Towards a Holistic Theory of Pattern Recognition: A Game-Theoretic Perspective

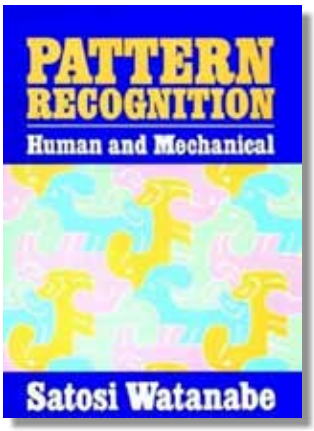
Marcello Pelillo
Ca' Foscari University, Venice

Keynote lecture, ICPRAM 2015, Lisbon, Portugal, 11 January 2015

Our essentialist assumption

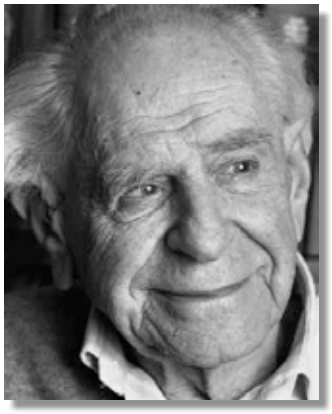
«Whether we like it or not, under all works of pattern recognition lies tacitly the Aristotelian view that the world consists of a discrete number of self-identical objects provided with, other than fleeting accidental properties, a number of fixed or very slowly changing attributes. Some of these attributes, which may be called "features," determine the class to which the object belongs.»

Satosi Watanabe
Pattern Recognition: Human and Mechanical (1985)



Essentialism and its discontents

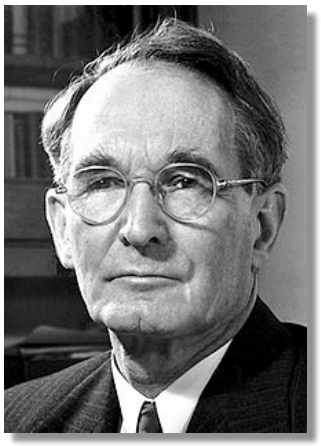
«The development of thought since Aristotle could be summed up by saying that every discipline, as long as it used the Aristotelian method of definition, has remained arrested in a state of empty verbiage and barren scholasticism, and that **the degree to which the various sciences have been able to make any progress depended on the degree to which they have been able to get rid of this essentialist method.**»



Karl Popper
The Open Society and Its Enemies (1945)

Definitions in Physics

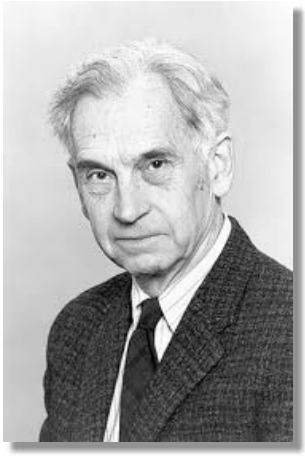
«**What do we mean by the length of an object?**
[...]
To find the length of an object, we have to perform certain physical operations. The concept of length is therefore fixed when the operations by which length is measured are fixed
[...]
In general, we mean by any concept nothing more than a set of operations; **the concept is synonymous with the corresponding set of operations.**»



Percy W. Bridgman
The Logic of Modern Physics (1927)

Can we be essentialist after Darwin?

«Essentialism [...] dominated the thinking of the western world to a degree that is still not yet fully appreciated by the historians of ideas. [...]
It took more than two thousand years for biology, under the influence of Darwin, to escape the paralyzing grip of essentialism.»



Ernst Mayr
The Growth of Biological Thought (1982)

Against "classical" categories

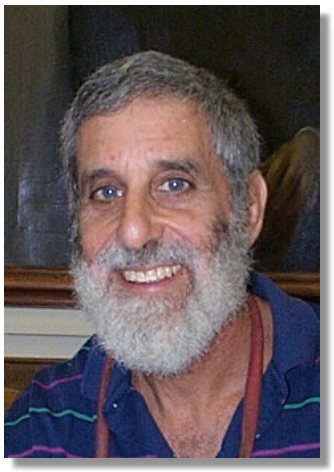
«Categorization is a central issue. The traditional view is tied to the classical theory that categories are defined in terms of common properties of their members.
But a wealth of new data on categorization appears to contradict the traditional view of categories. In its place there is a new view of categories, what Eleanor Rosch has termed *the theory of prototypes and basic-level categories*.»

George Lakoff
Women, Fire, and Dangerous Things (1987)

What is the subject-matter of math?

«In mathematics the primary subject-matter is not the individual mathematical objects but rather the structures in which they are arranged.»

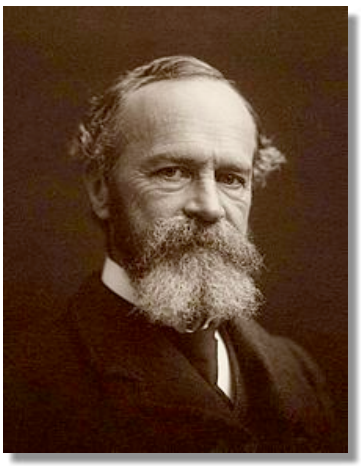
Michael D. Resnik
Mathematics as a Science of Patterns (1997)



"Signal" *vs.* "noise"

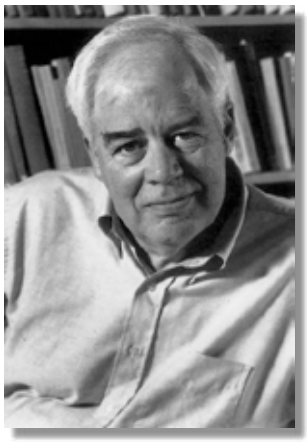
«*There is no property* ABSOLUTELY *essential to any one thing.* The same property which figures as the essence of a thing on one occasion becomes a very inessential feature upon another.»

William James
The Principles of Psychology (1890)



Epistemic anti-essentialism

«We antiessentialists would like to convince you that it [...] does not pay to be essentialist about tables, stars, electrons, human beings, academic disciplines, social institutions, or anything else. We suggest that you think of all such objects as resembling numbers in the following respect: **there is nothing to be known about them except an initially large, and forever expandable, web of relations to other objects**.



