



COPPE/UFRJ

**ANÁLISE DE COMPONENTES INDEPENDENTES PARA UMA FILTRAGEM
ONLINE BASEADA EM CALORIMETRIA DE ALTA ENERGIA E COM FINA
SEGMENTAÇÃO**

Eduardo Furtado de Simas Filho.

Tese de Doutorado 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 Doutor
em Engenharia Elétrica.

Orientadores: José Manoel de Seixas
Luiz Pereira Calôba

Rio de Janeiro
Junho de 2010

ANÁLISE DE COMPONENTES INDEPENDENTES PARA UMA FILTRAGEM
ONLINE BASEADA EM CALORIMETRIA DE ALTA ENERGIA E COM FINA
SEGMENTAÇÃO

Eduardo Furtado de Simas Filho.

TESE 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 DOUTOR
EM CIÊNCIAS EM ENGENHARIA ELÉTRICA.

Examinada por:

Prof. José Manoel de Seixas, D.Sc.

Prof. Luiz Pereira Calôba, Dr.Ing.

Prof. Luiz Wagner Pereira Biscainho, D.Sc.

Prof. Marley M. B. Rebuzzi Vellasco, Ph.D.

Prof. Leandro Salazar de Paula, D.Sc.

Prof. André Ponce de Leon F. de Carvalho, Ph.D

RIO DE JANEIRO, RJ – BRASIL

JUNHO DE 2010

Simas Filho., Eduardo Furtado de
Análise de Componentes Independentes para uma
Filtragem Online Baseada em Calorimetria de Alta Energia
e com Fina Segmentação/Eduardo Furtado de Simas
Filho.. – Rio de Janeiro: UFRJ/COPPE, 2010.

XVIII, 169 p.: il.; 29, 7cm.

Orientadores: José Manoel de Seixas

Luiz Pereira Calôba

Tese (doutorado) – UFRJ/COPPE/Programa de
Engenharia Elétrica, 2010.

Referências Bibliográficas: p. 105 – 127.

1. Análise de Componentes Independentes. 2.
Filtragem Online. 3. Calorimetria de Alta Energia. I.
Seixas, José Manoel de *et al.*. II. Universidade Federal do
Rio de Janeiro, COPPE, Programa de Engenharia Elétrica.
III. Título.

Aos meus amores Cléa e Letícia.

Agradecimentos

Agradeço ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pelo suporte financeiro concedido em parte do tempo no qual este trabalho foi desenvolvido.

Sou profundamente grato ao apoio, motivação e conhecimentos que me foram passados por meus orientadores, os professores José Manoel de Seixas e Luiz Pereira Calôba, sem os quais este trabalho não existiria.

Agradeço aos colegas, professores e funcionários do Laboratório de Processamento de Sinais (LPS/COPPE/UFRJ) pela inestimável ajuda na condução do trabalho, em especial aos amigos Natanael Nunes de Moura, Rodrigo Coura Torres e Danilo Enoque Ferreira.

Também gostaria expressar minha gratidão ao Instituto Federal de Educação, Ciência e Tecnologia da Bahia, pelo apoio concedido para a finalização deste trabalho, em especial aos professores e funcionários do Campus Simões Filho.

Finalmente preciso agradecer à minha família (especialmente a meus pais e avós) e meus amigos (do Roger) por suportarem minha ausência durante este período e por me incentivarem sempre a continuar. Não poderia deixar de agradecer aos amigos Frederico Wegelin, pela companhia e apoio durante o tempo em que morei no Rio de Janeiro, e Edson Pinto Santana, pelo constante incentivo (mesmo que à distância).

Em especial, dedico esse trabalho a minha esposa Cléa e minha filha Letícia pelo amor e apoio em todos os momentos.

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

ANÁLISE DE COMPONENTES INDEPENDENTES PARA UMA FILTRAGEM
ONLINE BASEADA EM CALORIMETRIA DE ALTA ENERGIA E COM FINA
SEGMENTAÇÃO

Eduardo Furtado de Simas Filho.

