TIJUANA INSTITUTE OF TECHNOLOGY

International Seminar on Computational Intelligence

IEEE CIS-MEXICO CHAPTER

IEEE - Computational Intelligence Society Mexico-Chapter
In cooperation with HAFSA Association

On the use of 'Long-Short Term Memory' neural networks for time series prediction

Pilar Gómez-Gil
National Institute of Astrophysics, Optics and Electronics
ccc.inaoep.mx/~pgomez

In collaboration with:
J. Carlos Galán Hernandez, Rigoberto Fonseca Delgado, Manuel Ramírez-Cortés, Vicente Alarcón-Aquino
PSIC group

This document is available at ccc.inaoep.mx/~pgomez/conferences/PggISCI14.pdf

Outline

- Prediction
- Recurrent neural networks
- Temporal Classification
- The LSTM network
- Applications of LSTM
- Results modeling sine function so far…
- Conclusions

About INAOE

www.inaoep.mx
ccc.inaoep.mx

- This is a public research center, with the aim of creating and spreading knowledge related to the areas of astrophysics, optics, electronics, computer science and similar fields.
- INAOE has a mission to contribute for the development of Mexico and the whole humanity, to solve real problems and to prepare advanced professionals in such areas.

Prediction

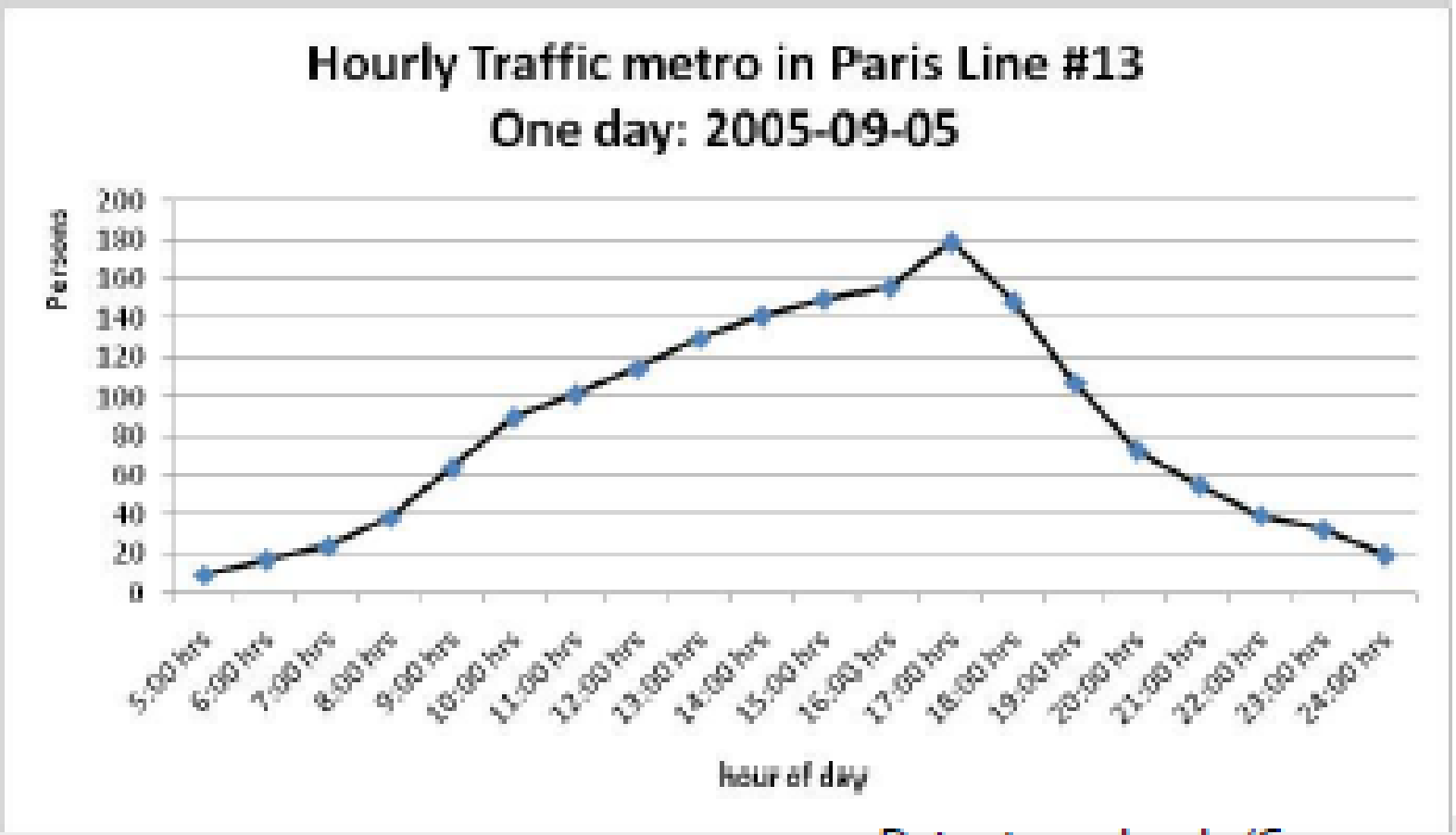
Why is it important to forecast a time series ?

- Because most of business and projects require some planning, which most of the time is performed with an uncertainty knowledge of future conditions,
- Because it is mandatory to measure the possible risks around future events,
- Because most off the time it is required to calculate some metric indices, which may be related to economy, politics, technology etc.

