

Particle Swarm Optimization

Algoritmos Bioinspirados

Heitor Lourenço Werneck
heitorwerneck@hotmail.com

1 Introdução

Otimização por Enxame de Partículas (Particle Swarm Optimization - PSO) é uma meta-heurística baseada em população, para otimização, sendo um dos algoritmos mais populares para otimização. A ideia geral é baseada em padrões da natureza, como por exemplo o enxame de pássaros. PSO é inspirado no comportamento social e cooperativo exibido na natureza por diversas espécies. Diferentemente de algoritmos genéticos, programação evolucionária e estratégias de evolução no PSO não há o operador de seleção [7].

As aplicações do PSO são diversas, assim como diversas outras meta-heurísticas, um exemplo delas são: clustering [3]; redes distribuídas [6] e outros [5].

Esse trabalho consiste na implementação e análise de um algoritmo de PSO para minimização das funções de Chung Reynolds e Rosenbrock [4].

2 Problema

Então como já dito anteriormente o trabalho era a minimização de funções, essas funções tem um valor ótimo e um espaço de busca, assim como um número de dimensões a ser definido. Mais especificamente a função de Chung Reynolds é definida como a seguir:

$$f(x) = \left(\sum_{i=1}^D x_i^2 \right)^2 \quad (1)$$

tal que $-100 \leq x_i \leq 100$ e o mínimo global está localizado em $x^* = f(0, \dots, 0)$, $f(x^*) = 0$. E a função de Rosenbrock:

$$f(x) = \sum_{i=1}^{D-1} [100(x_{i+1} - x_i^2)^2 + (x_i - 1)^2] \quad (2)$$

tal que $-30 \leq x_i \leq 30$ e o mínimo global está localizado em $x^* = f(1, \dots, 1)$, $f(x^*) = 0$. O D é o número de dimensões, neste trabalho será usado $D = 20$ (20 dimensões).

3 Solução

Para otimização das funções anteriores será utilizado o PSO, como já dito anteriormente. Este é um modelo estocástico e baseado em população, nele nos temos um número de partículas e cada partícula é representada em forma vetorial pela posição no espaço de busca, assim como a velocidade. Cada partícula irá armazenar o melhor "pessoal" dela, isto é, a melhor posição que ela chegou até então (com a melhor aptidão) [2].

As partículas irão se mover por um processo que incorpora velocidade. A velocidade então será computada através do uso da melhor posição que essa partícula já teve p_i e a posição da melhor partícula vizinha a i . Isto pode ser formalizado como:

$$v_i^{t+1} = wv_i^t + r_1^t c_1 (p_i^t - x_i^t) + r_2^t c_2 (g_i^t - x_i^t) \quad (3)$$

Sendo w um fator de inércia, esse fator quando grande ajuda na procura global enquanto um valor pequeno ajuda na busca local [7], r_1 e r_2 são D amostras no domínio $[0, 1]$, c_1 é uma constante que dá o peso ao fator cognitivo, c_2 é uma constante que dá o peso ao fator social. O componente social e cognitivo são responsáveis pela intensificação e a primeira parte, velocidade multiplicado com inércia, é responsável pela diversificação. x_i é a posição da partícula, v_i a velocidade, p_i a melhor posição da partícula i , g_i a melhor posição na vizinhança da partícula i (irá depender da topologia).

Com essa velocidade é possível computar a próxima posição da partícula, dado pela equação abaixo:

$$x^{t+1} = x^t + v^{t+1} \quad (4)$$

Para manutenção de uma boa busca foi utilizado uma restrição no máximo e mínimo da posição de cada partícula de acordo com a função a ser minimizada. A restrição sobre a velocidade também foi testada, porém para simplificação foi utilizado a não restrição, pois generalizou melhor para as funções.

