Perpetual Motion, Evolutionary Computation in Industry and other Chimeras

Anna I Esparcia Alcázar

Software Production Methods Research Centre

Universitat Politècnica de València, Spain

UNIVERSITAT POLITÈCNICA DE VALÈNCIA

PROS

Centro de Investigación en Métodos de Producción de Software

The Chimera

My great-grandfather Juan

Famous quotes

Oh ye seekers after p
how many vain chime
pursued? Go and take
the alchemists.

— Leonardo da Vinci, 1494

Famous quotes

In theory, theory and same, but in practise

Attributed t

Albert Einstein

Yogi Berra

Objectives of this talk

- To show that it is **possible** to do Computational Intelligence in an SME
- To show that it is **impossible** to do Computational Intelligence in an SME

- …likely, to destroy the Universe in the process

The Engineer’s enemies

1. Money
2. Time
3. Space

0.People

People you should fear

- Your boss
- Your colleagues
- Yourself

Profiles of EC practicioners

- Edison:

  “If it’s technically good, people will buy it”

- Merlin:

  “Gi’es yer requirements, we’ll put them in a cauldron with all sorts of algorithms and…voilà!”

- Archimedes

  “Who cares if they buy it?”

6 popular academic misconceptions

1. Companies need what you've got
2. Companies want what you've got
3. Companies understand what you've got (and if they don't they must be blind/stupid/not worth talking to)
4. You don't need to do development, just show them proof of the pure research and they'll love/buy it
5. Just give it to the PhD student/R.A., s/he'll manage

But… EC has been applied in “the REAL world”, hasn’t it?

Things people have done with EC

- Design of diesel engine cylinders for French car manufacturer Peugeot
- Chevron has a patent for predicting oil well production
- Find patterns in big data: Nutonian's tool, Eureqa has been used to predict… basketball results
- …plus other things they won't tell you about

Meanwhile, in España

My History

Chapter 1

When I was in the "other side"

Success stories (i)

1. Logistics and transport
   - Minimization of transport and inventory costs
   - Allocation of products to shelves
2. Finance
   - Bankruptcy prediction
   - Portfolio optimisation
3. Modelling & optimisation
   - Placement of pheromone dispensers in agriculture
4. Bio-signal classification
   - Brain-computer interface for the disabled
   - Neuromarketing, usability
5. *Bot* evolution in computer games

Success story #1 - EVITA

An IRP aims to find the answer to three questions:

1. **When** to serve each customer?
2. **How much** to deliver to each customer each time it's served?
3. **What routes** to use for the delivery?

EVITA's Objective

Minimise global costs, c_g

$$c_g = c_h + c_t$$

where:

- c_t : transport costs (delivery)

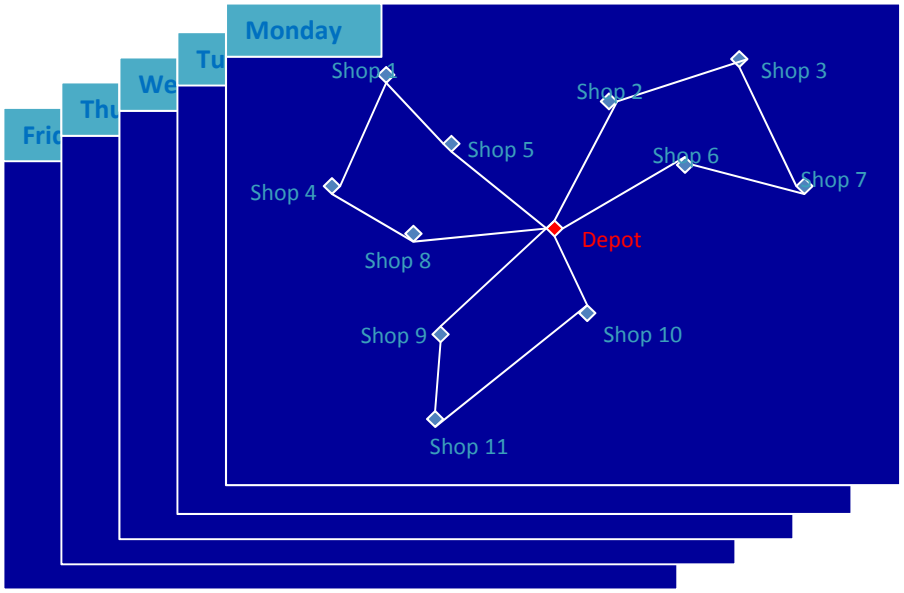
$$c_t = \sum_{H} \sum_{k \in K} \sum_{(i,j) \in E} c_{i,j} x_{ijk}$$

- c_h : inventory holding costs

$$c_h = \sum_{i \in V} h_i$$

My Inventory & Routing Problem

- Given a depot that supplies a retail chain, determine:
  - For each shop
    - Optimal delivery **frequency** (number of days a week).
    - Optimal delivery **pattern** (which days of the week)
  - Once both are known
    - Optimal set of **delivery routes** (for each day of the week)



…
Shop 9:
Delivery: Mon, Wed, Fri (pattern 21)
Holding cost: 968 €
…
Shop 10:
Delivery: Mon, Thurs (pattern 18)

Restrictions & simplifications

- Each shop must be served by only one vehicle.
- Maximum delivery time per vehicle = 8 hours.
- Each shop has its own set of admissible frequencies.
- The demand of each shop depends on the frequency it is supplied.
- *Time windows*: fixed time intervals in which shops can be served
- Unlimited number of vehicles
- Not all patterns are admissible for each frequency.

Pattern	Frequency (days per week)	L	M	X	J	V
18	2	✓			✓	
17	2	✓				✓
10	2		✓		✓	
9	2		✓			✓
5	2			✓		✓
21	3	✓		✓		✓
11	3		✓		✓	✓
13	3		✓	✓		✓
23	4	✓		✓	✓	✓
29	4	✓	✓	✓		✓
31	5	✓	✓	✓	✓	✓

Inventory holding costs and demands

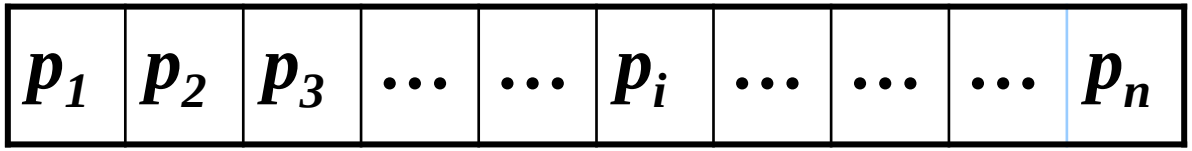
Shop type	Inventory cost (€)				Delivery size (roll containers)			
	Frequency (days)				Frequency (days)			
	2	3	4	5	2	3	4	5
A	-	-	336	325	-	-	2	2
B	-	-	327	317	-	-	2	2
C	330	311	303	301	4	2	2	1
D	310	292	285	283	3	2	2	1
E	293	276	269	267	3	2	2	1
F	277	261	255	-	2	2	1	-
G	268	253	-	-	2	1	-	-

Shop type codes

A - Valencia city centre
B - Valencia other + shopping malls
C - Valencia suburbs
D - Coastal villages
E - La Ribera county
F - Other villages
G - Requena-Utiel

Solving with CI (in 2 levels)

- Higher level: finds the pattern for each shop
  - Genetic algorithm



Where p_i = pattern for shop i

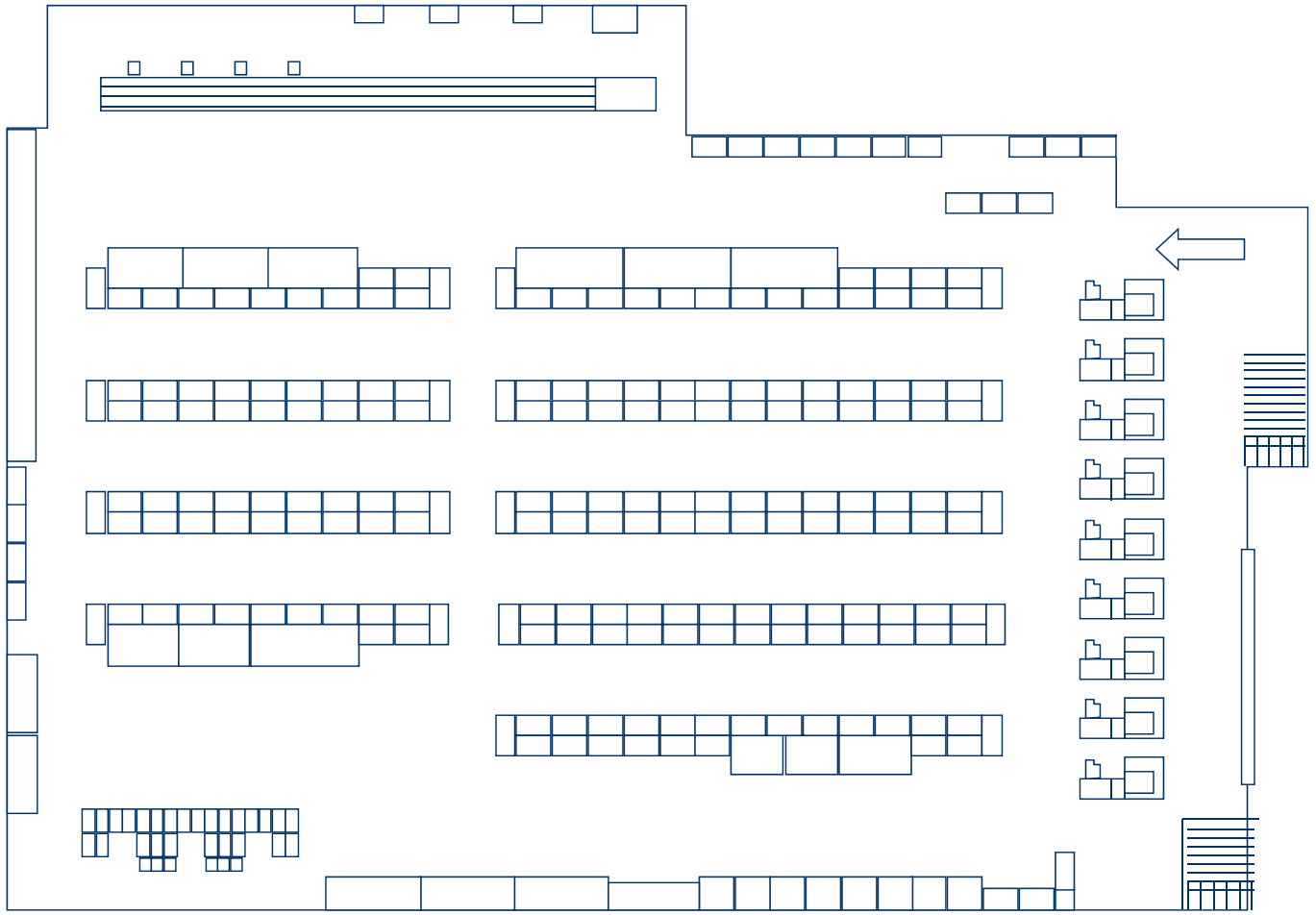
- Lower level: finds the routes for each day
  - Vehicle Routing Problem (VRP)

