ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 3, March 2015.

A Hierarchal Tensor Based Approach with Boosted Regression Trees

Deeba M[1], J.Shyamala Devi[2]

Student, M.Tech, SRM University, Chennai, India[1]

Assistant Professor (OG), Department of Computer Science, SRM University, Chennai, India[2]

Abstract: Within the split-and-merge paradigm for the compressed, continuously updating and data querying of multidimensional spatial data, the original multidimensional spatial tensor data are divided into small blocks according to their spatial data references. Each block is represented in regression tree structure and at each level, all leaf nodes of regression tree are boosted then compressed hierarchically. Then the blocked hierarchical tensor representation combined into a single hierarchical tree as the representation of original data. With a buffered regression tree data structure and boosting, the corresponding optimized operation algorithms, the original multidimensional spatial field data can be continuously compressed, appended, and queried. The new approach with above used regression trees instead of buffered trees and using boosting algorithms to improve the fit ratio, so that the quality of the original data with much low storage costs and faster computational performance.

Keywords: Geo Spatial Data, Tensors, Regression Trees, Gradient Boosting,

I. INTRODUCTION

The spatial data have large volumes (eg. several gigabyte) and high dimensionality (eg. hundreds of attributes). The spatial data are often compressed for storage and the newly arrived data also be continuously compressed and appended to the existing data. The updated process has to be completed in short time and repeatedly applied for next set of data. The storage maintenance and query processing are a challenged one because of continuous changes in data volumes and dimensions. Different data structures are used to improve the processing throughput. Tensor is one of the data structure used for high dimensional data manipulation. Recently hierarchical Tensor representations (HTR) provide rich reference for high dimensional data. The most reference computational Tensor tools used are htucke r- Toolbox, Tensor-train Tool box and Tensor Calculus Library. Several issues are notified with HRT as the imbalance of dimension splitting and null space problems.

The above issues of HRT will be reduced with applying Boosted Regression trees in split-and merge paradigm. In spatial data all the attributes are continuous data. The construction of regression trees used the values of continuous variables in block for splitting, then boosting is applied at each level nodes. Boosting is a general method for increasing accuracy of any given algorithm. Ada boost is the example for boosting binary categorization. With a buffered regression tree data structure, corresponding optimized operation algorithms and types of boosting methods, and the original multidimensional spatial field data can be continuously compressed, appended, and queried. The result suggests the blocked hierarchical tensor representation with boosted Regression trees provides an effective splitting for effective storage and computation of multidimensional spatial field data.

This paper explicitly proposes the boosted regression tree construction in Hierarchical Tensor based approach in spatial temporal data. The intention of this proposal is to represent a new algorithm for blocked hierarchical tensor representation with boosted Regression trees provides an effective splitting for effective storage, retrieval and computation of multidimensional spatial field data. This paper is organized as follows: In section 2, related work while section 3 gives detail background explanation on theoretical aspects of the description of necessary notations for HTR , regression tree and boosting. Section4 demonstrate the overall framework, In section5 Algorithm description. Finally, conclusion is discussed in section6.

II. RELATED WORK

Hierarchical Tensor Representation for spatial data is a recent trends in geospatial data. So many approaches are already in use to compress, updating and data querying of multidimensional spatial field data. The Sketch technique is used for approximate spatial join of two spaces, Low rank Tensor approximation techniques, Low Rank Tensors and Fixed Rank filtering for the spatial temporal data used various techniques for efficient storage and computation of multidimensional spatial field data. Each technique used its own representation of spatial data. Our algorithm falls into HTR. Hierarchical Tensors are a flexible generalization of the well known Tucker representation. Tensors decomposition can be computed via Singular Value Decomposition, limited do not extent matrix to Tensors.

III. BACKGROUND

In this section we review the theory of HTR, regression trees and Boosting concept.

3.1 Blocked HTR representation.

DOI 10.17148/IJARCCE.2015.43109

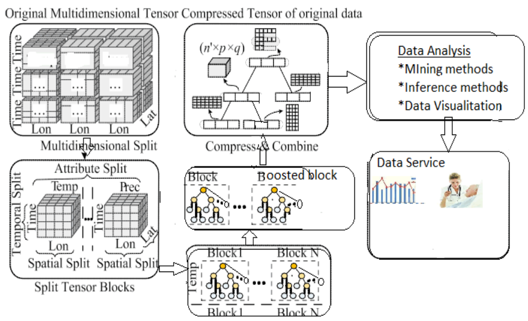
With the subspace split of spatial temporal representation and attribute dimension, the multidimensional temporal field data are split into blocks, Each block has its own temporal references. A function Split to Blocks(T,b), where T is the original tensor and b is the block size is designed to split the original tensor into sub blocks.

3.2 Regression Tree
Regression analysis is a statistical tool widely used for prediction and forecasting. Regression tree construction in classification done in repeated expansion of nodes until the given depth is met. Initially, all data are assigned to root node. The general approach to derive nodes from few simple if-then condition can be applied to regression problems as well. The more advantage of using regression is much easier to evaluate just one or two logical conditions than some elaborate equations.

