

# 基于特征加权 FCM 算法的消防火灾等级研究

杜 漫<sup>1</sup> 侯忠辉<sup>2</sup> 刘亮亮<sup>3</sup>

## 0 引言

随着经济社会的高速发展,公共安全与应急领域越来越受到国家与相关政府部门的重视,借助信息化手段,该领域已经取得了不少进步,多个省市已经实现了消防一体化的建设。对于消防管理来说,防火灭火是核心,而火灾等级也影响了灭火救援过程以及灭火预案的制定。

火灾等级是衡量火灾大小的一个标准,反映的是火灾危险性的规模,火灾的大小是相对的,只有通过等级才能确定火灾的规模,才能进行救援力量的调配。目前,在世界范围内,对于火灾等级的划分,没有一个统一的划分标准。通常,火灾等级的划分是根据影响火灾发展的多个因素以及专家的经验来确定的,而往往现实是,靠专家的经验多一点,没有科学性。如今,随着数据挖掘技术快速发展,利用大量火灾警情数据与灭火预案数据,可以结合有效的数据模型,科学的计算出火灾等级,以此来加强消防救援管理的科学性,减少人工干预,使得消防事业更加信息化和智能化。

## 1 背景研究

### 1.1 国内火灾等级划分

根据 2007 年 6 月 26 日公安部下发的《关于调整火灾等级标准的通知》,新的火灾等级标准由原来的一般火灾、重大火灾、特大火灾三个等级调整为一般火灾、较大火灾、重大火灾、特别重大火灾四个等级<sup>[1]</sup>,参考某市消防局的火灾业务数据,火灾等级是从高到低由数字级别标示的。

一级火灾:一般的火灾。火灾造成 3 人以下人员死亡,或者 10 人以下人员受伤,或者 1000 万元以下财产损失;

二级火灾:较大的火灾。火灾造成 3 人以上 10 人以下人员死亡,或者 10 人以上 50 人以下人员受伤,或者 5000 万元以上 1 亿元以下财产损失;

三级火灾:重大火灾。火灾造成 10 人以上 30 人以下人员死亡,或者 50 人以上 100 人以下人员受伤,或者 5000 万元以上 1 亿元以下财产损失;

四级火灾:特别重大火灾。火灾造成 30 人以上人员死亡,或者 100 人以上人员受伤,或者 1 亿元以上财产损失。

对于本文火灾数据中直接财产损失、死亡人数、受伤人数等特征值的处理也尽可能采用这种规则,以减少模型计算与实际可操作性之间的误差。

### 1.2 火灾等级划分存在的问题

城市火灾属于突发事件的一种,近年来,关于突发事件的分级问题取得了不少成果,对突发事件的分类分级法通常分三种类型<sup>[2]</sup>:一是根据事件发生后的影响和严重程度来划分,二是以主观经验打分为基础,评价级别标准和级别数一成不变;三是分类分级随着突发事件的发展不断进行调整。分析来看,第一种属于事后评估,第二种属于静态分级评估,第三种属于动态分级评估,但在实际进行分级决策时,由于突发事件的类型不同,各类型的评估因素也不相同,因此,分级决策仍然主要依靠决策者的直觉和经验来确定。决策者依靠直觉和经验所做的决策主观性和随意

性较强,不能客观地反映突发事件的级别特征。

细化到算法模型研究方面,韩兴等人基于 Petri 网提出火灾救援指挥过程模型<sup>[3]</sup>,朱霁平等人基于 GIS 和 Petri 网的火灾扑救调度决策<sup>[4]</sup>依据文献<sup>[5]</sup>制定火灾等级规则。应急等级划分研究方面,文献<sup>[6]</sup>通过单因素模糊评价矩阵和等级影响因素的权重集进行模糊综合评判,其中权重集没有做出定量分析,只是通过专家问卷并统计得出的。杨静等人的研究<sup>[7]</sup>从系统的角度,结合聚类分析和判别方法提出了一套突发事件动态分类分级的研究思路和方法,没有对其提出的方法进行实现和验证,对于突发事件在发生、发展以及对其处置过程中,类别和级别如何转移以及转移概率的确定没有涉及。考虑等级的模糊性因素,基于模糊数学理论,徐海量<sup>[8]</sup>等人采用模糊聚类算法(Fuzzy c-Means Clustering)实现了洪水灾害等级划分的应用;利用模糊聚类算法,龚雪<sup>[9]</sup>实现了火灾早期探测系统的应用。

