模糊聚类算法分析及程序实现

山东工业职业学院 彭丽英 董佳佳

摘要：本文详细介绍了模糊聚类算法的产生过程以及应用模糊聚类算法进行模糊分析的 ASP 代码，最后通过一个具体的实例，对模糊聚类算法进行了验证。

关键词：普通聚类　模糊聚类　最大最小法　代码

由于客观事物之间的界限往往是不清晰的，用模糊数学的聚类分析处理具有模糊性事物的聚类问题是十分合适的。所以，近几年，模糊聚类分析在模式识别、数据挖掘、计算机视觉以及模糊控制等领域的应用越来越广泛，所以它成为研究的热点。但在一般的资料中对聚类算法的介绍都比较粗略，更缺少代码，本文除了对算法进行了详细介绍外，还提供了相关代码。

在模糊聚类分析中，首先要计算模糊相似矩阵，即建立样本间的模糊关系，而模糊聚类是在普通聚类的基础上产生的。

1、普通聚类[1]

设被分类对象的集合为 X＝$\{x_1,x_2,\cdots,x_n\}$，其中每一个元素 x_i，i =1,2,...n，都有 m 个特性指标，即 x＝$\{x_{i1},x_{i2},\cdots,x_{im}\}$

如果要把 X 分成 C 类，则它的每一个分类结果都对应一个 c×n 阶 0-1 矩阵 R=$\{r_{ij}\}$,其中 $r_{ij}=\begin{cases}1 & x_i \in j\\ 0 & x_i \notin j\end{cases}$

矩阵 R 具有如下性质：

(1)R=$\{0,1\}$，即 R 是一个布尔矩阵；

(2) $\sum_{i=1}^{c} r_{ij}$ =1，即每一列有且仅有一个元素为 1，这一性质保证了每一个样本只能划归到其中的一类中去；

(3) $\sum_{j=1}^{n} r_{ij}$ ＞0，即每一行的元素之和大于 0，这保证了每一类不空，且一类中可以有多个样本。

2、模糊聚类

在普通聚类的基础上，人们提出了模糊聚类(软聚类)的概念，认为被分类对象集合 X 中的样本 x_i ，i=1,2, ⋯,n 以一定的隶属度隶属于某一类，也就是说，所有的样本都分别以不同的隶属度隶属于某一类。因此，每一类就认为是样本集合 X 上的一个模糊子集，于是每一种这样的分类结果所对应的分类矩阵，就是一个模糊矩阵 R。

该分类矩阵满足下列三个条件：[1]

(1) $x_{ij}\in[0,1]$ ，即分类矩阵元素在 0 和 1 之间取值；

(2) $\sum_{i=1}^{c} r_{ij}$ =1，即每一列中分别属于各类的隶属度之和为 1，对一个样本而言，它对各类的隶属度之和为 1；

(3) $\sum_{j=1}^{n} r_{ij} > 0$，即每一行的元素之和大于 0，这保证了每一类都必须有样本，即总有一些样本不同程度的隶属于各类。

3、数据规格化

聚类分析输入的是一组未分类记录，并且这些记录应分成几类事先也不知道，通过分析数据库中的记录数据，根据一定的分类规则，合理地划分记录集合，确定每个记录所在类别。它所采用的分类规则是由聚类分析工具决定的。采用不同的聚类方法，对于相同的记录集合可能有不同的划分结果。由于在一般情况下，所分不同类事物之间的界限不是分明的，用传统的分类方法具有一定的局限性，而用模糊聚类方法确定分类样品的亲疏关系，从而得到更为合理的分类。由于特性指标的量纲和数量级都不相同，致使对各特性指标的分类缺乏一个统一尺度，为消除影响，需要对个各指标值实行数据规格化，从而使每一指标值统一于某种共同的数值特性范围。[2]

应用相似系数法、距离法、最大最小法标定被分类对象的模糊相似关系矩阵。相似关系矩阵反映了样本与样本之间的相似关系。假设应用最大最小法：[3]

$$r_{ij}=\frac{\sum_{k=1}^{m}(x_{ik} \wedge x_{jk})}{\sum_{k=1}^{m}(x_{ik} \vee x_{jk})}$$

