Equilibrium Concepts for Pattern Recognition

Project acronym	SIMBAD
Project full title	Beyond Features: Similarity-Based Pattern Analysis and Recognition
Deliverable Responsible	Dipartimento di Informatica Università Ca' Foscari di Venezia Via Torino 155 30172 Venezia Mestre Italy http://www.dsi.unive.it/~pelillo
Project web site	http://simbad-fp7.eu
EC project officer	Teresa De Martino
Document title	Equilibrium Concepts for Pattern Recognition
Deliverable	D5.1
Document type	Report
Dissemination level	Public
Contractual date of delivery	M 12
Project reference number	213250
Status & version	Definitive version
Work package, deliverable responsible	WP 5, UNIVE
Author(s)	Marcello Pelillo and Andrea Torsello
Additional contributor(s)	-

Equilibrium Concepts for Pattern Recognition

1. Introduction

The objective of workpackage WP5 is to develop novel, general learning models which do not require the (geo)metric assumption, thereby working directly on the original data. Game theory offers an attractive and unexplored perspective that serves well our purpose.

In task WP5.1 we aimed at developing a game-theoretic framework based on a formalization of the competition between the hypotheses of class membership. According to this perspective, the focus shifts from optima of objective functions to equilibria of (non-cooperative) games. The lines of research in the first 6 months of the workpackage have concentrated on four lines of investigation

1. **Grouping and Matching:** extensions to a game-theoretic framework for grouping have been studied, allowing for the enumerationand extraction of overlapping clusters, and a general matching framwork which provides very robust parameter estimation.
2. **Algorithms:** new and efficient algorithms for extracting clusters have been proposed and analyzed.
3. **High order and contextual grouping:** generalizations that allow k-way interactions among players have been proposed and studied.
4. **Equilibrium concepts for clustering:** new game-theoretic concepts that generalize the notion of Nash equilibrim have been studied in relation to classification problems.

In the following we will report the main results from each line of investigation.

2. Grouping and matching

We explored and extended the grouping framework introduced in (A. Torsello, S. Rota Bulò, and M. Pelillo, 2006), which saw the introduction of a new framework for grouping and clustering derived from a game-theoretic formalization of the competition between the hypotheses of group membership. The basic idea behind the proposal is as follows: the hypotheses that each object belongs to the figure compete with one-another, each obtaining support from compatible edges and competitive pressure from all the other. Competition will reduce the population of individuals that assume hypotheses that do not receive strong support from the rest, while it will allow populations assuming hypotheses with strong support to thrive. Eventually all inconsistent hypotheses will be driven to extinction, while all the surviving hypotheses will reach an equilibrium. At equilibrium all surviving hypotheses receive the same average support, hence exhibiting the internal coherency of a group, while all the extinct hypotheses must have a lower support, hinting to external incoherency. With such formalization the group corresponds to the evolutionary stable strategies of a non-cooperative grouping game, which are found using replicator dynamics, a classic formalization of a natural selection process (J. W. Weibull, 1995).

The grouping game is defined as follows: Assume a preexisting set of objects *O* and a (possibly asymmetric) matrix of affinities *A* between the elements of *O*. Two players with complete knowledge of the setup play by simultaneously selecting an element of *O*. After both have shown their choice, each player receives a payoff, monetary or otherwise, proportional to the affinity that the chosen element has with respect to the element chosen by the opponent. Clearly, it is in each player's interest to pick an element that is strongly supported by the elements that the adversary is likely to choose.

2.1 Preliminaries

Let us give some notation and review a few fundamental concepts of Game Theory (J. W. Weibull, 1995). Let $O=\{1,\dots, n\}$ be the set of available elements (pure strategies in the language of game theory) and, $A=(a_{ij})$ be the $n\times n$ element-affinity matrix, also called *payoff* or *utility* matrix in game theory. Specifically, for each pair of strategies $i,j \in O$, a_{ij} represents the payoff of an individual playing strategy i against one playing strategy j. A mixed strategy is a probability distribution $x=(x_1,\dots,x_n)^T$ over the available strategies *O*. Clearly, mixed strategies are constrained to lie in the standard simplex of the n-dimensional Euclidean space $\mathbb{R}^n$.

$$\Delta = \left\{ \mathbf{x} \in \mathbb{R}^n \ : \ x_i \geq 0 \text{ for all } i \in O, \ \sum_{i \in O} x_i = 1 \right\} .$$

The support of a mixed strategy x, denoted by $\sigma(x)$, is defined as the set of elements chosen with non-zero probability: $\sigma(x)=\{i \in O \mid x_i > 0\}$. The expected payoff received by a player choosing element i when playing against a player adopting a mixed strategy x is $(Ax)_i=\sum_j a_{ij} x_j$, hence the expected payoff received by adopting the mixed strategy y against x is y^TAx. The best replies against mixed strategy x is the set of mixed strategies

$$\beta(\mathbf{x}) = \{\mathbf{y} \in \Delta \mid \mathbf{y}' A\mathbf{x} = \max_{\mathbf{z}}(\mathbf{z}' A\mathbf{x})\}$$

while the best pure replies against mixed strategy x, denoted with $\Omega(x)$, is the set of pure strategies that are best replies to x. It can be shown that, if y is in $\beta(x)$, then each strategy in $\sigma(y)$ is in $\Omega(x)$. A strategy x is said to be a Nash equilibrium if it is a best reply to itself, i.e., $\forall y \in \Delta,\ x^TAx \geq y^TAx$. It is easy to show that this implies that $\forall i \in \sigma(x)$ we have $(Ax)_i=x^TAx$; that is, the payoff of every strategy in the support of x is constant. Furthermore, note that, in general, we have $\sigma(x) \subseteq \Omega(x)$.

Within our setting, Nash equilibria abstracts well the main characteristics of a group: internal coherency, that is, a high mutual support of all elements within the group, and external incoherency, or low support from elements of the group to elements that do not belong to the group. In fact, any element $i \in \sigma(x)$ of a Nash equilibrium x receives from x the same expected payoff $(Ax)_i=x^TAx$, while elements not in $\Omega(x)$ receive a lower or equal support from the elements of the group.

