基于特征加权的案例推理方法在消防车辆调派中的应用

**摘要**：案例推理模型被广泛应用于应急管理等领域，目前对于案例推理（CBR）的研究，大多集中在案例的表示和检索中，应用较广泛的是KNN的检索策略，而KNN一般使用均权法，不同的属性对案例的表示通常具有不同的权重，会影响计算结果的可靠性。本文提出了一种新的特征权重计算方法和一种新的案例合理性评估机制，对占有权重较大的属性赋予较合理的权重，并应用于案例推理的相似度检索上，检索结果的准确率与KNN算法相比有了一定的提高，同时能够快速的对消防车辆案例推理系统的推荐方案合理性做出合理的评估。

**关键字**：案例推理；应急管理；合理性评估；权重计算

0.引言

CBR(case-based reasoning)案例推理是由AI衍生而来的一项技术，它的工作方式是案例的推理和学习，这个与人类思维和处理问题的方式一致，从一个已知案例到另一个新案例的类比过程。而KNN算法具有易于实现、无需训练、支持增量学习、能对超多边形的复杂空间建模训练等优点，与案例推理的增量学习方法非常契合，可以使案例库随着案例的增加而逐渐增加所覆盖的范围，逐步提高解决实际问题的能力，从而提高案例库的实用性和易用性。在突发事件发生后利用一定的智能化方法快速检索出相似历史案例进而辅助决策者参照历史经验快速制定应急方案对于在应急状态下节省宝贵的时间、最大程度地减少灾害带来的损失具有重要的意义。

构造基于案例推理的系统的关键问题包括案例表示、案例检索和案例修改与学习，本文主要研究特征加权相似度算法在提高案例检索精度中的应用。

1.KNN算法

K最近邻法（KNN，K－nearest　neighbor　approach）是一种从用户案例库中寻找与当前目标案例最近距离的方法。通过计算目标案例与原案例的属性距离和来检索案例库中匹配的案例。由于该算法比较成熟并且简单易于实现，别广泛应用于机器学习分类和模式识别中，算法的基本思想如下：  
 如果一个实例（样本）在特征空间中 的最近的K个实例属于某一种类别，那么这个实例也属于这个类别。该算法设定所有的实例对应于N维欧式空间中的点，通过计算该目标点与其他所有点之间的距离，取出与该点最近的K个点，然后统计这K个点里面比例最大的类别属性，则该目标点属于这个分类。

输入训练数据集：

　　　　　　　(1)

其中，为实例的特征向量，为实例类别，；输出实例所属的类。

(1)根据给定的距离度量，在训练集T中找出与距离最近的个点，包含这个点的邻域记作；

(2)在中根据决策规则确定的类别：

　　　(2)

其中当时为1，否则为0。

在本文中，取特殊情况1，将训练集中与距离最近的类作为的类。

1.2 距离度量

特征空间中两个实例点的距离直接反映出两个案例的相似度。本文中的特征空间为n维实数向量空间，使用的距离是欧氏距离，但是也可以使用其他距离如距离或者距离。

设和是n维实数向量空间的两个示例，，，那么的距离定义为：

(3)

；

当p=2时，称为欧氏距离即：

(4)

当p=1时，称为曼哈顿距离：

(5)

当时，是各个坐标距离的最大值：

(6)

1.3 K值的选择

在常规的KNN算法中，k值的选择会对分类结果产生重大的影响。如果k值选取较小，预测结果会对附近的实例点非常敏感，模型整体会变得复杂，容易发生过拟合；如果选取较大的k值，虽然可以减少学习评估的误差，但是会把与目标实例较远不太相似的源案例也划分进来，预测的近似误差会增大，可能会使预测发生错误。因此需要通过交叉验证等方法选取一个合适的k值。

但是在案例推理模型中，我们可以根据自己的需要选取k值，让系统同时输出与目标案例最相似的k个案例，并找到对应的k个解决方案。将检索到的结果与专家经验相结合，从中选取最合适的解决方案，辅助完成应急决策。

2. 加权KNN模型

大多数的案例推理模型都是采用的KNN检索方法实现案例的一对一属性匹配。但由于KNN一般采用均权法，无法体现案例属性对案例的贡献度，一定程度上对相似度检索会产生偏差，本文在KNN的基础上对属性权重重新赋值，对类别相关性高的属性给予更高的权重。加权KNN的相似度计算公式如下：

(7)

