



INTELIGÊNCIA COMPUTACIONAL NA AVALIAÇÃO DE CÓDIGOS EM UM SISTEMA COMPLEXO DE DETECÇÃO COM DESENVOLVIMENTO COLABORATIVO

Andressa A. Sivolella Gomes

Dissertação de Mestrado apresentada ao Programa de Pós-graduação em Engenharia Elétrica, COPPE, da Universidade Federal do Rio de Janeiro, como parte dos requisitos necessários à obtenção do título de Mestre em Engenharia Elétrica.

Orientador: José Manoel de Seixas

Rio de Janeiro
Março de 2016

INTELIGÊNCIA COMPUTACIONAL NA AVALIAÇÃO DE CÓDIGOS EM UM
SISTEMA COMPLEXO DE DETECÇÃO COM DESENVOLVIMENTO
COLABORATIVO

Andressa A. Sivolella Gomes

DISSERTAÇÃO SUBMETIDA AO CORPO DOCENTE DO INSTITUTO
ALBERTO LUIZ COIMBRA DE PÓS-GRADUAÇÃO E PESQUISA DE
ENGENHARIA (COPPE) DA UNIVERSIDADE FEDERAL DO RIO DE
JANEIRO COMO PARTE DOS REQUISITOS NECESSÁRIOS PARA A
OBTENÇÃO DO GRAU DE MESTRE EM CIÊNCIAS EM ENGENHARIA
ELÉTRICA.

Examinada por:

Prof. Aluizio Fausto Ribeiro Araújo, D.Sc.

Prof. Afonso de Bediaga e Hickman, D.Sc.

Pesquisadora Carmen Lúcia Lodi Maidantchik, D.Sc.

RIO DE JANEIRO, RJ – BRASIL
MARÇO DE 2016

Sivolella Gomes, Andressa A.

Inteligência Computacional na Avaliação de Códigos em um Sistema Complexo de Detecção com Desenvolvimento Colaborativo /Andressa A. Sivolella Gomes. – Rio de Janeiro: UFRJ/COPPE, 2016.

XI, 14 p.: il.; 29, 7cm.

Orientador: José Manoel de Seixas

Dissertação (mestrado) – UFRJ/COPPE/Programa de Engenharia Elétrica, 2016.

Referências Bibliográficas: p. 13 – 13.

1. Mineração de códigos. 2. Métodos ensemble com árvores. 3. Plataforma colaborativa. I. Manoel de Seixas, José. II. Universidade Federal do Rio de Janeiro, COPPE, Programa de Engenharia Elétrica. III. Título.

A todo mundo, geralção.

Agradecimentos

Gostaria de agradecer a todos.

Resumo da Dissertação apresentada à COPPE/UFRJ como parte dos requisitos necessários para a obtenção do grau de Mestre em Ciências (M.Sc.)

INTELIGÊNCIA COMPUTACIONAL NA AVALIAÇÃO DE CÓDIGOS EM UM
SISTEMA COMPLEXO DE DETECÇÃO COM DESENVOLVIMENTO
COLABORATIVO

Andressa A. Sivolella Gomes

Março/2016

Orientador: José Manoel de Seixas

Programa: Engenharia Elétrica

Apresenta-se, nesta tese, ...

Abstract of Dissertation presented to COPPE/UFRJ as a partial fulfillment of the requirements for the degree of Master of Science (M.Sc.)

COMPUTATIONAL INTELLIGENCE IN SOURCE CODE ASSERTION IN A
COMPLEX SYSTEM IN A COLLABORATIVE DEVELOPMENT
ENVIROMENT

Andressa A. Sivolella Gomes

March/2016

Advisor: José Manoel de Seixas

Department: Electrical Engineering

In this work, we present ...

