

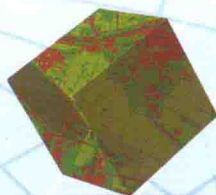
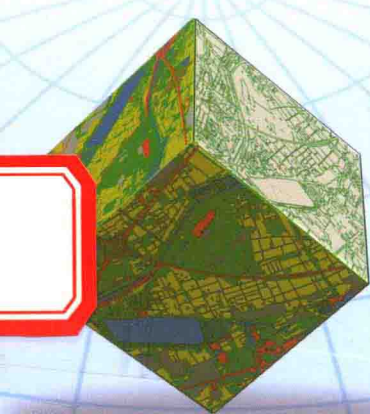


测绘地理信息科技出版资金资助
CEUI DILI XINXI KEJI CHUBAN ZIJIN ZIZHU

遥感数据可靠性 分类方法研究

Study on Reliable Classification Methods
Based on Remotely Sensed Imagery

张 华 史文中 汪云甲 著



测绘出版社

江苏高校优势学科建设工程
测绘地理信息科技出版资金资助

遥感数据可靠性分类方法研究

Study on Reliable Classification Methods
Based on Remotely Sensed Imagery

张 华 史文中 汪云甲 著

测绘出版社

· 北京 ·

© 中国矿业大学环境与测绘学院 2016
所有权利(含信息网络传播权)保留,未经许可,不得以任何方式使用。

内容简介

本书总结了作者近年来在遥感数据分类方面的研究成果,以遥感数据分类过程为出发点,从训练样本数据对遥感数据分类精度的影响、遥感数据可靠性分类模型和分类精度评价的可靠性样本抽样方法等方面系统地研究遥感数据分类过程中主要环节的不确定性对分类结果精度的影响,并提出遥感数据可靠性分类方法,以提高最终分类结果精度。研究成果将为提高遥感数据分类结果的精度提供一整套新的解决方案。

本书在内容上力求做到深入浅出、通俗易懂,不仅具有一定的深度和广度,而且反映学科的新动向、新问题。本书可作为高等学校遥感相关专业研究生的参考书,也可供遥感相关科技人员阅读参考。

图书在版编目(CIP)数据

遥感数据可靠性分类方法研究 / 张华, 史文中,
汪云甲著. — 北京: 测绘出版社, 2016. 3
ISBN 978-7-5030-3917-1

I. ①遥… II. ①张… ②史… ③汪… III. ①遥感数
据—可靠性—分类—研究 IV. ①TP701

中国版本图书馆 CIP 数据核字(2016)第 011105 号

责任编辑	贾晓林	封面设计	李伟	责任校对	曹平	责任印制	陈超
出版发行	测绘出版社			电 话	010—83543956(发行部)		
地 址	北京市西城区三里河路 50 号				010—68531609(门市部)		
邮政编码	100045				010—68531363(编辑部)		
电子邮箱	smp@sinomaps.com			网 址	www.chinasmp.com		
印 刷	三河市博文印刷有限公司			经 销	新华书店		
成品规格	169mm×239mm						
印 张	11.75			字 数	225 千字		
版 次	2016 年 3 月第 1 版			印 次	2016 年 3 月第 1 次印刷		
印 数	0001—1000			定 价	58.00 元		
书 号	ISBN 978-7-5030-3917-1						

本书如有印装质量问题,请与我社门市部联系调换。

前 言

遥感数据分类技术是从遥感数据中提取专题类别数据的一个重要手段,但由于自然环境的复杂性、遥感传感器及分类算法的局限性等原因,不确定性伴随遥感数据分类的整个过程,在利用遥感数据分类获取的信息中含有较大的不确定性。如何了解这些不确定性的本质,降低其对分类结果精度的影响,提高分类精度,建立可靠的遥感数据分类方法,是遥感数据分类研究的重要问题。为此,本书在国家自然科学基金重点项目“可靠性遥感影像分类与空间关联分析研究”(41331175)、国家自然科学基金项目“基于模糊拓扑及多特征融合的遥感影像亚像元定位”(41201451)、“江苏高校优势学科建设工程(测绘科学与技术)”和国家测绘地理信息局公益行业科研专项“面向地理国情服务的资源开发地表沉降监测”(201412016)等项目的资助下,以遥感数据分类过程为出发点,研究分类过程中主要环节的不确定性对分类精度的影响,提出遥感数据可靠性分类方法,提高最终分类结果精度。研究内容包括训练样本数据对遥感数据分类精度的影响、遥感数据可靠性分类模型和遥感数据分类精度评价三个部分。研究成果将为提高遥感数据分类结果的精度提供一整套新的解决方案。

