漂移。现有的技术为了消除过时的样本的影响，一般简单地丢弃了过时的样本，损失了隐含在这些样本中的信息，同时也导致可以利用的样本数量变小，从而引起误差的增加。

**发明内容**

针对现有技术中的缺陷，本发明的目的是提供一种适用于概念漂移的医疗方案的推荐系统及方法，其能够考虑到了概念漂移现象对结果产生的影响，兼顾了样本的采集时间等因素，能检测到概念漂移的发生，并且修正了过时的样本以适应新的规律，使得此后的预测更加准确。

根据本发明的一个方面，提供一种适用于概念漂移的医疗方案的推荐系统，其特征在于，包括：

用户界面模块，与推荐模块相连，用于读取一开始的训练样本；

推荐模块，与工作流系统相连，根据历史病例的数据，自动智能地为当前的病例计算出一个最合适的医疗方案；

外部数据库，与推荐模块相连，用来读取一开始的训练样本；外部数据库用于存放每一个病例的详细信息，例如姓名身高体重等个人信息，以及具体的各项医疗指标信息；

工作流数据库，与工作流系统相连，工作流数据库用来存储和读取每一个病例的状态；

工作流系统，工作流系统用于控制每一个病例的被整个系统处理的基本流程。

优选地，所述推荐模块包括方案推荐模块、方案制定模块，其中：方案推荐模块是推荐方法给待预测的样本推荐一个方案 的过程；方案制定模块是确定实际采用的方案的过程。

本发明还提供一种用于概念漂移的医疗方案的推荐方法，其特征在于，其包括以下模块：

步骤一，开始；

步骤二，输入待推荐方案的病例信息x；

步骤三，读取历史病例数据Da，从历史病历样本集合中读取历史病历数据Da；

步骤四，同时计算对概念敏感的探测器；

步骤五，输出自适应分类器的推荐结果Yac，自适应分类器的结果将作为此病例的推荐医疗方案输出；

步骤六，确定时间采用的方案y，待医生根据推荐结果进行讨论分析，并将病人个体的特征作为参考，最终得出实际采用的诊疗方案之后；

步骤七，将病例信息x和步骤六的方案y加入历史并样本集合，将此例病人的信息加入历史数据样本集合中，以供将来使用；

步骤八，判断对比自适应分类器的结果、CSD分类器的结果和实际采用的方案是否相同，是则转步骤九，否则转步骤十；

步骤九，更新每个历史样本Dt的冲突列表Ct，自适应分类器预测错误，而CSD分类器预测正确，则需要对历史病例样本集合中的相关数据进行更新，使旧的样本适应新的规律，从而更准确地预测当前规律下的医疗方案；

步骤十，结束。

优选地，所述步骤二使用病例信息以及各项临床指标等属性作为特征建立模型，使用模型逐条计算新病例与现有的所有病例记录的相似度，并根据相似度排序，根据相似度最高的几条结果，综合产生推荐方案。

优选地，所述步骤三中历史病例数据Da等于训练样本集合，每一条历史病例数据，就是一个训练样本。

与现有技术相比，本发明具有如下的有益效果：本发明能够考虑到了概念漂移现象对结果产生的影响，兼顾了样本的采集时间等因素，能检测到概念漂移的发生，并且修正了过时的样本以适应新的规律，使得此后的预测更加准确。

**附图说明**

通过阅读参照以下附图对非限制性实施例所作的详细描述，本发明的其它特征、目的和优点将会变得更明显：

图1为本发明适用于概念漂移的医疗方案的推荐系统的模块框图。

**具体实施方式**

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。

如图1所示，本发明适用于概念漂移的医疗方案的推荐系统包括：

用户界面模块，与推荐模块相连，用于读取一开始的训练样本；

推荐模块，与工作流系统相连，根据历史病例的数据，自动智能地为当前的病例计算出一个最合适的医疗方案。

外部数据库，与推荐模块相连，用来读取一开始的训练样本；外部数据库用于存放每一个病例的详细信息，例如姓名身高体重等个人信息，以及具体的各项医疗指标信息。

工作流数据库，与工作流系统相连，工作流数据库用来存储和读取每一个病例的状态，例如某一个病例当前处于什么状态(已经确定医疗方案了，还是在等待系统推荐方案，还是具体信息不足等等)

工作流系统，工作流系统用于控制每一个病例的被整个系统处理的基本流程。

所述推荐模块包括方案推荐模块、方案制定模块，其中：方案推荐模块是推荐方法给待预测的样本推荐一个方案 的过程；方案制定模块是确定实际采用的方案的过程。

本发明适用于概念漂移的医疗方案的推荐方法包括以下步骤:

步骤一，开始；

步骤二，输入待推荐方案的病例信息x；

步骤三，读取历史病例数据Da，从历史病历样本集合中读取历史病历数据Da；

步骤四，同时计算CSD分类器推荐方案Ycsd和自适应分类器推荐方案Yac，分别使用自适应分类器和CSD分类器进行方案推荐，Yac表示自适应分类器的分类结果，Ycsd表示分类器CSD的分类结果；

