# Scenario Generation by CART

Huaguang Feng

Advisor: Dr. Nong Shang

Rensselaer Polytechnic Institute

Key Words: Time Series, Wavelet, CART

# The Problem

Given a historical trace, how to generate high quality time series from it. The new trace must be functionally indistinguishable from independent samples of the same random process.

# Problems of the Existing Methods

- Can not deal with the co-existence of long-range dependence and short-range dependence.
- Too complex to be implemented
- Computationally too expensive to generate long series. (For Long Range Dependent (LRD) data, the computational complexity of generating time series with length N is in the order of $N^2$ )

# Moving Blocks Bootstrap (MBB)

- Divide the given series into overlapping blocks
- Take bootstrap sample from these blocks
- Combine the sample blocks to form a new series

# What we need?

A simple,  unified approach to model a mixture of long- and short-range dependence efficiently and accurately

# Solution: Wavelet Domain Modeling

- As opposed to existing methods which model random process in the time domain, we model it in the wavelet domain.
- By doing this, we can reduce the long dependence in the data to the short dependence in its wavelet coefficients.

# Major Advantages of Wavelet Domain Modeling

- Wavelet transformations de-correlate the temporal correlation significantly.
- Wavelets only require O(N) (N is the length of the trace) computational complexity.

# Wavelets

- Wavelets are complete orthogonal basis functions.
- Wavelet coefficients are random variables.
- There is a one-to-one correspondence between a random process and its wavelets coefficients.

# Tree Representation of Wavelet Basis Functions

- We can use a tree structure to represent wavelet basis functions.
- For example, if we use Harr wavelet and N = 8, then we have the following figures:

# Classification And Regression Tree (CART)

- CART is an efficient tool for tree-structured data analysis. It presents its results in the form of decisions trees.
- CART can automatically searches for important relationships and uncover hidden structures in complex data.

# Time Series Generation by CART(1): General Description

- Recall our problem: Generate new time series from a given historical trace
- The old trace is composed of two parts: the signal part and the noise part.
- We use CART to separate these two parts, from which we generate synthetic traces.

# Time Series Generation by CART(2): the Algorithm

- Generate the biggest tree
- Prune the tree back, stop at a certain level
- Regenerate the signals by a Markov model
- Add noise to get the new trace

# Time Series Generation by CART(3): Examples

- Short Range Dependence
- Long Range Dependence
- Mixture of SRD and LRD

# Byproduct – Complexity parameters

- CART's pruning algorithm provides us a group of complexity parameters when we prune the tree from the bottom to the top.
- Useful properties of the complexity parameters:

  -- Distinguish LRD data from SRD data

  -- Tell us when we should stop pruning the tree

  -- Other useful properties needed to be explored

# Applications

- Computer Network Traffic Modeling
- Trace driven simulations
- Automatic Generation of lifelike scenarios in the field of finance