

文章编号: 1004 - 4574(2011)06 - 0081 - 06

Kappa 系数在干旱预测模型精度评价中的应用

——以关中平原的干旱预测为例

许文宁¹, 王鹏新¹, 韩萍², 严泰来¹, 张树誉³

(1. 中国农业大学信息与电气工程学院, 北京 100083; 2. 中国农业大学理学院, 北京 100094;
3. 陕西省农业遥感信息中心, 陕西 西安 710015)

摘要: Kappa 系数较多地用在评价遥感影像分类精度中。通过对分类影像和参考影像逐个像元统计, 并建立误差矩阵, 可以较准确地验证遥感影像分类的精度。将 Kappa 系数引入到关中平原地区加权马尔可夫和自回归移动平均两种干旱预测模型的精度评价中, 基于标准降水化指数和条件温度植被指数两种干旱指标, 对干旱监测数据和模型预测数据建立误差矩阵, 得到了错估误差、漏估误差、总体精度和 Kappa 系数。综合应用 4 种评价指标分析模型的预测结果表明, 错估误差和漏估误差能够验证预测模型的局部适用性, 总体精度在一定范围内不能够直接反映模型的预测精度, Kappa 系数可以较精确地评价不同时空间尺度的干旱预测模型精度。当参与预测的样本数目增加到一定程度时, Kappa 系数和总体精度基本相等, 可以更准确地评价模型预测精度。

关键词: 精度评价; Kappa 系数; 误差矩阵; 干旱预测

中图分类号: TP79

文献标志码: A

Application of Kappa coefficient to accuracy assessments of drought forecasting model: a case study of Guanzhong Plain

XU Wen-ning¹, WANG Peng-xin¹, HAN Ping², YAN Tai-lai¹, ZHANG Shu-yu³

(1. College of Information and Electrical Engineering, China Agricultural University, Beijing 100083, China; 2. College of Science, China Agricultural University, Beijing 100094, China; 3. Remote Sensing Information Center for Agriculture of Shaanxi Province, Xi'an 710015, China)

Abstract: Kappa coefficient is widely used to assess the accuracy of classification on remote sensing images. The assessment is generally done through statistics of error matrix among the classification types on the images. In this paper, Kappa coefficient was used to assess accuracy of drought forecast for Guanzhong Plain, using an approach to integrate the coefficient with the weighted Markov model and the ARIMA model. Commission error, omission error, overall accuracy and Kappa coefficient based on SPI and VTCI were obtained by establishing error matrix between original data and forecast data. Above all, commission error and omission error could show the local applicability of the two models, overall accuracy could not figure the truly accuracy of the models at different spatial and temporal scales but Kappa coefficient could. However, when the number of the forecasting samples increased to a certain ex-

收稿日期: 2010 - 01 - 17; 修回日期: 2010 - 10 - 24

基金项目: 国家自然科学基金资助项目(40871159, A1071235, A0571111); 国家高技术研究发展计划课题(2007AA12Z139); 高等学校博士学科点专项科研基金项目(20100008110031)

作者简介: 许文宁(1983 -), 女, 博士生, 主要从事定量遥感、图像处理等研究。E-mail: xuwenning@sohu.com

通讯作者: 王鹏新(1965 -), 男, 教授, 主要从事定量遥感及其在农业中的应用研究。E-mail: wangpx@cau.edu.cn

tent, overall accuracy could be the same as Kappa coefficient, it could be used to assess drought forecasting accuracy.

Key words: accuracy assessment; Kappa coefficient; error matrix; drought forecast

基于误差矩阵的 Kappa 系数精度评价方法是一种多变量分析方法,能够在统计意义上反映分类结果在多大程度上优越于随机赋予各点某一类型的分类结果。有很多学者在 Kappa 系数的算法和应用方面做了大量工作^[1-4],使其逐渐发展成为比较成熟的遥感分类精度评价方法。比如 Pontius 等人在 Kappa 系数的基础上又进一步发展可以量化数量错误和位置错误的改进系数:标准 Kappa 系数、随机 Kappa 系数、位置 Kappa 系数和数量 Kappa 系数等。