全书共分六章。第1章为绪论,介绍了研究背景、研究进展、研究目标与研究内容。第2章通过对典型区域遥感数据分类实验的研究,得出训练样本数量、质量及抽样方法与不同分类方法分类精度的响应关系。第3章针对混合像元是造成分类精度低的根本原因,将模糊拓扑理论引入传统支持向量机,提出比传统支持向量机分类精度更高的模糊拓扑支持向量机遥感影像分类方法。第4章充分考虑不同分类器对每个像元的识别能力,提出了基于矩阵特征向量的多分类器组合方法,提高了分类精度和稳定性。第5章提出融合光谱与空间特征的遥感数据多种分类方法:改进后的模糊聚类图像分割模型(Markov random field based fuzzy c-means algorithm, MFCCM)、基于引力模型的马尔可夫模型(spatial attraction based Markov random field, SAMRF)、基于光谱、纹理和像元特征融合的遥感数据分类模型。第6章提出了基于空间均衡抽样和基于聚类的空间分层抽样方法进行检验样本选择,保证了检验样本的均匀性和代表性,从而提高了精度评价结果的可靠性。

本书以分类过程为线索进行组织,具有一定的深度和广度,反映学科的新动向

和问题,介绍学科前沿问题和新的成果。当然,遥感影像分类精度的影响因素很多,本书中阐述的某些学术观点仅为一家之言,限于作者的水平 and 经验有限,难免有不妥之处,恳请专家和广大读者批评指正。

作 者

2015 年 11 月

目 录

第 1 章 绪 论	1
1.1 研究背景	1
1.2 国内外研究进展	2
1.3 研究内容	14
第 2 章 训练样本数据对遥感数据分类精度的影响	17
2.1 训练样本数量对分类精度的影响	17
2.2 训练样本质量对分类精度的影响	31
2.3 训练样本抽样方法对分类精度的影响	40
2.4 小 结	43
第 3 章 基于模糊拓扑支持向量机的遥感数据分类	44
3.1 支持向量机	44
3.2 基于 FTSVM 的遥感数据分类模型	46
3.3 实验结果与分析	53
3.4 小 结	61
第 4 章 基于自适应权值的多分类器组合遥感数据分类	62
4.1 标准的多分类器组合	62
4.2 基于矩阵特征值自适应权值多分类器组合模型	65
4.3 实验结果与分析	69
4.4 小 结	76
第 5 章 融合光谱与空间特征的遥感数据分类	77
5.1 基于 MFCM 分割的遥感影像分类	77
5.2 基于空间引力模型的 MRF 遥感数据分类	95
5.3 基于光谱、纹理和像元形状特征融合的遥感数据分类	107
5.4 小 结	152

第 6 章 基于可靠性抽样方法的遥感数据分类精度评价.....	154
6.1 基于空间均衡抽样的遥感数据分类精度评价	154
6.2 基于聚类空间分层抽样的遥感数据分类精度评价	160
6.3 小 结	166
参考文献.....	167

CONTENTS

Chapter 1	Introduction	1
1.1	Research Background	1
1.2	Research Progress	2
1.3	Research Content	14
Chapter 2	Effect of Training Samples on Remote Sensing Image Classification Accuracy	17
2.1	Effect of Training Samples Number on Remote Sensing Image Classification Accuracy	17
2.2	Effect of Training Samples Quality on Remote Sensing Image Classification Accuracy	31
2.3	Effect of Training Samples Sampling Methods on Remote Sensing Image Classification Accuracy	40
2.4	Summary	43
Chapter 3	Fuzzy Topology-Integrated Support Vector Machine for Remote Sensing Image Classification	44
3.1	Support Vector Machine	44
3.2	FTSVM Model for Remote Sensing Image Classification	46
3.3	Experiment Results and Analysis	53
3.4	Summary	61
Chapter 4	Adaptive Weight Multiple Classifiers Combination for Remote Sensing Image Classification	62
4.1	Normal Multiple Classifiers Combination	62
4.2	Multiple Classifiers Combination Based on Adaptive Weight Matrix Eigenvalues	65
4.3	Experiment Results and Analysis	69
4.4	Summary	76

