Biological Models and Data Mining techniques to discover behaviours and dynamics in Complex Systems

Giovanni Ponti
DTE-ICT-HPC, ENEA Portici Research Center

joint work with Salvatore Cuomo[1] and Pasquale De Michele[1]

[1]Università degli Studi di Napoli Federico II

Italian national agency for new technologies, energy and sustainable economic development

16th June, 2016

Outline

Case Study

Real Data from Art exhibition within Maschio Angioino Castle, in Naples (Italy), of sculptures by Francesco Jerace promoted by DATABENC

1. User Interest
   - Clustering to tune the Integrate & Fire (I&F) model
   - Clustering VS Bayesian Model
   - From single neuron to a network of I&F neurons
   - Parallelization of the CA1 hippocampus neuron modeling the *depolarization block*
2. User Behaviours
   - Visitor behaviours compared to the ones of 4 typical animals
   - Metamorphosis of the visitor behaviours
   - Technique and modeling refinements
3. User Evaluation
   - Collaborative Reputation Systems and visitor behaviours

Outline

Case Study

Real Data from Art exhibition within Maschio Angioino Castle, in Naples (Italy), of sculptures by Francesco Jerace promoted by DATABENC

1. User Interest
   - Clustering to tune the Integrate & Fire (I&F) model
   - Clustering VS Bayesian Model
   - From single neuron to a network of I&F neurons
   - Parallelization of the CA1 hippocampus neuron modeling the *depolarization block*
2. User Behaviours
   - Visitor behaviours compared to the ones of 4 typical animals
   - Metamorphosis of the visitor behaviours
   - Technique and modeling refinements
3. User Evaluation
   - Collaborative Reputation Systems and visitor behaviours

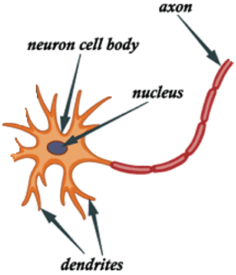
Problem background

The focus

In the cultural heritage context, we want to analyze **spectator** behaviours by means of a biological **neuron**.

A spectator

A neuron

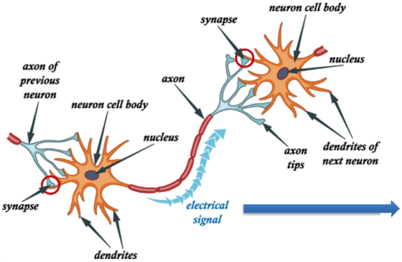


Biological neuron behaviour

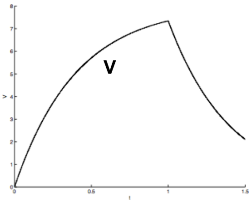
In a neural network:

- Each neuron is subject to electrical signals (i.e., stimuli) from other neurons.
- The electrical activity of a neuron is measurable by means of the potential difference (**V**).
- A neuron appropriately stimulated has a potential peak (i.e., spike).

Neural network

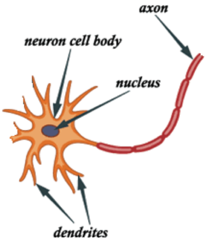


Spike (a potential peak)



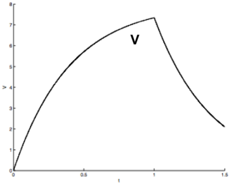
To measure the **interest** of a **spectator** w.r.t. an **artwork**.

A spectator (a neuron)



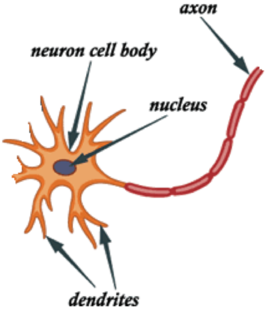
An artwork

User interest (spike)

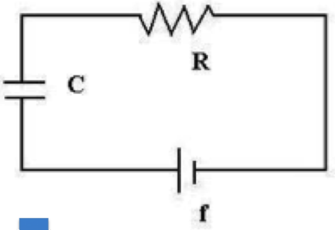


From biological to computational model

A neuron



Schematic behaviour (RC circuit)



Adopted mathematical model: Integrate & Fire

$$I(t) - \frac{V}{R} = C\frac{dV}{dt}$$

Our mathematical model: I&F

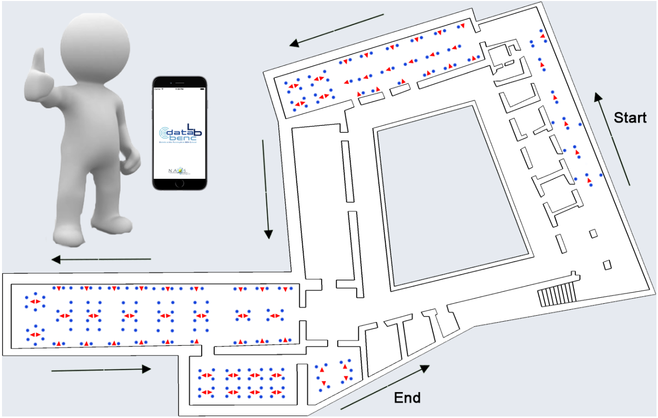
- The Integrate&Fire (I&F) is a simple model that considers the neuron as an electrical circuit, in which only the effects of the membrane capacitance are evaluated.

$$I(t) - \frac{V}{R} = C\frac{dV}{dt}$$

- C is the membrane **capacitance**, R is the **resistance**, V is the membrane **potential** and $I(t)$ is the ionic **current** of the neuron.