目前 Kappa 系数因其良好的适用性而被广泛应用在医学、教学、人口统计中。布仁仓等^[5]以辽宁中部城市群为例,用 Kappa 系数全面的揭示景观空间信息的变化,与传统方法相比,它们既能说明数量变化,又能说明位置和综合信息的变换;闫岩等^[6]给出了计算疾病分型不一致时的 Kappa 系数计算方法,能够较客观的用于评价新的医疗诊断方案在临床中的应用;韩婷等^[7]选取第二军医大学学生和教学督导组对全校教师的评分结果为样本,利用 Kappa 系数比较分析学生与教学督导组对教师教学质量评估的一致性,全面加强了对教学质量的监控。

本文以干旱预测模型为例,将 Kappa 系数引入到干旱预测模型的精度评价中,通过探讨 Kappa 系数作用和机理,以期获得更加客观可靠的精度评价方法,发展和丰富预测模型的精度评价体系。

1 研究方法和数据处理

1.1 Kappa 系数的机理和计算过程

Kappa 系数是由 Cohen 在 1960 年提出的用于评价遥感影像分类结果的一致性检验方法^[8-9]。通过对地面调查的地物类型数据和影像分类结果建立误差矩阵,其中矩阵列表示地面调查的地物类型,矩阵行表示影像数据的分类类型,每一个元素表示像元的数量。Kappa 系数计算公式为:

$$k = \frac{P_0 - P_e}{(1 - P_e)}, \quad (1)$$

$$P_0 = \frac{\sum_{i=1}^n P_{ii}}{N}. \quad (2)$$

式中: P_0 是分类的总体精度(overall accuracy),表示对每一个随机样本,分类结果与地面调查数据类型一致的概率; P_e 表示由于偶然机会造成的分类结果与地面调查数据类型相一致的概率; n 为分类的类型数量; N 为样本总数; P_{ii} 为第 i 类型的被正确分类的样本数目。当分类结果与实际类型完全吻合时, Kappa 系数的值即为 1。

错分误差(commission error) 是指不该属于某一类型的像元被分为该类型的误差,漏分误差(omission error) 是指属于某一类型的像元未被分为该类型的误差。

$$P_{c_i} = 1 - \frac{P_{ii}}{P_{i+}}, \quad (3)$$

$$P_{o_i} = 1 - \frac{P_{ii}}{P_{+i}}. \quad (4)$$

式中: P_{c_i} 是第 i 类型的错分误差, P_{o_i} 是第 i 类型的漏分误差, P_{i+} 是第 i 类型所在列的像元数目之和, P_{+i} 是第 i 类型所在的行的像元数目之和。通过以上各种描述性的精度值, Kappa 系数计算方法可以表示为

$$k = \frac{P_0 - P_e}{1 - P_e} = \frac{\frac{\sum_{i=1}^n P_{ii}}{N} - \frac{\sum_{i=1}^n (P_{i+} \times P_{+i})}{N^2}}{1 - \frac{\sum_{i=1}^n (P_{i+} \times P_{+i})}{N^2}}. \quad (5)$$

在分类的精度评价中,不同的精度评价方法有不同的划分标准和含义。本研究主要是探讨 Kappa 系数在预测精度评价中的作用,因此直接借用 Cohen 提出的 Kappa 系数分类评价标准^[1-2](表 1)。

表 1 Kappa 系数分类标准
Table 1 Classification criterion based on Kappa coefficient

Kappa	< 0.00	0.00 ~ 0.20	0.21 ~ 0.40	0.41 ~ 0.60	0.61 ~ 0.80	0.81 ~ 1.00
一致性程度	很差	微弱	弱	适中	显著	最佳