Sumário

Lista de Figuras	x
Lista de Tabelas	xi
1 Introdução	1
1.1 Motivação	1
1.2 Objetivos	1
1.3 Organização do documento	1
2 A Colaboração ATLAS do CERN	2
2.1 CERN e LHC	2
2.2 Experimento ATLAS	3
2.3 Calorímetro Hadrônico de Telhas (TileCal)	4
2.3.1 Análise <i>Online</i> e <i>Offline</i>	6
2.3.2 A colaboração TileCal	6
3 Plataforma web Tile-in-ONE	8
3.1 Fluxo de dados	8
3.2 Novo desenvolvimento	8
4 Mineração de códigos fonte para identificação de falhas	9
4.1 Seleção de categorias	9
4.1.1 Analisadores Estáticos	9
4.1.2 Medidas estatísticas e de qualidade	9
4.2 Analisadores estáticos e Mineração de códigos no CERN	9
5 Metodologia	10
5.1 Aquisição de dados	10
5.2 Seleção de categorias	10
5.3 Métodos <i>ensemble</i> e Árvore de Decisão	10
5.4 Avaliação de classificadores	10
6 Resultados	11

7	Conclusões	12
	Referências Bibliográficas	13
A	Alguns casos	14

Lista de Figuras

2.1	Representação aérea do LHC e seus detectores no CERN. Extraído de [1].	3
2.2	Esquema do detector ATLAS. Adaptado de [2]	4
2.3	Calorímetro de Telhas do ATLAS (com suas 4 partições na cor verde) envolvendo o calorímetro de argônio líquido. O Sistema Magnético e a Câmara de Múons não estão ilustrados nesta figura. Adaptado de [3].	5
2.4	Estrutura de absorção e amostragem do TileCal. Extraído de [4] . . .	5
2.5	Esboço das células do TileCal. Extraído de [3].	6

Lista de Tabelas

Capítulo 1

Introdução

1.1 Motivação

1.2 Objetivos

1.3 Organização do documento

Capítulo 2

A Colaboração ATLAS do CERN

2.1 CERN e LHC

Fundado em 1954, o CERN (em francês *Centre Européen pour la Recherche Nucléaire* [5]) é o maior centro de pesquisa na área de física de partículas de altas energias no mundo. Atualmente, conta com a participação de 38 países membros e outros países colaboradores, entre eles o Brasil.

O LHC (em inglês *Large Hadron Collider* [6]) é o maior colisor de partículas já construído e encontra-se atualmente em operação no CERN. Instalado a 175 metros abaixo do solo, consiste em um grande túnel em formato anelar, com 27 km de perímetro. Quando partículas são eletricamente carregadas e submetidas a pulsos eletromagnéticos no interior de tubos tem-se o processo de aceleração de partículas. Neste processo, as partículas adquirem aceleração ao serem envolvidas a campos eletromagnéticos variantes [7]. Em aceleradores circulares, partículas são injetadas e circulam no anel até atingirem a energia desejada. Diversos experimentos podem ser realizados em determinados pontos de colisão ao longo da circunferência. O perímetro do acelerador circular é diretamente proporcional a energia necessária na colisão. Ao atingir a energia desejada, os feixes de partículas aceleradas em sentidos opostos colidem e, nesse caso, detectores em formato cilíndrico são os mais utilizados. Cada colisão gera informações que devem ser analisadas por especialistas, com objetivos distintos, tais como: identificação de sub-partículas, observação de fenômenos e a comprovação ou a eliminação de teorias físicas. Nos pontos de colisão, detectores são montados para observação e coleta de dados. Existem, por exemplo, detectores de calorimetria, para identificar a energia depositada pelas partículas colisionadas; detectores de traços, para observar trajetórias após colisões; e detectores de múons, responsáveis por absorver tais partículas.

