Swarm Intelligence
Particle Swarm Optimization

Rafael Pérez-Torres
Selected Topics on Pattern Recognition

Cinvestav Tamaulipas

August 7, 2015

Table of contents

Introduction

- A major thrust in algorithmic development is the design of algorithmic models to solve increasingly complex problems.
- Enormous successes have been achieved through the modeling of biological and natural intelligence, resulting in so-called *intelligent systems*.
- Together with logic, deductive reasoning, expert systems, case-based reasoning and symbolic machine learning systems, these intelligent algorithms form part of the field of Artificial Intelligence (AI).

What is AI

The study of how to make computers do things at which people are doing better

Introduction

What is computational intelligence (CI)

The study of adaptive mechanisms to enable or facilitate intelligent behavior in complex and changing environments. These mechanisms include those AI paradigms that exhibit an ability to learn or adapt to new situations, to generalize, abstract, discover and associate

- (NN, Neural Networks)
- (EC, Evolutive Computing)
- (SI, Swarm Intelligence)
- (AIS, Artificial Immune Systems)
- (FS, Fuzzy Systems)

Each of the CI paradigms has its origins in biological systems.

Swarm intelligence

- You and a group of friends are on a treasure finding mission.
- You want that treasure, or at least some part of it.
- All who have taken part in the search will be rewarded: the person who found the treasure getting a higher reward; all others being rewarded based on distance from the treasure.
- Each one in the group has a metal detector and can communicate the strength of the signal and his current location to the nearest neighbors.
- What actions will you take?
  - Ignore your friends. If you find the treasure, it is all yours. However, if you do not find it first, you get nothing.
  - Make use of the information from your neighboring friends, and move in the direction of your closest friend with the strongest signal. You increase your chances of finding the treasure, or at least maximizing your reward.

Swarm intelligence

- Swarm intelligence (SI) refers to the problem-solving behavior that emerges from the interaction of such agents.
- Computational swarm intelligence (CSI) refers to algorithmic models of such behavior.
- More formally, swarm intelligence is the property of a system whereby the collective behaviors of unsophisticated agents interacting locally with their environment cause coherent functional global patterns to emerge. Swarm intelligence has also been referred to as *collective intelligence*.

Particle Swarm Optimization, PSO

- The particle swarm optimization (PSO) algorithm is a population-based search algorithm based on the simulation of the social behavior of birds within a flock.
- Individuals in a particle swarm follow a very simple behavior: to emulate the success of neighboring individuals and their own successes.
- The collective behavior that emerges from this simple behavior is that of discovering optimal regions of a high dimensional search space.
- In analogy with evolutionary computation paradigms, a *swarm* can be considered as a population and a *particle* an individual.

PSO

- The particles are *flown* through a multidimensional search space, where the position of each particle is adjusted according to its own experience and that of its neighbors.
- Let $\mathbf{x}_i(t)$ denote the position of particle i in the search space at time step t. The position of the particle is changed by adding a velocity, $\mathbf{v}_i(t)$, to the current position

$$\mathbf{x}_i(t+1) = \mathbf{x}_i + \mathbf{v}_i(t+1) \tag{1}$$

  with $\mathbf{x}_i(0) \sim U(\mathbf{x}_{\min}, \mathbf{x}_{\max})$.
- The velocity vector drives the optimization process, reflecting the experiential knowledge of the particle (*cognitive component*) and socially exchanged information from the particle's neighborhood (*social component*).
- Originally, two PSO algorithms have been developed which differ in the size of their neighborhoods: *gbest* (global) and *lbest* (local) PSO.

Global best PSO

- The neighborhood for each particle is the entire swarm (*star* topology).
- In this case, the social information is the best position found by the swarm, $\widehat{\mathbf{y}}(t)$.
- For *gbest* PSO, the velocity of particle i is calculated as:

$$v_{ij}(t+1) = v_{ij}(t) + c_1 r_{1j}(t)[y_{ij}(t) - x_{ij}(t)] + c_2 r_{2j}(t)[\hat{y}_j(t) - x_{ij}(t)] \tag{2}$$

- $v_{ij}(t)$ is the velocity of particle i in dimension $j = 1, \ldots, n_x$ at time step t, $x_{ij}(t)$ is the position of particle i in dimension j at time step t, c_1 and c_2 are positive acceleration constants used to scale the contribution of the cognitive and social components respectively, and $r_{1j}(t), r_{2j}(t) \sim U(0,1)$ are random values that introduce a stochastic element to the algorithm.