1.2 干旱预测模型和数据处理

研究采用的标准化降水指数(standardized precipitation index, SPI) 时间序列以关中平原 36 个气象站 39a (1965 - 2004 年) 逐月降水量为数据源,计算 5 个时间尺度(1 个月、3 个月、6 个月、9 个月和 12 个月)的 SPI 值。SPI 采用 Γ 分布概率来描述降水量的偏态分布变换,再计算给定时间尺度的累积概率,然后转化成标准正态分布而得,这样能够消除降水量在时空分布上的差异,使得 SPI 能够适用于反映不同地区、不同时间尺度的旱涝情况;条件植被温度指数(vegetation temperature condition index, VTCI) 时空变化序列以关中平原 1999 - 2004 年 NOAA - AVHRR 遥感影像为数据源,使用最大最小值合成算法得到 VTCI 干旱监测结果^[10-11],VTCI 可在理论上揭示为 NDVI 值相等时, LST 差异大小的比率, VTCI 的值越小,干旱程度越严重。

本研究应用 Kappa 系数对选取的 4 个干旱预测模型进行精度评价。模型 1 和模型 2 基于关中平原 1966 - 2003 年的 5 个时间尺度 SPI 序列建立加权马尔可夫^[12]和自回归移动平均(autoregressive integrated moving average, ARIMA)^[13]干旱预测模型,预测得到 2004 年 12 个月的干旱状态。模型 3 和模型 4 基于关中平原 1999 - 2004 年的 VTCI 监测数据建立加权马尔可夫和自回归移动平均干旱预测模型,预测得到 2005 - 2006 年 VTCI 干旱结果。加权马尔可夫预测的基本思想是将 SPI 和 VTCI 看作一系列相依的随机变量,其各阶自相关系数刻画了各种滞时状态间相关关系的强弱,分别依其以前若干时段指数值的状态进行预测,然后按前面各时段与该时段相依关系的强弱进行加权求和,最后根据各状态加权求和的结果来预测该时段可能所处的状态;自回归移动平均预测主要是针对平稳时间序列的 SPI 和 VTCI 建模,对于含有趋势性的非平稳时间序列,需先进行逐期差分运算转化为平稳序列,对于含有季节性变化的非平稳时间序列,需先进行季节性差分运算然后建立 ARIMA 模型,预测结果不仅与以前时刻的序列值相关,而且与以前时刻进入预测系统的随机扰动相关。

在获得 SPI 和 VTCI 的实际监测结果及模型预测结果后,将“实际监测结果”与遥感影像分类中的“实地调查的地物类型”相对应,“预测结果”与“分类类型”相对应,“模型的总体精度”与“分类的总体精度”相对应,建立相应的误差矩阵,通过计算 Kappa 系数评价模型的预测精度,同时得到各个等级的错估(分)误差、漏估(分)误差等各项精度评价指标。

4 个预测模型使用了两种不同时空尺度的数据源,两种不同物理意义的物理模型,获得了不同空间尺度、不同时间尺度的预测精度指标。VTCI 有两点不同于 SPI。首先,它是遥感影像“面”上的干旱监测结果,而 SPI 计算的是气象观测“点”上的干旱监测结果,两者所代表的空间尺度和时间尺度均不相同;其次, SPI 的旱情等级划分明确,而 VTCI 的旱情等级划分尚处在研究中,尚需进一步验证。

2 结果与分析

2.1 基于 SPI 序列的预测精度评价

根据《气象干旱等级》国家标准对 SPI 的划分标准,将干旱类型划分为无旱、轻旱、中旱、重旱和特旱 5 种类型。通过对气象站附近的 SPI 干旱监测结果和加权 Markov, ARIMA 模型预测 2004 年 12 个月的干旱结果建立误差矩阵,获得 Kappa 系数、总体精度、错估误差和漏估误差等精度评价指标如图 1。

从错估误差和漏估误差分析五种干旱等级的两种误差统计数据可以验证干旱预测模型的局部适用性。图 1(a) 和 (c) 显示加权 Markov 模型无旱等级的两种误差小于 0.26,并且随时间尺度增长误差越来越小,9 个月以上尺度重旱和特旱等级的两种误差为 0,表明该预测模型对无旱等级的预测效果较好,适合长时间尺度的干旱预测;图 1(b) 和 (d) 显示 ARIMA 模型对无旱以外其他干旱等级的错估误差为 0 而漏估误差为 1,说明该模型对

无旱以外其他干旱等级的预测能力较差。

对比两个模型的误差统计数据,加权 Markov 和 ARIMA 模型对无旱等级的预测能力较强,对其它干旱等级预测能力较差,基于同样的 SPI 序列,ARIMA 模型的错估误差和漏估误差变化较大,说明模型的稳定性较差。