A figura 2.1 ilustra uma representação aérea do LHC. Existem quatro detectores de partículas instalados em pontos de colisão ao longo da extensão do LHC, alta-

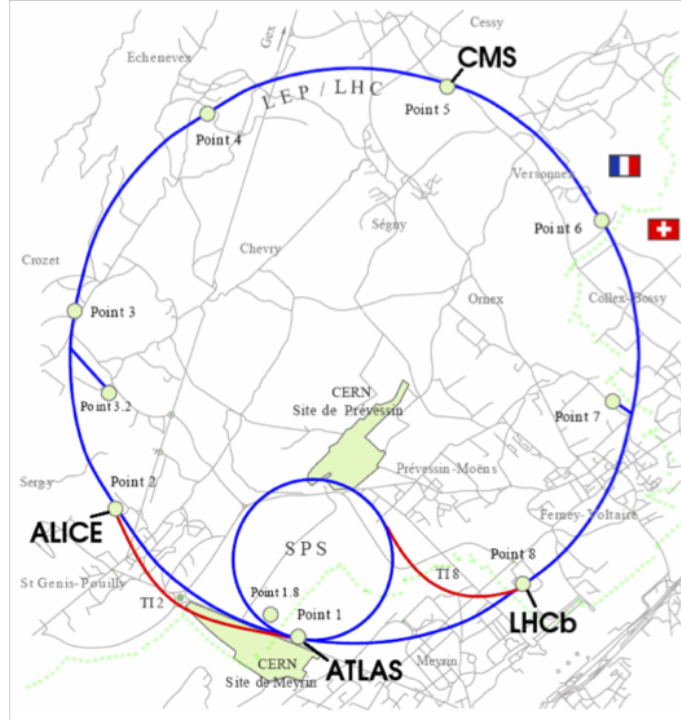


Figura 2.1: Representação aérea do LHC e seus detectores no CERN. Extraído de [1].

mente especializados: ALICE [8], ATLAS [3], CMS [9] e LHCb [10]. Os detectores ilustrados permitem obter informações detalhadas sobre a trajetória das partículas resultantes das colisões ocorridas e características energéticas, sendo possível identificar partículas. O ATLAS é o maior dos detectores do LHC.

2.2 Experimento ATLAS

O ATLAS (em inglês *A Toroidal LHC Apparatus*) [3] é um detector em formato cilíndrico. O detector pesa aproximadamente 7.000 toneladas, com 44 metros de comprimento por 24 metros de altura. A figura 2.2 ilustra os diversos subsistemas que compõem o detector ATLAS.

Cada subsistema apresenta uma característica específica de acordo com suas funcionalidades. São eles, da parte interna para a externa:

- Detector Interno (ID), responsável por registrar as trajetórias das partículas resultantes da colisão, bem como o momento e o sinal da carga, se for o caso;
- Sistema de Calorimetria, composto pelos calorímetros eletromagnético (LAr) e hadrônico de telhas (TileCal). Os calorímetros citados são responsáveis por medirem a energia das partículas resultantes;

- Sistema Magnético, responsável por desviar partículas carregadas para medir o momento;
- o Espectrômetro de Múons, responsável por medir a trajetória de uma determinada partícula denominada Múon.

