

基于感应线圈数据的城市道路 交通状态判别方法

姜桂艳, 郭海锋, 吴超腾

(吉林大学 交通学院, 长春 130022)

摘要:以 SCOOT 系统感应线圈检测器采集到的交通数据为基础,设计了一种基于模糊聚类的城市道路交通状态实时判别算法及其评价方法,并提出了交通状态判别时间间隔的确定方法。以 VISSIM 为工具,对上述方法进行了模拟。对比分析结果表明,所提出的算法能够提高城市道路交通状态实时判别的效果。

关键词:交通运输系统工程;城市道路;交通状态;感应线圈;模糊聚类

中图分类号:U491 **文献标识码:**A **文章编号:**1671-5497(2008)Sup.-0037-06

Identification method of urban road traffic conditions based on inductive coil data

Jiang Gui-yan, Guo Hai-feng, Wu Chao-teng

(College of Transportation, Jilin University, Changchun 130022, China)

Abstract: A real-time method identifying the urban road traffic conditions and its evaluation approach were proposed based on the fuzzy clustering algorithm and the data from the inductive coil of the SCOOT traffic signal control system. A method to determine the identification time interval was also presented. A simulation was performed by software VISSIM to confirm the proposed methodology. A comparative analysis show positive and encouraging results.

Key words: engineering of communication and transportation system; urban road; traffic conditions; inductive coil; fuzzy cluster

目前,国内外研究者提出的交通状态判别方法主要应用于高速公路的事件检测,而用于城市道路交通状态判别的方法相对较少^[1,2]。在城市道路中由于信号交叉口的影响,其交通流特性与高速公路相比较为复杂,对城市道路交通状态进行判别也增加了一定的难度。

作者以 SCOOT 信号控制系统感应线圈检测器采集到的交通数据为基础,对文献[3]提出的特

征指标进行了改进,在此基础上设计了一种基于模糊聚类的城市道路交通状态实时判别算法及其评价方法,分析了时间间隔长度对交通状态判别效果的影响,并以 VISSIM 为工具通过编程对上述方法进行了模拟试验,结果表明所建立的试验环境能够再现实际交通运行中对感应线圈检测器数据进行实时采集以及对交通状态进行实时判别的全过程。

收稿日期:2007-05-23.

基金项目:“863”国家高技术研究发展计划项目(2006AA11Z228);国家自然科学基金重点项目(50338030).

作者简介:姜桂艳(1964-),女,教授,博士生导师.研究方向:交通信息工程. E-mail:jljiangguiyan@126.com

1 交通状态特征指标的选取

本文所采用的检测器为感应线圈检测器,且其布设位置与 SCOOT 系统检测器布设位置相同,在路段的上游,能够提供流量、速度及占有率等交通数据。感应线圈检测器的交通数据分布图如图 1 所示。为了能使其在同一窗口中显示,对原数据进行了标准化处理。

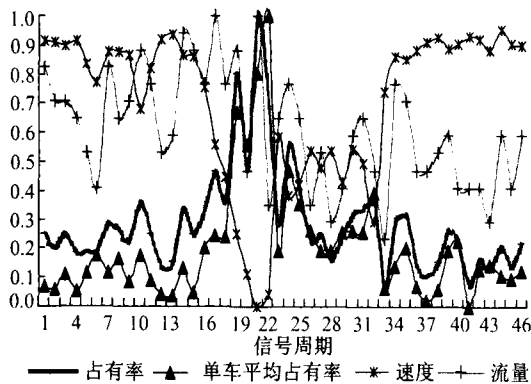


图 1 交通数据状态分布图
Fig.1 Profile of traffic data

从图 1 可以看出,在拥挤发生、持续及消散的过程中,感应线圈检测器采集到的占有率和速度对拥挤的时间敏感性和程度敏感性反应较强;而流量并没有太明显的波动。但以占有率除以流量计算的单车平均占有率则变化较为明显。因此,作者在选定交通状态特征指标时,保留了文献[3]中占有率和速度指标,去除了流量指标,而将其与占有率组合换算为单车平均占有率。以这 3 个指标作为交通状态的特征向量,即 $x_i = (y_1, y_2, y_3)^T$ 。特征向量中的各指标分别为