Note, however, that external incoherency is not strict: while strategies that are not in $\sigma(x)$ cannot have higher than average payoff, they can have a payoff equal to x^TAx like elements in the group. For this reason we will impose a more stringent requirement, namely that $\Omega(x)=\sigma(x)$. This, however, is still not enough, as we also require the solution to be stable and unambiguous, that is we require the solution to be isolated and unique in $\beta(x)$. To this end, here we undertake an evolutionary game theoretic analysis of the possible strategies available to each player. Evolutionary game theory (J. W. Weibull, 1995) considers an idealized scenario whereby pairs of individuals are repeatedly drawn at random from a large population to play a symmetric two-player game. In contrast to traditional game theoretic models, players are not supposed to behave rationally or to have complete knowledge of the details of the game. They act instead according to a pre-programmed behavior pattern, or mixed strategy, and it is supposed that some selection process operates over time on the distribution of behaviors. In our grouping-game setting, each player is preprogrammed to select each element in O with a certain probability and the evolutionary selection will allow players that select elements with high average support to thrive, while driving players that choose elements with low support to extinction. In our grouping setup, we expect the selective pressure to drive to extinction the players programmed to select elements that are not of the cluster selected by the adversary, converging to a population selecting elements of a single cohesive group.

In a evolutionary setting, a strategy x is said to be an evolutionary stable strategy (ESS) if it is a Nash equilibrium and

$$\forall y \in \Delta \ x\,Ax = y\,Ax \Rightarrow x\,Ay > y\,Ay.$$

This condition guarantees that any deviation from the stable strategies does not pay providing a constraint that forces the group to be non-ambiguous. Indeed, the fact that x is ESS implies that it is an isolated Nash equilibrium, or that there exists an open set U containing x with no other other Nash equilibrium within it. Hence, evolutionary stable strategies with $\sigma(x)=\Omega(x)$ satisfy all the conditions we posed for a cluster: internal coherency, external disomogeneity, stability and non-ambiguity.

In (A. Torsello, S. Rota Bulò, and M. Pelillo, 2006) was introduced a combinatorial characterization of the evolutionary stable strategies of the grouping game in two important cases: binary affinities and general continuous affinities.

A clustering problem with binary affinities can be described as a directed graph where the presence of a directed edge from node *i* to node *j* implies a positive compatibility of node *j* with node *i*. In this context our notion of cluster becomes a straightforward extension to directed graphs of the concept of clique. Let $G(V, E)$ be a directed graph with vertex set V and edge set $E \subseteq V \times V$, a $S \subseteq V$ is a doubly-linked clique if $\forall i, j \in S$, $(i, j) \in E$, and $(j, i) \in E$. Furthermore, if there is no $j \in (V \setminus S)$ such that $\forall i \in S$, $(i, j) \in E$, the doubly linked clique is said to be *saturated*. In a situation with a binary affinity matrix A, evolutionary stable strategies turn out to be in a one-to-one relationship with saturated doubly linked cliques of a graph with A as its adjacency matrix. In the continuous case, on the other hand, the evolutionary stable strategies were shown to be a directed generalization of the concept of dominant sets (M. Pavan and M. Pelillo, 2007).

It is interesting to note the role of the asymmetry in the selection of the cluster. First, the elements that belong to a group must all be mutually compatible, hence forcing, in the binary case, a strong symmetry within the cluster. In the general case the affinities within the cluster must not be completely symmetric, but there must be a strong mutuality between each pair of elements so that compatibility must be high in both directions. The asymmetry comes into play only in inside/outside relations through the condition that a doubly-linked clique must not be fully connected to an external node to be evolutionally stable. This condition allows the asymmetry to intervene in the selection of an equilibrium by dominating strategies belonging to a clique.

2.2 Enumeration and Overlapping Groups

The first extension regarded the possibility of extracting overlapping clusters in a pairwise context, and it is based on two important properties of the game theoretical approach: First, the approach functions as a multi figure/ground discrimination algorithm, extracting only cohesive groups, while leaving spurious entries unclustered. Second, the clusters are extracted as surviving strategies at an equilibrium, thus different equilibria can provide different, possibly overlapping, groups. Transforming the payoff matrix that drives the evolution of the selection process, we can render unstable previously extracted equilibria, while not affecting any other cluster. This guarantees that once found, a cluster will not be extracted again. The net result of this process is an approach to enumerate all possible groups approximately in order of relevance. This approach was used in (S. Rota Bulò, A. Torsello, M. Pelillo, 2009) to enumerate matches for shape recognition, while in (A. Torsello, S. Rota Bulò, M. Pelillo, 2008) the approach was generalized to continuous affinities and applied to perceptual grouping and image segmentation.

First we will focus our attention on the enumeration process using symmetric binary affinities. In this setting the approach results in an enumeration of cliques in a graph.

In order to render a given ESS *x* unstable, it is sufficient to drop the Nash condition for *x*. A simple way to do it without affecting other equilibria, is to add a new strategy *z* that is a best reply to *x*, but to no other ESS. This way, *x* will be no longer asymptotically stable.

Let *G=(V,E)* be an undirected graph and *G'=(V',E')* be its directed version obtained by replacing each edge with two directed edges: one for each direction. Hence, for each *(u,v)* ∈ *E, we have* (u,v) ∈ E' *and* (v,u) ∈ E'*.* Given a set ∑ of maximal cliques of *G*, we define the ∑extension G_Σ of *G* by adding new nodes to *G'* as follows. For each clique S ∈ ∑, we create a new vertex *v*, called ∑-vertex, and put edges from *v* to each vertex in *S* and from each vertex in *V\S* to *v*. After this, each ∑-vertex *v* dominates a particular clique *S* of ∑. Further, each vertex not in *S* dominates the ∑-vertex *v* so that it cannot form a new asymptotically stable strategy. Given a set of maximal cliques ∑ of an undirected graph *G*, there exists a one-to-one correspondence between the set of maximal cliques of *G* not in ∑ and the set of ESSs of a two-player symmetric game associated to the extended graph G_Σ.

Our enumerative algorithm uses this result in the following way. We iteratively find an asymptotically stable point through the replicator dynamics. If we have an ESS, then we have found a new maximal clique. After that, we extend the graph by adding the newly extracted clique to ∑, thus rendering its associated strategy unstable, and reiterate the procedure until we have enumerated the selected number of maximal cliques. The theoretical results and the way the extension is constructed guarantee that our algorithm is correct, i.e., each ESS corresponds to a maximal clique that has not already been enumerated, and complete, i.e., each remaining maximal clique still corresponds to an ESS.

