

基于最近邻的遥感影像单类信息提取

薄树奎¹ 荆永菊²

(郑州航空工业管理学院计算机科学与技术系 郑州 450015)¹

(郑州航空工业管理学院图书馆 郑州 450015)²

摘 要 遥感影像单类信息提取是一种特殊的分类,旨在训练和提取单一兴趣类别。研究了基于最近邻分类器的单类信息提取方法,包括类别划分和样本选择问题。首先分析论证了最近邻方法提取单类信息只与所选择的样本相关,而与类别划分无关,因此可以将单类信息提取作为二类分类问题进行处理。然后在二类分类问题中,根据空间和特征邻近性选择非兴趣类别的部分训练样本,简化了分类过程。实验结果表明,所提出的方法可以有效实现遥感影像单类信息的提取。

关键词 单类,信息提取,遥感,最近邻

中图法分类号 TP391 文献标识码 A

One-class Information Extraction from Remote Sensing Imagery Based on Nearest Neighbor Rule

BO Shu-kui¹ JING Yong-ju²

(Department of Computer Science and Application, Zhengzhou University of Aeronautics, Zhengzhou 450015, China)¹

(Library, Zhengzhou University of Aeronautics, Zhengzhou 450015, China)²

Abstract One-class extraction from remote sensing imagery is a special method of classification, where users are only interested in recognizing one specific land type. The extraction of a specific class was studied based on nearest neighbor rule in this paper. Two aspects were considered, class partitioning and sample selection for each class. Firstly, the effect of data distribution partitioning is analyzed theoretically based on nearest neighbor in one-class classification. It is confirmed that the nearest neighbor classifier requires the data distribution to be partitioned into only two classes, namely the class of interest and the remainder. Secondly, as a two-class problem, the classification process was simplified, and the sample selection in nearest neighbor classification was performed in terms of both the spatial and the feature space. The experiments show that the specific class of interest can be well extracted from the remote sensing image with the proposed method.

Keywords One-class, Information extraction, Remote sensing, Nearest neighbor

1 引言

遥感影像单类信息提取仅对其中的某个特定类别感兴趣,并将感兴趣的类别提取出来,在遥感影像处理中有重要应用^[1-5]。单类信息提取感兴趣的是一个特定类别的提取精度,而不用关心影像中其他的地物类型。单类信息提取可以通过多类分类器或者单类分类器来实现,采用多类分类器提取单类信息,将影像数据划分为多个地物类别,同时获取兴趣类别信息。例如为了能准确地提取水田信息,需要将原影像分成水田与林地、旱地、水体和城镇用地等不同的类别^[3]。而采用单类分类器进行单类信息提取,在不了解非兴趣类别的情况下,只需兴趣类别样本进行分类器的训练,并提取兴趣类别。例如 OCSVM(one-class support-vector machine, OCSVM)方法^[6]、最大熵方法^[7],以及 PUL(Positive and Unlabeled Learning)方法^[8]等。单类分类器可以很好地适应非兴趣类别样本难以获取的情况,而在遥感影像中非兴趣类别样本一

般是可以获取或者能部分获取的,因此,采用多类分类器进行遥感影像单类信息提取仍然是主要的方法。

虽然各种多类分类方法都可以用来提取感兴趣的单类别信息,但以往方法一般都是将原影像进行完全的划分,需要准确掌握影像中的各个地物类型的特征。而在单类信息提取中,一般对兴趣类别了解得比较充分,而对于其他(非兴趣)类别可能并不熟悉。例如,在全国范围的湿地提取与制图工作中^[1],对湿地类别了解比较充分,但是在大范围区域的影像中,湿地以外的非兴趣类别可能多种多样,而且各个地区有所区别,不易掌握。因此,要准确掌握所有类别的特征和划分情况,可能花费大量样本采集的时间和精力。对于同一遥感数据来说,兴趣类别不变,由于遥感影像信息的复杂性,其他非兴趣类别的划分有时难以确定,因此采用多类分类器进行单类信息提取,如何划分非兴趣类别,并有效提取单类信息是一个关键问题。本文基于最近邻分类器研究遥感影像单类信息问题中的类别划分及其样本选择问题。