Para a definição de g_i foram utilizadas algumas topologias básicas, ilustradas abaixo:

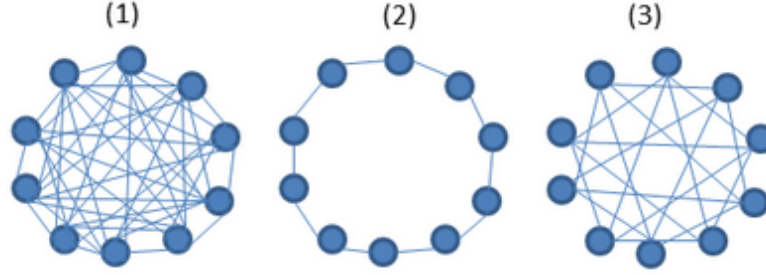


Figura 1: Representação das topologias: (1) completamente conectada, (2) anel, (3) Von Neumann.

No caso da topologia completamente conectada as partículas terão o mesmo g_i já que todas são vizinhas de todas. Na anel a vizinhança de uma partícula será duas outras partículas. Na Von Neumann (Toroidal) toda partícula é conectada com outras 4 partículas. Todas definições de vizinhança são feitas com base no índice, ou seja, não é pela informação espacial assim como no artigo [1]. As topologias podem ser facilmente definidas utilizando $x + imod \#Particulas$, com i sendo um valor limitado superiormente e inferiormente pela metade positiva e negativa do número de vizinhos possíveis em cada topologia. Na implementação otimizações foram feitas em cada topologia, não utilizando uma abordagem generalista que seria ineficiente.

É importante notar que em [1] as topologias são indicadas como são essenciais em PSO, então é importante pelo menos variar esse fator para ver se o problema é mais adequado a alguma topologia.

Para esse modelo teremos os seguintes parâmetros do PSO:

Tabela 1: Parâmetros.

Parâmetro	Descrição
#Iterações	Critério de parada do algoritmo, terminará após #Iterações
#Partículas	Número de partículas a serem utilizadas
w	Inércia (diversificação)
c_1	Fator que pesa o componente cognitivo (intensificação)
c_2	Fator que pesa o componente social (intensificação)
Topologia	A estrutura de vizinhança das partículas

Uma visão geral do algoritmo é descrita no algoritmo 1.

4 Resultados

Para a análise de resultados foi utilizado 10 execuções para cada conjunto de parâmetro, já que o modelo é estocástico é importante sempre sumarizar os resultados de varias execuções para ter uma ideia do resultado esperado.

Algoritmo 1 Algoritmo PSO

```
1: Inicia partículas em uma posição aleatória e velocidade zero
2: Inicia vizinhanças de acordo com a topologia escolhida
3: for  $t = 1$  to #Iterações do
4:   for  $i = 1$  to #Partículas do
5:     Computar a velocidade  $v_i^{t+1} \leftarrow wv_i^t + r_1^t c_1(p_i^t - x_i^t) + r_2^t c_2(g_i^t - x_i^t)$ 
6:     Computar a posição de acordo com a velocidade anterior  $x_i^{t+1} \leftarrow x_i^t + v_i^{t+1}$ 
7:     if  $f(x) < f(p_i)$  then
8:        $p_i^{t+1} \leftarrow x$ 
9:     Atualize  $g$  utilizando como informação  $i$ ,  $x$  e  $f(x)$  e de acordo com uma regra da topologia
```

Foram utilizadas as funções já mencionadas anteriormente, que estão também apresentadas na tabela a seguir [4]:

Tabela 2: Características das funções.

Função	Mínimo	Máximo	Solução	Valor
$f(x) = (\sum_{i=1}^D x_i^2)^2$	-100	100	$x^* = f(0, \dots, 0)$	$f(x^*) = 0$
$f(x) = \sum_{i=1}^{D-1} [100(x_{i+1} - x_i^2)^2 + (x_i - 1)^2]$	-30	30	$x^* = f(1, \dots, 1)$	$f(x^*) = 0$

4.1 Calibragem de parâmetros

Para o começo da análise um experimento fatorial completo será realizado em cada função. Os valores para busca estão indicados na tabela a seguir com o domínio de busca para cada parâmetro (tabela 3).

Tabela 3: Domínio de busca de cada parâmetro.

#Iterações	200
#Partículas	80
w	{0.4, 0.6, 0.8}
c_1	{0.5, 1, 1.5, 2}
c_2	{0.5, 1, 1.5, 2}
Topologia	{Totalmente conectada(TC), Von Neumann, Anel}

4 parâmetros foram escolhidos para variação como é possível ver na tabela anterior totalizando 144 combinações de conjuntos de parâmetros (1440 execuções). O número de iterações e partículas foram escolhidos por sucessos em execuções anteriores e eles são razoáveis para execução nas funções propostas para minimização.