Chapter 5 Remote Sensing Image Classification Based on the Fusion of Spectral and Spatial Features	77
5.1 Remote Sensing Image Classification Based on MFCM Segmentation	77
5.2 Remote Sensing Image Classification Based on Spatial Attraction Markov Random Field	95
5.3 Remote Sensing Image Classification Based on the Fusion of Spectral, Texture and Shape Features	107
5.4 Summary	152
Chapter 6 Remote Sensing Image Classification Accuracy Assessment by Using Reliable Sampling Methods	154
6.1 Remote Sensing Image Classification Accuracy Assessment by Using SBS Method	154
6.2 Remote Sensing Image Classification Accuracy Assessment by Using Cluster-based Spatial Stratified Sampling Method	160
6.3 Summary	166
References	167

第1章 绪论

1.1 研究背景

遥感技术因其覆盖范围广、多时相、现势性、多分辨率、多光谱和信息量丰富等优势使其成为获取专题类别数据(如土地覆被/利用数据等)的重要手段之一,是进行科学决策的重要依据,在国民经济发展中有着极其重要的作用。如何高效、快速、智能地从遥感数据中获取高精度的专题类别数据,成为遥感信息提取研究的核心问题之一。

遥感数据分类技术是从遥感数据中提取专题类别数据的一个重要技术。但同时,通过遥感数据分类获取的专题类别信息存在很多局限性,其中最大的问题就是通过遥感数据分类技术得到的专题类别数据的可靠性。由于自然环境以及自然环境与遥感波谱相互作用的复杂性,如同物异谱、同谱异物、时相变化和地物单元空间分布的相互交错关系等现象,造成了影像中大量的混合像元存在,使得从遥感影像光谱信息中提取的相关专题类别信息中总是存在着不确定性。在遥感数据分类过程中,数据转换、分析与处理、遥感数据尺度、分类算法选取、分类后处理,以及最终的精度评价等各个环节都会不可避免地引入不同程度与类别的不确定性,并在随后的处理过程中传播与累积(Lunetta et al, 1991)。因此,可以说遥感信息的本质决定了不确定性将伴随着遥感数据分类的整个过程。在利用遥感数据分类得到专题类别信息时,需要了解这些信息的不确定性的本质,并科学地表达和处理其中蕴含的不确定性,降低这些不确定性对分类精度的影响,建立可靠的遥感数据分类方法,提高分类精度与效率,从而获取高精度的专题类别信息。

从模式分类的角度可将遥感数据分类看作是通过某些具有类别标识的数据集进行统计学习,得到相同类别数据的特征数据,然后利用得到的类别特征将未标识数据进行区分分类,即构造一个分类器,其过程一般包括:①将已知类别的数据集分成训练和验证样本数据集;②采用某种算法对训练样本数据集进行学习,得到一个符合一定精度要求的分类模型;③通过验证样本数据集对训练得到的分类模型进行验证,若验证结果符合要求,则进行④,否则,返回②;④应用②得到的分类模型对待分类数据进行分类。因此,为获得可靠的分类结果必须要保证分类过程中每个环节具有可靠性,包括分类的训练样本数据、分类算法和分类结果精度评价的可靠性。

本书将从遥感数据分类器构造(分类)过程出发,提出遥感数据可靠性分类方法,在研究其不确定性传播机理的基础上,对分类过程的可靠性和结果的精度进行控制与评价,通过层层控制,提高分类过程中的主要环节的可靠性,降低其不确定性对最终结果精度的影响,提高遥感数据分类结果的精度,最终获得一个遥感数据的可靠性分类方法。研究成果将为降低遥感数据分类结果的不确定性提供一整套新的解决方案,在遥感影像信息提取、遥感数据应用和空间数据不确定性研究等方面具有重要理论意义与应用价值。