There are, so to speak, relations all the way down, all the way up, and all the way out in every direction: you never reach something which is not just one more nexus of relations.»

Richard Rorty
A World Without Substances or Essences (1994)

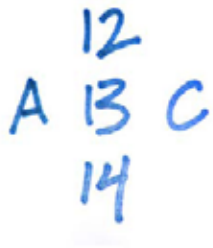
Two consequences of the essentialist assumption

Our essentialist attitude has had two major consequences which greatly contributed to shape the ML/PR fields in the past few decades.

- ✓ it has led the community to focus mainly on **feature-vector representations**, where, each object is described in terms of a vector of numerical attributes and is therefore mapped to a point in a Euclidean (geometric) vector space

- ✓ it has led researchers to maintain a **reductionist position**, whereby objects are seen in isolation and which therefore tends to overlook the role of contextual, or relational, information

Context helps ...



c → cat
→ circus

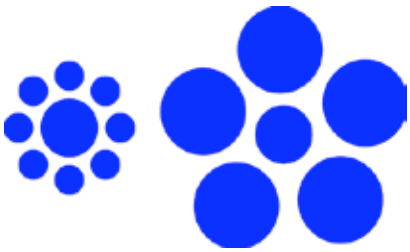
i → sin
→ fine

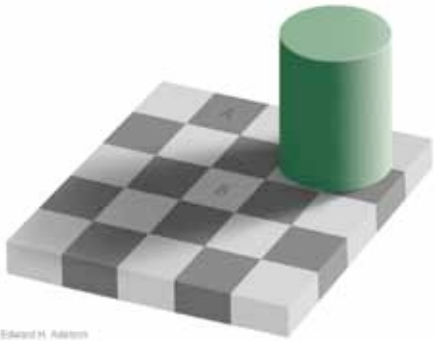
e → red
→ read

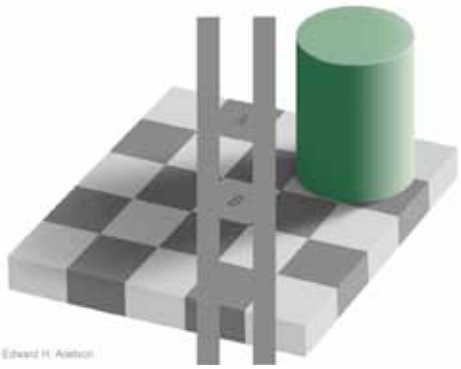
festival

graphics

... but can also deceive







What do you see?

Figure 2. The strength of context. The visual system makes assumptions regarding object identities according to their size and location in the scene. In this picture, observers describe the scene as containing a car and pedestrian in the street. However, the pedestrian is in fact the same shape as the car, except for a 90° rotation. The atypicality of this orientation for a car within the context defined by the street scene causes the car to be recognized as a pedestrian.

From: A. Oliva and A. Torralba, "The role of context in object recognition", Trends in Cognitive Sciences, 2007.

Context and the brain

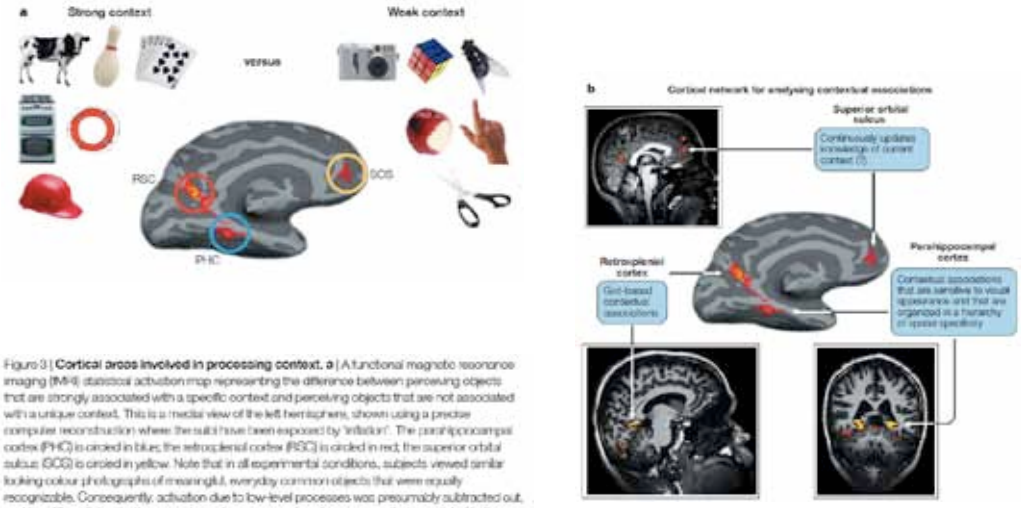


Figure 3 | **Cortical areas involved in processing context. a** | A functional magnetic resonance imaging (fMRI) statistical activation map representing the difference between perceiving objects that are strongly associated with a specific context and perceiving objects that are not associated with a unique context. This is a medial view of the left hemisphere, shown using a precise computer reconstruction where the cuts have been exposed by 'inflation'. The parahippocampal cortex (PHC) is circled in blue; the retrosplenial cortex (RSC) is circled in red; the superior orbital sulcus (SOS) is circled in yellow. Note that in all experimental conditions, subjects viewed similar looking colour photographs of meaningful, everyday common objects that were equally recognizable. Consequently, activation due to low-level processes was presumably subtracted out, and the differential activation map shown here represents only processes that are related to the level of contextual association. **b** | The cortical network for contextual associations among visual objects, suggested on the basis of existing evidence. Other types of context might involve additional regions (for example, hippocampus for navigation[31] and Broca's area for language-related context). Modified, with permission, from *NN* 12 © (2003) Elsevier Science.

From: M. Bar, "Visual objects in context", Nature Reviews Neuroscience, August 2004.