其中，T表示目标案例，S表示案例库中的源案例，n为案例特征属性，i表示案例的第i个特征，为目标案例与源案例在第i个特征上的相似函数，表示第i个特征的权重。

2.1 权值更新算法

Relief算法是由Kira最早提出来的一种特征权重算法，根据各个特征和类别的相关性赋予特征不同的权重，算法比较简单且运行效率高，但由于该算法仅局限于两类数据的分类问题，在1994年kononeill对其进行了扩展，得到了Relief-F算法处理多分类问题。每次从样本集中随机取出一个样本A，然后从与A同类的样本中找出与A近邻的k个样本，从A的不同类样本中找出k个近邻样本，更新每个特征的权重。但是该算法存在一个缺点：每次随机抽样可能会抽到同一类，计算出的特征权重在该类的属性上可能会被赋予较大的值，这对于分类来说显然是不合理的。

本文设计了一种新的权值更新算法，通过在训练集的每类样本中随机抽样n次计算特征权重后求平均，降低了样本分布差异对特征权重的影响，从而得到更有效的特征权重，使用特征加权后的案例推理检索出的结果更具有科学性。

样本特征权值的计算方法如下：

1. 使用K-Means对案例库样本聚类，运行速度快，结果也令人满意。根据火灾等级聚为m类；
2. 从每类样本中随机抽取1个样本，对于每个样本，从和它同类的样本中找出最近的j个样本（j=1,2,...k），从每个不同类样本集中找到j个近邻，然后更新每个特征的权重如下式所示：

初始化各特征权重W为0，重复抽取n次：

 (8)

其中，是特征属性A的权重，是特征属性A的初始化权重， 是从第i类案例样本库中随机抽取的第t个案例样本，是从第 i类案例样本库中找出的距离随机抽取的第t个案例样本的第j个 最 近案例样本，是从C类案例样本库中找出的距离从第i类案 例样本 库中随机抽取的第t个案例样本的第j个近邻案例样本，C类 案例样本 库不属于第i类案例样本库，表示案例样本 和案例 样本在特征属性A上的差，表示案 例样本和 案例样本在特征属性A上的差。

的计算如下：

 (9)

是从第i类案例样本库中随机抽取的第t个案例样本在特征属性 A上的取值，是从第i类案例样本库中找出的距离随机抽取的 第t个案例样本的第j个最近案例样本在特征属性A上的取值，max(A) 是指在收集的案例样本中特征属性A的最大值，min(A) 是指在收集的 案例样本中特征属性A的最小值。

整个算法流程图如下：



**图1 权值更新流程图**

3.案例推理模型

将计算的特征权重代入上式(7)，检索得到最相似的案例和相应的派车方案Y，若计算出的最相似案例存在多个，则分别对推荐出的多个案例进行合理性分析。案例推理步骤及流程图如下：

1. 从数据库中查找已有案例和处置方案并找到可以描述案例特的特征字段，建立案例库样本数据集。
2. 数据清洗、预处理、转化、归一化和编码等。
3. 利用上述改进的算法计算样本特征的属性权重，并代入距离公式检索得到最相似的几个案例和相对应的解决方案。
4. 将检索得到的方案与专家经验结合，评估方案的合理性并稍作改进得到最终解决方案。
5. 对上一步推荐的解决方案进行专家评价，如果符合要求并且能够顺利解决问题则将新案例加入到案例库，否则弃之。



**图2 案例推理模型流程图**

1. 推荐方案评估模型

结合消防火灾救援的实际情况，对推荐方案的合理性进行评估，评估方法如下：

1. 根据案例库不同类型车辆的数量对车辆排序，如数量最多的第一类车为泡沫 消防车（A），接着为压缩空气泡沫消防车（B），抢险救援消防车（C）， 器材消防车（D），E，F，G……
2. 根据推荐车辆类型及数量，将推荐方案转换为一维数组，索引对应车辆类型，例如：推荐方案 抢险救援消防车2辆，通讯指挥消防车1辆，泡沫消防车3辆，压缩空气泡沫消防车2辆，云梯消防车1辆，则对应的矩阵为
3. 根据消防救援标准定义派车数量扣分表，对于多派的情况，会造成资源浪费，相应扣除一部分分数；对于少派的车辆，可能会造成灭火救援效果不理想甚至人员伤亡和财产损失，所以扣除更多的分数，扣分表设计如下：