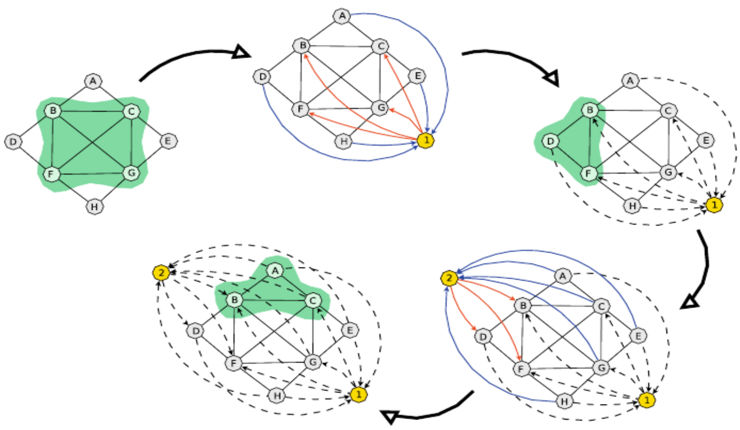

Figure 1: Example of ∑-extension and the enumeration of cliques

When applied to the problem of finding a clique of maximum cardinality the approach allows to improve on an initial possibly suboptimal solution. However, the major advantage of our enumerative method emerges from the enumeration of matches expressed as cliques in an association graph. In this context, the enumeration approach allows one to look for a set of matches that maximizes a more generic function that is only weakly correlated with the cardinality of the match and that cannot be expressed in terms of the weight of a clique.

In the continuous case we adopt a similar approach to iteratively render unstable all previously extracted ESSs by adding new strategies that are best replies to the previous ESSs, but to no other. This way the previous equilibria will no longer be asymptotically stable. Let Σ be a tuple of ESSs of a game with payoff matrix *A*. So for example if *x* and *y* are ESSs of a doubly symmetric game then Σ = (x, y) and with Σ_i we select the i-th ESS. . The Σ-extension $A_\Sigma = (a_\Sigma)$ of the payoff matrix *A* is defined as follows.

$$a_{ij}^{\Sigma} = \begin{cases} a_{ij} & \text{if } i,j \in [1,n] \\ \alpha & \text{if } j > n \text{ and } i \notin \sigma(\Sigma_{j-n}) \\ \beta & \text{if } i,j > n \text{ and } i = j \\ \frac{1}{|\Sigma_{i-n}|} \sum_{k \in \Sigma_{i-n}} a_{kj} & \text{if } i > n \text{ and } j \in \sigma(\Sigma_{i-n}) \\ 0 & \text{otherwise} \end{cases}$$

where $\alpha > \beta$ and $\beta = \max a_{ij}$. Let Φ be a two-player doubly symmetric game with payoff matrix A and Φ_Σ be a two-player game with payoff matrix A_Σ . If *x* is a mixed strategy of Φ then *x* is a mixed strategy of Φ_Σ obtained from *x* by setting the components relative to Σ-strategies to 0.

With these definitions we have the following: Let Φ be a two-player doubly symmetric game with payoff matrix *A* and let Σ be a tuple of ESSs of Φ. Furthermore let Φ_Σ be a two-player game with payoff matrix A_Σ . Then *x* is an ESS of Φ not in Σ if and only if *x* is an ESS of Φ_Σ .

We use this result to enumerate the clusters in the following way: We iteratively find new dominant sets by looking for an asymptotically stable point using the replicator dynamics. After that, we extend the graph by adding the newly extracted set to Σ, hence rendering its associated strategy unstable, and reiterate the procedure until we have enumerated all the groups and hence are unable to find new dominant sets.

Experiments on perceptual grouping problems showed that the approach is able to extract groups that overlap due to ambiguities.

2.3 Matching and Parameter Estimation

A further extension of the framework regarded the use of a game theoretic approach to matching and robust parameter estimation. In this framework matching can be formulated as a competition between correspondence hypotheses and the selection process leads to an equilibrium where only compatible correspondences survive. This matching process can then be used as inlier selection for robust parameter estimation.

This idea was first explored in (A. Albarelli, M. Pelillo, and S. Viviani 2008) for an application to symmetry estimation and then further refined in (A. Albarelli, S Rota Bulò, A. Torsello, and M. Pelillo, 2009).

The first parameter estimation approach (A. Albarelli, M. Pelillo, and S. Viviani 2008) concentrates on the extraction of symmetries from point-sets. The fundamental idea is to cast the robust estimation problem into one of the extraction of a consensus structure. Given a set of points $P=\{p_1, p_2,..., p_n\}$, their Symmetry Consensus Graph (SCG) is the undirected graph $G_B = (V_B , E_B)$ where $V_B = \{(p_a , p_b) \in P \mid a < b\}$ is the set of unordered pairs in P and where $((p_a , p_b), (p_c , p_d)) \in E_B$ if and only if the plane between p_a and p_b is exactly the same plane between p_c and p_d. Each vertex in such a graph represents a pair of points in the original object and is attributed with the plane that separates such points in a symmetric way. The key property of this graph is that two vertices are connected if and only if they share the same plane. As this is an equivalence relation the transitive property holds and it is easy to see that each set of vertices that share the same symmetry plane with each other forms a complete subgraph (clique). By building the SCG over a set of points we can thus cast the search for the most relevant symmetry planes into the search for the largest cliques over the Consensus Graph. However, the extraction of a maximum consensus graph is an NP-complete problem and the constraint of exact symmetries are very strict in practical scenarios. In fact noise, numerical errors and not perfect symmetries in the object itself make it unlikely that many pairs of points share exactly the same symmetry plane.

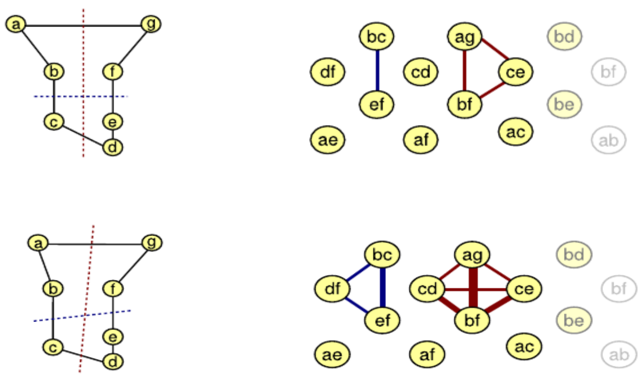

Figure 2: Symetry Consesus Graph (top) and Weighted Symmetry Consensus Graph (bottom) for a set of points

One possible solution to this problem could be the use of a threshold in order to state that two plane are equal. Unfortunately, this approach introduces two more hurdles: the first is the correct choice for such threshold, which is not different from the choice of the correct bin size, the other is that this way the transitive property would not be valid anymore, thus breaking the correspondence between cliques in the Consensus Graph and symmetries.

