Nonlinear Modeling of the Internet Delay Structure

Xiao Wang, Yang Chen, Beixing Deng and Xing Li
Tsinghua University

1. Motivation

• Why model the Internet delay structure?
  - Large-scale distributed systems applications:
    - overlay multicast
    - server selection
    - distributed query optimization
    - file-sharing via BitTorrent
    - compact routing

• Typical models – Euclidean
  - Existing systems: GNP, NPS, Vivaldi
  - A serious problem: Triangle Inequality Violation (TIV)

2. Solution: Kernel Methods (KM)

• What are Kernel Methods?
  - A class of algorithms for pattern analysis (e.g., Support Vector Machine (SVM)).

• General task: to find and study general types of relations in general types of data.

• Types of relations: clusters, rankings, principal components, correlations, classifications

• Types of data: sequences, text documents, sets of points, vectors, images, etc.

• Kernel
  - What is kernel?
    - instead of using a mapping \( \phi: X \to \mathcal{F} \) to represent \( x \in X \) by \( \phi(x) \in \mathcal{F} \)
    - using \( K: X \times X \to \mathbb{R} \) to represent Internet delay matrix by \( K(x_i, x_j) \)
  - Interpretation: a mapping exerted on Internet delay matrix
    - Isotrope stationary kernel: \( K(x, y) = K_S([x - y]) \)
    - Euclidean norms:\( \|x - y\| \)
    - The mapping: \( K_S(\cdot) \)
  - Typical kernels:
    - Polynomial kernel, Gaussian kernel, exponential kernel, etc.

3. Methodology

• How to choose kernels?
  - We define:
    - Measured Internet delay matrix: \( D_{Ma} \)
    - Kernel: \( K_S(\cdot) \)
    - Mapped matrix in feature space: \( D_S \), thus \( D_S = K_S(D_{Ma}) \)
    - Current assumption: If there are less TIVs in \( D_S \), each kernel \( K_S(\cdot) \) is a good kernel, since Euclidean models can be embedded in \( D_S \).

• Example: a Euclidean based Network Coordinate system

  Suppose: \( K_S(\cdot) = (\cdot)^2 \)

  \[ D_{Ma} = \begin{bmatrix} a_1 & a_2 & a_3 \\ a_2 & a_3 & a_1 \\ a_3 & a_1 & a_2 \end{bmatrix} = \begin{bmatrix} 0 & 9 & 36 \\ 9 & 0 & 16 \\ 36 & 16 & 0 \end{bmatrix} \]

  Here \( a_1 > a_2 + a_2 \) TIV

  \[ D_S = \begin{bmatrix} b_1 & b_2 & b_3 \\ b_2 & b_3 & b_1 \\ b_3 & b_1 & b_2 \end{bmatrix} = \begin{bmatrix} 0 & 3 & 6 \\ 3 & 0 & 4 \\ 6 & 4 & 0 \end{bmatrix} \]

  Here \( b_1 < b_2 + b_3 \) no TIV

• 2-D coordinates of 1, 2, 3:
  - 1: (0,0)
  - 2: (-1.375, 2.666)
  - 3: (4.0)

Can be embedded!

4. Framework Design

5. Evaluation

• Data sets:
  - PlanetLab: 226 nodes
  - Meridian: 2500 nodes

• Metrics:
  - TIV ratios: the number of triples of nodes violating triangle inequality to the proportion of all triples
  - TIV severity of edge AC: \( \sum_{B \in S} d(A,C) > d(A,B) + d(B,C) \)
  - S: the set of all nodes

6. Further Works

• Delay prediction performance:
  - Current results: not steady among different data sets;
  - Future work: tune adaptive parameters

• How to search for good kernels:
  - Current kernel: polynomial kernel
  - Current methodology:
    - less TIVs in mapped matrix, better kernel.
    - But no guarantee: “if there is no TIV, Euclidean space can be embed”
  - Future work: need further exploration

• Benefits: only Euclidean models?
  - Hyperbolic: spherical add kernels on them
  - Dot-product: add kernels and guarantee non-negativity
  - Future work: need further exploration

Acknowledgement: This work is supported by the National Basic Research Program of China (No.2007CB310806) and the National Science Foundation of China (No.60473087, No.60705032, No.60850003),