Time series

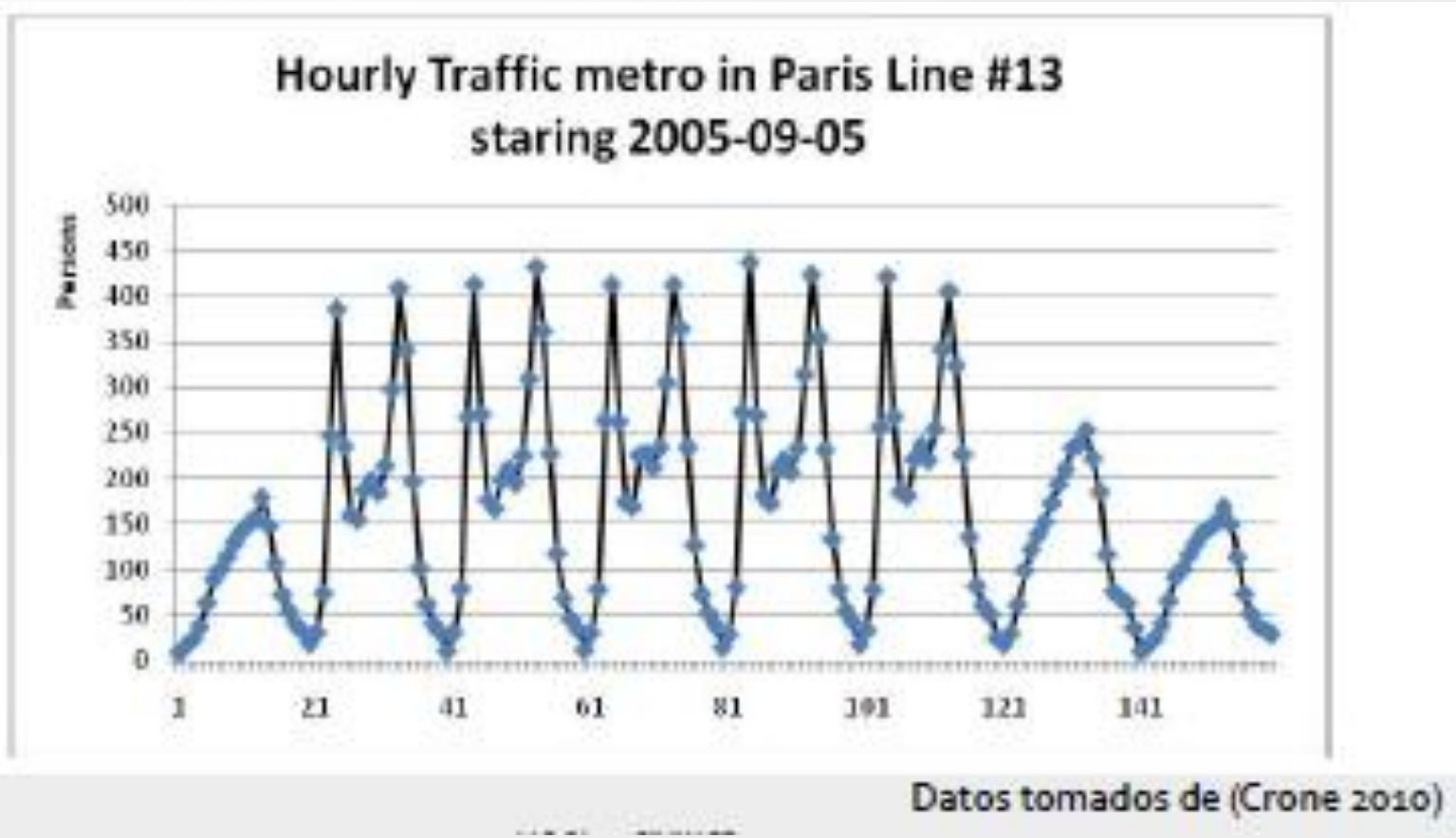
- A time series is a signal that is measured in regular time steps.
- The estimation of future values in a time series is commonly done using past values of the same time series.
- Notice that the time step may of a series may be of any length, for example: seconds, hours, days, years etc. This will bring on very different "looks" of the time series

Example: a time series measured each hour



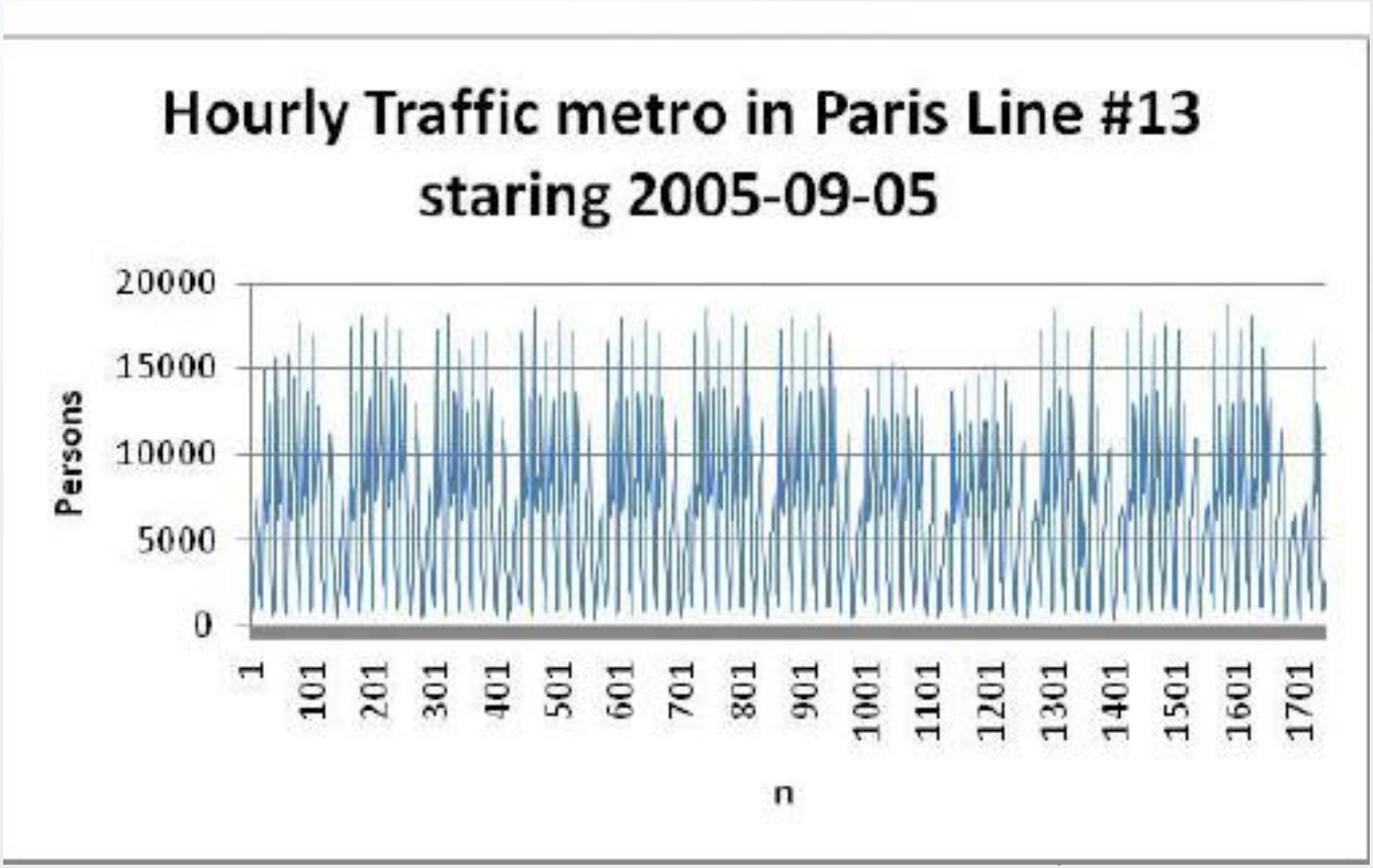
Datos tomados de (Crone 2010)