本文受国家自然科学基金(41001235),河南省高等学校青年骨干教师资助计划(2012GGJS-145)资助。

薄树奎(1976—),男,博士,副教授,主要研究方向为图像信息提取、模式识别, E-mail: bsk586@163.com; 荆永菊(1978—),女,主要研究方向为信息处理。

2 最近邻方法提取单类信息的类别划分

遥感影像单类信息提取只需要提取一个兴趣类别,因此涉及到如何划分类别空间的问题。以单类提取的错误率为标准,首先分析类别划分对单类信息提取的影响。

以最近邻方法提取单类信息,假设有来自 M 个类别的 N 个样本(每个样本的类别已知),从这 N 个样本中找出与给定样本距离最近的一个,称为最近邻,那么最近邻所属的类别就是未知样本的判别类别。给定未知样本 X ,其最近邻为 $X_n \in \omega_i$,假定 $X \in \omega_j$,如果 ω_j 不等于 ω_i ,则发生一次错误,不失一般性,假定兴趣类别为 ω_1 ,因此在给定 X 和 X_n 时的条件错误概率为:

$$\begin{aligned} r(X, X_n) &= \Pr\{\omega_i \neq \omega_j / X_n, X\} \\ &= P(\omega_1 / X_n) P(\omega_2 / X) + P(\omega_2 / X_n) P(\omega_1 / X) + \\ &\quad \cdots + P(\omega_1 / X_n) P(\omega_M / X) + P(\omega_M / X_n) P(\omega_1 / X) \end{aligned} \quad (1)$$

若取 N 值较大,使 X 及其最近邻 X_n 在位置上非常接近,则有以下近似关系:

$$P(\omega_i / X_n) \cong P(\omega_i / X) \quad (2)$$

于是,式(1)变成

$$\begin{aligned} r(X, X_n) &\cong \Pr\{\omega_i \neq \omega_j / X_n, X\} \\ &= 2P(\omega_1 / X) P(\omega_2 / X) + \cdots + 2P(\omega_1 / X) P(\omega_M / X) \\ &= 2P(\omega_1 / X) \sum_{j=2}^M P(\omega_j / X) \\ &= 2P(\omega_1 / X) (1 - P(\omega_1 / X)) \end{aligned} \quad (3)$$

容易证明,当 $N \rightarrow \infty$ 时,式(1)以概率 1 收敛于式(3)。

由式(3)可见,最近邻方法的错误概率只与兴趣类别(ω_1)的条件概率相关,而与其他类别如何划分无关。因此,采用最近邻方法进行单类信息提取时,只需将类别划分成两类,即兴趣类别和非兴趣类别。

3 最近邻方法提取单类信息的样本选择

在遥感分类中,训练样本选择是一个耗时费力的工作,多类分类中要选择每个类别的代表点,而在二类分类中,也可以选择类别之间的边界点,基于此边界进行类别的划分。由前面分析可知,基于最近邻的单类信息提取精度与类别划分无关,因此可以将单类信息提取看做二类分类问题,即兴趣类别和非兴趣类别。在遥感影像单类信息提取应用中,一般对兴趣类别了解充分,训练样本选择也较容易。而由于对研究区域熟悉程度的限制,可能对非兴趣类别了解并不充分,甚至不能确定其中的类别划分情况。在单类信息提取中,我们只关心兴趣类别的提取,不考虑其他的类别,这也是单类提取与多类分类的主要区别。如果在特征空间中能找到一个边界,将兴趣类别和非兴趣类别区分开来,则不需要对遥感影像进行完全的类别划分和提取。

在单类信息提取训练样本选择中,首先选择兴趣类别的代表点,而对非兴趣类别,则选择与兴趣类别邻近的部分样本进行训练,目的是找到兴趣类别与非兴趣类别的边界,通过边界来区分兴趣类别与非兴趣类别。由于遥感影像的空间临近效应,相互邻近的像素特征往往相近,因此选择兴趣类别样本的同时,在图像空间中,选择与兴趣类别邻近的非兴趣类别样本,而不需要对非兴趣类别代表点进行完全选择,使样本选择过程变得简单。此外,在实际应用中,将图像空间和特征空间的邻近性结合起来进行训练样本的选择,在无需了解影像中

