

## 第 3 章 灾害节点属性的量化及分级方法

第 2 章我们给出了灾害复杂网络的整体结构，在这种结构下，网络灾害节点是与具体的历史灾害记录相对应的，这就需要对这些历史灾害记录进行合理的属性定义和量化才能转化成有效数据。历史灾害记录数量庞大并且以描述性语言为主，只有确定合理的量化方案才能使得生成的数据使用起来科学、高效。本章首先针对灾害记录量化中比较关键的等级划分问题提出一种基于蚁群聚类的历史灾害分级方法；以地震灾害为例，基于混合模糊算法，针对灾种特殊属性量化问题提出一种判定历史地震烈度的方法。

### 3.1 以蚁群聚类为基础的历史灾害分级方法

#### 3.1.1 研究现状分析

我国具有悠久的历史文化和丰富的长序列资料库，各朝代正史都记载了各种灾难的产生和危害，为后人在灾害识别、防治方面提供了很好的理论实践参考，也得到了国内外研究专家的高度好评。国家减灾委指出，历史灾害信息平台是我国灾害空间建设的重要内容组成，随后与之相关的《国家自然灾害空间信息基础设施总体构思》也在 2011 年 5 月正式提出[6]。

由于科技发展的局限性，古代历史灾害记录大多以语言描述为主，缺乏具体的属性值和可比性。以“十二月戊戌，开封府陈留等六县水灾，诏免其田租”为例，这条较为典型的历史灾害记录仅简略交代了灾难种类、发生的时间和地点、政府反映等信息，缺乏关键的数据如具体伤亡和经济损失数据。由于涉及较多的人为观念，其数据量化难度高，不足以使人信服。根据现代翔实灾害记录衍生出的各类灾害分级方法与描述性的历史灾害记录兼容性很低，在灾害的还原选择以及量化上具有一定难度。与历史灾害记录相比，现代灾害分级方法的主要依据是客观准确的灾害属性统计结果，如伤亡人数、受灾面积、倒房数量、直接经济损失等，这些数据具有较高的量化程度和可比性。针对于此，神经网络[167]表现出了一定的优势，其能够有效地对非线性函数进行拟合，从而实现模式识别和聚类分级，在一定程度上降低了现代灾害记录分级方法[153-156]的局限性。但是历史

灾害分级没有准确先例，因此利用神经网络必须先通过专家进行人工分级，再对结果进行提炼，因此该阶段依然要受人的主观因素影响。

由此，可将历史灾害归结为以下分级问题：

(1) 分级标准问题。由于不同专家各自评判的标准不同，会对同一灾害记录产生不同的评级结果。针对历史灾害记录的描述性和主观性，采用聚类法，能够摆脱传统的具体分级标准，从而提高数据的客观性。

(2) 数据量化问题。由于历史灾害记录描述格式的不规范，要求研究者不仅要确定什么信息能够代表灾害的本质特征，还要结合灾害的其他个性信息对其进行调整。

(3) 批量处理问题。处理数据量的大小是与分级方法的运算速度直接相关的，大样本会产生较低的运算效率，因此算法的时效性是历史灾害记录分级处理必须要考虑的因素。

(4) 专家监督问题。聚类结果离不开历史专家的认可，缺乏表述性的历史灾害数据系统更需要专家的意见。历史灾害记录具有主观性，如果专家之间无法达成一致，则一味依赖专家意见会导致分级过程缺乏客观性。针对这一缺陷，可以在大方向上采取专家的指导意见，对其加以权衡。

与此相关新的灾害聚类分级方法有模糊聚类[158,159]、灰色聚类[160,161]等，但仍受主观参数和高度量化的灾害记录影响，如模糊聚类必须手动设置合适的  $\alpha$  值。鉴于此，本章针对历史灾害分级问题采用先聚类再分级的方法。为了有效减免灾害分级流程中的主观因素，我们采用一种蚁群聚类算法自动聚类历史灾害记录，而灾害分级则是通过判断各聚类中心的大小实现，最终通过实验证明它的有效性。

### 3.1.2 蚁群聚类思想

按有关特性的相似程度，将物理或抽象的数据集合进行分组，这一过程被称为聚类。灾害分级问题就是一种层次聚类问题。聚类能够提高同一组中各数据属性的拟合率和不同组数据间属性的相斥率，这一点与灾害分级对等级内和等级间灾害数据属性的要求是一致的。

1991年，蚁群算法[162,163]首次被提出，从此研究者根据观察蚂蚁行为提出了多个对应模型并被广泛应用于多个领域。孵化分类(brood sorting)行为是一种蚂蚁常见的群居行为，巢穴孵化区的中心放置着成束的蚂蚁卵和小幼虫，孵化束外围排列着最大的幼虫。1991年，Deneubourg等人按照蚂蚁根据环境物品数量进行取舍的原理，对这一现象建立了一个模型[164-166]。例如，一只蚂蚁很有可能

