RESEARCH PROPOSAL

Title: Research and Application of Spatially and Temporally Weighted Regression Models under High-Performance Computing Strategies

1. Project Origin:

Spatially and temporally weighted regression analysis is an effective method to detect the non-stationary spatial and temporal characteristics of phenomena, and it is an essential component of spatiotemporal data mining techniques. By utilizing the spatially and temporally weighted regression method, regression equations between characteristic variables and dependent variables can be established, the spatiotemporal correlation between phenomena can be analyzed, and the spatiotemporal variation patterns of phenomena can be summarized. This approach is of great significance for exploring the potential value of spatiotemporal data.

With the continuous development of methods and technologies for acquiring spatiotemporal data, the time interval for data acquisition has gradually shortened, and the amount of data collected has continuously expanded, resulting in the rapid exponential growth of spatiotemporal data. The increase in data volume provides a solid analytical foundation for exploring the interrelationships between phenomena and understanding their spatiotemporal variation patterns. However, the ability to spatially analyze and model massive spatiotemporal data is the key and core of geographic information system research in the new era, and improving the efficiency of local regression and model calibration has always been a significant challenge in geographic information science. Research has found that traditional geographically weighted regression (GWR) analysis requires more than two weeks for over 100,000 data points, and geographically and temporally weighted regression (GTWR), which incorporates the temporal factor in addition to the traditional GWR method, has a more complex weighting calculation method and consumes even more time. Therefore, how to quickly calculate the parameters and fitted values of the spatially and temporally weighted regression model is a crucial issue that needs to be addressed in the methodology of spatially and temporally weighted regression.

2. Practical research significance:

Research on applications at large regional and global scales has become an important trend in the development of geospatial analysis, such as global population flow simulation, carbon cycle, dynamic spatial and temporal monitoring of environmental changes, which puts higher demands on the rapid processing and expression of multidimensional geospatial data. The huge computational load and multi-source heterogeneous data types involved in the geospatial analysis process in the era of big data have challenged the computational efficiency of traditional processing methods in the field of data mining. Currently, most scholars are concerned about the improvement and application promotion of spatiotemporal geographically weighted regression methods, while there is less research on the computational efficiency of GTWR. However, improved spatiotemporal geographically weighted regression methods may increase the complexity of algorithms and consume more computing time. By utilizing high-performance computing resources and strategies, it is possible to solve the challenges of low computational efficiency and multi-source heterogeneous data in spatiotemporal geographically weighted regression. Exploring spatiotemporal geographically weighted regression models under high-performance computing strategies can not only improve the computational efficiency and data processing capabilities of geospatial analysis, but also provide important support and guarantee for solving complex problems in global-scale geospatial analysis.

3. Current Status, Level, and Development Trends of Practical Research at Home and Abroad:

In dealing with large data volumes and highly complex spatiotemporal computing issues, scholars at home and abroad have conducted a series of studies to gain a deeper understanding and expression of the problems themselves and their solution processes, exploring the feasibility of high-performance parallel processing. Lv Jiangbo implemented a parallel vehicle trajectory information preprocessing program based on MapReduce [1]. The results showed that the preprocessing program could meet the preprocessing needs of large-scale taxi travel trajectories. Cao Jie et al. proposed a Graphics Processing Unit (GPU)-accelerated differential evolution particle filter algorithm to address the issue of slow computation caused by the complex computational process in particle filter systems. This algorithm reduces computational complexity while accelerating the computational speed [2]. Experimental results showed a significant reduction in computational time using the GPU-accelerated algorithm. Zhou Chen et al. proposed a hybrid heterogeneous spatial analysis adaptive load balancing method utilizing Central Processing Unit (CPU) and GPU to address computational challenges posed by vast geographic data [3]. Experimental results indicated that this method can better address the issue of load balancing in spatial analysis under parallel computing modes. Wu Weimin achieved real-time stitching of high-resolution images using the Compute Unified Device Architecture (CUDA) [4]. Experimental results showed that using the CUDA framework to transfer image feature point matching to the GPU significantly shortened the matching time and significantly increased the speed of video stitching.

