Game Theory and Natural Language
Origin, Evolution and Processing

Marcello Pelillo  Rocco Tripodi

European Centre for Living Technology
Ca’ Foscari University, Venice, Italy
{pelillo, rocco.tripodi}@unive.it

Tutorial at ACL 2016, Berlin, August 7, 2016
Web page: http://www.dsi.unive.it/~tripodi/tutorial-acl16/
Course Outline

Part I (14:00 – 15:30) – M. Pelillo

✓ Introduction to game theory
✓ Game-theoretic models of machine learning

Coffee break

Part II (16:00 – 17:30) – R. Tripodi

✓ Origin and evolution of language: A game-theoretic perspective
✓ Game theory and natural language processing
Introduction to Game Theory
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, \( P_1, \ldots, P_n \), play a given game \( \Gamma \), 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 of game theory

**1921–1928:** Emile Borel and John von Neumann give the first modern formulation of a mixed strategy along with the idea of finding minimax solutions of normal-form games.

**1944, 1947:** John von Neumann and Oskar Morgenstern publish *Theory of Games and Economic Behavior*.

**1950–1953:** In four papers John Nash made seminal contributions to both non-cooperative game theory and to bargaining theory.

**1972–1982:** John Maynard Smith applies game theory to biological problems thereby founding “evolutionary game theory.”

**late 1990’s –:** Development of algorithmic game theory…
Normal-form Games

We shall focus on finite, non-cooperative, simultaneous-move games in *normal form*, which are characterized by:

- A set of **players**: \( I = \{1, 2, \ldots, n\} \) (\( n \geq 2 \))

- A set of **pure strategy profiles**: \( S = S_1 \times S_2 \times \ldots \times S_n \) where each \( S_i = \{1, 2, \ldots, m_i\} \) is the (finite) set of pure strategies (actions) available to the player \( i \)

- A **payoff function**: \( \pi : S \rightarrow \mathbb{R}^n \), \( \pi(s) = (\pi_1(s), \ldots, \pi_n(s)) \), where \( \pi_i(s) \) (\( i=1\ldots n \)) represents the “payoff” (or utility) that player \( i \) receives when strategy profile \( s \) is played

Each player is to choose one element from his strategy space in the absence of knowledge of the choices of the other players, and “payments” will be made to them according to the function \( \pi_i(s) \).

Players’ goal is to maximize their own returns.
Two Players

In the case of two players, payoffs can be represented as two $m_1 \times m_2$ matrices (say, $A$ for player 1 and $B$ for player 2):

\[
A = (a_{hk}) \quad \quad a_{hk} = \pi_1(h,k)
\]

\[
B = (b_{hk}) \quad \quad b_{hk} = \pi_2(h,k)
\]

Special cases:

- Zero-sum games: $A + B = 0$ ($a_{hk} = -b_{hk}$ for all $h$ and $k$)
- Symmetric games: $B^T = A$
- Doubly-symmetric games: $A = A^T = B^T$
Example 1: Prisoner’s Dilemma

<table>
<thead>
<tr>
<th></th>
<th>Prisoner 2</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Confess (defect)</td>
<td>Deny (cooperate)</td>
</tr>
<tr>
<td>Prisoner 1</td>
<td>-10, -10</td>
<td>-1, -25</td>
</tr>
<tr>
<td></td>
<td>-25, -1</td>
<td>-3, -3</td>
</tr>
</tbody>
</table>
### What Would You Do?

**The Prisoner's Dilemma**

**Prisoner 1**
- Confess (defect): 
  - (Prisoner 1: -10, Prisoner 2: -10) 
  - (Prisoner 1: -25, Prisoner 2: -3)
- Deny (cooperate): 
  - (Prisoner 1: -25, Prisoner 2: -1) 
  - (Prisoner 1: -3, Prisoner 2: -3)

**Prisoner 2**
- Confess (defect): 
  - (Prisoner 1: -10, Prisoner 2: -10) 
  - (Prisoner 1: -25, Prisoner 2: -3)
- Deny (cooperate): 
  - (Prisoner 1: -25, Prisoner 2: -1) 
  - (Prisoner 1: -3, Prisoner 2: -3)

