Measuring similarities in contextual maps as a support for handwritten classification using recurrent neural networks

Pilar Gómez-Gil, PhD

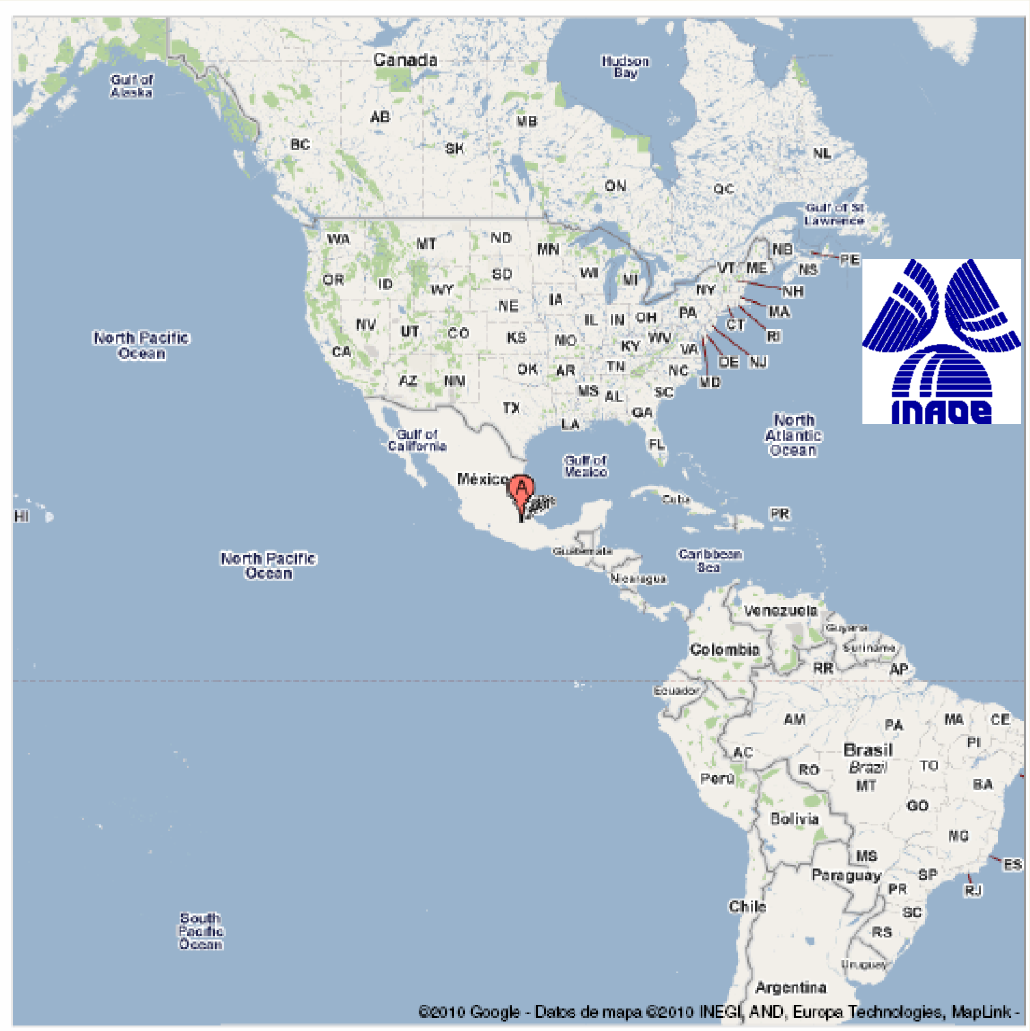
National Institute of Astrophysics, Optics and Electronics (INAOE)

Tonantzintla, Puebla

ISCI'2012

IEEE Computational Intelligence Society

Tonantzintla



inaoep.gob.mx

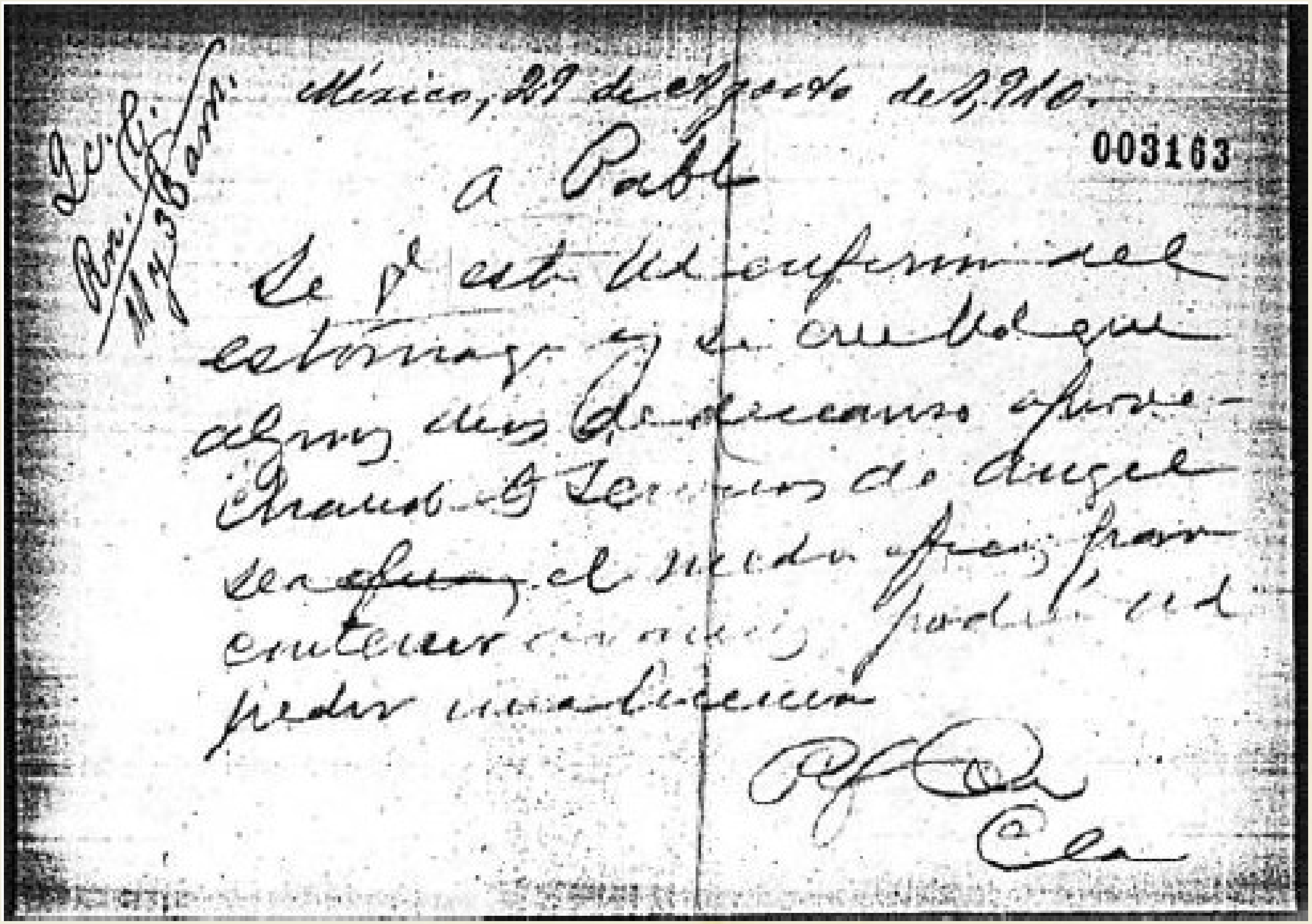
A view from office…

Outline

1. Few words about the problem of handwritten recognition
2. A system for temporal classification
3. Self organizing map
4. Simple recurrent network
5. Experiments
6. Conclusions and future work
7. References /future readings

¿Is it possible to automatically read this?