Hume's similarity principle

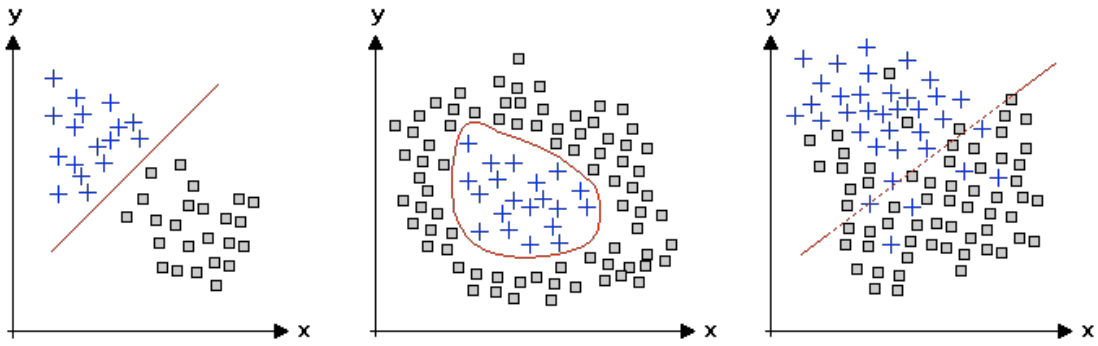
«I have found that such an object has always been attended with such an effect, and I foresee, that other objects, which are, in appearance, similar, will be attended with similar effects.»

David Hume
An Enquiry Concerning Human Understanding
(1748)

See also the "homophily" principle in social network analysis.

Today's view: Similarity as a by-product

Traditional machine learning and pattern recognition techniques are centered around the notion of **feature-vector,** and derive object similarities from vector representations.



Limitations of feature-vector representations

There are situations where either it is not possible to find satisfactory feature vectors or they are inefficient for learning purposes.

This is typically the case, e.g.,

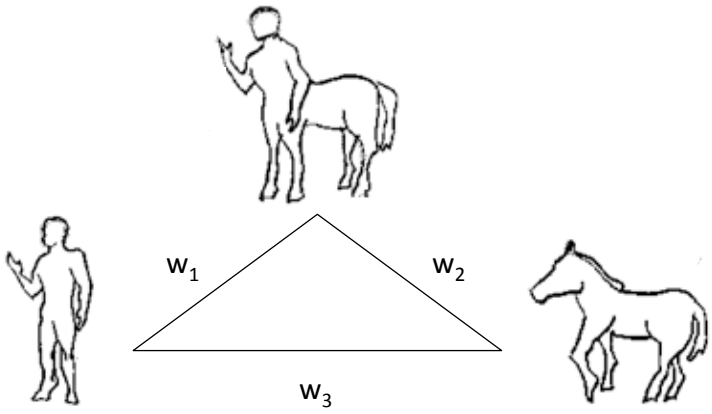
- ✓ when data are high dimensional (e.g., images)
- ✓ when features consist of both numerical and categorical variables
- ✓ in the presence of missing or inhomogeneous data
- ✓ when objects are described in terms of structural properties, such as parts and relations between parts, as is the case in shape recognition
- ✓ in the presence of purely relational data (graphs, hypergraphs, etc.)
- ✓ …

Application domains: Computational biology, adversarial contexts, social signal processing, medical image analysis, social network analysis, document analysis, network medicine, etc.

The need for non-metric similarities

«Any computer vision system that attempts to faithfully reflect human judgments of similarity is apt to devise non-metric image distance functions.»

Jacobs, Weinshall and Gdalyahu, 2000



$w_3 > w_1 + w_2$

Adapted from: D. W. Jacobs, D. Weinshall, and Y. Gdalyahu. Classication with non-metric distances: Image retrieval and class representation. PAMI 2000.

The symmetry assumption

«Similarity has been viewed by both philosophers and psychologists as a prime example of a symmetric relation. Indeed, the assumption of symmetry underlies essentially all theoretical treatments of similarity.

Contrary to this tradition, the present paper provides empirical evidence for asymmetric similarities and argues that **similarity should not be treated as a symmetric relation**.»

Amos Tversky
Features of Similarities (1977)

Examples of asymmetric (dis)similarities:

- ✓ Kullback-Leibler divergence
- ✓ Directed Hausdorff distance
- ✓ Tversky's contrast model

Towards a Paradigm Shift?

The field is showing an increasing propensity towards anti-essentialist/ relational approaches, e.g.,

- ✓ Kernel methods
- ✓ Pairwise clustering (e.g., spectral methods, game-theoretic methods)
- ✓ Graph transduction
- ✓ Dissimilarity representations (Duin et al.)
- ✓ Theory of similarity functions (Blum, Balcan, …)
- ✓ Relational / collective classification
- ✓ Graph mining
- ✓ Adversarial learning
- ✓ Contextual object recognition
- ✓ …

See also the parallel development of "network science" …

A Game-Theoretic Perspective

What is Game Theory?

«The central problem of game theory was posed by von Neumann as early as 1926 in Göttingen. It is the following: If n players, P_1,..., P_n, play a given game Γ, how must the i^{th} player, P_i, play to achieve the most favorable result for himself?»

Harold W. Kuhn
Lectures on the Theory of Games (1953)

A few cornerstones in game theory

1921–1928: Borel and Von Neumann give the first modern formulation of a mixed strategy along with the idea of finding minimax solutions

1944, 1947: Von Neumann and Morgenstern publish *Theory of Games and Economic Behavior*

1950–1953: Nash made seminal contributions to non-cooperative game and bargaining theory

1972–1982: Maynard Smith applies game theory to biological problems, thereby launchig "evolutionary game theory"

late 1990's –: Development of algorithmic game theory...

"Solving" a game

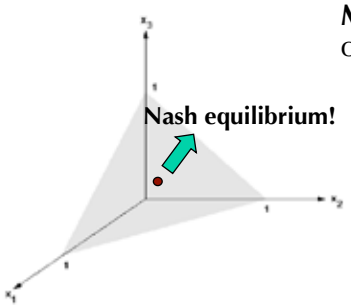
Nash equilibrium: no player has an incentive to deviate unilaterally from it.

