



Universidade Federal de Pernambuco  
Centro de Informática

Graduação em Ciência da Computação

**Aprimoramento de um algoritmo de  
clustering através de programação em  
GPU**

Mário Henrique Santos da Silva

Proposta de Trabalho de Graduação

Orientador: Hansenclever de França Bassani

Recife  
13 de abril de 2017

# Resumo

À medida que aplicações computacionais geram um volume cada vez maior de dados, aumenta a demanda por algoritmos de aprendizagem de máquina para processar e extrair conhecimento desses dados. Para tornar possível a utilização em larga escala desses algoritmos, são utilizadas diversas técnicas e otimizações para melhorar sua eficiência. Uma das mais importantes é o uso de GPUs (Graphics Processing Units) para executar múltiplas operações em paralelo. O algoritmo precisa ter sua implementação adaptada para tirar proveito do ganho de desempenho provido pelo paralelismo em massa. Tal esforço de otimização tem sido mais voltado para algoritmos supervisionados, como os de deep learning, mas também seria benéfico para algoritmos não-supervisionados. Este trabalho tem o objetivo de reimplementar um algoritmo de agrupamento de dados utilizando processamento paralelo de GPU, e avaliar sob que condições se obtém um ganho de desempenho.

**Palavras-chave:** aprendizagem de máquina, aprendizagem não-supervisionada, agrupamento, gpu, mapas auto-organizáveis, otimização, torch

# Sumário

<b>1</b>	<b>Introdução</b>	<b>1</b>
<b>2</b>	<b>Objetivo</b>	<b>2</b>
<b>3</b>	<b>Cronograma</b>	<b>3</b>
<b>4</b>	<b>Possíveis avaliadores</b>	<b>4</b>
<b>5</b>	<b>Assinaturas</b>	<b>5</b>

## CAPÍTULO 1

# Introdução

Nos últimos anos a computação vem se tornando cada vez mais presente e importante em várias áreas da sociedade, como educação, comércio, entretenimento e comunicação. Junto com reduções do custo de tecnologias de armazenamento, esta tendência gera um volume cada vez maior de dados. Há uma demanda crescente de técnicas e algoritmos para processar e extrair conhecimento destes dados automaticamente e em larga escala, com o objetivo de aprimorar ou criar novas aplicações e serviços para satisfazer as necessidades dos usuários.

A aprendizagem de máquina é a área de pesquisa que contempla boa parte de tais técnicas e algoritmos[1]. Ela geralmente é dividida em dois paradigmas principais: aprendizagem supervisionada e não-supervisionada. A aprendizagem supervisionada tem o objetivo de aproximar funções específicas a partir de conjuntos de dados e respostas esperadas correspondentes, e generalizar essas funções para casos ainda não vistos[1]. Embora seja capaz de obter resultados poderosos, este paradigma tem a desvantagem de precisar que sejam fornecidas as respostas esperadas para todos os dados usados no treinamento, o que pode ser inviável para bases de dados muito grandes ou quando a função desejada não for fácil de resolver.

Algoritmos de aprendizagem não-supervisionada extraem informações e características do conjunto de dados sem nenhuma orientação específica ou conhecimento especializado quanto aos resultados desejados. Eles não tentam reproduzir ou aproximar resultados específicos para cada caso, mas podem utilizar métricas mais abstratas sobre o resultado geral[2]. Tais métricas dependem da tarefa do algoritmo: agrupar instâncias em grupos coerentes (clustering), estimar distribuições de densidade dos atributos dos dados, detectar quais atributos são mais importantes ou descritivos, ou detectar padrões anômalos ou divergentes do resto do conjunto.

Recentemente há um grande esforço da comunidade de aprendizagem de máquina para desenvolver implementações eficientes de algoritmos para viabilizar o uso deles em larga escala. Um dos avanços mais importantes é a utilização do processamento paralelo de GPUs (Graphics Processing Units) no treinamento e execução dos algoritmos. A paralelização dos cálculos realizados tem causado ganhos de desempenho significativos, principalmente em redes neurais artificiais. Esse esforço de otimização tem sido mais voltado para métodos supervisionados, utilizados em tarefas específicas. Métodos não-supervisionados não têm recebido tanta atenção nesse sentido, apesar de sua importância para a tarefa de inferir informações sobre conjuntos de dados.

## CAPÍTULO 2

# Objetivo

O objetivo deste trabalho é dar continuidade a um projeto anterior[3], que visava implementar algoritmos de agrupamento de dados com ganhos de desempenho ao aproveitar o processamento paralelo de GPUs. Esse trabalho lidou com uma extensão do Self-Organizing Map (SOM)[4], o Dimension Selective Self-Organizing Map (DSSOM)[5], e comparou os tempos de execução da implementação em GPU e da implementação original em CPU usando conjuntos de dados de diferentes dimensionalidades.

Neste trabalho será criada uma implementação voltada para GPUs de outra extensão do SOM, mais atual e de implementação mais complexa e que obteve resultados melhores que os do DSSOM: o Local Adaptive Receptive Field Dimension Selective Self-Organizing Map (LARFDSSOM)[6], e também será avaliado o ganho de desempenho em relação à implementação original em CPU. Também será avaliado como o tamanho, dimensionalidade ou características do conjunto de dados afetam o desempenho do algoritmo, e sob que condições o algoritmo original pode ser modificado para melhorar o desempenho sem prejudicar a qualidade dos resultados.

CAPÍTULO 3

# Cronograma

	Março				Abril				Maio			Junho				Julho	
Revisão bibliográfica	X	X	X	X	X	X	X	X									
Implementação dos algoritmos			X	X	X	X	X	X	X	X							
Avaliação da corretude							X	X	X	X							
Testes e experimentos									X	X	X	X	X	X			
Escrita do relatório											X	X	X	X	X		
Preparação da apresentação															X	X	

## CAPÍTULO 4

# Possíveis avaliadores

- Aluizio F. R. Araújo
- George D. C. Cavalcanti
- Tsang Ing Ren

CAPÍTULO 5

**Assinaturas**

---

Mário Henrique Santos da Silva  
Aluno

---

Hansenclever de França Bassani  
Orientador



## Referências Bibliográficas

- [1] T. M. Mitchell, *Machine Learning*. New York, NY, USA: McGraw-Hill, Inc., 1 ed., 1997.
- [2] E. Müller, S. Günemann, I. Assent, and T. Seidl, “Evaluating clustering in subspace projections of high dimensional data,” in *Proc. 35th International Conference on Very Large Data Bases (VLDB 2009)*, Lyon, France, *PVLDB Journal*, Vol. 2, No. 1., pp. 1270–1281, VLDB Endowment, 2009.
- [3] E. P. Macedo, “Agrupamento em subespaço com mapas auto-organizáveis em gpu,” Bachelor’s thesis, Universidade Federal de Pernambuco, Dec. 2016.
- [4] T. Kohonen, “The self-organizing map,” *Proceedings of the IEEE*, vol. 78, pp. 1464–1480, Sep 1990.
- [5] H. F. Bassani and A. F. R. Araújo, “Dimension selective self-organizing maps for clustering high dimensional data,” in *The 2012 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–8, June 2012.
- [6] H. F. Bassani and A. F. R. Araújo, “Dimension selective self-organizing maps with time-varying structure for subspace and projected clustering,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 26, pp. 458–471, March 2015.