

doi:10.3969/j.issn.1001-358X.2016.04.004

## 多分类器融合与单分类器影像分类比较研究<sup>\*</sup>

牛明昂<sup>1,2</sup>, 王 强<sup>2</sup>, 崔希民<sup>2</sup>, 赵康年<sup>1</sup>, 柴鹏辉<sup>2</sup>

(1. 青海大学地质工程系, 西宁 810016;

2. 中国矿业大学(北京)地球科学与测绘工程学院, 北京 100083)

**摘要:**针对单一分类器的不足,文中采用最短距离分类器、马氏距分类器、K-均值分类器对多光谱遥感影像进行分类,并在测量级的融合方法下进行多种分类器融合分类实验,最后采用混淆矩阵进行分类结果精度评价。实验结果表明:多分类器融合的遥感影像分类方法在精度上高于单一分类器分类。

**关键词:**多分类器融合;混淆矩阵;Kappa 系数;精度

中图分类号:TP751 文献标识码:A 文章编号:1001-358X(2016)04-0011-05

## A study on the classification and comparison of the multi-classifier fusion and the single classifier

Niu Mingang<sup>1,2</sup>, Wang Qiang<sup>2</sup>, Zhao Kangnian<sup>1</sup>, Chai Penghui<sup>2</sup>

(1. Department of Geology and engineering, Qinghai University, Xining, 810016,

2. College of Geoscience and Surveying Engineering, China University of Mining & Technology, Beijing, 100083;)

**Abstract:** Because of the limitations of the single classifier, the paper classifies the multi-spectral remote sensing image with the minimum distance classifier, Mahalanobis distance classifier and K-means classifier and makes the multiple classifier fusion experiment in the measurement level, based on which the paper makes precision accuracy of the experimental results with Confusion Matrix. It shows that the remote sensing classification method of multi-classifier is more accurate than the single classifier.

**Key words:** Multi-classifier fusion; confusion matrix; Kappa coefficient; precision

在遥感分类应用研究方面,经典模式识别分类方法如最小距离、马氏距离、ISODATA 故聚类等分类法<sup>[1-2]</sup>经常被使用。但是由于同物异谱、异物同谱现象的存在,这些方法很容易出现分类错误。故单一分类器的分类结果是有限的、有缺陷的。

多分类器融合算法的研究从最开始的模式识别领域<sup>[3]</sup>,到现在的图像领域,形成了许多比较成熟的多分类器融合的模型和算法<sup>[4-8]</sup>。众多学者研究成果表明,若单分类器之间差异越大,则多分类融合的优势反而会越加突出。因此,如果将各单分类器综合起来使用,则融合后的多分类器优点凸显,缺点压制,那么分类的精度和运算速度都会有所提升<sup>[9]</sup>。

本文以几个主要的单分类器理论及算法为依

托,进而探究了多分类器融合相关方法理论,并通过实验数据分析得出多分类器融合分类处理的精度在一定程度上优于单分类器。多分类器能够获得更好的分类效果,对分类作业提供了有益的指示。

### 1 遥感影像分类

遥感影像分类比较经典的应用是将模式识别技术引入到遥感技术领域,对影像中的地物属性进行识别与归类,从而实现从图像信息辨别出实际地物的目的。分类方法一般分为监督分类和非监督分类,本文也将研究基于这两类分类方法的多分类器融合方法。

<sup>\*</sup> 基金项目:国家自然科学基金面上项目(51474217)

引用格式:牛明昂,王强,崔希民,等.多分类器融合与单分类器影像分类比较研究[J].矿山测量,2016,44(4):11-15.

### 1.1 监督分类方法

监督分类以训练样本为前提需求,其重要原则是训练区域具有典型性和代表性<sup>[10]</sup>。本文中采用了最小距离法和马氏距法这两种常用的分类方法进行融合实验。

#### 1.1.1 最小距离法

最小距离法是用特征空间中的距离表示未知类别向量到要识别各类别代表向量中心点的相似程度,把未知类别的像元分配到相似程度最小的类别中去。具体原理操作过程如下:

(1)首先,统计每一类别的训练样本数据,计算出特征向量,将其作为这一类别特征空间中的中心位置。

(2)逐个像元计算到各自类别中心位置的欧氏距离。公式如式(1)所示:

$$D_{ei} = \sqrt{\sum_{j=1}^N (x_j - M_{ij})^2} \quad (1)$$

式中: $N$ 为波段数; $x_j$ 为像元 $X$ 在 $j$ 波段的光谱灰度值; $M_{ij}$ 为第 $i$ 类在第 $j$ 波段的光谱灰度均值。

(3)统计计算,像元按照最小距离归类。

(4)计算每个像元的隶属度,如式(2):

$$P_E(X \in C_i | X) = \frac{1/d_k(C_i | X)}{\sum_{i=1}^M 1/d_k(C_i | X)} \quad (2)$$

式中: $d_k(C_i | X)$ 为第 $k$ 个分类器中像元 $X$ 与类别 $C_i$ 的光谱矢量距离。

#### 1.1.2 马氏距法

马氏距法是一种广义距离法。其定义为:假定 $\Sigma$ 表示样本的协方差阵,即:

$$\Sigma = (\delta_{ij})_{P \times P} \quad (3)$$

$$\delta_{ij} = \frac{1}{n-1} \sum_{a=1}^n (x_{ai} - \bar{X}_i)(x_{aj} - \bar{X}_j) \quad (4)$$

$$\bar{x}_i = \frac{1}{n} \sum_{a=1}^n x_{ai} \quad \bar{x}_j = \frac{1}{n} \sum_{a=1}^n x_{aj} \quad (5)$$

如果 $\Sigma^{-1}$ 存在,则两样本之间的马氏距离可以表示为:

$$d_{ij}^2(M) = (x_i - x_j)' \Sigma^{-1} (x_i - x_j) \quad (6)$$

式中: $x_i$ 为 $x_i$ 的 $P$ 个指标组成的一维向量。这里定义样品 $X$ 到总体 $W$ 的马氏距离为:

$$d^2(X, W) = (X - \mu)' \Sigma^{-1} (X - \mu) \quad (7)$$

式中: $\mu$ 为总体样本均值, $\Sigma$ 表示协方差阵。

### 1.2 非监督分类方法

非监督分类也叫聚类分析,是一种不需要先验

知识的一种盲目的分类,分类的结果只能确定出来分出的类别,不能确定类别的属性。常见的非监督分类方法有:ISODATA 分类法<sup>[11]</sup>  $K$ -均值分类法<sup>[12]</sup>等。本文选取 $K$ -均值分类方法进行单独分类和与最小距离、马氏距离的融合实验。

$K$ -均值分类法重点在均值上,它需要给样本的空间聚类中心设置一个初始值和起始个数。先通过初始值形成一个集群,然后根据集群的统计参数不断调整某些参量,然后再聚类、再调整,不断迭代直到有关参数达到一定的限值,最终根据误差平方和的准则实现聚类空间中所有样本到聚类中心的距离平方和最小。

#### 1.3 基于混淆矩阵的精度评价

分类的精度评定一般选用混淆矩阵。混淆矩阵也称误差矩阵,作为模式识别领域中常用的一种精度评价标准格式,它能够表现出像元被分为某一类别与其真实代表地物类别的差异情况。我们一般使用用户精度、生产者精度、总体精度和 $Kappa$ 系数<sup>[13]</sup>来作为精度评价的指标。

#### 1.4 测量级多分类器融合

测量级多分类器融合使用的是 Bayesian 平均方法,其融合规则是:首先,以某种规则将不同分类器得出的所有表示像元属于各类别的级别测量值糅合起来,以糅合后的测量值为依据进行判别像元的重新归类。所以,测量级多分类器融合实质就是对各单分类器的后验概率估计进行平均,并由此计算得到待分类像元的最终类别<sup>[14]</sup>。

针对某个分类器 $\xi$ , $P_k$ 代表将模式 $X$ 划分为类别 $C$ 的后验概率测量,其中 $i=1,2,\dots,M, k=1,2,\dots,K$ 。Bayesian 平均结合规则如式(8)所示:

$$P_Z(X \in C_i) = \frac{1}{k} \sum_{k=1}^k P_k(X \in C_i) \quad i=1, \dots, M \quad (8)$$

最后的分类决策如式(9)所示:

$$E(X) = \begin{cases} j, & \text{当 } P_Z(X \in C_j) = \max_{i \in \Lambda} P_Z(X \in C_i) \\ M+1 & \text{其他} \end{cases} \quad (9)$$

针对 Bayesian 平均方法,一般比较常规的方法是直接融合各种不同的分类器,但是这需要一个前提,那就是:各个分类器输出的结果为同一个级别上的量级;而在分类器所使用的矢量距离之间容易出现差异大、不可比较的情况。所以,往往将不可比较的距离矢量测量转换为与概率可比的概率矢量测