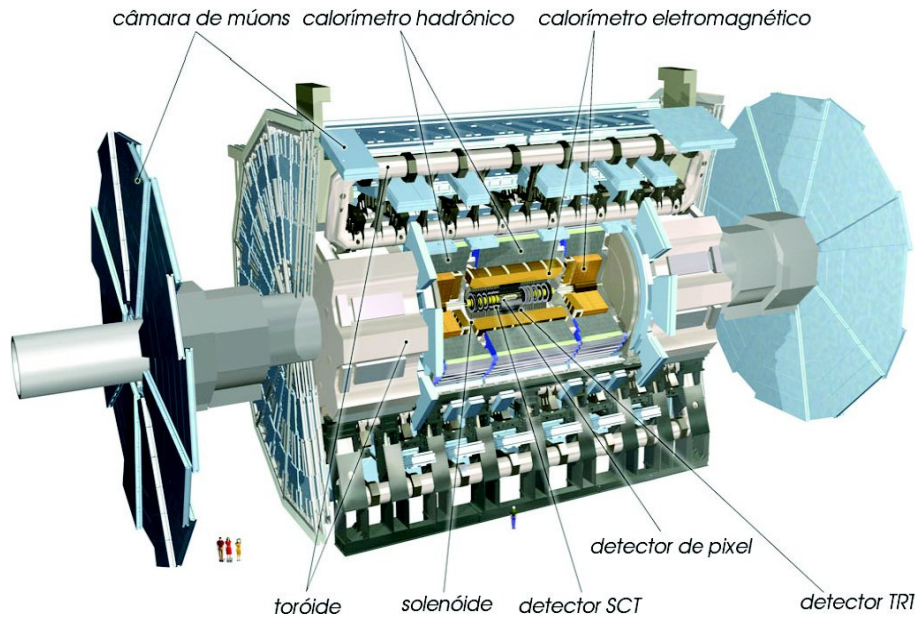


Figura 2.2: Esquema do detector ATLAS. Adaptado de [2]

O sistema de calorimetria do LHC foi projetado para absorver a energia das partículas que cruzam o detector, sendo o calorímetro hadrônico de telhas (em inglês *Tile Calorimeter* ou TileCal) é o foco deste mestrado.

2.3 Calorímetro Hadrônico de Telhas (TileCal)

O TileCal [4] é composto por três barris: um central (que se divide em duas partições: LBA e LBC) e dois estendidos (cada um correspondendo a mais duas partições, EBA e EBC, respectivamente), como ilustrado na figura 2.3. Cada partição é igualmente dividida em 64 partes (módulos). Os módulos do TileCal são compostos de ferro como mediadores passivos e telhas cintilantes como material ativo. A estrutura absorvedora é coberta de placas de aço de várias dimensões, conectadas a uma estrutura maciça de suporte. A principal inovação deste detector é a posição perpendicular das telhas em relação aos feixes do LHC.

A passagem de partículas ionizadas pelo TileCal produz luz no espectro. Sua intensidade é proporcional à energia depositada pela partícula. A luz produzida se propaga através das telhas para suas bordas onde é absorvida por fibras óticas e deslocada, por reflexão interna total, até os canais de leitura eletrônica ou fotomultiplicadores (em inglês *PhotoMultipliers Tubes* ou PMTs), onde é convertido em

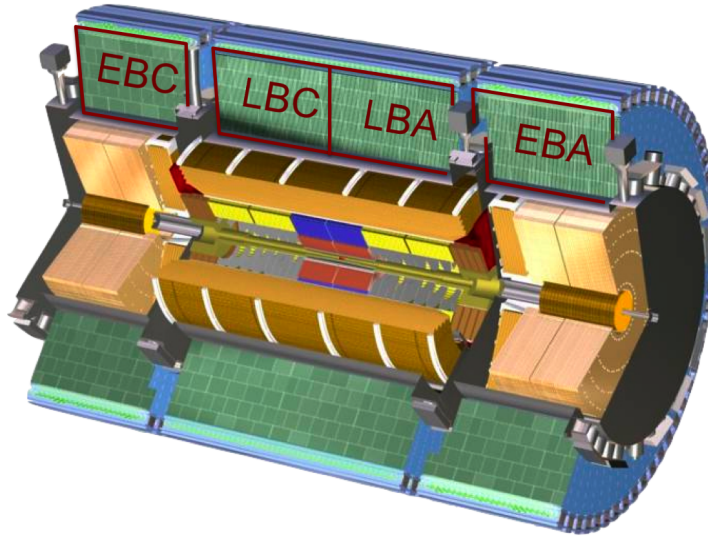


Figura 2.3: Calorímetro de Telhas do ATLAS (com suas 4 partições na cor verde) envolvendo o calorímetro de argônio líquido. O Sistema Magnético e a Câmara de Múons não estão ilustrados nesta figura. Adaptado de [3].

um sinal de corrente. Um PMT pode ser idealizado como um pulso de corrente. A figura 2.4 ilustra a estrutura de absorção e amostragem do TileCal sendo possível identificar tubos fotomultiplicadores.