The Druni network

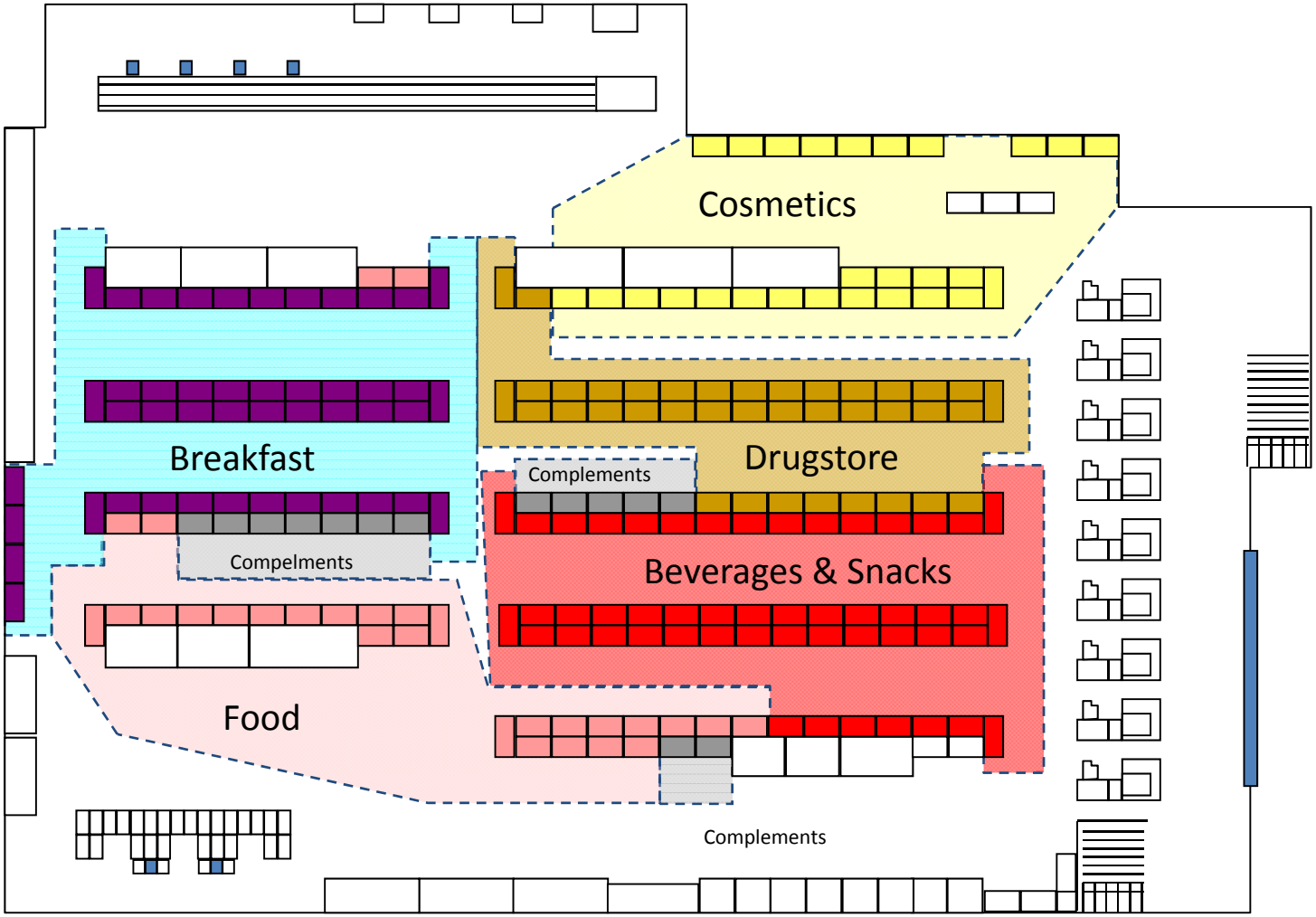
Success story #2 – MELiSSA

- Given the description of a new shop:
  - Shelves placement
  - Capacity (in modules) of each shelf
- Determine:
  - **Optimal number of modules** to allocate to each product group.
  - **Optimal location** of the modules occupied for each product group

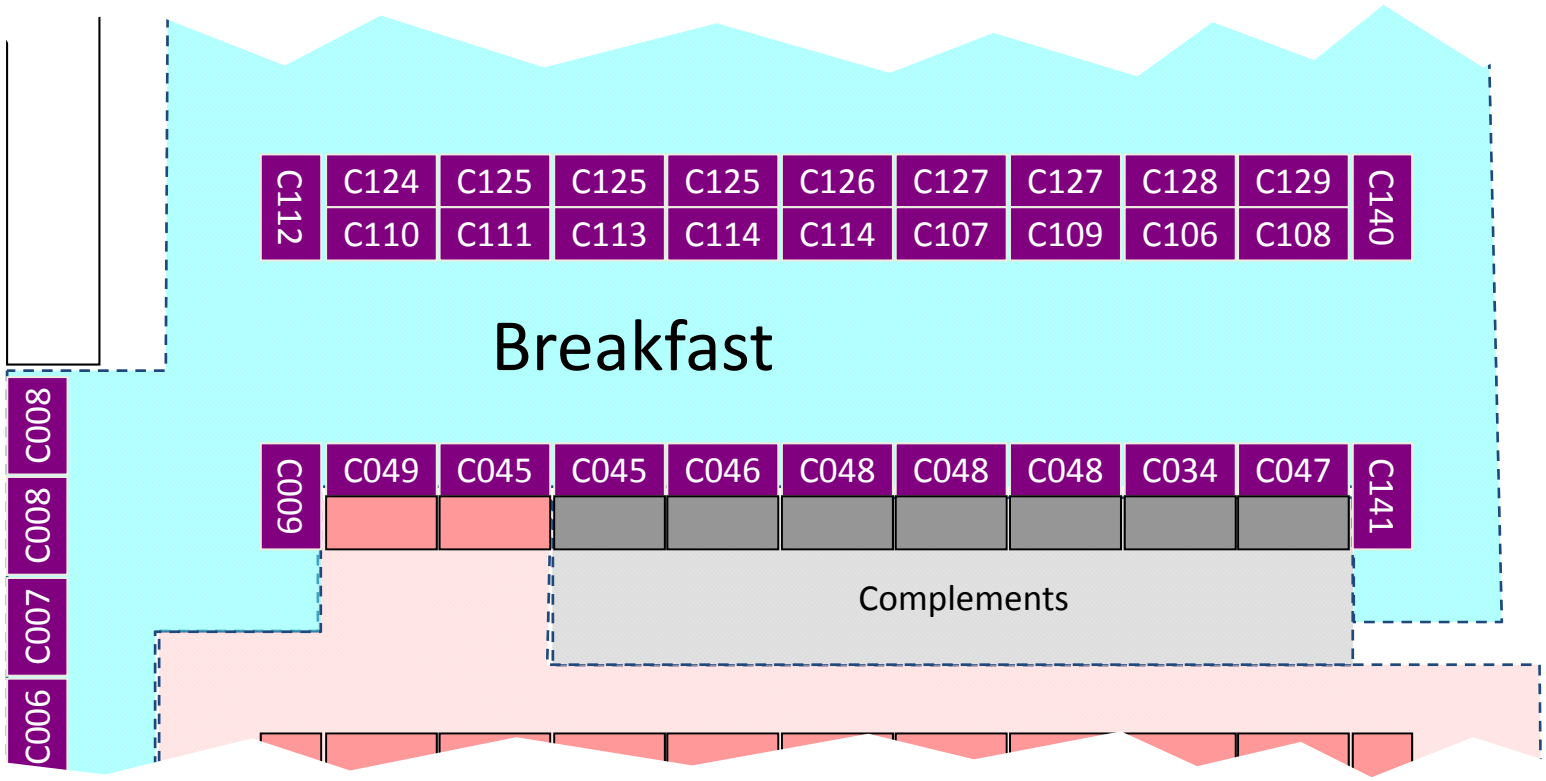
i. e. I have this :