从总体精度和 Kappa 系数来分析,图 1(e) 显示加权 Markov 模型在 9 个月尺度总体精度 0.85, Kappa 系数 0.58,一致性程度显著,表明模型在该时间尺度预测效果较理想;模型在 1 个月尺度总体精度 0.74, Kappa 系数 0,一致性程度微弱,表明模型在该时间尺度预测效果不理想。随着时间尺度增大,模型的总体精度和 Kappa 系数不断增大,在 9 个月尺度时 Kappa 系数达到最大值,虽然模型的总体精度没有达到最大值,但是预测结果和监测数据的一致性程度达到最好。图 1(f) 显示 ARIMA 模型对 5 个干旱等级的总体精度在 0.66~0.88, Kappa 系数基本为零,一致性程度微弱,表明该模型的预测精度非常低。这是因为关中大部分气象站点属于无旱等级,模型对无旱等级预测能力较强,所以总体精度较高,但是模型在其他干旱等级漏估误差及基本为 1, Kappa 系数机理分析表明模型预测精度较低。因此,总体精度较高并不能直接说明模型预测精度较高^[14], Kappa 系数能够从整体上验证预测结果和监测数据的一致性程度,只有 Kappa 系数一致性程度达到适中以上时,才能够证明模型预测结果较理想。