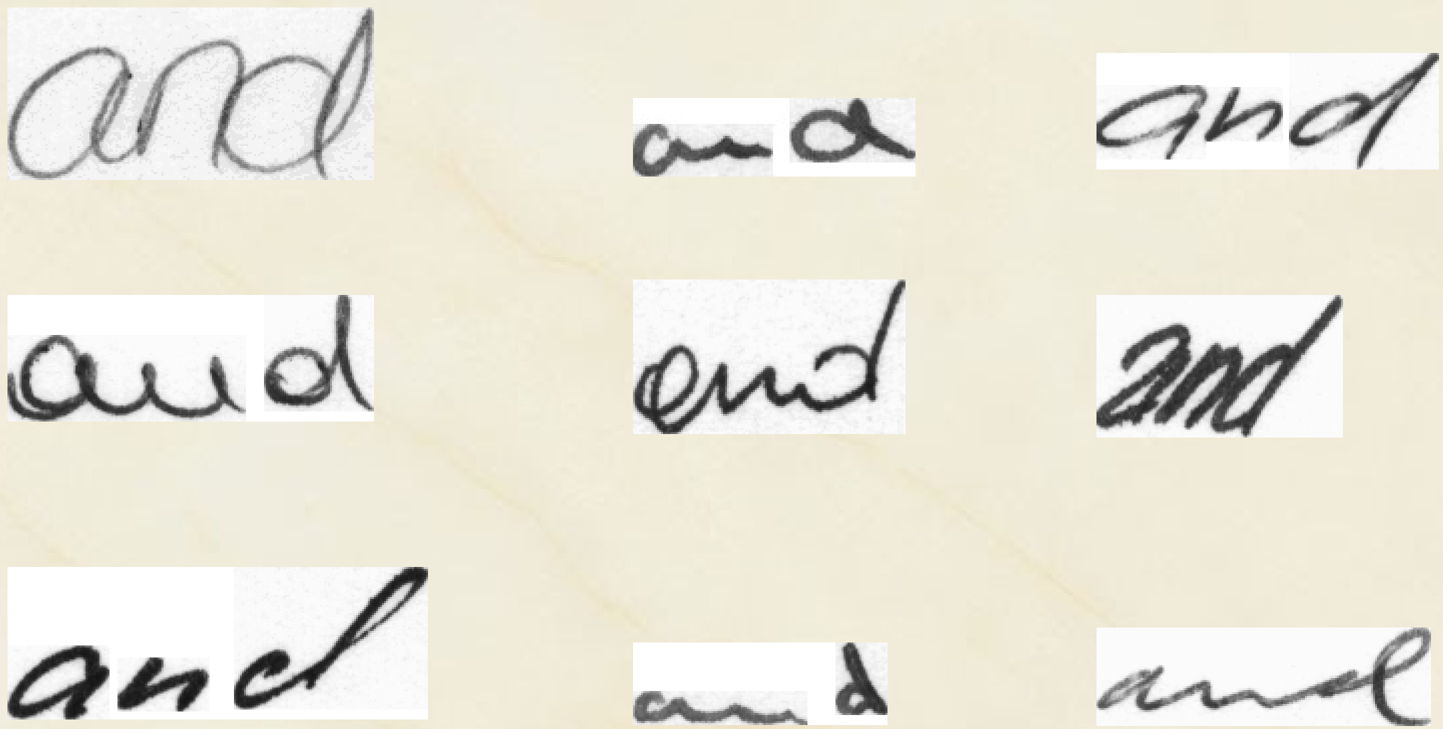
(Gomez-Gil et al. 2007)


003163

Few words about handwritten recognition

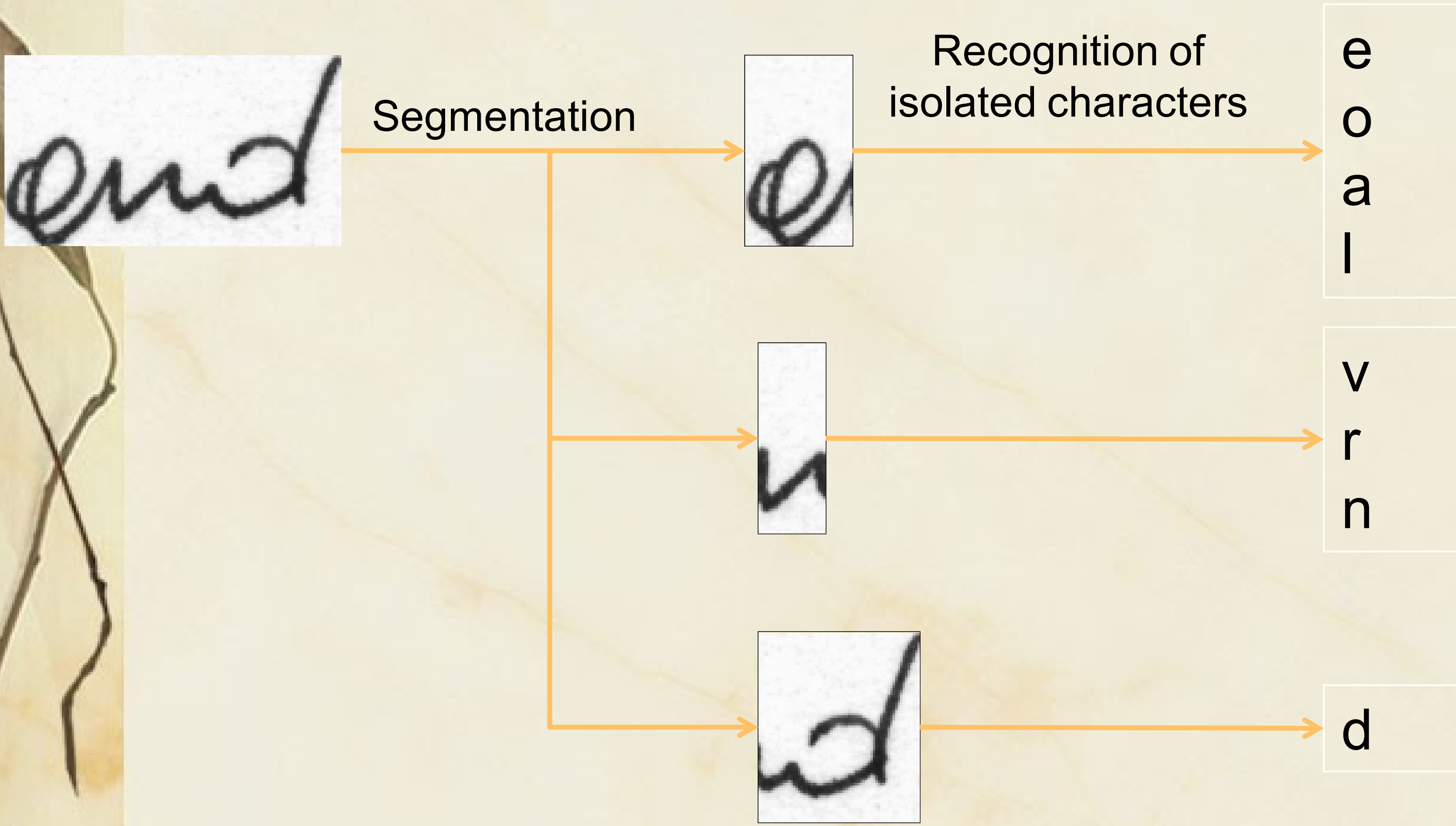
- There is a huge amount of non-digital manuscript documents that are required to be read and translated to digital form.
- Important problems are faced during off-line, write-independent manuscript recognition, as:
  - Different writing styles
  - Segmentation issues
- Automatic reading of manuscripts may be handled by:
  - Character recognition
  - Word recognition
  - Text line recognition

Examples of different writing styles for the word “and”



From IAM database (Marti & Bunke 2002)