**Dominated strategy!**
Example 2:  
**Battle of the Sexes**

<table>
<thead>
<tr>
<th></th>
<th>Wife</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Soccer</td>
<td>2,1</td>
<td>0,0</td>
</tr>
<tr>
<td>Ballet</td>
<td>0,0</td>
<td>1,2</td>
</tr>
</tbody>
</table>

**Husband**

![Image of husband and wife pulling a rope](image_url)
Example 3: Rock-Scissors-Paper

<table>
<thead>
<tr>
<th></th>
<th>You</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Rock</td>
<td>Scissors</td>
<td>Paper</td>
</tr>
<tr>
<td>Me</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rock</td>
<td>0, 0</td>
<td>1, -1</td>
<td>-1, 1</td>
</tr>
<tr>
<td>Scissors</td>
<td>-1, 1</td>
<td>0, 0</td>
<td>1, -1</td>
</tr>
<tr>
<td>Paper</td>
<td>1, -1</td>
<td>-1, 1</td>
<td>0, 0</td>
</tr>
</tbody>
</table>
Mixed Strategies

A mixed strategy for player $i$ is a probability distribution over his set $S_i$ of pure strategies, which is a point in the standard simplex:

$$\Delta_i = \left\{ x_i \in R^{m_i} : \forall h = 1…m_i : x_{ih} \geq 0, \text{ and } \sum_{h=1}^{m_i} x_{ih} = 1 \right\}$$

The set of pure strategies that is assigned positive probability by mixed strategy $x_i \in \Delta_i$ is called the support of $x_i$:

$$\sigma(x_i) = \left\{ h \in S_i : x_{ih} > 0 \right\}$$

A mixed strategy profile is a vector $x = (x_1,...,x_n)$ where each component $x_i \in \Delta_i$ is a mixed strategy for player $i \in I$.

The mixed strategy space is the multi-simplex $\Theta = \Delta_1 \times \Delta_2 \times \ldots \times \Delta_n$. 
Standard Simplices

$m_i = 2$

$m_i = 3$

Note: Corners of standard simplex correspond to pure strategies.
Mixed-Strategy Payoff Functions

In the standard approach, all players’ randomizations are assumed to be independent.

Hence, the probability that a pure strategy profile \( s = (s_1, \ldots, s_n) \) will be used when a mixed-strategy profile \( x \) is played is:

\[
x(s) = \prod_{i=1}^{n} x_{is_i}
\]

and the expected value of the payoff to player \( i \) is:

\[
u_i(x) = \sum_{s \in S} x(s) \pi_i(s)
\]

In the special case of two-players games, one gets:

\[
u_1(x) = \sum_{h=1}^{m_1} \sum_{k=1}^{m_2} x_{1h} a_{hk} x_{2k} = x_1^T A x_2 \quad \quad u_2(x) = \sum_{h=1}^{m_1} \sum_{k=1}^{m_2} x_{1h} b_{hk} x_{2k} = x_1^T B x_2
\]

where \( A \) and \( B \) are the payoff matrices of players 1 and 2, respectively.
Best Replies

Notational shortcut. If $z \in \Theta$ and $x_i \in \Delta_i$, the notation $(x_i, z_{-i})$ stands for the strategy profile in which player $i \in I$ plays strategy $x_i$, while all other players play according to $z$.

Player $i$’s best reply to the strategy profile $x_{-i}$ is a mixed strategy $x_i^* \in \Delta_i$ such that

$$u_i(x_i^*, x_{-i}) \geq u_i(x_i, x_{-i})$$

for all strategies $x_i \in \Delta_i$.

Note. The best reply is not necessarily unique. Indeed, except in the extreme case in which there is a unique best reply that is a pure strategy, the number of best replies is always infinite.
Nash Equilibria

A strategy profile $x \in \Theta$ is a **Nash equilibrium** if it is a best reply to itself, namely, if:

$$u_i(x_i, x_{-i}) \geq u_i(z_i, x_{-i})$$

for all $i = 1 \ldots n$ and all strategies $z_i \in \Delta_i$.

