

基于遥感数据的空间目标融合预测模型

韩波^{1,2}, 康立山¹, 陈毓屏¹, 宋华珠³

(1. 武汉大学计算机学院, 武汉 430072; 2. 坦普尔大学信息科学技术中心, 费城 19122;

3. 武汉理工大学计算机科学与技术学院, 武汉 430070)

摘要:提出了一种融合预测模型,把基于全局数据的径向基函数、人工神经网络与基于区域数据的空间采样插值相结合,并通过优化权重组合和修正误差,使得预测误差最小化。利用了多角度成像光谱辐射仪 MISR,采集了从 2002 年—2003 年美国大陆 26 个局部区域的辐射数据,对模型进行了 2 组气溶胶光学厚度的预测实验。地表特征因为反射能力的不同,导致了预测模型的不同复杂度。结果显示,融合模型能显著地减少陆地上的预测均方差。

关键词:空间目标;融合预测模型;全局预测模型;局部预测模型;气溶胶光学厚度

A Fusion Prediction Model for Spatial Target Based on Remote Sensing Data

HAN Bo^{1,2}, KANG Lishan¹, CHEN Yuping¹, SONG Huazhu³

(1. School of Computer, Wuhan University, Wuhan 430072; 2. Center for Information Science and Technology, Temple University, Philadelphia 19122;

3. School of Computer Science and Technology, Wuhan University of Technology, Wuhan 430070)

【Abstract】This paper proposes a fusion approach, which combines radius-basis-function artificial neural network model learning from global data with spatial interpolation model sampling from local data. The integration model minimizes the mean squared prediction errors by optimizing weights of their outputs and correcting the errors. It evaluates the fusion approach by cross-year-validation based on multi-angle imaging spectra radiometer (MISR) collected from January 2002 to December 2003 over 26 local regions in the continental US. It is noticed that fusion model can significantly reduce prediction error on land and desert.

【Key words】Spatial target prediction; Fusion model prediction; Global model prediction; Local prediction model; Aerosol optical thickness

1 概述

1999 年 12 月美国航空航天局(NASA)发射了地球观测卫星 Terra, 安装在其上面的多角度成像光谱辐射仪(Multi-angle Imaging Spectro Radiometer, MISR)主要用于拍摄大量从地球发出或反射的射线的辐射率^[1],以便估算全球范围内太阳光透过大气层时被阻挡的程度和大气层中的悬浮小微粒气溶胶的光学厚度(Aerosol Optical Thickness)。这项研究对于监测水文周期、环境和气候变化等都有非常重要的意义^[2]。然而,由于遥感数据属性间相互关系复杂,并涉及很多未知领域,使得在许多实际应用中基于遥感数据的空间目标预测准确率一般不高,因此研究人员希望利用信息融合技术建立不同的预测模型来提高预测性能。

通过对大量空间数据的分析,我们注意到一个空间区域既含有全局普遍特征,又含有局部所特有的性质^[3]。对于全局特征,以往采用的数据挖掘技术一般是以数据间独立分布关系为假设,建立基于整个数据集的全局学习模型;对于局部性质,根据地理第一定律:距离近比距离远的数据表现出更多的相关性,因此空间统计学利用区域权重插值建立局部学习模型^[4]。前者能从大量全局数据中发现较为复杂的普遍关系,后者从一定数量的区域插值中捕捉特定时空属性。本文提出了一种融合预测模型,它采用RBF人工神经网络构造全局模型,使用反距离采样插值构造局部模型,通过优化加权两个模型的输出并进行空间误差校正来进行预测。

利用 MISR 光学辐射观察数据,以气溶胶光学厚度为空间目标的实验结果表明本融合模型比独自使用局部模型或全

局模型能更准确地作出预测。

2 融合模型的建立

信息融合技术通常在数据层次、特征层次和模型层次综合信息上提高预测性能。对于 MISR 数据源,我们希望使用不同的方法建立独立模型,然后再利用线性组合和优化求快速得到最佳组合权重。