The main problem using character recognition

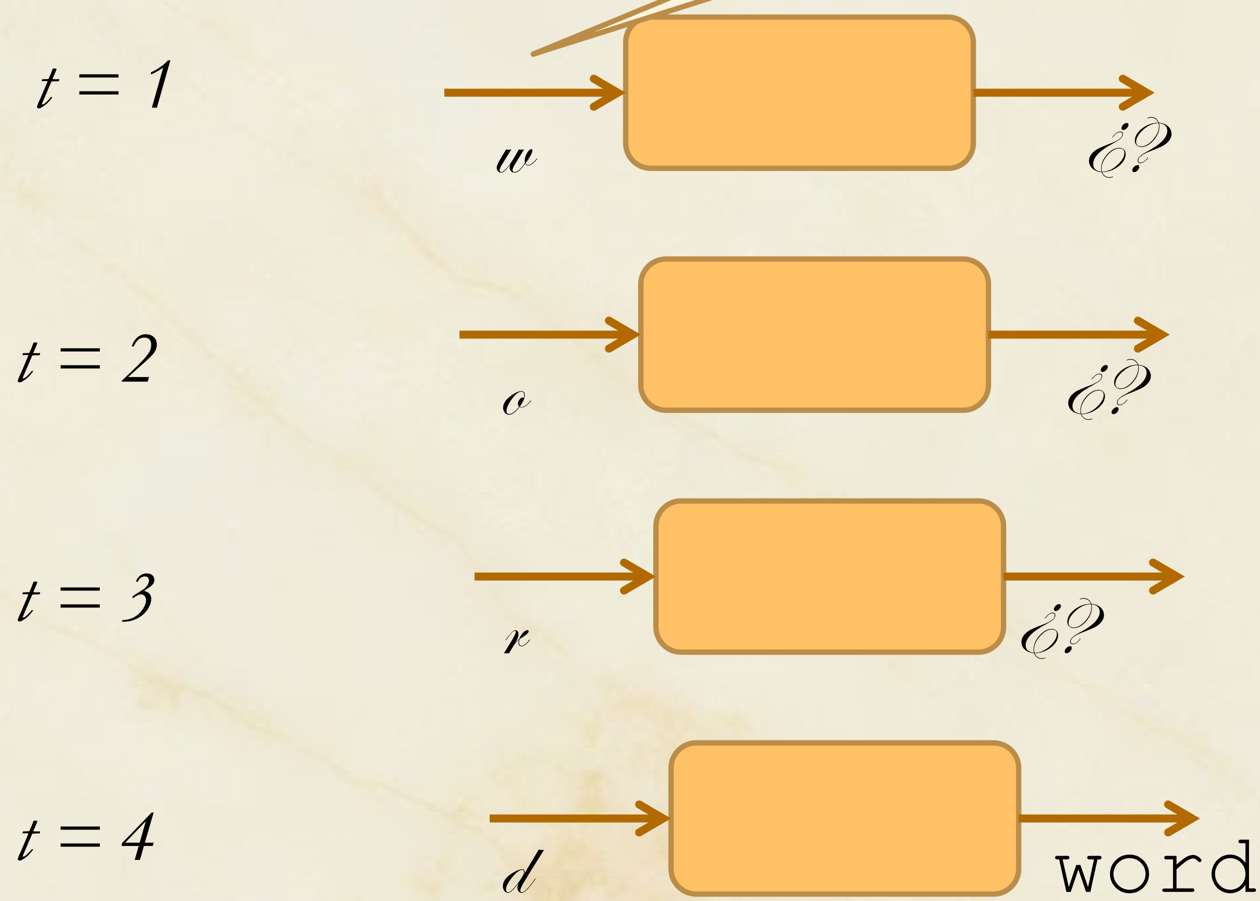


About Word recognition

- Word recognition consists of finding the word that is most compatible to a specific image, with respect to a previously defined lexical set (Vinciarelli 2002).
- Word recognition may be treated as a temporal classification problem, that is, a problem where the assigned class depends upon a sequence of events occurring in the past. (cont.)

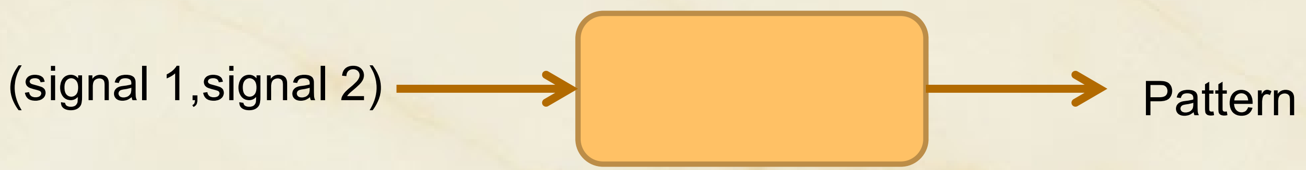
A system for temporal classification

Each event is fed one at the time



Identifying patterns in sequences

Signal 1	1	2	2	1	2	1	1	1	1	2	1	2	2	1	1	2	2	1	2
Signal 2	2	1	2	2	2	2	1	1	2	1	2	2	1	2	2	1	2	2	1
Pattern	0	1	0	0	0	0	1	0	0	1	0	0	0	0	0	1	0	0	1



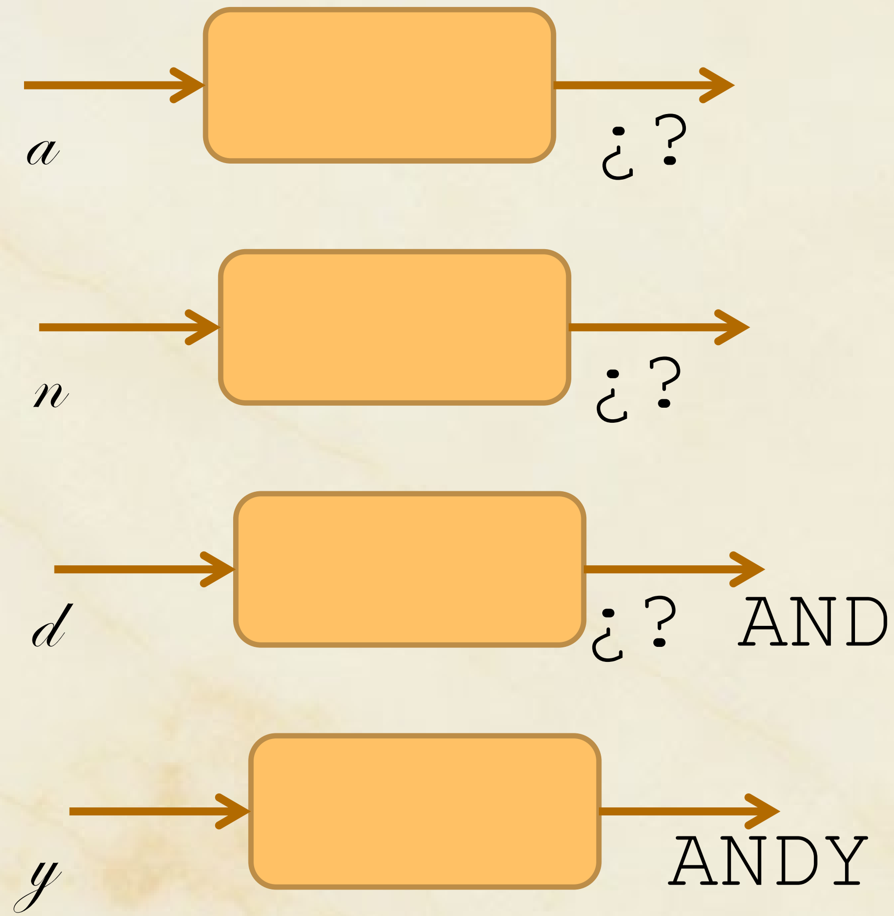
What about this?