The ICT system framework

The communication technology is based on Bluetooth.



- In blue: the **artworks**.
- In red: the **smart crickets**.

The data collecting: booklet contents

- It is necessary to feed the artwork fruition.
- It requires a particular structure to ensure that the artworks start to talk and interact with the people.
- It is implemented with an xml schema diagram that collects the spectator behaviour w.r.t. the attributes of an artwork.

```xml
<?xml version="1.0" encoding="UTF-8"?>
 <USER ID='UI001'>
        <STEREOTYPE_USER>2</STEREOTYPE_USER>
        <START_SESSION></START_SESSION>
        <END_SESSION></END_SESSION>
  <TRANSACTION>
   <REQUEST>
     <HTTP_METHOD>GET</HTTP_METHOD>
     <PATH_INFO>/opera</PATH_INFO>
     <REQUEST_PARAMETERS>
        <CODEARTWORK>ART0224VICTA</CODEARTWORK>
        <DATE>13/05/2013</DATE>
          </REQUEST_PARAMETERS>
        <REMOTE_ADDRESS>192.168.1.6</REMOTE_ADDRESS>
  </REQUEST>
    <PARAMETERS_LOG>
        <HOUR_LISTEN_START>13/05/2013 13:58:12</
             HOUR_LISTEN_START>
        <HOUR_LISTEN_END>13/05/2013 14:05:42</
             HOUR_LISTEN_END>
        <AUDIOS>
    <TOT_NUMBER>3</TOT_NUMBER>
    <AUDIO ID='AU1111'>
           <HOUR_END>13/05/2013 14:00:42</HOUR_END>
```

The biological inspired model behaviours

- Collected data are organized as **tuples** representing spectator's **fruition** w.r.t. media contents:
  - Audio
  - Image
  - Text
- A spectator enjoyed an artwork (emits a spike) if it exploits:
  - the **whole content** of **at least one** of the **available media contents**
  - **more than the 66%** of the **total contents**

Bayesian Classifier [SEBD '14] (1)

- To have a comparison metric, we firstly choose to analyze the data through a **Bayesian Classifier** (**BC**):

$$CM(x_1, x_2, \ldots, x_n) = \arg\max_z p\big(Z = z\big) \prod_{i=1}^{N} p(X_i = x_i | Z = z)$$

  - Z is a dependent class variable;
  - $x_1, x_2, \ldots, x_n$ are feature variables.

- The BC computes individual conditional probabilities for each values of the class variable Z and for each feature $p(X_i|Z_j)$.

Bayesian Classifier [SEBD '14] (2)

- We have organized the log file structure in a Weka's ARFF file format.

```
@RELATION ARTWORK
@ATTRIBUTE audios {p02, p04, p06, p08, p1}
@ATTRIBUTE images {p02, p04, p06, p08, p1}
@ATTRIBUTE texts {p02, p04, p06, p08, p1}
@ATTRIBUTE spike {yes, no}
@DATA
p02,p02,p02,No
p04,p02,p02,No
...
p1,p04,p02,Yes
...
```

- p02,p02,p02,No → **Tuple that does not generate a spike**
- p04,p02,p02,No → **Tuple that does not generate a spike**
- p1,p04,p02,Yes → **Tuple that generates spike**

Bayesian Classifier [SEBD '14] (3)

- To evaluate the quality of the classification, we define two metrics:
  - Precision: $\frac{TP}{(TP+FN)}$
  - Recall: $\frac{TP}{(TP+FP)}$

Metric	**Value**
#of classified elements	125
True Positive (TP)	65
False Positive (FP)	0
True Negative (TN)	52
False Negative (FN)	8
Precision	**1**
Recall	**0.89**

Clustering [KomIS'14] [CCIS'15] (1)

- Then, we choose to analyze the data through a clustering-based approach.
- We have organized the log file structure in a novel Weka's ARFF file format.

```
@RELATION ARTWORK
@ATTRIBUTE audios NUMERIC [0..1]
@ATTRIBUTE images NUMERIC [0..1]
@ATTRIBUTE texts NUMERIC [0..1]
@ATTRIBUTE spike {0,1}
@DATA
0.1,0.4,1.0,1
0.3,0.6,0.4,0
...
0.5,1.0,0.7,1
...
```

Tuple that generates spike

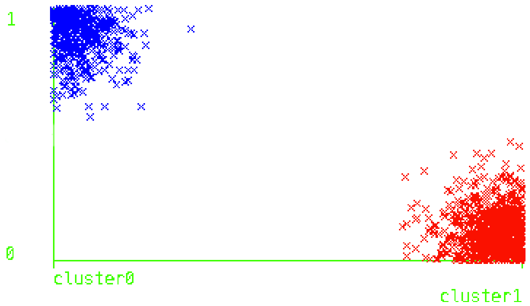
Tuple that does not generate a spike

Tuple that generates spike

Clustering [KomIS'14] [CCIS'15](2)

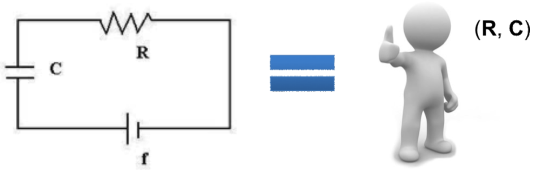
- The clustering approach consists in:
  - Combining the values of the attributes **audios**, **images** and **texts**:
    - $N = 1,331$ tuples.
  - Resorting to the well-known K-means clustering algorithm, with $K = 2$ number of clusters:
    - "Cluster0 = Spike", "Cluster1 = No-spike"

  we obtain:



The biological inspired model: proposed approach (1)

- We have exploited firsly the BC results [SEBD '14], then the clustering results [KomIS '14] in order to tune the values of the **Resistance** (R) and **Conductance** (C) of the circuit that **represents** the **spectator** (the neural model).