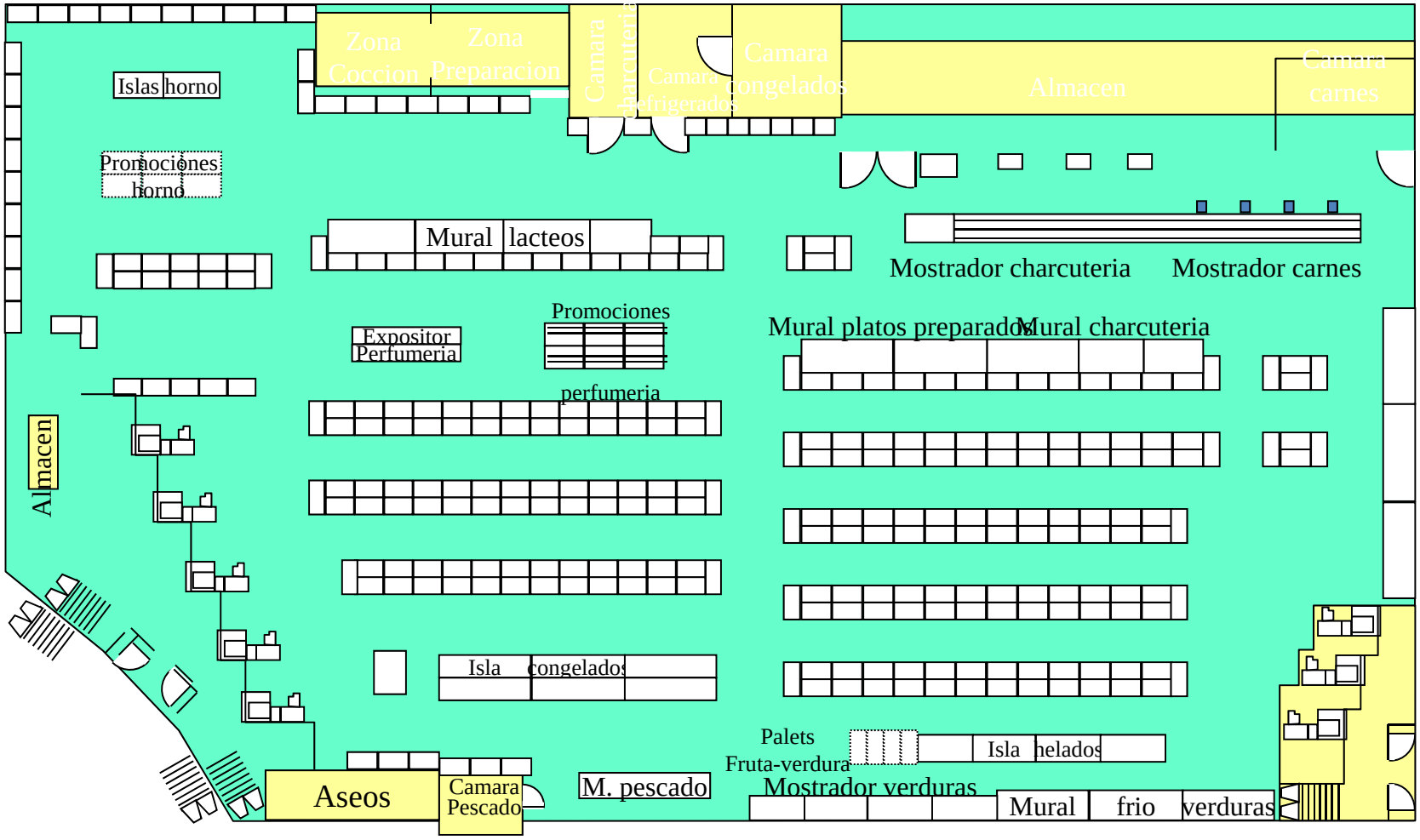
And I want something like this:

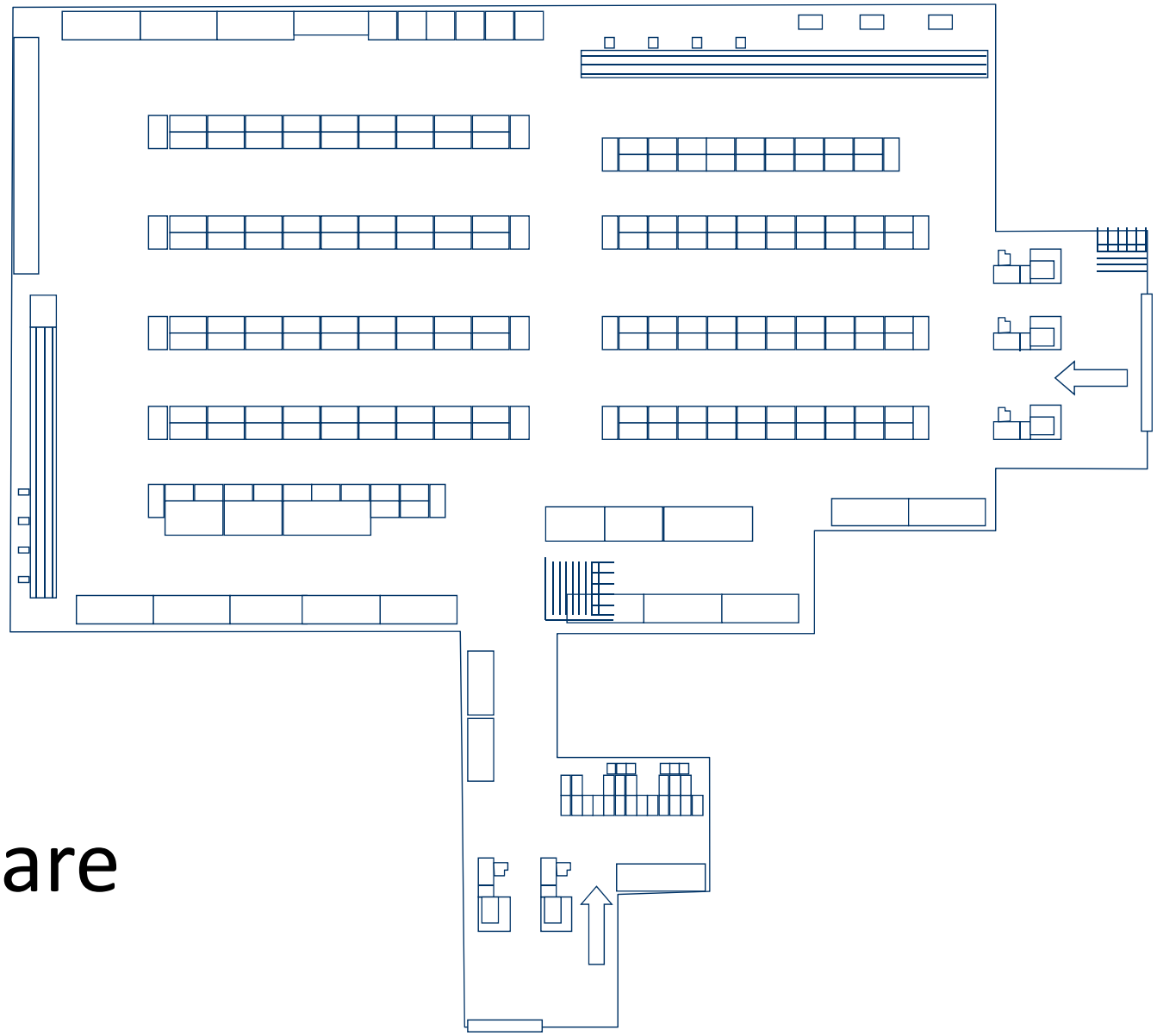


Or this :

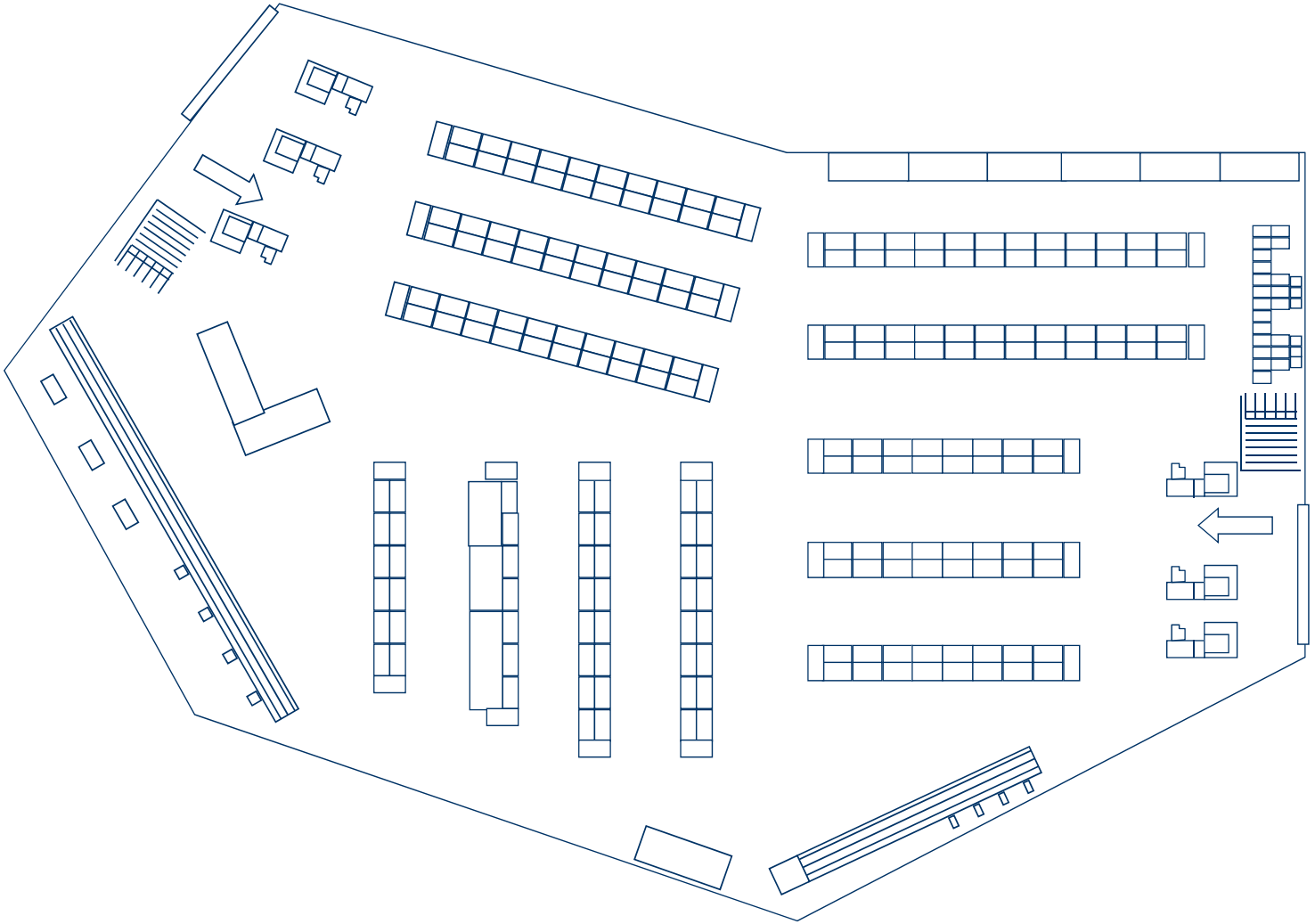

C112
C124
C125
C125
C125
C126
C127
C127
C128
C129
C140
C110
C111
C113
C114
C114
C107
C109
C106
C108
Breakfast
C008
C008
C007
C006
C009
C049
C045
C045
C046
C048
C048
C048
C034
C047
C141
Complements

There can be as many distributions…





… as there are shops



“Hard” restrictions

Space and capacity of each shelf.