We extend the Symmetry Consensus Graph to a weighted graph where each edge is weighted according to the similarity between the planes associated to the vertices connected: Given a set of points $P=\{p_1, p_2,..., p_n\}$, their Weighted Symmetry Consensus Graph (WSCG) is the edge weighted undirected graph $G_W = (V_W , E_W , \omega)$ where $V_W = \{(p_a , p_b) \in P \,|a < b\}$ is the set of unordered pairs in P , $E_W = V_W \times V_W$ is the set of edges and $\omega : E_W \rightarrow \mathbb{R}_+$ is a weight function that assigns to each edge a positive real value $\omega((p_a, p_b), (p_c, p_d))$ proportional to the similarity between the symmetry plane of p_a, p_b and the symmetry plane of p_c, p_d.

We are interested in finding large sets of vertices that are all associated to very similar symmetry planes. In other words, since the edges reflect the similarity between planes we are interested in finding sets of vertices that present a high pairwise similarity among them and a low pairwise similarity with respect to vertices external to the set itself. These two properties (high internal similarity and low external similarity) are commonly used to define the notion of a cluster. For this reason we formulate the extraction of the symmetry pair forming the consensus graph as one of extracting a cluster of similar symmetry pairs.

The idea of using a clustering approach on the correspondences to attain robust parameter estimation was further generalized in (A. Albarelli, S Rota Bulò, A. Torsello, and M. Pelillo, 2009), where a generic game-theoretic matching approach was presented. The proposed approach is quite general since it can be applied to any formulation where both the objective function and the feasible set can be defined in terms of unary and pairwise interactions. The main idea is to model the set of possible correspondences as a set of game strategies. Specifically, we formulate the matching problem as a non-cooperative game where the potential associations between the items to be matched correspond to strategies, while payoffs reflect the degree of compatibility between competing hypotheses. Within this formulation, the solutions of the matching problem correspond to evolutionary stable states (ESS's), a robust population-based generalization of the notion of a Nash equilibrium. A distinguishing feature of the proposed framework is that it allows one to deal with general many-to-many matching problems even in the presence of asymmetric compatibilities.

Let O_1 and O_2 be the two sets of features that we want to match, we define the set of feasible associations $A \subseteq O_1 \times O_2$ the set of relations between O_1 and O_2 that satisfy the unary constraints. Hence, each feasible association represents a matching hypothesis. We assume that we can compute a set of pairwise compatibilities $C : A \times A \rightarrow \mathbb{R}_+$ that measure the support that one association gives to the other. Here, the self compatibilities, i.e., the compatibilities that an association gives to itself, are assumed to be zero.

In this formulation, a sub-match (or simply a match) is intuitively a set of associations, which satisfies the pairwise feasibility constraints, and two additional criteria: high internal compatibility, i.e. the associations belonging to the match are mutually highly compatible, and low external compatibility, i.e. associations outside the match are scarcely compatible with those in it. This definition of match allows us to abstract from the specific problem, since domain-specific information is confined to the definition of the compatibility function. Further, we are able to deal with many-to-many, one-to-many, many-to-one and one-to-one relations in an uniform way, as we do not impose restriction on the how the associations are selected, but incorporate the constraints into the compatibilities.

Following our game-theoretic approach, we define a matching game: Assume that we have two sets of objects O_1 and O_2 , and a compatibility function C. Two players with complete knowledge of the setup play by simultaneously selecting an association. After both have shown their choices, each player receives a payoff, monetary or otherwise, proportional to the compatibility of the selected association with respect to the association chosen by the opponent. Clearly, it is in each player's interest to pick an association, which is strongly supported by the association that the adversary is likely to choose and, assuming no prior knowledge of the inclination of the adversary, the best strategy for a player becomes the selection of associations belonging to strongly supported matches.

Within our matching setting, Nash equilibria are good candidates for a match, as they satisfy both the internal and external compatibility criteria. In particular, the internal compatibility criteria allows for a more robust match as it guarantees that we pick only association that we are confident belong to the same matching.

A main characteristic of the proposed approach is that association pairs that have zero compatibility cannot be in the same selected sub-match. This means that pairwise constraints can be enforced by forcing to zero the compatibility between associations that do not satisfy the constraints.

The approach was tested both on matching image regions obtained with an automatic segmentation algorithm, and by estimating an image transformation by matching features extracted from affinely-transformed images. The latter application is particularly interesting, as the robustness induced by the internal compatibility criteria induces an improved inlier selection mechanism that allows the approach to outperform RANSAC in parameter estimation.

3. Algorithms

We explored new efficient algorithms to extract Nash equilibria as a tool to achieve efficient classification. The replicator dynamics used in previous work, and in general all the payoff-monotonic dynamics have serious drawbacks for the use in automatic classification. First, the simplex and its faces are invariant under imitation dynamics (J. W. Weibull, 1995). This observation has two implications: fixed points under imitation dynamics may not be Nash equilibria, and every trajectory never reaches the boundary of the face from which it started in finite time. This problem forces the need of approximating the support by setting a completely arbitrary threshold to decide whether a strategy has non-zero support. Second, each iteration of the dynamics is quadratic in the number of elements to be classified, leading to computation times that are too high for large scale problems, thus requiring the definition of out-of-sample approaches.

Building upon the invasion barrier paradigm, we proposed an Infection and Immunization Dynamics (InImDyn), modelling a plausible adaptation process in a large population. This dynamics exhibits a better asymptotic behaviour compared to other popular procedures like Replicator Dynamics, and can establish support separation in finite time, which can never be achieved by any interior-point method or any other evolutionary game dynamics. (S. Rota Bulò and I. Bomze, 2009). This last property is particularly interesting as it eliminates the need for an arbitrary threshold to extract the members of a cluster.