If strict inequalities hold for all $z_i \neq x_i$ then $x$ is said to be a **strict Nash equilibrium**.

**Theorem.** A strategy profile $x \in \Theta$ is a Nash equilibrium if and only if for every player $i \in I$, every pure strategy in the support of $x_i$ is a best reply to $x_{-i}$.

It follows that every pure strategy in the support of any player’s equilibrium mixed strategy yields that player the same payoff.
## Finding Pure-strategy Nash Equilibria

<table>
<thead>
<tr>
<th>Player 1</th>
<th>Player 2</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Top</strong></td>
<td>Left</td>
</tr>
<tr>
<td></td>
<td>3, 1</td>
</tr>
<tr>
<td><strong>High</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4, 5</td>
</tr>
<tr>
<td><strong>Low</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2, 2</td>
</tr>
<tr>
<td><strong>Bottom</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5, 6</td>
</tr>
</tbody>
</table>

*Nash equilibrium!*
Multiple Equilibria in Pure Strategies

<table>
<thead>
<tr>
<th>Husband</th>
<th>Wife</th>
</tr>
</thead>
<tbody>
<tr>
<td>Soccer</td>
<td>2,1</td>
</tr>
<tr>
<td>0,0</td>
<td></td>
</tr>
<tr>
<td>Ballet</td>
<td>0,0</td>
</tr>
<tr>
<td>1,2</td>
<td></td>
</tr>
</tbody>
</table>

Nash equilibrium!
No Equilibrium in Pure Strategies

<table>
<thead>
<tr>
<th></th>
<th>You</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Rock</td>
<td>Scissors</td>
<td>Paper</td>
</tr>
<tr>
<td>Me</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rock</td>
<td>0, 0</td>
<td>1, -1</td>
<td>-1, 1</td>
</tr>
<tr>
<td>Scissors</td>
<td>-1, 1</td>
<td>0, 0</td>
<td>1, -1</td>
</tr>
<tr>
<td>Paper</td>
<td>1, -1</td>
<td>-1, 1</td>
<td>0, 0</td>
</tr>
</tbody>
</table>

Nash equilibrium!
Existence of Nash Equilibria

Theorem (Nash, 1951). Every finite normal-form game admits a mixed-strategy Nash equilibrium.

Idea of proof.

1. Define a continuous map \( T \) on \( \Theta \) such that the fixed points of \( T \) are in one-to-one correspondence with Nash equilibria.

2. Use Brouwer’s theorem to prove existence of a fixed point.

Note. For symmetric games, Nash proved that there always exists a symmetric Nash equilibrium, namely a Nash equilibrium where all players play the same (possibly mixed) strategy.
The Complexity of Finding Nash Equilibria

“Together with factoring, the complexity of finding a Nash equilibrium is in my opinion the most important concrete open question on the boundary of P today.”

Christos Papadimitriou


At present, no known reduction exists from our problem to a decision problem that is NP-complete, nor has it been shown to be easier.

**Theorem** (Daskalakis *et al.* 2005; Chen and Deng, 2005, 2006). The problem of finding a sample Nash equilibrium of a general-sum finite game with two or more players is PPAD-complete.
Theorem (Gilboa and Zemel, 1989). The following are $NP$-complete problems, even for symmetric games.

Given a two-player game in normal form, does it have:

1. at least two Nash equilibria?
2. a Nash equilibrium in which player 1 has payoff at least a given amount?
3. a Nash equilibrium in which the two players have a total payoff at least a given amount?
4. a Nash equilibrium with support of size greater than a given number?
5. a Nash equilibrium whose support contains a given strategy?
6. a Nash equilibrium whose support does not contain a given strategy?
7. etc.
“We repeat most emphatically that our theory is thoroughly static. A dynamic theory would unquestionably be more complete and therefore preferable. But there is ample evidence from other branches of science that it is futile to try to build one as long as the static side is not thoroughly understood.”

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
Evolutionary Games

Introduced by John Maynard Smith and Price (1973) to model the evolution of behavior in animal conflicts.