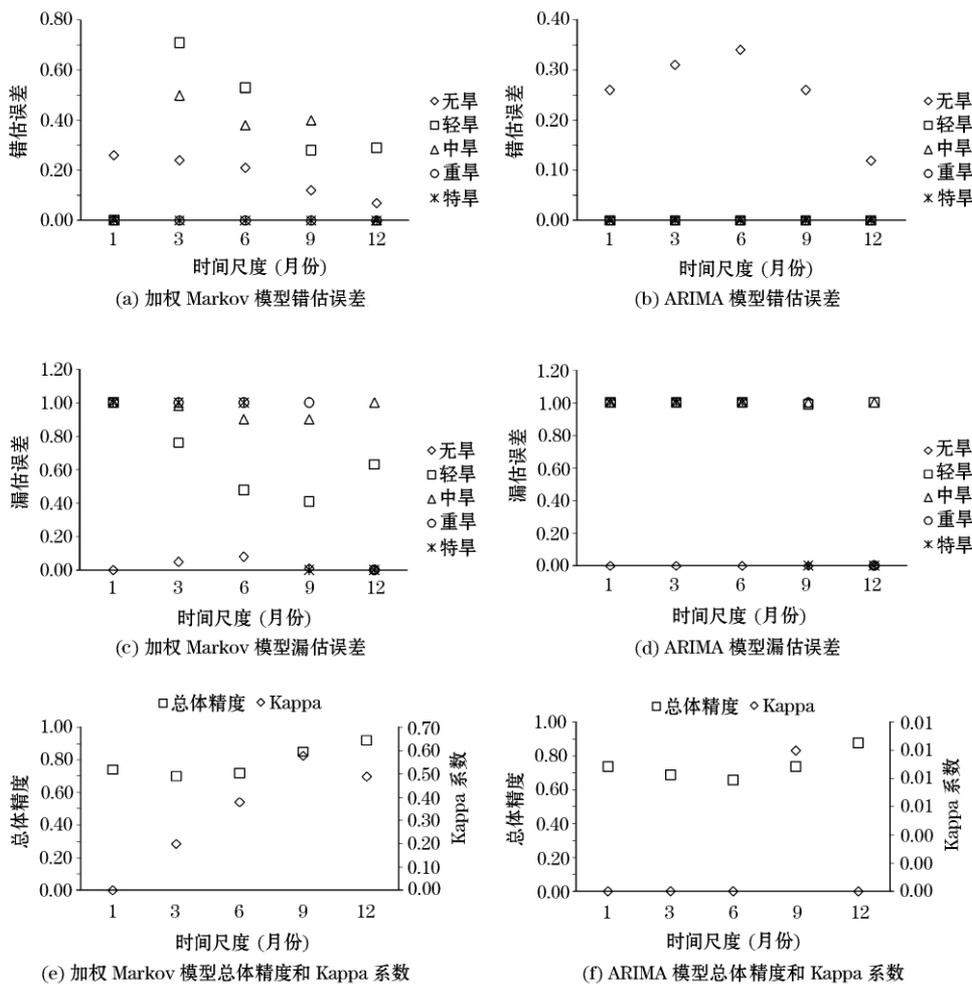


图 1 基于 SPI 两种模型的精度评价

Fig. 1 Accuracy estimation of two models based on SPI

加权 Markov 模型和 ARIMA 模型的总体精度较高,但仅从这一点不能对比两个模型的预测精度, Kappa 系数统计分析表明,加权 Markov 模型的预测结果和监测数据的一致性程度明显高于 ARIMA 模型,前者预测精度较高,更适合基于 SPI 序列的干旱预测建模。

2.2 基于 VTCI 序列的预测精度评价

根据 VTCI 的干旱监测结果,将 VTCI 划分等级划分为 5 个等级(0.00~0.20, 0.21~0.40, 0.41~0.60, 0.61~0.80, 0.81~1.00)进行精度评价。通过对 1999-2004 年关中地区的 VTCI 序列建模,以旬为单位步长预测 1 步和 2 步,应用加权 Markov 模型和 ARIMA 模型得到 2005-2006 年关中地区的 20 幅预测 VTCI 序列。逐像元统计监测 VTCI 序列和预测 VTCI 序列并建立误差矩阵,得到 Kappa 系数、总体精度、错估误差和漏估误差等精度评价指标,其中总体精度和 Kappa 系数的变化范围、平均值和标准差如表 2。

表 2 基于 VTCI 两种模型的精度评价

Table 2 Accuracy estimation of the two models based on VTCI

	总体精度				Kappa 系数			
	最小值	最大值	平均值	标准差	最小值	最大值	平均值	标准差
Markov 1 步预测	0.12	0.73	0.37	0.17	-0.04	0.18	0.03	0.06
Markov 2 步预测	0.05	0.67	0.35	0.20	-0.04	0.09	0.02	0.05
ARIMA 1 步预测	0.07	0.83	0.46	0.21	-0.12	0.28	0.04	0.11
ARIMA 2 步预测	0.01	0.82	0.38	0.26	-0.10	0.21	0.03	0.08

以 1 步预测为例分析 2 个模型的预测精度,表 2 统计显示,加权 Markov 模型的总体精度 0.37, Kappa 系数 0.03,一致性程度微弱,表明该模型预测精度较低; ARIMA 模型总体精度 0.46, Kappa 系数 0.04,一致性程度微弱,表明该模型预测结果也不好。仅从总体精度分析,两个模型预测精度较低的原因是在逐像元建立误差矩阵的过程中,单个像素代表的地物差别较大;此外,根据 Kappa 系数机理分析,两个模型错估误差和漏估误差较大,导致 Kappa 系数很低。因此评价模型预测精度不能够仅考虑模型的总体精度,使用 Kappa 系数进行评价能够更好的验证模型预测结果和原始监测数据的一致性程度。

以 20 幅监测 VTCI 序列(20×341×161 像素)和 20 幅预测 VTCI 序列各自作为一个整体,逐像元建立误差矩阵,得到加权 Markov 模型 1 步预测和 2 步预测的总体精度和 Kappa 系数分别为 0.37, 0.37 和 0.35, 0.36, ARIMA 模型 1 步预测和 2 步预测的总体精度和 Kappa 系数分别为 0.46, 0.46 和 0.38, 0.38。由此可见,整体上模型的 1 步预测较好于 2 步预测, Kappa 系数随样本数目增加明显增大,预测序列和监测序列的一致性程度明显提高,模型的预测精度也越来越高;根据极限和 Kappa 系数的定义,样本数目的增加将使得 Kappa 系数不断逼近于总体精度,二者均可以用于评价模型的预测精度。