1.2 国内外研究进展

遥感数据从成像到分类产品输出大约需要经过以下几个过程:遥感数据获取,数据处理,信息提取、融合、集成,精度评价和产品输出。在这一过程中,由于操作和环境等诸多因素的影响而引入了各种误差:①数据获取误差;②数据转换误差;③数据分析与处理误差;④分类和信息提取误差;⑤误差评价产生的不确定性(Lunetta et al,1991;史文中 等,2012;Congalton,1991;Shi,2009),即每一个环节都会不可避免地引入不同类型及程度的不确定性,并在随后的各种处理过程中传播与累积(图 1.1)。

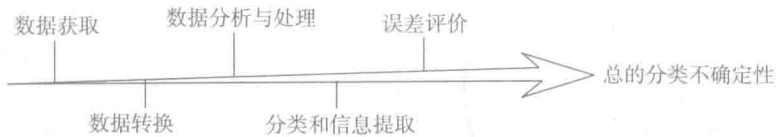


图 1.1 遥感数据分类过程中的不确定性累积

下面分别从训练样本数据、分类算法、多分类器组合、多特征遥感数据分类、分类精度评价对最终遥感数据分类精度的影响等方面论述国内外相关研究现状与发展趋势,指出存在的问题,并提出解决方法。

1.2.1 训练样本数据

监督分类是根据一定数量且具有代表性的训练样本数据集,经过特征选择,计算类别特征参数并作为判断依据,建立相应的决策函数对待分数据进行分类(延森等,2007;Han et al,2001)。研究表明,训练样本对遥感数据分类精度的影响比分类方法本身对分类精度的影响还要大(Hixson et al,1991),训练样本的质量、数量和空间分布都会对遥感数据分类结果精度产生重要的影响,要获得高精度分类结果,一般要求足够数量的且具有很好的代表性的样本数据集(Jackson et al,2001)。为此,下面将从样本量、样本质量及样本抽样方法(空间分布)三个方面论述样本数据对分类结果精度的影响。

1. 样本量方面

Nelson 等(1984)和 Van 等(2005)从统计学的角度给出了遥感数据分类所需要的最小样本数量的计算公式。Zhuang 等(1994)分别抽取了研究区 Landsat TM 影像总体像元个数的 5%、10%、15% 和 20% 的样本数据作为训练样本集,并利用神经网络方法对影像进行了分类,结果表明,要获得一个较好的分类精度大约需要 5%~10% 的样本量。Foody 等(1997)研究了影响神经网络遥感数据分类精度的主要因素,得出样本量是影响分类精度的一个重要因素。阎静等(2001)利用神经网络对不同样本训练分类的结果进行了研究。Foody 等(2004a;2004c)针对支持向量机分类算法构建了智能样本选取方法,实验结果表明并不是所有的训练样本对提高支持向量机分类精度都有作用,利用辅助信息可以大大减少训练样本的数量。Foody 等(2006a)研究了某个类别分类时所需的最小样本数量的问题,并基于支持向量分别用四种方法减少了标准训练数据样本量,给出了样本量的减少与生产精度及用户精度之间的变化关系。朱秀芳等(2007)探讨了训练样本量和分类方法对农作物种植面积遥感监测精度的影响,分别利用支持向量机和最大似然等六种分类方法在不同训练样本量下进行了多次实验,实验结果表明不同样本量下各种分类方法的分类精度都存在着波动,而且地块越破碎波动性越大,但随着样本量的增加波动逐渐变小;在相同样本量下,不同分类方法对同一区域内小麦的识别率是不相同的,同种方法对不同地区小麦的识别能力也不一样。薄树奎等(2010)基于统计学理论给出了训练样本数量确定的理论依据,并通过实验验证了训练样本数量与分类精度的关系,得出面向对象分类中的样本量比基于像元分类的样本量要小得多,而且与地物的复杂程度相关。综上,这些方法研究了特定的分类方法和遥感数据下的样本量与分类精度之间的关系,但对于不同分类方法针对不同的遥感数据还有待于进一步研究。

