

Aplicação de otimização por enxame de partículas na estimativa de parâmetros de itens calibrados com teoria da resposta ao item

Victor H. W. Freire^{1,2}, Rubens E. Carneiro^{1,2}, Rodrigo E. Carneiro^{1,2}

¹Escola Politécnica de Pernambuco – Universidade de Pernambuco (UPE)
Recife – PE – Brasil

²Mídias Educativas Ltda.
Recife – PE – Brasil

vhwf@poli.br, {rec2,rec}@ecom.poli.br

Abstract. *This paper defines an approach using a population-based optimization algorithm, the Particle Swarm Optimization (PSO), to obtain parameters used by the Item Response Theory (IRT), which is the official assessment method for the Exame Nacional do Ensino Médio (ENEM), the standardized test used for college admissions in Brazil. The proposed technique consists in extracting a representative sample from test-takers and then applying PSO and IRT to the extracted dataset to obtain each student's grade which are later used to calculate the error against the official published grades. The results showed that it is viable to use Swarm Intelligence to obtain parameters for items calibrated by IRT while achieving equivalent or less error than traditional methods.*

Resumo. *Este trabalho faz uso da aplicação de uma técnica de otimização por enxame, o Particle Swarm Optimization (PSO), para obtenção dos parâmetros utilizados pela Teoria de Resposta ao Item (TRI) no Exame Nacional do Ensino Médio (ENEM). O método proposto consiste na extração de uma amostra representativa de respondentes do ENEM, aplicação do PSO juntamente com a TRI para obtenção das notas e posterior cálculo dos erros em relação às notas oficiais divulgadas. O resultado mostrou a viabilidade da utilização de inteligência de enxames para obtenção de parâmetros de questões já calibradas pela TRI, com erros equivalentes, ou até menores que os obtidos pelos métodos tradicionais.*

1. Introdução

Com a utilização do Exame Nacional do Ensino Médio (ENEM) e do Sistema de Avaliação do Ensino Básico (SAEB) como principais instrumentos de avaliação da educação brasileira, estes exames passaram a ser referência, tanto como índices de ensino, como para construção de avaliações. De acordo com o Ministério da Educação (MEC), o ENEM utiliza a Teoria de Resposta ao Item (TRI) para fazer análise de questões e de proficiências dos respondentes.

Neste presente artigo, a partir dos vetores-resposta obtidos dos dados disponibilizados pelo ENEM, foi utilizada a técnica do Particle Swarm Optimization (PSO), juntamente com a TRI, para obtenção dos parâmetros das questões da prova de Matemática e suas Tecnologias do ano de 2009 do ENEM.

A importância da obtenção destes parâmetros está na possibilidade de criação de novos itens calibrados com a mesma métrica dos itens originais, o que permite uma comparação entre resultados de testes com a prova do ENEM ou qualquer outra que faça uso da TRI. Tais testes podem potencializar sistemas inteligentes por aproximar os resultados e dificuldades reais dos aprendizes [Carneiro, Lima Neto, & Silveira, 2014].

2. Fundamentação teórica

2.1 Teoria de Resposta ao Item (TRI)

A Teoria de Resposta ao Item (TRI) é uma teoria estatístico-matemática que pertence a um ramo da teoria da medida que propõe modelos para obter características do indivíduo que não conseguem ser observadas diretamente. Tais variáveis devem ser inferidas a partir da observação de variáveis secundárias que estejam relacionadas a ela. O que a TRI sugere são formas de relacionar a habilidade latente de um indivíduo a probabilidade deste mesmo indivíduo de dar uma resposta correta a um item [Andrade, Tavares, & Valle, 2000].

Recorrentemente usa-se a Teoria Clássica dos Testes, ou TCT, baseada em resultados obtidos em avaliações a partir dos escores brutos, porém esta metodologia contém diversas limitações, tal como ser altamente dependente dos itens que compõem o instrumento de medida. Ou seja, a TCT analisa o conjunto de itens que formam a avaliação.