$$y_1 = occu(i) \tag{1}$$

$$y_2 = \frac{occu(i)}{q(i)} \tag{2}$$

$$y_3 = speed(i) \tag{3}$$

式中: $occu(i)$ 为第 i 时间间隔的占有率; $q(i)$ 和 $speed(i)$ 分别为第 i 时间间隔的交通流量和交通流平均速度。

2 基于模糊聚类的交通状态判别算法

2.1 模糊 C 均值聚类的步骤

模糊 C 均值聚类是一种基于划分的聚类算法,令 $X = \{x_i, i = 1, 2, \dots, n\}$ 是 n 个样本组成的样本集合, c 为预定的类别数目, $m_j, j = 1, 2, \dots, c$ 为每个聚类的中心, $\mu_j(x_i)$ 是第 i 个样本对于第 j

类的隶属度函数。用隶属度函数定义的聚类损失函数见式(4)。

$$J = \sum_{j=1}^c \sum_{i=1}^n [\mu_j(x_i)]^b \|x_i - m_j\|^2 \tag{4}$$

式中: $b > 1$ 是一个可以控制聚类结果的模糊指数。

在不同的隶属度定义方法下最小化式(4)的损失函数,就得到不同的模糊聚类方法。其中最具有代表性的是模糊 C 均值方法,它要求一个样本对于各个类聚类的隶属度之和为 1,即

$$\sum_{j=1}^c \mu_j(x_i) = 1, i = 1, 2, \dots, n \tag{5}$$

在条件式(5)下求式(4)的极小值,令 J 对 m_j 和 $\mu_j(x_i)$ 的偏导数为 0,可得必要条件

$$m_j = \frac{\sum_{i=1}^n [\mu_j(x_i)]^b x_i}{\sum_{i=1}^n [\mu_j(x_i)]^b}, j = 1, 2, \dots, c \tag{6}$$

$$\mu_j(x_i) = \frac{(1/\|x_i - m_j\|^2)^{\frac{1}{b-1}}}{\sum_{k=1}^c (1/\|x_i - m_k\|^2)^{\frac{1}{b-1}}} \tag{7}$$

用迭代方法求解式(6)和式(7),当算法收敛时,就得到了各类的聚类中心和各个样本对于各类的隶属度值,从而完成了模糊聚类的划分。

模糊 C 均值聚类算法中需要确定的两个主要参数是聚类数目 c 和模糊指数 b 。本文将交通状态分为 3 类,即顺畅、阻滞和拥堵,因此 $c = 3$ 。下面对 b 值的确定进行分析。

模糊指数 b 的取值对聚类结果有较大影响,选择合理的 b 值非常重要。文献[4]通过实验数据分析,给出 b 值的最佳选择区间为 $[1.5, 2.5]$ 。文献[5]通过理论证明指出,模糊指数 b 的选择依赖于数据样本 X 的一个统计量 $\lambda_{\max}[C_X]$

$$C_X = \sum_{i=1}^N \frac{[x_i - \bar{x}][x_i - \bar{x}]^T}{n \|x_i - \bar{x}\|^2}, \|x_i - \bar{x}\| \neq 0 \tag{8}$$

其中, λ_i 为矩阵 C_X 的特征根,当 $\lambda_{\max}[C_X] < 0.5$ 时,则应该有 $1 \leq b \leq \frac{1}{1 - 2\lambda_{\max}[C_X]}$,否则模糊 C 均值聚类算法的聚类结果不够理想;当 $\lambda_{\max}[C_X] \geq 0.5$ 时,则应该有 $b > 1$ 。综上,本文确定 b 值的基本步骤为:

(1) 根据感应线圈检测器采集到的历史数据样本 X 计算 $\lambda_{\max}[C_X]$;

(2) 如果 $\lambda_{\max}[C_X] < 0.5$, 则 $1 \leq b \leq$

$$\frac{1}{1-2\lambda_{\max}[C_X]};$$

(3)如果 $\lambda_{\max}[C_X] \geq 0.5$, 则 $1.5 \leq b \leq 2.5$;

(4)采用启发式方法,以步长为 0.1 对步骤(2)或(3)中 b 所在的区间进行渐进寻优,计算与不同 b 值对应的各类类心间距离,寻找使分类的各类类心间距离最大的 b 值。