Para os outros parâmetros foram escolhidos domínios representam aproximadamente valores de sucesso em execuções na literatura, variando tanto intensificação quanto diversificação.

Depois da execução o top-15 parâmetros, de cada instância, são descritos nas tabelas 4 e 5, onde μ denota a média e σ denota o desvio padrão das 10 execuções. Nessa tabela os dados são obtidos pela ultima iteração de cada execução.

A coluna "Melhor aptidão global" descreve o melhor valor obtido na ultima iteração, de todas soluções obtidas durante toda execução do PSO.

"Melhor aptidão" descreve a melhor valor obtido nas partículas da última iteração (o valor objetivo da posição atual, não a melhor posição conhecida pela partícula), as outras colunas seguem a mesma lógica.

4.2 Experimento fatorial

Primeiro, começando pela função visivelmente mais simples, a função de Chung Reynolds. Podemos ver que a maior parte dos parâmetros se deram melhor com a topologia totalmente conectada (abreviada para TC), assim como valores extremamente baixos para a função foram obtidos, porém não chegou no ótimo que é 0. Um ponto geral a se notar é que a aptidão média e mediana são bem próximas o que mostra uma

pouca diversidade e alto convergimento para um ponto por todas partículas, assim como esperado dessa topologia, e até mesmo o pior valor é extremamente próximo do melhor, tendo poucas variações.

É possível ver que o melhor conjunto de parâmetros tem um direcionamento maior na solução global já que $c_2 = 1.5$ e um pouco menos em solução local pelo c_1 , porém o w é o menor dos valores disponíveis para esse parâmetro indicando que ele serviu para balancear e favorecer mais busca local, já que só observando os parâmetros anteriores há mais favorecimento na busca global, então com isso há mais balanceamento de ambas intensificações.

Um ponto interessante que os melhores conjuntos de parâmetros são bem similares entre si, indicando que esse problema precisa de certas características na busca que foram descritas anteriormente.

	c_1	c_2	w	Topologia	Melhor aptidão global		Melhor aptidão		Aptidão média		Aptidão mediana		Pior aptidão	
					μ	σ	μ	σ	μ	σ	μ	σ	μ	σ
1	1.0	1.5	0.4	TC	7.143e-43	1.404e-42	7.145e-43	1.404e-42	2.390e-42	4.733e-42	1.976e-42	3.861e-42	8.513e-42	1.617e-41
2	1.5	1.5	0.4	TC	1.488e-41	4.215e-41	1.493e-41	4.214e-41	4.294e-41	1.022e-40	3.038e-41	7.892e-41	4.622e-40	8.401e-40
3	1.5	1.0	0.6	TC	1.448e-36	3.754e-36	1.454e-36	3.752e-36	9.319e-36	2.447e-35	6.181e-36	1.619e-35	5.312e-35	1.227e-34
4	2.0	1.5	0.4	TC	7.497e-33	1.765e-32	7.540e-33	1.764e-32	3.874e-31	8.812e-31	2.637e-32	5.331e-32	2.617e-29	7.062e-29
5	0.5	1.5	0.4	TC	6.369e-31	1.137e-30	6.448e-31	1.145e-30	1.212e-30	2.153e-30	1.008e-30	1.778e-30	4.282e-30	9.406e-30
6	2.0	1.0	0.6	TC	3.241e-30	8.003e-30	3.254e-30	7.999e-30	1.461e-27	2.170e-27	2.268e-29	5.780e-29	1.053e-25	1.677e-25
7	1.0	1.0	0.6	TC	9.620e-30	3.041e-29	9.620e-30	3.041e-29	4.287e-29	1.355e-28	3.419e-29	1.081e-28	2.274e-28	7.192e-28
8	1.0	1.5	0.6	TC	1.214e-23	2.709e-23	1.217e-23	2.708e-23	1.173e-22	2.030e-22	7.121e-23	1.520e-22	1.214e-21	1.421e-21
9	1.0	2.0	0.4	TC	1.576e-23	2.017e-23	1.590e-23	2.038e-23	1.033e-22	1.326e-22	7.221e-23	9.490e-23	7.663e-22	1.047e-21
10	1.0	0.5	0.8	TC	4.388e-23	8.911e-23	4.410e-23	8.900e-23	6.339e-19	1.786e-18	1.514e-22	2.322e-22	5.066e-17	1.428e-16
11	1.5	1.5	0.6	TC	3.489e-21	4.503e-21	3.679e-21	4.978e-21	7.242e-20	7.058e-20	2.398e-20	2.499e-20	1.278e-18	1.094e-18
12	0.5	1.0	0.6	TC	1.741e-20	4.401e-20	1.741e-20	4.401e-20	3.117e-20	7.354e-20	3.056e-20	7.218e-20	7.303e-20	1.693e-19
13	1.5	0.5	0.8	TC	3.264e-20	4.051e-20	3.327e-20	4.036e-20	2.641e-04	8.347e-04	1.317e-19	1.544e-19	2.113e-02	6.678e-02
14	0.5	0.5	0.8	TC	1.717e-18	5.408e-18	1.717e-18	5.408e-18	4.381e-18	1.376e-17	4.008e-18	1.260e-17	1.543e-17	4.850e-17
15	0.5	2.0	0.4	TC	4.025e-18	1.034e-17	4.025e-18	1.034e-17	1.161e-17	2.991e-17	9.646e-18	2.510e-17	7.805e-17	2.105e-16