所有类别划分的情况下,实现感兴趣类别的有效提取。

为了提高分类精度采用面向对象影像分析方法,面向对象方法最基本的特点就是以图像分割后获得的图像对象为基本的操作单元。图像对象的生成综合考虑了邻近像素的光谱和空间信息,将邻近像素合并成匀质的斑块,克服了基于像素方法的不足。通过图像分割获得的图像对象具有一定的属性,包括光谱、形状等。面向对象方法中,训练样本的选择也是基于影像对象的。在图像空间中,选择相互邻近的兴趣类别和非兴趣类别图像对象,提高了学习效率和可靠性。在特征空间中,由于分割后形成的图像对象都有属性值,根据属性值选择与兴趣类别相近的非兴趣类别图像对象作为训练样本。两种方法都不考虑影像中的类别划分情况,只根据空间和特征邻近性进行样本选择,二者结合起来实现兴趣类别和非兴趣类别的有效区分。

4 实验结果

实验图像为两幅 TM 遥感影像,如图 1 所示,其中图 1(a)包含不透水表面、草地、林地和水体 4 个类别,图 1(b)包含不透水表面、水体、耕地和村庄 4 个类别。

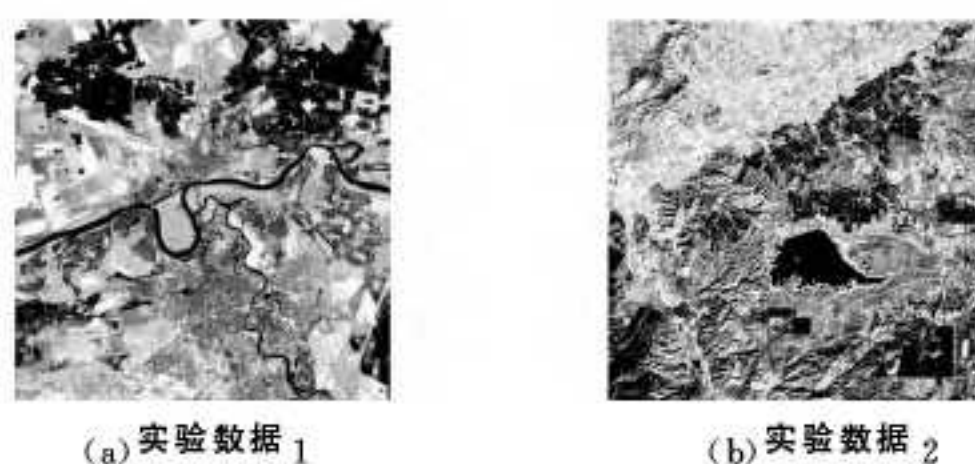


图 1 TM 遥感影像

采用最近邻方法提取单类信息,在兴趣类别样本的周围选择非兴趣类别样本,这种方法减少了样本选择的工作量。对图 1(a)中的 TM 实验影像,以不透水面为兴趣类别,训练样本的选择如图 2(a)所示,其中白色区域为兴趣类别样本,周围的黑色区域为非兴趣类别样本,图 2(b)为分类结果。单类信息提取的精度评价指标一般为生产精度和用户精度,生产精度指兴趣类别的正确分类数占参考数据中该类别象素总数的比例,而用户精度指兴趣类别正确分类数占分为该类象素总数的比例。实验数据 1 中不透水面单类提取的生产精度和用户精度分别为 85.3% 和 90.9%。