		Player 2		
		Left	Middle	Right
Player 1	Top	3 , 1	2 , 3	10 , 2
	High	4 , 5	3 , 0	6 , 4
	Low	2 , 2	5 , 4	12 , 3
	Bottom	5 , 6	4 , 5	9 , 7

Nash equilibrium!

Mixed-Strategy Solutions

Mixed strategy = probability distribution $\mathbf{x}=(x_1,\ldots,x_n)^T$ over the set of "pure" strategies

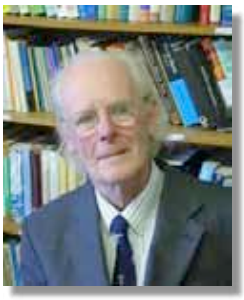


		You		
		Rock	Scissors	Paper
Me	Rock	0 , 0	1 , -1	-1 , 1
	Scissors	-1 , 1	0 , 0	1 , -1
	Paper	1 , -1	-1 , 1	0 , 0

Evolutionary Game Theory

«We repeat most emphatically that our theory is thoroughly static. A dynamic theory would unquestionably be more complete and therefore preferable.»

John von Neumann and Oskar Morgenstern
Theory of Games and Economic Behavior (1944)



«Paradoxically, it has turned out that game theory is more readily applied to biology than to the field of economic behaviour for which it was originally designed.»

John Maynard Smith
Evolution and the Theory of Games (1982)

Evolutionary Game Theory

Assumptions:

- ✓ A large population of individuals belonging to the same species which compete for a particular limited resource
- ✓ This kind of conflict is modeled as a two-player game, the players being pairs of randomly selected population members
- ✓ Players do not behave "rationally" but act according to a pre-programmed behavioral pattern (pure strategy)
- ✓ Utility is measured in terms of Darwinian fitness, or reproductive success

Key notion:

Evolutionary Stable Strategies (ESS's) = "stable" version of Nash equilibria.

Finding ESS's: Replicator Dynamics

Replicator dynamics are a populare way to find ESS's and are motivated by Darwin's principle of natural selection:

$$x_i(t+1) = x_i(t)\frac{A(x(t))_i}{x(t)^T Ax(t)}$$

where $x_i(t)$ is the population share playing strategy i at time t, and A is the payoff matrix.

MATLAB implementation

```
distance=inf;
while distance>epsilon
    old_x=x;
    x = x.*(A*x);
    x = x./sum(x);
    distance=pdist([x,old_x]');
end
```

The Clustering Problem

Given:

- a set of n "objects"
- an $n \times n$ matrix A of pairwise similarities

} = an edge-weighted graph

Goal: Group the the input objects (the vertices of the graph) into maximally homogeneous classes (i.e., clusters).

What is a Cluster?
A Game-Theoretic Perspective

No universally accepted (formal) definition of a "cluster" but, informally, a cluster should satisfy two criteria:

Internal criterion
all "objects" inside a cluster should be highly similar to each other

External criterion
all "objects" outside a cluster should be highly dissimilar to the ones inside

An answer from game theoy

The classical notion of ESS equilibrium provides a general and elegant answer to the question above.

The Clustering Game

In the (pairwise) clustering game we have:

- ✓ Two players (because we have pairwise affinities)
- ✓ Pure strategies = objects to be clustered
- ✓ Payoff matrix = similarity matrix

It is in each player's interest to pick an element that is similar to the one that the adversary is likely to choose.

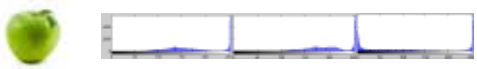
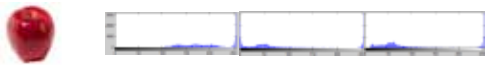
ESS's abstract well the main characteristics of a cluster (Pavan and Pelillo, 2007; Pelillo et al., 2013):

- ✓ **Internal coherency**: High mutual support of all elements within the group
- ✓ **External incoherency**: Low support from elements of the group to elements outside the group

Special cases:

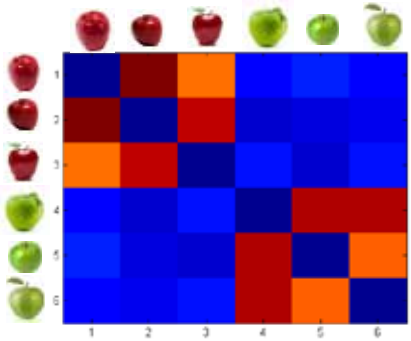
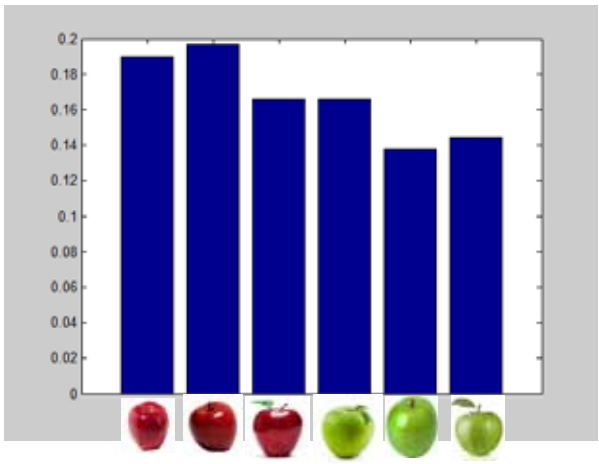
- ✓ *Binary similarities:* ESS-clusters = maximal cliques
- ✓ *Symmetric similarities:* ESS-clusters = optima of global coherency function

A Toy Example

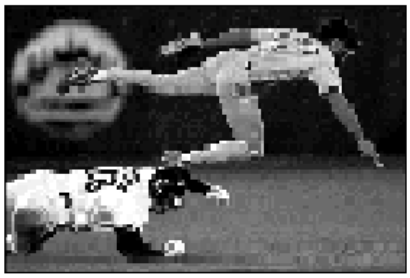



Payoff between two pair of apples is computed as the distance between the two RGB histogram (EMD, ChiSq, Euc…).