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 symmetric 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)

✓ Reproduction is assumed to be asexual

✓ Utility is measured in terms of Darwinian fitness, or reproductive success
A strategy is **evolutionary stable** if it is resistant to invasion by new strategies.

Formally, assume:

- A small group of “invaders” appears in a large populations of individuals, all of whom are pre-programmed to play strategy $x \in \Delta$.
- Let $y \in \Delta$ be the strategy played by the invaders.
- Let $\varepsilon$ be the share of invaders in the (post-entry) population ($0 < \varepsilon < 1$).

The payoff in a match in this bimorphic population is the same as in a match with an individual playing mixed strategy:

$$w = \varepsilon y + (1 - \varepsilon)x \in \Delta$$

and the (post-entry) payoffs got by the incumbent and the mutant strategies are $u(x,w)$ and $u(y,w)$, respectively.
Evolutionary Stable Strategies

**Definition.** A strategy $x \in \Delta$ is said to be an **evolutionary stable strategy** (ESS) if for all $y \in \Delta\setminus\{x\}$ there exists $\delta \in (0,1)$, such that for all $\varepsilon \in (0, \delta)$ we have:

$$u[x, \varepsilon y + (1 - \varepsilon)x] > u[y, \varepsilon y + (1 - \varepsilon)x]$$

**Theorem.** A strategy $x \in \Delta$ is an ESS if and only if it meets the following first- and second-order best-reply conditions:

1. $u(y,x) \leq u(x,x)$ for all $y \in \Delta$ (Nash equilibrium)

2. $u(y,x) = u(x,x) \Rightarrow u(y,y) < u(x,y)$ for all $y \in \Delta\setminus\{x\}$

**Note.** From the conditions above, we have:

- $\Delta^{ESS} \subseteq \Delta^{NE}$
- If $x \in \Delta$ is a strict Nash equilibrium, then $x$ is an ESS
Existence of ESS’s

Unlike Nash equilibria existence of ESS’s is not guaranteed.

✓ Unique Nash equilibrium $(1/3, 1/3, 1/3)^T$
✓ Hence, all $\Delta$s are best replies to $x$
✓ Let the “mutant” be $y = (1, 0, 0)^T$
✓ But $u(y, y) = u(x, y)$, hence $\Delta_{yyy} = \emptyset$

<table>
<thead>
<tr>
<th>You</th>
<th>Rock</th>
<th>Scissors</th>
<th>Paper</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rock</td>
<td>0, 0</td>
<td>1, -1</td>
<td>-1, 1</td>
</tr>
<tr>
<td>Scissors</td>
<td>-1, 1</td>
<td>0, 0</td>
<td>1, -1</td>
</tr>
<tr>
<td>Paper</td>
<td>1, -1</td>
<td>-1, 1</td>
<td>0, 0</td>
</tr>
</tbody>
</table>
Complexity Issues

Two questions of computational complexity naturally present themselves:

✓ What is the complexity of determining whether a given game has an ESS (and of finding one)?

✓ What is the complexity of recognizing whether a given $x$ is an ESS for a given game?

**Theorem (Etessami and Lochbihler, 2004).** Determining whether a given two-player symmetric game has an ESS is both NP-hard and coNP-hard.

**Theorem (Nisan, 2006).** Determining whether a (mixed) strategy $x$ is an ESS of a given two-player symmetric game is coNP-hard.
Replicator Dynamics

Let $x_i(t)$ the population share playing pure strategy $i$ at time $t$. The state of the population at time $t$ is: $x(t)= (x_1(t),\ldots,x_n(t))\in\Delta$.

Replicator dynamics (Taylor and Jonker, 1978) are motivated by Darwin’s principle of natural selection:

$$\frac{\dot{x}_i}{x_i} \propto \text{payoff of pure strategy } i - \text{average population payoff}$$

which yields:

$$\dot{x}_i = x_i\left[u(e^i, x) - u(x, x)\right] = x_i\left[(Ax)_i - x^T Ax\right]$$

Notes.