2. 样本质量方面

训练样本的质量是影响遥感数据分类精度的另一个重要因素(Kuo et al, 2002)。为获取高精度的分类结果,一般需要尽量选择一些高质量(纯)的、信息含量大的样本以提高样本的代表性。由于训练样本中可能含有大量的混合像元,Arai(1992)针对最大似然法分类,提出了一种纯化训练样本的方法,并利用纯化后的训练样本进行训练分类,分类结果精度得到了很大的提高。吴健平等(1996)将计算出的图像像元灰度值与类别灰度值平均值之差,以及每一像元与周围像元的总的方差大小作为准则进行了像元的纯化,实验结果表明经纯化后的训练样本可以提高类别间的发散度、高斯分布的拟合度和分类结果的精度。Foody 等(2006b)分别利用含有混合像元的样本训练集和纯像元训练样本集对支持向量机分类算法进行研究,实验结果表明使用含有混合像元的训练样本集与纯净样本集获得的分类精度相差很小,但当训练样本量较小时,含有混合像元的训练样本集获得的分类

精度要高于纯净样本集获得的精度。上述方法是针对某种特定的分类方法进行研究的,那么对于不同质量的训练数据与不同的分类器之间的响应关系到底如何,有没有一个数据质量标准来衡量训练样本的质量等。目前,主要是通过样本间的分离度对样本质量进行分等定级(Schmidt et al, 2003;杜红艳等,2004),分离度指标主要包括欧氏光谱距离、分离度(连石柱,1996)、转换分离度(张治英等,2004)和J-M距离(Key et al, 2001;Koukoulas et al, 2001b;Van et al, 2005),一般认为样本的分离度大于1700时就可以将两个类别区别开,但在某些情况下,即使分离度很大,分类精度也不会很高,像元错分现象出现得比较多。Foody等(2006b)指出影响支持向量机分类精度的训练样本像元主要是那些支持向量像元,这些像元通常分布在类别边界处。而且对于地形复杂区域,混合像元大量存在的遥感影像数据,其类别间的分离度本身就很低,因此仅依靠分离度指标去衡量训练样本的质量是不够完善的。不同的分类器是否有相同的样本质量标准?相同质量的训练样本对不同的分类器响应关系是怎样的?不同质量等级的训练样本对相同的分类器的响应关系是怎样的?这些问题都值得进一步研究。

3. 样本抽样方法方面

大样本量数据一般会提高分类结果的精度和可靠性,但同时分类成本会增加,所以一般选取一些具有代表性的样本作为训练样本集(Scott, 1998)。为保证分类结果的精度和可靠性,一般要求抽样方法必须确保抽样结果的无偏性,抽样方法决定样本的空间分布,样本的空间分布又会影响采样成本,从而影响精度评价的成本(Dunn et al, 1993)。研究表明,当样本之间存在较强相关性时,样本点值在一定范围内是很相似的,特别容易造成样本点信息的重叠,当采用经典抽样方法进行抽样时,由于经典抽样方法假设样本之间是完全独立的,忽略了因样本之间的相关性而导致的信息损失,使得所抽取样本的代表性降低(曹志冬等,2008)。统计学中给出了常用的如简单随机采样、系统采样、层次随机采样、聚集采样及层次系统采样等抽样方法(冯士雍,2007;冯士雍等,1996)。在这些传统抽样方法中,抽样对象一般不具有空间特性,不考虑数据的空间分布特征,但实际上,大部分数据是具有空间特征的,若改变抽样对象的位置,将会影响最终的抽样结果。李连发等(2004)提出了空间抽样框架,以及如何在地理数据上进行抽样调查的方法和理论,实验表明提高了空间抽样的效率。Wang等(2002)提出了依据抽样对象的特征进行空间分层抽样的“三明治”空间抽样模型,此方法充分利用了抽样对象的自然特征,同时也考虑了实际工作中需要按行政区统计结果的要求,为此类抽样问题提供了一个好的解决方案。国外学者在样点布局、选择及优化等相关研究成果比较丰富,主要有基于前向/后抽样的最优样本布局、双向抽样的最优样本布局(Webster et al, 2006;Simbahan et al, 2006)、随机选择法(Brus et al, 1997)、枚举法(Rogerson et al, 2004)和序贯法(Journel, 1988)等。样本点选择方法包括模拟退火法(Van et

al, 1999; Van et al, 2000)、空间均衡布样(Stevens, 1997)、适应性抽样(Smith et al, 1995)、Kriging 方差最小化(Bertolino et al, 1983)、WM(Van et al, 1999)、极大熵(Nunes, 2006)、分形维(Zio et al, 2004)和均方距离(Stevens, 2006)等样本优化选择准则。这些研究使得样本点选择逐渐由主观知识过渡到优化决策模型自动选择,保证了抽样结果的科学性,避免了基于主观判断的误差,合理选择样本点,使样本点所含信息量最大,具有最大的代表性。

