**Tutorial on particle swarm optimization algorithm and its combinations to other evolutionary algorithms**

**Abstract**

**1. Introduction to particle swarm optimization (PSO) algorithm**

The main idea of particle swarm optimization (PSO) algorithm is based on social intelligence when it simulates how a flock of birds search for food. Given a target function known as *cost function* *f*(***x***), recall that the optimization problem is to find out the minimum point ***x***\* known as minimizer or optimizer so that *f*(***x***\*) is minimal. In PSO theory, *f*(***x***) is also called *fitness function* and thus, when *f*(***x***) is evaluated at *f*(***x***0) then, *f*(***x***0) is called fitness value. As a convention, the optimization problem is global minimization problem. For maximization, it is simple to change a little bit our viewpoint.

Traditional local optimization methods such as Newton-Raphson and gradient descent along with global optimization methods require that *f*(***x***) is differentiable. Alternately, PSO does not require existence of differential. PSO scatters a population of candidate solutions (candidate optimizers) for ***x***\* and such population is called swarm whereas each candidate optimizer is called particle in the swarm. PSO is an iterative algorithm running over many iterations in which every particle is moved at each iteration so that it approaches the global optimizer ***x***\*. Movement of all particles is attracted by ***x***\*. In other words, such movement is attracted by minimizing *f*(***x***) so that *f*(***x***) is small enough. In PSO, ***x*** is considered as position of particle. It is focused that the movement of each particle is affected by its best position and the best position of the swarm. Note, the closer to ***x***\*, the better the position is.

As a formal definition, let be the swam of particles and let ***x****i* and ***p****i* be current position and best position of particle *i*. Moreover, the movement speed of particle *i* is specified by its velocity ***v****i*. Let ***p****g* be the *global best position* of entire swarm. Note, ***p****i* is called *local best position*. The closer to ***x***\*, the better the positions ***p****i* and ***p****g* are. It is expected that ***p****g* is equals to ***x***\* or is approximated to ***x***\*. The ultimate purpose of PSO is to determine ***p****g*.

Of course, ***x****i*, ***p****i*, and ***p****g* are *n*-dimensional points and ***v****i* is *n*-dimensional vector because *f*(***x***) is from real *n*-dimensional space ***R****n* to real space ***R***.. Following is pseudo-code of PSO (Wikipedia, 2017).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Input: the swam of particles along with their initialized positions and velocities.  Output: the global best position ***p****g* of entire swarm with expectation that ***p****g* is equal or approximated to the global minimizer ***x***\*.  Let ***lb*** and ***ub*** be lower bound and upper bound of particles in their search space. They are vectors.  All current positions ***x****i* of all particles are initialized randomly. Moreover, their best positions are set to be their current positions such that ***p****i* = ***x****i*. Note, all particles are randomized in the range [***lb***, ***ub***] as closed sphere.  All current velocities ***vi*** of all particles are initialized randomly. Because each ***vi*** is vector, its elements are randomized in the range [–|***ub*** – ***lb***|, |***ub*** – ***lb***|] from –|***ub*** – ***lb***| to |***ub*** – ***lb***| where the notation |.| denotes distance between two vectors or two points.  The global best position ***p****g* is assigned by the local best position ***p****i* such that *f*(***p****i*) is smallest among particles.  While terminated condition is not met do  For each particle *i* in *S*  Velocity of particle *i* is updated as follows:   |  |  | | --- | --- | |  | (1.1) |   Position of particle *i* is updated as follows:   |  |  | | --- | --- | |  | (1.2) |   If *f*(***x****i*) < *f*(***p****i*) then  The best position of particle *i* is updated: ***p****i* = ***x****i*  If *f*(***p****i*) < *f*(***p****g*) then  The best position of swarm is updated: ***p****g* = ***p****i*  End if  End if  End for  End while |

**Table 1.1.** Basic particle swarm optimization (PSO) algorithm

Equation 1.1 is the heart of PSO, which is called *velocity update rule*. Equation 1.2 is called *position update rule*. There are two most popular terminated conditions:

1. The cost function at ***p****g* which is evaluated as *f*(***p****g*) is small enough. For example, *f*(***p****g*) is smaller than a small threshold.
2. Or PSO ran over a large enough number of iterations.