郑华利等人提出了改进的 FCM 算法<sup>[10]</sup>,应用于图像分割并提高了图像分割的精度,并且对于图像分割应用,杨红颖等人<sup>[11]</sup>引入 Relief 技术进行特征加权,在 FCM 图像分割算法基础上,有效地提高了图像分割效果,充分考虑了特征矢量对样本贡献不均的情况。对于引入特征加权算法的 FCM 算法方面,国内外都有大量的研究成果,张国锁等人<sup>[12]</sup>通过引入特征加权算法对模糊 FCM 聚类算法进行优化提高了聚类精度;李金秀等人对传统 FCM 算法与基于特征加权的 FCM 进行了试验对比<sup>[13]</sup>,试验结果说明了基于特征加权的 FCM 算法具有更低的误分率;袁正午等人利用基于特征加权的 FCM 算法实现在基因表达数据分析上的应用<sup>[14]</sup>,获得了较高精度的聚类结果。

综合考虑火灾等级的动态模糊性因素,火灾历史数据特性,本文以客观历史数据作为分析的样本数据,借助特征加权算法区分火灾的特征权重,结合模糊聚类分级算法,研究出消防火灾等级模型,并采集某市消防局 2013-2015 年的火灾警情数据进行仿真实验,验证算法模型的有效性 with 准确性。

## 2 基于特征加权 FCM 算法

基于特征加权的 FCM 算法是传统的模糊聚类算法的扩展,在新算法中通过一定的特征选择方法为每维特征进行自动赋权,从而实现特征的自动加权和加权后样本的聚类分析。本文利用 ReliefF 算法实现各属性的特征值提取。

### 2.1 FCM 算法

FCM 算法是一种区别于硬性划分(Nonfuzzy)的聚类算法,最早由 Duda 和 Hart 在 1973 年提出,之后,由 Bezdek 在 1981 年进行了优化<sup>[15]</sup>,其算法思想是使得被划分到同一类的对象之间相似度最大,而不同类之间的相似度最小。设样本  $Y = \{y_1, y_2, \dots, y_N\} \subset R^n$ ,  $Y = \{y_1, y_2, \dots, y_N\} \subset R^n$ ,  $c$  为样本的类别数,且  $c$  满足  $2 \leq c < n$ ,则 FCM 算法实现的目标函数如下:

$$J_m = \sum_{k=1}^c \sum_{i=1}^N (u_{ik})^m (d_{ik}^2) \quad (1)$$

其中  $m$  为加权指数,取值范围为  $1 < m < \infty$ ,  $U$  为样本  $Y$  的隶属度矩阵,  $v$  为聚类中心,  $u_{ik}$  为样本数据  $y_k$  对第  $i$  个聚类中

心的隶属函数，并且满足  $u_{ik} \in [0,1], 1 \leq i \leq c, 1 \leq k \leq N$   
 $u_{ik} \in [0,1], 1 \leq i \leq c, 1 \leq k \leq N$ 。式中  $d_{ik}^2, d_{ik}^2$  为样本数据  $y_k, y_k$  与聚类中心  $v_i, v_i$  的距离，表示如下：

$$d_{ik}^2 = \|y_k - v_i\|_A^2 \quad (2)$$

$v_i, v_i$  为类别  $i$  的聚类中心。对于任意的  $j$  和  $k$ ，当  $m > 1$ ， $y_k \neq v_j$ ， $j_m m > 1$ ， $y_k \neq v_i$  收敛的必要条件是：

$$v_i = \frac{\sum_{k=1}^N (u_{ik})^m y_k}{\sum_{k=1}^N (u_{ik})^m}, 1 \leq i \leq c \quad (3)$$

$$u_{ik} = \left( \frac{d_{ik}^2}{\sum_{j=1}^c (d_{ik}^2)^{2/(m-1)}} \right)^{-1}, 1 \leq k \leq N; 1 \leq i \leq c \quad (4)$$