As vantagens da aplicação do TRI em comparação à TCT são amplas, como por exemplo: A TRI envolve a possibilidade de construir várias avaliações constituídas por alguns itens exclusivos e alguns itens compartilhados, e ainda assim garantir que as dificuldades dessas avaliações em questões sejam equivalentes, ou ainda, a comparação entre indivíduos da mesma população que tenham sido submetidos a provas totalmente diferentes. Isto porque a TRI usa como principais elementos centrais os itens, e não a prova como um todo.

A TRI é um conjunto de modelos matemáticos que tem por objetivo representar a probabilidade do avaliado em questão dar uma resposta certa a um item tendo como função os parâmetros dos itens e a habilidade de quem responde. Tal relação é dada de forma que quanto maior a habilidade do indivíduo que responde o item, maior a probabilidade que este tem de responder corretamente à questão. Entre os modelos utilizados pela TRI, destaca-se o Modelo Logístico de 3 Parâmetros (ML3P), onde os itens são tratados de forma dicotômica (corrigidos como certo ou errado) [Andrade et al., 2000].

No ML3P, cada questão é representada por uma curva característica e a probabilidade de um indivíduo com habilidade θ responder corretamente ao item é dada por (1):

$$P_i(\theta) = P(\theta) = c_i + \frac{1-c_i}{1+e^{-Da_i(\theta-b_i)}} \quad (1)$$

Onde:

U_{ij} : é uma variável dicotômica, ou seja, assume o valor 1 quando o item i é respondido corretamente pelo indivíduo j e o valor 0 quando o item i é respondido erroneamente pelo indivíduo j .

θ_j : habilidade do respondente j . Normalmente se encontra no intervalo $[-4,4]$.

B_i : é o parâmetro de dificuldade do item i , medido na mesma escala da habilidade.

A_i : é o parâmetro de discriminação do item i . Normalmente se encontra no intervalo $[0,8, 2,5]$.

C_i : é o parâmetro do item i que representa a probabilidade do aluno de baixa habilidade responder corretamente ao item em questão (muitas vezes chamado de parâmetro que representa a probabilidade de acerto casual). Como se trata de uma probabilidade, seu valor flutua entre 0 e 1.

D : é o fator de escala, é constante e de valor, normalmente igual a 1. Pode assumir também o valor 1,7 quando se deseja que a função logística retorne um valor semelhante ao da função ogiva normal.

$P(U_{ij} = 1 | \theta_j)$: é a probabilidade de um indivíduo j responder corretamente ao item i dada sua habilidade θ [Spenassato & Kinas, 2010]. É chamada de função resposta do item [Andrade et al., 2000].

2.2 Particle Swarm Optimization (PSO)

O Particle Swarm Optimization (PSO) é uma técnica de otimização estocástica criada em 1995 por James Kennedy e Russell Eberhart para simular graficamente o comportamento de um bando de pássaros [Kennedy & Eberhart, 1994]. Um enxame pode ser caracterizado como um conjunto estruturado de organismos interativos, em que cada componente é demasiadamente simples, mas o comportamento coletivo em questão passa a ser complexo. Portanto, o comportamento global do enxame de organismos sociais surge de maneira não linear através do comportamento dos indivíduos pertencentes ao enxame, de forma que haja uma forte relação entre o comportamento de cada componente e o comportamento geral do enxame.

Assim, a conduta coletiva dos indivíduos caracteriza a conduta do enxame. Porém, o comportamento do enxame não é unicamente ditado pelo comportamento dos indivíduos, independentemente de outros indivíduos, mas a interação entre cada componente do enxame executa função imprescindível para a composição do comportamento do enxame.

O princípio que coordena o PSO é a interação entre os indivíduos que compõem o enxame, cada um deles aprende uns com os outros e, de acordo com o conhecimento adquirido, cada componente se movimenta de forma a se aproximar de seus melhores vizinhos.

Cada partícula armazena suas próprias coordenadas no espaço problema, que estão associadas com a melhor solução (fitness) que foi encontrada até o momento e é chamado de Pbest. Além disso, outro valor que é armazenado é a posição que retorna o melhor valor global do enxame, ou seja, é a localização da partícula que encontrou o melhor resultado, que é chamado de Gbest. O PSO consiste, então, em a cada etapa, mudar a velocidade de cada partícula em direção do Pbest e Gbest [Eberhart & Yuhui Shi, 2002].