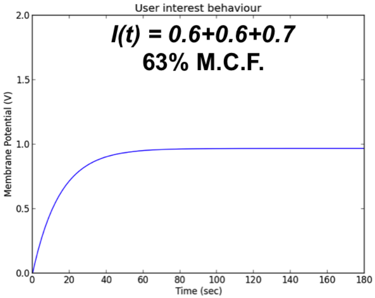


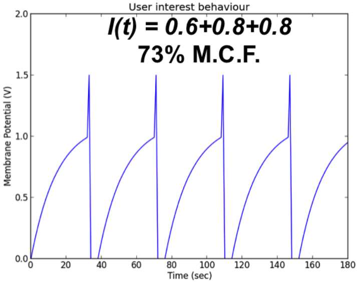
- The couple (R, C) represents the spectator sensitivity w.r.t. an artwork:

$$(R, C) = (0.51kOhm, 30\mu F)$$

The biological inspired model: proposed approach (2)

- The **Current** $I(t)$, which represents the **Media Content Fruition** (**M.C.F.**) related to an artwork, is a linear combination of the values of the attributes in the dataset.





The biological inspired model: results

- Experimental results for the clustering and the proposed I&F model, with

$$(R, C) = (0.51kOhm, 30\mu F)$$

Tuples	M.C.F. (%)	Cluster	# spikes
0.2,0.2,0.2	20%	cluster1	0
0.2,0.2,0.4	27%	cluster1	0
0.4,0.2,0.2	27%	cluster1	0
0.6,0.6,0.7	63%	cluster1	0
0.6,0.8,0.8	73%	cluster0	4
0.7,0.9,0.5	70%	cluster0	4
0.8,0.9,0.3	67%	cluster0	2
0.8,0.9,0.6	76%	cluster0	5
1.0,0.2,0.1	43%$^{(*)}$	cluster0	5
1.0,0.8,0.9	90%$^{(*)}$	cluster0	10
1.0,1.0,0.6	86%$^{(*)}$	cluster0	13
1.0,1.0,1.0	100%$^{(*)}$	cluster0	16

(*) = Tuple combinations that contain the information about the **fully fruition of at least one media content**.

From single neuron to neural network model [VICTA'14]

- Another challenge is the building of computational neural network able to model the dynamics of users in a community that shares interests on cultural heritage topics:
  - The **neural network** represents a **social network**.
  - The **neurons** represent the spectators that are social network **users**.
  - The **synapses** (i.e., connection) among the neurons are intended such as **user friendships**:
    - Connections may be more (or less) strong according to users share (or not) a particular interest.

Modeling the (social) network topology

- User U_i has a friendship with user U_j by means of a synapse that is a positive weight $w_{i,j}$ ($w_{i,j} = w_{j,i}$).
- Then, the network topology is modeled by means of a matrix

$$W = \begin{cases} w_{i,j} > 0 & if \quad U_i \quad is \quad linked \quad to \quad U_j \\ 0 & otherwise. \end{cases}$$

- Connections may be more strong according to users share a particular interest:
  - E.g.: users belonging to the same *circle* (or *fan page*).

Liking degree and Sociality level (1)

- The weights $w_{i,j}$ are a measure of the **liking degree** of the users w.r.t. an assigned feature:
  - E.g., the interest on the art history, on an artistic period or on an artwork.
- In this context, the (R, C) couple gives information about the **sociality level** of a user.
- The measure of the sociality level is expressed by the product $\tau = R \times C$, which is the tendency of the user to interact on a social network, sharing, liking or commenting a social content.

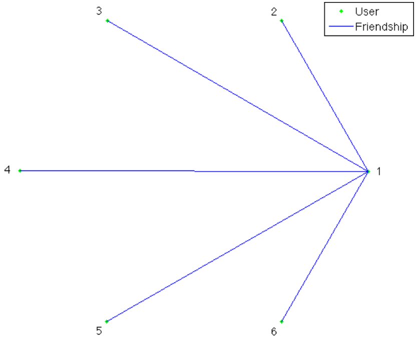
- A **small** value of τ corresponds to an **high** sociality level.
- A **large** value of $w_{i,j}$ gives that liking degree between users U_i and U_j is **high**.
- The neuron spikes give information about the social activity.

Experiments

- Only one neuron is subject to an external stimulus I_{ext} and we are interested to observe how his neural activity influences other neurons.
- By the social network point of view, this means that we observe how the social activity of a single user influences other users.

Experiments: simple interaction (1)

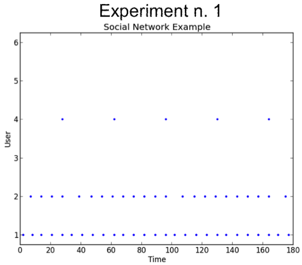
- $N = 6$ users.
- U_1 with $(R, C) = (0.51kOhm, 10\mu F) \rightarrow \tau_1 = 5.1$.
- User U_1 is connected with each other users.