$t = 1$

$t = 2$

$t = 3$

$t = 4$

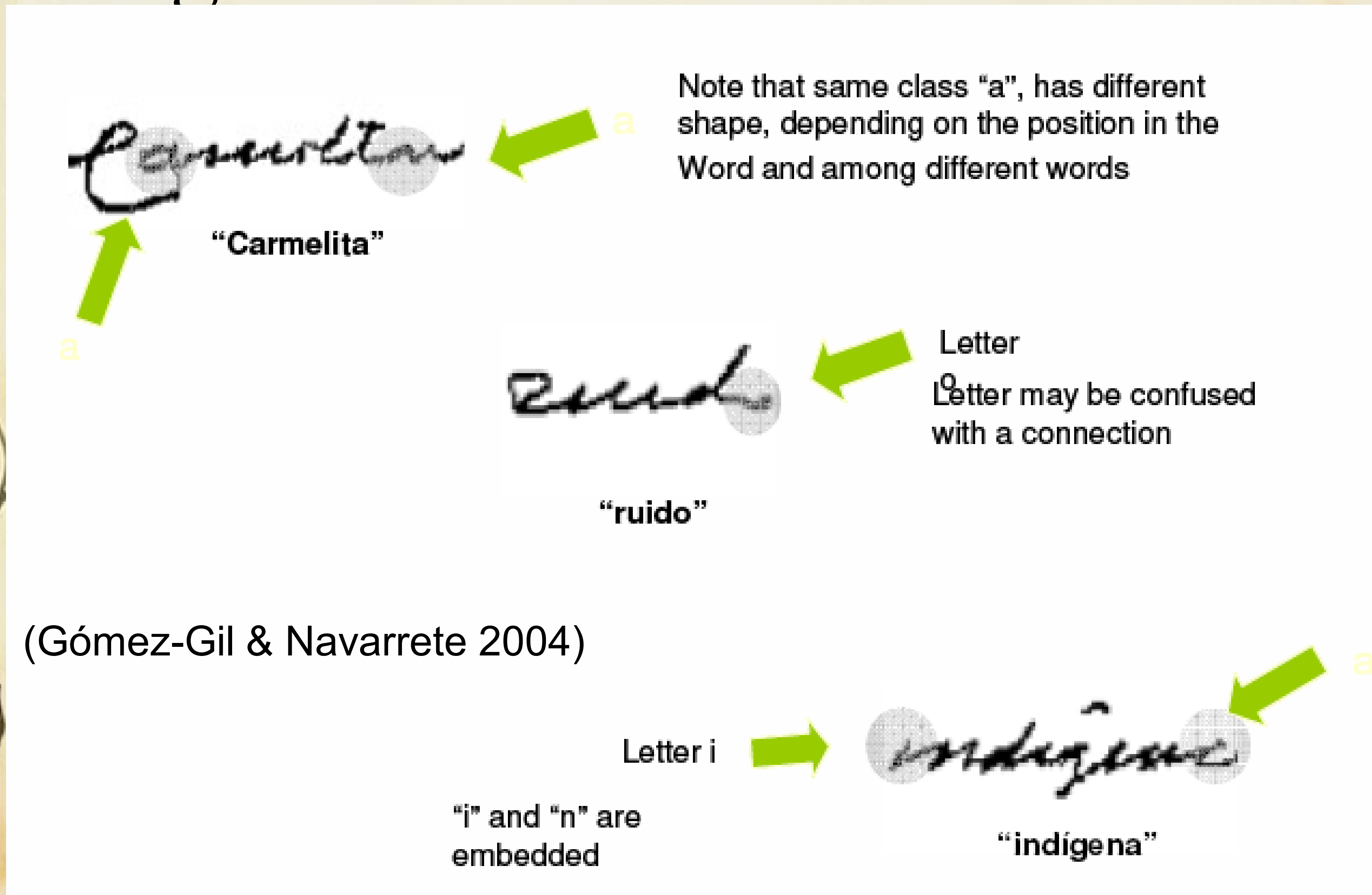


Words as sequences

- To identify words as sequences, each element of the sequence has to be defined. This may be a character, but…
- Sayre paradox:

"*a character can not be segmented before being recognized, and it cannot be recognized before being segmented*"

Segmentation issues

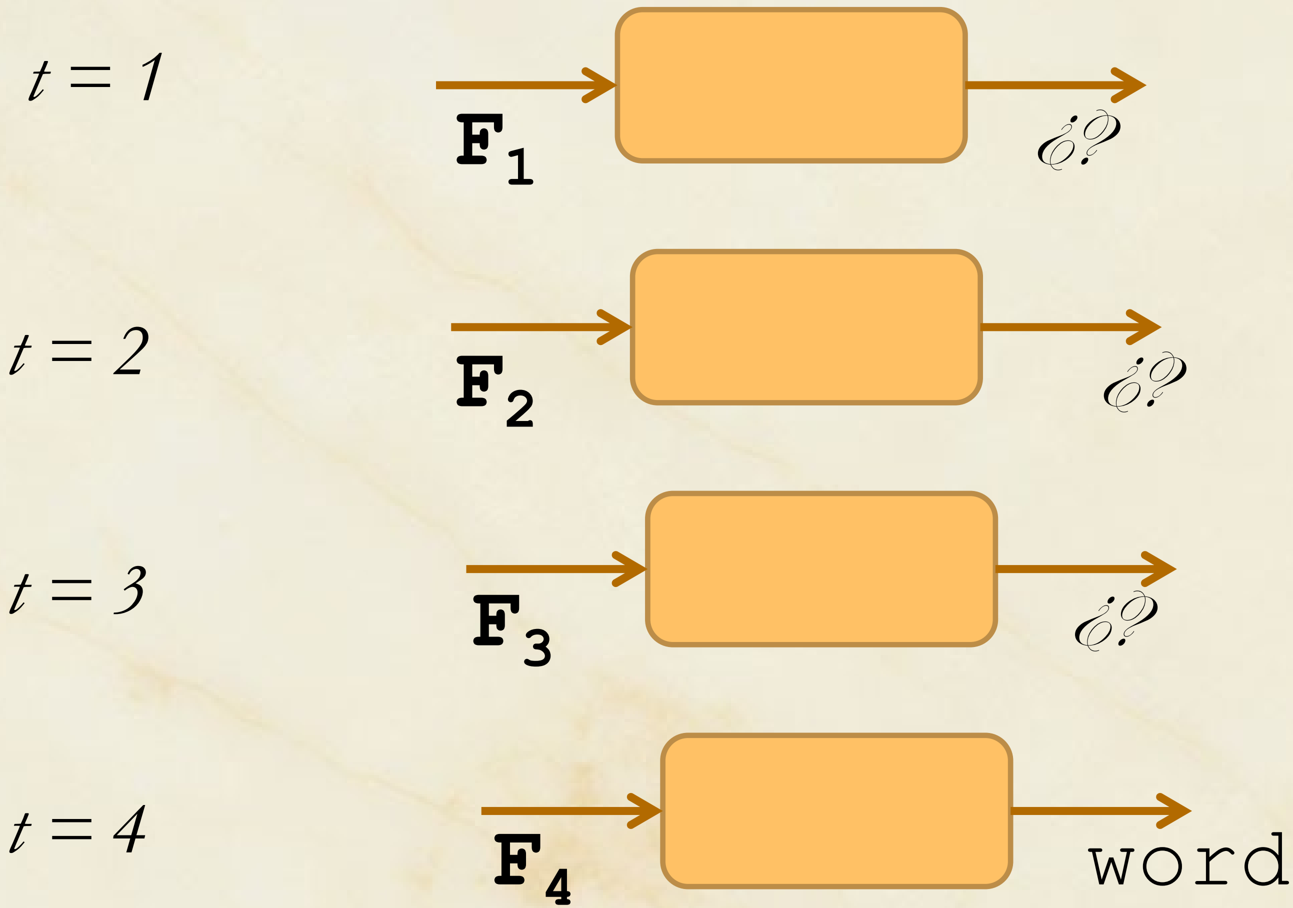


(Gómez-Gil & Navarrete 2004)

Segmentation issues (cont.)

- We may try to segment a word in ¨chunks” that may or may not be characters, but that are fairly consistent in a word.
- Chunks of different writers using different words, may be clustered in a unsupervised way, creating prototypes.
- Given a chunk, we can identify their k-most similar prototypes and measure these relationships.
- This information will create feature vectors $\mathbf{F}_i$ to be used in temporal classification

