Embedding of Time Series and Data Streams

Rodrigo Fernandes de Mello

Invited Professor at Télécom ParisTech (until July 2019)

Associate Professor at Universidade de São Paulo, ICMC

mello@icmc.usp.br

March 13th, 2019

DigiCosme, Paris-Saclay

Some information on the University of São Paulo, Brazil

- Founded in 1934 in the State of São Paulo
  - 46 million people
  - 12 million in the city of São Paulo (capital)
    - 21 million in the capital and neighborhoods
- 11 Campi
  - Annual budget 1.2 billion Euros
- 96.364 students
  - 58.823 Undergraduation
  - 14.106 Masters
  - 15.894 PhD

DigiCosme, Paris-Saclay

Some information on the University of São Paulo, Brazil

- Students:
  - 52.28% men and 47.72% women
- Staff members
  - 5.844 Professors
  - 14.866 Administrative positions
- World University Ranking of the Times Higher Education

Rank	Name	Overall	Teaching	Research	Citations	Industry Income	International Outlook
251–300	University of São Paulo Brazil	46.4–49.4	55.9	53.5	37.0	39.5	32.7

Some Definitions

- What is a data stream?
  - Unbounded sequence of observations collected along time

Some Definitions

- What is a data stream?
  - Unbounded sequence of observations collected along time
- What is a time series?
  - Sequence of observations along time
    - Each observation is typically single dimensional, but that is not a law
      - Ex: The Lorenz System with 3 partial derivatives
    - They can be uniformly or non-uniformly spaced along time
    - It does not assume anything about being unbounded

Some Definitions

- So let us take advantage of Time Series tools
  - From the Statistical branch
    - AR, MA, IMA, ARMA and ARIMA models by Box&Jenkins
    - Auto-Correlation Function to analyze time dependencies

Some Definitions

- So let us take advantage of Time Series tools
  - From the Statistical branch
    - AR, MA, IMA, ARMA and ARIMA models by Box&Jenkins
    - Auto-Correlation Function to analyze time dependencies
  - From the Dynamical Systems branch
    - It attempts to find the set of partial derivatives to represent a **generating process**
    - The generating process is the set of equations that produces a next observations
    - Subdivision:
      - Stochastic approaches
      - Deterministic approaches

Some Definitions

- Today, we will discuss a little about
  - Dynamical Systems >> Deterministic approaches

- Main results:
  - Whitney (1936), Differentiable manifolds. The Annals of Mathematics, v. 37, n. 3, p. 645-680
  - Takens (1980), Detecting strange attractors in turbulence. Dynamical Systems and Turbulence, p. 366-381

Some Definitions

- In differential topology, there are two Whitney embedding theorems:
  - **The strong Whitney embedding theorem** states that any smooth real m-dimensional manifold can be smoothly embedded in the real 2^m-space, if $m > 0$.
  - **The weak Whitney embedding theorem** states that any continuous function from an n-dimensional manifold to an m-dimensional manifold may be approximated by a smooth embedding provided $m > 2^n$.

Some Definitions

- Takens extended Whitney's studies to propose an embedding theorem
  - It provides the conditions under which a smooth attractor can be reconstructed from the observations made with a generic function

$$x(t+(m-1)\tau) = f(x(t), x(t+\tau), x(t+2\tau), \ldots, x(t+(m-2)\tau))$$

τ : is the time lag or separation dimension
m : is the embedding dimension

  - What is that in practice?
    - Using the Logistic map

Some Definitions

- Today, we will discuss a little about
  - Dynamical Systems >> Deterministic approaches

- Main results:
  - Kennel, M.; Brown, R.; Abarbanel, H. (1992a). Determining embedding dimension for phase space reconstruction using the method of false nearest neighbors. Institute for Nonlinear Science and Department of Physics, University of California.
  - Fraser, A. M.; Swinney, H. L. (1986). Independent coordinates for strange attractors from mutual information. Phys. Rev. A , v.33, n.2, p.11341140.

Some Definitions

- They basically define ways of estimating the embedding dimension and the time delay
- Let's see in practice:
  - Using the Lorenz system

Some Definitions

- Recently, we came up with a ML approach to estimate both dimensions in conjunction
  - Pagliosa and Mello (2017), Applying a kernel function on time-dependent data to provide supervised-learning guarantees, Expert Systems with Applications.
  - It builds up some embedding and accesses it using a predictor
- **Most important:** It ensures observations in the reconstructed space are iid, so we can employ supervised learning!!

Conclusions

- What happens if data is not deterministic?
  - We can decompose it, as we will see on the next talk
- Why not taking advantage of time series tools to perform the same with Data Streams?
- Important result:
  - Phase spaces provide iid observations
  - The Statistical Learning Theory is respected and learning is then ensured

Embedding of Time Series and Data Streams

Rodrigo Fernandes de Mello

Invited Professor at Télécom ParisTech (until July 2019)

Associate Professor at Universidade de São Paulo, ICMC

mello@icmc.usp.br

March 13th, 2019