可以发现,针对训练样本的不确定性对遥感数据分类精度影响的研究相对较少,本书将系统地研究训练样本数量、质量及空间分布(抽样方法)对遥感数据分类精度的影响的定量或定性关系,为训练样本和分类方法的选择提供一定的理论和方法基础。

1.2.2 分类算法

遥感数据分类是将图像中每个像元根据像元灰度值、空间上下文及其他分类信息依据相应的判断规则划分到预设的类别(Lu et al, 2007)。早期的遥感数据分类工作主要依靠人的经验和知识,通过目视解译来完成。目视解译技术发展已经很成熟,但由于其效率低、成本高、周期长、需要解译经验且解译结果中含有较大的不确定性等缺点,难以满足海量遥感数据的快速、自动、智能分类要求。而计算机技术的发展为遥感数据自动分类提供了坚实的基础,目前,计算机自动分类技术已成为遥感数据自动分类的研究重点。近年来,国内外学者致力于研究发展先进的遥感数据分类技术来获取高精度分类结果(Gong et al, 1992; Foody, 1996; 李祚泳, 1998; Gallego, 2004; 术洪磊 等, 1997; 余凡 等, 2012; Zhong et al, 2012),如神经网络(李祚泳, 1998; Kavzoglu et al, 2003)、支持向量机(support vector machine, SVM)(Huang et al, 2002; Zhu et al, 2002; Pal, 2006; Pal et al, 2004; Melgani et al, 2004)、专家系统(Hodgson et al, 2003)等算法广泛应用于遥感数据自动分类,但遥感数据分类精度受多种因素的影响,其中主要因素有地物的空间分布结构、遥感数据的尺度问题、遥感数据预处理方法、分类算法本身差异等,分类结果的可靠性仍需进一步深入研究,当然,合适的分类方法是保证分类精度的关键。现已出现了多种遥感数据分类方法,分类方法的分类体系很多,本书按参数和非参数的分类体系来论述遥感数据分类方法的国内外研究进展。

1. 参数分类器

参数分类器一般假设待分类数据满足高斯正态分布,首先选取一定量的训练样本,然后统计这些样本的特征参数(如均值和方差等),依此建立决策函数对待分类数据进行分类。常用的参数分类器主要有最大似然分类(maximum likelihood classification, MLC)、平行多面体、最小距离(minimum distance, MinD)和马氏距离(Mahalanobis distance, MaD)等(杜培军, 2006)。但由于实际中的遥感数据的

分布基本不符合正态分布,因此,参数分类器的分类结果具有较大的不确定性,且参数分类器对高维数据分类的效果不是很好,不利于融合其他分类知识等。随着遥感技术的不断发展与提高,遥感影像数据的光谱、空间及时间分辨率都得到了很大的提高,一般的仅基于像元光谱信息的参数分类器已难以满足日益增长的分类需求,需要发展高效智能的新分类器来满足遥感数据分类要求。

2. 非参数分类器

非参数分类器不要求待分类数据满足高斯正态分布,而且能方便地将其他辅助分类信息融合到分类中以提高最终的分类精度。研究表明,在对含有比较复杂的地物分布的遥感数据进行分类时,非参数分类器的分类精度要高于参数分类器的分类精度(Foody et al, 2004b)。目前,广泛用于分类的非参数分类器主要有神经网络分类、决策树分类、支持向量机分类和专家系统分类等。