Global best PSO II

- The personal best position, $\mathbf{y}_i$, associated with particle i is the best position the particle has visited since the first time step.
- Its best position in time $t+1$ is calculated as:

$$\mathbf{y}_i(t+1) = \begin{cases} \mathbf{y}_i(t) & \text{if } f(\mathbf{x}_i(t+1)) \geq f(\mathbf{y}_i(t)) \\ \mathbf{x}_i(t+1) & \text{if } f(\mathbf{x}_i(t+1)) < f(\mathbf{y}_i(t)) \end{cases} \tag{3}$$

  where $f : \mathbb{R}^{n_x} \to \mathbb{R}$ is the fitness function.
- The global best position, $\widehat{\mathbf{y}}(t)$, in time step t is calculated as:

$$\widehat{\mathbf{y}}(t) \in \{\mathbf{y}_0(t), \ldots, \mathbf{y}_{n_s}(t)\} \,||\, f(\widehat{\mathbf{y}}(t)) = \min\{f(\mathbf{y}_0(t)), \ldots, f(\mathbf{y}_{n_s}(t))\}$$

  where n_s is the number of particles in the swarm.

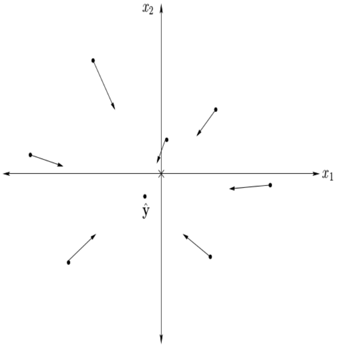
Global best PSO III

Algorithm 1 PSO gbest

Crear e inicializar un enjambre de n_x dimensiones
repeat
  for cada partícula i $= 1, \ldots, n_s$ **do**
    // Establecer la mejor posición personal
    if $f(x_i) < f(y_i)$ **then**
      $y_i = x_i$
    end if
    // Establecer la mejor posición global
    if $f(y_i) < f(\hat{y})$ **then**
      $\hat{y} = y_i$
    end if
  end for
  for cada partícula i $= 1, \ldots, n_s$ **do**
    Actualizar la velocidad utilizando 2
    Actualizar la posición utilizando la ecuación 1
  end for
until se cumpla la condición de paro

Global best PSO IV

gbest algorithm behavior



Local best PSO

- Uses a ring social network topology where smaller neighborhoods are defined for each particle.
- The social component reflects information exchanged within the neighborhood of the particle, reflecting local knowledge of the environment
- velocity is calculated as:

$$v_{ij}(t+1) = v_{ij}(t) + c_1 r_{1j}[y_{ij}(t) - x_{ij}(t)] + c_2 r_{2j}(t)[\hat{y}_{ij}(t) - x_{ij}(t)] \tag{4}$$

where $\hat{y}_{ij}$ is the best position found in the neighborhood of particle i in dimension j.

Local best PSO II

- Then $\widehat{\mathbf{y}}_i$ is the best position found in neighborhood $\mathcal{N}_i$ and is defined as:

$$\widehat{\mathbf{y}}_i(t+1) \in \{\mathcal{N}_i | f(\widehat{\mathbf{y}}_i(t+1)) = \min\{f(x)\}, \forall \mathbf{x} \in \mathcal{N}_i\}$$

- Neighborhood $\mathcal{N}_i$ is defined as:

$$\mathcal{N}_i = \left\{ y_{i-n_{\mathcal{N}_i}}(t), y_{i-n_{\mathcal{N}_i}+1}(t), \ldots, y_{i-1}(t), y_{i+1}(t), \ldots, y_{i+n_{\mathcal{N}_i}}(t) \right\}$$

- Global best PSO is a special case of local best PSO with $n_{\mathcal{N}_i} = n_s$