2.2 交通状态判别时间间隔的选择

对交通状态判别算法进行设计时,一个很重要的问题就是判别时间间隔的确定。尤其是在城市道路中,交通状态判别的时间间隔不能像高速公路一样根据需要任意设定,因为城市道路的交通状态受信号交叉口信号周期的影响较大。对于城市路网而言,以各交叉口信号周期为交通状态判别的时间间隔是不现实的。首先,对城市路网中各路段的交通状态进行实时判别时,判别的时间间隔要一致,否则无法以统一的频率对外发布交通状态信息。其次,路网中各信号交叉口信号配时方式不尽相同,信号周期也不一样,即使是同一信号交叉口,其配时方案在不同时段也可能有所不同,尤其是自适应信号控制方式,其信号周期更是不断变化。因此,无法以信号周期作为判别的时间间隔。

如果交通状态判别的时间间隔小于信号周期,则感应线圈检测器采集到的交通数据具有较强的随机性,难以准确体现交通状态变化的规律性。因此,对城市道路交通状态进行判别时,判别的时间间隔应大于全路网或局部区域内信号交叉口最大的信号周期时长,一般选择为 5~15 min。

2.3 交通状态判别算法设计

对感应线圈检测器采集到的包含各种交通状态的历史交通数据,通过模糊聚类将其分为代表不同拥挤程度的 3 类,并得到各类的中心点。然后根据各类的类心及公式(7)建立判别函数,以隶属度最大为原则对城市道路交通状态进行实时判别。该算法的具体步骤如下:

(1)按照指定的时间间隔 t 获取感应线圈的历史数据,采样数据中要包括各种交通状态;

(2)计算交通状态特征向量 $y_1(t)$ 、 $y_2(t)$ 和 $y_3(t)$,并进行标准化处理,所有数据均在 0~1 之间;

(3)进行模糊聚类,将所有历史数据样本分为 3 类,得到聚类中心和隶属度函数;

(4)按照时间间隔 t 获取感应线圈的当前数

据,并进行标准化变换;

(5)计算该组数据距离各类中心的隶属度数值;

(6)按照隶属度最大原则判定当前的交通状态,并提供给相关用户;

(7)令 $t=t+1$,转步骤(4)。

2.4 交通状态判别算法的评价

上述算法能够将城市道路的交通状态分为顺畅、阻滞和拥堵 3 个拥挤等级,其本质是在高维空间上对交通状态特征指标进行相对分类,但难以直接获得评价算法效果的交通状态真值,这一点与交通事件检测算法的评价不同。后者可以通过现场观测直接获得有无事件发生的真实结论,从而完成对其有效性的评价。为了分析和比较交通状态判别算法的效果,作者提出了一种误判率交叉估计评价方法。

误判率交叉估计方法的基本原理是在运用感应线圈的历史数据集完成初始聚类分析后,记录每个样品的类别归属。然后,剔除聚类分析数据集中的—个样品,利用其余 $n-1$ 个样品构成的数据集再次进行聚类分析,求得分类中心和隶属度函数。之后运用新的判别准则对所剔除的样品做判别,记录其所属类别。对原始聚类分析数据集中的每个样品作上述分析,以原始分类结果为标准,通过交叉计算误判比例得到误判概率的估计值,误判率较小的交通状态判别算法性能较好。

经过初始聚类分析,本文将感应线圈的历史数据集 X 分为 3 类: C_1 、 C_2 和 C_3 。假设 C_1 中有 n_1 个样品, C_2 中有 n_2 个样品, C_3 中有 n_3 个样品,则误判率交叉估计评价方法的具体步骤如下:

(1)从 C_1 中剔除其中一个样品,用剩余的 n_1-1 个样品和 C_2 中 n_2 个样品及 C_3 中 n_3 个样品构成样本集 X_1 ,并以此样本集进行聚类,得到聚类中心点与隶属度函数;

(2)用(1)中的新聚类中心点及隶属度函数对剔除的那个样品进行判别,记录其所属类别;

(3)重复(1),(2),直到 C_1 中的 n_1 个样品依次被剔除,与原始分类结果相比较,将 n_1 中误判的个数记为 n_1^* ;

(4)对 C_2 、 C_3 中的样品重复上述步骤,并将其误判个数分别记为 n_2^* 和 n_3^* ;