3.2 Boosting:
Boosted Regression trees combined the strength of two algorithms: regression trees and boosting. Boosting is an adoptive method for combining many simple models to give improved predictive performance. Boosting is the technique for improving the accuracy of a predictive function by applying the function repeatedly in a series and combining the output of each function with weighted so that the total error of the prediction is minimized. Relative techniques – including bagging and model averaging also build then merge results from multiple models , but boosting is unique because it is sequential. The original boosting algorithms such as AdaBoost(Freund & Schapire 1996) were developed.

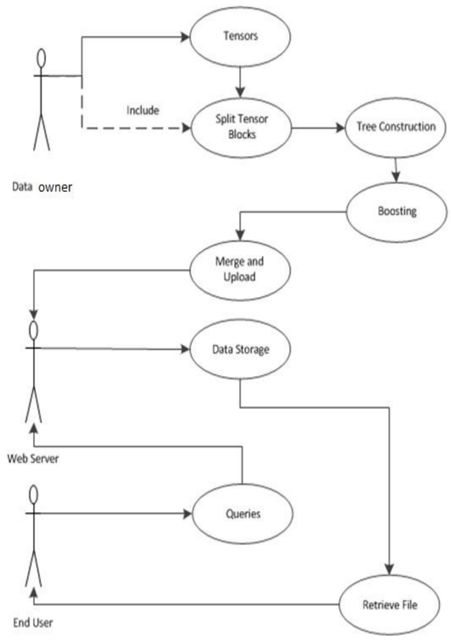
IV. OVERALL FRAMEWORK

The overall framework of the hierarchical tensor representation with boosted regression trees and the updating of the geospatial tensor are depicted in the following figure.



The overall idea is the original tensors are divided into sub-tensors called blocks.In each block, the regression tree is formed based on the continuous attribute values and are boosted the compressed. Then all blocks are combined to form the hierarchical representation of original tensor. Then the query process can be done effectively. Since we used boosted regression in splitting it is more efficient and accurate.

The use case diagram for the above framework is given in following figure.



V. IMPLEMENTATION OF ALGORITHM

The following algorithm has been developed for hierarchical Tensor Block representation.
Algorithm.
1.Input:Original Tensor T.
2.Parameter block size bk.
3.Tree←Unlabeled node.
4.output: Hierarchical Blocked Representation.
5.begin
6.Ti=SplitToBlock(T,bk)
7.For each block bi, where 1<i<k{
8.While(tree depth < threshold) do{
9.create n number of leaf nodes based on attribute values.
10.for each leaf node l(i) where 1<i<n{
11. AdaBoost(l(i))
12.compress(l(i))} } }
13.combine blocks(b(i)}
15.Hierblockrepof blocks(bi, where 1<i<k)
16.End.

Algorithm description:
The above algorithm starts by giving input of original Tensor which is sub divided into k number of blocks of size bk. For each block the regression tree is constructed and at each level the leaf node is boosted and it extent to next level if depth is less than maximum depth. After that each block is compressed and updated, then all blocks are combined together to form original Tensor.

VI. CONCLUSION

Geospatial data set has high subject settings .Recent years have brought excellent progress in usability. The main objective of this study is to implement the following in Hierarchical Tensor Representation(HTR)
1.Effective Compression and
2.Efficient Query processing.
The above algorithm provide adoptability of regression trees with boosting mechanism in HTR

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 3, March 2015.

The individual researcher cannot easily keep up with the literature in this domain. Up to our knowledge this is the first algorithm Boosted Regression Trees(BRT) with HTR. This set up will be a efficient one. The algorithm could be extended into several directions, so that the efficiency can be improved in future.

REFERENCES

[1] "Tensor Based approach for Big Data Representation and dimensionality reduction". Liwei Kuang, Fei Hao,Laurence T.Yang 2014 *Digital object idendifier10.1190/tech.2014.2330516*
[2] "Tensor methods for machine learning, Computer graphics, and computer vision".Amnon Shashua.
[3] Graham Coremod "Sketch Techniques for Approximate Query processing" Foundation and Trends in sample vol.xx,no,xx(xxxx)1-64.
[4] Machine Learning, 47, 153–200, 2002_c 2002 Kluwer Academic Publishers. Manufactured in The Netherlands."Boosting Methods for Regression"∗Computer Science Department, University of California, Santa Cruz, Santa Cruz, CA 95064, USA
[5] "Fixed rank filtering for spatio-temporal data"Noel Cressie, .au,Tao Shi,*Ohioy*,Emily, L. Kang,Pub:Cressie, N., Shi, T. & Kang, E. L. (2010). Fixed rank filtering for spatio-temporal data. Journal of Computational and Graphical Statistics, 19 (3), 724-745.
[6] " Fast Multivariate Spatio-temporal Analysis via Low Rank Tensor Learning",Mohammad Taha Bahadori_,Qi (Rose) Yu_,Yan Liu
[7] "High Performance spatial indexing for location based services"., Jussi Myllymalei,James Kaufman, IBM Almenden Research Center.
[8] "Tensor completion in hierarchical tensor Representations",Holger Rauhut, Reinhold Schneider and ˇ Zeljka Stojanac.
[9] Stephen Tyree, Killan Q. Weinberger, KunalAgarwal, "Parallel Boosted Regression Trees for Web Search Ranking", Department of Computer Science and Engineering, Washington University in St.Louis.