A posição da partícula é modificada adicionando o vetor velocidade ao seu posicionamento, vetor este que é dado por:

$$\vec{v}_i(t+1) = \omega \vec{v}_i(t) + c_1 r_1 [\vec{P}_{best\ i} - \vec{x}_i(t)] + c_2 r_2 [\vec{G}_{best} - \vec{x}_i(t)] \quad (2)$$

onde $\vec{v}_i(t+1)$ será a nova velocidade da partícula, enquanto $\vec{v}_i(t)$ e $\vec{x}_i(t)$ são, respectivamente a velocidade e a posição atual. Na equação (1.2), ω representa o fator de inércia, c componentes sociais e r coeficientes aleatórios. O t em questão são as iterações.

A nova posição da partícula, então, é dada por:

$$\vec{x}_i(t+1) = \vec{v}_i(t+1) + \vec{x}_i(t) \quad (3)$$

A técnica se mostra eficiente na resolução de problemas onde o espaço de busca é contínuo, em consequência da forma que a velocidade e a posição da partícula é atualizada [Eberhart, 2006].

O algoritmo do PSO segue o pseudo-código abaixo:

Algoritmo 1. Pseudo-código PSO

```
PARA CADA partícula
    Inicializa partícula
FIM
FAÇA
    PARA CADA partícula
        Calcula valor de aptidão (fitness)
        SE o valor de aptidão for o melhor já ocorrido na partícula (pBest)
            Atualiza o valor o pBest
        FIM
    Escolha a partícula com o melhor valor de aptidão de todas, como a gBest
    PARA CADA partícula
        Calcule a velocidade da partícula de acordo com a equação (2)
        Atualize a posição das partículas de acordo com a equação (3)
    FIM
ENQUANTO não atender o critério de número máximo de interações ou erro mínimo
RETORNE o gBest
```

A técnica do PSO foi escolhida pois é uma meta heurística capaz de otimizar parâmetros em espaço de busca contínuo. Por conseguinte, a forma de atualização das posições das partículas tem eficiência em domínios contínuos. [Bonyad & Michalewicz, 2016]

3. Metodologia

Esta pesquisa foi realizada a partir dos MICRODADOS do ENEM disponibilizados pelo INEP. Do conjunto de dados disponibilizados, foram utilizados os vetores resposta dos participantes, seus respectivos números de inscrição, a situação de conclusão, o código da prova respondida e os estados de realização do exame.

O tratamento dos dados foi feito a partir da linguagem Python e banco de dados PostgreSQL. Foram obtidos 904.998 concluintes dos quais 228.221 realizaram o tipo de prova que foi chamado de “tipo 1”, cujo código no MICRODADOS é 62. Da amostra de 228.221 respondentes foi utilizado, para treinamento do modelo, 30% dos dados, visto o grande custo computacional da simulação, conforme mostra a Tabela 1.