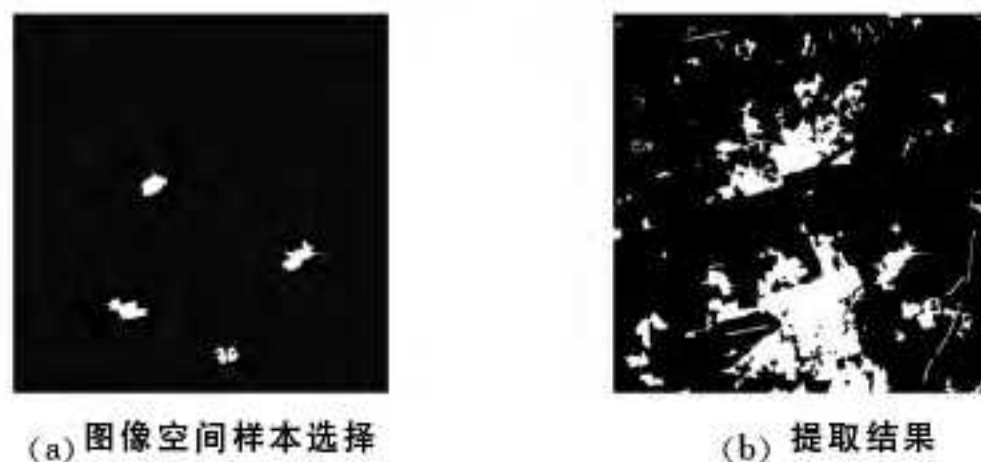


图 2 不透水表面提取

而对于有些情况,这种训练样本的选择方法不能很好地将特征相近的兴趣类别与非兴趣类别分开。例如在图 1(b)所示的实验影像中提取水体类别,图 3(a)为水体类别及其周围选择的训练样本,白色区域为水体,黑色区域为其他类别。图 3(b)为分类结果,可以看出,由于阴影的特征与水体相似,有些阴影区域被错分为水体。

(下转第 259 页)

2005). New York: ACM Press, 2005; 276-283

- [2] Fall K. A delay-tolerant network architecture for challenged internets [C]//Proc of Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications. New York: ACM Press, 2003; 27-34
- [3] Wang Y, Wu H Y. Delay/Fault-Tolerant mobile sensor network (DFT-MSN): A new paradigm for pervasive information gathering[J]. IEEE Tran. on Mobile Computing, 2006, 6(8): 1021-1034
- [4] Shah R C, Roy S, Jain S, et al. Data Mules: Modeling a three-tier architecture for sparse sensor networks [C]//Ekici E, ed. Proc. of the 1st Int'l Workshop on Sensor Network Protocols and Applications. Anchorage: IEEE Computer Society Press, 2003; 30-41
- [5] Zhao W, Ammar M, Zegura E. A message ferrying approach for data delivery in sparse mobile Ad hoc networks[C]//Proc. of the ACM Mobihoc 2004. Roppongi: ACM Press, 2004; 187-198
- [6] Vahdat A, Becker D. Epidemic routing for partially connected ad hoc networks; Technical Report, CS-200006[R]. Durham: Duke University, 2000
- [7] Lindgren A, Doria A, Schelén O. Probabilistic routing in intermittently connected networks[J]. SIGMOBILE Mobile Computing Communications Review, 2003, 7(3): 19-20
- [8] Wang Y, Wu H Y. Replication-Based efficient data delivery scheme (RED) for delay/fault-tolerant mobile sensor network (DFT-MSN)[C]//Gregori E, ed. Proc. of the 4th Annual IEEE

Int'l Conf. on Pervasive Computing and Communications Workshops. Washington: IEEE Computer Society Press, 2006; 485-489

- [9] Wang Y, Wu H Y, Dang H, et al. Analytic, simulation, and empirical evaluation of delay/fault-tolerant mobile sensor networks [J]. IEEE Trans. on Wireless Communications, 2007, 1(11): 3287-3296
- [10] 朱金奇, 刘明, 龚海刚, 等. 迟容忍移动传感器网络中基于选择复制的数据传输[J]. 软件学报, 2009, 20(8): 2227-2240
- [11] Xu Fu-long, Liu Ming, Cao Jian-nong, et al. A motion tendency-based adaptive data delivery scheme for delay tolerant mobile sensor networks[C]//The 51st Annual IEEE Global Telecommunications Conference(GLOBECOM'09). Hawaii, USA, 2009; 1-5
- [12] Spyropoulos T, Psounis K, Raghavendra CS. Efficient routing in intermittently connected mobile networks; The single-copy case [J]. IEEE/ACM Trans. on Network, 2008, 16(1): 63-76
- [13] Spyropoulos T, Psounis K, Raghavendra C S. Efficient routing in intermittently connected mobile networks; The multiple-copy case[J]. IEEE/ACM Trans. on Network, 2008, 16(1): 77-90
- [14] Johnson DMaltz D. Dynamic source routing in ad hoc wireless networks [M]//Imelinsky T, Korth H, eds. Mobile Computing. Kluwer Academic Publishers, 1996; 153-181
- [15] 杨奎武, 郑康锋, 杨义先, 等. 基于运动状态的延迟容忍移动传感器网络数据传输策略[J]. 通信学报, 2010, 31(11): 138-146

