

TP2 - Análise de hiper-parâmetros no PSO

Guilherme Castro Silva

Abstract—Neste trabalho é apresentado como o método do enxame de partículas e suas variações podem ser utilizados aplicados na minimização de funções unimodais e multi-modais como a Rastrigin. Ademais, é mostrado as estratégias adotadas para ambas as configurações Gbest e Lbest. Por fim há uma seção de testes com variações dos hiper-parâmetros que compõe o PSO.

I. INTRODUÇÃO

O método do enxame de partículas (*particle swarm optimization* ou PSO) é um ramo da inteligência artificial também classificado por alguns autores como um ramo da computação evolucionária, que otimiza um problema iterativamente ao tentar melhorar a solução candidata com respeito a uma dada medida de qualidade. O método do enxame de partícula foi proposto por Kennedy e Eberhart[1] em 1995.

O PSO compartilha diversos conceitos com a Programação Evolucionária em geral, como os Algoritmos Genéticos (AG). Ambos, AG e PSO, possuem uma população inicial de indivíduos (como possíveis soluções) e uma função objetivo (que mede o quanto cada indivíduo é apto a solucionar o problema). No entanto, ao contrário do AG, no PSO não temos os conceitos explícitos de cruzamento, mutação nem seleção. Uma diferença conceitual importante entre PSO e AE é que o princípio nestes é a sobrevivência dos melhores indivíduos, os quais competem para sobreviver e gerar um novo descendente. No PSO, o princípio consiste na interação social e a troca de conhecimento sobre o espaço de busca.

Como operam sobre uma população de candidatos em paralelo, a busca é realizada em diferentes áreas do espaço de solução ao mesmo tempo, possibilitando uma busca mais diversa sobre o espaço. Esse classe de algoritmos é robusta, genérica e facilmente adaptável, além disso, tem como vantagem a capacidade de encontrar boas soluções para problemas que não podem ser resolvidos em tempo polinomial.

O objetivo deste trabalho foi testar a abordagem do algoritmo PSO que contém (momento de inércia e fator de constrição) combinadas a dois tipos de topologia de vizinhança (gbest e lbest) para as funções *Rastrigin* e *Sphere* além de uma análise sobre seus hiper-parâmetros.

II. PSO

O método de *Particle Swarming Optimization*, encontra-se, atualmente, entre as meta-heurísticas de algoritmos de otimização, baseadas em padrões da natureza (como a representação do movimento de cada indivíduo dentro de um bando de pássaros ou de um cardume de peixes), mais populares nesta área e surge como sendo um dos algoritmos

mais promissores para a resolução dos mais diversificados problemas de otimização.

PSO é inspirado pelo comportamento social e cooperativo exibido por várias espécies por forma a realizar as suas necessidades no espaço de pesquisa ("Search-space"). Além disso, como se trata de uma meta-heurística, isto é, realiza poucas ou nenhuma premissa sobre o problema que está a ser otimizado e pode procurar soluções candidatas em espaços de grandes dimensões. No entanto, o PSO não garante que uma solução ideal será encontrada. Em termos gerais, o algoritmo guia-se por experiência pessoal (*Pbest*), experiência geral (*Gbest*) e o movimento das partículas atual para decidir as posições seguintes no espaço de pesquisa.

O seu funcionamento é dado pela evolução de um enxame de partículas (análogo a uma população de indivíduos), inicialmente aleatórias, e movendo estas partículas em torno do espaço de pesquisa. A cada iteração, cada partícula é atualizada, seguindo dois melhores valores: o melhor fitness da população e o melhor fitness encontrado pela partícula (considerando suas gerações passadas), com o intuito de mover o enxame em direção da melhor solução. Portanto, cada partícula possui uma posição x (ou um vetor de posições) e uma velocidade v (ou um vetor de velocidades). A posição é a representação da solução do problema, e a velocidade é o peso do direcionamento para o deslocamento das partículas.

Um equacionamento comumente utilizado na literatura para a atualização das velocidades é dado pela equação a seguir.

$$v_i(t+1) = X[wv_i(t) + c_1r_1 * (p_i(t) - x_i(t)) + c_2r_2 * (g_i(t) - x_i(t))]$$

Neste modelo:

v_i é a velocidade da partícula i ;

c_1 e c_2 são os coeficientes de aceleração (cognitivos e social);

Os valores r_1 e r_2 são números aleatórios uniformemente distribuídos no intervalo $[0, 1]$;

x_i é a posição atual da partícula i ;

p_i é a melhor posição que a partícula i já obteve durante a busca;

g_i é a melhor posição encontrada na vizinhança da partícula i ;

X é o parâmetro de constrição, consiste em um fator de amortecimento baseado nos parâmetros cognitivo e social para limitar a velocidade da partícula;

A inércia da partícula, representada por W , introduz a preferência da partícula em continuar se movendo na mesma direção que seguia na interação anterior. Um valor de inércia

alto facilita uma exploração mais global, enquanto um valor baixo facilita uma exploração mais local na busca pelo ótimo. A Figura 1 apresenta o fluxograma de um PSO padrão.