Given by a table of distances and the description of the relative positions between shelves.

Standard shop

Values of an “ideal” shop, as defined by the mgt of the chain

(it may not coincide with the shop at hand)

Each group has a Standard/minimum/maximum:

→ ideal/minimum /maximum number of modules to assign.

Affinities & Adversities

Some products must be placed near or far from others, or near a reference point (oven, checkout line etc.)

Adverse: baby products & pet food

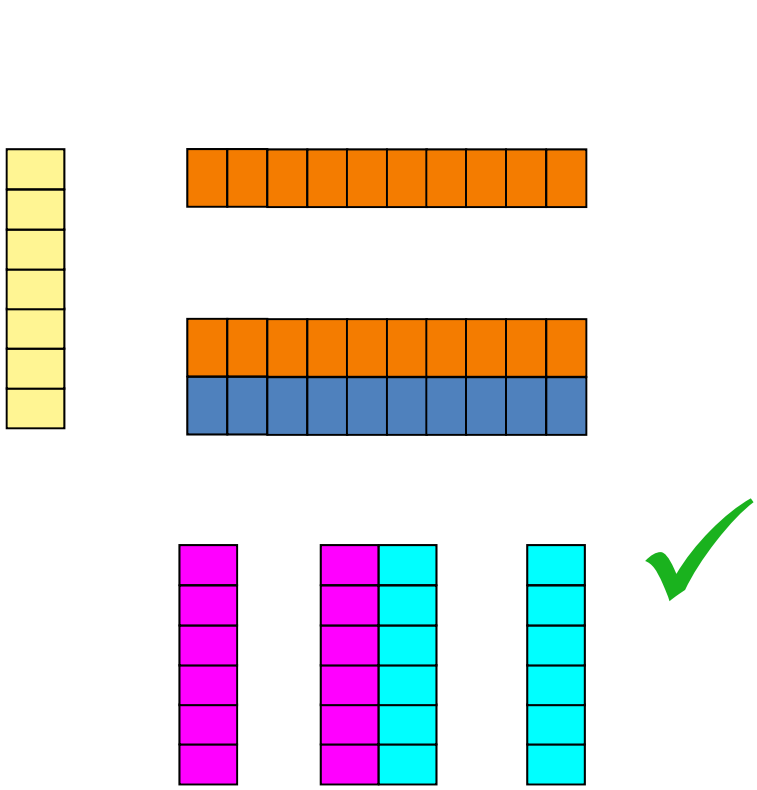
Indifferent: baby products & sauces

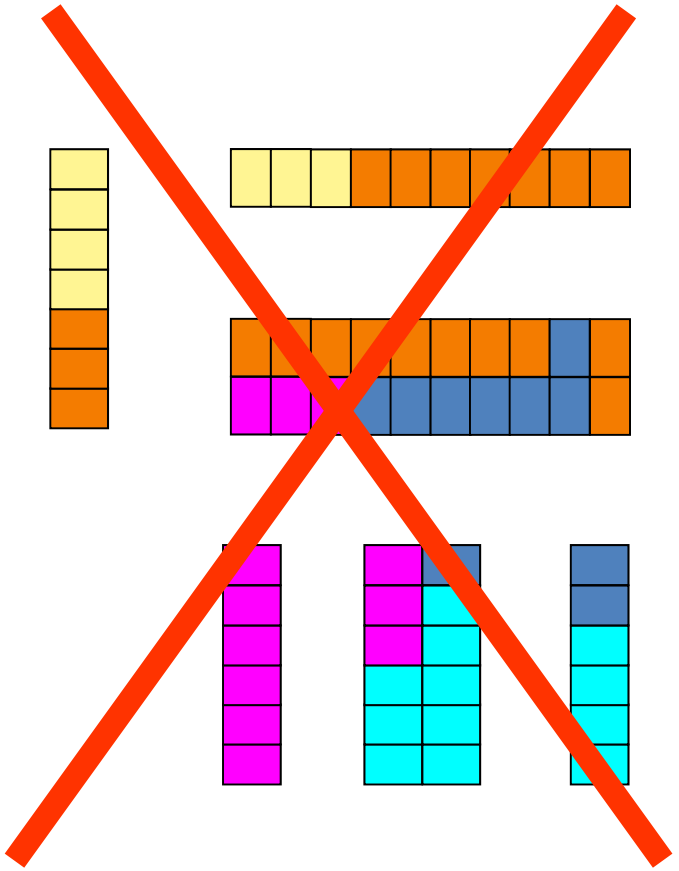
Affine: snacks& beverages

Affine to a reference point: biscuits& the oven

“Soft” restriction

Cohesion between groups





Multiobjective Evolutionary Algorithm (MOEA)

Optimal number of modules: **minimal deviation**

- If the size of the Standard Shop is less than the actual shop, all the groups must at least be in their standard
- If the size of the Standard Shop is more than the actual shop, adjust as best as possible.

Relative position of groups: **maximal affinity**

- Affine/adverse groups must be placed near/far.
- Groups affine to a reference point must be placed as near of it as possible

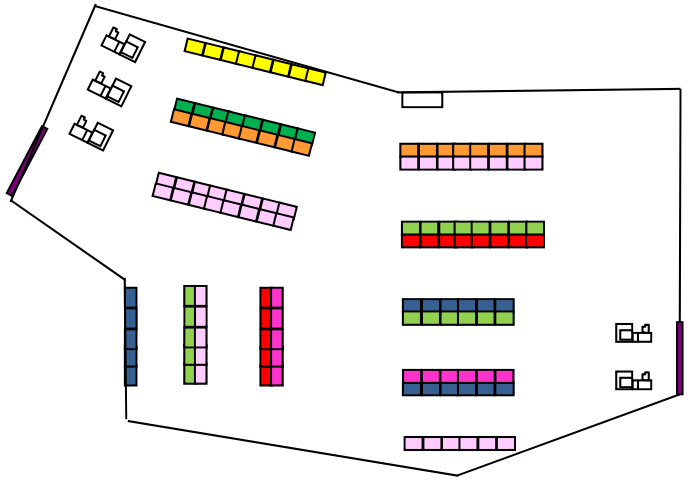
Maintain cohesion between groups: **minimal dispersion**

- Avoid “lost” modules
- Aim at making groups occupy full shelves

The problem

Group		Standard shop			Affine to	Adverse to
		std	mín	máx		
1	Food in general	14	11	19	3, 4, 7	2, 8
2	Drugstore	40	34	52	-	1, 3, 4, 5, 6, 7
3	Snacks	7	7	9	1	2, 8
4	Sauces & condiments	9	7	11	1,7	2
5	Bakery	21	15	28	Oven	2, 8
6	Baby Products	13	8	19	-	2,8
7	Beverages	17	9	21	1, 4	2, 8
8	Pet products	9	5	11	-	1, 3, 5, 6, 7
Size of standard shop		130				