满足公式 (3) 和公式 (4) 的目标公式 (1) 是一个不断迭代过程，直到收敛则停止继续迭代，FCM 算法具体步骤描述如下：

- (1) 确定聚类数目  $c$ ，初始化  $m$ ；
- (2) 第  $k$  轮迭代，通过公式 (3) 计算聚类中心向量；
- (3) 计算并更新隶属度矩阵  $U^{k+1} = [u_{ik}^{(k+1)}] U^{k+1} = [u_{ik}^{(k+1)}]$ ；
- (4) 如果  $\|U^{(k+1)} - U^{(k)}\| < \varepsilon$  或  $\|U^{(k+1)} - U^{(k)}\| < \varepsilon$ ，则停止迭代，否则回到步骤 2) 继续迭代。

## 2.2 基于 ReliefF 的 FCM 算法

由于火灾等级的影响因素较多，各个因素的影响程度也各不相同，为提高算法的实际可用性，采用 ReliefF 算法对火灾各个影响因素进行加权分析。ReliefF 算法最初是由 K Kira 和 LA Rendell 在 1992 年提出<sup>[16]</sup>，用来解决多维度的分类问题，在 1994 年，由 Kononenko 进行了改进和扩展<sup>[17]</sup>，提出 ReliefF 算法。ReliefF 算法支持更多类别的分类计算，同时对数据类型没有限制，属于一种特征权重算法。算法的主要原理是采用赋予和类别相关性高的特征属性以更高的权重值，相应的会赋予和类别相关性低的特征属性以更低的权重值。

设训练数据集  $D$ ，样本抽样次数  $m$ ，任意选择样本实例  $dR_q$ ， $R_q$ ， $C$  类样本出现的概率  $p(C)$ ，则 ReliefF 算法如式 (5) 所示：

$$W(S) = W(S) - \sum_{j=1}^k \text{diff}(S, R_q, H_j) / (mk) + \sum_{C \in \text{class}(R)} \frac{p(C)}{1 - p(\text{Class}(R))} \sum_{j=1}^k \text{diff}(S, R_q, M_j(C)) / (mk) \quad (5)$$

式中  $W(S)$  为特征  $S$  的特征权重值； $H_j, H_j$  为  $D$  中与  $R_q, R_q$  同类的  $k$  个最近邻中的第  $j$  个样本实例， $M_j(C), M_j(C)$  为  $D$  中与  $R_q, R_q$  不同类的  $k$  个最近邻中的第  $j$  个  $C$  类样本实例，且  $j$  均满足  $j \leq k, j \leq k$ 。式中  $\text{diff}(S, R_1, R_2), \text{diff}(S, R_1, R_2)$  函数表示样本  $R_1, R_1$  与样本  $R_2, R_2$  在特征  $S$  上的差异，其函数计算方法如下：

$$\text{diff}(S, R_1, R_2) = \begin{cases} \frac{|R_1[S] - R_2[S]|}{\max(S) - \min(S)} & S \text{ 为连续型特征属性} \\ 0 & S \text{ 为离散型特征属性，且 } R_1[S] = R_2[S] \\ 1 & S \text{ 为离散型特征属性，且 } R_1[S] \neq R_2[S] \end{cases} \quad (6)$$

基于 ReliefF 的 FCM 算法表示如下：

$$I_m = \sum_{k=1}^N \sum_{i=1}^c (u_{ik})^m W(S) (d_{ik}^2) \quad (7)$$

其中  $W(S)$  为样本数据  $y_k, y_k$  的权重矩阵。与非特征加权相比，特征加权模糊聚类算法解决了不同差异特征属性的贡献率问题，也摒弃了专家打分的不科学性问题。