Tabela 1. Forma de coleta de dados. Fonte: Autores, 2019

TOTAL CONCLUINTES			PROVA TIPO 1		
UF	INSCRITOS	% DO TOTAL	INSCRITOS	SELECIONADOS	% DO TOTAL PROVA 1
AC	3.015	0,33%	755	217	0,32%
AL	9.631	1,06%	2.440	717	1,06%
AM	21.123	2,33%	5.215	1.595	2,36%
AP	2.772	0,31%	726	198	0,29%
BA	58.377	6,45%	14.766	4.369	6,47%
CE	33.829	3,74%	8.536	2.520	3,73%
DF	13.659	1,51%	3.360	1.020	1,51%
ES	28.572	3,16%	7.362	2.105	3,12%
GO	27.960	3,09%	7.084	2.108	3,12%
MA	28.790	3,18%	7.199	2.154	3,19%
MG	104.641	11,56%	26.519	7.778	11,52%
MS	14.440	1,60%	3.673	1.081	1,60%
MT	20.696	2,29%	5.248	1.548	2,29%
PA	26.295	2,91%	6.669	1.921	2,85%
PB	13.676	1,51%	3.503	1.008	1,49%
PE	44.829	4,95%	11.146	3.338	4,95%
PI	19.522	2,16%	5.000	1.461	2,16%
PR	58.197	6,43%	14.648	4.357	6,45%
RJ	75.009	8,29%	19.011	5.596	8,29%
RN	15.838	1,75%	3.997	1.203	1,78%
RO	8.477	0,94%	2.158	622	0,92%
RR	1.786	0,20%	450	130	0,19%
RS	53.707	5,93%	13.426	4.004	5,93%
SC	26.133	2,89%	6.557	1.997	2,96%
SE	8.669	0,96%	2.148	647	0,96%
SP	176.845	19,54%	44.462	13.169	19,51%
TO	8.510	0,94%	2.163	637	0,94%
TOTAL	904.998	100%	228.221	67.500	30%

A prova de Matemática e suas Tecnologias foi escolhida para a análise.

Os parâmetros do modelo logístico de 3 parâmetros (ML3) referente à TRI, das questões da prova de matemática, de acordo com os vetores respostas dos 67.500 respondentes citados acima, foram então, calculados a partir do algoritmo PSO [Clerc, 2006].

Para testar os parâmetros obtidos foram selecionados, aleatoriamente e distintamente dos 67.500 utilizados para treinamento, 10 grupos de teste com 2200 respondentes cada. As notas dos grupos de teste foram calculadas, também de acordo com a TRI, com os parâmetros obtidos a partir grupo de treinamento, e comparadas com as notas divulgadas pelo INEP.

As métricas escolhidas para comparar os resultados foram:

- Erro relativo percentual médio (ERP):

$$\frac{\sum \frac{|(p-y)| * 100\%}{y}}{n} \quad (4)$$

Em que:

p: nota calculada
y: nota real fornecida pelo ENEM
n: número de notas utilizadas para o cálculo

- Erro médio (EM):

$$\frac{\sum |(p-y)|}{n} \quad (5)$$

Em que:

p: nota calculada
y: nota real fornecida pelo ENEM
n: número de notas utilizadas para o cálculo

- Root Mean Squared Error (RMSE):

$$\sqrt{\frac{\sum (y-p)^2}{n}} \quad (7)$$

Em que:

p: nota calculada
y: nota real fornecida pelo ENEM
n: número de notas utilizadas para o cálculo

3.1 Modelagem do problema

A prova escolhida para o experimento possui 45 itens e cada item é caracterizado por 3 parâmetros, a, b e c, como mostrado na equação 1. Dessa forma, foi definido um vetor com 135 elementos (45 itens x 3 parâmetros) para ser otimizado pelo PSO. O tamanho do exame utilizado foi 33 partículas definidos pela equação 8, onde *par* é o vetor de parâmetros.

$$\text{floor}(10 + 2 * \text{sqrt}(\text{length}(\text{par}))) \quad (8)$$

A função de aptidão utilizada calcula a nota de cada aluno a partir de suas respostas para cada item e os parâmetros da melhor partícula do PSO. A diferença entre a nota computada e a nota do ENEM de todos os respondentes é o erro que o pso otimiza.

Por conta da inicialização aleatória das partículas do PSO executamos 6 vezes o experimento com *seeds* diferentes e comparamos os resultados obtidos com o do TRI.

Os experimentos foram realizados utilizando a linguagem R [Team, 2018] com as bibliotecas *mirt* [Chalmers, 2012] e *pso* [Bendtsen, 2012] para TRI e PSO respectivamente.