Few days of the same time series…

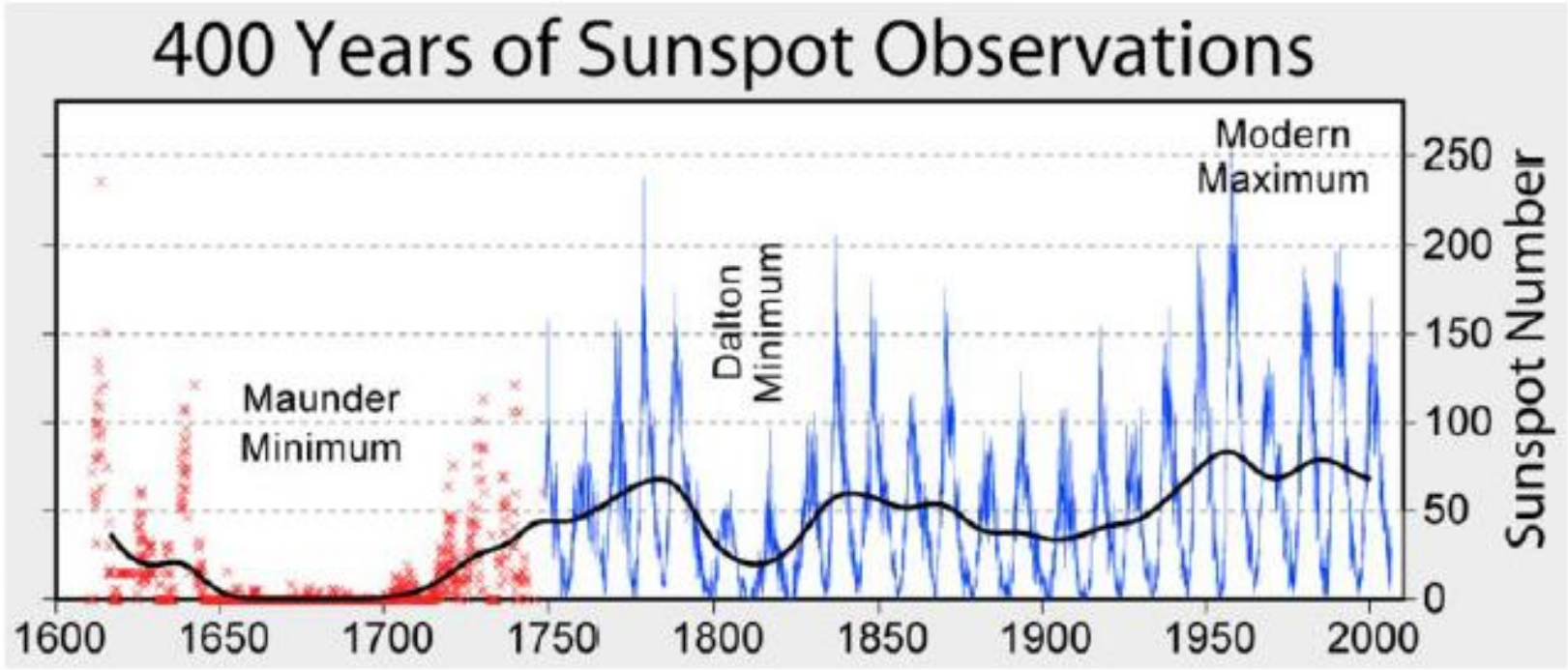


Datos tomados de (Crone 2010)

A few months of the same time series….



Four hundred years of measuring sunspots…



Fuente: http://www.carolmoore.net/articles/sunspot-cycle.html

Examples of Prediction Tools

- Linear models:
  - ARMA
  - ARIMA
  - Kalman filters
- Non-linear models:
  - Neural networks
  - Support vector machines
  - Fuzzy systems
  - Bayesian estimators

Prediction and Chaos

- A chaotic system presents some special characteristics (Kaplan & Cohen, 90):
  - A trajectory created by a chaotic system is non-periodic and deterministic,
  - It is highly dependent on initial conditions,
  - It is bounded by strange attractors (An attractor is a point or set of point where a trajectory is conducted, when the transient of such system ends).