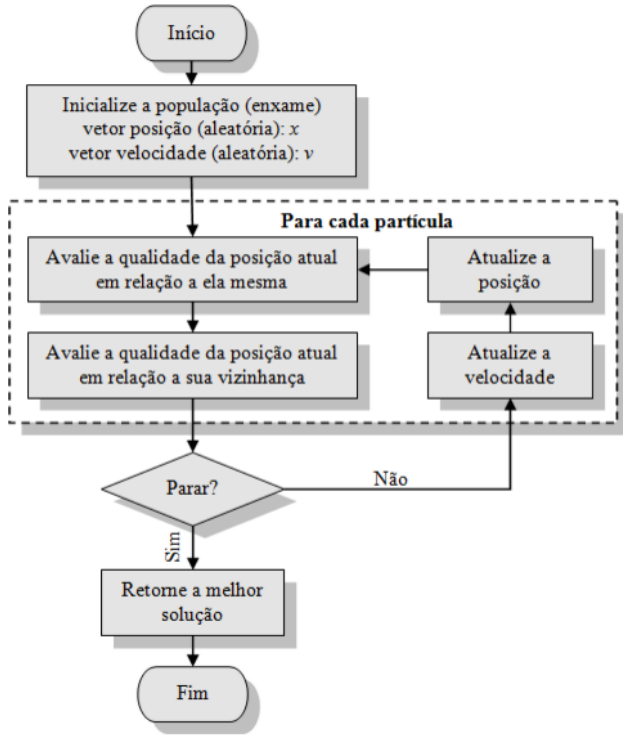


Fig. 1. Fluxograma de um PSO padrão.

Para resolver um dado problema de otimização usando o algoritmo PSO deve-se, primeiramente, definir:

- o tamanho da população;
- os valores de X e W ;
- os valores dos coeficientes de aceleração;
- uma estrutura de vizinhança entre os indivíduos;
- pelo menos uma função de teste para avaliar a qualidade de uma solução candidata;
- uma condição de término, por exemplo, número de iterações ou um valor de fitness.

A. Tamanho da População

O número de partículas para a população foi definido como 50, pois há um trade-off entre quantidade de partículas e número de avaliações de funções.

B. Valores de X e W

O PSO será executado para as seguintes quatro configurações para os parâmetros X e W :

- $X = 1$, $W = 1$;
- $X = 1$, $0 < W < 1$;
- $0 < X < 1$, $W = 1$;
- $0 < X < 1$, $0 < W < 1$.

C. Valores de $C1$ e $C2$

Para ambos $C1$ e $C2$ o valor adotado será 2.05 pois os valores $r1$ e $r2$ são números distribuídos no intervalo $[0, 1]$ que fará o equilíbrio entre a parte cognitiva e social.

D. Estrutura de Vizinhança

Definiu-se para cada abordagem do algoritmo PSO, gbest e lbest, duas topologias de vizinhança: estrela e anel (Figura 2).

- Estrela: Todas as partículas estão conectadas.
- Anel: Cada partícula comunica com seus N vizinhos mais próximos. O caso mais simples e comum é quando o número de vizinhos é igual a 2.

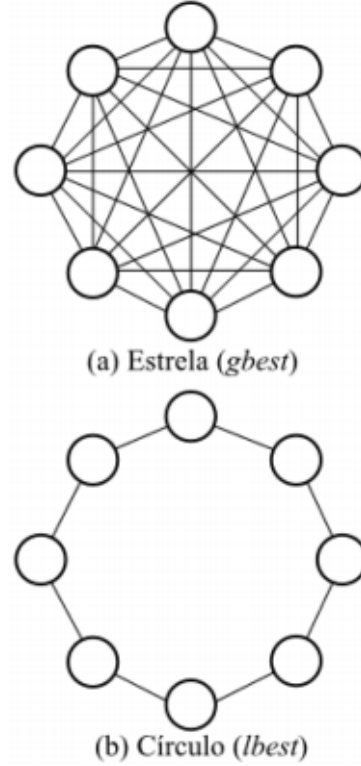


Fig. 2. Topologias de vizinhança utilizadas.