会将携带的一个小卵存放在相同卵密集的区域；与此不同的是，单独位于一堆小卵中的大幼虫被未携带任何物品的蚂蚁带走的概率会比较大。除此之外，蚂蚁几乎不会对物品进行取放操作。

基于上述蚂蚁行为，本章从中提炼出一种蚁群聚类模型，以进行历史灾害分级。首先，依据专家意见，划分出四个灾害等级，分别为轻灾、中灾、重灾、特重灾，换言之，将历史灾害聚为四类，然后随机分配历史灾害数据到四个组中，引导蚂蚁进入组，将随机混乱的数据进行分门别类的整理和调换。同时仿照上述蚂蚁孵化模型，蚂蚁无论执行何种操作都不是确定性的，而是按照概率进行的。当蚂蚁面对多个组时，往往会针对被其认为最不整齐的组进行整理，但这一过程仍存在着一定概率，因此可以避免算法产生局部最优的问题，可以保证算法的公正性和客观性。K-means 聚类算法在很大程度上会受到初始聚类中心选取质量的影响，蚁群聚类算法则打破了这一局限。蚂蚁会依据各组混乱程度进行聚类中心的调整，能够很好地解决大样本问题。数量众多的样本可以产生层次明确的聚类中心，其产生的聚类结果也更有意义。

### 3.1.3 历史灾害记录的属性定义

结合历史专家的意见，同时分析了大量历史灾害记录中的信息特征，并对所有灾害记录进行了标准化编码，整合出各条记录的灾害信息内容，主要包括以下4项属性：

(1) 受灾范围：指遭受灾害影响的地区，以县为单位，为了精确计量范围，我们通过 GIS 系统实现了现代县制转化，量化程度较高，该项属性是不可缺省的。

(2) 物情：指灾害对房屋、庄稼等实物的影响，如“伤民田”、“水决三十里”和“颗粒无收”等。

(3) 民情：指灾难过后当地人民的状况信息描述，如“溺死人畜”、“民饥”等。

(4) 政情：指灾难过后政府的作为情况，如“祈祷雨泽”、“诏免其租”、“发仓赈济”等。

“物情”、“民情”和“政情”可以大致涵盖每条历史灾害记录的基本灾情信息，但是并不是不可缺省的，实际上历史记录一般也无法全部具备上述4项属性。为了使这三项属性能够正确发挥区别等级的作用，必须科学地定制出各项属性的编码方案，从而将这些历史灾害记录量化并批量载入灾害数据库。以“政情”属性为例，其可分为无记录、祈祭、遣官、减租税粮、免租税粮、赈济六个属性级别，若将其分别设置为0、1、2、3、4、5六个编码，则可将历史灾害记录中的诏

免盐粮、诏免税粮、免粮、免征秋粮、免征田粮等描述归纳为“免租税粮”，以“4”进行编码。另外，实际灾情严重程度的上升幅度与0、1、2、3、4、5并不是完全符合的，因此还需要结合专家的评估意见和聚类结果适当调整各等级的对应编码。

属性值通过历史灾害的4个等级和4个数据属性实现了初步量化，为实现历史灾害数据的聚类提供了处理前提。在对实际历史灾害进行聚类时，4项属性的权重是有所差别的。为了保证聚类的高质量，需要根据专家反馈意见和聚类结果进行权值的修改，换言之，对历史专家专业意见的汲取主要体现在灾害信息量化的过程上，专家的指导监督作用存在于大方向上，而非存在于快速自动化处理的聚类过程中。

### 3.1.4 蚁群聚类算法描述

设历史灾害数据集为 $S$ ，数据属性为 $p$ （本问题中 $p=4$ ）项， $S$ 中的数据被随机分配为 $k$ （此处 $k=4$ ）个组，那么每组有 $t_n$ （ $n \in [1, k]$ ）个灾害数据。算法具体流程如下：

#### （1）数据预处理。

前面已经介绍过，由于历史灾害记录的简略性，通常具有缺失属性值问题，因而必须对这些空值进行合理处理。可结合灰色关联系数定义，针对数据聚合性差这一问题进行处理和归一。具体如公式（3.1）所示。

用 $x_{ni}$ 代表第 $n$ 个灾害数据的第 $i$ 个初始属性值，初次处理后，生成 $y_{ni}$ ：

$$y_{ni} = \frac{1}{2 - (x_{ni} - x_{\min i}) / (x_{\max i} - x_{\min i})} \quad (3.1)$$

其中， $x_{\max i}$ 、 $x_{\min i}$ 分别为灾害数据集 $S$ 中第 $i$ 个属性的最大值和最小值。

#### （2）选出被整理组。

蚂蚁按照公式（3.2）评估各个灾害数据组的不整齐度，用 $chaos_n$ （ $n \in [1, k]$ ）来表示，此组越不整齐那么相应 $chaos$ 值越大，所以蚂蚁对该组进行操作的可能性就越大。第 $n$ 个组的不整齐度表示为：

$$chaos_n = \sqrt{\frac{\sum_{i=1}^{t_n} (x_i - avg_n)^2}{t_n}} \quad (3.2)$$

其中， $x_i = \sum_{j=1}^p y_j^2$ 表示第 $n$ 组中第 $i$ 条数据各属性的平方和， $avg_n = \left( \sum_{i=1}^{t_n} x_i \right) / t_n$ 表示第 $n$ 组各 $x_i$ 的平均值，即 $chaos_n$ 表示第 $n$ 组中各灾害数据属性平方和的标