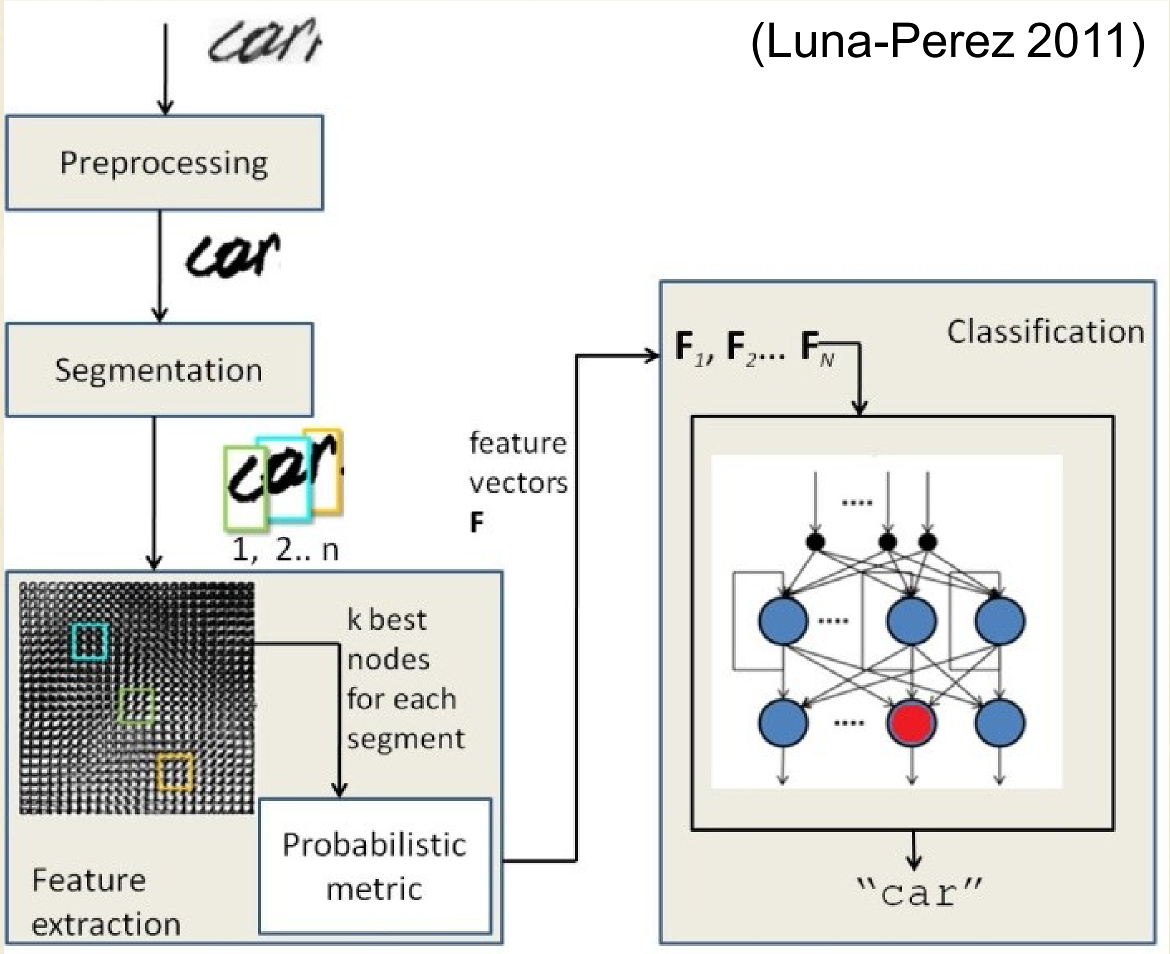
2. A system for temporal classification



Components of such classifier

- Chunks (segments) information represented in a self-organizing map (SOM) trained with segments of words,
- A metric to represent the probability of a segment to belong to a specific cluster defined by the SOM and to its neighbors,
- A simple recurrent network (SRN), which memorizes the temporal relationships among all segments

Architecture of this classifier



Preprocessing

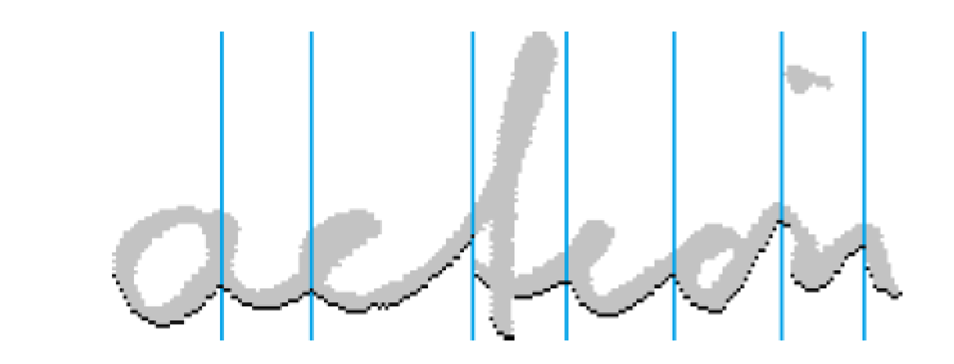
- Noise elimination
- Binarization
- Slant correction

before after

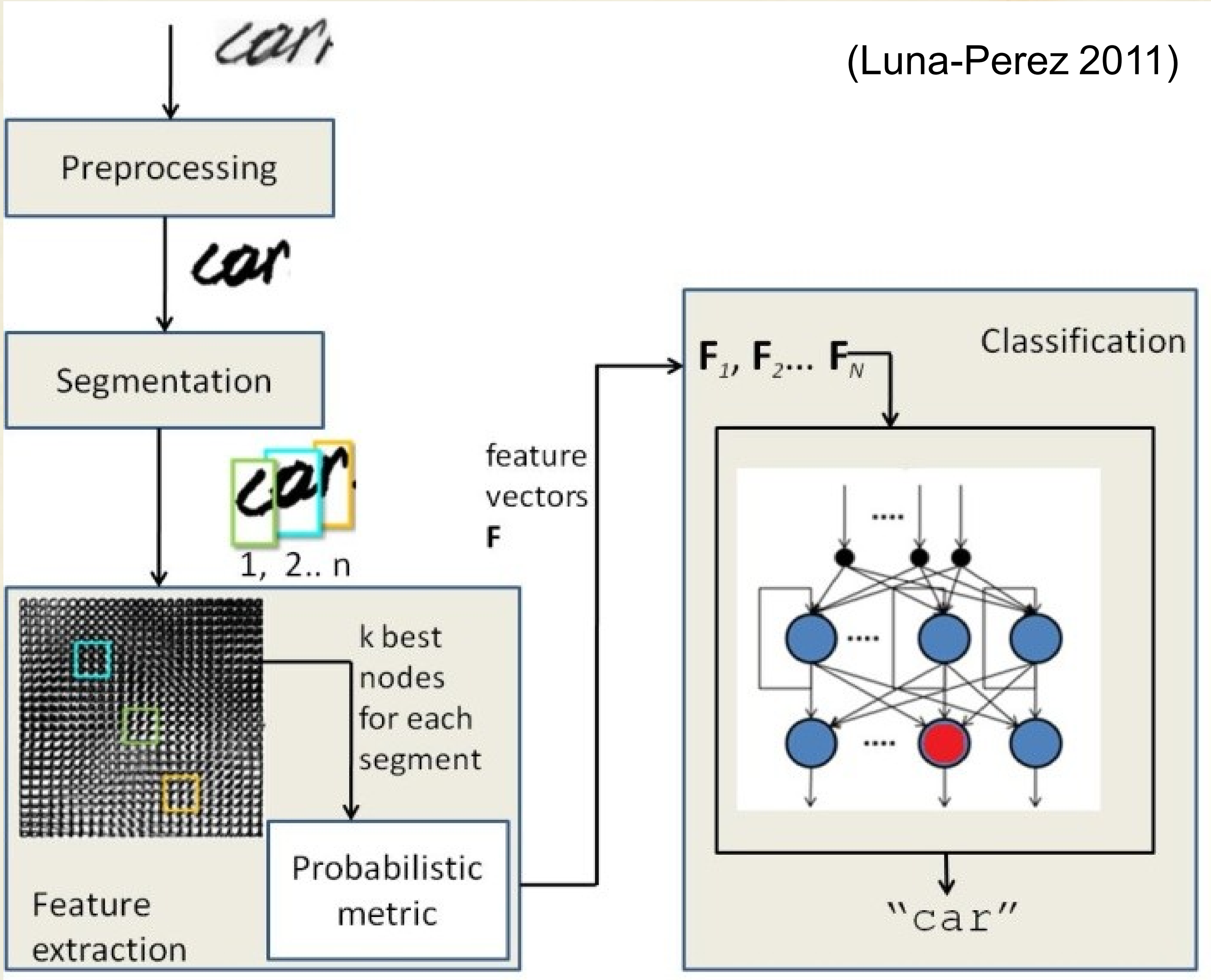
(Luna-Perez 2011)

Segmentation

- The lowest pixel in each column is found
- A cut point is located where the pixel is higher than their neighbors



(Luna-Perez 2011)


(Luna-Perez 2011)
Preprocessing
Segmentation
1, 2.. n
k best
nodes
for each
segment
Probabilistic
metric
Feature
extraction
feature
vectors
F
Classification
F_1, F_2... F_N
"car"