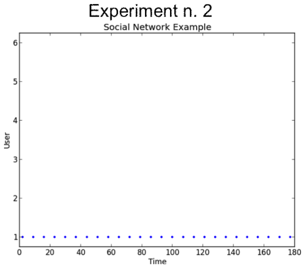


Experiments: simple interaction (2)

- Experiment n. 1:
  - $w_{1,i} = 1.3$ $(i = 2, \ldots, N)$;
  - $\tau_i = i \times x_i \times \tau_1$ $(i = 2, \ldots, N)$;
  - x_i is a random value between 0 and 1.
- Experiment n. 2:
  - $w_{1,i} = 0.5$ $(i = 2, \ldots, N)$.

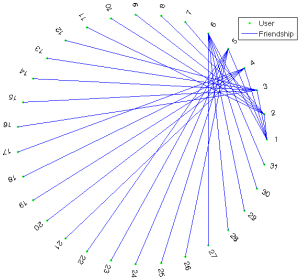
User	τ_i $(w_{1,i} = 1.3)$	τ_i $(w_{1,i} = 0.5)$
U_1	5.10	5.10
U_2	1.68	2.36
U_3	7.94	1.04
U_4	5.49	1.67
U_5	8.24	7.55
U_6	22.85	21.15

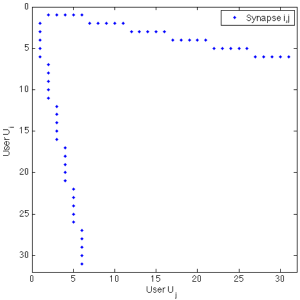




Experiments: viral interaction (1)

- $N = 31$ users, $M = 5$ (*friendship cardinality*).
- U_1 with $(R, C) = (0.51kOhm, 10\mu F) \rightarrow \tau_1 = 5.1$.
- User U_i is connected with users U_j ($j = M*(i-1)+2, \ldots, (M+1)*i+1-i$).



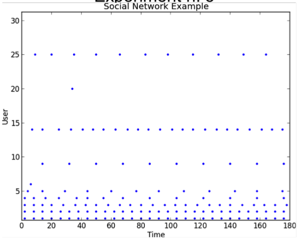


Experiments: viral interaction (2)

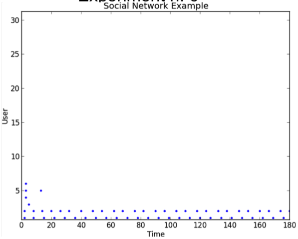
- Experiment n. 5:
  - $w_{i,j} = 1.3$ $(j = M * (i-1) + 2, \ldots, (M+1) * i + 1 - i)$;
  - $\tau_i = i \times x_i \times \tau_1$ $(i = 2, \ldots, N)$;
  - x_i is a random value between 0 and 1.
- Experiment n. 6:
  - $w_{i,j} = 0.5$ $(j = M * (i-1) + 2, \ldots, (M+1) * i + 1 - i)$.

User	τ_i $(w_{i,j} = 1.3)$	τ_i $(w_{i,j} = 0.5)$
U_1	5.10	5.10
U_2	0.40	0.36
U_3	1.78	6.28
U_4	3.47	4.07
U_5	13.56	3.18
U_6	24.79	3.99
U_9	3.32	23.66
U_{14}	1.59	53.58
U_{20}	7.65	43.35
U_{25}	2.32	78.98

Experiment n. 5

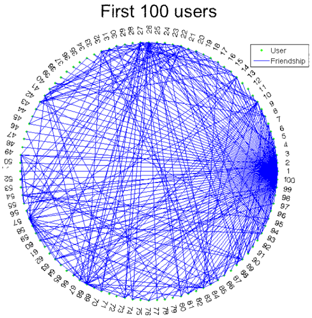


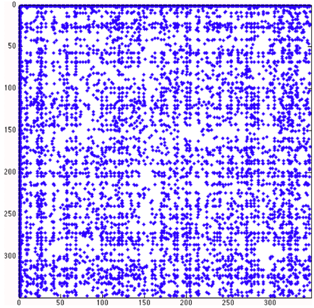
Experiment n. 6



Experiments: real interaction (1)

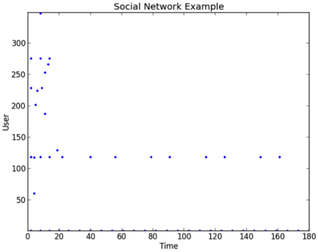
- $N = 348$ users connected on Facebook.
- The dataset coming from the Stanford Network Analysis Platform (SNAP) (`https://snap.stanford.edu`).
- U_1 with $(R, C) = (0.51kOhm, 10\mu F) \rightarrow \tau_1 = 5.1$.
- User U_1 is connected with each other users.
- User U_i may have connections with user U_j (with $i, j \neq 1$).