2.1 全局模型 GRBF

利用 3 层结构的径向基函数,人工神经网络(RBF ANN)从整个数据集 $D=\{t_i\}(i=1,\dots,N)$, 每 1 个样本 t_i 用 (X_i, Y_i) 表示,其中 Y_i 是目标属性, $X_i=[x_{i1}, x_{i2}, \dots, x_{iM}]$ 是从光学辐射信息中提取的 M 维属性向量)中建立全局模型 GRBF。所有 M 维属性首先经过主成分分析变换成为 M' ($M' < M$) 个输入,通过输入层传给隐含层;隐含层包含 M' 个神经元(Radius Basis Function Unit)。用高斯密度公式作为激励函数训练神经网络,每个隐含神经元对应一个按属性 $\langle \text{center}_j, \text{width}_j \rangle (j=1,\dots,M')$ 构成的聚类,其中 center_j 为第 j 个聚类的 M' 维中心向量, width_j 为影响范围;输出层将隐含层中的各个聚类输出,通过加权线性组合构成。对于某个采样点的光学辐射向量 X_i , 它的气溶胶光学厚度预测公式如下:

$$\text{GRBF}(X_i) = \sum_{j=1}^{M'} w_{ij} \Phi_j(X_i) = \sum_{j=1}^{M'} w_{ij} e^{-\frac{(X_i - \text{center}_j)^2}{2 \text{width}_j^2}} \quad (1)$$

作者简介:韩波(1972—),男,博士生,主研方向:数据挖掘,并行计算;康立山,博导、教授;陈毓屏,副教授;宋华珠,博士
收稿日期:2006-02-28 **E-mail:** bhmilinfo@163.com

模型参数利用训练数据通过最小化预测误差来决定。

2.2 局部模型 LIDSJ

根据地理第一定律：距离近比距离远的数据表现出更多的相关性。把数据D根据空间上的连续性分为L个空间区域，针对每个区域构造一个局部预测模型。我们标记局部区域j中的数据子集为D_j，它含有的样本个数为N_j，相应建立的局部预测模型为LIDSJ_j。

本模型可以使用前向神经网络来构造。但在研究中发现，建立只包含有限数量的区域神经网络，比较耗费时间并且容易造成过度训练(Overfitting)。因此采用了空间统计学中的反距离空间插值预测。

给定位于 1 个局部区域j的训练数据集{(lat_i, lon_i, Y_i), i = 1,...,K} 的集合，用反距离空间插值预测邻近的位置的AOT值。在特定位置 (lat, lon)的AOT的计算公式如下：

$$LIDSJ_j = \frac{\sum_{i=1}^K Y_i / d_i^p}{\sum_{i=1}^K 1 / d_i^p} \quad (2)$$

其中，d_i 是从(lat_i, lon_i)到(lat, lon)的距离，p是参数用于确定最近邻居强调的范围。p 值大则分配大的权值给近邻；当p=0时，把平均值分配给近邻；当p = ∞时，此模型等价于最近邻居预测模型。

2.3 空间目标融合预测模型 OPWM

空间目标融合预测模型 OPWM 需要考虑两个问题：

- (1)对 1 个区域，如何既考虑其具有的全局普遍特性又考虑具有的局部特殊性，并将全局模型和局部模型有效地结合起来。
- (2)与一般预测模型中预测误差大体符合正态分布不同，空间目标预测误差仍然具有空间相关性，需要对预测误差进一步修正。

因此本模型定义如下：

$$OPWM_j(X_i) = Fusion_j(X_i) + Error_j(X_i) \quad (3)$$

其中，Fusion_j(X_i) = α_jGRBF(X_i) + (1 - α_j)LIDSJ_j(X_i)

融合参数α_j用于对全局模型和局部模型进行优化加权平均，使得在一个特定局部区域j中的预测均方误差

$$MSE_j = \sum_{i=1}^{N_j} [Y_i - (\alpha_j GRBF(X_i) + (1 - \alpha_j) LIDSJ_j(X_i))]^2$$

最小化，优化参数为

$$\alpha_j = \frac{\sum_{i=1}^{N_j} (Y_i - LIDSJ_j(X_i)) \cdot (GRBF(X_i) - LIDSJ_j(X_i))}{\sum_{i=1}^{N_j} (GRBF(X_i) - LIDSJ_j(X_i))^2}$$