Tabela 4: Resultados da função de Chung Reynolds.

Para a função de Rosenbrock, ligeiramente mais complexa valores grandes foram obtidos, distantes do ótimo comparado com a outra função, porém de acordo com a literatura são valores bons [7]. O primeiro ponto que se nota é que a topologia de Von Neumann foi a melhor, ou seja, comprovando que a topologia depende da função a ser minimizada e de acordo com a literatura [1]. O melhor conjunto de parâmetro possui uma melhor aptidão global significativamente melhor que a melhor solução da topologia TC, mostrando a efetividade da topologia nesse cenário.

A aptidão média e mediana se diferenciam significativamente, indicando a presença de outliers que são confirmados pela coluna de pior aptidão, isso é uma característica dessa topologia já que grupos vão buscar em locais diferentes o que pode gerar grande diversidade em soluções encontradas assim gerando outliers, porém pode se ver que esse comportamento ocorreu na topologia TC, o que indica alguma característica do problema, que talvez seja extremamente restrito e pequenos passos levem a grandes diferenças no valor da solução, também outro ponto que o ajudaria a ter soluções tão diversas são seus parâmetros que indicam uma grande diversificação e pouco foco no componente social($c_1 = 1, c_2 = 0.5, w = 0.8$). Isso indica que o TC tentou fazer uma diversificação que não é seu forte, enquanto isso a topologia Von Neumann conseguiu convergir muito mais organicamente sem outliers tão extremos, porém sempre com a população diversa.