Chaotic time series prediction

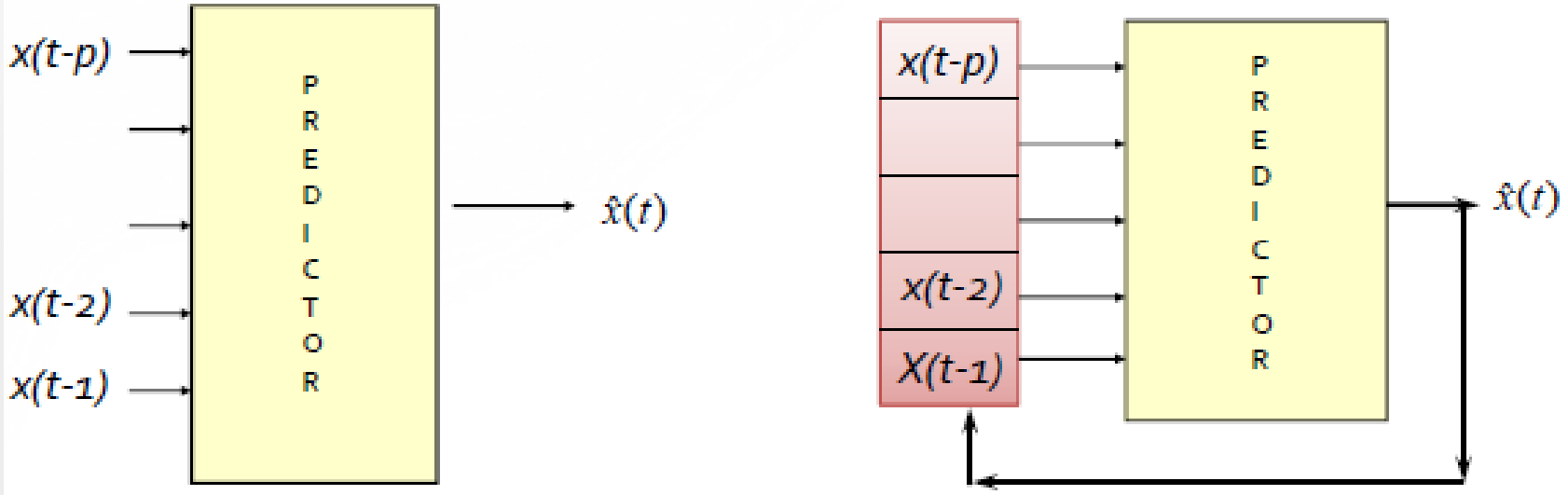
- To predict the behavior of long-trajectories created by a chaotic system is mathematically impossible.
- Even though, it is required for many applications to estimate in a reasonable way, the possible behavior of chaotic time series.

What is forecasting?

- Given a time series, forecasting refers to the process of calculating one of several values ahead, using just the information given by the past values of the time series.
- If no external values are used to calculate a time series, then it is supposed that all required information is located into the time series itself

Types of time series forecasting

- One-step prediction
- Several-steps prediction or long-time prediction



$$\bar{x}_t = \varphi(x_{t-1}, x_{t-2} \ldots x_{t-p})$$

Recursive prediction

- One-step prediction is calculated using one or several measured past values.
- To calculate several steps ahead, a predictor may use measured past values. However, if several future values are required to be calculated, then "recursive prediction" is used.
- Recursive prediction eventually uses values already predicted, instead of measured past values. This produces an accumulation of errors, which may grow very fast.
- In highly non-linear systems, this accumulation of errors may be an important problem

Recurrent Neural Networks

Types of Recurrent Neural Networks

- According to Kremer (2001),recurrent neural networks can be categorized into two classes:
  - Networks with a one-time input signal designed to enter an stable state.
  - Networks with time-varying inputs, designed to provide outputs in different points in time, known as dynamic neural networks.
    - These networks can be applied to the problem of identifying a subset of a language (sequence) in a string of discrete values

Sequence labeling

- It encompasses all tasks where sequences of data are transcribed with sequences of discrete labels (Graves 2012).
- Examples:
  - Speech
  - Handwritten recognition
  - Protein secondary structure prediction
- In these kind of problems, individual data points cannot be assumed to be independent.
- Both the inputs and the labels form strongly correlated sequences.

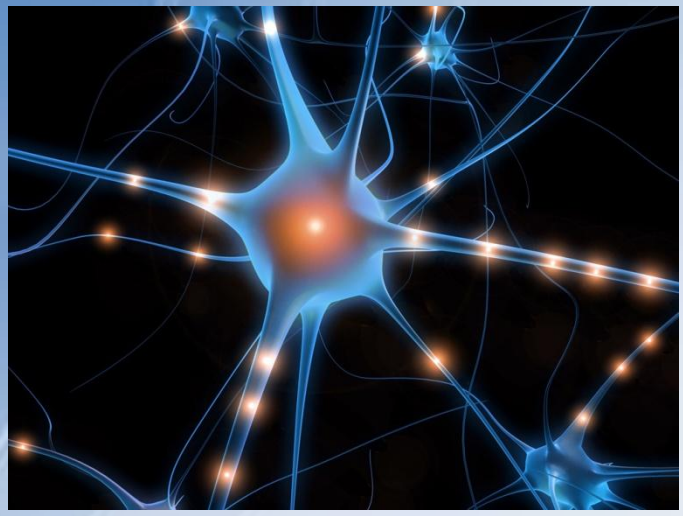
Example

- Identifying the sequence "1010" in a string

1	0	0	0	1	0	1	0	0	0	1	1	1	0	1	0	1	1	0	0
0	0	0	0	1	1	1	1	0	0	0	0	1	1	1	1	0	0	0	0

RNN and Sequence Labeling

- RNN are an attractive choice for sequence labeling because (Graves, 2012):
  - They are flexible in their use of context information, due to the fact that they can learn what to store and what to forget).
  - They accept many types and representations of data
  - They can recognize sequential patters in the presence of sequential distortions.
- The main drawback of RNN is that it is very difficult to get them to store information for long periods of time.