步骤五，输出自适应分类器的推荐结果Yac，自适应分类器的结果将作为此病例的推荐医疗方案输出；

步骤六，确定时间采用的方案y，待医生根据推荐结果进行讨论分析，并将病人个体的特征作为参考，最终得出实际采用的诊疗方案之后；

步骤七，将<x，y>加入历史并样本集合，将此例病人的信息加入历史数据样本集合中，以供将来使用；

步骤八，判断对比自适应分类器的结果、CSD分类器的结果和实际采用的方案是否相同，是则转步骤九，否则转步骤十；

步骤九，更新每个历史样本Dt的冲突列表Ct，自适应分类器预测错误，而CSD分类器预测正确，则需要对历史病例样本集合中的相关数据进行更新，使旧的样本适应新的规律，从而更准确地预测当前规律下的医疗方案；

步骤十，结束。

所述步骤二使用病例信息以及各项临床指标等属性作为特征建立模型，使用模型逐条计算新病例与现有的所有病例记录的相似度，并根据相似度排序，根据相似度最高的几条结果，综合产生推荐方案，这样使用方便。

所述步骤三中历史病例数据Da等于训练样本集合，每一条历史病例数据，就是一个训练样本，训练样本集合D如下式（1）所示：

......（1）

其中，N表示训练集合的样本容量，n表示每个样本的属性个数，在训练集合中，样本按照采集的时间顺序排列，换言之序号i越小，样本的采集时间越早，对于训练集合中的第i个历史样本Di如下式（2）所示而言：

......（2）

其中分别表示第i个样本n个不同属性的值，yI表示第i个样本的所属类别。

所述步骤四中AC表示使用训练集合中所有样本的自适应分类器，其分类结果综合考虑了所有历史样本，因而更为稳定；CSD表示使用训练集合中最近采集样本的分类器，其分类结果对新的概念更加敏感，用作概念漂移的检测；为了检测概念漂移的发生，需要使用一组使用不同样本的kNN分类器对目标样本进行分类。

所述步骤七，对于一个当前要预测的样本<x（now）,y（NOW）>，如果分类器AC的结果错误，而CSD的结果正确，那么意味着AC产生这个结果的过程中使用的历史数据和当前的样本，由于服从的规律不同而产生了不一致。

所述步骤八，判断Yac和Ycsd是否一致，如果Yac和Ycsd一致，则认为没有概念漂移的发生；反之，如果Yac不等于Ycsd，说明近期采集的样本服从的规律和历史样本有所不同，可能有概念漂移的发生，需要进一步进行检测。

所述步骤九中，Dt如上式（2）所示满足t＞N，N为当前训练集合的样本容量，要预测Dt所属的类别yt，就要利用训练集合中的其他样本；Ct下式（3）所示用来记录和Dt不一致的样本，对AC分类器使用到的每一个历史样本Dt，需要标记它和当前预测的样本有冲突，即将加入（now,y（now））中Ct，对任意一个历史样本Dt，如果Ct中的项目达到了一定的阈值，则认为概念漂移已经对该历史样本产生了很大的影响，用最新的值更新历史样本中的yt以适应最新的规律，更准确地为接下来的预测和推荐服务，

......（3）

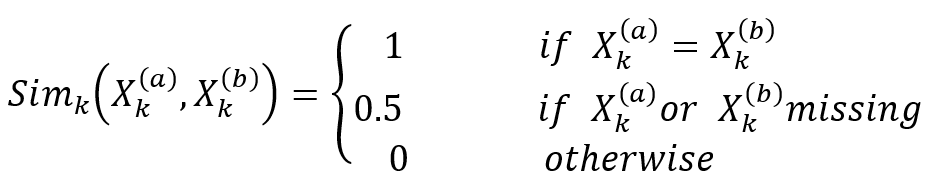
Ct是一个集合，用来记录和Dt不一致的样本。这个公式是一个数学的集合表达式，意思是：这个集合里面每一个元素的形式是（i，y(i))这样的，其中i是不一致的样本的序号，要求比t大；x(i)是这个样本的特征属性，要求满足它属于“距离Dt最近的k个样本”之中。

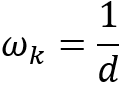
其中nearest（最近的）代表和当前待预测的样本Xi相似度最高的k个样本的集合如下式（4）所示。

......（4）

a 和 b 分别是两个样本Da和Db的序号, d表示训练样本中的样本总数。

对于类别型的离散属性，其值之间并没有特定的大小关系，所以不能像数值型的属性一样简单地使用马氏距离来直接度量样本之间的相似度，通常，为了计算样本和之间的相似度，需要先分别定义每个属性上的距离，最后各个属性加权综合计算得出最终的结果，在本发明中在具有离散属性的样本上定义了一种新的相似度度量方法，如果两个样本和在第k个属性上的值相同，则样本和在这个属性上的相似度为1；如果两个样本和任何一个在第k个属性上的值未知或者缺失，相似度为0.5；否则相似度为0，具体如下式（5）和（6）所示：

......（5）

......（6）

以上对本发明的具体实施例进行了描述。需要理解的是，本发明并不局限于上述特定实施方式，本领域技术人员可以在权利要求的范围内做出各种变形或修改，这并不影响本发明的实质内容。

**说 明 书 附 图**



图1