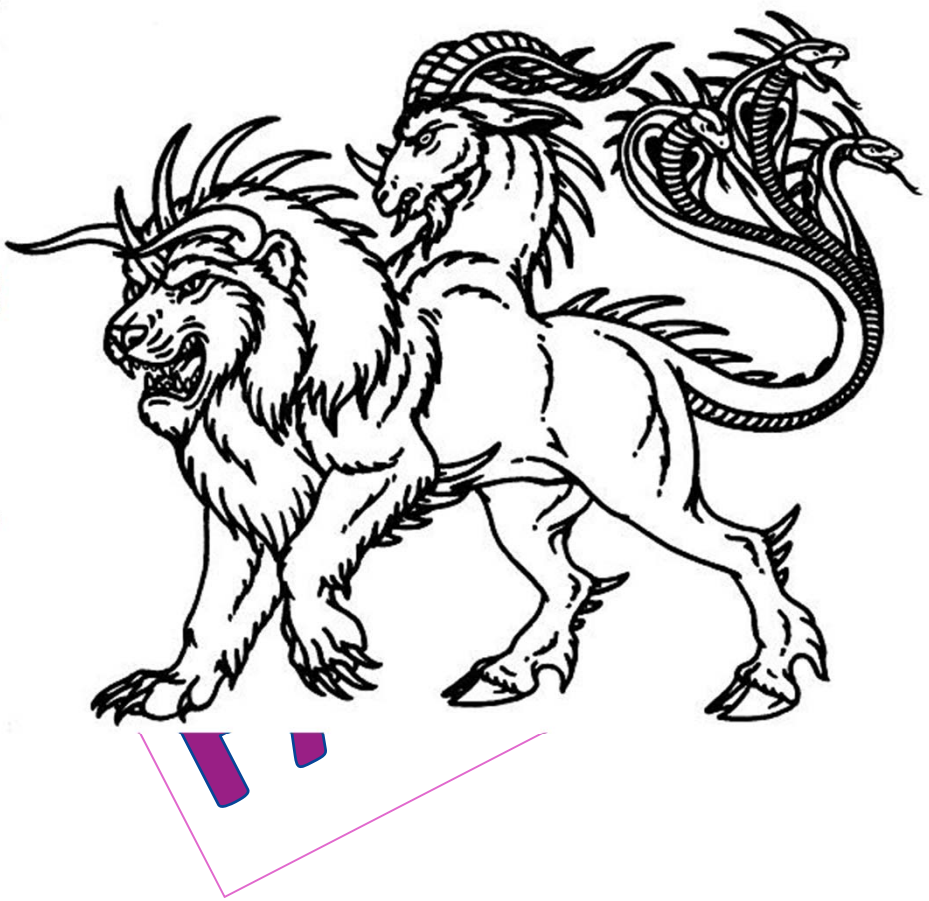
A solution



(In)Famous quotes

I don't know how to s

A certain Commer



Success story # 3
Bankruptcy prediction

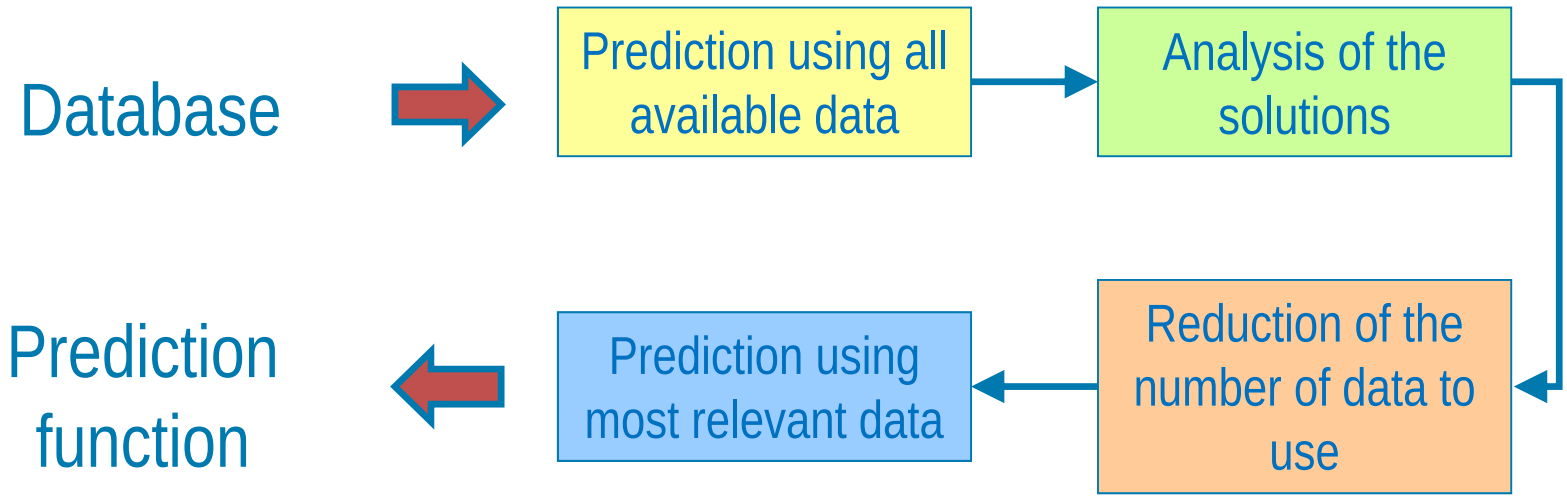
- Given a database of companies data for a given year k, predict:
  - If the company will go **bankrupt** in year $k+n$
  - If the company will **suffer loses** in year $k+n$

- Imbalanced classification problem: there are many more companies that go bankrupt than "healthy" ones

… luckily!

Why use genetic programming?

- It generates **comprehensible** & **analysable** solutions.
- It allows to determine which data are relevant for the prediction.
- The reduction in the number of data brings about the generation of simpler models.



One solution

$y \equiv$ SI $\dfrac{x_0}{x_1 - x_2} \leq x_1 \exp(x_3)$

ENTONCES DEVUELVE $if_0 - x_3 - x_0 - \exp(x_4) - \exp(x_2)$

SI NO DEVUELVE -1

$if_0 \equiv$ SI $x_2 \leq x_0$

ENTONCES DEVUELVE if_1

SI NO DEVUELVE x_1

$if_1 \equiv$ SI $x_2 \leq x_3 + 43.45$ $if_1 \equiv$ SI $x_2 \leq x_3 + 43.45$

ENTONCES DEVUELVE $\dfrac{x_3}{\ln(x_1) + x_2} - 7.43$

SI NO DEVUELVE $\ln(x_0)$

$y > 0 \Rightarrow$ **THE COMPANY WILL GO BANKRUPT**

$y \leq 0 \Rightarrow$ **THE COMPANY WILL NOT GO BANKRUPT**

Famous quotes

Prediction is very difficult, especially of the future

Niels Bohr

(attributed also to Yogi Berra)

Success story #4
Pest control in agriculture

- **Sexual confusion** is a technique aimed at substituting pesticides.
- It consists on diffusing a high amount of female sexual pheromone in order to confuse the males and avoid coupling.
- How? → Using **pheromone dispensers**

Ecology in figures

- 1 kg of pheromone costs 1000 €
- 1 dispenser contains 200 mg of pheromone

  → One dispenser costs 20 cent. (+ manufacturing)
- In 1Ha we need 500 or 1000 dispensers (depending on the pest)

  → The cost is 100 or 200 € / Ha (+ hand work)

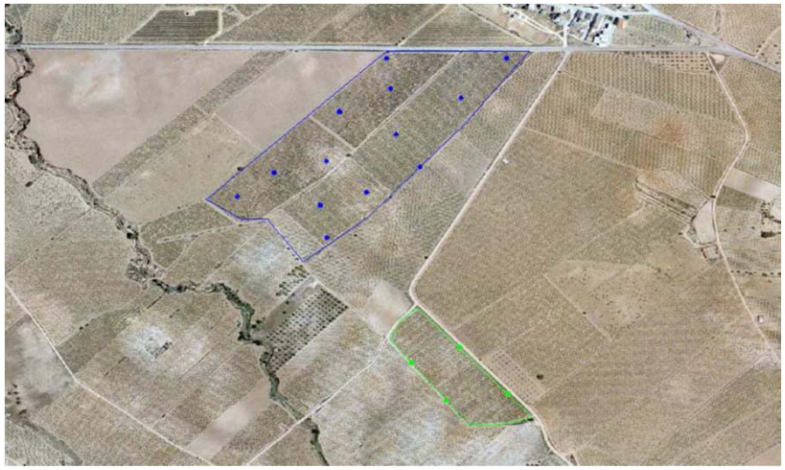
  (real commercial price is 115 or 300 €/Ha)

On the other hand,

- Spraying a traditional pesticide (e.g. Malathion) costs 20-30 €/Ha

“Fields” of application

Vineyards, 500 dispenser per Ha.



Apples and pears,

1000 dispensers per Ha.

Photos courtesy of CEQA - IAM - UPV

CEQA
Centro de Ecología Química Agrícola

Modeling & optimisation for sexual confusion systems

- The price of pheromone dispensers is a limitation for their massive application
  - Aim is to **optimise their number and location** to ensure maximal crop protection at an affordable price
  - Besides, **models** are required that allow making the costs of registration of the products cheaper

Aplication of Genetic Programming

Work carried out: Finding models of release kinetics

Let the residual r be the percentage of product not released into the atmosphere

For a given dispenser, find a function r= r(t) where t = time

Data: a sequence of points (r,t) obtained

in field conditions

Measuring r is costly → few measures,

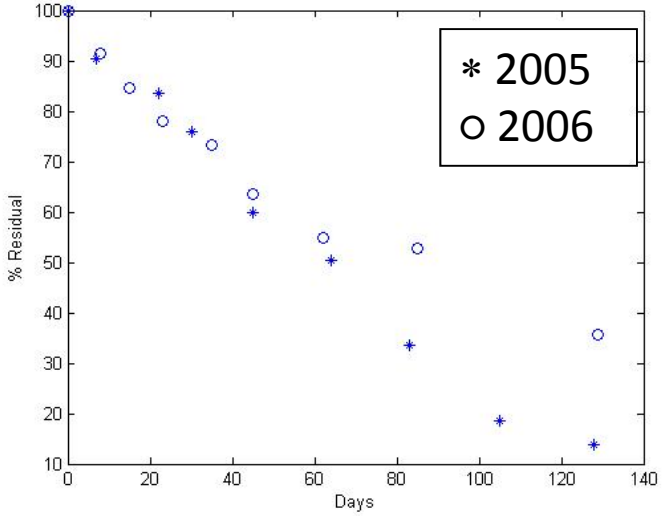
Measures unequispaced → more at the

beginning, when the release is faster)

Resolution with MOGP

Cost function ("leave-one-out"):

$$Obj_i = |r_{calculated\ (i)} - r_{measured\ (i)}|$$



Further work:

Find models of pheromone distribution in the environment

Use these models to optimise the placement of dispensers and make cheaper the registration costs

Famous quotes

Us people from Madrid go to the countryside to ensure cows are not purple

J. Ignacio Hidalgo, 2009

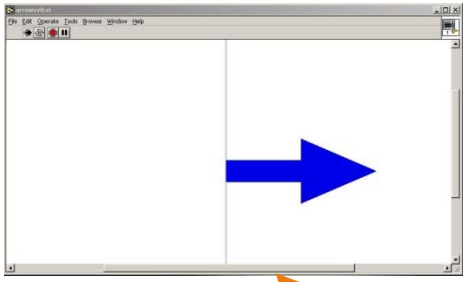
Success story #5
Bio signal Classification

Applications in:

- BCI for comunication & mobility of disabled persons,
- Neuromarketing
- Usability

