

Minerando dados em grandes quantidades - SURVEY

Othon L. T. Oliveira
Escola Politécnica de Pernambuco
Engenharia de Sistemas
Universidade de Pernambuco
Recife, Pernambuco 50720-001 - Benfica
Email: olto@ecomppoli.br

Fernando B. L. Neto
Universidade de Pernambuco
PhD - UK
Recife, Pernambuco 50720-001 - Campus: Benfica
Email: fbln@ecomppoli.br

Abstract—The abstract goes here.

1. Introdução

As técnicas tradicionais de busca não são eficientes para resolver muitos problemas com grande complexidade, i.e. ordens de grandeza dantes inimagináveis, especialmente quando possuem complexidade exponenciais, geradas por explosões combinatórias, felizmente, muitos desses problemas não-triviais são eficientemente resolvidos por soluções naturais [?].

As heurísticas tornam os custos computacionais mais aceitáveis, na medida que não exploram todas as possibilidades, somente as mais prováveis de se encontrar a solução.

A natureza tem contribuído como inspiração de soluções para vários problemas de buscas, muitos, de maneira eficaz e elegante. Essas soluções são encontradas nas mais diversas espécies de seres vivos. Por exemplo, as formigas quando vão em busca de alimentos, facilmente encontram um caminho mais curto entre o ninho e a fonte de alimentos. Os cardumes de peixes executam movimentos aparentemente aleatórios mas, quando em grupo, são precisos para fuga dos predadores. Nuvens de insetos quando em busca de novos locais para ninhos ou de alimento, inspiram os mais diversos algoritmos inteligentes baseados em populações de animais sociais, demonstrando que há uma inteligência coletiva nessas populações, desenvolvida ao longo do tempo e das interações entre essas espécies sociais e o meio ambiente.

A classe de algoritmos proposta, metaforiza o comportamento de tais populações, e promoveu o desenvolvimento de uma área que hoje é conhecida como computação bioinspirada ou computação natural. Essa área investiga a relação entre a computação e a biologia (e mesmo a sociologia), estudando soluções de buscas e otimização, modelando problemas mais eficientemente, baseado nas elegantes soluções encontradas pela natureza pensou-se que outras inspirações podem e devem originar novos algoritmos, daí a presente proposta.

A essa classe de algoritmos bioinspirados chamamos de meta-heurísticas.

2. Grande volume de dados

Inferir sobre algum assunto agora pode ser coisa do passado. Com o grande volume de informações existentes no Big Data, astrônomos atualizam suas descobertas numa base de dados disponíveis para outros utilizarem, as ciências biológicas agora têm tráfego em depositar seus avanços científicos em repositórios públicos [?]. Redes sociais estão focadas na Web; Facebook, LinkedIn, Microsoft, Tweeter e Yahoo, livros eletrônicos; Amazon, Submarino, *et cetera*, compras online; Americanas.com, Magazine Luiza, *et cetera*, artigos científicos dos mais variados assuntos, alimentam, todos os dias, o que se passou a chamar Big Data.

”Em 2010 empresas e usuários armazenaram mais de 13 exabytes de novos dados” ([?]) [?].

Assim que a Internet das Coisas for uma realidade, a explosão combinatória de coisas conectadas fará com que esse Big Data cresça de maneira assustadora.

Essas informações são dados estruturados e não estruturados (citar)

3. Enxame de partículas

Em 1989, G. Beni e J. Wang cunharam a expressão *Swarm Intelligence*, no seu trabalho em *Robotic Swarm* [?]. O estudo do reino animal aprofundou-se no estudo comportamental e possibilitou o melhor entendimento de como cooperam indivíduos dentro de um grupo e quais os mecanismos usados para controlar o enxame e condicionar o indivíduo, tais como a estigmergia. Por enxame, pode entender-se manada, alcateia, bando, colônia, entre outras designações conforme o animal ou inseto e, a partir daqui, qualquer referência a um grupo de agentes passa a ser feita por enxame, e.g., um enxame de passaros. Os 5 princípios da inteligência de enxame segundo [?] [?], são:

- Proximidade: os agentes têm que ser capazes de interagir
- Qualidade: os agentes devem ser capazes de avaliar seus comportamentos
- Diversidade: permite ao sistema reagir a situações inesperadas

- Estabilidade: nem todas as variaes ambientais devem afetar o comportamento de um agente
- Adaptabilidade: capacidade de se adequar as variaes ambientais

3.1. Particle Swarm Optimization - (PSO)

Kennedy e Eberhart (1995), criaram a popular Otimizao por Enxame de Partculas do ingls *Particle Swarm Optimization* - (PSO), que na verdade uma tcnica de otimizao de funes no-lineares baseado em populaes. Foi inspirado no comportamento social em bando de psaros, essa tcnica uma das mais conhecidas e investigadas hoje [?]. Esse algoritmo (e todos os outros) tm um fator de convergncia, para fazer com que encontre mais rapidamente a resposta procurada. Para acontecer isso foi introduzido um "poleiro" virtual. Com essa simples abordagem foi inaugurado uma nova famlia de algoritmos baseados em enxames.