(上接第 218 页)

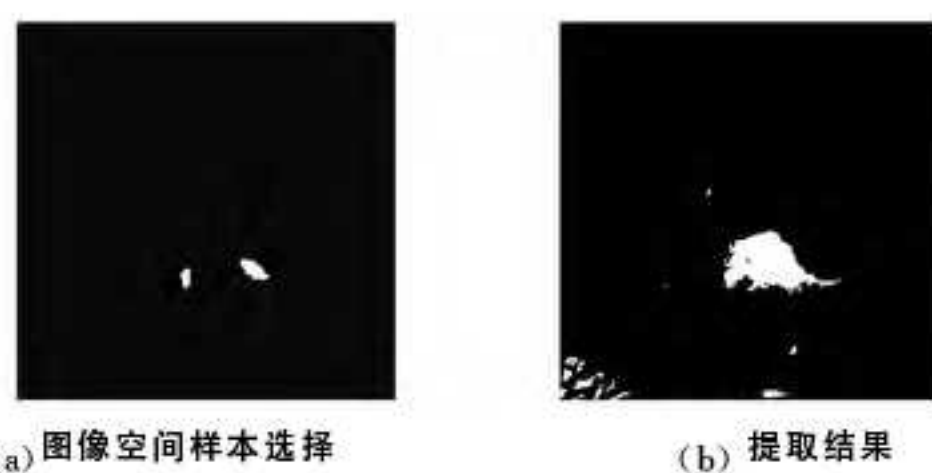


图3 水体提取

为了避免基于图像空间的样本选择产生特征相近目标的错分,将图像空间和特征空间结合起来。由于在面向对象方法中,图像分割后每个对象都有各自的属性特征值,从而可以选择与兴趣类别光谱特征相似的非兴趣类别训练样本。如图4(a)所示,增加图像下方的阴影区域作为非兴趣类别样本,图4(b)为分类结果。可以看出,这种特征空间和图像空间样本选择相结合的方法可以克服单一方法的不足,错分区域明显减少,本实验中水体的单类提取生产精度和用户精度分别为99.0%和98.0%,分类效果好。



图4 水体提取

结束语 本文基于最近邻方法,研究了遥感影像单类信息提取问题。根据最近邻方法的特点,将类别划分为兴趣类

别和非兴趣类别即可,并根据图像空间和特征空间邻近性选择非兴趣类别的部分训练样本,简化了分类过程。在实验中对两幅 TM 遥感影像进行了单类信息提取,结果表明该方法是有用的。遥感影像单类信息提取与多类分类问题既有区别也有联系,随着研究的不断深入,单类信息提取方法会更加完善并得到广泛应用。

参考文献

- [1] 宫鹏,牛振国,程晓,等. 中国 1990 和 2000 基准年湿地变化遥感[J]. 中国科学 D 辑:地球科学,2010,40(6):768-775
- [2] Tooke T R, Coops N C, Goodwin N R, et al. Extracting Urban Vegetation Characteristics Using Spectral Mixture Analysis and Decision Tree Classifications[J]. Remote Sensing of Environment, 2009, 113(2): 398-407
- [3] 张友水,原立峰,姚永慧. 多时相 MODIS 影像水田信息提取研究[J]. 遥感学报,2007,11(2):282-288
- [4] 徐涵秋. 利用改进的归一化差异水体指数(MNDWI)提取水体信息的研究[J]. 遥感学报,2005,9(5):589-595
- [5] 闫柏琨,王润生,甘甫平,等. 热红外遥感岩矿信息提取研究进展[J]. 地球科学进展,2005,20(10):1116-1126
- [6] Schölkopf B, Platt J C, Shawe-Taylor J, et al. Estimating the Support of a High-dimensional Distribution [J]. Neural Computation, 2001, 13(7): 1443-1471
- [7] Li W K, Guo Q H. A Maximum Entropy Approach to One-class Classification of Remote Sensing Imagery [J]. International Journal of Remote Sensing, 2010, 31(8): 2227-2235
- [8] Li W K, Guo Q H, Elkan C. A Positive and Unlabeled Learning Algorithm for One-class Classification of Remote-sensing Data [J]. IEEE Transactions on Geoscience & Remote Sensing, 2011, 49(2): 717-725