## 3 特征加权 FCM 算法在消防火灾等级的应用

### 3.1 数据处理

本文数据样本集来自某市消防局 2013-2015 年的火警数据。这些数据包括火灾的接处警明细、火灾发生的单位、火场信息、火灾死伤人员信息、火灾等级、救援力量出动情况等。通过数据仓库等数据处理技术获取火灾的建筑类别、建筑耐火等级、建筑结构、起火场所、起火物、火灾原因分类、直接财产损失、过火

面积、死亡人数、受伤人数以及火灾等级，分别对所有的特征字段进行编号，数据相关属性值类型与说明如表 1 所示。

表 1 属性值类型说明

属性名称	属性说明	特征属性编号
火灾序号	数据集编号，不属于特征值	
建筑类别	离散值。建筑物类别	1
建筑耐火等级	离散值。五个等级	2
建筑结构	离散值。六种建筑结构	3
起火场所	离散值。起火发生的场所	4
起火物	离散值。	5
火灾原因分类	离散值。	6
直接财产损失	连续值。单位为万元	7
过火面积	连续值。单位为百平米	8
死亡人数	整数值	9
受伤人数	整数值	10
火灾等级	离散值。分类属性，1-4 四个等级	

抽取的样本集中建筑类别属性分为高层、民用、小于 50 米、大于等于 50 米且小于 100 米、大于 100 米、工业、多层、老式居民住宅楼、单层、地下等；建筑结构属性分为木结构、砖木结构、砖混结构、钢砼结构、钢结构等；起火物属性分为建筑构件、建筑材料、家具、易燃、易爆物品等；火灾原因分类有电气火灾、电气线路故障、短路、负荷、接触不良、断路、漏电、配电盘故障等。这些属性值均用自然数进行标准化处理。

根据《关于调整火灾等级标准的通知》采取表 2 规则对数据进行标准化方式处理。

表 2 直接财产损失数据处理规则

财产损失值（万元）	属性值
小于 1000	1
小于 5000	2
小于 10000	3
大于等于 10000	4

### 3.2 火灾等级计算评定

为了消除数据不同量纲和数量级上的差异对聚类效果的影响以达到目标函数快速收敛的目的，在实验开始前对上述赋值处理后的数据通过 Matlab 规格化处理。基于规格化处理后的火灾数据，使用 Matlab 工具对上述 ReliefF 算法进行编程，其中取抽样次数  $m=200$ ，类临近值  $k=3$ ，计算权重值，每次结果计算均经过 200 次以上迭代，由公式 (5) 和公式 (6) 计算 5 次得出上述 10 个特征属性权重值矩阵为：

$$W = \begin{bmatrix} 0.2716 & 0.3084 & 0.3265 & 0.4703 & 0.1412 & 0.2055 & 0.5325 & 0.7265 & 0.3944 & 0.5124 \\ 0.2131 & 0.1124 & 0.3023 & 0.4472 & 0.1438 & 0.1801 & 0.5634 & 0.7437 & 0.3621 & 0.4531 \\ 0.2449 & 0.1477 & 0.3244 & 0.4420 & 0.1301 & 0.1947 & 0.5765 & 0.6911 & 0.3167 & 0.4202 \\ 0.2181 & 0.1246 & 0.3293 & 0.4288 & 0.1380 & 0.1751 & 0.5181 & 0.6630 & 0.3527 & 0.4276 \\ 0.2531 & 0.3006 & 0.3459 & 0.4312 & 0.1725 & 0.1832 & 0.5146 & 0.7201 & 0.3750 & 0.4708 \end{bmatrix}$$