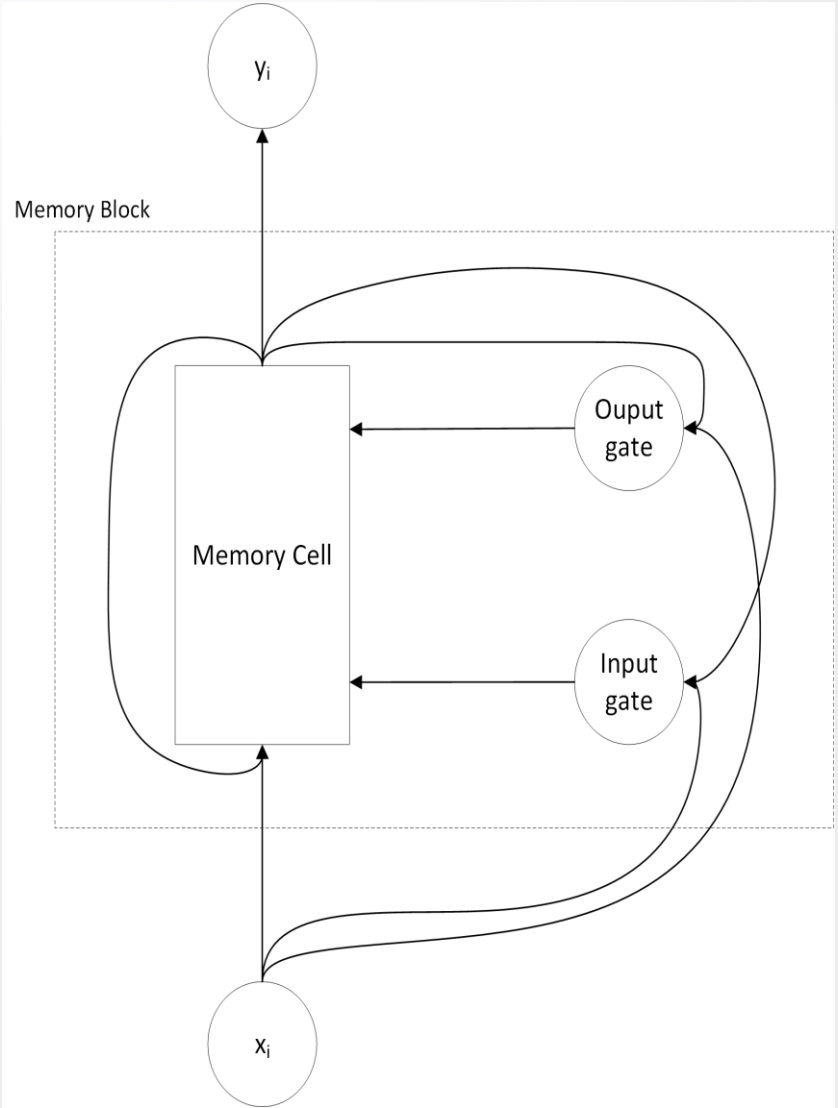


The LSTM network

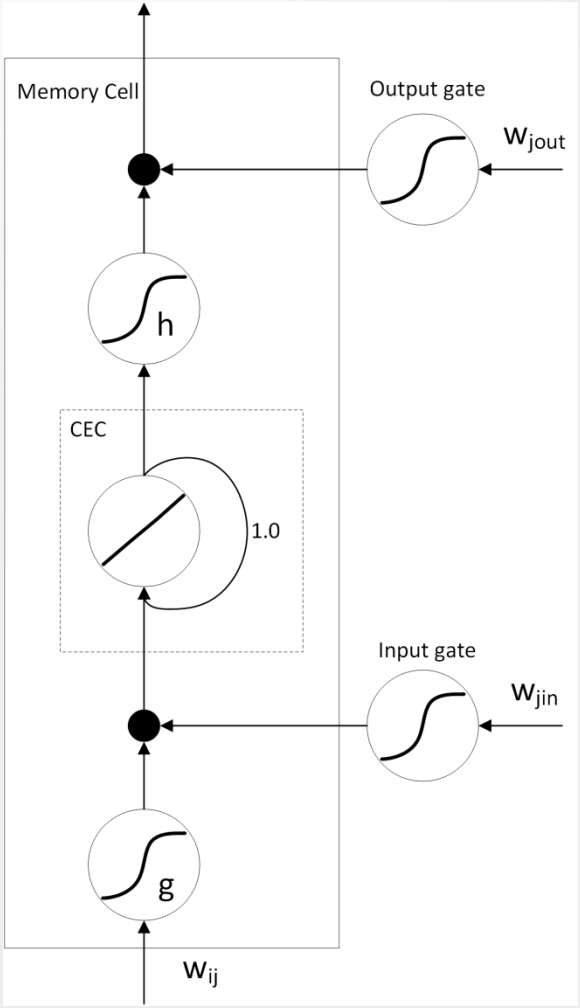
Long short-term memory (LSTM) network

- They were proposed by (Hochreiter and Schmidhuber, 1997)
- Long-Short Time Memory architecture consists of a set of recurrently connected subnets.
- The objective of the LSTM architecture is to overcome the problem known as "vanishing error problem".
- The vanishing error problem refers to how the influence of past inputs decays quickly over time.
- LSTM networks aims to solve this problem using memory cells.