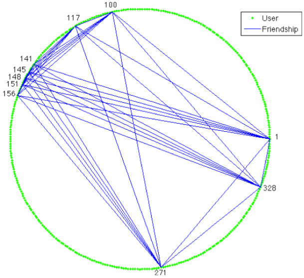




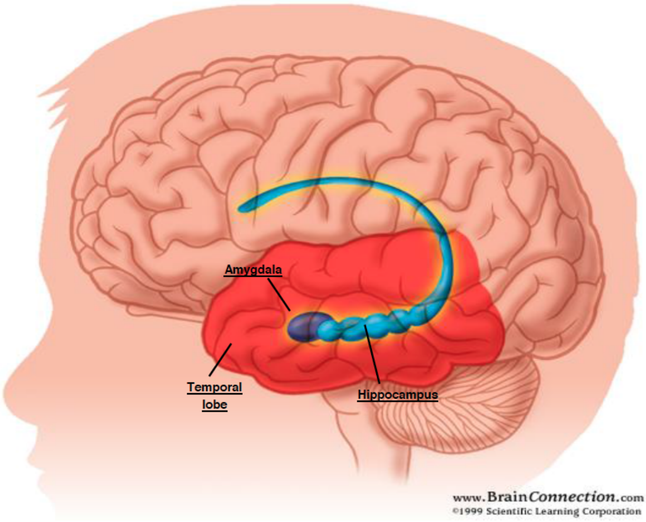
- Experiment n. 8:
  - $w_{i,j} = 0.2$;
  - $w_{1,j} = 0.4$ (sharing some interest with user 1)
  - $\tau_i = i \times x_i \times \tau_1$ $(i = 2, \ldots, N)$;
  - x_i is a random value between 0 and 1.

Experiment n. 8



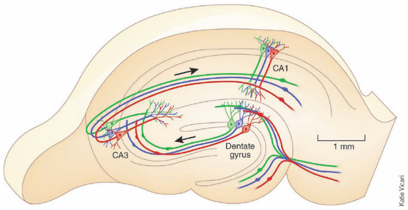


Neuroscience: study of the human brain



The hippocampus brain region

- One of the most widely studied of brain regions is the *hippocampus*:
  - It is involved in the intermediate term storage of memories which can be consciously recalled: *declarative memory*.
  - It is divided in many areas: CA3, CA1 and Dentate Gyrus (DG):
    - regions CA3 and CA1 have been proposed to be auto and heteroassociative memories, respectively, for the storage of declarative memories.



Goal of the work [ICCS'15]

Starting Point

The single CA1 neuron model for the hippocampus [Bianchi '12]

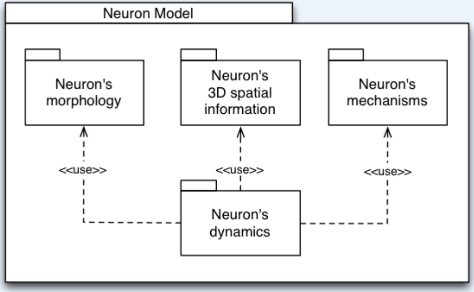
- a computational model (i.e., the simulator) able to reproduce the *depolarization block*, a characteristic biological behaviour that occurs when, under sustained input current of increasing strength, neurons stop firing.

Focus

- formalization of the computational framework underlying the model for CA1 neuron;
- parallel definition and implementation of the model;
- performance analysis

Main packages of the CA1 computational model

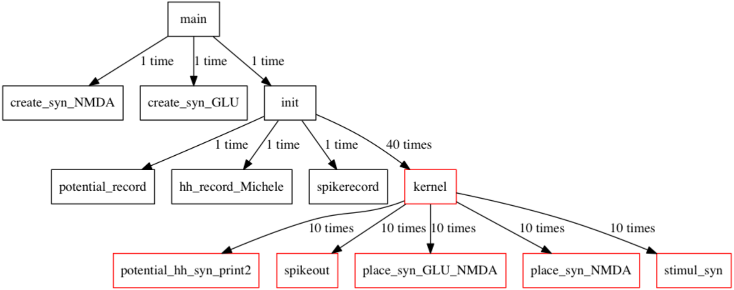
4 main packages:



- *Neuron's morphology*
- *Neuron's 3D spatial information*
- *Neuron's mechanisms*
- *Neuron's dynamics*

Implementation scheme of the CA1 simulator

- Computational resolution tree of task *neuron's dynamics* of the CA1 model.



- The most expensive part of this task is the procedure `kernel`.

The solution is to resort to the parallel methodology available on the problem solving environment NEURON: *multisplit*

- NEURON provides a complete suite of parallel tools, based on the Message Passing Interface (MPI).

Computing infrastructures

- S.Co.P.E. Grid infrastructure at University of Naples Federico II, Naples, Italy:
  - 304 blade servers
  - each of which equipped with 2 "Intel Xeon E5410" 4−cores CPU (8−cores for each node) at 2.33GHz (2432 cores in total);
  - 10Gbps Infiniband links, with 8 or 16GB/RAM per node.
- CRESCO3 infrastructure of the ENEA Research Center, Portici, Naples, Italy:
  - 84 blade servers
  - each of which equipped with 2 "AMD Opteron 6234TM" 12−cores sockets (24−cores for each node) at 2.4GHz (2016 cores in total);
  - 40Gbps Infiniband links, with 64GB/RAM per node.