4. Resultados

A partir dos parâmetros obtidos pelo PSO, foram calculadas as notas dos participantes por grupo e comparadas com as notas fornecidas pelo ENEM. Os erros obtidos estão expostos nas Figuras 1, 2 e 3 a seguir:

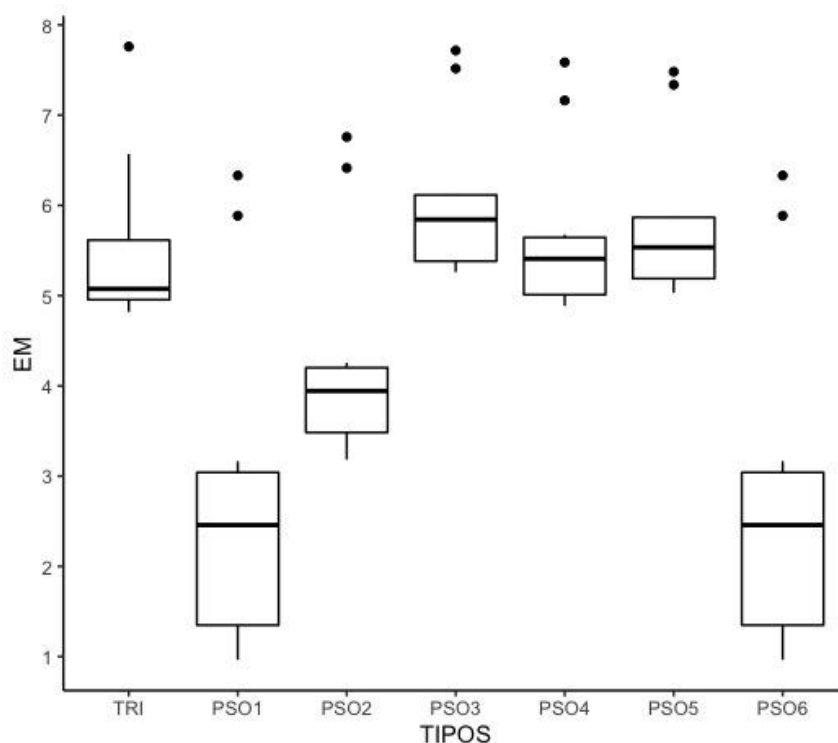


Figura 1. Erro médio dos conjuntos de teste.

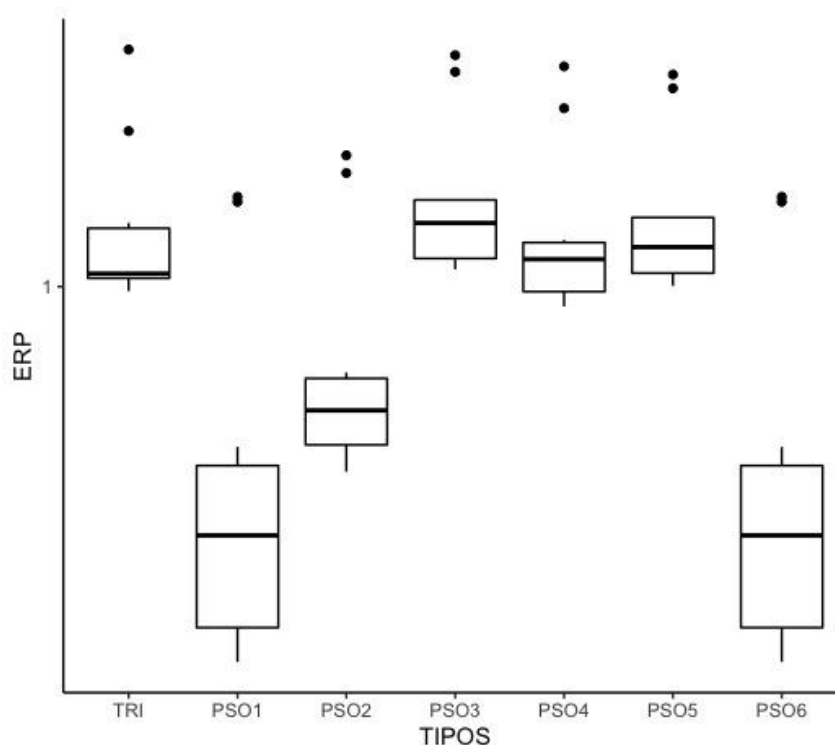


Figura 2. Erro relativo percentual.