Let $x \in \Delta$ be the incumbent population state, y be the mutant population invading x and let $z = (1 - \varepsilon)x + \varepsilon y$ be the population state obtained by injecting into x a small share of y-strategists. Then the score function of y versus x is given by

$$h_{\mathbf{x}}(\mathbf{y}, \varepsilon) = \pi(\mathbf{y} - \mathbf{x}|\mathbf{z}) = \varepsilon\pi(\mathbf{y} - \mathbf{x}) + \pi(\mathbf{y} - \mathbf{x}|\mathbf{x}) .$$

The (neutral) invasion barrier $b_x(y)$ of $x \in \Delta$ against any mutant strategy y is defined as the largest population share ε of y-strategists such that for all smaller positive population shares ε, x earns a higher or equal payoff than y in the post-entry population z. Formally

$$b_{\mathbf{x}}(\mathbf{y}) = \inf\{\varepsilon \in (0,1) : h_{\mathbf{x}}(\mathbf{y}, \varepsilon) > 0\} \cup \{1\} .$$

Now x is neutrally stable if and only if it is protected by a positive invasion barrier (I. M. Bomze and J. W. Weibull 1995): x is neutrally stable if and only if $b_x(y) > 0$ for all $y \in \Delta$.

Given populations $x, y \in \Delta$, we say that x is immune against y if $b_x(y) > 0$. Trivially, a population is always immune against itself. Note that, x is immune against y if and only if either $\pi(y - x|x) < 0$ or $\pi(y - x|x) = 0$ and $\pi(y - x) \leq 0$. If $\pi(y - x|x) > 0$ we say that y is infective for x. Hence, the set of infective strategies for x is given by

$$\Upsilon(\mathbf{x}) = \{\mathbf{y} \in \Delta : \pi(\mathbf{y} - \mathbf{x}|\mathbf{x}) > 0\} .$$

Consider $y \in \Upsilon(x)$; clearly, this implies $b_x(y) = 0$. If we allow for invasion of a share ε of y-strategists as long as the score function of y versus x is positive, at the end we will have a share of $\delta_y(x)$ mutants in the post-entry population, where

$$\delta_{\mathbf{y}}(\mathbf{x}) = \inf\{\varepsilon \in (0,1) : \; h_{\mathbf{x}}(\mathbf{y},\varepsilon) \leq 0\} \cup \{1\} \,.$$

Note that if y is infective for x, then $\delta_y(x) > 0$, whereas if x is immune against y, then $\delta_y(x) = 0$. Further note that all the above concepts can be straightforwardly extended to contests with more than two participants and/or correlated individual behavior, where the score functions may be nonlinear in ε *(I. M. Bomze and J. W. Weibull 1995)*.

Given these definitions we have the following result: Let $y \in \Upsilon(x)$ and let $z = [1 - \delta y\ (x)]x + \delta_y(x)y$,then z is immune against y.

The core idea of the proposed approach consists in selecting a strategy y which is infective for the current population x. By allowing for invasion we obtain a new population z which is immune to y. This idea suggests the following class of new dynamics which for evident reasons is called Infection and Immunization Dynamics (InImDyn):

$$\mathbf{x}^{(t+1)} = \delta_{\mathcal{S}(\mathbf{x}^{(t)})}(\mathbf{x}^{(t)})[\mathcal{S}(\mathbf{x}^{(t)}) - \mathbf{x}^{(t)}] + \mathbf{x}^{(t)} \,,$$

where $S : \Delta \rightarrow \Delta$ is a strategy selection function, which returns an infective strategy for x if it exists, or x otherwise:

$$\mathcal{S}(\mathbf{x}) = \begin{cases} \mathbf{y} & \text{for some } \mathbf{y} \in \Upsilon(\mathbf{x}) \;\; \text{if } \Upsilon(\mathbf{x}) \neq \emptyset \,, \\ \mathbf{x} & \text{otherwise} \,. \end{cases}$$

By reiterating this process of immunization we aim at reaching a population state x that cannot be infected by any other strategy. If this is the case then x is a Nash strategy.

Depending on how we choose the function $S(x)$, we may obtain different dynamics. One in particular, which is simple and leads to nice properties, consists in allowing only infective pure strategies or their respective co-strategies. This way, our equilibrium selection process closely resembles a vertex-pivoting method, as opposed to interior-point approaches like replicator dynamics. Let x be the current population and let y be a strategy. The co-strategy of y with respect to x is given by

$$\bar{\mathbf{y}}_{\mathbf{x}} = (1 - \bar{\varepsilon})\mathbf{x} + \bar{\varepsilon}\mathbf{y} \,,$$

where

$$\bar{\varepsilon} = \min\{\varepsilon \in \mathbb{R} : \; (1 - \varepsilon)\mathbf{x} + \varepsilon\mathbf{y} \in \Delta\} \leq 0 \,.$$

Equilibrium Concepts for Pattern Recognition

Consider the strategy selection function $S_{\text{Pure}}(x)$, which finds a pure strategy i maximizing $|\pi(ei - x|x)|$, and returns e^i, e^i_x, or x according to whether i is an increasing or decreasing direction. In particular, let M(x) be a (randomly or otherwise selected) pure strategy such that

$$\mathcal{M}(\mathbf{x}) \in \arg\max\{\pi(\mathbf{e}^i - \mathbf{x}|\mathbf{x}) : i \in \tau_+(\mathbf{x})\} \cup \{\pi(\mathbf{x} - \mathbf{e}^i|\mathbf{x}) : i \in \tau_-(\mathbf{x}) \cap \sigma(\mathbf{x})\}.$$

where $\tau_-(\mathbf{x}) = \{i \in S : \pi(\mathbf{x} - \mathbf{e}^i|\mathbf{x}) > 0\}$ and $\tau_+(\mathbf{x}) = \{i \in S : \pi(\mathbf{x} - \mathbf{e}^i|\mathbf{x}) < 0\}$.