$\vee$为取大运算，如：$(0.9 \vee 0.7)$=0.9；

$\wedge$为取小运算，如：$(0.8 \wedge 1.0)$=0.8。

建立模糊相似矩阵的 ASP 代码：

```
<%
Function mMatrix(a,b)
        dim multiplyMatrix(),s,s1
        redim multiplyMatrix(UBOUND(a),UBOUND(a,2))
        for i = 0 to UBOUND(a)
            for j = 0 to UBOUND(a,2)
               s=0
                s1=0
                if(i=j) then multiplyMatrix(i,j)=1
                for k = 1 to UBOUND(a,2)
                    s1=s1+min(a(i,k),a(j,k))
                    s=s+max(a(i,k),a(j,k))
                next
                multiplyMatrix(i,j)=s1/s
            next
        next
                mMatrix = multiplyMatrix
```

```
End Function
Function min(a,b)
  dim t
  if(a>=b) then t=b else t=a
  min=t
End Function
Function max(a,b)
  dim t
  if(a>=b) then t=a else t=b
    max=t
End Function
Sub DisplayMatrix(multiplyMatrix,title)
                    response.write "Matrix("&title&"):<BR>"
          for i=0 to UBOUND(multiplyMatrix)
               for j=0 to UBOUND(multiplyMatrix,2)
                    response.write multiplyMatrix(i,j) & " "
               next
               response.write " <BR>"
          next
End Sub
dim a()
redim a(10,10)
……  '从数据库中读取数据到数组
call DisplayMatrix(a,"A")
dim c
c=mMatrix(a)
call DisplayMatrix(c,"C")
%>
```

4、求相似关系矩阵 R 的传递闭包 R*

一般采用平方自成法。取 R 的乘幂：R^2，R^4，R^8， …，若在某一步有 $R^k=R^{2k}=R^*$，…，则 R*便是一个模糊等价关系，已具有传递性。这里 $R^2=R*R$,（其它乘幂类似）遵守模糊矩阵复合运算中的先取小后取大规则。

模糊方阵幂的定义：若 R 为 n 阶方阵，定义 $R^2 = R\circ R$，$R^3 = R^2\circ R$，…，$R^k = R^{k-1}\circ R$。

其定义的根据是模糊矩阵的合成运算：设 $A = (a_{ik})_{m\times s}$，$B = (b_{kj})_{s\times n}$，定义模糊矩阵 A 与 B 的合成为：$A\circ B = (c_{ij})_{m\times n}$，其中 $c_{ij} = \vee\{(a_{ik}\wedge b_{kj}) \mid 1\leqslant k\leqslant s\}$。

对于模糊方阵来说是同一个矩阵，即 A 和 B 是同一个矩阵，那么式 $cij = \vee\{(aik\wedge bkj) \mid 1\leqslant k\leqslant s\}$可以写为：

$rij = \vee\{(a_{ik}\wedge a_{kj}) \mid 1\leqslant k\leqslant s\}$

其中 i 和 j 的变化为从 0～n-1，k 的变化从 1～n-1。

其主要 JavaScript 代码：

```
<script language="JavaScript">
var aEg = new Array(
    [1,0.97,0.97,0.98,0.97,0.98,0.97,0.97,0.97,0.98],
[0.97,1,0.97,0.96,0.97,0.96,0.97,0.97,0.96,0.96],
[0.97,0.97,1,0.96,0.97,0.96,0.97,0.97,0.96,0.96],
[0.98,0.96,0.96,1,0.96,0.99,0.96,0.96,0.97,0.99],
[0.97,0.97,0.97,0.96,1,0.96,0.98,0.97,0.96,0.96],
[0.98,0.96,0.96,0.99,0.96,1,0.96,0.96,0.97,0.99],
[0.97,0.97,0.97,0.96,0.98,0.96,1,0.97,0.96,0.96],
[0.97,0.97,0.97,0.96,0.97,0.96,0.97,1,0.96,0.96],
[0.97,0.96,0.96,0.97,0.96,0.97,0.96,0.96,1,0.97],
[0.98,0.96,0.96,0.99,0.96,0.99,0.96,0.96,0.97,1]);
var i, j, k, m, b;
var c=new Array(
[0,0,0,0,0,0,0,0,0,0],[0,0,0,0,0,0,0,0,0,0],
[0,0,0,0,0,0,0,0,0,0],[0,0,0,0,0,0,0,0,0,0],
[0,0,0,0,0,0,0,0,0,0],[0,0,0,0,0,0,0,0,0,0],
[0,0,0,0,0,0,0,0,0,0],[0,0,0,0,0,0,0,0,0,0],
[0,0,0,0,0,0,0,0,0,0],[0,0,0,0,0,0,0,0,0,0]);
for (i=0; i < aEg.length; i++)
{ for (j=0; j < aEg[i].length; j++)
   { if(i==j)
      {c[i][j]=1;}
        else
          {   m=0.0;
              for (k=1; k < aEg.length; k++)
                 {b=Math.min(aEg[i][k],aEg[k][j]);
                  m=Math.max(m,b);
                 }
              c[i][j]=m;
           } } }
for (i=0; i < aEg.length; i++)
{       for (j=0; j < aEg[i].length; j++)
     {     document.write(aEg[i][j]);
```

```
            document.write('    ');
        }
        document.write('<BR>');
    }
    document.write(aEg.length);
    document.write('<BR>');
    for (i=0; i < c.length; i++)
    {       for (j=0; j < c[i].length; j++)
        {       document.write(c[i][j]);
                document.write(',');
        }
        document.write('<BR>');
    }
    </script>
```