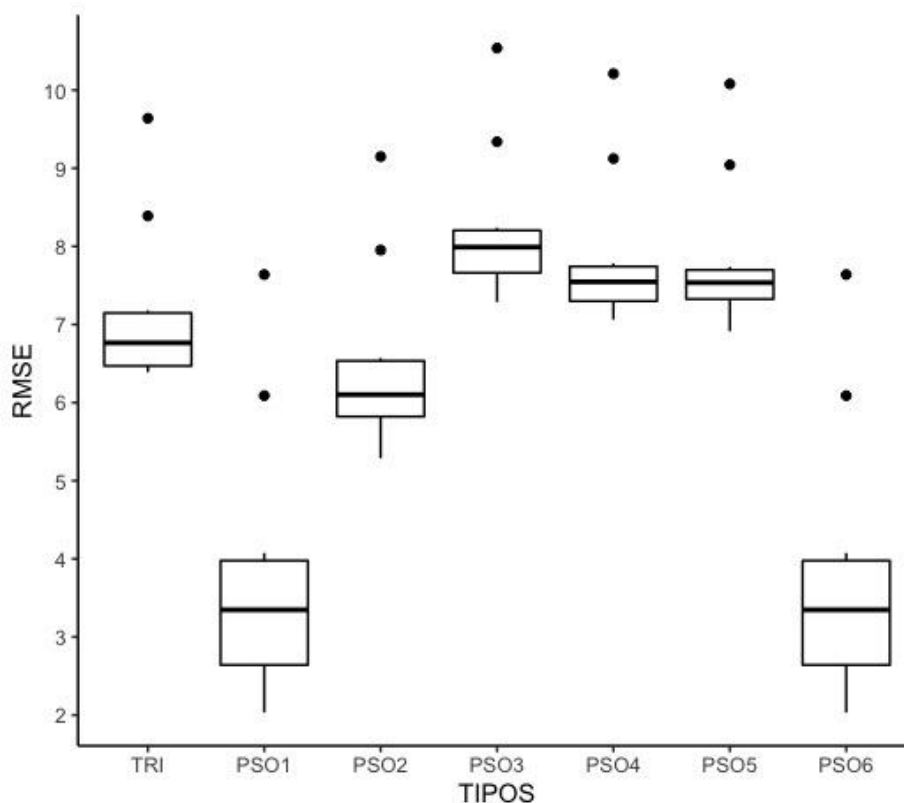


Figura 3. Raiz do erro médio quadrático.

Nas Figuras 1, 2 e 3, estão dispostos, a partir de gráficos de tipo *boxplot*, os erros médios, erros relativos percentuais e as raízes dos erros médios quadráticos, respectivamente. Em sequência, da esquerda para direita, são observados os tipos de erros obtidos a partir do cálculo tradicional do TRI (a partir da biblioteca MIRT), PSO1, PSO2, PSO3, PSO4, PSO5 e PSO6, respectivamente. O erro médio calculado com a biblioteca Mirt foi de apenas 5,52 pontos em relação às notas divulgadas pelo ENEM o que representou erro percentual de 1,12% e raiz do erro médio quadrático igual a 7,15. Os resultados obtidos a partir do PSO obtiveram erros médios que variaram entre 2,81 e 6,06 pontos (2,99, 4,29, 6,06, 5,67, 5,81 e 2,81), os erros percentuais variaram entre 0,54% e 1,19% (0,56%, 0,81%, 1,19%, 1,11%, 1,15% e 0,54%) e a raiz do erro médio quadrático variou entre 3,81 e 7,83 (4,27, 6,48, 8,23, 6,96, 7,83 e 3,81).

O resultado, então, mostrou a viabilidade da utilização de inteligência de enxames para obtenção de parâmetros de questões já calibradas pela TRI mas que não são conhecidos. Exames nacionais, como o ENEM (Exame Nacional do Ensino Médio) e SAEB (Sistema de Avaliação do Ensino Básico), utilizam-se da Teoria de Resposta ao Item para cálculo das proficiências dos respondentes, porém, os parâmetros dos itens utilizados por essas provas não estão amplamente disponíveis e não são de fácil acesso. Portanto, como os erros obtidos foram equivalentes ou até menores que os obtidos pelos métodos tradicionais, o PSO pode ser utilizado para obter tais parâmetros.