Function *U*(0, *ϕ*1) generates a random vector whose elements are random numbers in the range [0, *ϕ*1]. Similarly, function *U*(0, *ϕ*2) generates a random vector whose elements are random numbers in the range [0, *ϕ*2]. For example,

Note, the super script “*T*” indicates transposition operator of vector and matrix. The operator denotes component-wise multiplication of two points (Poli, Kennedy, & Blackwell, 2007, p. 3). For example, given random vector *U*(0, *ϕ*1) = (*r*11, *r*12,…, *r*1*n*)*T* and position ***x****i* = (*xi*1, *xi*2,…, *xin*)*T*, their component-wise multiplication is:

Two components and are considered as attraction forces that push every particle to move. Sources of force and force are the particle *i* itself and its neighbors. Thus, two most important parameters of PSO are *ϕ*1 and *ϕ*2 which represent the two attraction forces. They are in the interval (0, 1), which means that 0 < *ϕ*1 < 1 and 0 < *ϕ*2 < 1. Parameter *ϕ*1 along with the force express the exploitation of PSO whereas parameter *ϕ*2 along with the force express the exploration of PSO (Poli, Kennedy, & Blackwell, 2007, p. 4). The larger parameter *ϕ*1 is, the faster PSO converges but it trends to converge at local minimizer. In opposite, if parameter *ϕ*2 is large, convergence to local minimizer will be avoided in order to achieve better global optimizer but convergence speed is decreased. Parameters *ϕ*1 and *ϕ*2 are also called acceleration coefficients or *attraction coefficients*. Especially, *ϕ*1 is called *cognitive weight* and *ϕ*2 is called *social weight* because *ϕ*1 represents cognitive attraction force and represents *ϕ*2 social attraction force.

Because any movement has inertia, inertial force is added to the two attraction forces. Hence, the inertial force is represented by a so-called *inertial weight* *ω* where 0 < *ω* < 1. Equation 1.1 becomes (Poli, Kennedy, & Blackwell, 2007, p. 4):

|  |  |
| --- | --- |
|  | (1.3) |

The larger inertial weight *ω* is, the faster particles move because its inertial is high, which leads PSO to explore global optimizer. Note that moving fast does not imply fast convergence. In opposite, the smaller *ω* leads PSO to exploit local optimizer. In general, large *ω* expresses exploration and small *ω* expresses exploitation. The inverse 1–*ω* is known as friction coefficient.

Pioneers in PSO (Poli, Kennedy, & Blackwell, 2007, p. 5) recognized that if velocities ***v****i* of particles are not restricted, their movements can be out of convergence trajectories at unacceptable levels. Therefore, they proposed a so-called *restriction coefficient* *χ* to damping dynamics of particles. Note, *χ* is also called *restriction weight* or *damping weight*. With support of restriction coefficient, equation 1.1 becomes (Poli, Kennedy, & Blackwell, 2007, p. 5):

|  |  |
| --- | --- |
|  | (1.4) |

It is easy to recognize that equation 1.2 is special case of equation 1.3 when the expression *χ****v****i* is equivalent to the expression *ω****v****i*.

Because ***p****g* is the best position of entire swarm, the attraction force indicates the movement of each particle is affected by all other particles, which means that every particle connects to all remaining particles. In other words, neighbors of a particle are all other particles, which is known as fully connected *swarm topology* or *population topology*. For easily understandable explanation, suppose particles are vertices of a graph, fully connected swarm topology implies that such graph is fully connected, in which all vertices are connected together. Alternately, swarm topology can be defined in different way so that each particle *i* only connects with a limit number *Ki* of other particles. In other words, each particle has only *Ki* neighbors. With custom-defined swarm topology, equation 1.4 is written as follows (Poli, Kennedy, & Blackwell, 2007, p. 6):

|  |  |
| --- | --- |
|  | (1.5) |

Where ***q****k* is the best position of the *k*th neighbor of particle *i*. Of course, ***q****k* is ***p****j* of some particle *j*.