5、适当选取λ截割传递闭包，对被分类对象进行动态聚类。λ是 R 中的隶属度，选择不同的隶属度使样本分为不同的 C 类。

聚类就在已建立的模糊等价关系矩阵上，给定不同的λ水平进行截取，从而得到不同的分类。λ越小，分的类就越少、越粗；λ越大，分的类就越多、越细。当取最优的λ值时，得到最合理的分类体系。

6、下面以 10 名教师某学期各种评价指标的得分组建数字模型

教师编号	学生评价	督导组	系部评价	同行评价	减分	加分	总分
T1	90	80	80	88	90	91	519
T2	70	65	68	69	80	85	437
T3	85	90	78	90	95	90	528
T4	65	70	66	70	80	82	433
T5	80	90	80	82	87	88	507
T6	96	90	90	89	90	88	543
T7	80	70	65	70	85	90	460
T8	75	70	68	72	81	82	448

应用最大最小法求 R

$$\begin{bmatrix} 1 & 0.805 & 0.946 & 0.784 & 0.925 & 0.926 & 0.843 & 0.843 \\ 0.805 & 1. & 0.793 & 0.932 & 0.819 & 0.745 & 0.934 & 0.954 \\ 0.946 & 0.793 & 1 & 0.773 & 0.957 & 0.934 & 0.831 & 0.831 \\ 0.784 & 0.932 & 0.773 & 1 & 0.798 & 0.726 & 0.93 & 0.93 \\ 0.925 & 0.819 & 0.957 & 0.798 & 1 & 0.91 & 0.858 & 0.858 \\ 0.926 & 0.745 & 0.934 & 0.726 & 0.91 & 1 & 0.781 & 0.781 \\ 0.843 & 0.934 & 0.831 & 0.93 & 0.858 & 0.781 & 1 & 0.966 \\ 0.843 & 0.954 & 0.831 & 0.93 & 0.858 & 0.781 & 0.966 & 1 \end{bmatrix}$$

求相似关系矩阵 R 的传递闭包，根据模糊矩阵的合成运算式 $r_{ij} = \vee\{(a_{ik}\wedge a_{kj}) \mid 1\leqslant k$

≤s}，可以计算出 $R^{*}=R^{4}=R^{8}$。

下面的矩阵为 R^{4}：

$$\begin{bmatrix} 1 & 0.858 & 0.946 & 0.858 & 0.946 & 0.934 & 0.858 & 0.858 \\ 0.805 & 1. & 0.858 & 0.932 & 0.858 & 0.858 & 0.954 & 0.954 \\ 0.946 & 0.858 & 1 & 0.858 & 0.957 & 0.934 & 0.858 & 0.858 \\ 0.858 & 0.932 & 0.858 & 1 & 0.858 & 0.858 & 0.932 & 0.932 \\ 0.946 & 0.858 & 0.957 & 0.858 & 1 & 0.934 & 0.858 & 0.858 \\ 0.934 & 0.858 & 0.934 & 0.858 & 0.934 & 1 & 0.858 & 0.858 \\ 0.858 & 0.954 & 0.858 & 0.932 & 0.858 & 0.858 & 1 & 0.966 \\ 0.858 & 0.954 & 0.858 & 0.932 & 0.858 & 0.858 & 0.966 & 1 \end{bmatrix}$$

适当选取λ截割传递闭包，根据λ截集的意义，在 R 中 $r_{ij}\geq\lambda$，便换为 l，否则换为 0。取$\lambda=0.97$，有 R_{λ} 为：

$$\begin{bmatrix} 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix}$$

可以得出分类：{T6}{T1,T3,T5}{T2,T7,T8}{T4}

通过聚类分析可以看出，成绩相近的分在同一类别中，并且根据教师考核要求，评为“优秀”的 1 人，“良好”的 3 人，“合格”的 3 人，“不合格”的 1 人，基本上符合常情。实际上算法还可以再进行优化，可以再进行二次模糊聚类分析得到更加合理的分类。

参考文献：

1. 李士勇. 工程模糊数学及应用　哈尔滨工业大学出版社，2004
2. 冯　梅. 基于模糊聚类分析的教师课堂教学质量评价　数学的实践与认识，2008
3. 许 哲. 基于模糊聚类分析的用户身份认证方法研究　延边大学硕士论文，2005
4. 周海洋. 模糊聚类分析在教师综合考评中的应用　中国科技信息，2007(22)