E. Função Objetivo

As funções escolhidas são a *Rastrigin* e *Sphere*. Dentre as seções seguintes haverá um detalhamento sobre ambas as funções.

F. Condição de Parada

100 mil avaliações de função (orçamento computacional).

III. SPHERE

A função *Sphere* possui um mínimo global. É contínua, convexa e unimodal. Na Figura 3 mostra sua forma bidimensional. É descrita pela seguinte equação:

$$\min f(x) = \sum_{i=1}^n x_i^2$$

usualmente com $x_i \in [-5.12, 5.12]$, para todo $i=1, \dots, n$. O mínimo global ocorre em $x^* = 0$, onde $f(x^*) = 0$.

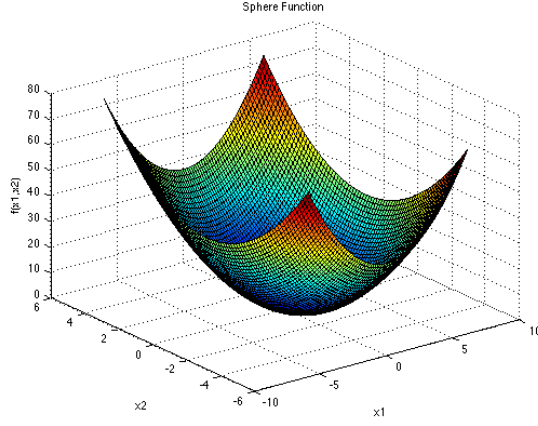


Fig. 3. Função Sphere.

IV. RASTRIGIN

A função Rastrigin é uma função não convexa utilizada para avaliar o desempenho de algoritmos de otimização. Ela foi proposta inicialmente por Rastrigin [2] como uma função bidimensional e posteriormente generalizada por Muhlenbein et al. [3]. Esta função tem sua importância devido ao grande número de mínimos que apresenta. Na Figura 4 mostra sua forma bidimensional. É descrita pela seguinte equação:

$$\min f(x) = 10n + \sum_{i=1}^n (x_i^2 - 10\cos(2\pi x_i))$$

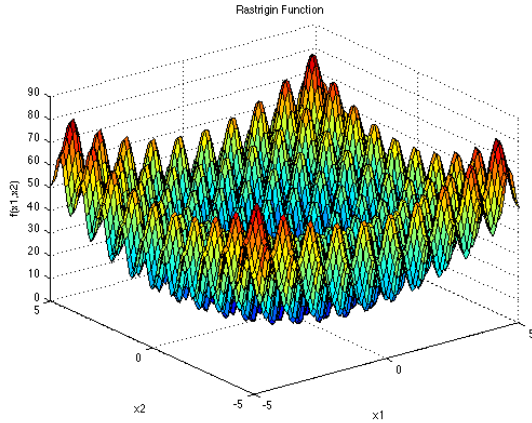


Fig. 4. Função Rastrigin.

com $x_i \in [-5.12, 5.12]$, onde n é a dimensão do espaço de busca onde a função é definida (adotaremos n igual a 10). O mínimo global ocorre em $x^* = 0$, onde $f(x^*) = 0$.

Com a utilização de um método determinístico, seria praticamente impossível otimizar a função Rastrigin. Isto acontece devido à natureza multimodal da função.

V. TESTES

O algoritmo foi testado utilizando duas configurações, sendo elas Lbest e Gbest. Para ambas, considerou-se as

configurações descritas na seção "Valores de X e W " para realizar os testes.

Número de partículas por geração: 50.

C1 e C2 : 2.05.

Estruturas de Vizinhança: Anel (Lbest) e Estrela (Gbest).

A. Sphere

Segue a baixo a Tabela I e II que mostra o comparativo entre o valor médio de função objetivo encontrado pelo algoritmo para cada configuração ao variar os valores de X e W .