Recurrent Neural Network with a hidden layer based on LSTM.



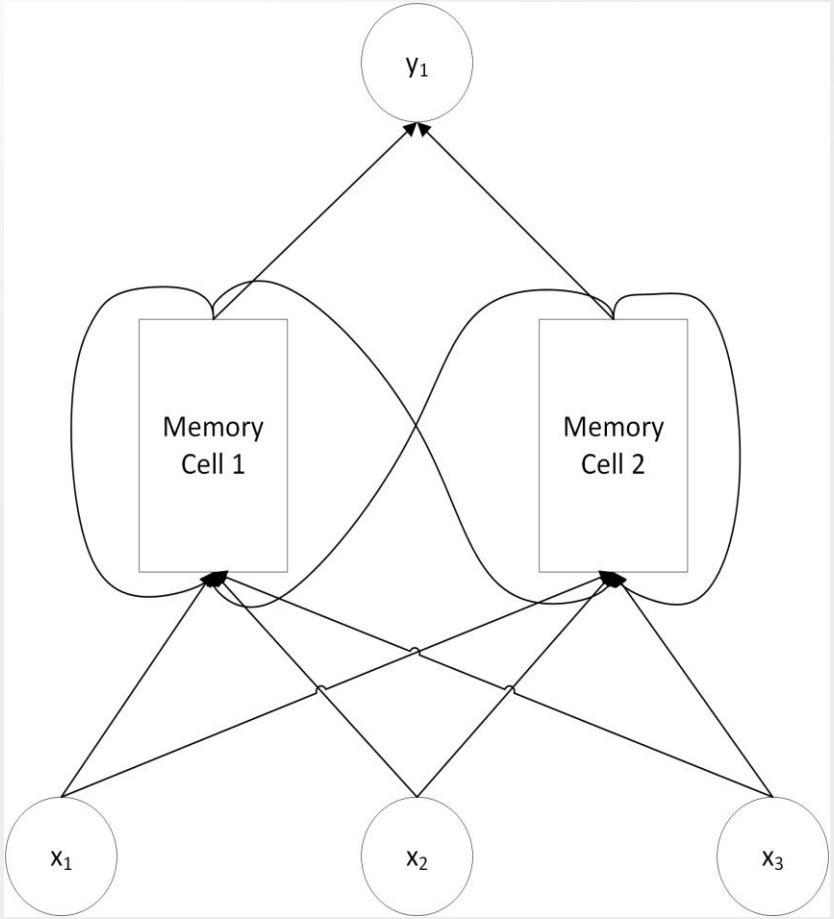
Basic structure of a LSTM cell



Calculation of LSTM output

- Because of the several recurrent connections, the update of the state of the network must be done in a particular way:
  1. Input gate activation
  2. Forget gate activation
  3. Cell input and cell state
  4. Output gate activation
  5. Cell output

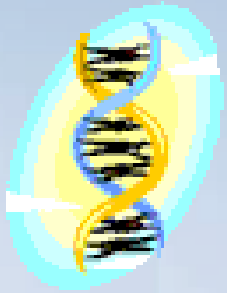
RNN with one hidden layer



Training algorithm

- The recommended training algorithm for LSTM is the un-truncated Back Propagation Through Time (BPTT) [6]. BPTT requires calculating the gradients of each cell and gate and the error associated with each cell output. This is known as the Backward Pass. The equation for calculating the error of the cell output is given in [2] and is defined as

$$\epsilon_{c_j}^t = \sum_u w_{kc_j}^v \delta_k^t + \sum_k w_{kc_j}^v \delta_h^{t+1}$$



Applications of LSTM

Applications of LSTM (1/2)

- Prediction of Reber's Grammar. The aim of this task was to predict the following character after a given sequence.
- Prediction of the next symbol of a noisy free and noisy sequences. LSTM is able to learn each sequence successfully
- Discrimination a sequence from an input with the sequence mixed with a noisy sequence.
- Adding a given sequence of real values between -1 and 1. The input size of the network was 1 as well as its output.

Applications of LSTM (2/2)