准差。

然后基于赌轮选择法[164]，根据各个组的 *chaos* 值选择一个出数据组。

(3) 选出一个灾害数据。

按照公式 (3.3)，蚂蚁在遴选出的组中，进行每个灾害数据对该组不整齐度的影响评估，可以用 *influence* 来代表，*influence* 值的大小和此灾害数据的影响度成正比，也就是说值越大的数据其被蚂蚁取出的机率越大。第 *i* 个数据在组中的影响力公式为：

$$influence_i = (x_i - avg_n)^2 \quad (3.3)$$

其中求得  $x_i$ 、 $avg_n$  的方法与第 2 步一样。

然后基于赌轮选择法，依据组中各个灾害数据的 *influence* 值挑选出单个数据进行重新归属。

(4) 选出归属组。

按照公式 (3.4) 评估，用  $similarity_n$  ( $n \in [1, k]$ ) 来表示蚂蚁所选出的灾害数据与各个组的拟合度，较大的 *similarity* 值表示较高的拟合度，会增加该灾害数据放入对应组的概率。与第 *n* 个组的拟合度表示为：

$$similarity_n = \frac{1}{(x_i - avg_n)^2} \quad (3.4)$$

其中求得  $x_i$ 、 $avg_n$  的方法与第 2 步一样。

然后基于赌轮选择法，就各个组的 *similarity* 值蚂蚁可以把选出的灾害数据添加到此组中。

(5) 选择蚂蚁。

为了达到算法收敛的目的，单只蚂蚁是不够的，这里通过多只蚂蚁分别进行 2~4 步来对比效果，然后以公式 (3.5) 对每只蚂蚁整理的质量进行评估 (用 *fitness* 表示)。*fitness* 值与蚂蚁整理的质量成正比，值越高代表质量越好。然后将 *fitness* 值最大的蚂蚁的整理方案定义原聚类。第 *n* 只蚂蚁的整理质量公式为：

$$fitness_n = \frac{1}{t_n * chaos_n} \quad (3.5)$$

其中  $t_n$  为第 *n* 个组的大小， $chaos_n$  为此组的不整齐值。这些值都是蚂蚁操作后更新的值。

(6) 算法收敛。

对上述多只蚂蚁的整理行为进行迭代，直到达到算法终止条件，原则上此时蚂蚁选出的灾害数据不会再放到其他组，亦即聚类结果达到稳定状态。同时各组聚类中心的大小可以作为等级评定标准，即根据其大小顺序划分为轻灾、中灾、

重灾、特重灾四个等级。

蚂蚁的数据结构及算法流程伪代码能够更为方便地理解上述算法的实现过程。

可将蚂蚁的数据结构定义如下：

```

structure single_ant          % 定义数据结构 single_ant
begin
integer from                 % 数据出组
integer in                   % 被整理的灾害数据
integer to                   % 数据归属组
real fitness                 % 适应值
end
single_ant ant[n]           % 将蚂蚁数量设为 n

```

以下是算法的程序伪代码，将数据集分为  $m$  个组，分别用  $S_1, S_2, \dots, S_m$  表示。

```

Procedure AntClustering
InitializeData()             % 灾害数据预处理
while (not terminate) do    % 未收敛会一直循环
    best  $\leftarrow$  0.0      % 标记适应值最高的蚂蚁
    for k = 1 to n do        % 使用 n 只蚂蚁
        ant[k].from  $\leftarrow$  WheelFrom( $S_1, S_2, \dots, S_m$ ) % 赌轮法选择数据出组
        ant[k].in  $\leftarrow$  WheelIn(ant[k].from)           % 赌轮法选择数据出组中的数据
        ant[k].to  $\leftarrow$  WheelTo( $S_1, S_2, \dots, S_m$ ) % 赌轮法选择数据入组
        ant[k].fitness  $\leftarrow$  ComputeFitness(ant[k])    % 评估蚂蚁 k 的操作质量
    if ant[k].fitness > best then
        best  $\leftarrow$  k          % 记录表现最好的蚂蚁序号
    end-if
    end-for
    Updata(ant[best])         % 用表现最好的蚂蚁的操作更新聚类
end-while
end-procedure

```

### 3.1.5 性能分析

神经网络和模糊聚类算法被用来与这里的算法进行性能对比，用来测试的历史灾害为从历史灾害数据库中随机抽取的 500 条明代旱灾记录，请了 3 位历史灾害专家进行人工评级。评级结果作为进行性能比较的神经网络算法训练及测试集的同时，也检验和判断了本章算法的准确率。