量。转换公式如式(10)所示:

$$P_k(X \in C_i | X) = \frac{1/d_k(C_i | X)}{\sum_{i=1}^M 1/d_k(C_i | X)} \quad (10)$$

式中: $d_k(C_i | X)$ 表示在第  $k$  个分类器中,模式  $X$  与类别  $C_i$  的矢量距离。

## 2 实验与分析

实验源数据为 TM 多光谱遥感影像,共 7 个波段,选择其中的第 3、4、5 波段进行实验。影像中区域颜色定义如下:深绿色代表林地,黄绿色代表为耕地,灰褐色代表裸地,白色代表沙地。首先利用 ENVI4.7 进行样本采集;然后通过 MATLAB 代码进行分类。

实验运用的具体技术路线如图 1 所示。

### 2.1 样本采集

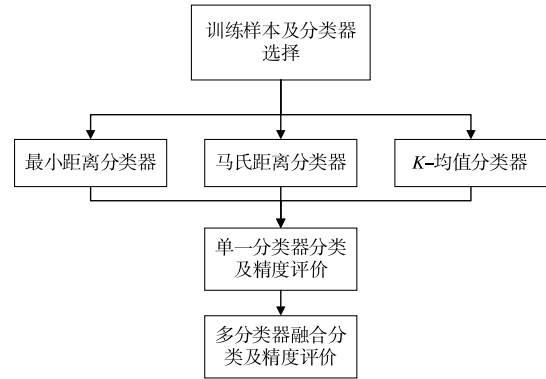


图 1 技术路线图

作为同一训练样本的选择原则是:地物样本的光谱特征均一、色调一致。首先利用 ENVI 软件的点取功能,对四种类型的每个样本类型采集 12 个样本点,则总共可以得到 48 个样本点,将各样本点在 3、4、5 波段的灰度值记录在表 1 中。

表 1 四种类别样本像元原始值

样本编号	类别 1			类别 2			类别 3			类别 4		
	各波段号像元值			各波段号像元值			各波段号像元值			各波段号像元值		
	5	4	3	5	4	3	5	4	3	5	4	3
1	85	117	51	146	255	85	158	127	142	207	223	193
2	67	106	56	182	255	136	188	165	170	231	234	233
3	73	95	34	158	255	96	182	154	159	207	223	216
4	79	106	62	170	255	102	170	127	130	255	255	255
5	54	106	51	213	234	113	176	175	176	243	255	255
6	54	106	34	188	255	119	213	154	147	237	223	233
7	85	127	62	195	234	119	182	154	153	255	255	250
8	60	106	51	176	245	130	225	165	159	255	255	255
9	36	74	39	140	255	68	152	143	125	231	234	238
10	30	63	17	164	255	108	140	106	96	255	255	255
11	60	106	34	140	255	73	152	127	119	213	223	233
12	54	95	45	140	255	79	146	117	102	255	234	255

### 2.2 单分类器分类结果

一般情况下,按照光谱特征的分类在分类结果中会产生一定的图斑,实验中在各个分类器程序中,采用中值滤波平滑初分类产生的噪声。由于是四种类别,将三种方法的分类结果影像统一划分为 0,0.45,0.8,1 四个等级,分类结果如图 2~图 4。其中,白色表示沙地;浅白色表示耕地;灰白色表示裸地;黑色表示林地。