(5)以式(9)作为误判率的估计值。

$$a^* = \frac{n_1^* + n_2^* + n_3^*}{n_1 + n_2 + n_3} \times 100\% \quad (9)$$

3 模拟验证

3.1 模拟试验设计

本文以长春市自由大路与亚泰大街交叉口为对象,如图 2 所示。该交叉口为双向八车道(单向四车道)十字型交叉口,采用固定配时信号控制方式,信号周期时长为 188 s。运用连续两周对该交叉口进行调查获得的数据对 VISSIM 模拟软件进行基础参数标定,模拟路网中流量输入方案见表 1,交通流构成包括大车和小车两类,其中小车比例 92%,大车比例 8%,左转比例 10%,直行比例 70%,右转比例 20%。交叉口的信号配时方案如表 2 和图 3 所示。并仿照 SCOOT 系统检测器的布设方式,在西-东方向每一车道上游位置分别设置交通检测器,用于采集该方向路段上游截面的交通流参数数据。

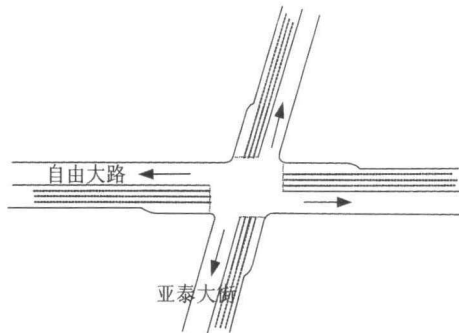


图 2 自由大路-亚泰大街交叉口
Fig. 2 Intersection of Ziyou-Yatai

表 1 模拟路网中流量输入方案(veh/h)
Table 1 Traffic volumes of road network

时间段/s	流量输入位置			
	东进口	北进口	南进口	西进口
0—1800	1350	1400	1320	1540
1801—3600	1440	1600	1400	1650
3601—5400	1620	1800	1510	1740
5401—7200	1760	1900	1640	1850
7201—9000	1940	2110	1720	1940
9001—10800	2010	2230	1850	2160
10801—12600	2150	2350	1750	2050
12601—14400	1940	2180	1660	1930
14401—16201	1850	1920	1630	1820
16201—18000	1680	1830	1520	1730
18001—19800	1530	1740	1410	1630
19801—21600	1420	1620	1300	1550

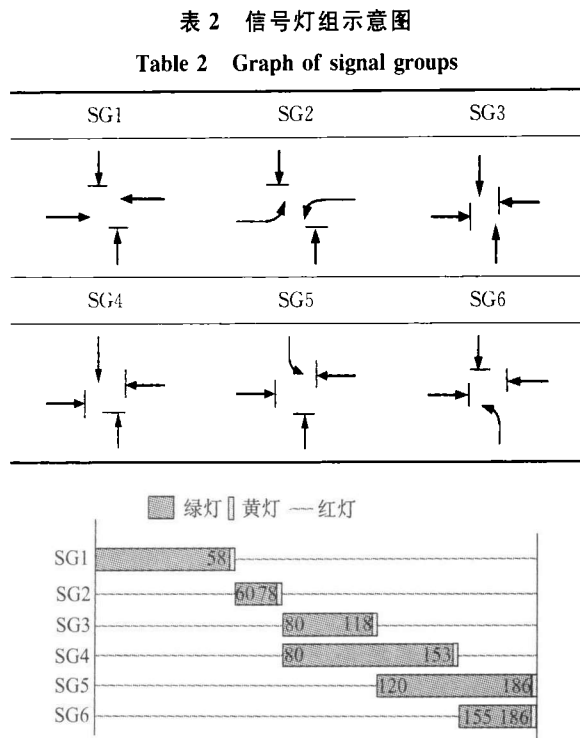


图 3 交叉口信号配时方案

Fig. 3 Signal assignment of the intersection

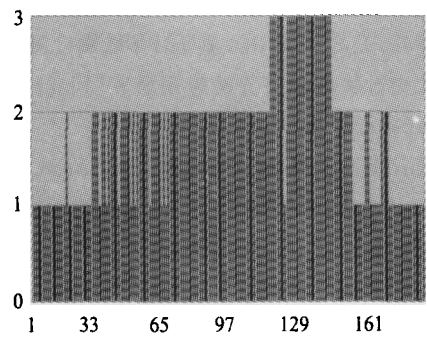
在设计试验方案时,为了体现交叉口交通信号配时的影响,本文将交通状态分析的时间间隔分别设定为 120 s、188 s 和 300 s,即小于信号周期、等于信号周期和大于信号周期 3 种试验方案。运用 VISSIM 模拟软件对每一种方案进行模拟试验,模拟时设定 11 个不同的随机数种子,每一个种子的模拟时间为 21 600 s。前 10 个随机数种子模拟产生的交通数据构成历史数据集 X,用于模糊聚类;第 11 个随机数种子模拟产生的交通数据用于对交通状态进行实时判别试验。