✓ Invariant under positive affine transformations of payoffs

✓ Standard simplex $\Delta$ is invariant under replicator dynamics, namely, $x(0)\in\Delta \Rightarrow x(t)\in\Delta$, for all $t > 0$ (so is its interior and boundary)
Replicator Dynamics and ESS’s

Theorem (Nachbar, 1990; Taylor and Jonker, 1978). A point $x \in \Delta$ is a Nash equilibrium if and only if $x$ is the limit point of a replicator dynamics trajectory starting from the interior of $\Delta$.

Furthermore, if $x \in \Delta$ is an ESS, then it is an asymptotically stable equilibrium point for the replicator dynamics.

The opposite need not be true.

$$A = \begin{bmatrix} 0 & 6 & -4 \\ -3 & 0 & 5 \\ -1 & 3 & 0 \end{bmatrix}$$

- The point $m = (1/3, 1/3, 1/3)^T$ is asymptotically stable (its eigenvalues have negative parts).
- But $e_1 = (1,0,0)^T$ is an ESS.
- Hence $m$ cannot be an ESS (being in the interior, it would have to be the unique ESS).
Doubly Symmetric Games

In a doubly symmetric (or partnership) game, the payoff matrix $A$ is symmetric ($A = A^T$).

**Fundamental Theorem of Natural Selection (Losert and Akin, 1983).**

For any doubly symmetric game, the average population payoff

$$f(x) = x^T A x$$

is strictly increasing along any non-constant trajectory of replicator dynamics, namely, $d/dt f(x(t)) \geq 0$ for all $t \geq 0$ (with equality if and only if $x(t)$ is a stationary point).

**Characterization of ESS’s (Hofbauer and Sigmund, 1988)**

For any doubly symmetric game with payoff matrix $A$, the following statements are equivalent:

- $a)$ $x \in \Delta^{ESS}$
- $b)$ $x \in \Delta$ is a strict local maximizer of $f(x) = x^T A x$ over the standard simplex $\Delta$
- $c)$ $x \in \Delta$ is asymptotically stable under the replicator dynamics
A well-known discretization of replicator dynamics, which assumes non-overlapping generations, is the following (assuming a non-negative $A$):

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

which inherits most of the dynamical properties of its continuous-time counterpart (e.g., the fundamental theorem of natural selection).

MATLAB implementation
References

Texts on classical and evolutionary game theory


Computationally-oriented texts


On-line resources

http://gambit.sourceforge.net/ a library of game-theoretic algorithms

http://gamut.stanford.edu/ a suite of game generators for testing game algorithms
Game-theoretic Models of Machine Learning
The Standard ML Paradigm

From: Duda, Hart and Stork, Pattern Classification (2000)
Limitations

There are cases where it’s not easy to find satisfactory feature-vector representations.

Some examples

✓ when experts cannot define features in a straightforward way
✓ when data are high dimensional
✓ 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
✓ …
Tacit Assumptions

1. Objects possess “intrinsic” (or essential) properties
2. Objects live in a vacuum

*In both cases:*  
*Relations are neglected!*
«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)

«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)
The Many Types of Relations

- Similarity relations between objects
- Similarity relations between categories
- Contextual relations

Application domains: Natural language processing, computational biology, adversarial contexts, social signal processing, medical image analysis, social network analysis, network medicine, …
Context helps ...

c → cat
    → circus

i → sin
    → fine

e → red
    → read

12
A
B
C
14

testival
graphics
... but can also deceive!
Context and the Brain

A labeling problem involves:

- A set of \( n \) objects \( B = \{b_1, \ldots, b_n\} \)
- A set of \( m \) labels \( \Lambda = \{1, \ldots, m\} \)

The goal is to label each object of \( B \) with a label of \( \Lambda \).

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 \( \lambda \)” and “\( b_j \) is labeled \( \mu \)”.
In a now classic 1976 paper, Rosenfeld, Hummel, and Zucker introduced the following 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 \(\lambda\)”. 

See (Pelillo, 1997) for a rigorous derivation of this rule in the context of a formal theory of consistency.
Since their introduction in the mid-1970’s 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
- ...

Further, intriguing similarities exist between relaxation labeling processes and certain mechanisms in the early stages of biological visual systems (see Zucker, Dobbins and Iverson, 1989)
In 1983, Hummel and Zucker developed an elegant theory of consistency in labeling problem.

