A Game-Theoretic Approach for Similarity-Based Data Clustering

Samuel Rota Bulò

June 23, 2015
Clustering

To organize a set of objects into maximally homogeneous groups (clusters)

Ill posed problem
Clustering

To organize a set of objects into maximally homogeneous groups (clusters)

Spatial homogeneity
Clustering

To organize a set of objects into maximally homogeneous groups (clusters)

Color homogeneity
Data representation

Feature-based
- Objects are explicitly described with attributes
- e.g., $\bullet = \{x: 0, y: 0, \text{color: green}\}$

Similarity-based
- Tuples of two (or more objects) are given a similarity value
- e.g., similarity ( $\bullet$, $\bullet$ ) = ?
Clustering representation

Partitions for majority of approaches
Limitations of partitional approaches

No overlapping clusters

This solution cannot be represented
Limitations of partitional approaches

Clusters are not mutually independent entities
  - implied by the constraints of having a partition

A fixed number of clusters is required by some algorithms
  - difficult to represent outliers
A more expressive representation

Families of clusters (non-empty subsets)
Our desiderata

Similarity-based clustering
  - with no restrictions on the similarities

Well-defined notion of cluster
  - not a by-product of the clustering process

Clusters independent of each other
  - a cluster can be extracted without knowing the other clusters

Objects belong to an unrestricted number clusters
  - also none, e.g., in case of outliers
A game-theoretic approach
The clustering game
The clustering game

At each round

1. each player selects an object (simultaneously)
The clustering game

At each round
1. each player selects an object (simultaneously)
2. each player is rewarded based on the similarity of the selected objects

\[ \text{\$} = \text{similarity}(\bullet, \bullet) \]  
\[ \text{\$} = \text{similarity}(\bullet, \bullet) \]

(A. Torsello, S. Rota Bulò and M. Pelillo, IEEE CVPR 2006)
The clustering game

At each round
1. each player selects an object (simultaneously)
2. each player is rewarded based on the similarity of the selected objects
   $\text{\$} = \text{similarity}(\bigcirc, \bigcirc)$
   $\text{\$} = \text{similarity}(\bigcirc, \bigcirc)$
3. each player is penalized if both select the same object

(A. Torsello, S. Rota Bulò and M. Pelillo, IEEE CVPR 2006)
Players are prone to pick distinct objects from the same cluster

- higher reward
Intuition

Players are prone to pick **distinct** objects from the same cluster
- higher reward

The distributions of selections encode player's **cluster hypotheses**
Players are prone to pick distinct objects from the same cluster
- higher reward

The distributions of selections encode player's cluster hypotheses

Cluster discovery
1. the cluster hypotheses of the two players coincide (symmetry)
2. no player finds convenient to change hypothesis (equilibrium)
Intuition

Players are prone to pick **distinct** objects from the same cluster
- higher reward

The distributions of selections encode player's **cluster hypotheses**

**Cluster discovery**
1. the cluster hypotheses of the two players coincide (**symmetry**)
2. no player finds convenient to change hypothesis (**equilibrium**)

**Every cluster is a (symmetric) equilibrium of the game**
Formalization of the clustering game

Clustering problem

Objects: \( \mathcal{O} = \{1, \ldots, n\} \)

Similarities: \( S \in \mathbb{R}^{n \times n} \)

Clustering game

- Non-cooperative, symmetric, two-player game

Strategies: \( \mathcal{O} \)

Affinities: \( A_{ij} = \begin{cases} S_{ij} & \text{if } i \neq j \\ \alpha & \text{if } i = j \end{cases} \)

- \( \alpha \) is a penalization term
Equilibria of the clustering game

Mixed strategy $x \in \Delta$
- a distribution of strategies in $\Theta$
- $x_i$ is the probability of playing strategy $i \in \Theta$

Nash equilibrium

A mixed strategy $x$ is a Nash equilibrium if

$$y^T A x \leq x^T A x \quad \forall y \in \Delta$$

Evolutionary Stable Strategy (ESS)

A mixed strategy $x$ is an ESS if it is a Nash equilibrium satisfying the following stability condition for all $y \in \Delta \setminus \{x\}$:

$$y^T A x = x^T A x \implies x^T A y > y^T A y.$$
ESS-clusters

Given an ESS equilibrium $x$ of a clustering game, the support $\sigma(x)$ of $x$ is a cluster under our framework.