	c_1	c_2	w	Topologia	Melhor aptidão global		Melhor aptidão		Aptidão média		Aptidão mediana		Pior aptidão	
					μ	σ	μ	σ	μ	σ	μ	σ	μ	σ
1	1.5	1.0	0.6	Von Neumann	15.424	3.9693	15.427	3.971	54.405	54.6226	16.710	4.915	1256.218	1973.643
2	2.0	1.5	0.4	Von Neumann	15.862	3.3902	15.874	3.411	991.89	1047.22	36.341	29.932	39974.04	54332.32
3	2.0	1.0	0.6	Von Neumann	16.582	2.6474	16.585	2.647	195.33	229.285	27.141	20.743	9866.966	14652.23
4	1.0	0.5	0.8	Von Neumann	17.751	2.3579	17.816	2.394	100.34	162.104	20.659	4.332	3512.359	8764.795
5	1.0	0.5	0.8	TC	19.598	23.450	19.599	23.450	199.84	412.606	20.048	23.287	9513.162	22921.10
6	1.0	1.5	0.4	Von Neumann	19.780	18.866	19.780	18.866	126.44	235.322	29.599	30.807	5074.415	12906.29
7	1.0	1.5	0.4	TC	19.813	19.271	19.813	19.271	65.821	82.4565	19.974	19.210	2314.998	4811.549
8	0.5	0.5	0.8	TC	19.879	17.107	19.879	17.107	20.004	17.0584	19.961	17.075	20.68030	16.82604
9	2.0	1.0	0.6	TC	21.116	23.152	21.120	23.151	52.627	36.8753	29.321	24.849	560.3266	664.6859
10	0.5	1.0	0.6	Von Neumann	21.134	20.599	21.135	20.598	28.348	22.0904	21.581	20.480	186.2836	309.7003
11	0.5	1.0	0.8	TC	21.573	23.697	21.607	23.799	159.26	347.769	25.203	33.885	9125.846	26975.33
12	1.0	1.5	0.6	Von Neumann	21.673	17.430	21.690	17.423	292.41	507.186	30.224	28.802	10792.01	24763.06
13	0.5	0.5	0.8	Von Neumann	22.597	11.343	22.598	11.343	41.617	23.9151	31.109	20.021	660.4136	983.2309
14	0.5	1.5	0.4	Von Neumann	23.399	27.555	23.399	27.555	39.189	46.3824	24.448	30.144	493.8112	938.5562
15	1.5	1.0	0.4	TC	24.098	19.245	24.098	19.245	40.546	62.1953	25.330	21.092	738.5723	2266.834

Tabela 5: Resultados da função de Rosenbrock.

Um ponto a se acrescentar é que o Anel não apareceu no top-15 nenhuma vez, isso indica que ele não se encaixa tanto nesses problemas, o que é normal, talvez em trabalhos futuros com outras funções ele possa ser o melhor.

4.3 Análise das melhores soluções

Na figura 2 é possível ver que o PSO converge rapidamente para um ponto muito baixo, isso mostra que a intensificação feita está sendo muito forte e consegue convergir facilmente para um valor bem baixo, como a função a ser minimizada é extremamente simples então não há muita necessidade de diversificação para procura muito dispersa no espaço de soluções e a topologia completamente conectada é ideal e domina por grande margem as outras duas topologias. É possível ver que não há muita diferença na convergência entre execuções também.

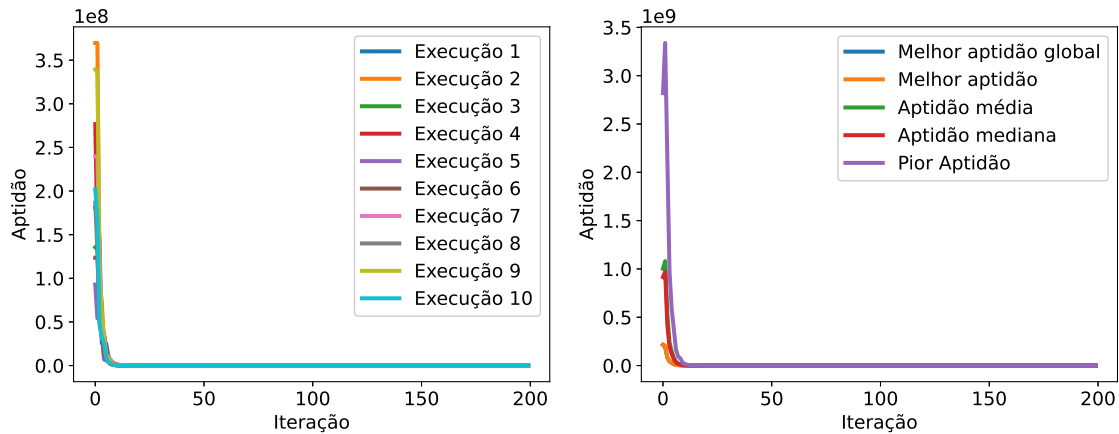


Figura 2: Execuções do melhor conjunto de parâmetros da função Chung Reynolds.