By analogy with the unambiguous case, which is easily understood, they define a weighted labeling assignment $p$ consistent if:

$$\sum_{\lambda} p_i(\lambda)q_i(\lambda) \geq \sum_{\lambda} v_i(\lambda)q_i(\lambda) \quad i = 1 \ldots n$$

for all labeling assignments $v$.

If strict inequalities hold for all $v \neq p$, then $p$ is said to be strictly consistent.

**Geometrical interpretation.**

The support vector $q$ points away from all tangent vectors at $p$ (it has null projection in $\mathbb{R}$K).
Relaxation Labeling and Non-cooperative Games

As observed by Miller and Zucker (1991) the consistent labeling problem is equivalent to a polymatrix 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
- Strictly consistent labeling = strict Nash equilibrium

Further, the RHZ rule corresponds to discrete-time multi-population “replicator dynamics” used in evolutionary game theory.
Application to Semi-supervised Learning

Graph Transduction

Given a set of data points grouped into:

- labeled data: \( \{(x_1, y_1), \ldots, (x_\ell, y_\ell)\} \)
- unlabeled data: \( \{x_{\ell+1}, \ldots, x_n\} \quad \ell \ll n \)

Express 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
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 Transduction Game

Given a weighted graph $G = (V, E, w)$, the graph transduction game is as follow (Erdem and Pelillo, 2012):

- 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

Can be solved used standard relaxation labeling / replicator dynamics.

**Applications:** NLP (see part 2), interactive image segmentation, content-based image retrieval, people tracking and re-identification, etc.
References


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).
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 theory

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:

- **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

ESS’s are equivalent to dominant sets (Pavan and Pelillo, 2007).
Given a symmetric real-valued matrix $A$ (with null diagonal), consider the following Standard Quadratic Programming problem (StQP):

$$\text{maximize } f(x) = x^T Ax$$
$$\text{subject to } x \in \Delta$$

**Note.** The function $f(x)$ provides a measure of cohesiveness of a cluster (see Pavan and Pelillo, 2007; Sarkar and Boyer, 1998; Perona and Freeman, 1998).

ESS’s are in one-to-one correspondence to (strict) local solutions of StQP.
Suppose the similarity matrix is a binary (0/1) matrix.

Given an unweighted undirected graph $G=(V,E)$:

A *clique* is a subset of mutually adjacent vertices
A *maximal clique* is a clique that is not contained in a larger one

**ESS’s are in one-to-one correspondence to maximal cliques of $G$**
A Toy Example: Grouping Apples

Payoffs between apples is computed using a distance between RGB histograms.
Measuring the 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.

See Rosch’s *prototype theory* of categorization (e.g., Lakoff, 1987).
Some Applications of Game-Theoretic Clustering

**Image segmentation and grouping**

*Pavan and Pelillo* (CVPR 2003, PAMI 2007); *Torsello and Pelillo* (EMMCVPR’09)

**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)

**Security and video surveillance**

Detection of anomalous activities in video streams (*Hamid et al.*, CVPR’05; AI’09)
Detection of conversational groups (*Hung and Kröse*, 2011)

**Content-based image retrieval**

*Wang et al.* (Sig. Proc. 2008); *Zemene et al.* (2016)

**Analysis of fMRI data**

*Neumann et al.* (NeuroImage 2006); *Muller et al.* (J. Mag Res Imag. 2007)

**Video analysis, object tracking, human action recognition**

*Torsello et al.* (EMMCVPR’05); *Gualdi et al.* (IWVS’08); *Wei et al.* (ICIP’07)

**Feature selection**

*Hancock et al.* (GbR’11; ICIAP’11; SIMBAD’11)
In a Nutshell...

The game-theoretic approach to clustering:

- makes no assumption on the underlying (individual) data representation
- makes no assumption on the structure of the similarity function, 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)
- extends to hierarchical clustering (ICCV’03)
- generalizes naturally to hypergraph clustering problems (PAMI’13)

M. Pavan and M. Pelillo. Dominant sets and pairwise clustering. *PAMI* 2007