为了更好地确认效果，我们根据 3 位专家的评级结果分别训练了 3 个神经网络，并用每个神经网络分别对 3 位专家的评级结果进行了验证，实验结果如表 3.1 所示。

表 3.1 算法匹配率比较

算法	与专家 1 结果 匹配率	与专家 2 结果 匹配率	与专家 3 结果 匹配率	平均匹配率
蚁群聚类	92.3%	89.5%	94.1%	92.0%
神经网络（专家 1）	92.5%	78.9%	80.2%	83.9%
神经网络（专家 2）	78.4%	91.4%	76.0%	81.9%
神经网络（专家 3）	79.1%	82.8%	90.3%	84.1%
模糊聚类	73.0%	70.8%	75.6%	73.1%

我们将历史灾害分级准确率的参考设置为专家分级结果的匹配率，由此可见：

（1）模糊聚类算法的性能是最差的。神经网络的高匹配率只存在于对应专家的评级结果中，而在其他情况下匹配率则不能让人满意。这一结果体现了神经网络需要依赖训练集、专家评级具有主观任意性等问题。

（2）较稳定的性能是蚁群聚类算法的优点，该算法具备了较高的专家匹配率和准确率。

（3）各算法的运算效率比较如图 3.1 所示。相对于神经网络算法的高次数训练，蚁群聚类算法具有较快的收敛速度，在大样本量的历史灾害分级上占有关键优势，具有高时效性。

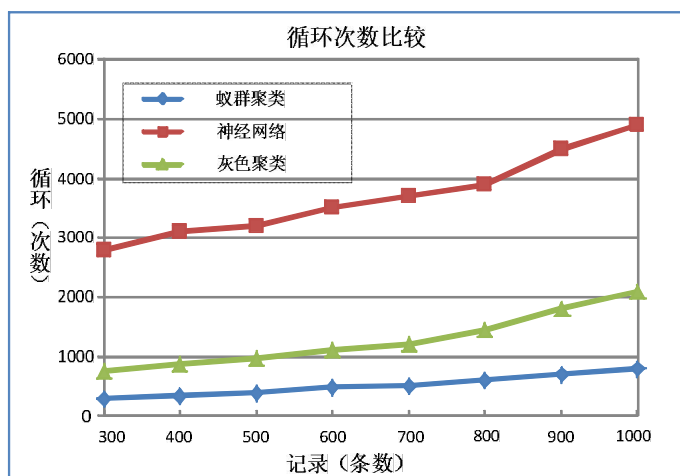


图 3.1 算法循环次数比较

在具体的蚁群聚类分级结果中，这 500 条明代旱灾记录中轻灾为 137 次，约占记录总数的 27%，中灾为 118 次，约占记录总数的 24%，重灾为 196 次，约占

记录总数的 39%，特重灾 49 次，约占记录总数的 10%，可见明代期间遭受旱灾的影响较大，这也印证了历史专家提出的“明代干旱期长，大旱多”的看法。同时需要注意的是，若专家结果明显差异于这里的聚类算法结果或专家对算法聚类结果提出较大意见时，这里算法的聚类效果可以通过调整算法参数及编码权重来加以进一步完善。另外，历史灾害分级的模糊性是实际存在的不可忽视的问题，即便是历史灾害专家的分级也无法就认定是绝对准确的，换言之，当聚类结果能达到和多位专家有较高的匹配率时，我们就可以认为该聚类结果已经具备了一定的指导作用和实用意义。

## 3.2 基于混合模糊算法的历史地震烈度判定方法

在我们建立的复杂网络模型中，灾害节点不仅具有统一的等级、类别、时间、地点等默认属性，针对不同灾种的特点我们还设置了特殊属性字段以更好地还原历史记录，本节就以历史地震的烈度属性为例，利用一种混合模糊算法进行相关的判定工作。

### 3.2.1 研究现状分析

一次地震发生后，由于地震能量的扩散性和衰减性，不同地区所受到的地震影响也不相同，用以描述地震影响强弱程度的标尺就是地震烈度[167]。地震烈度是地震研究中一个至关重要的参数，以建筑物损坏度和地表改变状况为依据，能够描绘出一次地震不同区域的等烈度线，从而确定地震中心和地震级别。

在地震烈度判定上，烈度表[168]是最基本也是最简便的参照标准，目前国际上通用的各种 12 度烈度表以及我国最新的《中国地震烈度表》（全国地震标准化技术委员会，2004）都缺乏对历史地震特征的考虑，仅适用于研究现代地震，对历史地震烈度评定兼容性明显不足。为此我国的地震研究者又根据历史地震记录[169]的特点制定了相应的《历史地震烈度表》，从而为历史地震烈度判定提供了很好的参照规范。

但是，无论是历史地震记录，还是《历史地震烈度表》，都是定性描述的，量化度很低，因此现在的历史地震烈度判定还是依靠历史专家参照《历史地震烈度表》，并结合各种时空背景因素，从而对历史地震记录进行评定，尽管专业性较强，但仍不可避免地会加入个人的主观判断，而且历史地震记录数量繁多，人工评定



