Available online at www.sciencedirect.com

ScienceDirect

IERI Procedia 10 (2014) 138 – 143

Procedia IERI

www.elsevier.com/locate/procedia

2014 International Conference on Future Information Engineering

Window Query and Analysis on Massive Spatio-Temporal Data

Huan Wang, Junhui Deng, Guodong Yuan*

Department of Computer Science and Technology, Tsinghua University, Beijing,100084, China

Abstract

Along with the expansion of computer-based climate simulations, efficient visualization and analysis of massive climate data are becoming more important than ever. In this paper, we try to explore the factors behide climate changes by combining window query and time-varying data mining techniques. With constant query time and acceptable storage cost, the algorithms presented support various queries on 3d time-varying datasets, such as average, min, and max value. A new time-varying data analysis algorithm is given, which is especially suitable for analyzing big data. All these algorithms have been implemented on and integrated into a visual analysis system, with tiled-LCD ultra-resolution display. Experimental results on several datasets from practical applications are presented.

Selection and peer review under responsibility of Information Engineering Research Institute

Keywords: Window Query; Visual Analytics; Spatio-Temporal Data Mining; Visualization; Earth System Model;

1. Introduction

In recent years, with the fast growing of computational capabilities and storage capacities, much more scientific data has been generated by advanced observation instruments and simulation experiments, such as the global ocean real-time observing system. We will take the earth system model (ESM) as an example. The

* Corresponding author. Tel.: +86-158-0140-2071.
E-mail address: whthu168@163.com.
This paper is supported by the National High Technology Research and Development Program of China (863 Program) under grant No. 2010AA012402.

2212-6678
Selection and peer review under responsibility of Information Engineering Research Institute
doi:10.1016/j.ieri.2014.09.068

ESMs are sets of equations describing processes within and between the atmosphere, ocean, sea-ice and the terrestrial and marine biosphere [1]. According to the simulation of earth climate and environmental changes, scientists can get a better understanding on the ecological environment and furthermore interpret the earth evolution mechanism in more details.

Big data of size up to TB is everywhere in earth science. Typically, even a single scan of such a dataset will cost dozens of minutes. That is why the visualization techniques and tools are so important for scientists [2]. On the other hand, however, the way human beings intend to interpret data is much different from that of computers. With higher-level visual and perceptual capabilities, for example, we can find patterns from images and animations more quickly and precisely.

In fact, climate visualization has been an important part of VISC for the last two decades. Usually, 2D scalar data can be visualized using algorithms such as color mapping, while 3D scalar data can be visualized using algorithms such as isosurfaces, contouring extractions, volume rendering, etc.. A number of visualization toolkits have also been developed, such as VTK [3] and visIt [4]. In recent years, visualization techniques are moving fast in the direction of large-scale, interactive, and real-time.

Visual analytics (or, visual data mining) is a combination of visualization and data mining techniques. Wong et al. gave a list of top challenges in extreme-scale visual analytics [5]. In this paper, we present a comprehensive visual analysis platform for massive spatio-temporal data based on a combination of window query techniques and time-varying data mining methods.

2. Background

2.1. Basic Concepts

The problem of range query is one of the core problems in computational geometry. Suppose Ω is a system consisting of several subsets of a d-dimensional Euclidean space R^d . P is a point set consisting of n elements in R^d . The problem of range query can be defined as: for any given region $R \in \Omega$, to design efficient algorithms that can find out what elements belong to R (fig. 1(a)). If P and R are given at the same time, the problem can be transformed into a problem of solving $P \cap R$. We can make a one by one test on the elements of P , which is obviously time-consuming [6].

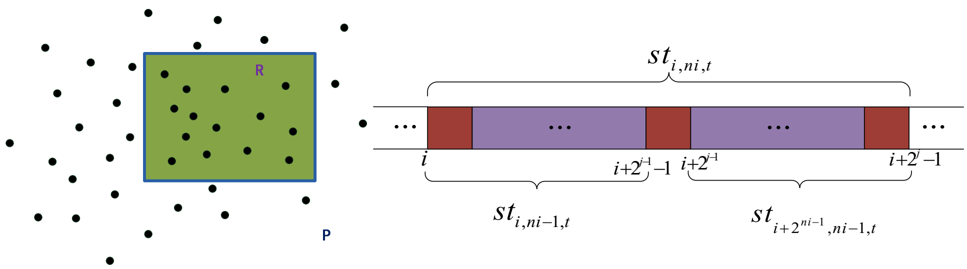

Fig. 1. (a) Left: an example for range query; (b) Right: the computation of sparse table in 1d situation

In real applications, P is often given in advance and relatively unchangeable, which means it can be solved by a more efficient algorithm. In this situation, it is abstracted into a precise mathematical model. For any given point $p \in P$, a weight $w(p) \in S$ is assigned. It can be proved that $(S,+)$ is a commutative semigroup [6]. Then this problem is mathematically expressed to compute the sum of all the weights of p belonging to $P \cap R$.