Then $S_{\text{Pure}}(x)$ can be written as

$$\mathcal{S}_{\text{Pure}}(\mathbf{x}) = \begin{cases} \mathbf{e}^i & \text{if } i = \mathcal{M}(\mathbf{x}) \in \tau_+(\mathbf{x}) \\ \overline{\mathbf{e}^i}_{\mathbf{x}} & \text{if } i = \mathcal{M}(\mathbf{x}) \in \tau_-(\mathbf{x}) \cap \sigma(\mathbf{x}) \\ \mathbf{x} & \text{otherwise}. \end{cases}$$

A possible interpretation of Pure InImDyn is as follows: as time passes by, an advertisement on the basis of the aggregate behavior of the population, tells the agents that a certain pure strategy is trendy or is out-of-fashion. A strategy is trendy, if it is the best performing one in terms of payoff in the population, whereas it is out-of-fashion, if it is the worst still alive performing strategy in the population. The choice among them depends on which strategy at most deviates from the average payoff. Note that if x is the current population, k is trendy if and only if $S_{\text{Pure}}(x) = e_k$, whereas k is out-of-fashion if and only if $S_{\text{Pure}}(x) = e^k_x$.

Consider a scenario where agents can gather informations only about the announced strategy k which in turn keeps its trendy (out-of-fashion) status as long as its score function with respect to the population is positive (negative). As long as a strategy remains trendy, agents playing other strategies will switch to it when they receive the possibility of re-evaluating their strategy (changing this way the population state). On the other hand, as long as a strategy is out-of-fashion, agents playing that strategy will switch randomly to another strategy, if allowed for the strategy switch. Once a strategy looses its status, a new advertisement will be done on the basis of the current population aggregate behaviour.

Of particular interest is the behaviour for symmetric two-player games, i.e., games where the payoff function is characterized by a symmetric payoff matrix A. In this case we can show a result which generalizes the Theorem of Natural Selection proving that the InImDyn generates a growth transformation on Δ for the population payoff $\pi(x)$.

Despite the fact that this is a quadratic form in x, we cannot invoke general convergence principles for continuous growth transformations since, unlike many familiar evolutionary game dynamics, the transformation given by InImDyn may exhibit discontinuities. However, we can still show that the dynamics is globally convergent. Further, the pure-strategy variant has proven to be a very efficient approach for computing Nash equilibria, each iteration of the dynamic being computable in time linear with the number of strategies.

4. High order and contextual grouping

The game-theoretic framework naturally generalizes to allow k-way interactions among players, which is equivalent to using high-order similarity relations (hypergraphs).

To follow this direction of investigation we generalized the Motzkin-Straus theorem (T. S. Motzkin and E. G. Straus,1965) relating cliques of a graph to the optima of a quadratic problem on the standard simplex, which is strongly related to the evolutionary stable strategies of our game theoretic clustering formulation. Our generalization links cliques of k-uniform hypergraphs to the minimizers of a polynomial optimization problem on the standard simplex (S. Rota Bulò, M. Pelillo, 2009). The problem is then optimized using a dynamical system that can be seen as a high-degree (or contextual) form of the calssical replicator dynamics developed by Baum and Eagon in the late '60s (L. E. Baum and J. A. Eagon, 1967).

A k-uniform hypergraph, or simply a *k*-graph, is a pair *G = (V, E)*, where *V = {1, ..., n}* is a finite set of vertices and $E \subseteq \binom{V}{k}$ is a set of *k*-subsets of *V*, each of which is called a hyperedge. With this definition 2-graphs correspond to the classical notion of graphs. The complement of a *k*-graph *G* is given by $\bar{G} = (V, \bar{E})$ where $\bar{E} = \binom{V}{k} \setminus E$. A subset of vertices *C* ⊆ *V* is called a hyperclique if $\binom{C}{k} \subseteq E$. A clique is said to be maximal if it is not contained in any other clique, while it is called maximum if it has maximum cardinality. The clique number of a *k*-graph *G*, denoted by *ω(G)*, is defined as the cardinality of a maximum clique.

Given a *k*-graph *G* with *n* vertices, the Lagrangian of *G* is the following homogeneous multilinear polynomial in *n* variables:

$$L_G(\mathbf{x}) = \sum_{e \in E} \prod_{i \in e} x_i$$

With a view to provide a new proof of Turan's theorem, in 1965 Motzkin and Straus (1965) established a remarkable connection between the clique number of a graph *G* with *n* vertices and the maxima of its Lagrangian over the standard simplex of $\mathbb{R}^n$. Namely, they proved that if x* is a maximizer of L_G over Δ, then

$$L_G(\mathbf{x}^*) = \frac{1}{2}\left[1 - \frac{1}{\omega(G)}\right]$$

Moreover, the optimizer is in the form of a *characteristic vector* of the clique *S* ⊆ *V:*

$$\mathbf{x}_i^S = \frac{1_{i \in S}}{|S|}$$

where |S| denotes the cardinality of *S* and 1_P is an indicator function returning *1* if property *P* is satisfied and *0* otherwise.

We generalized the Motzkin-Straus Theorem to *k*-graphs. Specifically, we presented a continuous characterization of maximal cliques in *k*-graphs in terms of minimizers of a particular (parametrized) homogeneous polynomial over the standard simplex.

Given a *k*-graph *G*, consider the following non-linear program.

$$\begin{array}{ll} \text{minimize} & h_{\overline{G}}(\mathbf{x}) = L_{\overline{G}}(\mathbf{x}) + \tau \sum_{i=1}^{n} x_i^k, \\ \text{subject to} & \mathbf{x} \in \Delta \end{array} \quad (1)$$

where $\tau \in \mathbb{R}$.

We have shown that if *G* is a *k*-graph and $0 < \tau \leq \frac{1}{k(k-1)}$ (with strict inequality for k = 2), then a vector $x \in \Delta$ is a local (global) solution of the Program (1) if and only if it is the characteristic vector of a maximal (maximum) clique of *G*.

In order to grasp the intuition behind the choice of Program (1), let us investigate some elementary properties of the minimizers of the first term and the second one if considered separately. If we take any vector *x* in the simplex whose support is a clique of *G*, then trivially $L_G(x)$ attains its global minimum at 0. Vice versa, for any clique *C* of *G*, every vector in the simplex with support *C* is a global minimizer of L_G over Δ. Hence, the role of the first term is to force the minimizers to have a clique as support. As for the second term, trivially, the minimizer of $\sum_{i=1}^{n} x_i^k$ over Δ is the simplex barycenter, i.e. the characteristic vector of V . Therefore, ideally, the role of the second term is to enforce the minimizers to have a maximal support and the form of a characteristic vector. By linearly combining the two terms and by adequate choices of τ, we achieve a continuous characterization of maximal (maximum) cliques in *k*-graphs which consists in a one-to-one correspondence between the set of maximal (maximum) cliques of a *k*-graph *G*, and the local (global) solutions of (1).