In the research of geographically weighted regression methods, scholars have also proposed some strategies to improve the efficiency of parameter estimation. Liu Zhentao leveraged the excellent parallel computing capabilities of GPUs to meet the speed requirements of edge extraction in dynamic photogrammetry [5]. Experimental results showed that the parallel computing-based edge extraction method improved the computational speed by 11.6 times compared to the serial computing mode. Liu Zhentao also implemented a CUDA-based parallel acceleration of geographically weighted regression to address issues such as excessive memory usage and slow computation speed in traditional geographically weighted regression models for big data processing. This approach decomposed a large serial computing task into multiple independently executable subtasks and optimized storage methods [6]. Danlin Yu addressed the issue of traditional GWR's inability to perform parameter estimation on large data volumes by extracting a subset of observations for analysis [7]. This method produced results by reducing the data volume but did not fully utilize the observation data, resulting in a loss of non-stationarity in local regions. Thierry Feuillet et al. divided the study area into multiple sub-regions to address the issue of traditional GWR's inability to calculate all observation points at once. They first calculated the estimated parameters for each sub-region separately and then aggregated them [8]. This strategy fully utilized the observation data, but calculating by sub-regions may alter the non-stationarity at sub-region boundaries, causing local estimation biases. Alexei Pozdnoukhov et al. proposed a regression analysis efficiency optimization strategy based on the MapReduce parallel framework [9]. This strategy significantly improved the computational efficiency of linear regression models but still limited the data volume when processing spatial data. Hung Tien Tran et al. proposed a distributed GWR computation strategy based on the Spark framework [10]. This method effectively improved the computational efficiency of traditional GWR, but it required setting up a cluster computing environment, which raised the application threshold for GWR. Ziqi Li et al. proposed a FastGWR processing strategy based on Python and the Message Passing Interface (MPI), which achieved parallel processing of regression coefficients and distance matrices on a single computer, thus improving computational efficiency [11]. While many studies have been conducted on parallel processing strategies for parameter estimation in traditional GWR methods, no effective efficiency optimization solutions have been proposed for traditional GTWR methods. Therefore, how to improve the computational efficiency of GTWR based on high-performance computing strategies is a worthy research question.

References:

[1] 吕江波,张永忠.Hadoop 支持下海量出租车轨迹数据预处理技术研究[J].城市勘测,2016(03):46-49.

[2] 曹洁,黄开杰,王进花.GPU 加速的差分进化粒子滤波算法[J].计算机应用研究,2018,35(07):1965-1969.

[3] 周琛. 面向 CPU/GPU 混合架构的地理空间分析负载均衡并行技术研究[D].南京大学,2018.

[4] 吴为民,刘新,江先伟.应用 CUDA 进行交通监控视频实时拼接[J].福建电脑,2021,37(11):110-112.

[5] 刘振涛,燕必希,董明利,孙鹏,王君.并行计算在动态摄影测量边缘提取算法中

应用[J].计算机工程与设计,2019,40(01):97-102.
［6］刘振涛,杨毅,王东超,谢晓尧. 一种基于 CUDA 的大数据量地理加权回归并行加速算法[J]. 测绘通报,2020,(12):1-5.
［7］Danlin Yu. Modeling owner-occupied single-family house values in the city of Milwaukee: a geographically weighted regression approach[J].GIScience & Remote Sensing,2007,44(3):267-282.
［8］Thierry Feuillet, Hadrien Commenges, Mehdi Menai, et al. A massive geographically weighted regression model of walking-environment relationships[J]. Journal of Transport Geography, 2018,68:118-129.
［9］Alexei Pozdnoukhov, Christian Kaiser. Scalable local regression for spatial analytics[C]. New York: ACM, 2011,361-364.
［10］Hung Tien Tran， Hiep Tuan Nguyen， Viet-Trung Tran. Large-scale geographically weighted regression on Spark[C], IEEE, 2016:127-132.
[11] Ziqi Li, A. Stewart Fotheringham, Wenwen Li, et al. Fast geographically weighted regression(FastGWR):a scalable algorithm to investigate spatial process heterogeneity inmillions of observations[J].International Journal of Geographical Information Science,2019,33(1): 155-175.