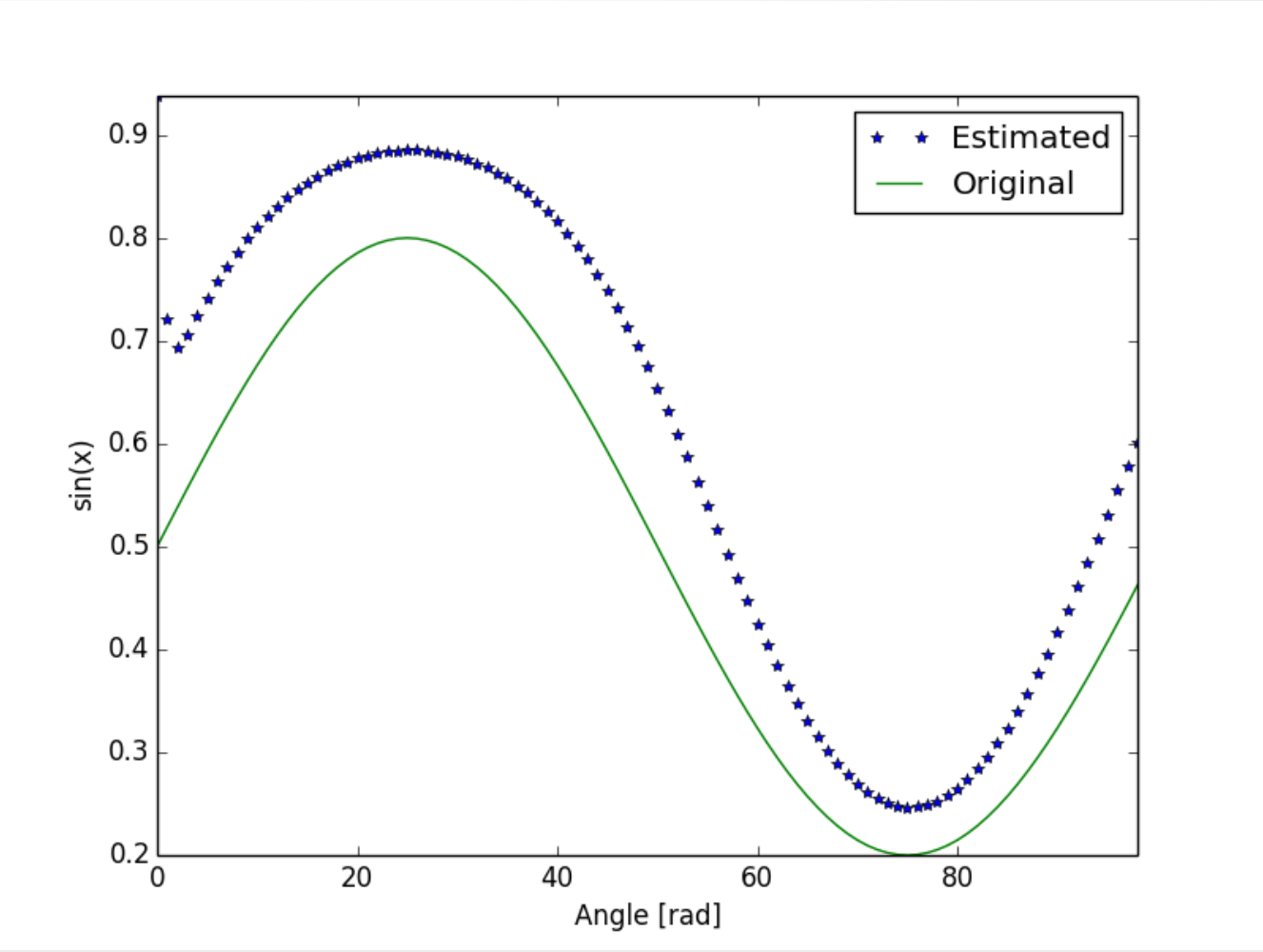
- Multiply all inputs on a given sequence and outputs the result.
- Learn temporal relations of particular inputs that are far apart on a given sequence. The LSTM networks can learn to predict which symbol is next to a given sequence of inputs that are related on time.
- Handwriting Recognition. LSTM showed good performance with Arabic handwriting recognition.
- Protein Localization of specific sub-cellular types known as Eukaryotic proteins on a given sequence.

Results modeling sine function so far…

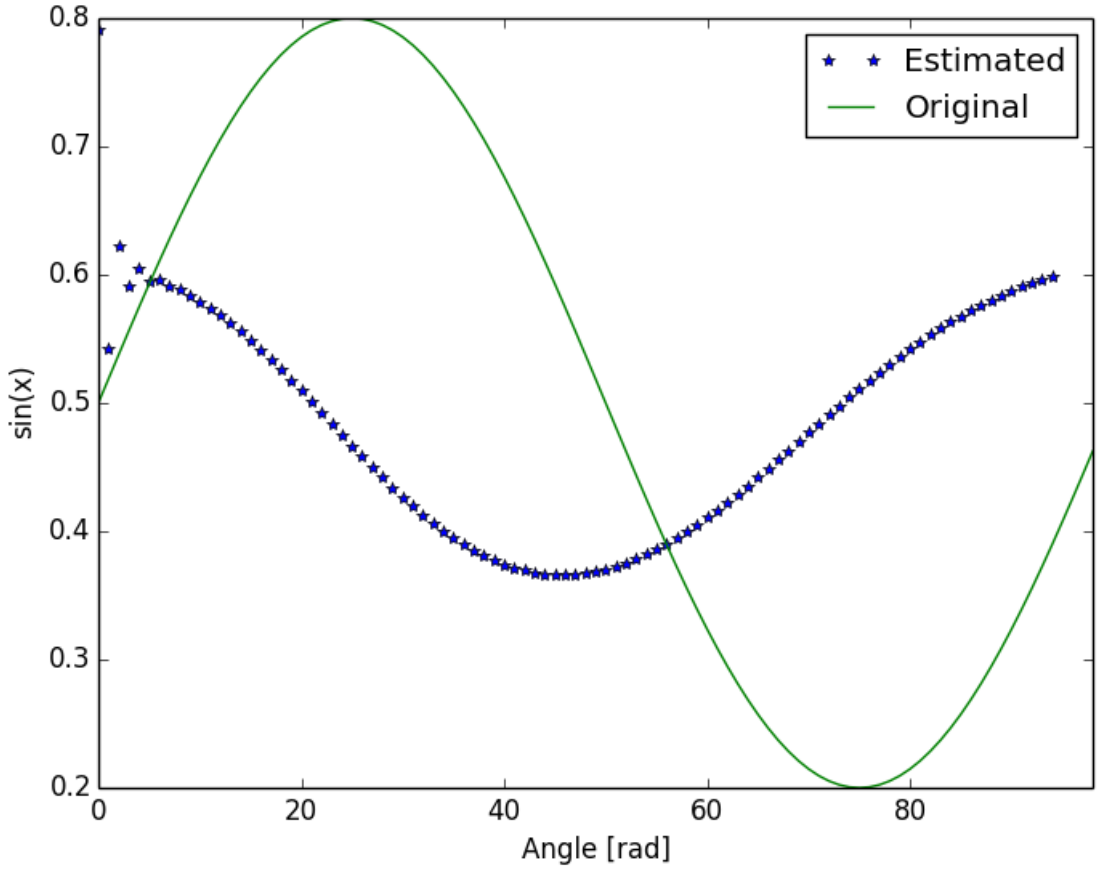
PyBrain

- PyBrain is a Python Language library for the use of LSTM and other neural networks implementations
- It requires a CPython implementation with a few extra modules.
- To have a functional PyBrain environment on any platform is recommended to install the Anaconda distribution of CPython (http://docs.continuum.io/anaconda/index.html).
- More information: http://pybrain.org