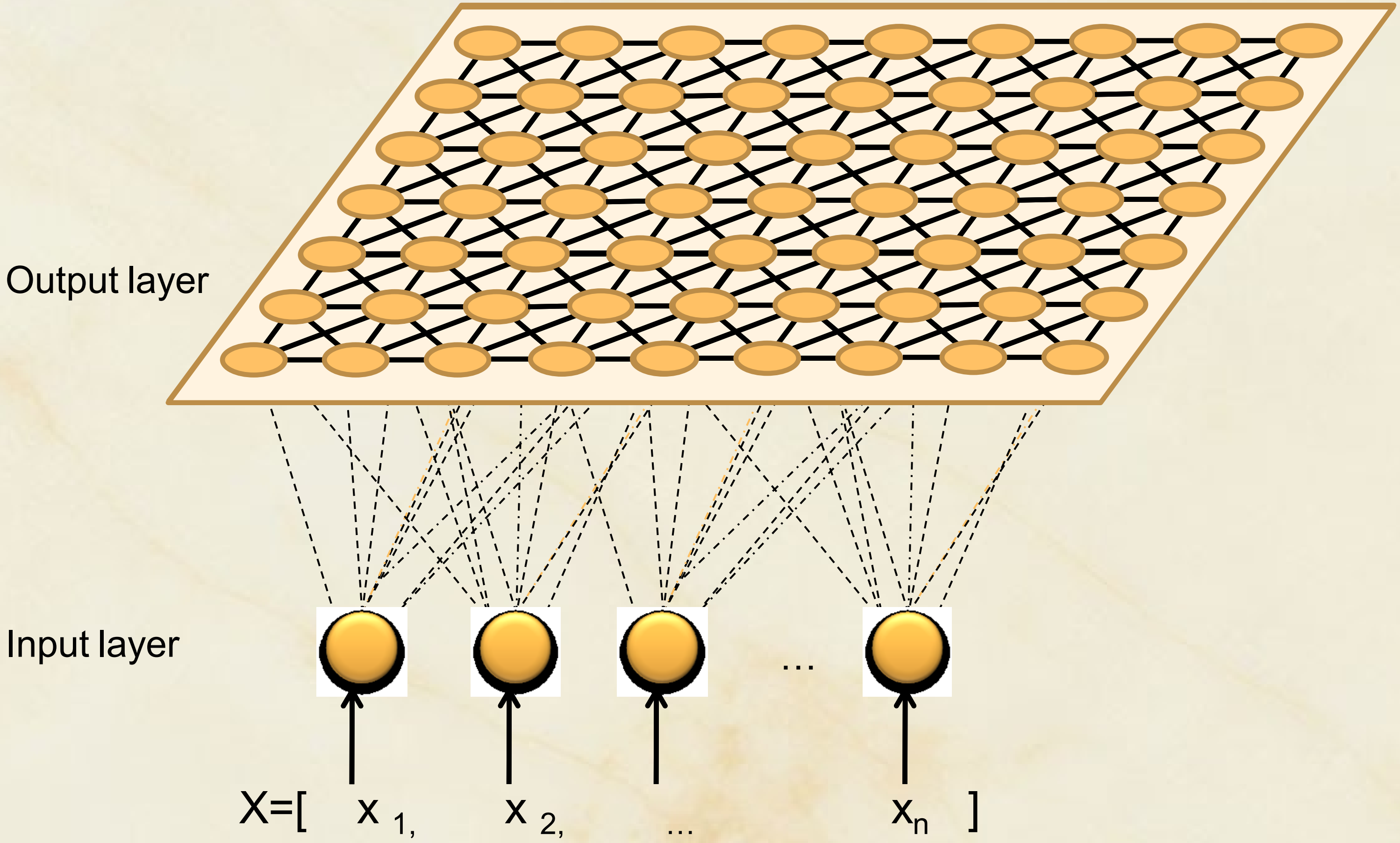
3. Self-Organizing Map (1/2)

(Kohonen 2001)

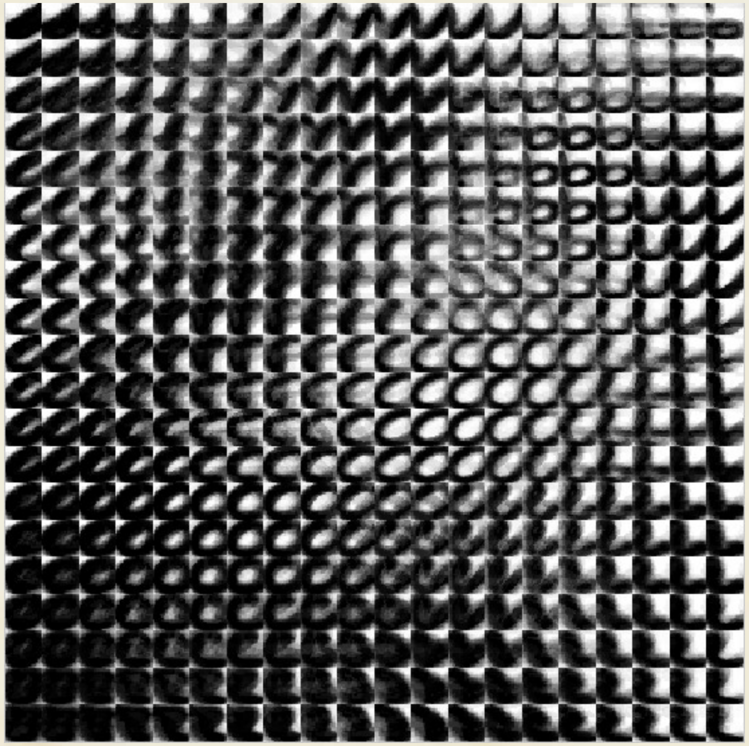
- SOM is able to generate groups with no supervision.
- Once trained, the output layer of a SOM forms a two-dimensional map, where each node contains a prototype of a cluster.
- Each neuron represents a cluster similar to clusters represented by neuron's neighbors.

Self-Organizing Map (2/2)

(Kohonen 2001)



Clustering segments



Feature map generated by a SOM using several “types” of segments (Luna-Perez, 2010)

Generation of feature vectors $\mathbf{F}_i$ *(1/3)*

1. Each segment t *is fed to SOM, getting a winning* neuron c_t , a 2-D vector containing the coordinates of the winning neuron
2. the k-1 *neurons, each identified as* m_{ti} $i=2..k$, with highest activations in the map are also identified.
3. Let $m_{t1} = c_t$ *(neuron with highest activation)*

More…

Generation of feature vectors $\mathbf{F}_i$ *(2/3)*

4. A measure of the probability of each of these neurons to represent best the segment is calculated as:

where: $$p_t(\mathbf{m_{ti}}) = \frac{\exp(-\|\mathbf{m_{ti}} - \mathbf{c_t}\|)}{\sum_{j=1}^{k} \exp(-\|\mathbf{m_{tj}} - \mathbf{c_t}\|)}$$

c_t is a 2D vector defined by the coordinates of the winning neuron at SOM

m_{ti} is the 2D vector defined by the coordinates of the *i-neuron with highest activation at the SOM, i= 1..k, obtained* when segment *t* is applied.

Generation of feature vectors $\mathbf{F}_i$ *(3/3)*

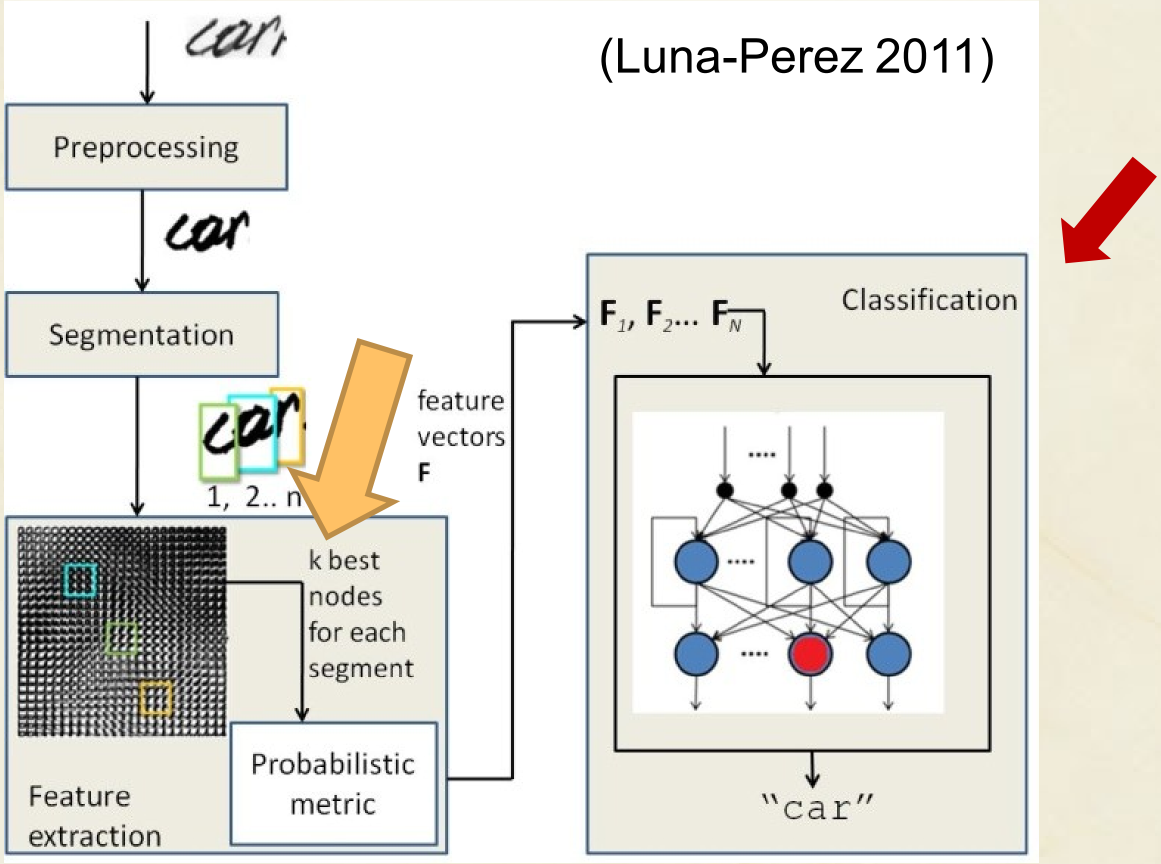
- Then feature vector F_t for each t- *segment is* defined as:

$$F_t=(m_{t11}, m_{t12}, p(m_{t1}), m_{t21}, m_{t22}, p(m_{t2}), \ldots, m_{tk1}, m_{tk2}, p(m_{tk}))$$