的方法难以实现大批量处理，因此，设计出一种准确客观高效的历史地震烈度判定方法是十分必要的。

历史地震记录是我国地震研究中非常宝贵和重要的基础资料，研究地震规律、测定区域地震衰减关系、分析地震危险性、检验重大工程场地地震安全性都离不开历史地震信息调查和烈度评定。但是由于当时科技的局限、记录的主观性和格式的混乱，很难归纳和量化地震烈度特征，并且很多记录存在特征信息缺失问题，这使得基于现代翔实地震记录数据的地震分级及烈度判定方法[170-173]难以套用于历史记录。

因此，针对历史记录特点制定出专门的《历史地震烈度表》并以此作为判定或量化的基础是十分必要的。早在1960年，李善邦在进行历史地震相关研究时，针对评定历史地震烈度和震级，在《中国地震目录》的编制中加入“实用地震烈度与震级简表”，1983年，顾功叙等将其修改补充为《历史地震烈度—震级简表》（简称83表）。李群（1989年）在整理近30年地震史料的基础上，提出了改进的《历史地震烈度表》方案（简称89表）。上述历史烈度表均采用了烈度的12度分档，以保持与现代地震烈度表的对应，但是对其中某些烈度做了归并没有细分，鄢家全等（2010年）在83表和89表的基础上做了进一步的补充完善，给出了最新的《2010历史地震烈度表》（简称2010表），由于烈度I和II仅反映人的轻微感觉，历史地震记录未予记载，所以历史烈度表从III度开始划分。我们即以2010表（如表3.2所示）为基础对历史地震记录进行烈度特征量化，并对烈度III-XII进行判定。

表 3.2 2010 历史地震烈度表

烈度	人与社会反响	房屋建筑	构筑物	地表现象
III	地微震			
IV	地震，地震有声	房屋动		
V	地大震，有声如雷，山谷响应	房屋响动或震落檐瓦	坏土城垛	偶有山石滚落
VI	惊逃户外，偶有伤亡（个别）；	坏民居，屋瓦落，墙壁圯损，间或有倒塌者	坏城堞或城垣，石牌坊轻微损坏，塔顶损坏或坠落，石桥栏杆有损坏	山石滚落
VII	杀人或压杀人畜（数十）	民居多坏，房屋倾圮，房屋倾圮无算或无数，坏官民庐舍，庙宇两庑有损坏者	坏城垣、城楼，或城垣局部倒塌数丈，土城崩，墙垣多圮，塔顶部分震圮，石牌坊顶部有震塌，桥面及桥体震裂	崖岸崩塌，山石崩落，地裂涌水

续表

烈度	人与社会反响	房屋建筑	构筑物	地表现象
VIII	伤亡人畜(数十至数百); 赈济	官民庐舍尽坏, 民居、公廨多倾圮, 官署、学府、监仓等多坏或部分倒塌, 坏宙宇、殿堂, 或庙宇两庑倾圮	大坏城廓, 城垣倒塌数十丈至百余丈或多倒塌, 城楼震圮, 塔崩裂或上部塌落, 石碑坊部分倒塌或个别倒塌, 桥梁震坏或部分震塌	山裂、山崩, 平地开裂, 沙水涌出, 出现新泉, 泉水干涸
IX	压杀人众(数百至上千); 赈济、减免税赋	民居、公廨、官署、学府、监仓等倾圮殆尽或存一二, 庙宇、殿堂多坏或部分倒塌	城垣、边墙、墩台多坍塌或砖城墙崩, 城楼倾圮, 石碑坊大部分倒塌, 塔倾圮, 桥倾圮	山裂、山崩, 阻塞道路或河流, 地坼裂, 黑沙水涌出, 井水外溢或井泉干涸
X	伤亡甚众(数千至万); 赈济、免税赋	民居、公廨、官署、学府、监仓等倒塌殆尽或仅存个别, 庙宇、殿堂多倾圮或仅存一二	城垣、城楼多崩塌, 砖城墙崩数十丈, 石碑坊尽行倒塌, 塔崩颓, 桥崩圮	山多裂崩, 阻绝道路或积水成湖, 地坼裂普遍, 涌水冒沙, 井水喷溢或井泉干涸, 有地陷积水或陆陷成海者
XI	伤亡甚众(数万或过半); 赈济、免税赋	民居、公廨、官署、学府、监仓、庙宇、殿堂等倒塌殆尽或少有存者	城垣、城楼尽倾塌, 存者无多, 砖城墙多崩塌, 石碑扭转或翻倒, 塔倾颓, 桥摧崩	山劈阜移, 地貌改观, 地裂地陷普遍, 涌泉喷高丈许, 夹沙石土等
XII	压杀惨重(死亡者过半); 赈济、免税赋	民居、公廨、官署、学府、监仓、庙宇、殿堂等俱倒如平地或仅有孑存	城垣、城楼倾塌殆尽, 砖城墙尽崩毁, 碑、桥尽毁	山川改易, 地形地貌剧烈变化, 地裂地陷规模宏大, 宽不可越, 深不可视, 地表水和地下水发生剧烈变化