References

- **SEBD'14**: Cuomo, S., De Michele, P., Posteraro, M.: *A biologically inspired model for describing the user behaviors in a Cultural Heritage environment*. SEBD 2014, June 16-18 2014, Sorrento Coast, Italy.
- **KomIS'14**: Cuomo, S., De Michele, P., Ponti, G., Posteraro, M.: *A clustering-based approach for a finest biological model generation describing visitor behaviours in a Cultural Heritage scenario*. International Conference on Data Management Technologies and Applications, Vienna, Austria, 29-31 August, 2014
- **VICTA'14**: Cuomo, S., De Michele, P., Galletti, A., Ponti, G.: *A Biologically Inspired Model for Analyzing Behaviours in a Social Network Community and Cultural Heritage Scenario*. International Conference on Signal-Image Technology and Internet-Based Systems, Marrakech, Morocco, November 23-27, 2014
- **CCIS'15**: Cuomo, S., De Michele, P., Ponti, G., Posteraro, M.: *Validation Approaches for a Biological Model Generation Describing Visitor Behaviours in a Cultural Heritage Scenario*. Communications in Computer and Information Science 2015
- **ICCS'15**: Cuomo, S., De Michele, P., Galletti, A., Ponti, G.: *Parallel tools for simulating the depolarization block on a neural model*. International Conference on Computational Science, Reykjavík, Iceland,

Outline

Case Study

Real Data from Art exhibition within Maschio Angioino Castle, in Naples (Italy), of sculptures by Francesco Jerace promoted by DATABENC

1. User Interest
    - Clustering to tune the Integrate & Fire (I&F) model
    - Clustering VS Bayesian Model
    - From single neuron to a network of I&F neurons
    - Parallelization of the CA1 hippocampus neuron modeling the *depolarization block*
2. User Behaviours
    - Visitor behaviours compared to the ones of 4 typical animals
    - Metamorphosis of the visitor behaviours
    - Technique and modeling refinements
3. User Evaluation
    - Collaborative Reputation Systems and visitor behaviours

Visitor behaviours (1)

To classify the **visitor behaviours** w.r.t. the **usage of mobile devices** in an exhibit. In Zancaro et al. (2007), visitor behaviours are compared to that of four typical animals:

- ANT (A);
- FISH (F);
- BUTTERFLY (B);
- GRASSHOPPER (G).

We adapt this classification to find **how visitors interact with the ICT** technology and **how a lot they are interested in the exhibition**.

Visitor behaviours (2)

ANT (A)

tends to follow a specific path and spends a lot of time using the furnished technology

FISH (F)

moves around in the centre of the rooms and usually avoids looking at media content details

BUTTERFLY (B)

does not follow a specific path and stops frequently to look for more media content

GRASSHOPPER (G)

seems to have a specific preference for some preselected artworks and spends a lot of time observing the related media contents

The data collecting: JSON logs

- It is necessary to feed the artwork fruition.
- It is implemented with a JSON schema that collects the spectator behaviour w.r.t. the attributes of an artwork.


```
{
"USER":{
 "SESSION":[
  {
   "START_SESSION":"29/12/2014 16:15:10:540",
   "IDUSER":"fdb9eab819aa3f791419866070461",
   "PLATFORM":"Android",
   "START_LANG":"it",
   "ID_EVENT":10,
   "RANK_EVENT":2.5,
   "NUMBER_ARTWORK":"271",
   "TRANSACTION":[
```

```
      "ACTION":[
       {
        "TYPE":"AUDIO",
        "ID":"734",
        "TIME_START":"29/12/2014 16:16:00:448",
        "TIME_END":"29/12/2014 16:16:13:973",
        "TOT_TIME":"32.922"
       },
       {
        "TYPE":"IMAGE",
        "ID":"/Media/79/128/Image/2014-10-08_17-50-34.jpg"
       },
       ...
```

The interpretation of data (1)

We have tracked the visitor behaviours by using a suitable Extrapolation Algorithm (EA) that has the JSON file as input data.

```
IDUser :  e7a5774700c1e88e1417618582735
# of artworks:  271
# of viewed artworks:  44
% of viewed artworks :  17.5%

...
------------------------------------
i-th viewed artwork :  2
ID artwork :  128
Available audio (sec.) :  32.922
Listen audio (sec.) :  32.922
Available images :  3
Viewed images :  0
Available text : True
Viewed text :  False
Interaction time (sec.) :  58.259
Path is followed :  True
------------------------------------
...
------------------------------------
i-th viewed artwork :  6
ID artwork :  17
Available audio (sec.) :  85.141
Listen audio (sec.) :  85.141
Available images :  4
Viewed images :  2
Available text : True
Viewed text :  True
Interaction time (sec.) :  103.141
Path is followed :  False
```

The data interpretation (2)

- For each visitor, collected data are organized as **tuples** representing **fruition** percentages of the media contents of any artwork:
  - Audios
  - Images
  - Texts
- Moreover, relating to the entire exhibit, there is *another information collected*:
  - **Paths**

Paths [KomIS'15] [VICTA'15]

- Assume that a visitor follows this path:

$$1 \rightarrow 4 \rightarrow 5 \rightarrow 2 \rightarrow 3 \rightarrow 6 \rightarrow 7$$

- The count here is 4, as we have that:
  - the count starts from 1;
  - $1 \rightarrow 4$ and $3 \rightarrow 6$ are consecutive but not strictly consecutive (the contribute is 0.5 for each of these);
  - $4 \rightarrow 5$, $2 \rightarrow 3$ and $6 \rightarrow 7$ as strictly consecutive paths (the contribute is 1 for each of these);
  - $5 \rightarrow 2$ are not consecutive (the contribute is -1), because the user goes back from section 5 to section 2.
- Hence, **the percentage of followed path** is expressed as the ratio:

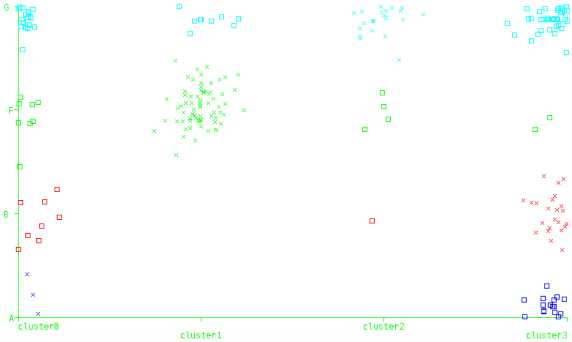
$$\frac{count}{\#\ \ of\ \ sections} = \frac{4}{7} = 0.571428$$

Clustering (1)

- The clustering approach consists in:
  - Combining the values of the attributes **audios**, **images**, **texts** and **paths**:
    - $N = 253$ tuples.
- Resorting to the well-known K-means clustering algorithm, with $K = 4$ number of clusters, for the first two sections of the exhibit, we obtain:

Animals	Cluster0	Cluster1	Cluster2	Cluster3
A	14	11	4	22
B	4	0	1	27
F	7	72	4	2
G	16	9	30	30

- Clustering with $K = 4$ for all the exhibit:
  - "Cluster0 = A", "Cluster1 = F ", "Cluster2 = G", "Cluster3 = B"



Assigning a score to a path: refinement (1) [CCIS'16]

To assign a value $v(p)$ to the quality of a path p, we can consider the $M-1$ movements $S_{j-1} \to S_j$, for $j = 2, \ldots, M$, and we can assign, at each of them, the value m_j in the following way

$$m_j = \begin{cases} 1, & \text{if } S_j = S_{j-1} + 1 \quad \text{(forward movement, consecutive)} \\ 0.5, & \text{if } S_j > S_{j-1} + 1 \quad \text{(forward movement, not consecut} \\ S_j - S_{j-1}, & \text{if } S_j < S_{j-1} \quad \text{(backward movement)} \end{cases}$$

Finally, in this way, the value of the path will be

$$v(p) = \max\left\{0, \frac{1}{N-1}\sum_{j=2}^{M} m_j\right\}.$$

The idea underlying is to consider correct the forward movements ($m_j = +1$ for $S_j = S_{j-1} + 1$, or $m_j = +0.5$ for $S_j > S_{j-1} + 1$) and to penalize the backward ones ($m_j = -(S_{j-1} - S_j)$ for $S_j < S_{j-1}$).

Assigning a score to a path: refinement (2)

By using the following formula

$$m_j = \begin{cases} 1, & \text{if } S_j = S_{j-1} + 1 & \text{(forward movement, consecutive)} \\ 0.5, & \text{if } S_j > S_{j-1} + 1 & \text{(forward movement, not consecut} \\ S_j - S_{j-1}, & \text{if } S_j < S_{j-1} & \text{(backward movement)} \end{cases}$$

let us to consider the paths p_1 and p_2:

- p_1 : $1 \to 2 \to 3 \to \cdots N-1 \to N$, is an *optimal path* that receives the score

$$v(p_1) = \max\left\{0, \frac{N-1}{N-1}\right\} = 1$$

- p_2 : $1 \to 2 \to 3 \to \cdots N-1 \to N \to 1$, is a *quasi-optimal path*, in the sense that it *contains* the optimal path and an unique repeated room, that receives the score

$$v(p_2) = \max\left\{0, \frac{N-N}{N-1}\right\} = 0$$

Assigning a score to a path: refinement (3)

Hence, a generic path may contain $L \geq 1$ **cycles**, which are sub-paths consisting in sequences of increasing steps interspersed with a backward steps. E.g.:

- p_1 : $1 \to 2 \to 3 \to \cdots N-1 \to N$, consists of only one cycle.
- p_2 : $1 \to 2 \to 3 \to \cdots N-1 \to N \to 1$, consists of two cycles.

Let us to consider p_3, containing two optimal cycles and another one:

$$p_3 : \quad \underbrace{1 \to 2 \to \cdots \to N}_{\text{1-st cycle } (N-1 \text{ movements})} \quad \to \quad \underbrace{1 \to 2 \to \cdots \to N}_{\text{2-nd cycle } (N-1 \text{ movements})} \quad \to \quad \underbrace{1 \to 2}_{\text{3-rd cycle (1 movem}}$$

In this way, p_3 receives a score $v(p_3) = 1/(N-1)$.

Assigning a score to a path: a further refinement (1) [KES'16]

The phenomena observed for $v(p_2) = 0$ and $v(p_3) = 1/(N-1)$ are due to the presence of (too) penalized backward steps.

These drawbacks in assigning the score are probably one of the main causes in some troubles of our classification scheme in previous works.

Then, to enhance these classification results, here we propose a new way to assign the score of a path which does not excessively penalize paths containing optimal sub-paths (cycles).

We denote the generic path consisting in $L \geq 1$ cycles, with the notation

$$p: \quad \underbrace{S_{c_0+1} \to \cdots \to S_{c_1}}_{\text{1-st cycle}} \to \underbrace{S_{c_1+1} \to \cdots \to S_{c_2}}_{\text{2-st cycle}} \to \cdots \to \underbrace{S_{c_{L-1}+1} \to \cdots \to S_{c_L}}_{L\text{-th cycle}}$$

where $c_0 = 0$ and $c_L = M$.