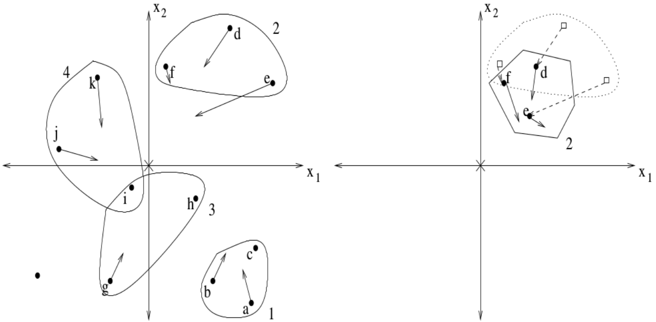
Local best PSO III

Algorithm 2 Algoritmo PSO lbest

Crear e inicializar un enjambre de n_x dimensiones
repeat
    for cada partícula i $= 1, \ldots, n_s$ **do**
        // Establecer la mejor posición personal
        if $f(x_i) < f(y_i)$ **then**
            $y_i = x_i$
        end if
        // Establecer la mejor posición global
        if $f(y_i) < f(\hat{y})$ **then**
            $\hat{y} = y_i$
        end if
    end for
    for cada partícula i $= 1, \ldots, n_s$ **do**
        Actualizar la velocidad utilizando 4
        Actualizar la posición utilizando la ecuación 1
    end for
until se cumpla la condición de paro

Local best PSO IV

lbest algorithm behavior



Stop criteria

- A maximum number of iterations.
- An acceptable solution is found: $f(x_i) \leq |f(x^*) - \epsilon|$.
- No improvement after a certain number of iterations.
- Swarm normalized radio is close to 0.
- Slope of objective function is approximately 0.

Global vs local best PSO

- Due to larger particles connectivity, *gbest* PSO converges faster, but it comes with the cost of less diversity than *lbest* PSO.
- In this regard, *lbest* PSO covers larger regions of search space becoming less sensible to local optimum. In general, neighborhood structures like star topology offer the best performance.
- Terminar cuando la pendiente de la función objetivoes aproximadamente 0.

- **Swarm size,** n_s: the more amount of particles, the more initial diversity and a larger region of search space can be covered. Also, it may improve convergence but also increase computational complexity.

- **Neighborhood size**: the smaller the neighborhood the smaller interaction and slower convergence; however it will improve quality of solutions. It is possible to start with small neighborhoods and then increase their size.

- **Iterations**: depending on scenario; a few iterations may produce premature search stop; many iterations migh cause unnecesary computations.

- **Acceleration coefficients**, c_1 y c_2: control the stochastic influence over cognitive and social components. c_1 expresses how much confidence has the particle on itself, while c_2 expresses the confidence the particle has in its neighbors.

PSO variants I

Velocity clamping

To control the global exploration of particles, velocities are clamped to stay within boundary constraints. If a particle's velocity in a dimension d exceeds a specified maximum velocity, the particle's velocity is set to the maximum velocity.
This clamping is done jus after calculatin the new speed in the PSO algorithm.

Inertia weight

Is a mechanism to control the exploration and exploitation abilities of the swarm, aimed at eliminating the need for velocity clamping. Basically, ω controlls how much memory of the previous flight direction will influence the new velocity.
In global PSO the velocity equation changes to

$$v_{ij}(t+1) = \omega v_{ij}(t) + c_1 r_{1j}(t)[y_{ij}(t) - x_{ij}(t)] + c_2 r_{2j}(t)[\hat{y}_j(t) - x_{ij}(t)] \tag{5}$$

PSO variants II

Constriction Coefficient

An approach very similar to the inertia weight to balance the exploration–exploitation trade-off, where the velocities are constricted by a constant $\mathcal{X}$, referred to as the constriction coefficient. The velocity update equation changes to:

$$v_{ij}(t+1) = \mathcal{X}[v_{ij}(t) + \phi_1(y_{ij}(t) - x_{ij}(t)) + \phi_2(\hat{y}_j(t) - x_{ij(t)})] \tag{6}$$

where

$$\mathcal{X} = \frac{2\kappa}{|2 - \phi - \sqrt{\phi(\phi - 4)}|}$$

y $\phi = \phi_1 + \phi_2$, $\phi_1 = c_1 r_1$ y $\phi_2 = c_2 r_2$. According to a formal analysis of eigenvalues of swarm dynamics, $\phi \geq 4$ and $\kappa \in [0, 1]$.

PSO variants III

Synchronous versus Asynchronous Updates

Original PSO algorithms update best positions isolated from particles position updates. It is possible to perform the update of best positions after updating each particle, improving thee performance of *lbest*.