This result permits to approach clique problems on k-graphs using continuous optimization techniques. We decided to locally solve Program (1), by turning it into the equivalent maximization of a homogeneous polynomial P with nonnegative coefficients over the standard simplex, where

$$P(\boldsymbol{x}) = \frac{1}{k(k-1)} \left[\left(\sum_{i=1}^{n} x_i \right)^k - \sum_{i=1}^{n} x_i^k \right] - L_{\bar{G}}(\boldsymbol{x}).$$

Equilibrium Concepts for Pattern Recognition

By means of the Baum-Eagon theorem (L. E. Baum and J. A. Eagon, 1967), we obtain a growth transformation that accomplishes the maximization task, leading to the update rule $x_j \leftarrow \alpha x_j\ \partial_j P(x)$ where ∂_j denotes partial derivative with respect to x_j and α is a normalizing constant that projects x on Δ. By unfolding the partial derivative we yield

$$x_j \leftarrow \alpha x_j \left[\frac{1 - x_j^{k-1}}{k-1} - \partial_j L_{\bar{G}}(\boldsymbol{x}) \right] .$$

This result was used in (Rota Bulò, A. Albarelli, M. Pelillo, A. Torsello, 2008) to achieve robust estimation of high-order parameters of an affine transformation between two images. Assume that we have extracted two sets of Euclidean features F_1 and F_2 from the images for which we are estimating the affine transformation. Then we can build an auxiliary structure that is a *4*-graph $H = (V, E)$ having associations as vertices, i.e., $V \subseteq F_1 \times F_2$. The edge set E consists of sets of 4 associations/vertices encoding one-to-one correspondences between features and reflecting an affine transformation within a given tolerance. In this way, a maximal/maximum clique of H becomes a maximal/maximum set of image features that were distorted following approximately, within a desired precision, the same affine transformation, and thereby the affine parameter estimation problem becomes a maximum clique problem on the so defined auxiliary *4*-graph. The method adopted to test whether an edge $e \in E$ internally reflects an affine transformation can be defined in many ways. Our solution is to calculate for every association $(x, y) \in e$ the affine transformation obtained from the remaining 3 associations and calculates the transformation error on x. If the 4 distances are all below the desired threshold ε, then the edge is kept. Note that by using this method to select edges, the user defined tolerance parameter ε is expressed in pixels, which is more intuitive than, for example, a quantization scheme for the model parameters needed by generalized Hough transforms. Once the auxiliary hypergraph H has been initialized, it suffices to find the maximum or in general a large maximal clique in it, in order to obtain a feature correspondence from which the best affine transformation is estimated in a least square sense. Clearly, for the way we constructed H, the associations in the clique will all agree on the found affine transformation within an error of ε pixels.

The approach was further generalized in (S. Rota Bulò and M. Pelillo, 2009) to clustering with continuous high-order affinities. The basic idea behind our approach is that the hypergraph clustering problem can be considered as a multi-player non-cooperative ``clustering game”.

Let $H=(V,s)$ be a k-graph modeling a hypergraph clustering problem, where $V=\{1,\dots,n\}$ is the set of objects to cluster and $s(\{i_1,\dots,i_k\})$ is the similarity function providing the similarity among k objects $i_1,\dots,i_k$. We can build a game involving k players, each of them having the same set

of (pure) strategies, namely the set of objects to cluster *V*. Under this setting, a population x ∈ Δ of agents playing a clustering game is to all intents and purposes a representation of a cluster, where x_i is the probability for object *i* to be part of it. Indeed, any cluster can be modeled as a probability distribution over the set of objects to cluster.
The payoff function of the clustering game is defined in a way as to favour the evolution of agents supporting highly coherent objects. Intuitively, this is accomplished by rewarding the *k* players in proportion to the similarity that the *k* played objects have.

Hence, assuming $(v_1, \ldots, v_k) \in V^k$ to be the tuple of objects selected by *k* players, the payoff function can be simply defined as

$$\pi(v_1, \ldots, v_k) = \begin{cases} \frac{1}{k!} s(\{v_1, \ldots, v_k\}) & \text{if } \{v_1, \ldots, v_k\} \in \binom{V}{k}, \\ 0 & \text{else}, \end{cases}$$

Within this context, the notion of a cluster turns out to be equivalent to a classical equilibrium concept from (evolutionary) game theory, namely Evolutionary Stable Strategies, as the latter reflects both the internal and external cluster conditions of a cluster, i.e., *internal coherency* condition, which asks that the objects belonging to the cluster have high mutual similarities, and an *external incoherency* condition, which states that the overall cluster internal coherency decreases by adding to it any external object.

We also show that there exists a correspondence between these equilibria and the local solutions of the following polynomial, linearly-constrained, optimization problem:

$$\text{maximize} \quad f(\mathbf{x}) = \sum_{e \in \binom{V}{k}} s(e) \prod_{i \in e} x_i \,, \quad \text{subject to} \quad \mathbf{x} \in \Delta \,.$$

Additionally, we provide an algorithm for finding them, which derives straightforwardly from the Baum-Eagon inequality (L. E. Baum and J. A. Eagon, 1967).
Finally in (A. Erdem and A. Torsello 2009) we investigated the idea to of learning using contextual-dependent similarities. In particular, our game-theoretic approach was used to learn both the categories present in the data and the specific intra-category similarities that emerged from the context.

Typically in similarity-based approaches, the similarity between two shapes is a measure of how well the primitives forming the shapes and/or their spatial organizations agree, and the assessmet of whether a shape belongs to a particular class is performed by comparing in isolation the shape to one or more prototype and by applying the nearest neighbor rule.

One problem with these approaches is that they assume the existence of a single universal measure of similarity between shapes, often requiring metric properties as well, while the human perception of similarity is not only non-metric, but also strongly dependent on the surrounding context. In particular, the observed variation within a shape-class is fundamental for determining the perception of the similarities of the shapes belonging to that class.

Recently, there has been a growing interest in trying to characterize the modes of variation of shape abstracted in terms of graphs, which in turn induces a measure of the similarity of a shape to the whole class. However, it does not help in determining the perceptual similarity to any particular shape belonging to the class, a problem that is central in query by example approaches in content based image retrieval systems.