目前,VISSIM 提供的占有率、速度、流量等参数数据只能在模拟结束后获得,这些数据可以用于离线的模糊聚类,但不能用于对交通状态进行在线判别。为了建立对交通状态进行在线判别的试验环境,本文使用 Visual Basic 编程语言,通过 COM 接口控制 VISSIM 的运行,并在 VISSIM 模拟期间通过 COM 接口实时获取检测器的数据。在每一时间间隔结束时实时计算所获得的交通数据距离各类中心的隶属度数值,并按照隶属度最大原则判定当前的交通状态。

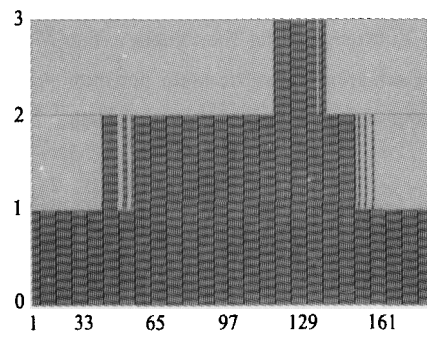
3.2 模拟结果分析

在本文建立的试验环境下,分别用文献[3]中的算法(记为算法 1)和本文提出的算法(记为算法 2),针对 3 种试验方案对该交叉口西—东方向路段的交通状态进行判别。图 4、图 5 和图 6 为