对比两个模型的 Kappa 系数和总体精度不难发现,无论是 20 幅预测 VTCI 序列的平均值,还是样本数目增加后作为一个整体的 VTCI 序列, ARIMA 模型比加权 Markov 模型预测精度较高,更适合基于 VTCI 的干旱预测建模。

3 结论

通过建立误差矩阵,把用于评价遥感分类精度的 Kappa 系数引入到干旱预测模型的预测精度评价中,实验结果证明,错估误差和漏估误差统计能够用来评价模型的局部适用性,总体精度较高不能够充分验证模型的预测精度较高, Kappa 系数可以从整体上评价模型的预测精度,因此该方法在干旱模型的预测精度评价中具有一定的适用性。

基于 SPI 序列的加权 Markov 模型干旱预测结果证实了该模型预测精度较高且稳定,适合较长时间尺度无旱等级的干旱预测,基于同样数据源的 ARIMA 模型预测结果则不理想;基于 VTCI 序列的两个模型干旱预测结果预测精度较低,但是一定程度上样本数目的增加下可以提高两个模型的预测精度。

目前利用 Kappa 系数等评价指标在预测模型精度的评价中的工作只是开始, VTCI 等级的划分具有一定的主观性,加权 Markov 和 ARIMA 模型的算法也需要改进, Kappa 系数评价指标可以充分考虑到标准 Kappa 系数、随机 Kappa 系数、位置 Kappa 系数和数量 Kappa 系数等评价因子,这些问题尚需进一步的探讨和尝试。

参考文献:

- [1] Feinstein A R, Cicchetti D V. High agreement but low Kappa: I. The problems of two paradoxes[J]. *Journal of Clinical Epidemiology*, 1990, 43(6): 543-549.
- [2] Cicchetti D V, Feinstein A R. High agreement but low Kappa: II. Resolving the paradoxes[J]. *Journal of Clinical Epidemiology*, 1990, 43(6): 551-558.
- [3] Pontius J. Quantification error versus location error in comparison of categorical maps[J]. *Photogrammetric Engineering and Remote Sensing*, 2000, 66(8): 1011-1016.
- [4] Naesset E. Use of the weighted Kappa coefficient in classification error assessment of thematic maps[J]. *International Journal of Geographical Information Systems*, 1996, 10(5): 591-604.
- [5] 布仁仓, 常禹, 胡远满 等. 基于 Kappa 系数的景观变化测度 - 以辽宁省中部城市群为例[J]. *生态学报*, 2005, 25(4): 778-784.
- [6] 闫岩, 华琳, 张建. 对诊断一致性 Kappa 系数及评价指标的探讨[J]. *中国卫生统计*, 2007, 24(3): 313-315.
- [7] 韩婷, 李慧梁, 张宏 等. Kappa 统计量评价教学督导组与学生评教的一致性[J]. *医学教育探索*, 2007, 6(12): 1117-1119.
- [8] Cohen J. A coefficient of agreement for nominal scales[J]. *Educational and Psychological Measurement*, 1960, 20(1): 37-46.
- [9] Cohen J. Weighted Kappa: nominal scale agreement with provision for scaled disagreement or partial credit[J]. *Psychological Bulletin*, 1968, 70(4): 213-220.
- [10] 王鹏新, 龚健雅, 李小文. 条件植被温度指数及其在干旱监测中的应用[J]. *武汉大学学报: 信息科学版*, 2001, 26(5): 412-418.
- [11] 孙威, 王鹏新, 韩丽娟 等. 条件植被温度指数干旱监测方法的完善[J]. *农业工程学报*, 2006, 22(2): 22-26.
- [12] 王彦集, 刘峻明, 王鹏新 等. 基于加权马尔可夫模型的标准化降水指数干旱预测研究[J]. *干旱地区农业研究*, 2007, 25(5): 198-203.
- [13] 韩萍, 王鹏新, 王彦集 等. 多尺度标准化降水指数的 ARIMA 模型干旱预测研究[J]. *干旱地区农业研究*, 2008, 26(2): 212-218.
- [14] Anthony J, Viera M D, Joanne M. Understanding interobserver agreement: the Kappa statistic[J]. *Family Medicine*, 2005, 37(5): 360-363.