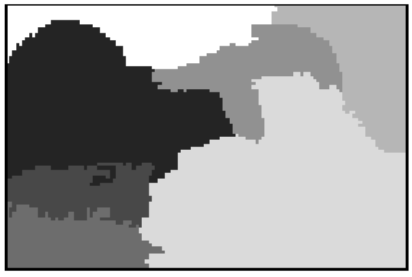
$$a_{i,j} = e^{-\frac{\|i-j\|}{\sigma}} \qquad x_i(t+1) = x_i(t)\frac{(Ax(t))_i}{x(t)^T Ax(t)}$$

An Example Application: Image Segmentation

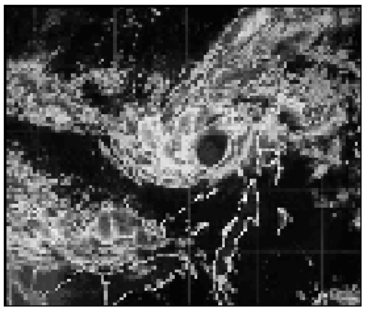


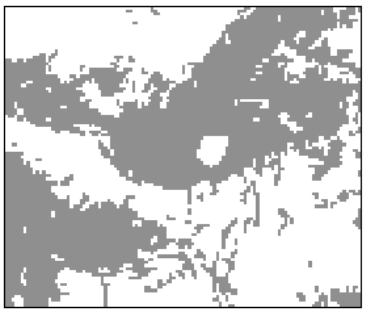
Dominant sets



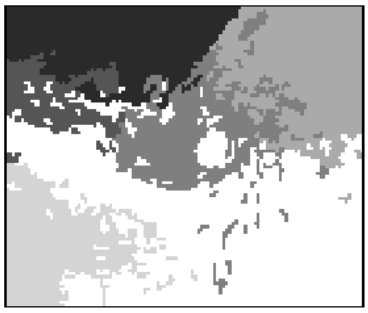
Ncut

An Example Application: Image Segmentation



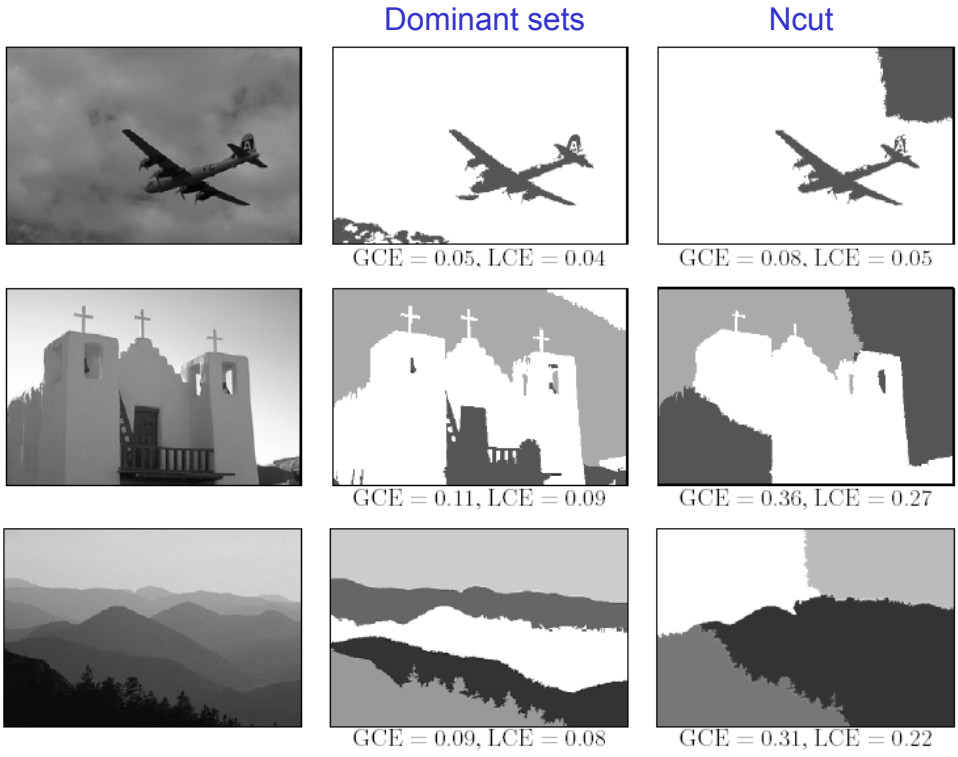


Dominant sets

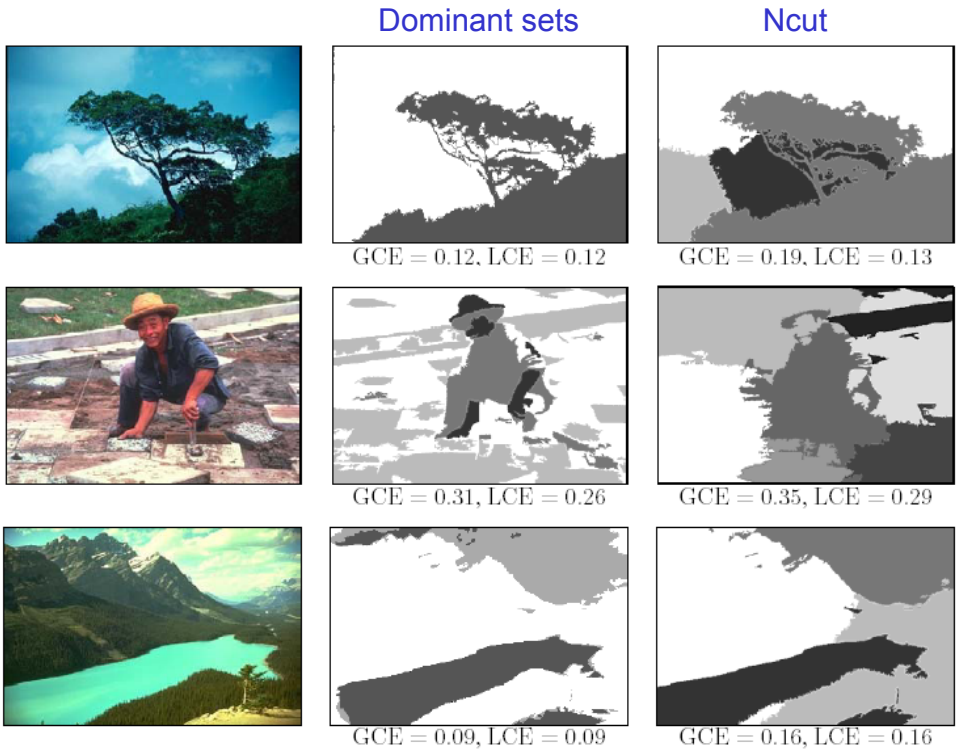


Ncut

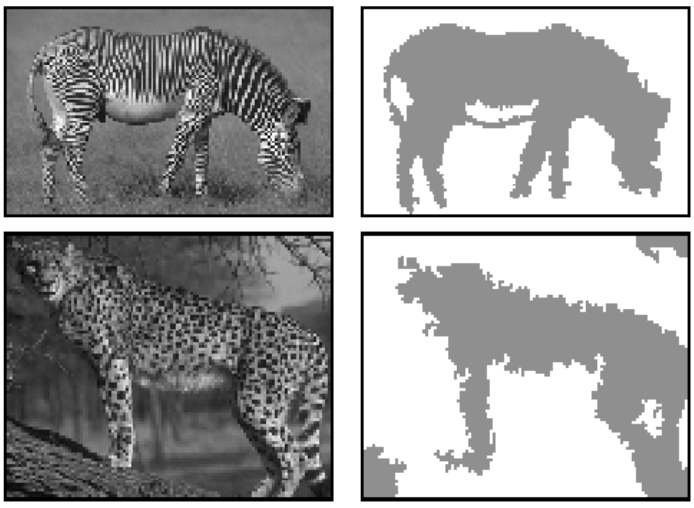
Results on Berkeley Dataset



Results on Berkeley Dataset

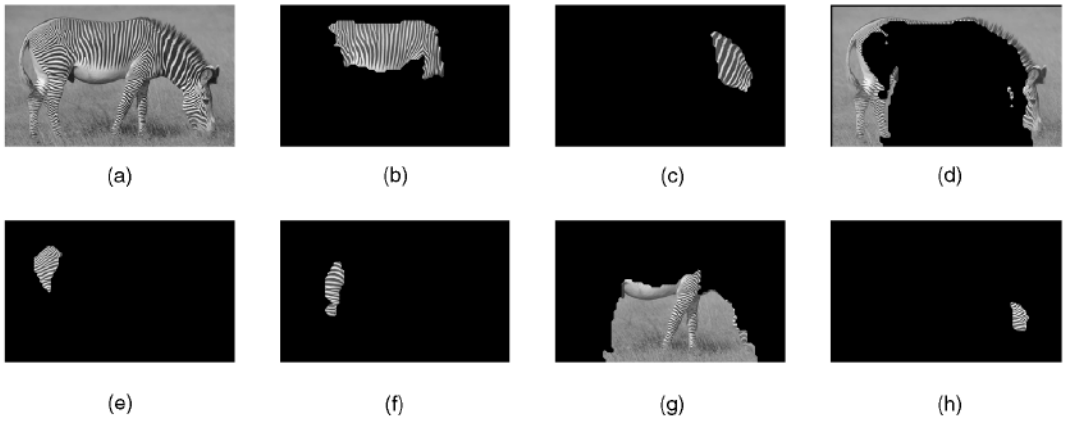


Results on Berkeley Dataset



Dominant sets

Results on Berkeley Dataset

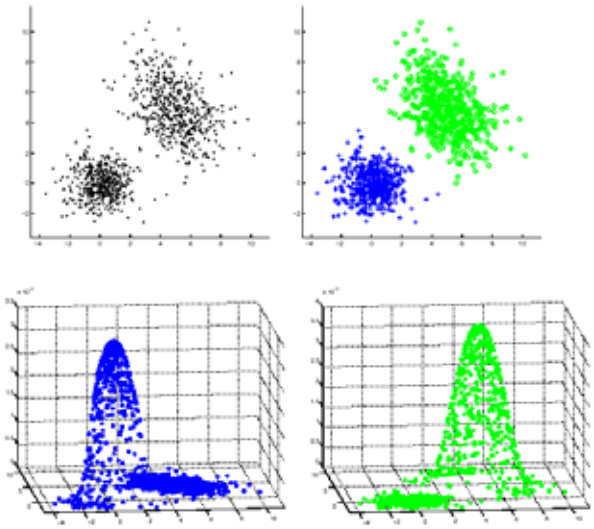


NCut

Measuring Degree of Cluster Membership

The components of the converged vector give us a measure of the participation of the corresponding vertices in the cluster, while the value of the objective function provides of the cohesiveness of the cluster.

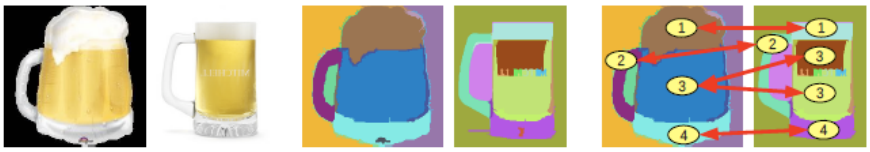
Remember Rosch's *prototype theory* of categorization….



Graph Matching and Related Problems

Formulate the (graph) matching problem as a game-theoretic clustering problem and use replicator game dynamics to solve it.

The framework can easily deal with many-to-many matching problems.



Idea: build an "association graph" where nodes correspond to correspondences, edges encode the matching (e.g., isomorphism) constraints, and edge-weights reflect similarities between correspondences.

References: ICCV 2009; IJCV 2012; CVPR 2012, etc.