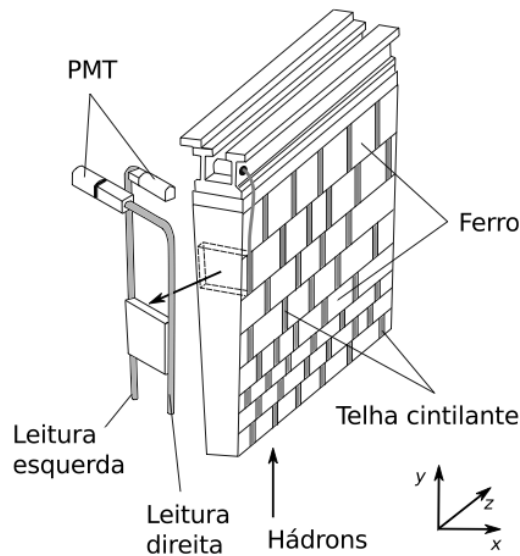


Figura 2.4: Estrutura de absorção e amostragem do TileCal. Extraído de [4]

Os módulos do barril central do detector comportam 45 canais cada, enquanto os módulos dos barris estendidos comportam 32 canais. Radialmente, cada módulo do TileCal ainda é segmentado em três camadas (A, BC e D), conhecidas como células. Estas são formadas pelo agrupamento de fibras até cada PMT. Um esboço da geometria das células é apresentado na figura 2.5. Cada módulo contém 11 linhas

transversas de telhas.

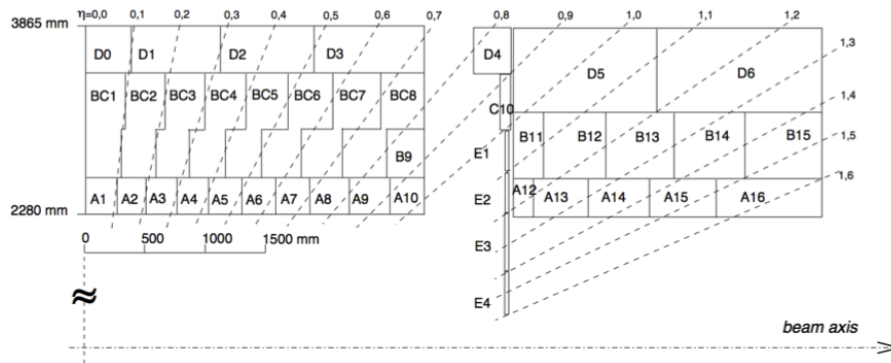


Figura 2.5: Esboço das células do TileCal. Extraído de [3].

2.3.1 Análise *Online* e *Offline*

Existem dois tipos de análises, comuns a todos os experimentos do LHC. O experimento ATLAS, sozinho, gera um volume de dados na ordem de 3,2 PetaBytes ao ano [11], que precisam ser processados e analisados. Nesse caso, é impraticável armazenar os dados adquiridos pelo experimento em sua totalidade para posterior análise. Por esse motivo, análises referentes a atividades anteriores ao armazenamento em mídia permanente são definidas como *online*. Já as análises que ocorrem a partir do acesso de informações em bases de dados são definidas como sendo *offline*.

As soluções abordadas neste mestrado estão contextualizadas no processo de análise *offline* do TileCal.

As análises *online* e *offline* serão descritas detalhadamente na seção ??.