Signal sampler and digitaliser

Electrodes cap

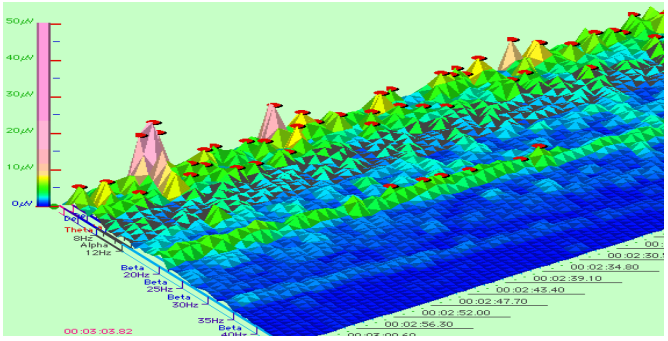


Active electrodes

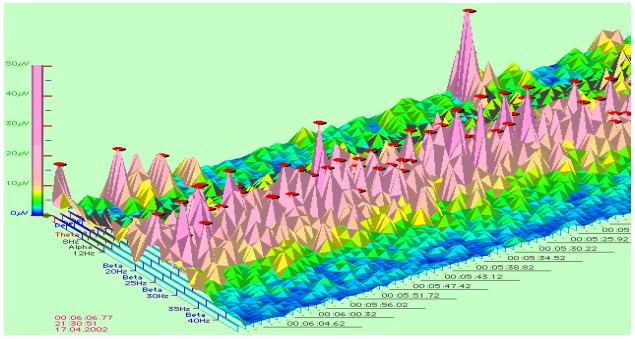
Training Programme

Brain-computer interface (BCI)

- Objective: **Control of devices by thought** by means of the measurement and analysis of the electrical activity in the brain using electroencefalogram (EEG)
- Based on the fact that **different thoughts originate diferent EEG patterns**

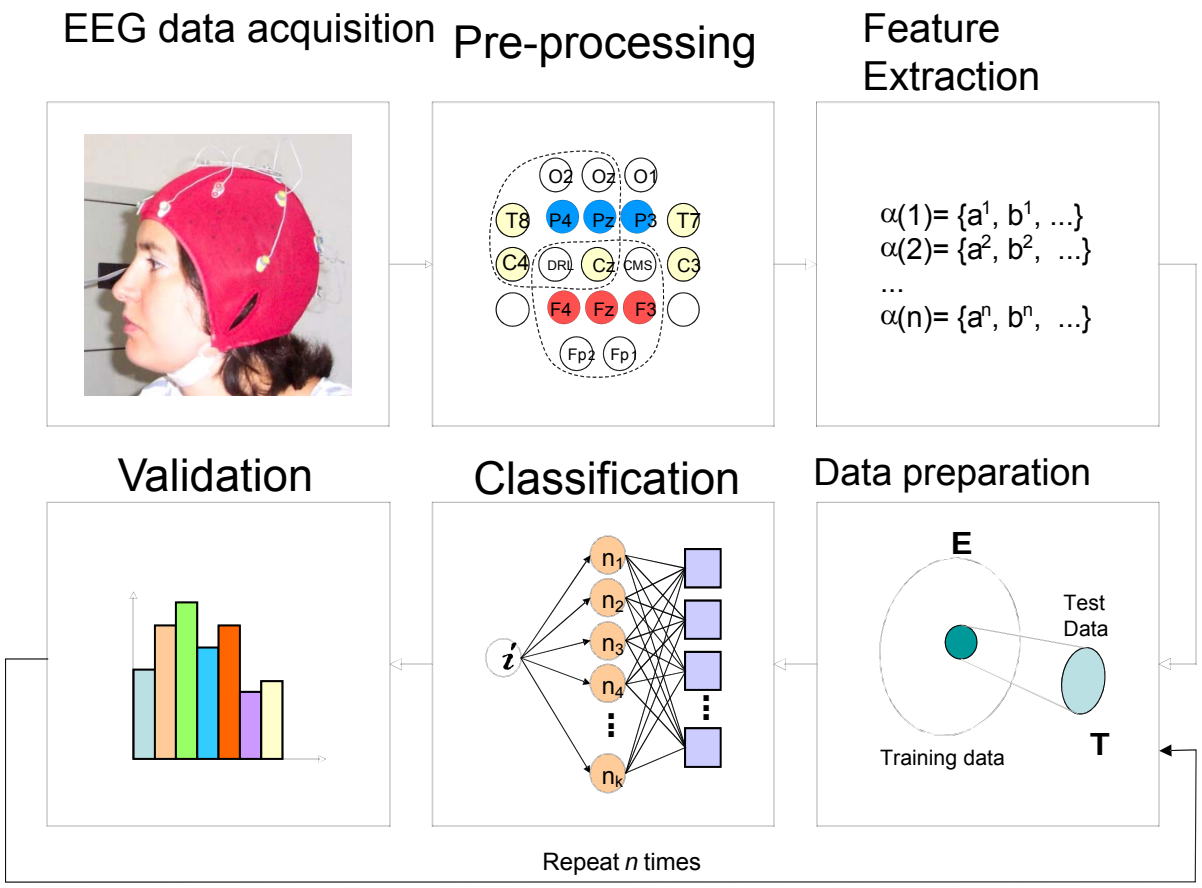


User relaxed, with eyes closed



User waching TV

Biosignal classification using CI



(In)Famous quotes

Who told you we wanted you to publish?

A Certain Manageress at a Certain Technological Institute, 2008

My History

Chapter n+1:

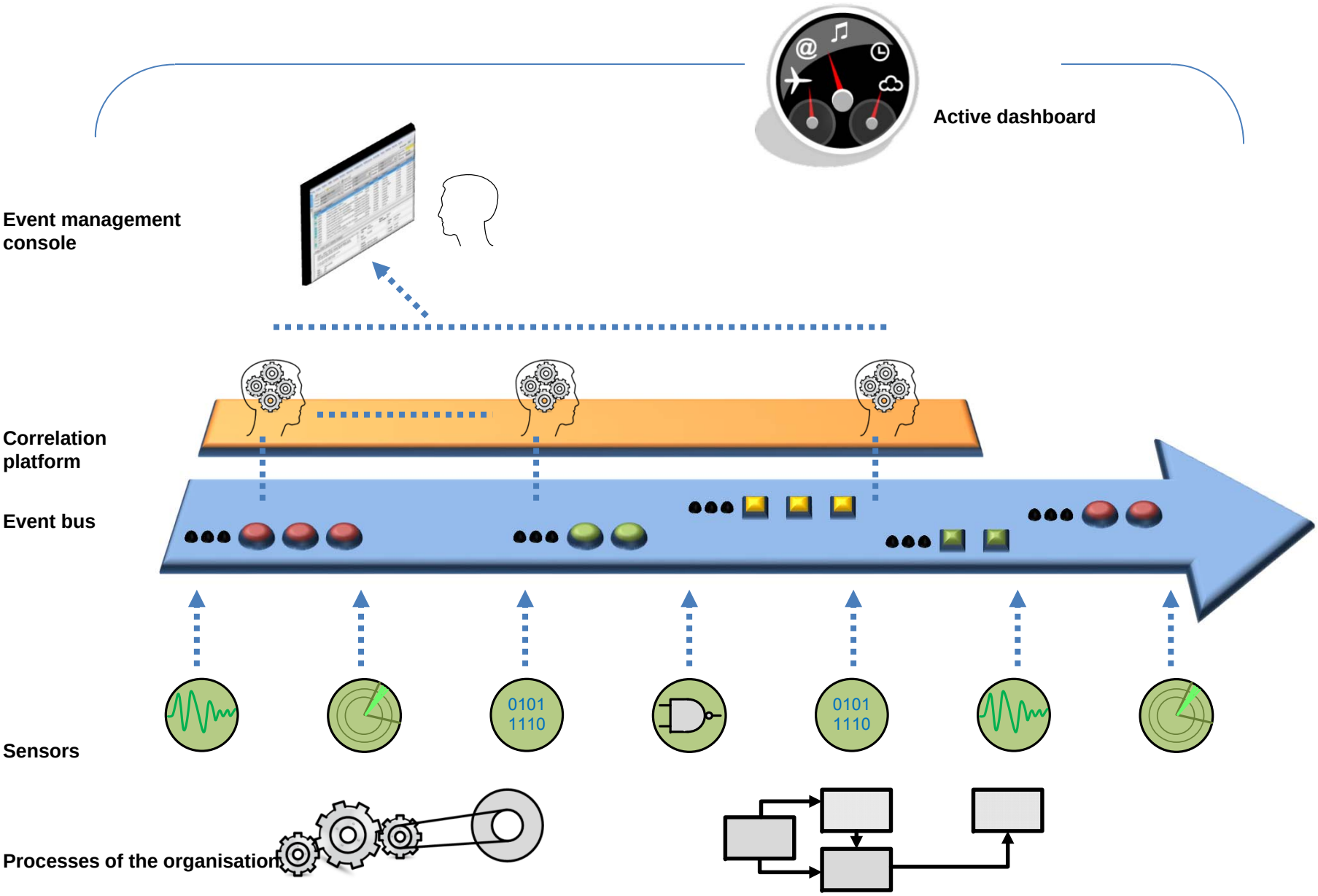
When I was in the “other” other side

Success stories (ii)

6. Evolving rules for a correlation system
7. Detection of Advanced Persistent Threats

A model company

- An SME specialising in cybersecurity that has developed its own product suite, which includes:
  - Sensors to measure physical, logical and business variables
  - A rule-based correlation platform, which triggers higher level events
  - An event management console, where technicians handle events


Active dashboard
Event management console
Correlation platform
Event bus
0101
1110
0101
1110
Sensors
Processes of the organisation

Beware of dangerous questions!

Is this an expert system?