Other Applications of Game-Theoretic Clustering

Security, video surveillance, analysis of social interacting behavior
- Detection of anomalous activities in video streams (Hamid et al., CVPR'05; AI'09)
- Detection of malicious activities in the internet (Pouget et al., J. Inf. Ass. Sec. 2006)
- Detecting F-formations as dominant sets (Hung and Kröse, ICMI'11)

Analysis of fMRI data
- Neumann et al (NeuroImage 2006); Muller et al (J. Mag Res Imag. 2007)

Object tracking, human action recognition
- Torsello et al. (EMMCVPR'05); Gualdi et al. (IWVS'08); Wei et al. (ICIP'07)

Multiple instance learning
- Erdem and Erdem (SIMBAD'11)

Feature selection
- Hancock et al. (GbR'11; ICIAP'11; SIMBAD'11)

Image matching and registration
- Torsello et al. (ICCV'09, IJCV 2011, CVPR'10, ECCV'10)

Bioinformatics
- Identification of protein binding sites (Zauhar and Bruist, 2005)
- Clustering gene expression profiles (Li et al, 2005)
- Tag Single Nucleotide Polymorphism (SNPs) selection (Frommlet, 2010)

In a Nutshell …

The game-theoretic approach:

- ✓ makes no assumption on the underlying (individual) data representation
- ✓ makes no assumption on the structure of the affinity matrix, being it able to work with asymmetric and even negative similarity functions
- ✓ does not require *a priori* knowledge on the number of clusters (since it extracts them sequentially)
- ✓ leaves clutter elements unassigned (useful, e.g., in figure/ground separation or one-class clustering problems)
- ✓ assigns a measure of "centrality" to the cluster's elements (prototype theory)
- ✓ allows extracting overlapping clusters *(ICPR'08)*
- ✓ generalizes naturally to hypergraph clustering problems, i.e., in the presence of high-order affinities, in which case the clustering game is played by more than two players *(IEEE T-PAMI'13)*

The (Consistent) Labeling Problem: A Game-Theoretic Perspective

A **labeling problem** involves (Hummel and Zucker, 1983):

- ✓ A set of n **objects** $B = \{b_1,\dots,b_n\}$
- ✓ A set of m **labels** $\Lambda = \{1,\dots,m\}$

The goal is to label each object of B with a label of Λ.

To this end, two sources of information are exploited:

- ✓ Local measurements which capture the salient features of each object viewed in isolation
- ✓ Contextual information, expressed in terms of a real-valued $n^2 \times m^2$ matrix of **compatibility coefficients** $R = \{r_{ij}(\lambda,\mu)\}$.

The coefficient $r_{ij}(\lambda,\mu)$ measures the strenght of compatibility between the two hypotheses: "b_i is labeled λ" and "b_j is labeled μ".

Relaxation Labeling Processes

In a now classic 1976 paper, Rosenfeld, Hummel, and Zucker introduced the following heuristic update rule: (assuming a non-negative compatibility matrix):

$$p_i^{(t+1)}(\lambda) = \frac{p_i^{(t)}(\lambda) q_i^{(t)}(\lambda)}{\sum_{\mu} p_i^{(t)}(\mu) q_i^{(t)}(\mu)}$$

where

$$q_i^{(t)}(\lambda) = \sum_j \sum_{\mu} r_{ij}(\lambda,\mu) p_i^{(t)}(\mu)$$

quantifies the support that context gives at time t to the hypothesis "b_i is labeled with label λ".

See (Pelillo, 1997) for a rigorous derivation of this rule in the context of a formal theory of consistency.

Applications

Since their introduction relaxation labeling algorithms have found applications in virtually all problems in computer vision and pattern recognition:

- ✓ Edge and curve detection and enhancement
- ✓ Region-based segmentation
- ✓ Stereo matching
- ✓ Shape and object recognition
- ✓ Grouping and perceptual organization
- ✓ Graph matching
- ✓ Handwriting interpretation
- ✓ ...

Intriguing similarities exist between relaxation labeling processes and the mechanisms of the early stages of biological visual systems (see Zucker, Dobbins and Iverson, 1989, for physiological and anatomical evidence).

The "Labeling Game"

As observed by Miller and Zucker (1991) the consistent labeling problem is equivalent to a non-cooperative game.

Indeed, in such formulation we have:

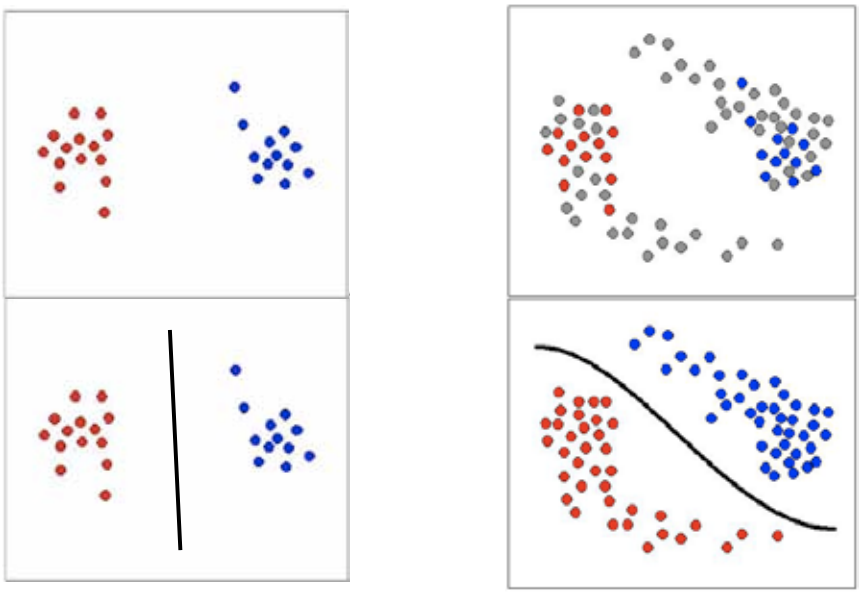
- ✓ Objects = players
- ✓ Labels = pure strategies
- ✓ Weighted labeling assignments = mixed strategies
- ✓ Compatibility coefficients = payoffs

and:

- ✓ Consistent labeling = Nash equilibrium

Further, the Rosenfeld-Hummel-Zucker update rule corresponds to discrete-time multi-population replicator dynamics.

Application to Semi-supervised Learning



Adapted from: O. Duchene, J.-Y. Audibert, R. Keriven, J. Ponce, and F. Ségonne. Segmentation by transduction. CVPR 2008.