(1)神经网络算法被广泛用于遥感数据分类,国内外学者分别提出并应用了一系列的不同类型的神经网络分类算法,如 BP 神经网络(李祚泳, 1998; Paola et al, 1995)、Hopfield 神经网络(李强 等, 2000)、径向基函数神经网络(骆剑承 等, 2000)和小波神经网络(王耀南, 1999)等遥感数据分类算法,这些分类算法在遥感数据自动分类中得到了广泛的应用,并取得了良好的效果。研究表明,神经网络分类方法的精度要高于其他的传统统计分类方法(Benediktsson, 1990),目前应用和研究最多的是 BP 神经网络分类算法, BP 神经网络的学习训练过程由正反向传播组成,在正向传播中,信息通过输入层输入,然后经隐含层,依次逐层处理,最后由输出层输出结果,如果输出的结果不能满足要求,则进行反向传播,将误差返回,并通过控制各层间的权值,最终达到训练误差最小(Heerman et al, 1992)。神经网络算法是基于经验风险最小化原则的一种学习算法,有其固有的缺陷,如层数和神经元个数难以确定、容易陷入局部极小和过学习现象等,但由于其对噪声数据的鲁棒性及分类精度较高等优点,使得神经网络分类在遥感数据自动分类中得到了广泛的重视和迅速发展。但其在对海量遥感数据进行分类时,其算法效率不太高,需要与其他方法相结合来达到理想效果。如张利等(2009)提出了基于粗糙集和遗传算法的改进 BP 神经网络算法,此改进算法降低了 BP 神经网络的输入层节点数目,在保证分类精度的前提下提高了网络收敛速度,并利用遗传算法优化了 BP 网络的初始权重值。

(2)支持向量机算法通过求解最优化问题,找出高维特征空间中的最优分类超平面,然后利用分类超平面进行数据的分类(Vapnik, 1995)。支持向量机分类算法在高维数据分类、小样本学习和抗噪声等方面具有较大的优势。Camps-Valls 等(2004)分别利用支持向量机分类算法和神经网络分类算法从 HyMAP 数据中提取农作物类别信息,结果表明支持向量机分类算法可以在高维特征空间中直接对遥感数据进行分类,而神经网络无法处理高维数据,必须先要降维处理,且在对

高维数据降维后,支持向量机对不同波段组合数据集的分类结果精度都是很好的,几乎在所有实验都表明支持向量机分类可以得到比神经网络分类精度高的分类结果。基于 Landsat TM 遥感数据,Melgani 等(2004)和 Dixon 等(2008)分别利用了支持向量机分类算法、最大似然和神经网络分类方法提取了土地覆被/利用信息,结果表明支持向量机分类算法获得了最高的分类精度。支持向量机分类算法能获得比其他分类方法更高的分类精度,而且在学习效率、自适应性及处理高维特征空间数据等方面具有明显的优势,是一种优秀的分类算法。但是支持向量机分类算法也存在着一些不可避免的缺点,如核函数选择优化及多类分类策略问题等。在支持向量机分类算法中,核函数最常用的是线性核、小波核、多项式核及径向基核等,其中径向基核用得最广泛。现有的最常用的选择支持向量机核函数的方法是分别采用不同的核函数对支持向量机模型进行训练,然后对比采用不同核函数的分类结果,选择分类精度最高的核函数作为支持向量机分类的最终训练的核函数,同样地,对于核函数中的参数确定方法也是通过取不同的参数值进行训练,选择分类精度最高的参数为核函数的参数。由此可发现,核函数及其参数选择优化都是依赖经验完成的,缺乏一定的理论依据。针对特定的分类问题,如何选择核函数及其参数并没有一个统一的准则,而且不同核函数与分类精度之间的关系如何等问题,都有待深入研究。如 Roli 等(2000)的研究表明采用径向基核的支持向量机分类结果的精度高于采用多项式核的支持向量机分类精度,采用线性核的支持向量机分类精度最低,而在 Huang 等(2002)的研究成果中,采用多项式核的支持向量机分类精度高于采用径向基核的支持向量机分类。在多类分类问题方面,由于支持向量机算法的初衷是求解两类分类的问题,所以不能直接用于多于两类的分类问题,而实际中,遥感数据分类一般都是多于两类的,针对于此,支持向量机多类分类中通常采取两种策略来处理多类分类问题:一是将多类分类问题分解为多个两类分类问题,然后将多个两类分类器组合起来以实现多类分类,如一对一方法(Hsu et al,2002)、一对多方法(Hsu et al,2002)和有向无环图(Platt et al,2000)等;另一种解决方法是在优化公式中直接求解多类分类问题,通过优化求解多个分类面的参数,一次性地解决多类分类问题。针对不同的分类对象,如何合理选择多类分类策略是支持向量机多类分类中的一个重要研究内容。此外,针对支持向量机分类算法本身的缺点,许多学者对支持向量机分类算法进行改进,如 Bruzzone 等(2006)基于传统支持向量机分类算法提出了直推式支持向量机算法,并利用其对 TM 影像进行了分类研究,发现直推式支持向量机的结果更稳定。Demir 等(2007)和 Foody(2008)使用相关向量机算法对高光谱遥感影像进行了分类研究,得出相关向量机可以大大缩短分类时间,提高分类效率。Zhang 等(2008)提出了带先验知识的支持向量机算法并用于 ASTER 和 ADS40 影像数据的分类,实验结果表明 P-SVM 在时间和精度上都要优于传统的支持向量机分类算法。这些改进