其中，α_j的取值范围是[0, 1]，它能揭示一个地区的特征。如果α_j接近 1，表明本地区的属性与全局普遍性质相似；相反，如果α_j接近 0，则表明局部区域特性突出。

对于Error_j(X_i)，再次利用训练数据预测误差Error_j(X_i) = OPWM_j(X_i) - Fusion_j(X_i)和反距离空间插值式去修正测试数据预测误差。

3 实验

气溶胶光学厚度用于反映太阳光透过大气层时被阻挡的程度，它是研究环境和气候变化的一个重要因素，也是地理学家普遍关注的一个重要问题^[3]。本实验以气溶胶光学厚度为例，研究融合模型对空间目标的预测准确率。

3.1 实验数据

我们获得了 2002 年—2003 年的有关 MISR 光学辐射观测的 3 个产品：辐射数据集 MI1B2E，云层分类数据集 MIRCCM 和气溶胶光学厚度数据集 MIASAE。这 26 个局部区域如图 1 所示。根据信息，抽取 M=73 个属性(见表 1)。

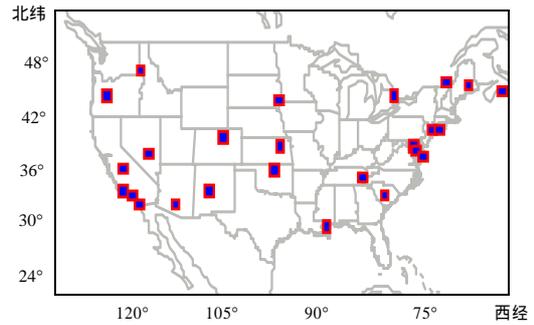


图 1 美国大陆 26 个局部区域数据

表 1 数据属性描述

属性索引	含义
1~36	36 个辐射属性(9 个相机，每个相机有 4 个波长)
37	太阳光角度
38~46	9 个相机角度
47~55	9 个相机相对的太阳光角度
56~64	9 个相机的散射角度
65~73	9 个相机的闪光角度

3.2 预测实验及结果

采用交叉年份检验方法来验证融合预测模型的性能。先根据给定某一年获得的各个采样点的 73 个属性和相应的气溶胶光学厚度值建立空间预测模型，然后根据另外一年的采样点属性预测其气溶胶光学厚度。

这样训练数据和测试数据分开，能比较真实地反映全局模型、局部模型和融合模型的预测效果。表 2 给出了各个实验的训练数据、测试数据以及 3 种模型的整体预测均方差 MSE。融合模型的整体预测均方差比全局模型和局部模型分别减少 19%和 16%。

表 2 3 种模型整体预测均方差 MSE 比较

	训练数据	测试数据	全局模型	局部模型	融合模型
实验 1	2003 年	2002 年	0.008 4	0.009 6	0.007 8
实验 2	2002 年	2003 年	0.008 4	0.006 5	0.005 8
总体平均			0.008 4	0.008 1	0.006 8

由图 2 可以看出，局部模型在第 1 季度、第 4 季度的预测误差比全局模型相对减少，而融合模型在 4 个季度中的预测准确率比其他 2 个模型显著提高。图 3 比较了 3 个模型在不同地表条件下的预测误差：3 个模型在水边及沿海地区的预测效果不如在陆地/沙漠地区的预测准确，而融合模型较大地提高了在陆地/沙漠地区的预测准确率。

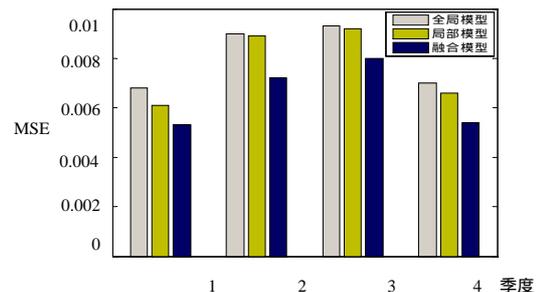


图 2 3 种模型 4 个季度的预测误差

(下转第 39 页)