In (A. Erdem and A. Torsello 2009) we propose a game theoretic approach to compute shape categories in an unsupervised way. There is a chicken and egg problem here: class knowledge is required to determine perceived similarities, while the similarities are needed to extract class knowledge. We solve this by iteratively clustering the shapes recomputing the similarities based on the extracted class information. Central to the approach is the ability of the framework to provide both the cluster information needed to extract the categories, and the relevance information needed to compute the category model and, thus, the similarities in a robust way. Further, the resulting contextual similarity is not symmetric, making the ability of the game-theoretic approach to deal with asymmetric affinities of fundamental importance.

5. Equilibrium concepts for clustering

An initial investigation of classic game-theoretic concepts of equilibria showed that the vast majority of the method proposed in the literature is a refinement of the Nash equilibrium, where additional constraints are added to offer stronger guarantees.

Experiments on perceptual grouping and image segmentation clearly showed that the Nash equilibrium is already overly-restrictive, often leading to over-segmentation of the data. Motivated by this observation we decided to start our investigation on relaxations rather than refinements of the Nash equilibrium. To this end we developed the concept of *maximal good* which is defined as the set of strategies whose face is completely contained in the basin of attraction of an evolutionary stable strategy for all payoff-monotonic evolution dynamics.

In the case of symmetric payoffs this coincides with the face of the simplex that satisfy the additional ESS conditions

$$\forall y \in \Delta \; x\,Ax = y\,Ax \Rightarrow x\,Ay > y\,Ay,$$

which is equivalent to finding the maximal face where the payoff matrix *A* is negative-definite on the space tangent to the simplex.

With this intuition, we developed an algorithm to compute it in the case of symmetric payoffs, based on the triangulation of the payoff matrix. Let O be the set of available strategies and S⊆O, the standard trace test (D. G. Luenberger, 1984) allows us to give a condition for the matrix A to be negative definite in the tangent space of the face Δ_S of the simplex linked with strategies in *S*. In order to check the property we need to verify that the principal minors from size *2* to size *n+1* of the matrix

$$B_S = \left(\begin{array}{c|c} 0 & \mathbf{1}^T \\ \hline \mathbf{1} & A_S \end{array}\right)$$

have alternating sign. This can be verified by iteratively triangulating the matrix B_S using the following identities:

$$B_S^{(t+1)} = L_S^{(t+1)} D_S^{(t+1)} L_S^{(t+1)^T} = \left(\begin{array}{c|c} L_S^t & 0 \\ \hline {l_S^t}^T & 1 \end{array}\right) \left(\begin{array}{c|c} D_S^t & 0 \\ \hline 0 & d_{tt} \end{array}\right) \left(\begin{array}{c|c} {L_S^t}^T & l_S^t \\ \hline 0 & 1 \end{array}\right)$$

from which we obtain the recurrence

$$L_S^1 = D_S^1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \qquad l_S^{(t+1)} = {D_S^t}^{-1} {L_S^t}^{-1} (A_S)_t \qquad d_{t,t} = -{l_S^{(t+1)}}^T D_S^t l_S^{(t+1)}$$

In this way, starting from an ESS *x* we greedily extend the good in the order given by *Ax* accepting only entries that give negative values of d_{tt}, until no further expansion is possible. The resulting set of elements form a maximal face *S* such that *A* is negative-definite in the tangent space of Δ_S.

Further, it can be shown that in the case of symmetric discrete 0-1 payoffs, i.e., the case described by the Motzkin-Straus theorem, the concept of cliques, dominant sets, and maximal good coincide.

6. References

A. Albarelli, M. Pelillo, and S. Viviani (2008). Consensus Graphs for Symmetry Plane Estimation. In joint IAPR International Workshops on Structural, Syntactical, and Statistical Pattern Recognition, pp. 197-206.

A. Albarelli, S Rota Bulò, A. Torsello, and M. Pelillo (2009). Matching as a Non-Cooperative Game. To appear in IEEE International Conference on Computer Vision, IEEE Computer.

L. E. Baum and J. A. Eagon (1967). An inequality with applications to statistical estimation for probabilistic functions of markov processes and to a model for ecology. Bull. Amer. Math. Soc., 73:360–363.

I. M. Bomze and J. W. Weibull (1995). Does neutral stability imply Lyapunov stability? Games and Econ. Behaviour 11:173–192.

A. Erdem and A. Torsello (2009). A Game Theoretic Approach To Jointly Learn Shape Categories and Contextual Similarities. Submitted.

D. G. Luenberger (1984). Linear and nonlinear programming. Addison Wesley.

T. S. Motzkin and E. G. Straus (1965). Maxima for graphs and a new proof of a theorem of Turán. Canad. J. Math. 17:533–540.

M. Pavan and M. Pelillo (207). Dominant sets and pairwise clustering. IEEE Trans. Pattern Anal. Machine Intell. 29(1):167–172.

S. Rota Bulò, A. Albarelli, M. Pelillo, A. Torsello (2008). A Hypergraph-based Approach to Affine Parameters Estimation. In 19th International Conference on Pattern Recognition.

S. Rota Bulò, I. Bomze (2009). Infection and Immunization: a new class of Evolutionary Game Dynamics. Submitted to Games and Economic Behaviour.

S. Rota Bulò, M. Pelillo (2009). A generalization of the Motzkin-Straus Theorem to Hypergraphs. Optimization Letters 3(2):287-295.

S. Rota Bulò and M. Pelillo (2009). Clustering with higher-order similarities: a game-theoretic approach. Submitted to Advances in Neural Information Processing Systems.

S. Rota Bulò, A. Torsello, M. Pelillo, (2009). Game theoretic approach to partial clique enumeration. Image and Vision Computing 27:911-922.

A. Torsello, S. Rota Bulò, M. Pelillo (2006). Grouping with Asymmetric Affinities: A Game-Theoretic Perspective. In IEEE International Conference on Computer Vision and Pattern Recognition, IEEE Computer Society, vol. 1, pp. 292-299.

A. Torsello, S. Rota Bulò, and M. Pelillo (2008). Beyond Partitions: Allowing Overlapping Groups in Pairwise Clustering. In 19th International Conference on Pattern Recognition.

J. W. Weibull (1995). Evolutionary game theory. MIT Press, Cambridge, MA.