5. Conclusão

Obtendo-se as notas (proficiência) e os vetores resposta dos respondentes, a técnica do PSO pode ser utilizada para obtenção dos parâmetros da TRI. Parâmetros esses que permitem a avaliação de cada item quanto a sua discriminação, ou seja, o quanto aquele item distingue os respondentes, a sua dificuldade e o seu parâmetro de acerto ao acaso. Pode-se utilizar uma amostra representativa dos dados, objetivando aperfeiçoar o

tempo computacional da técnica. O atingimento dos parâmetros da TRI pode ser de grande valia para avaliação de respondentes, de forma a construir avaliações equivalentes e, assim, comparar resultados obtidos a partir de diferentes exames [Andrade et al., 2000].

O teste de outras versões mais recentes do PSO além de outros algoritmos de otimização por enxames como o ABC [Karaboga & Basturk, 2007] e o FSS [Bastos Filho, de Lima Neto, Lins, Nascimento, & Lima, 2008] podem trazer uma melhora significativa no resultado obtido.

6. Referências

- Andrade, D. F. De, Tavares, H. R., & Valle, R. da C. (2000). *Teoria da Resposta ao Item : Conceitos e Aplicações*. Associação Brasileira de Estatística.
- Bastos Filho, C. J. a., de Lima Neto, F. B., Lins, A. J. C. C., Nascimento, A. I. S., & Lima, M. P. (2008). A novel search algorithm based on fish school behavior. *2008 IEEE International Conference on Systems, Man and Cybernetics*, 2646–2651. doi:10.1109/ICSMC.2008.4811695
- Bendtsen, C. (2012). pso: Particle Swarm Optimization. Retrieved from <https://cran.r-project.org/package=pso>
- Carneiro, R. E., Lima Neto, F. B., & Silveira, D. S. (2014). Modelagem de Usuários Baseada em Estilo de Aprendizagem, Teoria da Resposta ao Item e Lógica Fuzzy para Sistemas Adaptativos Educacionais (p. 21). doi:10.5753/cbie.wcbie.2014.21
- Chalmers, R. P. (2012). mirt : A Multidimensional Item Response Theory Package for the R Environment. *Journal of Statistical Software*, 48(6). doi:10.18637/jss.v048.i06
- Bonyadi, M. R., & Michalewicz, Z. (2016). Particle Swarm Optimization for Single Objective Continuous Space Problems: A Review. *Evolutionary Computation*, 25(1), 1–54. doi.org/10.1162/EVCO_r_00180
- Clerc, M. (2006). *Capítulo 15. On Parallelism. Particle Swarm Optimization*.
- Eberhart, R. (2006). Particle Swarm Optimization Particle Swarm Optimization The inventors : Particle Swarm Optimization Particle Swarm Optimization. *Science*, 1–12.
- Eberhart, & Yuhui Shi. (2002). Particle swarm optimization: developments, applications and resources, 81–86. doi:10.1109/cec.2001.934374
- Karaboga, D., & Basturk, B. (2007). A powerful and efficient algorithm for numerical function optimization: artificial bee colony (ABC) algorithm. *Journal of Global Optimization*, 39(3), 459–471. doi:10.1007/s10898-007-9149-x
- Kennedy, J., & Eberhart, R. (1994). Prognostic evaluation of abdominal echography in typhoid fever. *Giornale di Malattie Infettive e Parassitarie*, 46(10), 1942–1948. doi:10.1109/ICNN.1995.488968
- Spenassato, D., & Kinas, P. G. (2010). Teoria da Resposta ao Item : estimação dos parâmetros pelos métodos de Amostragem por Importância Adaptativa e Monte Carlo via Cadeias de Markov. In *CNMAC* (pp. 1004–1005).
- Team, R. C. (2018). R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <https://www.r-project.org/>