由权重矩阵  $W$  可以看出特征属性 8 拥有最大的权重，由该算法可知过火面积是最重要的特征属性，其次是特征属性 7 和特征属性 10；特征属性 2、特征属性 5 与特征属性 6 的权重可以忽略不计。为了优化分类的计算，在进行聚类时，去掉上述 2（建筑耐火等级）、5（起火物）和 6（火灾原因分类）三个特征属性。为了验证算法公式 (7) 的有效性，在样本集中随机抽取 10 条火灾灾情数据训练集，训练集覆盖四个等级的火灾数据，另外抽取 5 条待评定火灾测试集，测试集火灾等级初始化为零，去除特征属性 2、特征属性 5 与特征属性 6 后并进行上述相关数据处理后

的数据如表 3 所示。

表 3 火灾样本数据集

火灾序号	建筑类别	建筑结构	起火场所	直接财产损失	过火面积	死亡人数	受伤人数	火灾等级
1	5	3	4	4	12000	58	71	4
2	10	3	119	3	40	3	0	2
3	11	2	119	3	100	3	0	2
4	5	3	4	4	16000	78	98	4
5	11	3	119	3	28	3	0	2
6	10	3	119	4	40	3	0	2
7	11	5	78	4	11000	0	1	3
8	8	3	2	1	1	0	0	1
9	10	3	21	1	2	0	0	1
10	11	5	78	4	11000	0	1	3
11	11	5	78	2	8000	0	1	0
12	11	3	1	1	20	0	0	0
13	4	4	4	1	1	0	0	0
14	11	2	21	1	30	0	0	0
15	5	3	4	4	11000	28	91	0

对去除低权重属性后的数据集进行规格化处理，同样按照上述处理步骤取  $m=200$ ， $k=3$ ，通过公式（5）与公式（6）计算一次的特征权重为：

$$W=[0.1616\ 0.5098\ 0.5784\ 0.4756\ 0.6828\ 0.1665\ 0.3498]$$

特征权重  $W$  对应的权重属性分别为建筑类别、建筑结构、起火场所、直接财产损失、过火面积、死亡人数、受伤人数。利用 Matlab 对公式（7）进行编程，取公式（7）中聚类数目  $m=4$ ，迭代 20 次左右后目标公式（7）收敛，得出隶属度矩阵为：

$$U_{1-17} = \begin{bmatrix} 0.0335 & 0.0324 & 0.0454 & 0.0144 & 0.0199 & 0.0888 & 0.9398 \\ 0.0314 & 0.0204 & 0.0388 & 0.0092 & 0.0116 & 0.0429 & 0.056 \\ 0.9150 & 0.0137 & 0.0245 & 0.9648 & 0.0065 & 0.0344 & 0.0140 \\ 0.0301 & 0.9335 & 0.8913 & 0.0116 & 0.9620 & 0.8339 & 0.0306 \end{bmatrix}$$
  
$$U_{8-15} = \begin{bmatrix} 0.0350 & 0.0201 & 0.8318 & 0.7892 & 0.0253 & 0.1475 & 0.0693 & 0.2082 \\ 0.9040 & 0.9465 & 0.0410 & 0.0810 & 0.9376 & 0.6434 & 0.8082 & 0.1406 \\ 0.0215 & 0.0099 & 0.0480 & 0.0428 & 0.0127 & 0.0901 & 0.0395 & 0.5173 \\ 0.0395 & 0.0235 & 0.0792 & 0.0071 & 0.0244 & 0.1190 & 0.0830 & 0.1339 \end{bmatrix}$$

由上述隶属度矩阵和模糊聚类理论可知当前数据集火灾的四类等级分类情况为：{7 10 11}、{8 9 12 13 14}、{1 4 15}、{2 3 5 6}，对比样本集中 1-10 条已知火灾等级由此可知火灾 1 1 评定为 3 级火灾，火灾 12、13 以及 14 评定为 1 级火灾，火灾 15 评定为 4 级。通过历史数据给定的等级值可知最后的计算值有效，并且基本符合 2007 年给定的火灾等级划分规则。