### 2.3 单分类器分类精度对比

按照平均一种地物用 30 个样本点来验证,四种地物类型需要随机选取 120 个特征点,将原始影像的这 120 个特征点坐标记录下来,找到 3 张分后的

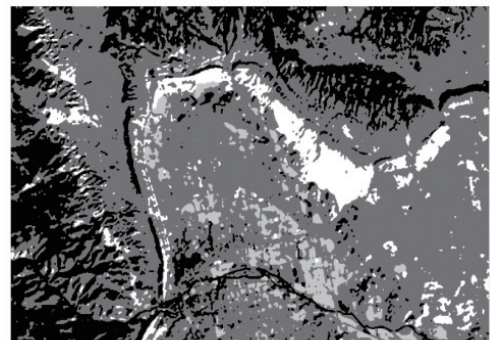


图 2 最小距离分类结果

影像图上对应的点与之进行比较,看分类正确与否的程度。

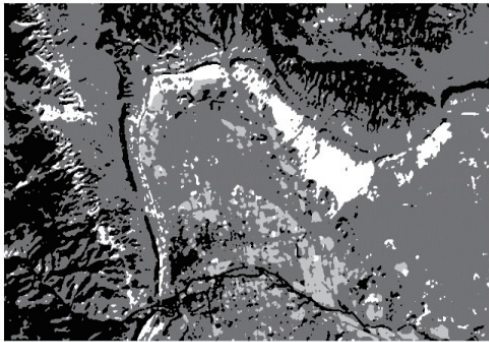


图 3 马氏距离分类结果

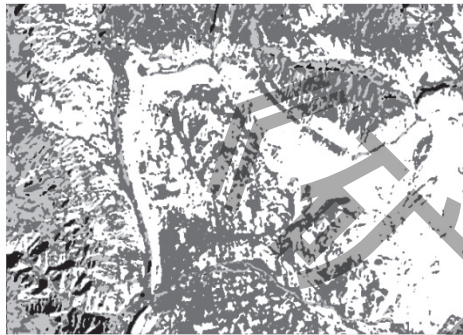


图 4 K-均值分类结果

根据分类正确与否程度可以计算得出每种分类器的混淆矩阵,进而得到各类分类器的各个类型精度和  $Kappa$  系数,见表 2、表 3。 $K$ -均值分类情况不佳,将耕地和大部分裸地都划分为沙地,林地被划分为耕地和裸地,总体分类精度和  $Kappa$  系数分别只有 16.53% 和 0.268。

表 2 最短距离分类精度

分类类型	林地	耕地	裸地	沙地
生产者精度/%	74.07	73.08	77.27	91.30
用户精度/%	80	90.48	77.27	70
总体精度/%	78.33			

$Kappa = 0.704$

表 3 马氏距离分类精度

分类类型	林地	耕地	裸地	沙地
生产者精度/%	77.78	80.77	75	95.65
用户精度/%	80.77	80.77	86.84	73.33
总体精度/%	80.83			

$Kappa = 0.741$

#### 2.4 多分类器融合影像分类

测量级融合分类执行流程如图 5 所示。

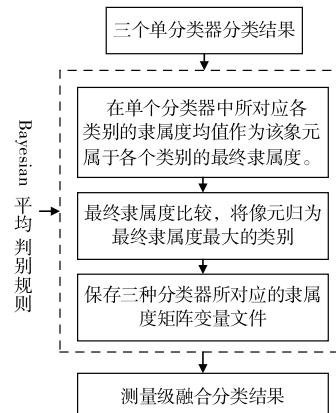


图 5 测量级融合分类流程

图 6 是测量级融合处理的分类结果,表 4 展示了测量级多分类器融合的分类精度。

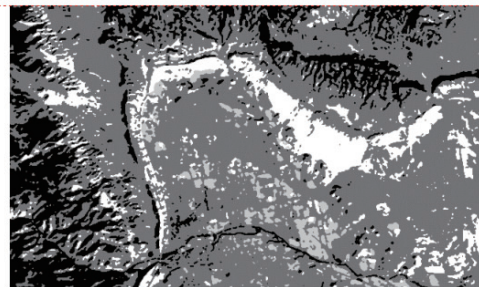


图 6 测量级融合分类结果

表 4 测量级分类精度

分类类型	林地	耕地	裸地	沙地
生产者精度/%	81.48	80.77	84.09	95.65
用户精度/%	84.62	91.30	86.05	78.57
总体精度/%	85.0			

$Kappa = 0.796$

#### 2.5 精度评价分析

将上述分类结果放在一起做比较,可得统计图 7。

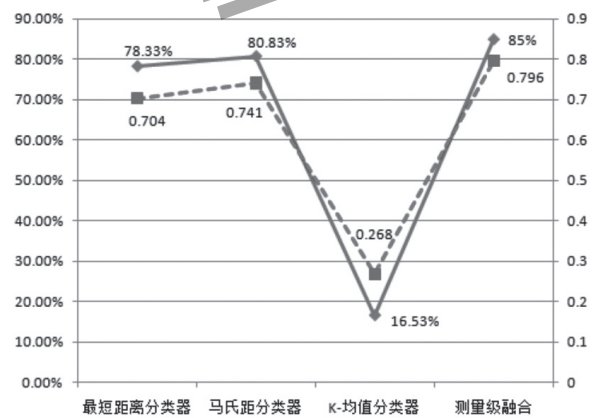


图 7 不同分类器精度统计图

从图7以及上述各表数据可以看出:

(1)最小距离和马氏距离分类器对于本影像的分类效果比较好,而K-均值分类情况比较差。说明了单一分类器是有缺陷的,而且不同分类器之间分类效果可能会有很大的差别;

(2)从表2到表4中可以看出:在一定相同条件下,相对于单分类器而言,测量级多分类器融合下的生产者精度和用户精度要高一些;

(3)测量级多分类器融合对总体分类精度有一定的提高。融合后的遥感影像不仅弥补了K-均值分类结果差的缺点,而且分类精度比前两种单一分类器分类精度高了约五个百分点,最高的总体分类精度可以达到85%。同时,从Kappa系数上比较可以看到,测量级融合分类比其它单分类的Kappa系数都高,精度可达0.796,将近0.80。0.80可以这么理解:相对于随机分配像素类别而言,测量级融合分类以80%的程度将像元归划到正确的类别中。

### 3 结 语

本文研究了兼顾监督分类和非监督分类的多分类器融合的理论,并通过实验采用混淆矩阵对分类后的影像进行精度评价。实验表明,多分类器融合方法能够在一定程度上弥补单一分类器的缺点,从而能取得较高的分类精度。

在精度评价内容中我们仅仅采集了120个点进行对比试验,样本点相对较少,会导致评定结果有一定的误差。另外,单一分类器有很多种,倘若大量分类器之间存在一定的相关性,如何选择合适的分类器组合也是需要进一步研究的内容。不过,对于一般性大部分遥感影像分类而言,多分类器融合分类针对单一分类器在分类精度提高方面是一个应该着重研究的方向。

#### 参考文献:

- [1] SCHOWENGERDT R A. Remote sensing: models and methods for image processing [M]. Academic press, 2006.
- [2] Richards J A, Richards J A. Remote sensing digital image analysis[M]. Berlin et al.: Springer, 1999.
- [3] GHOSH J, TUMER K, BECK S, et al. Integration of

neural classifiers for passive sonar signals [J]. Control and Dynamic Systems, 1996, 77: 301 - 338.

- [4] SAMADZADEGAN F, BIGDELI B, RAMZI P. A multiple classifier system for classification of LIDAR remote sensing data using multi-class SVM[M]. Multiple Classifier Systems. Springer Berlin Heidelberg, 2010.
- [5] 李明诗,彭世揆,周林,等.基于ASTER数据的决策树自动构建及分类研究[J].国土资源遥感,2006(3): 33 - 36,42.
- [6] 张秀英,冯学智,刘伟.基于多分类器结合的IKONOS影像城市植被类型识别[J].东南大学学报(自然科学版),2007,37(3):399 - 403.
- [7] 韩建峰,杨哲海.组合分类器及其在高光谱影像分类中的应用[J].测绘科学技术学报,2007,24(3):231 - 234.
- [8] 张丹.基于多分类器融合的遥感影像分类方法研究[D].阜新:辽宁工程技术大学,2009.
- [9] 柏延臣,王劲峰.结合多分类器的遥感数据专题分类方法研究[J].遥感学报,2005,9(5):555 - 563.
- [10] 钱乐祥,等.遥感数字影像处理与地理特征提取[M].北京:科学出版社,2004.
- [11] 张红,舒宁,刘刚.多时相组合分类法在土地利用动态监测中的应用[J].武汉大学学报(信息科学版),2005,30(2):131 - 134.
- [12] 任江涛,孙婧昊,施潇潇,等.一种用于文本聚类的改进的K均值算法[J].计算机应用,2006(S1):73 - 75.
- [13] 章孝灿.遥感数字影像处理[M].杭州:浙江大学出版社,1997.
- [14] XU L, KRZYZAK A, SUEN C Y. Methods of combining multiple classifiers and their applications to handwriting recognition[J]. Systems, Man and Cybernetics, IEEE Transactions on, 1992, 22(3):418 - 435.
- [15] 王筱君.高光谱遥感在植被分类中的研究[J].矿山测量,2014,42(1):18 - 22.

作者简介:牛明昂(1991-),男,河南许昌人,青海大学在读硕士研究生,研究领域为遥感图像预处理,遥感数字图像处理与分析,图像分类研究,遥感地质解译,资源环境遥感监测。

(收稿日期:2016-05-26)