No PSO, a populao chamada enxame e os indivduos, partculas. Cada partcula se move no espao de busca, procura de regies promissoras; cada partcula dessas representa uma soluo candidata a resolver nosso problema. A equao utilizada para encontrar uma partcula no espao de busca foi emprestada da cinemtica:

$$x_i(t+1) = x_i(t) + v_i(t+1) \quad (1)$$

Onde $x_i(t)$ a posio da partcula num determinado momento "t". O $x_i(t+1)$ a posio atual da partcula.

A velocidade da partcula de acordo com a equao:

$$v_i(t+1) = v_i(t) + c_1 r_1 j(t) [Nbest - x_i(t)] + c_2 r_2 j(t) [Lbest - x_i(t)] \quad (2)$$

Onde $v_i(t)$ a velocidade num momento "t" qualquer e $v_i(t+1)$ a velocidade atual da partcula. Os coeficientes $c_1 r_1 j(t)$ e $c_2 r_2 j(t)$ so nmeros que variam entre 0 e 1 para mudar a posio da partcula. O coeficiente $Nbest - x_i(t)$ a melhor posio da partcula na vizinhana e $Lbest - x_i(t)$ a melhor posio numa regio que engloba essa vizinhana.

3.2. Ant Colony Optimization - (ACO)

A otimizao por colnia de formigas ou *Ant Colony Optimization* - (ACO) uma tcnica de otimizao que foi introduzida desde os anos 90's [?] baseado no comportamento forrageiro de colnia de formigas. O comportamento forrageiro de diversas espcies [?] objeto de estudo das cincias biolgicas pois os animais predadores procuram otimizar seu ganho de protena, ao comer sua presa, minimizando o gasto de energia, ou minimizando o esforo para caar, capturar e comer essa presa. Esse comportamento explorado pelo ACO para buscar solues aproximadas para um problema de otimizao discreto, para problemas de otimizao contnuos e para problemas de roteamento em telecomunicaes.

No caminho da busca por alimentos as formigas deixam no ambiente uma marca chamado de feromnio. Esse feromnio evapora com o passar do tempo, sendo assim, a medida

que mais formigas sigam um determinado caminho, mais intenso o feromnio se far presente.

A equao da evaporaao do feromnio no ambiente segundo a frmula:

$$p(i, j) = \frac{[\tau(i, j)]^\alpha \cdot [\eta(i, j)]^\beta}{\sum [\tau(i, j)]^\alpha \cdot [\eta(i, j)]^\beta} \quad (3)$$

3.3. Fish School Search - (FSS)

Para contornar o problema explosivo combinatria Carmelo e Buarque propuseram a meta-heursticas da busca por cardume de peixes ou o *Fish School Search* - (FSS) [?].

Na busca FSS, cada peixe representa uma possvel soluo do problema ([?]). Em busca por enxame de partculas h o problema da degradaao do exame, quando aparentemente as partculas encontram um mnimo local (poderia ser mximo - depende da natureza do problema) "pensando" terem encontrado o mnimo global. Para contornar esse problema da degradaao do enxame introduz-se operadores que faam com que o exame saia desses "fossos" de busca. O FSS possui operadores para evitar o problema da perda de qualidade, dentre esses podemos citar o operador de volatilidade, que faz com que o enxame expanda quando o enxame se concentra por muito tempo.

A equao que faz isso a seguinte:

$$Bari(t) = \frac{\sum_{i=1}^N x_i(t) W_1(t)}{\sum_{i=1}^N x_i(t)} \quad (4)$$

3.3.1. Subsubsection Heading Here. Subsubsection text here.

4. Conclusion

The conclusion goes here.

Acknowledgments

The authors would like to thank...

References

- [1] H. Kopka and P. W. Daly, *A Guide to L^AT_EX*, 3rd ed. Harlow, England: Addison-Wesley, 1999.
- [2] H. Ahmed, and J. Glasgow, *Swarm intelligence: concepts, models and applications*, School of Computing, Queen's University, Citeseer, 2012.