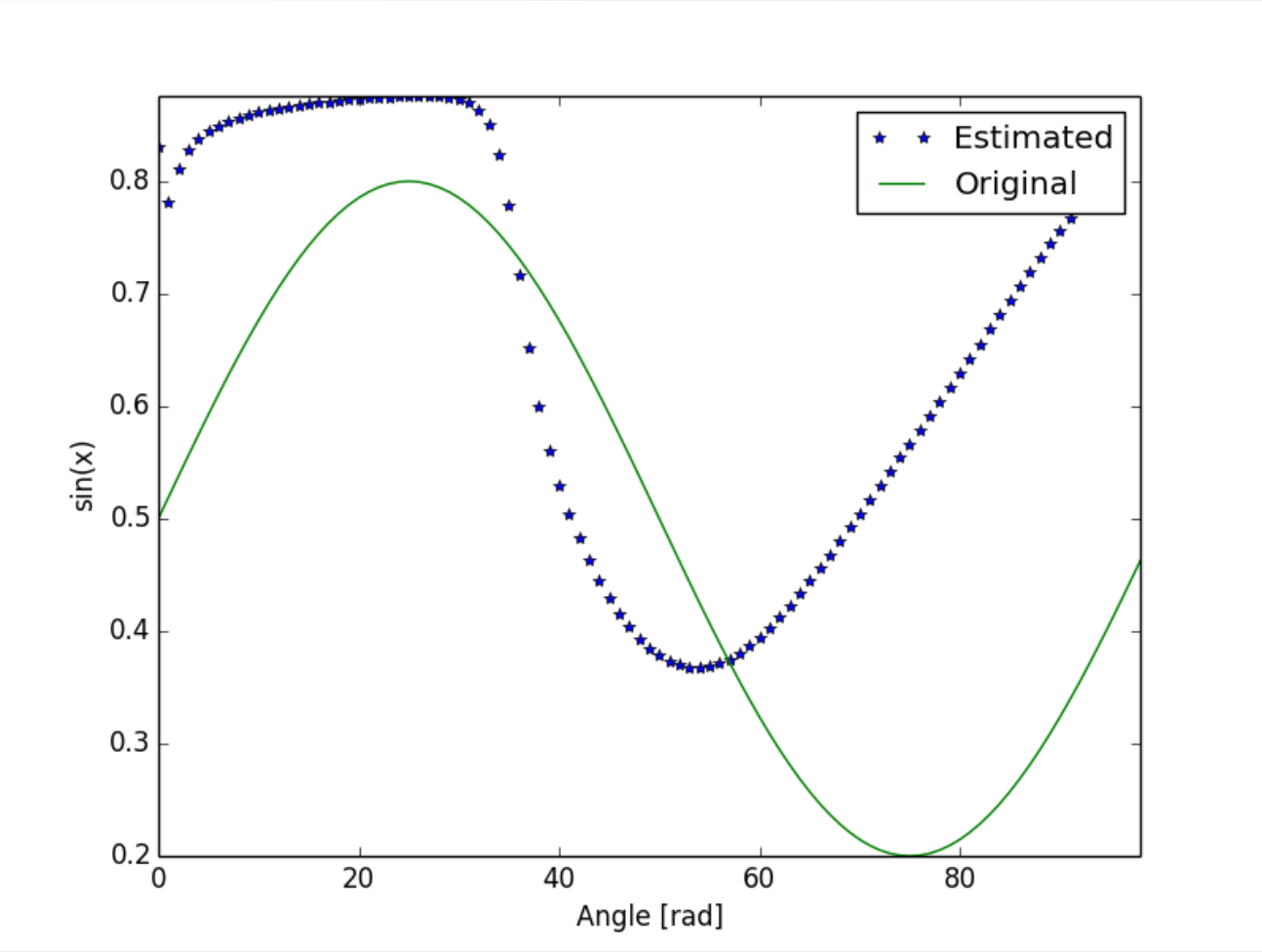
Results modeling sine function so far… (1/3)



Results modeling sine function so far… (2/3)



Results modeling sine function so far… (3/3)



Conclusions

- LSTM is a powerful tool that has showed be useful for sequence labeling and other time-related identifications
- LSTM is a complex RNN to program and to train for an specific task
- The use of LSTM for time series prediction may be too complicated to work in real problems,
- The use of "Pbrain" for LSTM is not straightforward.
- More experimentations is required, however, results so far show that other recurrent neural networks are more efficient that LSTM on learning a sine function, as the HCNN or the HWRN, or even plain RNN.

References

- Hochreiter, s., Schmidhuber, J. “Long Sort-term Memory.” Neural Computation, (1997), 9 (8), 1735-1780.
- Kremer, Stefan C. "Lessons from language learning." *Recurrent Neural Networks, Design and Appl* (2001): 179-204.
- Galán-Hernandez, C. “LSTM- Technical report project CONACYT-CB-2010-155250.” Oct.05, 2014. INAOE. México
- Gaves, Alex. “Supervised sequence labelling with recurrent neural networks”. Springer-Verlag, Berlin Heidelberg, 2012

TIJUANA INSTITUTE OF TECHNOLOGY

International Seminar on Computational Intelligence

IEEE CIS-MEXICO CHAPTER

IEEE - Computational Intelligence Society Mexico-Chapter

In cooperation with HAFSA Association

Thank you!

Pilar Gómez-Gil

pgomez@acm.org

ccc.inaoep.mx/~pgomez