但是由于历史地震记录本身的模糊性, 使得历史烈度表的烈度划分也是以定性描述为主, 烈度表主要以描述性词语为主, 这种模糊性使得基于烈度表的人工评定不可避免会以经验为主, 因而经验性、主观性成为人工评定的主要问题。而单纯利用烈度表进行机器判定, 又存在着生搬硬套、缺乏监督的情况, 而且由于记叙文明的进步, 历史灾害记录也随之剧增, 先秦时多记录大灾, 且叙述简略, 而明清时州志、县志普及, 灾害统计来源大为丰富, 且叙述较详, 但若因此认为先秦灾害少而轻, 明显是不客观的, 因此对烈度表的应用必须参入一定的动态修正, 这也是机器判定烈度需要解决的问题。

针对上述问题, 利用模糊理论[172-174]进行历史地震烈度判定是个很好的解决思路。模糊理论的特点是, 所得出的结论并不表征对象绝对归属于哪一类, 而

是采用相对理论, 通过隶属函数进行程度分析, 具有定量、客观、不受评估要素限制等优点。因此本章以最新的《2010 历史地震烈度表》作为地震烈度特征量化的基本参照, 结合历史地震记录的特点利用高斯隶属度函数和灰色关联理论来更好地生成模糊判定模型, 同时混合使用了指数拟合和熵权法对量化数据进行修正, 以使其能更好地还原出历史地震情况, 最后通过实验对比证明了该方法的有效性。

### 3.2.2 混合模糊历史地震烈度判定模型

设历史地震记录有  $s$  条, 每个记录有  $m$  个烈度特征, 则其初始特征矩阵可设为  $R$ , 按下列步骤构建历史地震烈度模糊判定模型。

$$R = \left[ x_{ij} \right]_{s \times m} \quad (i \in [1, s], j \in [1, m]) \quad (3.6)$$

#### (1) 初始量化。

根据最新的历史地震烈度表(2010表), 将地震烈度特征分为人与社会反响、房屋建筑、构筑物 and 地表现象四类(即  $m=4$ ), 并按照 2010 表的特征区间对历史地震记录进行量化, 以  $0 \sim n$  编码来对应烈度表中不同严重程度的特征描述, 其中 0 表示该特征记录缺失,  $1 \sim n$  分别表示该特征各个严重程度。在编码时利用命名实体识别技术对相似的特征描述进行了整合归类, 对于一条历史地震记录中出现多个特征描述的严重程度产生交叉时, 取较大值进行编码。

#### (2) 动态修正。

由于历史文明的发展以及纸张的逐渐普及, 较远时代的地震灾害记录非常简洁, 较近时代的记录则相对详细, 呈现出距今时代越近, 地震记录数量和内容也越丰富的特点(如表 3.3 所示), 所以对各个历史时期的地震灾害记录也应该区别对待, 若只根据 2010 表对这些历史地震记录进行统一处理, 那么较远时代的因为时代特点而描述简单的地震灾害记录就很可能和较近时代的因为灾情较轻而描述简单的记录相混淆, 从而脱离了实际历史情况。因此, 还必须考虑到历史地震记录中的社会因素, 对数据加以进一步修正。

表 3.3 中国各历史时期地震数量分布

时期	秦及秦以前	汉—三国	魏晋—南北朝	隋唐五代	宋	元	明	清
频次	47	132	220	356	509	759	2663	3838

图 3.2 为用 SPSS 软件对历史地震频次做的多项式拟合, 但是对于其中的波动我们认为是因为自然性的灾害频发或少发而产生的噪音, 不能反映出由社会发展原因而引起的记录剧增态势, 而且根据图 3.2 可以看出其上升趋势很符合指数特

点，这也与历史专家的结论相符，因此我们利用指数函数来拟合历史地震记录的时代特征（如图 3.3 所示），并以得出的指数函数来模拟历史记录的时代动态特性，对各个记录的历史地震烈度特征值进行修正，方法如下：

设通过 SPSS 软件拟合得出的指数函数为  $F(x)$ ，如公式 (3.7) 所示，其中  $x$  表示历史年份， $a$ 、 $b$  为参数。

$$F(x) = a * \exp(bx) \quad (3.7)$$

接着以各个朝代划分为区间，引入灰色关联分析法，设距今最近的清代的关联度为 1（即以清代为加权基数），按公式 (3.8) 得出各个朝代的权系数  $w_{epoch}$ ，其中  $x_{begin}$  和  $x_{end}$  分别表示清代的起始年份和结束年份， $x_{begini}$  和  $x_{endi}$  分别表示  $w_{epoch}$  所在朝代的起始年份和结束年份。

$$w_{epoch} = 2 - \frac{1}{2 - \frac{\ln F(x_{endi}) + \ln F(x_{begini})}{\ln F(x_{end}) + \ln F(x_{begin})}} \quad (3.8)$$