的支持向量机分类算法主要通过以下三个方面来提高传统支持向量机分类算法的性能:①将传统的支持向量机分类算法拓展成一个半监督支持向量机分类算法,以达到提高分类精度和减少训练样本的目的;②构造一个全新的目标函数和约束条件,从而产生一个新的支持向量机分类算法;③将传统的支持向量机分类算法与其他算法在理论层次进行结合。

(3)决策树分类通过训练样本学习后得到一系列的判别函数,然后在不同的判别函数中取不同阈值建立相应的分类树分支,在每个分类树分支的基础上再依次建立结点及分类树的分支,最后形成一个完整的分类决策树。常用的决策树分类算法主要有迭代二叉树(iterative dichotomiser, ID)(Quinlan, 1986)、C4.5 和建树和修剪相结合的分类(pruning and building integrated in classification, PUBLIC)算法(Friedl et al, 1997, 1999)等。决策树分类算法具有计算效率高、能处理多尺度数据、易与其他辅助分类知识融合及分类精度较高等优点,在遥感数据分类领域中应用广泛(Hansen et al, 1996)。Belward 等(1987)分别利用决策树和最大似然分类方法进行了遥感数据农作物分类,实验结果表明决策树分类的分类效率要远高于最大似然法。Wardlow 等(2008)基于 MODIS NDVI 数据,采用决策树分类算法获取了农作物分类信息,且取得了较高的分类精度。Schneider 等(2001)基于城市 MODIS 数据,利用决策树分类算法获取了城市地物分类信息,也取得了良好的效果。但决策树分类也有其缺点,如算法复杂度及获取判断函数需要大量的训练样本等,所以针对小样本数据、特征简单数据集的分类,决策树分类精度并不一定比传统分类方法高。针对决策树分类中存在的缺点,很多学者对其进行了改进,如将粗糙集引入决策树分类中(吴成东 等, 2006; 乔梅 等, 2005; Asit et al, 2011)用来处理决策树分类过程中的不确定性信息。虽然决策树分类在遥感数据分类中得到了广泛的应用,并取得了较好的分类效果,但随着应用需要的变化,仍需进行深入研究和改进,以提高决策树分类算法的分类精度及应用范围、提高决策树构造方法的效率及发展新的决策树分类算法等。

(4)专家系统分类。由于遥感影像数据一般具有空间特性,在进行分类时,应该充分利用这些空间特性来提高分类结果的精度。而普通统计方法、神经网络及支持向量机等分类方法缺乏空间特征信息的支持,难以反映一些特殊类型的空间分布。因此,一些学者提出了遥感数据分类专家系统,尝试利用计算机模拟地学专家对遥感数据进行综合解译与决策分析。遥感数据分类专家系统的基本原理是模拟人类利用各种知识进行推理并得出结果(Stefanov et al, 2001),包括推理机和知识库,知识库是求解问题所需要的知识集合,推理机是根据知识的语义,按一定策略找到知识规则并执行解释。在遥感数据分类过程中,将遥感影像光谱和相关空间特征数据输入推理机中,然后根据已建立的知识规则库对输入的数据进行判断分类,最终将像元归入相应的预设类别(贾海峰 等, 2006)。遥感数据专家系统分