Cluster homogeneity
- From the Nash condition on $x$ it follows that
  \[(Ax)_i = x^T Ax, \quad \forall i \in \sigma(x)\]

Cluster maximality
- From the Nash condition on $x$ it follows that
  \[(Ax)_i \leq x^T Ax, \quad \forall i \notin \sigma(x)\]
- and the ESS's stability condition guarantees that no ESS-cluster can be a proper superset of $\sigma(x)$
ESS-clusters and graph theory

What are ESS-clusters of graph clustering games?

Graph clustering
- vertices $\mathcal{V}$ are objects to cluster
- edges $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ give binary similarities

$$S_{ij} = \begin{cases} 1 & \text{if } (i,j) \in \mathcal{E} \\ 0 & \text{else} \end{cases}$$

ESS-clusters are maximal cliques
- there exists a one-to-one correspondence between ESS-clusters and maximal cliques if $0 < \alpha < 1$
ESS-clusters and optimization theory

- Let $A$ be the symmetric payoff matrix of a clustering game
- Consider the following program

$$\max \{ x^T A x : x \in \Delta \} \quad (1)$$

**Interpretation of Nash equilibria**

A vector $x \in \Delta$ is a critical point of (1) if and only if $x$ is a Nash equilibrium of the clustering game.

**Interpretation of ESS equilibria**

A vector $x \in \Delta$ is a strict local maximizer of (1) if and only if $x$ is an ESS of the clustering game.

- ESS-clusters are supports of strict local solution of (1)
A combinatorial characterization

- ESS-clusters extend Dominant Sets (Pavan and Pelillo, 2007)

A non-empty subset of objects $C \subseteq \mathcal{O}$ is a (directed) dominant set if

1. $w_C(i) > 0$ for all $i \in C$ (cluster homogeneity)
2. $w_{C \cup \{i\}}(i) < 0$, for all $i \notin C$ (cluster maximality)

provided that $W(\mathcal{T}) > 0$ for any non-empty $\mathcal{T} \subseteq C$.

- $w_C(i)$ is the degree of similarity of $i$ to $C \setminus \{i\}$ relative to the overall mutual similarity of elements in $C \setminus \{i\}$

$$w_C(i) = \begin{cases} 1 & \text{if } |C| = 1 \\ \sum_{j \in C \setminus \{i\}} \left( a_{ij} - \frac{1}{|C| - 1} \sum_{h \in C \setminus \{i\}} a_{jh} \right) w_{C \setminus \{i\}}(j) & \text{else} \end{cases}$$

- $W(\mathcal{T}) = \sum_{i \in \mathcal{T}} w_\mathcal{T}(i)$ is a degree of the total similarity of $\mathcal{T}$

Every dominant set is an ESS-cluster. Every ESS-cluster $C = \sigma(x)$ is a dominant set provided that $(Ax)_i < x^T Ax$ for all $i \notin C$. 
Patterns of ESS-clusters

What is the maximum number of ESS-clusters?
- Exponentially many (M. Broom, 2000)

ESS-clusters form a Sperner family
- ESS-clusters cannot be nested

What patterns of ESS-clusters are attainable?
- Exact characterizations are still an open issue
- works from C. Cannings, G.T. Vickers and M. Broom
- a subset of attainable patterns can be recovered from maximal cliques patterns

Algorithmic issues
Finding ESS-clusters

Clustering goal

- to find all ESS-clusters of a clustering game
- clusters are extracted one at time