将各个历史地震记录特征值乘以各自所在朝代的权系数，就可得出修正后的量化结果。由于各个烈度特征所划分的  $n$  值并不相同，还需要进行归一化处理，如公式 (3.9) 所示，第  $i$  个地震数据的第  $j$  个特征的修正值设为  $x_{ij}$ ，经归一化处理后变为  $y_{ij}$ ，这里  $x_{maxj}$ 、 $x_{minj}$  分别为所有地震数据中第  $i$  个特征值的最大值和最小值。

$$y_{ij} = (x_{ij} - x_{minj}) / x_{maxj} - x_{minj} \quad (3.9)$$

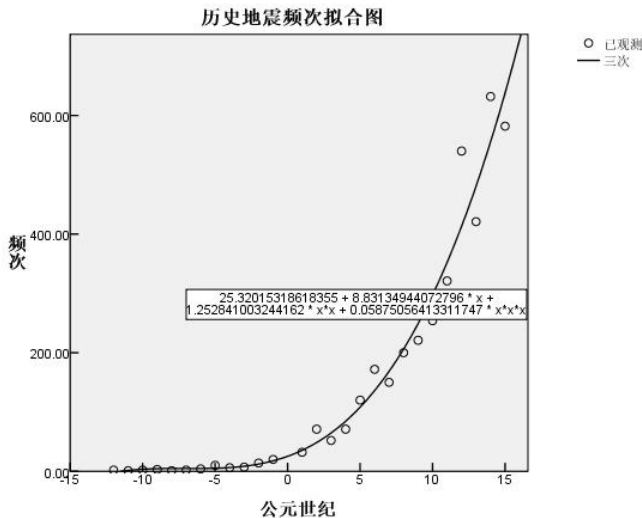


图 3.2 三次多项式拟合图

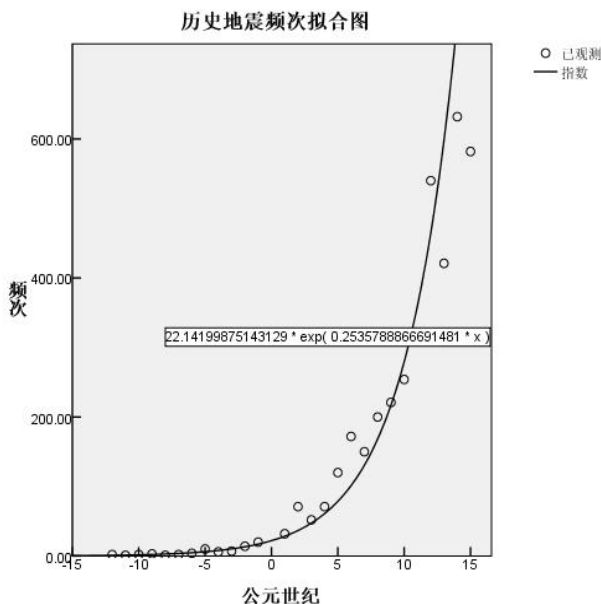


图 3.3 指数拟合图

### (3) 烈度特征赋权。

在进行模糊判定之前，还需要确定各个烈度特征的权重，这里我们使用熵权法[175-179]来求出各个特征权重。熵值在信息论中用来反映信息的无序化程度，信息量的大小可以由其来度量。指标携带的信息与其决策力成正比，信息多的指标对决策具有较大作用。系统无序度越大，熵值就越大，反之亦然。因此信息是否有序可以在很大程度上通过信息熵直接反映出来，这就可以不通过各种可能的人为因素而进行客观评估，使得评定结果更为科学可信。设经过上述 1、2 步处理后，得到归一化矩阵  $B$ ，各个地震烈度特征的熵权值的计算步骤如下：

根据公式 (3.10) 求得各烈度特征的熵：

$$H_j = - \left( \sum_{i=1}^s f_{ij} \ln f_{ij} \right) / \ln s \quad (3.10)$$

其中  $f_{ij}$  的求法如公式 (3.11) 所示，其中  $x_{ij}$  表示矩阵  $B$  中第  $i$  个数据的第  $j$  个特征值：

$$f_{ij} = (1 + x_{ij}) / \sum_{i=1}^s (1 + x_{ij}) \quad (3.11)$$

进而，第  $j$  个特征值的熵权值如公式 (3.12) 所示：

$$w_j = \frac{1 - H_j}{n - \sum_{j=1}^m H_j} \quad (3.12)$$

(4) 建立隶属函数。

由于历史烈度区间较多，且模糊性很强，常用的线性隶属函数分段生硬，去噪能力弱，难以满足需要，因此这里采用更为平滑的高斯隶属函数[180]来建立模糊判定模型，并参考历史专家的意见，建立出III-XII度的高斯隶属函数，如公式(3.13)至公式(3.15)所示，其中公式(3.14)中 $n \in [4, 11]$ 。

$$\mu_{12} = \begin{cases} 1 & x \geq 1 \\ \exp\left(-\frac{(x-0.95)^2}{0.02}\right) & 0.8 \leq x < 1 \\ 0 & x < 0.8 \end{cases} \quad (3.13)$$

$$\mu_n = \begin{cases} 0 & x \geq (n-2) \times 0.1 \\ \exp\left(-\frac{(x-(n-2) \times 0.1)^2}{0.04}\right) & (n-4) \times 0.1 \leq x < (n-2) \times 0.1 \\ 0 & x < (n-4) \times 0.1 \end{cases} \quad (3.14)$$

$$\mu_3 = \begin{cases} 0 & x \geq 0.15 \\ \exp\left(-\frac{(x-0.1)^2}{0.02}\right) & 0.05 \leq x < 0.15 \\ 0 & x < 0.05 \end{cases} \quad (3.15)$$