Na figura 3 já é possível ver diversas questões que são próprias, além da função que é complexa, a topologia usada, a pior aptidão demora um tempo a chegar próximo de todas outras soluções, que mostra um caso perfeitamente possível e que é comum de algum grupo de partículas ficarem um pouco abandonadas, porém com a transmissão de informação através da rede de vizinhos as partículas conseguem ir direcionando a boas soluções, para esse problema que é uma função mais complexa que a anterior, uma diversificação no espaço de soluções é o ideal e é isso que é possível observar nesses gráficos. Se formos prestar atenção aos pequenos detalhes é possível ver que a curva da melhor aptidão até virar uma reta é extremamente arredondada mostrando um nível de intensificação não tão grande. No problema anterior tivemos um nível maior de intensificação e a curva está mais para uma reta.

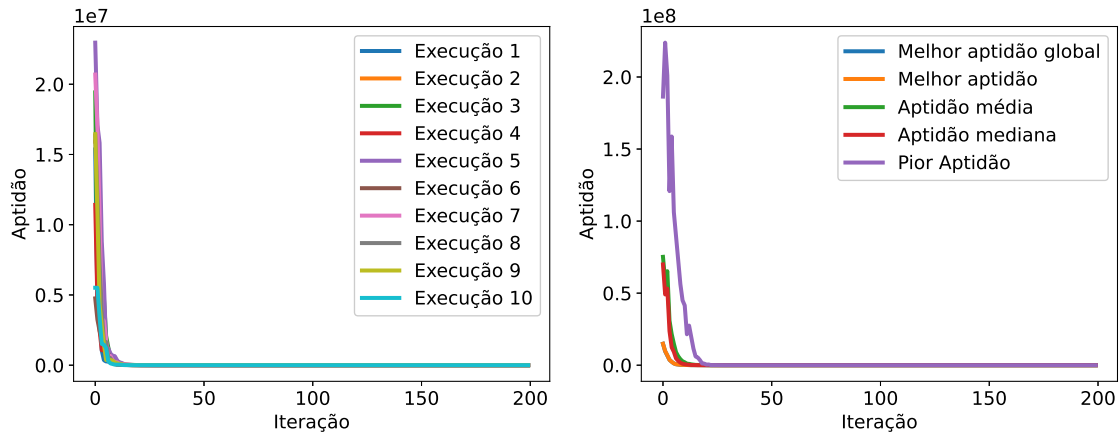


Figura 3: Execuções do melhor conjunto de parâmetros da função Rosenbrock.

5 Conclusão

Com esse trabalho foi possível ver na prática a efetividade do PSO para solucionar 2 funções da literatura, indicando sua efetividade.

Diversas análises foram feitas sobre os parâmetros, seus impactos, como eles funcionam e se comportam. Assim como características das topologias e outros.

Em trabalhos futuros métodos mais elaborados, com limitação de velocidade, atribuição dinâmica de alguns parâmetros e topologias dinâmicas com informação espacial podem ser feitas para comparar e gerar ainda mais ideias relevantes para a área de otimização.

Referências

- [1] A comparative study of neighborhood topologies for particle swarm optimizers. In *Proceedings of the International Joint Conference on Computational Intelligence*, page nil, - 2009.
- [2] Julio Barrera, Osiris Álvarez-Bajo, Juan J Flores, and Carlos A Coello Coello. Limiting the velocity in the particle swarm optimization algorithm. *Computación y Sistemas*, 20(4):635–645, 2016.
- [3] Ching-Yi Chen and Fun Ye. Particle swarm optimization algorithm and its application to clustering analysis. In *IEEE International Conference on Networking, Sensing and Control, 2004*, page nil, - nil.
- [4] JAMIL Momin and Xin-She Yang. A literature survey of benchmark functions for global optimization problems. *Journal of Mathematical Modelling and Numerical Optimisation*, 4(2):150–194, 2013.
- [5] Riccardo Poli. Analysis of the publications on the applications of particle swarm optimisation. *Journal of Artificial Evolution and Applications*, 2008(nil):1–10, 2008.
- [6] Ping Ren, Li-Qun Gao, Nan Li, Yang Li, and Zhi-Ling Lin. Transmission network optimal planning using the particle swarm optimization method. In *2005 International Conference on Machine Learning and Cybernetics*, page nil, - 2005.
- [7] Yuhui Shi and Russell C Eberhart. Empirical study of particle swarm optimization. In *Proceedings of the 1999 congress on evolutionary computation-CEC99 (Cat. No. 99TH8406)*, volume 3, pages 1945–1950. IEEE, 1999.