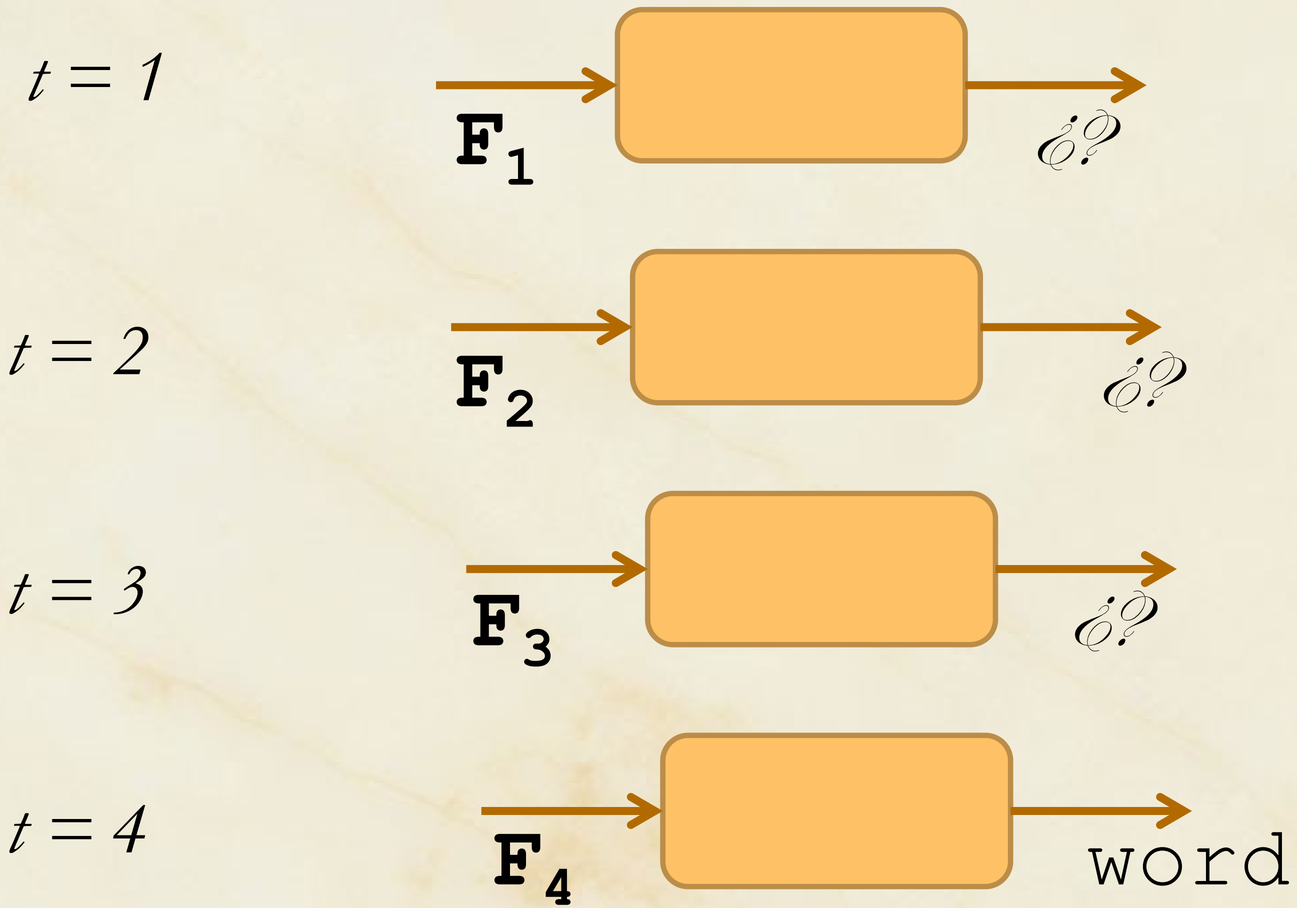
- F_t is $3k$-dimensional

  (k is a free parameter of this system, representing the number of best nodes in SOM involved)

The classifier



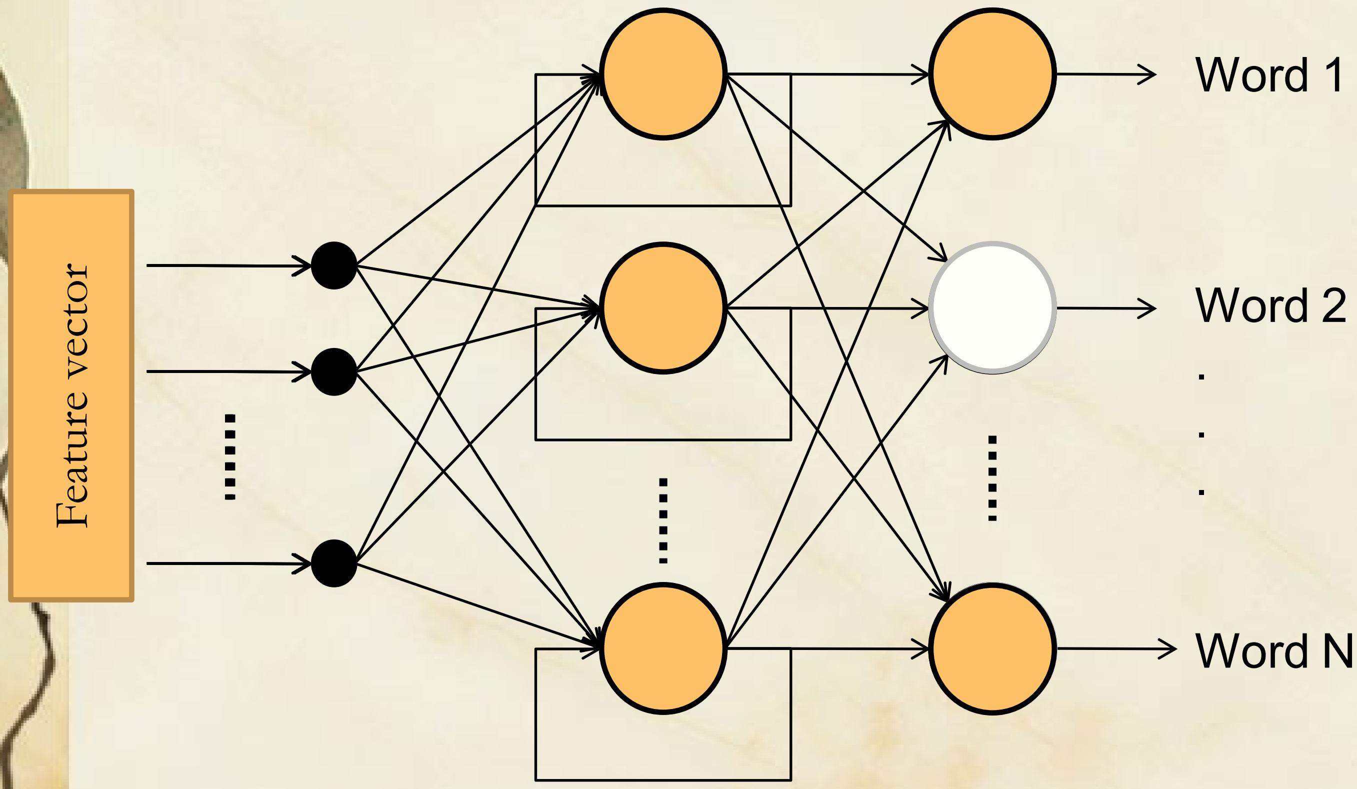
temporal classification



Classification

- It is carried out using a Simple Recurrent Network (SRN) (Elman 1990)
- The algorithm back propagation through time is used to train the SRN (supervised training).
- The SRN network used here consists of 3 layers:
  - an input layer with *3k neurons*
  - *a hidden layer with recurrent* connections
  - an output layer with as many nodes as the number of words to be recognized.
- Each feature vector, corresponding to each segment in the world, is fed to SRN once at the time, generating a network´s output once at the time

