

**Universidade de Pernambuco**  
**Programa de Pós-Graduação em Engenharia da Computação**  
**(PPGEC)**

**Proposta de Dissertação de Mestrado**

**Área: Inteligência Computacional**

**Título: Desenvolvimento de Novos Algoritmos de Swarm Intelligence para Data Mining em Big Data**

**Orientador – Carmelo José Albanez Bastos Filho ([carmelofilho@upe.br](mailto:carmelofilho@upe.br))**

**Co-orientador – Alexandre Magno A. Maciel ([amam@upe.br](mailto:amam@upe.br))**

**Descrição**

Um importante desafio na comunidade científica continua sendo a extração de informações relevantes a partir de gigantescas massas de dados. Essa imensa massa de dados provem da integração dos vários aplicativos de software via Internet e da interação desses aplicativos com seus usuários espalhados ao redor do mundo. Esse ecossistema entre usuários e aplicativos de software integrados em rede produz um grande volume de dados denominado genericamente de Big Data [1]. Sendo assim, um dos grandes desafios no contexto da Engenharia e da Ciência da Computação. Em particular, a Inteligência Artificial se destaca pela variedade de algoritmos capazes de extrair conhecimento útil de dados [1], em que a indústria possa, por exemplo, traçar perfis de compras para seus clientes, ou, até mesmo, perfis dos seus próprios usuários, e, assim fornecer/recomendar serviços/produtos mais personalizados.

Uma das principais tarefas utilizadas para solução desse tipo de problema é o Agrupamento de dados, que em inglês é chamado de Clustering [1] [2]. Clustering é o processo de agrupar dados em classes coerentes de tal maneira que dados em um mesmo grupo tenham máxima similaridade entre si, e, de modo que dados em grupos diferentes tenham alta dissimilaridade. Um dos mais conhecidos e conceituados algoritmos de Clustering são o K-means e o Fuzzy C-Means [1] [2] [3]. Contudo, esses dois algoritmos apresentam pontos negativos na sua aplicação. O K-means requer que o usuário conheça o número de grupos existentes na massa de dados, o que geralmente não é disponível. Além disso, o K-means possui dificuldade em separar massas de dados que apresentam sobreposição de características e são sensíveis a maneira em que é inicializado. O outro algoritmo, Fuzzy C-Means, não necessita que o usuário pré-determine o número de agrupamentos requeridos, no entanto, o mesmo necessita da determinação do grau de acurácia que o usuário deseja. Ademais, o Fuzzy C-Means não possui problemas graves na inicialização do sistema e na sensibilidade para inicialização. Uma outra forma de abordar este mesmo problema é modelá-lo como um problema de otimização, e usar algoritmos de Inteligência Computacional tais como a Otimização por Enxame de Partículas [3][4], a Colônia de Abelhas Artificiais [4], e o Algoritmo de Busca de Cardumes (FSS) [5]. Alguns trabalhos já foram desenvolvidos sobre tais algoritmos, porém, ainda não foi encontrado na literatura um algoritmo modelado para este problema que fosse inspirado no Algoritmo de Busca de Cardumes. O FSS [5] se torna interessante para este tipo de contexto pois possui uma capacidade superior de exploração em amplitude.

Este projeto de pesquisa tem como objetivo desenvolver um algoritmo inspirado nas qualidades desempenhadas pelo FSS para extração de conhecimento. Este projeto poderá impulsionar resultados interessantes para a indústria pois poderá acarretar em benefícios econômicos e sociais com a utilização do mesmo na extração de informações em base de dados.

**Referências Bibliográficas**

1. Steele, Brian, John Chandler, and Swarna Reddy. "Algorithms for Data Science." (2016).
2. Wu, Xindong, et al. "Top 10 algorithms in data mining." Knowledge and information systems 14.1 (2008): 1-37.
3. Cheng, Shi, et al. "Survey on data science with population-based algorithms." Big Data Analytics 1.1 (2016): 3.
4. Mernik, Marjan, et al. "On clarifying misconceptions when comparing variants of the Artificial Bee Colony Algorithm by offering a new implementation." Information Sciences 291 (2015): 115-127.
5. Bastos Filho, Carmelo JA, et al. "A novel search algorithm based on fish school behavior." Systems, Man and Cybernetics, 2008. IEEE International Conference on. IEEE, 2008.