TABLE I
CONFIGURAÇÕES - SPHERE - GBEST

Execuções	Configurações (X, W)			
	(1, 1)	(1, 0.5)	(0.5, 1)	(0.5, 0.5)
1	0.05	0.0	7.75e-264	0.0
2	0.16	0.0	6.38e-263	0.0
3	0.16	0.0	5.09e-266	0.0
4	0.06	0.0	4.97e-270	0.0
5	0.8	0.0	8.87e-267	0.0
6	0.98	0.0	5.19e-263	3.31e-132
7	0.51	0.0	7.819e-266	0.0
8	0.08	0.0	6.94e-268	5.21e-132
9	0.88	0.0	6.95e-269	0.0
10	0.04	0.0	6.58e-269	0.0
11	0	0.0	3.37e-265	0.0
12	0.01	0.0	4.07e-265	4.38e-168
13	1.29	0.0	2.73e-262	0.0
14	0.45	0.0	3.18e-267	0.0
15	0.23	0.0	1.53e-264	0.0
16	0.1	0.0	2.39e-269	8.46e-97
17	0.2	0.0	1.16e-268	0.0
18	0.49	0.0	5.37e-271	0.0
19	0.68	0.0	2.01e-265	2.63e-161
20	0.45	0.0	2.21e-268	0.0
21	0.38	0.0	1.61e-269	0.0
22	1.54	0.0	3.66e-264	0.0
23	0.24	0.0	1.39e-272	4.017e-304
24	1.72	0.0	3.62e-273	0.0
25	0.54	0.0	5.51e-265	0.0
26	0.27	0.0	7.01e-268	0.0
27	0.46	0.0	1.78e-269	0.0
28	1.08	0.0	1.13e-270	1.03e-306
29	0.17	0.0	1.09e-265	0.0
30	0.12	0.0	6.78e-269	1.60e-39
31	0.04	0.0	6.10e-267	0.0
Média	0.45	0.0	1.30e-263	5.16e-41
Std	0.45	0.0	0.0	2.82e-40

TABLE II
CONFIGURAÇÕES - SPHERE - LBEST

Execuções	Configurações (X, W)			
	(1, 1)	(1, 0.5)	(0.5, 1)	(0.5, 0.5)
1	0.25	1.57e-221	5.71e-132	0.53
2	0.56	4.37e-222	2.83e-136	0.64
3	0.81	2.58e-208	1.64e-134	0.12
4	1.43	3.95e-201	2.49e-134	0.74
5	0.36	4.14e-198	6.77e-133	0.00
6	0.2	1.86e-209	4.28e-126	0.07
7	0.55	1.67e-204	2.27e-136	0.00
8	0.36	1.84e-217	8.28e-129	0.50
9	0.49	3.07e-192	4.38e-132	1.25
10	0.45	5.49e-214	3.36e-131	0.40
11	1.04	2.06e-212	5.63e-139	0.04
12	0.66	6.11e-219	1.55e-132	1.08
13	0.99	1.04e-204	4.63e-139	0.00
14	0.3	7.58e-194	1.86e-140	1.63
15	1.76	5.47e-200	4.56e-131	2.28
16	0.02	6.37e-206	8.99e-140	0.09
17	0.54	3.28e-214	1.75e-135	0.06
18	0.35	3.89e-210	3.43e-135	10.02
19	0.98	1.51e-195	1.39e-134	0.25
20	2.02	6.93e-229	1.08e-133	4.96e-06
21	1.62	1.52e-219	6.82e-137	0.01
22	1.63	6.89e-208	2.45e-136	9.33
23	0.97	6.09e-212	8.66e-136	0.06
24	0.4	1.45e-198	3.17e-134	0.52
25	0.58	4.91e-211	1.40e-135	0.59
26	0.36	4.65e-199	5.78e-130	8.41e-06
27	0.38	1.61e-202	1.89e-131	1.28
28	0.26	4.84e-196	3.20e-140	1.20
29	1.3	2.35e-233	9.49e-138	3.16
30	0.67	7.18e-226	2.00e-137	0.00
31	1.05	2.60e-208	1.00e-136	0.20
Média	0.75	1.01e-193	1.38e-127	1.16
Std	0.50	0.0	7.56e-127	2.35

TABLE III
CONFIGURAÇÕES - RASTRIGIN - GBEST

Execuções	Configurações (X, W)			
	(1, 1)	(1, 0.5)	(0.5, 1)	(0.5, 0.5)
1	52.8	15.66	17.05	27.81
2	35.57	15.65	9.98	34.44
3	33.79	31.45	23.88	31.97
4	43.83	33.14	27.86	20.21
5	62.03	23.19	12.95	23.66
6	40.74	7.41	16.98	36.17
7	53.89	22.8	19.9	23.3
8	50.21	28.94	7.06	39.43
9	49.05	21.09	25.87	23.05
10	43.7	29.04	25.87	34.25
11	45.08	16.18	14.93	23.42
12	46.12	26.87	13.94	16.47
13	55.73	16.46	26.87	29.89
14	37.87	31.88	38.83	29.52
15	60.2	33.44	39.86	45.69
16	44.9	19.22	11.94	21.25
17	60.23	20.81	12.93	51.26
18	48	34.74	14.93	21.05
19	43.77	17.04	15.28	33.98
20	40.12	19.66	16.92	41.45
21	58.57	17.8	12.94	14.24
22	50.84	19.26	6.97	45.81
23	36.5	21.25	20.02	28.44
24	43.2	18.45	10.95	26.69
25	37.42	28.44	30.85	40.43
26	44.91	14.76	14.92	27.76
27	57.12	12.05	19.91	37.17
28	34.67	12.95	18.91	26.64
29	46.85	37.21	9.95	31.34
30	46.02	27.96	25.87	39.17
31	50.9	11.33	19.69	22.01
Média	46.92	22.13	18.86	30.57
Std	7.83	7.65	8.14	8.90