Anna Esparcia to her boss, 2011

(In)Famous quotes

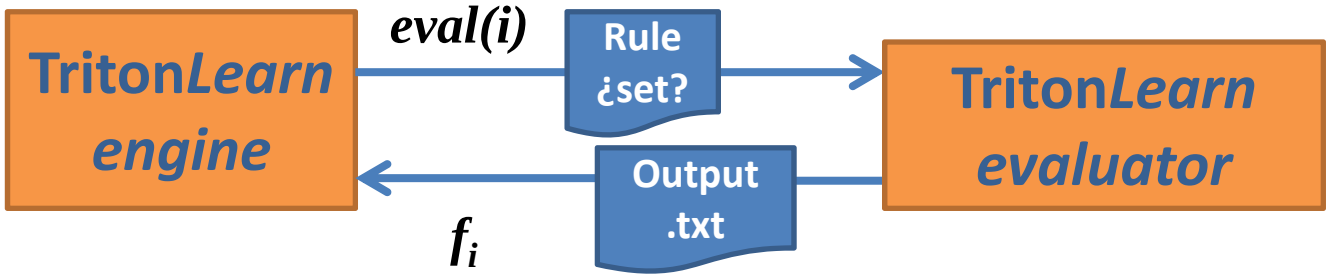
Expert systems have been obsolete for 10 years

Anna Esparcia to her boss, 2001

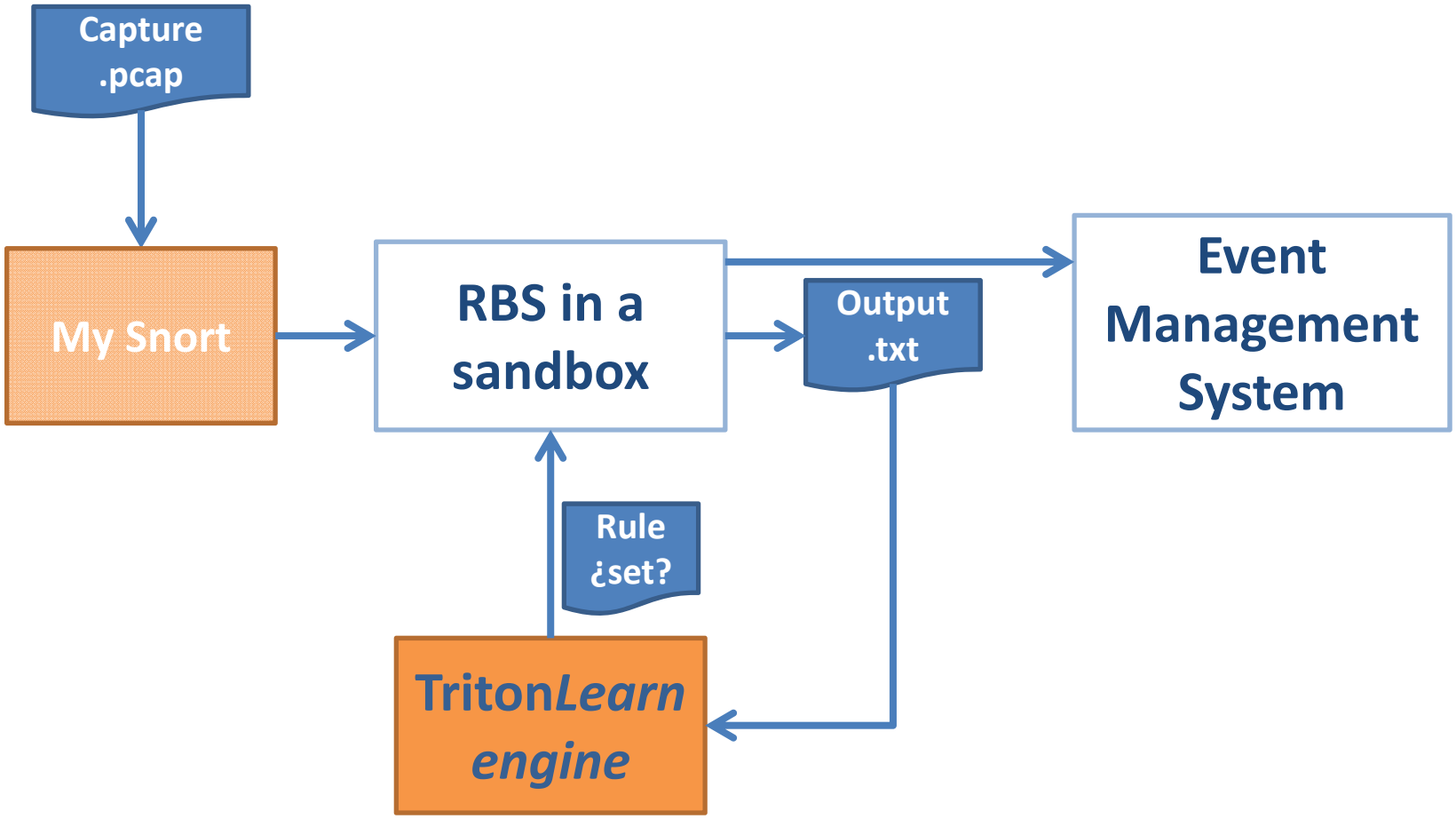
Problems with expert systems

- Technical Problems
  - Consistency: difficulty in detecting erroneous knowledge, handle differences in expert's opinion, verify the correct interaction between rules
  - Scalability: small increases in system size make knowledge management and especially acquisition become unfeasible
  - Static, no learning
  - Subjective way of representing knowledge: difficult to back up expert-defined rules with physical or statistical data
- Structural Problems
  - Cost of acquisition and maintenance of knowledge
- Human Problems
  - Knowledge Extraction from experts
  - Distribution of incompetence (Peter Principle applied to *pseudo-experts*)
  - Legal issues

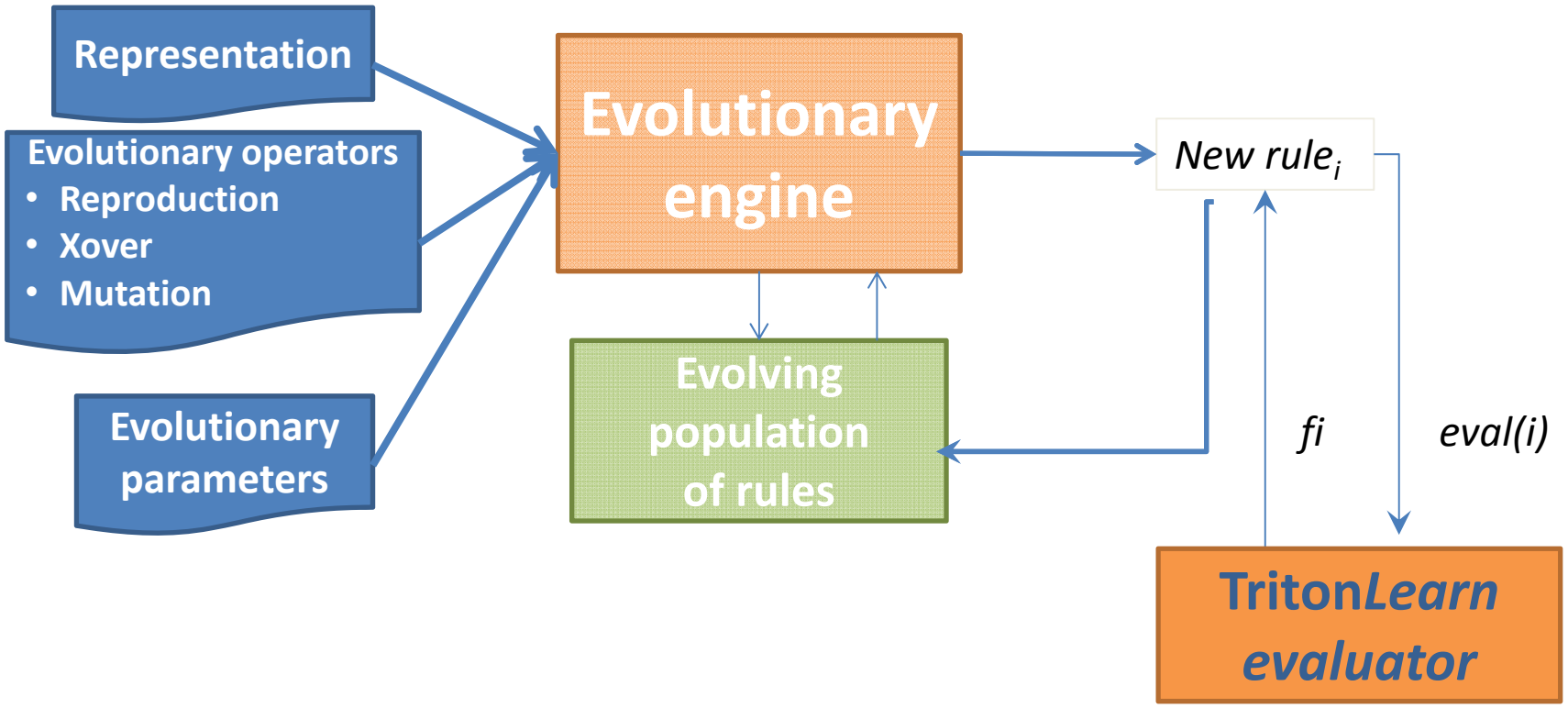
Success story #6: TritonLearn - Making a rule-based system “smart” using CI



Triton*Learn*

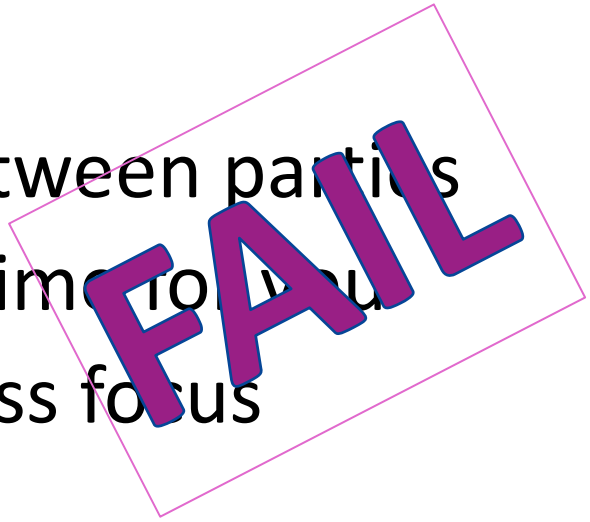


Triton*Learn*



TritonLearn problems

- Unconnected systems
- No logs
- Lack of understanding between parties
- Technicians do not have time for you
- Company changes business focus



Solution:

- Let’s do something else!