4 结束语

由于火灾等级作为火灾事后报告所用的特性，决定了其在灭火救援中的特殊应用场景。采用基于 ReliefF 特征加权的 FCM 聚类算法模型，计算出各个特征属性的权重，由于拥有更高权重值的特征属性具有更高的贡献度，去掉权重值低的特征权重，结合 FCM 算法，得到火灾等级的隶属度矩阵，以此达到等级分类的目的。

采用 Relief 算法计算权重的时候，我们发现火灾中建筑耐火等级、起火物、火灾原因分类在火灾等级计算中的作用是微乎其微的，在灭火救援的力量调派方面，需要多少救援车或救援人员，这三个属性相对来说可以不用考虑。相对的，过火面积因素占有最高的权重，而与实际也是相符的，在实际生活中，鉴于灭火救援的紧迫性与危险性，可以考虑适当增加过火面积因素等权重相对较高的权重值，在不失科学性的基础上使得更具可操作性。

最后实验结果表明，基于 ReliefF 特征加权的 FCM 算法在消防消灭等级的应用方面是可行的，并且达到了很好的效果，为进

一步在火灾等级与救援力量的量化分析方面提供了基础。

由于火灾警情的数据采集大部分由人工采集完成，火灾的现场状况也是动态变化的，数据的准确性和时效性也会影响最终的等级判定结果，因此可以考虑采用多次采集的方法。在此基础上，下一步将会在防火救援力量的调度方面进行重点研究。

引用：

[1]小方.公安部调整火灾等级死亡 30 人属于特大火灾[J].消防技术与产品信息, 2007.

[2]商丽媛, 谭清美.基于支持向量机的突发事件分级研究[J].管理工程学报, 2014.

[3]韩新, 李杰, 沈祖炎.城市火灾扑救调度指挥过程的 Petri 网模型研究[J].自然灾害学报, 1998.

[4]朱霁平, 范维澄, 薛海生, 廖光煌.一种基于 GIS 和 Petri 网的城市火灾扑救调度决策方法[J].中国科学技术大学学报, 2005.

[5]陈文贵.中国城市消防管理手册[M].中国人民公安大学出版社, 1992.

[6]王富, 李杰, 石永辉.城市事故灾难交通应急等级模糊综合评判模型及应用[J].湖北大学学报：自然科学版, 2013.

[7]杨静, 陈建明, 赵红.应急管理中的突发事件分类分级研究[J].管理评论, 2005.

[8]徐海量, 陈亚宁.洪水灾害等级划分的模糊聚类分析[J].干旱区地理, 2000.

[9] 龚雪.模糊聚类分析在火灾早期探测系统中的应用[J].消防科学与技术, 2008.

[10]郑华利, 周献中, 王建宇.FCM 图像分割算法的特征分析与改进[J].计算机工程, 2004.

[11]杨红颖, 王向阳, 王春花.基于特征加权的自适应 FCM 彩色图像分割算法[J].计算机科学, 2009.

[12]张国锁, 周创明, 雷英杰.改进 FCM 聚类算法及其在入侵检测中的应用[J].计算机应用, 2009.

[13]李金秀, 高新波, 高玉娥, 等.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.

[14]袁正午, 魏荣, 叶明星.一种适用于基因表达数据的特征加权 FCM 算法[J].计算机应用研究, 2010.

[15]Bezdek J C,Ehrlich R,Full W.FCM:The fuzzy c-means clustering algorithm[J].Computers & Geosciences,1984.

[16]Kira K,Rendell L A.The feature selection problem : traditional methods and a new algorithm[C]// Proceedings of the tenth national conference on Artificial intelligence AAAI Press,1992.

[17]Kononenko I.Estimating attributes:analysis and extensions of RELIEF[C]//Proceedings of the European conference on machine learning on Machine LearningSpringer-Verlag New York,Inc.,1994.

基金项目：上海市科学技术委员会专项基金资助项目（14511107600，15XD1521100）；上海市经济和信息化委员会专项资金资助项目（140413）；

- (1.上海迪爱斯通信设备有限公司 上海 200032；
- 2.电信科学技术第一研究所 上海 200032；
- 3.上海市消防局 上海 200051)