2.3.2 A colaboração TileCal

O TileCal é composto ainda por diferentes grupos, responsáveis por atividades específicas no experimento. Cada grupo desempenha uma etapa no processo de trabalho do calorímetro. São eles:

DAQ (do inglês, *Data Acquisition* ou aquisição de dados) responsável pela aquisição dos sinais resultantes das interações físicas no detector;

DCS (do inglês, *Detector Control System*, ou sistema de controle do detector) responsável pela manutenção das fontes de alta e baixa tensão do TileCal, placas mãe e sistema de refrigeração;

DQ (do inglês, *Data Quality*, ou qualidade de dados) responsável por avaliar o funcionamento do detector garantindo confiabilidade nos dados adquiridos que serão analisados;

Calibração (do inglês, *Calibration*) responsável pela calibração das células do calorímetro e validação da escala eletromagnética, a fim de corrigir desvios que ocorrem ao longo do tempo devido a desgastes ocorridos pela exposição do detector a altos níveis de radiação;

Computação e Software responsável por realizar a organização e desenvolvimento da simulação, reconstrução, digitalização e desenvolvimento de filtros de alto nível;

Dados e Processamento (do inglês, *Data and Processing*) responsável por coletar informações dos grupos de qualidade de dados e calibração a fim de decidir que alterações nos dados de condições serão realizadas. Responsável também por analisar os dados físicos.

Existem, portanto, seis grupos no TileCal onde o compartilhamento de informações entre eles faz-se necessário: se um PMT não está sendo corretamente alimentado, a qualidade dos dados adquiridos pode ficar comprometida. Tal situação exige o compartilhamento de informações do grupo e DQ, por exemplo.

Capítulo 3

Plataforma web Tile-in-ONE

3.1 Fluxo de dados

3.2 Novo desenvolvimento

Capítulo 4

Mineração de códigos fonte para identificação de falhas

4.1 Seleção de categorias

4.1.1 Analisadores Estáticos

4.1.2 Medidas estatísticas e de qualidade

4.2 Analisadores estáticos e Mineração de códigos no CERN

Capítulo 5

Metodologia

5.1 Aquisição de dados

5.2 Seleção de categorias

5.3 Métodos *ensemble* e Árvores de Decisão

5.4 Avaliação de classificadores

Capítulo 6

Resultados

Capítulo 7

Conclusões

Referências Bibliográficas

- [1] “CDS - CERN Document Server”,
Acessado em janeiro de 2016. cds.cern.ch.
- [2] “ATLAS Photos”,
Acessado em janeiro de 2016. <http://www.atlas.ch/photos/>.
- [3] ATLAS COLLABORATION, *ATLAS: Technical Proposal for a General-Purpose pp Experiment at the Large Hadron Collider at CERN*, Tech. rep., CERN, 1994, CERN/LHCC 94–43.
- [4] ATLAS/TILE CALORIMETER COLLABORATION, *Tile Calorimeter Calorimeter Technical Design Report*, Tech. rep., CERN, 1996, CERN/LHCC 96–42.
- [5] “The European Laboratory for Particle Physics”,
Acessado em janeiro de 2016. <http://www.cern.ch>.
- [6] “The Large Hadron Collider”,
Acessado em janeiro de 2016. <http://lhc.web.cern.ch>.
- [7] GRIFFITHS, D., *Introduction to Elementary Particles*. John Wiley & Sons, 1987.
- [8] THE ALICE COLLABORATION, “The ALICE experiment at the CERN LHC”, *Journal of Instrumentation*, v. 3, n. 08, Aug. 2008.
- [9] THE CMS COLLABORATION, “The CMS experiment at the CERN LHC”, *Journal of Instrumentation*, v. 3, n. 08, Ago 2008.
- [10] SZUMLAK, T., “The LHCb experiment”, *Acta Physica Polonica. Series B: Elementary Particle Physics, Nuclear Physics, Statistical Physics, Theory of Relativity, Field Theory*, v. 41, n. 7, pp. 1661–1668, 2010.
- [11] “ATLAS Factsheet”,
Acessado em janeiro de 2016.
http://www.atlas.ch/pdf/fact_sheet_1page.pdf.

Apêndice A

Alguns casos