PSO variants IV

Velocity models

Alternative velocity models differ in the components included in the velocity equation, and how best positions are determined.

- **Cognition-only model**: exluced the social component from velocity equation. This resembles nostalgia since particles may return toward their previous best positions.
- **Social only model**: Excludes the cognitive component from the velocity equation. All particles are then attracted towards the best position of their neighborhood.
- **Selfless model**: Similar to social model, but the neighborhood best solution is only chosen from a particle's neighborhood. The particle itself is not allowed to become the neighborhood best.

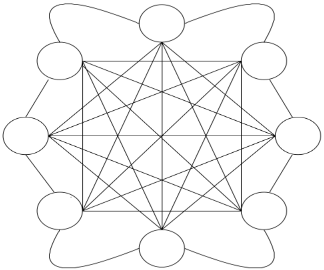
Topologies

- Social interaction guides PSO, and it is determined by the formation of overlapping neighborhoods.
- The information flow through a social network depends on:
  1. The degree of connectivity between the members.
  2. The amount of clustering (a node's neighbors are also neighbors to another).
  3. The average shortest distance from one node to another.
- Over a highly interconnected network, there is more communication and convergence is met faster.
- However, this also could lead to local optimum.

Topologies II

Star

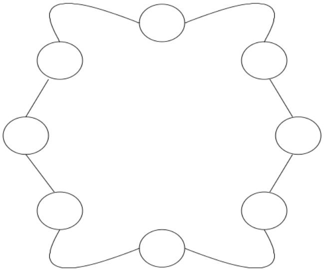
All particles are interconnected. It is employed by global PSO. Global PSO converges faster but is susceptible to local minima.



Topologies III

Ring

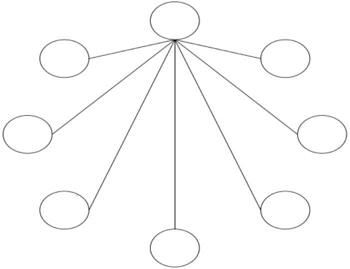
Each particle communicates with its $n_{\mathcal{N}}$ immediate neighbors. It is employed by *lbest* and takes more time to converge but covers a bigger area in search space.



Topologies IV

Wheel

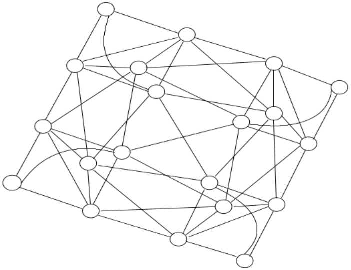
Individuals in a neighborhood are isolated from one another. One particle serves as the focal point, and all information is communicated through the focal particle. The focal particle adjusts its position towards the best neighbor. If the new position of the focal particle results in better performance, then the improvement is communicated to all the members of the neighborhood.



Topologies V

Pyramid

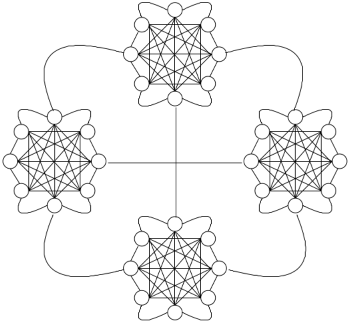
It forms a three-dimensional wire-frame



Topologies VI

Four clusters

Four clusters (or cliques) are formed with two connections between clusters. Particles within a cluster are connected with five neighbors.

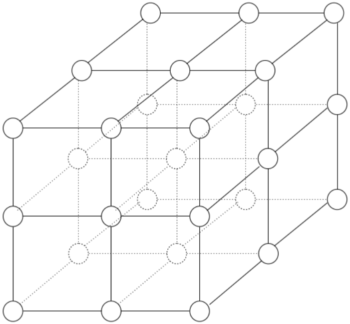


Topologies VII

Von Neumann

Particles are connected in a grid structure.

It has been shown in a number of empirical studies to outperform other social networks in a large number of problems



Conclusions

- PSO is a technique for optimization.
- It mimics the social behavior of organisms (birds).
- Particles fly toward the optimum, guided by its own knowledge (cognitive) and knoledge of the neighbors (social).
- The core component is the velocity.
- The topology of neighbors affect the communication flow.