

STATISTICAL ANALYSIS AND DATA MINING

Research Article

A scalable gaussian process analysis algorithm for biomass monitoringVarun Chandola , Ranga Raju Vatsavai

First published: 08 July 2011

<https://doi.org/10.1002/sam.10129>

Cited by: 13



About



Access

als



ire

Abstract

Biomass monitoring is vital for studying the carbon cycle of earth's ecosystem and has several significant implications, especially in the context of understanding climate change and its impacts. Recently, several change detection methods have been proposed to identify land cover changes in temporal profiles (time series) of vegetation collected using remote sensing instruments, but do not satisfy one or both of the two requirements of the biomass monitoring problem, that is, *operating in online mode* and *handling periodic time series*. In this paper, we adapt Gaussian process (GP) regression to detect changes in such time series in an online fashion. While GP have been widely used as a kernel - based learning method for regression and classification, their applicability to massive spatiotemporal data sets, such as remote sensing data, has been limited owing to the high computational costs involved. We focus on addressing the scalability issues associated with the proposed GP based change detection algorithm. This paper makes several significant contributions. First, we propose a GP based online time series change detection algorithm and demonstrate its effectiveness in detecting different types of changes in *Normalized Difference Vegetation Index* (NDVI) data obtained from a study area in IA, USA. Second, we propose an efficient Toeplitz matrix based solution which significantly improves the computational complexity and memory requirements of the proposed GP based method. Specifically, the proposed solution can analyze a time series of length t in $O(t^2)$ time while maintaining a $O(t)$ memory footprint, compared to the $O(t^3)$ time and $O(t^2)$ memory requirement of standard matrix manipulation based methods. Third, we describe a parallel version of the proposed solution which can be used to simultaneously analyze a large number of time series. We study three different parallel implementations: using threads, Message Passing Interface (MPI), and a hybrid implementation using threads and MPI. Experimental results show that the hybrid implementation scales better than the multithreaded and MPI based implementations. The application of the proposed scalable algorithm is demonstrated in analyzing massive remote sensing observation data. The hybrid implementation, using 1536 computing cores, can analyze an NDVI data set for the Iowa study area in nearly 5 s, while a serial algorithm, using standard Cholesky decomposition based routines, takes several days to process the same data set. © 2011 Wiley Periodicals, Inc. *Statistical Analysis and Data Mining* 4: 430–445, 2011



About Wiley Online Library



Help & Support



Opportunities



Connect with Wiley