***q****k* = ***p****j* such that particle *j* is the *k*th neighbor of particle *i*.

Please pay attention that, in equation 1.5, particle *i* is also its neighbor. In other words, in equation 1.5, the set of *Ki* neighbors includes particle *i*. The two parameters *ϕ*1 and *ϕ*2 are reduced into only one parameter *ϕ* where 0 < *ϕ* < 1, which implies the strengths of all attraction forces from all neighbors on particle *i* are equal. Equation 1.5 is known as Mendes’ fully informed particle swarm (FIPS) method. The topology in the basic PSO specified by equation 1.1, equation 1.3, and equation 1.4 is known *global best topology* because only one best position ***p****g* of entire swarm is kept track. However, equation 1.5 indicates that many best positions from groups implied by neighbors are kept track. Hence, FIPS method specifies a so-called *local best topology*, which converges slowly but avoids converging at local optimizer. In other words, local best topology aims to exploration rather than exploitation.

If we focus on the fact that the attraction force issued by the particle *i* itself is equivalent to the attraction force from the global best position ***p****g* and the other attraction forces from its neighbors ***q****k*, equation 1.5 is modified as follows:

|  |  |
| --- | --- |
|  | (1.6) |

In equation 1.6, the set of *Ki* neighbors does not include particle *i* and so, the three parameters *ϕ*1, *ϕ*2, and *ϕ* are co-existent. Inertial weight *ω* is kept intact too. It is easy to recognize that equation 1.6 is the most general form of velocity update rule. Equation 1.6 balances local best topology and global best topology with expectation that convergence speed is improved but convergence to local optimizer can be avoided. In other words, equation 1.6 aims to achieve both exploration and exploitation.

The topology from equation 1.1, equation 1.3, equation 1.4, equation 1.5, and equation 1.6 is static (Poli, Kennedy, & Blackwell, 2007, p. 6) because it is kept intact over all iterations of PSO. In other words, neighbors and neighborhood in *static topology* are established fixedly. We will later research dynamic topology in which neighbors and neighborhood are changed at each iteration.

**2. Variants of PSO**

The PSO shown in table 1.1 is basic PSO. Recently there are many PSO variants which aim to improve the basic PSO. Some of them are mentioned in this section.

Binary PSO (BPSO) is a simple version of PSO where positions ***x****i* and ***p****i* are binary (0 and 1). After velocity update rule (equation 1.1) is executed, the velocity ***v****i* = (*vi*1, *vi*2,…, *vin*)*T* is in turn squashed into range [0, 1] by squash function (logistic function) as follows (Poli, Kennedy, & Blackwell, 2007, p. 9):

|  |  |
| --- | --- |
|  | (2.1) |

Where *vij* is the *j*th element of ***v****i* = (*vi*1, *vi*2,…, *vin*)*T*. Of course, the squashed value *s*(*vij*) is in range [0, 1]. The main point of BPSO is to modified position update rule as follows:

|  |  |
| --- | --- |
|  | (2.2) |

Where *xij* is the *j*th element *xij* of ***x****i* = (*xi*1, *xi*2,…, *xin*)*T* and *r* is a random number in range [0, 1]. In general, position update rule in BPSO is specified by equation 2.2 instead 1.2.

Bare bones PSO (BBPSO) is also a simple version of PSO where velocity update rule (equation 1.1) is eliminated. In other words, positions ***x****i* are updated based on only previous position and previous best position. Given ***x****i* = (*xi*1, *xi*2,…, *xin*)*T*, ***p****i* = (*pi*1, *pi*2,…, *pin*)*T*, and ***p****g* = (*pg*1, *pg*2,…, *pgn*)*T*, BBPSO assumes that the *j*th element *xij* of ***x****i* follows normal distribution with mean (*pij*–*pgj*)/2 and variance (*pij*–*pgj*)2 (Poli, Kennedy, & Blackwell, 2007, p. 13).

|  |  |
| --- | --- |
|  | (2.3) |

Note, the sign “~” denotes distribution and denotes normal distribution. Thus, position update rule in BBPSO is modified as follows:

|  |  |
| --- | --- |
| Every *xij* is randomized according to normal distribution | (2.4) |

Obviously, position update rule in BBPSO is specified by equation 2.4 instead 1.2 and there is no velocity update rule.

Recall that the topology from equation 1.1, equation 1.3, equation 1.4, equation 1.5, and equation 1.6 is static topology because it is kept intact over all iterations of PSO. Here we research *dynamic topology* in which neighbors and neighborhood are changed at each iteration.

**3. PSO and other evolutionary algorithms**

Evolutionary algorithms (EA) are the ones which simulate natural activities in biological world. PSO is an EA because it simulates how a flock of birds search for food and so its combination to other EA is natural.

**4. Conclusions**

**References**