Success story #7: Advanced Persistent Threats, the new hype in security

A possible definition of APT:

A highly sophisticated targeted attack involving very skilled teams and significant financial resources

- One of the steps when they can be detected is called *exfiltration*
- During exfiltration it is **assumed** that the behaviour of the system will be anomalous

→we'll focus on **anomaly detection**

Example of http traffic

```
200 GET text/plain com 276 55490
notify9.dropbox.com
200 GET text/plain com 276 55246
notify5.dropbox.com
200 POST null com 197 307
twitterfall.com
200 GET application/octet-stream com
273 75962 su.ff.avast.com
```

… up to 5M instances

Available Data: HTTP session

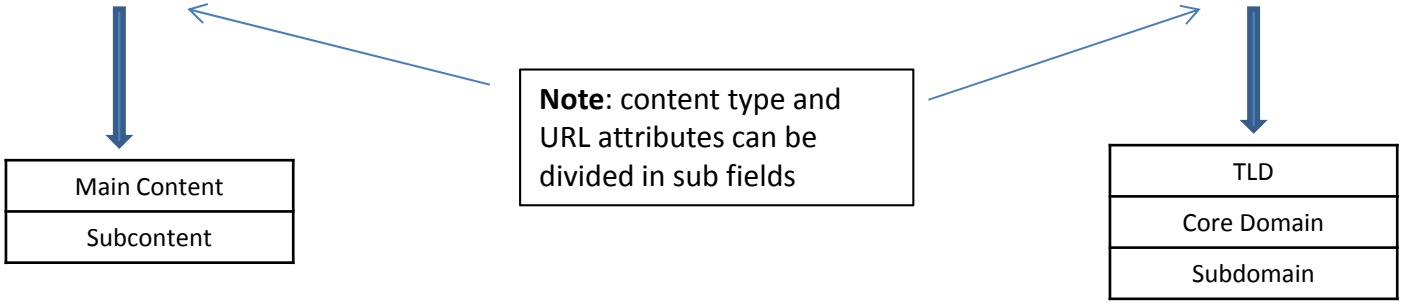
- Characteristics:
  - 5 Million instances
  - 10 attributes/fields per instance
  - 4 types: categorical [c], numerical [n], string [s], timel [t].

http reply code	c
http method	c
Duration (ms)	n
Content type	c
Server IP	c
Time	t
Squid hierarchy	c
Bytes	n
Client IP	c
URL	s

- General data properties: multi-attribute, several types of attributes, sequential and data from different users.

Data: Example

http reply_code	http method	Duration (ms)	content_type	server address	Time	Squid hierarchy	bytes	Client address	url (reduced)
200	GET	2014	application/javascript	192.168.4.4	05/03/2014 6:26:55	DEFAULT_PARENT	798	10.159.74.184	//syndication.twimg.com
200	GET	55581	text/plain	108.160.162.33	05/03/2014 6:26:57	DIRECT	276	192.168.4.4	//syndication.twimg.com
200	GET	55622	text/plain	192.168.4.4	05/03/2014 6:26:58	DEFAULT_PARENT	326	10.159.74.184	//notify5.dropbox.com
200	POST	135	application/vnd.google.safebrowsing-update	173.194.34.231	05/03/2014 6:27:02	DIRECT	885	192.168.4.4	//notify5.dropbox.com
200	POST	139	application/vnd.google.safebrowsing-update	192.168.4.4	05/03/2014 6:27:02	DEFAULT_PARENT	970	10.159.128.70	//safebrowsing.clients.google.com
200	GET	105	application/vnd.google.safebrowsing-chunk	173.194.34.232	05/03/2014 6:27:06	DIRECT	4133	192.168.4.4	//safebrowsing.clients.google.com
200	GET	111	application/vnd.google.safebrowsing-chunk	192.168.4.4	05/03/2014 6:27:07	DEFAULT_PARENT	4183	10.159.128.70	//safebrowsing-cache.google.com



Unbalanced binary classification

- Two clases: suspicious / non-suspicious
- Labels are assigned by experts
- To obtain the training/testing sets
  - Step 1: Filter all data using anomaly score
  - Step 2: Expert classifies anomalous data

Abnormal attributes and combinations

- Detection of **abnormal individual** values**:**
  - ***Infrequent values*** in categorical attributes/characteristics.
  - ***Values out of the normal range*** distribution in num attributes/characteristics.
  - Logs in ***abnormal time*** interval.

- Detection of **abnormal combinations**:
  - Of code, method, content and tld attributes.

Semi-Supervised Learning

- Design method to automatically detect suspicious instances
  - Hypothesis: instances not labelled are non-suspicious.
  - 2/3 lab instances for training and 1/3 for test
  - Categorical attributes converted to integers
- Methods applied
  - SVM
  - GP
  - treefit (CART tree)

Problems

- Technicians still do nc
- Company business fo
- Difficult to access rea

(In)Famous quotes

I just want a generic system that detects ***everything***

A Certain Director of Development in a Certain Company, 2013

Famous quotes

Success is the ability o
failure without losing

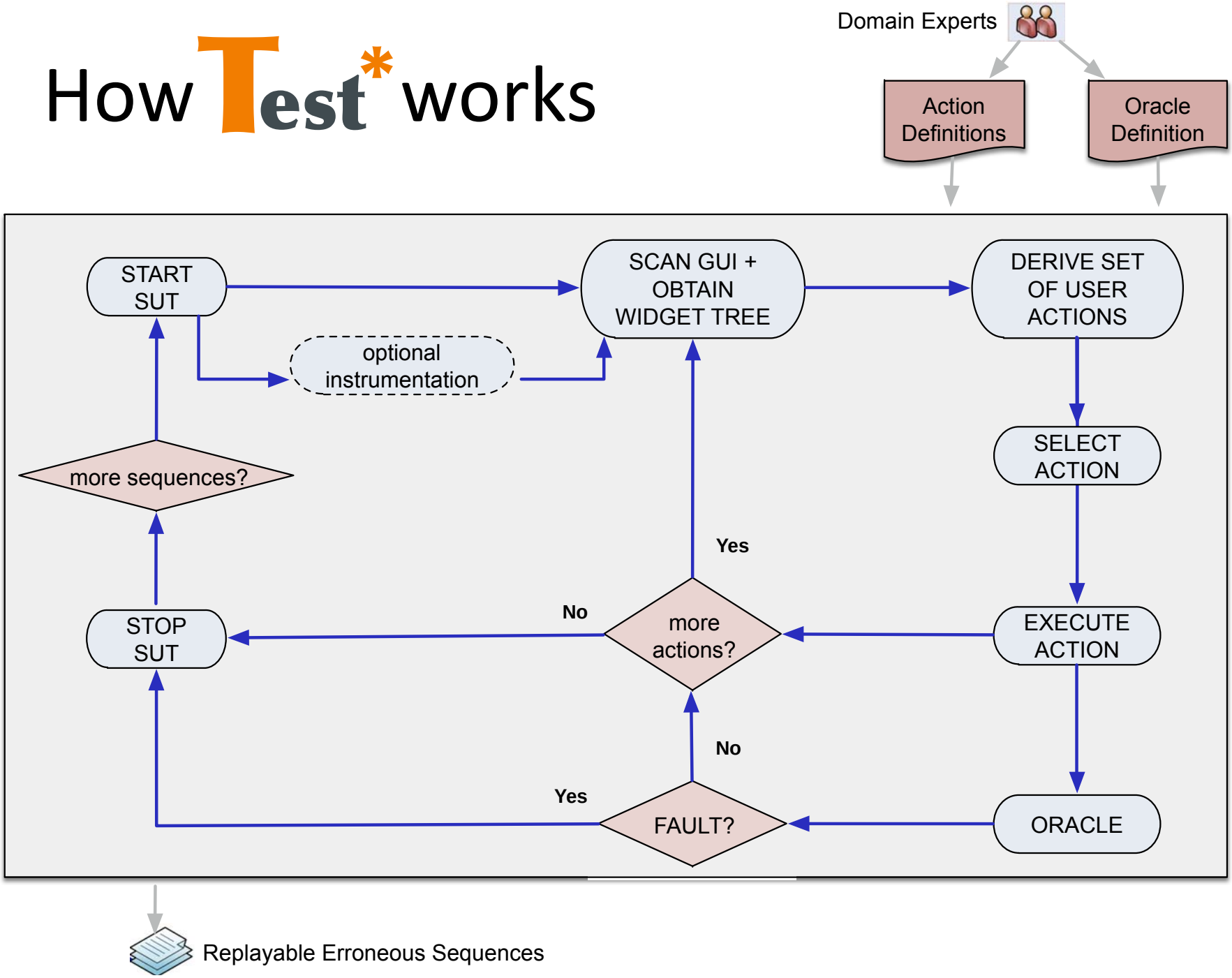
Now

Chapter n+2: Full circle

Success story #8: Automated software testing

TESTAR is a tool developed as part of European project FITTEST (2010-2013)

How Test* works



Test* in the “real world”

- In 2006 the UPV launched a programme called “Proof of concept” aimed at transferring results of research to companies
- As part of this programme, TESTAR has been deployed in 10 companies so far
- We have set up the Spanish Software Testing Innovation Alliance

THE SOFTWARE TESTING
INNOVATION
ALLIANCE

Some possible keys to success

- Optimism
  - Believe in what you do
  - Deal with negativity and resistance
- Determination
  - Keep trying
  - Follow up with companies
- Visibility
  - Never miss a chance to be an evangelist
  - Use social media


We are
* TEST *
stars
www.testar.org
We are
TEST*
stars
www.testar.org
ICT
INNOVATE
CONNECT
TRANSFORM
We are
* TEST *
stars
We are
* TEST *
We are
* TEST *
stars
www.testar.org

Famous quotes

In a good cause there are no failures, only delayed successes

Isaac Asimov, “The complete stories”

Take home ideas

Famous quotes

Sometimes I have believed six impossible things before breakfast

L. Carroll, "Through the looking glass and what Alice found there"

Six impossible things

- There is life beyond the h-index
- You can apply EC in a company – even an SME, even in a “lesser” country
- You can understand what companies want
- You can convince them that EC is not just a PR stunt, or “vapourware”
- Customers will *appreciate* (perhaps even understand) what you do
- Your boss will be happy

So… perhaps the Chimera exists, and is as fierce as you want to make it

Thanks to:

- Lewis Carroll: “Alice in Wonderland” and “Through the looking glass and what Alice found there”
- Lewis Wolpert: “Six impossible things before breakfast”
- Arthur Kordon: “Applied Computational Intelligence: How to create value”


Any
questions?