将隶属函数导入矩阵 $B$ 就可得出对应的隶属度矩阵 $L$ ，但由于历史地震记录特征缺失比较多，使得矩阵 $L$ 中存在大量0值，对模糊判定的性能有较大影响。在灰色系统[181]预测与决策中利用关联度来作为事物之间、因素之间关联性的量度，各烈度特征隶属值大小与某一标准隶属值的关系疏密程度也可以反映在关联度大小上，因此进一步使用灰色关联分析法将隶属度矩阵 $L$ 转化为关联系数矩阵 $G$ ，以去除原矩阵中的0值，如公式(3.16)所示，其中 $x_{ni}$ 表示原隶属度， $r_{ni}$ 表示转化后的关联系数， $x_{maxi}$ 、 $x_{mini}$ 分别表示第 $i$ 项隶属度区间上的最大值和最小值。

$$r_{ni} = \frac{1}{2 - (x_{ni} - x_{mini}) / (x_{maxi} - x_{mini})} \quad (3.16)$$

(5) 模糊判定。

设以 $t_{jv}$ 来表征数据第 $j$ 个特征与第 $v$ 个烈度( $v \in [III, XII]$ )的关联性，它决定

于特征权重  $w_j$  和关联系数  $r_{jv}$  的乘积大小, 如公式 (3.17) 所示。

$$t_{jv} = w_j \times r_{jv} \quad (3.17)$$

各个烈度的综合关联性(设为  $\Delta_v$ )则由对属于各烈度的特征关联性求和取得, 并选择取得最大综合关联性值  $\Delta$  的烈度作为历史地震烈度, 如公式 (3.18) 所示。

$$\Delta = \max \{ \Delta_v \} = \max \left\{ \sum_{j=1}^m t_{jv} \right\} \quad (3.18)$$

### 3.2.3 实例分析

为了验证本章提出方法的有效性, 我们选取了 10 条比较有代表性的历史地震记录进行分析, 并请历史专家分别根据 83 表、89 表和 2010 表对这 10 条记录进行了烈度评估, 由于专家评估过程中已充分参考了地震的历史背景因素, 所以准确性较高, 可以作为算法性能参照标准, 同时我们还选取了未使用本章混合方法的单一模糊模型作为对比算法, 同样以 2010 表为判定基础, 以验证我们的混合模型是否更有效。

对比结果如表 3.4 所示, 可以看到混合模糊模型的判定结果与 2010 表符合得很好, 3 个历史烈度表之间也有着较高的一致性, 而单一模糊模型则对较近时代的地震有不错的判定准确率, 而对较远时代的地震则出入较大, 如对非常著名的陕西岐山地震、甘肃表氏地震的烈度判定都明显过低, 反映出自适应性不足的缺陷, 可见混合模糊模型加入对历史地震记录时代特征的考量是很有必要的。

表 3.4 烈度对比

序号	地震时间	影响地点	(83) 简表	(89) 方案	(2010) 表	混合模型	单一模型
1	前 780 年	陕西岐山	IX~X	IX	IX	IX	VI
2	前 47 年	甘肃陇西	IX~X	IX	IX	IX	VIII
3	180 年	甘肃表氏	>X	X	XI	X	VIII
4	1067 年	广东潮州	IX~X	IX	IX	IX	VI
5	1474 年	云南鹤庆	VII	VII	VII	VII	VI
6	1556 年	陕西华县	>X	XI	XII	XII	XII
7	1605 年	海南琼山	$\geq$ X	X	X	X	IX
8	1654 年	甘肃天水	>X	XI	XI	XI	XI
9	1668 年	山东郯城	>X	XI	XII	XII	XII
10	1679 年	北京	VIII	VIII	VIII	VIII	VIII

### 3.3 本章小结

本章所提出的聚类分级算法有效避免了对初始聚类中心的敏感性，并利用蚂蚁算法使聚类结果不会陷入局部最优，整个聚类过程中受人为因素干扰少，使得聚类结果科学可信；所提出的混合模糊模型针对历史地震记录的长时间序列性，利用指数拟合法实现对烈度特征量化的动态调整，使量化结果更符合历史真实情况，并通过熵权法和灰色关联分析法对模糊模型进行了优化，使之更适应于历史地震记录的特点，最后通过实例比较证明了该方法的准确性。