Assigning a score to a path: a further refinement (2)

$$p: \quad \underbrace{S_{c_0+1} \to \cdots \to S_{c_1}}_{\text{1-st cycle}} \to \underbrace{S_{c_1+1} \to \cdots \to S_{c_2}}_{\text{2-st cycle}} \to \cdots \to \underbrace{S_{c_{L-1}+1} \to \cdots \to S_{c_L}}_{L\text{-th cycle}}$$

As before, we give a value $\bar{m}_j$ to each movement $S_{j-1} \to S_j$ but, in this case, no negative values are assigned.

Moreover, we consider the steps in the first cycle better than the steps in the other cycles, i.e. we set the $\bar{m}_j$ values decreasing along the cycles, as follow:

$$\bar{m}_j = 0.9^{c-1} \qquad \text{if } S_j \text{ is in the } c\text{-th cycle}$$

Notice that the basis 0.9 is chosen to weight less and less (90%) the steps in the new cycles.

Assigning a score to a path: a further refinement (3)

Our idea to take also into account the presence of optimal sub-paths, is to assign the score with the following rule:

$$\bar{v}(p) = \alpha \max_{i=1,\ldots,L} \frac{c_i - c_{i-1}}{N} + (1-\alpha)\frac{1}{N \cdot L}\sum_{j=1}^{M} \bar{m}_j \qquad (0 \leq \alpha \leq 1).$$

Using this formula, the score depends on the weights α and $1-\alpha$ and on the maximum length of the sub-paths.

The parameter α is crucial:

- a high value (close to 1) preserves path with at least an optimal sub-path;
- a small value (close to 0) would penalize *chaotic* paths regardless of they may contain good sub-paths.

In our experiments we have chosen $\alpha = 0.7$.

References

- **KomIS'15**: Cuomo, S., De Michele, P., Galletti, A., Pane, F., Ponti, G.: *Visitor Dynamics in a Cultural Heritage Scenario*. International Conference on Data Management Technologies and Applications, Colmar, Alsace, France, 20-22 July, 2015
- **VICTA'15**: Cuomo, S., De Michele, P., Galletti, A., Ponti, G.: *Visiting Styles in an Art Exhibition Supported by a Digital Fruition System*. International Conference on Signal-Image Technology and Internet-Based Systems, Bangkok, Thailand, November 23-27, 2015
- **CCIS'16**: Cuomo, S., De Michele, P., Galletti, A., Ponti, G.: *Classify Visitor Behaviours in a Cultural Heritage Exhibition*. Communications in Computer and Information Science 2016
- **KES'16**: Cuomo, S., De Michele, P., Galletti, A., Ponti, G.: *Influence of Some Parameters on Visiting Style Classification in a Cultural Heritage Case Study*. Smart Innovation, Systems and Technologies, 2016

Outline

Case Study

Real Data from Art exhibition within Maschio Angioino Castle, in Naples (Italy), of sculptures by Francesco Jerace promoted by DATABENC

1. User Interest
    - Clustering to tune the Integrate & Fire (I&F) model
    - Clustering VS Bayesian Model
    - From single neuron to a network of I&F neurons
    - Parallelization of the CA1 hippocampus neuron modeling the *depolarization block*
2. User Behaviours
    - Visitor behaviours compared to the ones of 4 typical animals
    - Metamorphosis of the visitor behaviours
    - Technique and modeling refinements
3. User Evaluation
    - Collaborative Reputation Systems and visitor behaviours

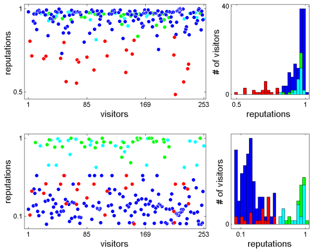
Evaluate users

Exploiting **Collaborative Reputation Systems** (CRS) tools to evaluate user experience in an exhibit

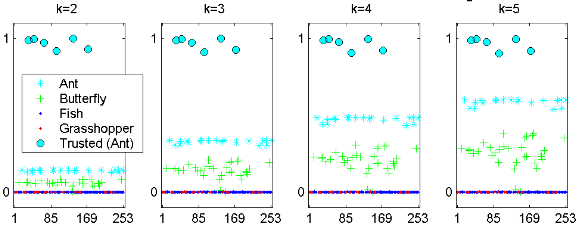
- Real data refers to artworks and digital media contents
- CRS ranking results are compared to the 4 behaviour categories

CRS results and categories

- approach without knowledge base for visitors [NUMTA'16]



- approach with info for **trusted users** [DAMIS'16]



References

- **NUMTA'16**: Cuomo, S., De Michele, P., Galletti, A., Ponti, G.: *Collaborative Reputation Systems in a Cultural Heritage Scenario*. International Conference on Numerical Computations: Theory and Algorithms, Pizzo Calabro, Calabria, Italy, 19-25 June, 2016
- **DAMIS'16**: Cuomo, S., De Michele, P., Galletti, A., Ponti, G.: *A numerical approach for assigning a reputation to users of an IoT framework*. International Workshop on Data Mining on IoT Systems, London, United Kingdom, 19-22 September, 2016

Thanks for the attention!

Giovanni Ponti, Ph.D.
ENEA DTE-ICT-HPC
Portici Research Center
giovanni.ponti@enea.it