Issue #1: how to find a single ESS-cluster

- dynamics from evolutionary game theory

Issue #2: how to enumerate ESS-clusters

- exponential worst-case complexity
- partial enumeration techniques
Replicator dynamics

Continuous-time

\[ \dot{x}_i = x_i \left[ (Ax)_i - x^T Ax \right] \]

Discrete-time

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

Properties

- \( x \) is a Nash equilibrium \( \iff \) \( x \) is the limit of a RD trajectory starting from the interior of \( \Delta \)
- ESSs are asymptotically stable
- Convergence guarantees if \( A \) is symmetric

(Taylor and Jonker, 1978)
Infection and Immunization dynamics

Idea

- If \( x \in \Delta \) is not a Nash equilibrium then \( \exists y \in \Delta \) (infective strategy) such that \( y^T A x > x^T A x \)
- We infect \( x \) with \( y \) by obtaining \( w = (1 - \epsilon)x + \epsilon y \)
- The extent \( \epsilon \) of the infection is up to "immunization", i.e. up to

\[
\delta_y(x) = \inf(\{ \epsilon \in (0, 1) : y^T A w \leq x^T A w \} \cup \{ 1 \})
\]
**Infection and Immunization dynamics**

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

- \( S(x) \) returns an infective strategy for \( x \) if it exists; otherwise it returns \( x \)

**Fixed points**
- Any fixed point of InImDyn is a Nash equilibrium and vice versa

**With particular choices of \( S(x) \)**
- linear space/time complexity per iteration
- if \( A \) is symmetric
  - convergence guarantees
  - ESSs are asymptotically stable and vice versa
  - support separation in a finite number of steps
Enumeration of ESS-clusters

Multi-start
- Run the dynamics multiple times with random initializations

Peeling-off
- Iteratively remove extracted clusters

Tabu-list based on InImDyn
- Find an ESS of the clustering sub-game with strategies not in the tabu-list
- Use the solution to initialize a cluster extraction on the original game
- If a new ESS-cluster is found, add to clustering result, otherwise extend the tabu-list with a new vertex and re-iterate

ESS destabilization
- Iteratively modify the payoff matrix with particular asymmetric extensions that preserve only non-extracted ESS
Enumeration of ESS-clusters

Beyond pairwise similarities
High-order similarities

Pairwise similarities have a limited expressiveness
Hypergraph clustering

**Setting**
- vertices $\mathcal{V}$ are objects to cluster
- edges $\mathcal{E}$ are non-empty subsets of vertices
- we assume edges of cardinality $k$
- $s : \mathcal{E} \rightarrow \mathbb{R}$ is the similarity function
Hypergraph clustering game

At each round
1. each player selects an object (simultaneously)
2. each player receives a reward

$= \text{similarity of selected subset of objects}$

(S. Rota Bulò and M. Pelillo, NIPS 2009, IEEE TPAMI 2013)
Hypergraph clustering game

At each round
1. each player selects an object (simultaneously)
2. each player receives a reward

Clusters are ESS equilibria

(S. Rota Bulò and M. Pelillo, NIPS 2009, IEEE TPAMI 2013)
Hypergraph clustering game

Strategies in $\mathcal{V}$

Payoff function

$$\pi(v_1, \ldots, v_k) = \begin{cases} s(\{v_1, \ldots, v_k\}) & \text{if } \{v_1, \ldots, v_k\} \in \mathcal{E} \\ 0 & \text{else} \end{cases}$$

Relation to optimization

$$\max_{x \in \Delta} P(x) = \sum_{e \in \mathcal{E}} s(e) \prod_{i \in e} x_i$$

- Nash equilibria are critical points and ESSs are strict local maxima

Dynamics to find a solution

- Baum-Eagon dynamics

$$x_i^{(t+1)} \propto x_i^{(t)} \frac{\partial}{\partial x_i} P(x^{(t)})$$
Conclusions