**表1 派车扣分表**

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 类型 | A | B | C | D | E | F | G | H | I | J | …… |
| 多派 |  |  |  |  |  |  |  |  |  |  |  |
| 少派 |  |  |  |  |  |  |  |  |  |  |  |

其中0<α<2，5<β<10。

1. 将专家方案和推荐方案做差值，定义阈值，若总扣分数，则该推荐方案合理；反之结合经验修正方案并返回第二步。若存在步骤3所述的多种推荐方案，则取推荐方案扣分数最小的方案。

整个评价过程如下图所示：



**图3 评价流程图**

5.仿真结果

5.１数据处理

本文的案例库来自上海市消防局近几年的消防案件，案件信息表的关键字段包括案件类型，建筑结构，立案时间，案件等级，处置对象，过火面积，烟雾情况，受伤人数，燃烧楼层和楼房层数，气象信息等。数据相关属性与说明见表1：

**表１　案件特征描述**

|  |  |  |  |
| --- | --- | --- | --- |
| 属性名称 | 属性说明 | 属性类型 | 备注 |
| 案件类型 | 3个类型 | Var char |  |
| 案件等级 | ５个等级 | Var char |  |
| 建筑结构 | 多种结构 | Var char |  |
| 处置对象 | 多种类型 | Var char |  |
| 烟雾情况 | 多种类型 | Var char |  |
| 受伤人数 | 整数 | int |  |
| 燃烧楼层 | 整数 | int |  |
| 楼房层数 | 整数 | int |  |
| 气象信息描述 | 多种类型 | Var char | 案发当天的天气情况 |
| 处置方案 | 多种类型 | Var char | 消防车辆的调派决策 |

为便于分类和案例的检索，我们将离散型的多类型特性使用一位有效编码（独热编码），数值型属性根据数值大小划分为“多”“少”“高”“低”等并编码。

４.２结果对比

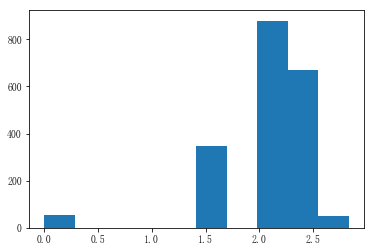
由于案例推理的准确度跟案例库的大小有关，并且根据推理结果无法直接定量确定推理的准确度，只能从原方案的基础上确定推理的合理性。本文使用了案例库的全部样本数据（约6W+），并选用其中的60条作为测试。在使用权重更新算法确定属性权重后应用案例推理方法计算目标案例和原案例库的相似度，得出的结果如下表2（仅选取前十几条数据）：

**表２　两种算法的结果对比**

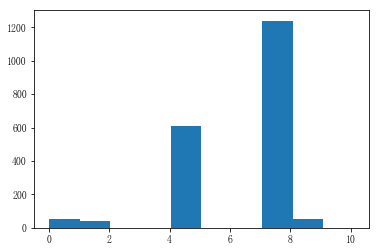
|  |  |  |  |
| --- | --- | --- | --- |
| **案例编号** | **原样本案例方案** | **均权推荐方案** | **加权推荐方案** |
| 1 | 泡沫消防车\*1 | 压缩空气泡沫消防车\*1 | 压缩空气泡沫消防车\*1 |
| 2 | 通讯指挥消防车\*1，抢险救援消防车\*2，泡沫消防车\*2，压缩空气泡沫消防车\*3，器材消防车\*1 | 泡沫消防车\*6，云梯消防车\*1，救援抢险消防车\*1 | 通讯指挥消防车\*1，云梯消防车\*1，泡沫消防车\*6，压缩空气泡沫消防车\*1 |
| 3 | 救援抢险消防车\*1，泡沫消防车\*1 | 救援抢险消防车\*1，压缩空气泡沫消防车\*1 | 救援抢险消防车\*1，压缩空气泡沫消防车\*1 |
| 4 | 泡沫消防车\*1 | 压缩空气泡沫消防车\*1 | 压缩空气泡沫消防车\*1 |
| 5 | 压缩空气泡沫消防车\*2，泡沫消防车\*2 | 泡沫消防车\*3，云梯消防车\*1 | 泡沫消防车\*3，压缩空气泡沫消防车\*1 |
| 6 | 泡沫消防车\*1 | 泡沫消防车\*1 | 泡沫消防车\*1 |
| 7 | 泡沫消防车\*1 | 泡沫消防车\*1，压缩空气泡沫消防车\*1 | 泡沫消防车\*1，压缩空气泡沫消防车\*1 |
| 8 | 压缩空气泡沫消防车\*2，泡沫消防车\*2 | 压缩空气泡沫消防车\*2，泡沫消防车\*1 | 压缩空气泡沫消防车\*2，泡沫消防车\*1 |
| 9 | 泡沫消防车\*1 | 救援抢险消防车\*1 | 救援抢险消防车\* |
| 10 | 救援抢险消防车\*1，泡沫消防车\*1 | 救援抢险消防车\*1，压缩空气泡沫消防车\*1 | 救援抢险消防车\*1，压缩空气泡沫消防车\*1 |
| 11 | 压缩空气泡沫消防车\*2，泡沫消防车\*2 | 泡沫消防车\*3，救援抢险消防车\*1，压缩空气泡沫消防车\*1 | 泡沫消防车\*3，压缩空气泡沫消防车\*1，救援抢险消防车\*1 |
| 12 | 泡沫消防车\*1 | 泡沫消防车\*1 | 泡沫消防车\*1 |
| 13 | 泡沫消防车\*2，压缩空气泡沫消防车\*1，救援抢险消防车\*1，云梯消防车\*1 | 压缩空气泡沫消防车\*1，泡沫消防车\*3，救援抢险消防车\*1 | 压缩空气泡沫消防车\*1，泡沫消防车\*3，救援抢险消防车\*1 |