4. Simple Recurrent Network (SRN) (Elman 1990)



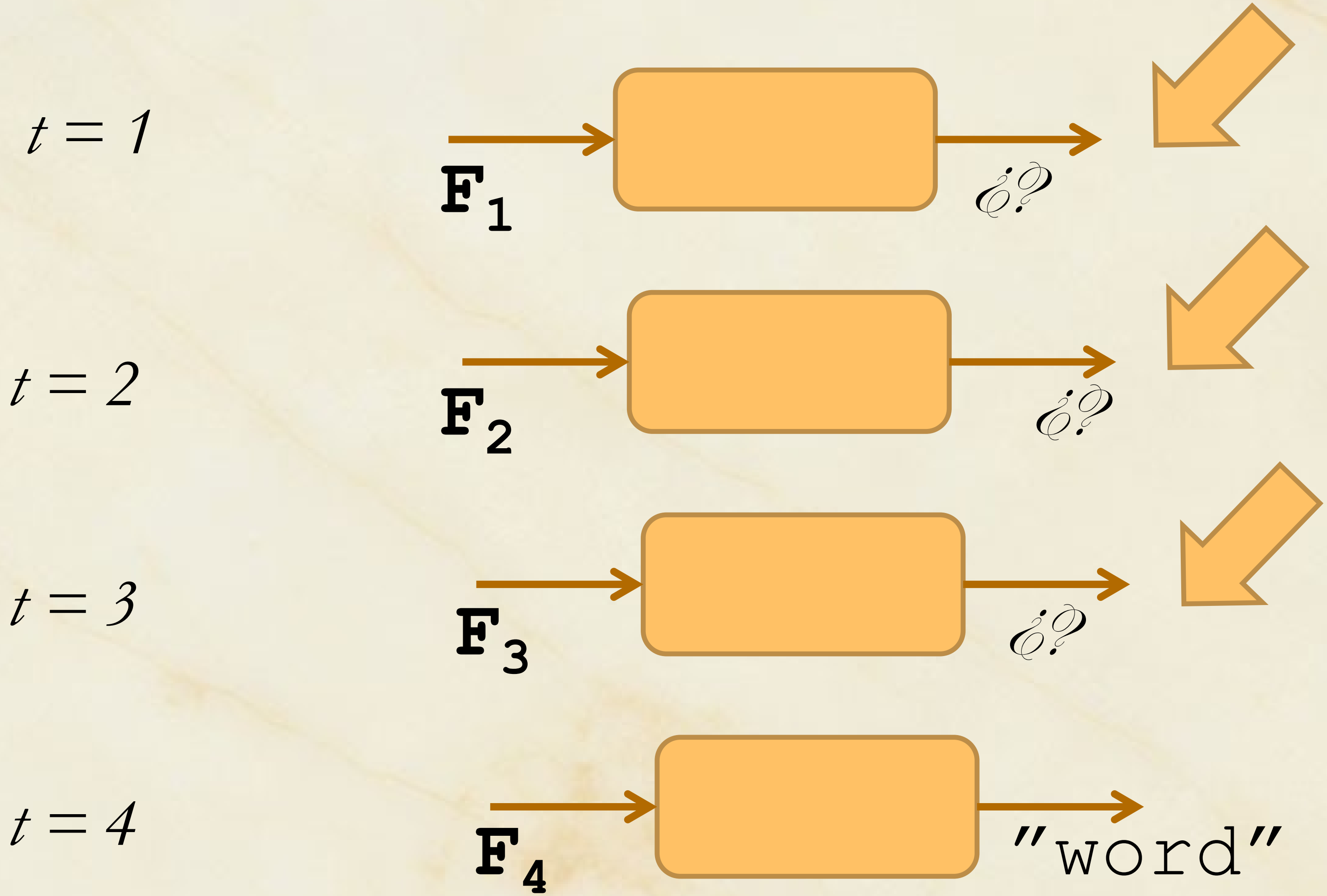
Other Characteristics of this SRN

- Output nodes use the activation function *Softmax:*

$$softmax(out) = \frac{\exp(out_i)}{\sum_{j=1}^{n} out_j}$$

- SRN is trained to receive each segment F_t of each word *w*.

Desired values for each sequence? (1/2)



Desired values for each sequence? (2/2)

1. The desired value of each output node i when segment F_t is input, is calculated as

$$d_i = \begin{cases} x/n & \text{if } i = \text{corresponding word of } F_t \\ 0 & \text{otherwise} \end{cases}$$

x is the position of segment F_t in the word and n is the number of segments in the word.

5. Experiments (1/2)

- The proposed method was tested using a lexical with 10 words taken from database IAM [5]

a	**and**	**are**	**as**	**at**	**be**	**but**	**bye**	**can**	**for**

- The proposed method (identified as SOM-SRN) was compared with other two neural classifiers:
  - *A feed-forward network (identified as FF in table)*
  - *FF-SOM network. (identified as SOM-FF in table)*

- 20 experiments were executed for each network case
- 150 different configurations for each network were tested in order to find the best configuration.

Experiments (1/2)

- Performance was measured in two ways:
  1. Classification Error (a percentage, best value = 0)

$$error = \frac{number\ of\ words\ incorrectly\ classified}{total\ number\ of\ words} \times 100$$

  2. Word accuracy metric (Graves et al. 2009) (best value = 1)

$$WA = 100 \times \left(\frac{insertions + substitutions + eliminations}{size\ set} \right)$$

Results (2/2)

Classification error obtained by the three cases

Case	Error in training set	Error in testing set
FF	46.95% ±27.32%	68.80% ±14.68%
SOM-FF	7.93% ±1.68%	36.80% ±5.13%
SOM-SRN	**5.75%±1.34%**	**24.30%±5.12%**

Word accuracy obtained by the three cases

Case	Word Accuracy using training set	Word accuracy using testing set
FF	55.52 ±27.14	32.92 ±15.51
SOM-FF	93.03 ±1.5	66.21 ±5.85
SOM-SRN	**95.42±1.21**	**78.25±3.25**

6 .Conclusions

- This classifier is based on the use of three main components:
  - a feature extractor based on non-supervised clustering,
  - measures of the probability that a segment of a word belongs to the *k most probable clusters,*
  - *a SRN* able to classify sequences of features representing the words.
- This method showed to overcome two other neural classifiers when tested over a set of 10 words, that were written by different people and showing very different styles.

Future work

- Programming all of these in a parallel system (GPU)
- Find a better way to define the “desired output”
- Try a more powerful RNN
- Consider “future chunks” to assign the best class

7. References

Elman, J. "Finding Structure in Time". *Congnitive Science , Vol. 14,* 1990 ,pp. 179-211.

Gómez-Gil, P. J. Navarrete-García "Analysis of a Neural-net-based Algorithm for the Segmentation of Difficult-to-read Handwritten Letters." WSEAS Transactions on Systems. Issue 4, Vol. 3. June 2004. ISSN: 1109-2777.pp. 1426 – 1429.

Gómez-Gil P, De-Los-Santos Torres G, Navarrete-García J, Ramírez-Cortés JM. "The Role of Neural Networks in the interpretation of Antique Handwritten Documents.". Hybrid Intelligent Systems. Analysis and Design Series: Studies at Fuzziness and Soft Computing. Vol 208. Editors: Castillo, O. Melin, P. Kacprzyk W. 2007 Springer. ISBN-10: 3-540-37419-1. pp. 269-281.

Kohonen, T. "Self-Organizing Maps", *Springer Series in Information Science, 3rd edition, 2001.*

Luna-Pérez R, Gómez-Gil P. "Unconstrained Handwritten Word Recognition Using a combination of Neural Networks." Lecture Notes in Engineering and Computer Science: Proceedings of the World Congress on Engineering and Computer Science 2010, WCECS 2010, 20-22 October, 2010,San Francisco, USA. pp 525 - 528. Eds. Ao SI, Douglas C, Grundfest WS, Burgstone J. International Association of Engineers. ISBN: 978-988-17012-0-6. ISNN: 2078-0958 (print) ISNN: 2078-0966 (online).

Luna-Perez, R. "Reconocimiento fuera de línea de palabras manuscritas, basado en un mapa auto-organizado" Tesis de Maestría en Ciencias Computacionales. Instituto Nacional de Astrofísica, Optica y Electrónica, Febrero 2011.

Marti U- and H.Bunke. *The IAM-database: an English sentence database or offline handwritten recognition.* 2002, International Journal on Document Analysis and Recognition, Vol. 5, pp. 39-46.

Vinciarelli, A "A survey on off-line Cursive Word Recognition". *Pattern Recognition, The journal off pattern recognition society, vol.* 35 (7), 2002, pp. 1433 - 1446.

Thank you!

ccc.inaoep.mx/~pgomez

This presentation is available at:
http://ccc.inaoep.mx/~pgomez/conferences/PggIsc12.pdf