各种试验方案判别结果的示意图,横轴为时间,纵轴为拥挤程度,其中 1 表示顺畅,2 表示阻滞,3 表示拥堵。



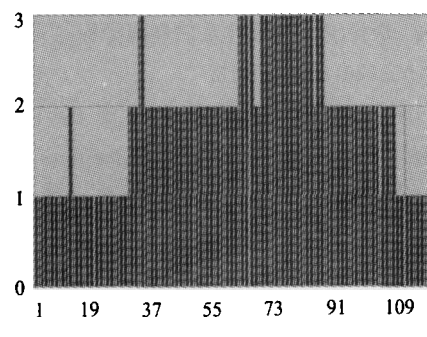
(a) 算法 1 判别结果



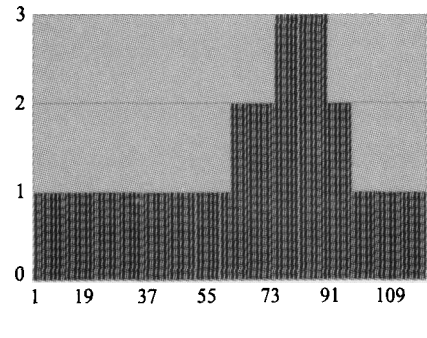
(b) 算法 2 判别结果

图 4 120 s 时间间隔判别结果

Fig. 4 Identified results of 120 seconds interval



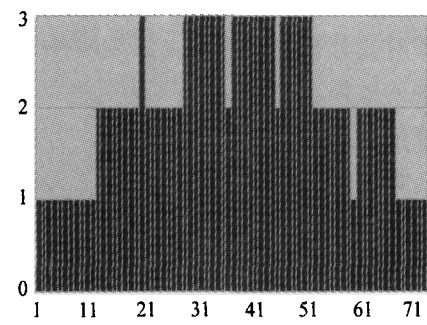
(a) 算法 1 判别结果



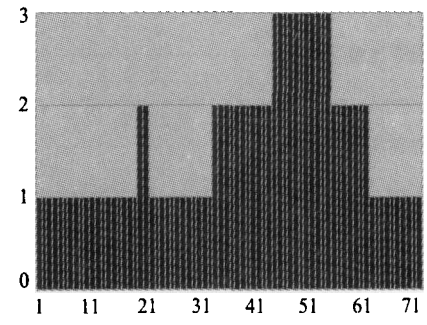
(b) 算法 2 判别结果

图 5 188 s 时间间隔判别结果

Fig. 5 Identified results of 188 seconds interval



(a) 算法 1 判别结果



(b) 算法 2 判别结果

图 6 300 s 时间间隔判别结果

Fig. 6 Identified results of 300 seconds interval

由图 4~图 6 可以看出,与算法 2 的判别结果相比,算法 1 的判别结果波动性相对较大,而且判别结果中的拥挤持续时间相对较长。这一结果与实际调查时所观测到的交通状况和模拟时所观察到的交通现象不是很吻合。其原因是流量指标在拥挤产生、持续至消散的整个过程中没有发生太大的变化,算法 1 以流量作为交通状态特征指标之一,且没有对数据的数量级进行标准化处理,弱化了拥挤期间其他参数所体现出的拥挤特征,因而将部分非拥挤区域数据划为了拥挤区域,导致拥挤期间较长。算法 2 将流量与占有率组合后转化为单车平均占有率,该指标对拥挤的时间敏感性和程度敏感性反应较强。

此外,由图 4~图 6 还可以看出,在交通状态分析的时间间隔小于交叉口的信号周期时长的条件下,交通状态判别的结果波动性较大;当分析时间间隔大于信号周期时长,交通状态的判别结果相对较为稳定;当分析时间间隔与信号周期时长一致时,交通状态判别的效果稳定性最好。然而,在实际应用中,由于各信号交叉口信号周期时长不同,尤其是自适应信号控制条件下各交叉口的信号周期时长不固定,无法以信号周期时长作为交通状态判别的时间间隔,因此交通状态分析的时间间隔应大于信号交叉口的最大信号周期时

长。

为了对比分析算法 1 和算法 2 两种交通状态判别算法的效果,本文以 300 s 为时间间隔,对算法 1 和算法 2 进行误判率交叉估计。由表 3 及式(9)可知,算法 1 的误判率为 $\alpha_1^* = 5.1\%$,算法 2 的误判率为 $\alpha_2^* = 2.9\%$ 。可以认为,本文所提出的算法的性能优于对比算法。

表 3 交通状态判别算法的性能比较
Table 3 Results of estimating identification algorithms

原始分类情况	n_1	n_2	n_3
	358	145	217
判别算法	误判个数		
	n_1^*	n_2^*	n_3^*
算法 1	17	7	13
算法 2	11	4	6

4 结束语

针对城市道路交通状态判别存在的难点,作者以 SCOOT 系统感应线圈检测器采集到的交通数据为基础,提取占有率、单车平均占有率和速度三个指标作为交通状态的特征向量,设计了一种基于模糊聚类的城市道路交通状态实时判别算法及其评价方法,分析了时间间隔长度对交通状态判别效果的影响。以 VISSIM 为工具,运用模拟手段对上述方法进行了模拟试验,并与已有方法进行了对比分析。结果表明,本文提出的判别算法能够提高城市道路交通状态实时判别的效果,

所提出的误判率交叉估计评价方法能够对城市道路交通状态判别算法的性能进行评价。

参考文献:

[1] 皮晓亮,王正,韩皓,等. 基于环形线圈检测器采集信息的交通状态分类方法应用研究[J]. 公路交通科技,2006,23(4):115-119.
Pi Xiao-liang, Wang Zheng, Han hao, et al. Application research of traffic state classification method based on collected information from loop detector [J]. Journal of Highway and Transportation Research and Development, 2006, 23(4):115-119.

[2] 劳云腾,杨晓光,云美萍,等. 交通状态检测方法的评价研究[J]. 交通与计算机,2006,24(6):74-78.
Lao Yun-teng, Yang Xiao-guang, Yun Mei-ping, et al. Evaluation of traffic state detection method[J]. Computer and Communications, 2006, 24(6):74-78.

[3] Jiang Gui-yan, Wang Jiang-feng, Zhang Xiao-dong, et al. The study on the application of fuzzy clustering analysis in the dynamic identification of road traffic state[C]// IEEE: Intelligent Transportation Systems Proceedings, 2003:408-411.

[4] Pal N R, Bezdek J C. On cluster validity for the fuzzy C-mean model [J]. IEEE Transactions on Fuzzy Systems, 1995, 3(3):370-379.

[5] 于剑. 论模糊 C 均值算法的模糊指标[J]. 计算机学报,2003,26(8):968-973.
Yu Jian. On the fuzziness index of the FCM algorithms[J]. Chinese Journal of Computers, 2003, 26(8):968-973.