Graph Transduction

Given a set of data points grouped into:

- ✓ labeled data:
- ✓ unlabeled data:

Represent data as a graph $G = (V,E)$

- ✓ V : nodes representing labeled and unlabeled points
- ✓ E : pairwise edges between nodes weighted by the similarity between the corresponding pairs of points

Goal: Propagate the information available at the labeled nodes to unlabeled ones in a "consistent" way.

Cluster assumption:

- ✓ The data form distinct clusters
- ✓ Two points in the same cluster are expected to be in the same class ("homophily" principle)

A Special Case

A simple case of graph transduction in which the graph G is an unweighted undirected graph:

- ✓ An edge denotes perfect similarity between points
- ✓ The adjacency matrix of G is a 0/1 matrix

The cluster assumption: Each node in a connected component of the graph should have the same class label.

The Graph Trasduction Game

Given a weighted graph $G = (V, E, w)$, the graph trasduction game is as follow:

- ✓ Nodes = players
- ✓ Labels = pure strategies
- ✓ Weighted labeling assignments = mixed strategies
- ✓ Compatibility coefficients = payoffs

The transduction game is in fact played among the unlabeled players to choose their memberships.

- ✓ Consistent labeling = Nash equilibrium

By assuming that only pairwise interactions are allowed, we obtain a game of strategy that can be solved used standard relaxation labeling / replicator algorithms (Erdem and Pelillo, 2012).

Applications: interactive image segmentation, content-based image retrieval, people tracking and re-identification, etc.

My Take-Home Message

- ✓ Today's machine learning and pattern recognition research is dominated by an essentialist attitude.
- ✓ Relational (similarity/contextual) information is relegated to a secondary role.
- ✓ Other fields made substantial progress by abandoning a purely essentialist position.
- ✓ **By analogy**… let's do the same!

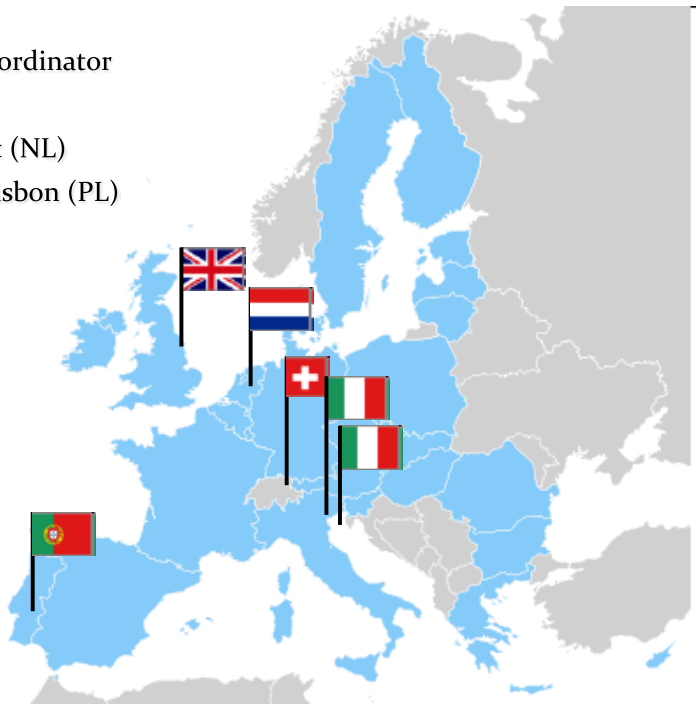
More concretely:

Starting from the observation that objects do not live in a vacuum, and that relational constraints do provide a rich source of information, to develop a holistic perspective to pattern recognition, with a view to overcome the limitations of today's approaches.

Is *game theory* the right conceptual framework?

The SIMBAD Project

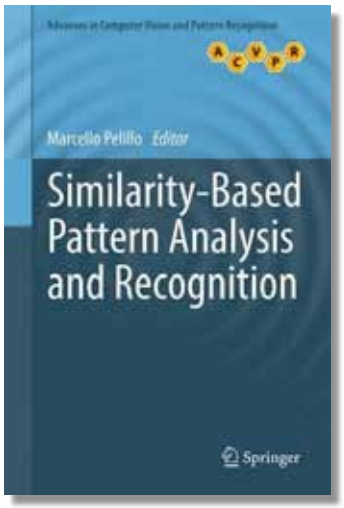
- ✓ University of Venice (IT), coordinator
- ✓ University of York (UK)
- ✓ Technische Universität Delft (NL)
- ✓ Insituto Superior Técnico, Lisbon (PL)
- ✓ University of Verona (IT)
- ✓ ETH Zürich (CH)



http://simbad-fp7.eu

Ad

M. Pelillo (Ed.)
Similarity-Based Pattern Analysis and Recognition 2013



Springer

The SIMBAD Workshop

http://www.dsi.unive.it/~simbad/2015/

Submission deadline: March 2015

References

M. Pavan and M. Pelillo. Dominant sets and pairwise clustering. *IEEE T-PAMI (2007).*

S. Rota Bulò and M. Pelillo. A game-theoretic approach to hypergraph clustering. *IEEE T-PAMI (*2013).

M. Pelillo. What is a cluster? Perspectives from game theory. *NIPS 2009 Workshop on "Clustering: Science or Art?"* (talk available on videolectures.net).

M. Pelillo et al. A game-theoretic approach to clustering and matching. In: M. Pelillo (Ed.), *Similarity-Based Pattern Analysis and Recognition*, Springer, London (in press).

A. Albarelli et al. Matching as a non-cooperative game. *ICCV 2009.*

R. A. Hummel and S. W. Zucker. On the foundations of relaxation labeling processes. *IEEE T-PAMI (*1983).

D. A. Miller and S. W. Zucker. Copositive-plus Lemke algorithm solves polymatrix games. *Operation Research Letters* (1991).

M. Pelillo. The dynamics of nonlinear relaxation labeling processes. *J. Math. Imaging and Vision* (1997).

A. Erdem and M. Pelillo. Graph transduction as a noncooperative game. *Neural Computation* (2012).