B. Rastrigin

Segue a baixo a Tabela III e IV que mostra o comparativo entre o valor médio de função objetivo encontrado pelo algoritmo para cada configuração ao variar os valores de X e W.

De acordo com as Tabelas 1 e 2 é possível perceber que a presença dos hiper-parâmetros se fazem necessário para um bom desempenho do algoritmo. Na primeira coluna, por exemplo, os valores de X e W são iguais a um, exemplificando a ausência dos parâmetros e consequentemente obtendo um resultado não muito satisfatório. Em contra partida, quando há a presença dos hiper-parâmetros, os algoritmos Gbest e Lbest apresentaram um resultado muito superior, dado que atingiram o ótimo global.

Já para a função Rastrigin, ambas as configurações propostas não foram encontrados o mínimo global antes que a condição de parada fosse acionada, o que mostra a alta complexidade desta função para 10 variáveis. Nota-se também que para cada combinação de X e W a média obtida é diferente, sendo necessário uma escolha melhor das constantes. Uma ideia seria aplicar o *Grid Search* para escolher os valores que mais se adequam na função.

TABLE IV
CONFIGURAÇÕES - RASTRIGIN - LBEST

Execuções	Configurações (X, W)			
	(1, 1)	(1, 0.5)	(0.5, 1)	(0.5, 0.5)
1	33.59	24.78	12.03	32.8
2	39.25	27.35	4.83	42.51
3	52.12	31.08	4.42	28.03
4	51.45	26.81	6.71	33.84
5	33.95	30.74	6.49	28.78
6	21.56	18.81	9.97	31.45
7	44.59	26.09	6.61	29.61
8	41.22	27.79	5.31	36.28
9	42.51	18.89	9.05	37.39
10	56.73	25.91	14.57	38.81
11	41.89	22.68	9.2	29.36
12	33.01	31.63	11.18	34.6
13	10.59	21.32	4.87	29.64
14	36.97	24.41	10.99	28.2
15	44.46	28.59	7.91	37.96
16	56.71	27.4	10.2	34.59
17	34.81	22.73	11.15	24.75
18	47.57	24.94	5.08	20.02
19	52.56	31.97	6.45	40.33
20	47.91	37.94	5.57	32.09
21	45.73	25.68	6.24	31.74
22	44.48	29.83	4.79	23.86
23	53.44	12.46	13.12	19.63
24	29.43	24.34	10.99	35.1
25	48.6	32	4.18	32.54
26	47.39	16.05	16.71	30.99
27	47.98	18.41	6.02	41.59
28	54.39	21.53	7.55	33.01
29	33.55	31.79	8	23.52
30	23.16	15.48	10.73	34.29
31	40.86	27.93	14.94	25.05
Média	41.69	25.39	8.57	31.68
Std	10.65	5.60	3.36	5.75

VI. CONCLUSÃO

Os Algoritmos de enxame de partículas se propuseram como uma eficiente solução para resolver problemas unimodais e também para otimização de funções multi-modais. Isso pode ser constatado devido aos valores coletados durante as simulações que apresentaram resultados bem próximos do esperado. Contudo, essa classe de algoritmos requer uma grande atenção na escolha dos hiper-parâmetros pois estes são cruciais para o bom desempenho do algoritmo.

REFERENCES

- [1] Kennedy, J.; Eberhart, R. (1995). Particle Swarm Optimization. Proceedings of IEEE International Conference on Neural Networks. IV. pp. 1942–1948. doi:10.1109/ICNN.1995.488968
- [2] L. A. Rastrigin. Systems of extremal control. Nauka, 1974.
- [3] Scientia Forestalis, volume 40, n. 96 p.547-556, dezembro de 2012. Uso da Meta-Heurística otimização por exame de partículas no planejamento Florestal.