由上表数据可以看到，使用上述算法更新权重计算加权相似度后的推荐方案与均权法的推荐方案存在一定的差别，加权相似度的案例推理推荐方案更接近于原方案。

下图是两种方法计算的相似度直方图，选取了一个目标案例计算与源案例库各案例的相似度对比，横坐标为相似度距离，纵坐标为数量：



**图２　均权法计算的相似度距离直方图**

****

**图３　特征加权相似度距离直方图**

对比图可知，使用加权KNN计算的相似度距离分布更加分散，特别是在计算出的相似度距离很小的时候能够有一个明显的区分，确保了一定的检索精度。

５.结论和改进

本文通过对案例检索中的相似度算法进行研究，利用改进的权重更新算法计算出案例库中各案例的特征权重并将其应用于案例推理的相似度计算中，实验结果表明使用加权相似度的案例推理方法具有更高的检索精度，验证了所提出方法的真实有效性，进一步为应急救援提供了有力的支撑。

本文还结合消防救援规则设计了一种评估案例推理方案合理性的扣分表，能够快速合理的对模型推荐出的救援派车方案进行打分，结合专家经验对推荐方案迅速进行判断，能够有效节省消防救援辅助决策的时间。

案例推理模型只是起到一个辅助决策的作用，还需要经过专家系统的验证和评估才能将修改后的方案应用于应急决策中。另一方面，由于案例库的案例种类较少和案例表示的方法有待提高，往往会对案例推理的结果产生一定的影响。在以后的工作中还会考虑到提取案例的文本信息并对案例进行分类，更精确的表示案例的各个特征属性，并结合神经网络和案例推理模型提高检索精度和速度。

参考文献

［１］李航. 统计学习方法[M]. 清华大学出版社, 2012.

［２］杜渂, 侯忠辉, 刘亮亮. 基于特征加权FCM算法的消防火灾等级研究[J]. 网络安全技术与应用, 2016(4).

［３］杨健, 赵秦怡. 基于案例的推理技术研究进展及应用[J]. 计算机工程与设计, 2008, 29(3):710-712.

［４］郭泳亨, 卢兴华, 刘云. 基于案例库的应急决策支持系统研究[J]. 微计算机信息, 2006, 22(24):148-150.

［５］石浩. 基于案例推理的城市应急决策支持系统的研究[D]. 浙江工业大学, 2005.

［６］王凯. 基于案例推理的应急管理案例库构建方法研究[D]. 上海交通大学, 2009.

［７］张凯岚. 基于案例推理与神经网络的建筑成本预测研究[D].

［８］李金秀, 高新波, 高玉娥, et al. 基于特征加权的模糊聚类算法研究[J]. 北京电子科技学院学报, 2007, 15(2):74-76.

［９］Ricci F, Avesani P, Perini A. Cases on Fire: Applying CBR to

Emergency Management[J]. New Review of Applied Expert System, 1996, 6(1): 175-190.