Junho/2010

Orientadores: José Manoel de Seixas
Luiz Pereira Calôba

Programa: Engenharia Elétrica

O ATLAS é o maior detector do acelerador de partículas LHC. Nas colisões, uma enorme quantidade de informação é produzida, porém, apenas uma pequena parcela é importante para a caracterização dos fenômenos físicos de interesse, tornando necessário aos detectores um eficiente sistema para detecção (filtragem) *online* de eventos. Os elétrons são partículas extremamente importantes para o experimento, e aparecem mascarados por um intenso ruído de fundo composto de jatos hadrônicos, pois esses últimos podem apresentar perfil de deposição de energia nos calorímetros semelhante ao de elétrons (os calorímetros são medidores de energia com fina segmentação e no ATLAS são divididos em sete camadas, totalizando mais de 100.000 sensores). Neste trabalho, é proposta a utilização do modelo não-linear da análise de componentes independentes no processo de extração de características, visando otimizar o desempenho do sistema neural de filtragem *online* de elétrons no ATLAS (*Neural Ringer*). Para explorar toda a segmentação e granularidade disponíveis, a extração de características foi efetuada a nível de cada camada do calorímetro. Diversos algoritmos foram utilizados na estimativa dos componentes independentes. Através da abordagem proposta, foi possível alcançar alta eficiência de discriminação, gerando dados mais limpos para a análise *offline*.

Abstract of Thesis presented to COPPE/UFRJ as a partial fulfillment of the requirements for the degree of Doctor of Science (D.Sc.)

INDEPENDENT COMPONENT ANALYSIS FOR ONLINE FILTERING
BASED ON HIGH-ENERGY AND HIGHLY SEGMENTED CALORIMETRY

Eduardo Furtado de Simas Filho.

June/2010

Advisors: José Manoel de Seixas

Luiz Pereira Calôba

Department: Electrical Engineering

Here goes the abstract ...

Sumário

Lista de Figuras	xii
Lista de Tabelas	xv
Símbolos e Abreviaturas	xvi
1 Introdução	1
1.1 Contexto	1
1.2 Motivação	2
1.3 Trabalhos Anteriores	4
1.4 Objetivos	5
1.5 Metodologia	6
1.6 Conteúdo do trabalho	7
2 Física de Partículas e o detector ATLAS	8
2.1 Panorama geral da física de partículas elementares	8
2.2 O acelerador LHC	13
2.3 Características gerais do detector ATLAS	15
2.4 Principais objetos de interesse no ATLAS	19
2.5 O sistema de calorimetria do ATLAS	21
2.5.1 Breve introdução à calorimetria	21
2.5.2 Características dos calorimetros do ATLAS	23
2.5.3 Desempenho esperado dos calorímetros	27
3 Filtragem Online no ATLAS	30
3.1 Introdução aos sistemas de filtragem em HEP	30
3.1.1 Aplicações de redes neurais artificiais	32

3.2	O sistema de filtragem <i>online</i> do ATLAS	36
3.2.1	Primeiro nível de filtragem	39
3.2.2	Filtragem de alto nível	42
3.2.3	Plataforma de <i>Software</i> do Sistema de Filtragem	45
4	Deteção de Elétrons a partir de Informações de Calorimetria no ATLAS	47
4.1	Filtragem de elétrons no L1	48
4.2	Filtragem de elétrons no L2 - Algoritmo T2Calo	50
4.3	Neural Ringer - Alternativa para Filtragem de Elétrons no L2	53
4.3.1	Extração de Características - Anelamento	53
4.3.2	Normalização	56
4.3.3	Teste de Hipóteses - Classificador Neural	58
4.3.4	Tempo de Execução	59
4.4	Extensões ao Neural Ringer	60
4.4.1	Importância do pré-processamento para classificadores neurais	61
4.4.2	Pré-processamento linear	62
4.4.3	Pré-processamento não-linear	63
5	Análise de Componentes Independentes	64
5.1	Modelo linear da ICA	64
5.2	ICA não-linear	67
5.2.1	Unicidade da Solução em NLICA	70
5.2.2	Modelos com restrições estruturais	70
5.2.3	Algoritmos sem restrições estruturais	73
5.2.4	ICA Local	77
5.3	Aplicações de ICA e NLICA para extração de características	78
5.4	Aplicações em Física de Altas Energias e Áreas Correlatas	81
5.5	Utilizando Informação das Classes na Estimação dos Componentes Independentes	83
5.5.1	Componentes Principais de Discriminação	84
5.5.2	Utilizando os rótulos de classe como sinais de entrada para os algoritmos de ICA	85

5.5.3	Componentes Independentes para Cada Classe	87
5.5.4	Proposta de algoritmo para estimação de componentes independentes e discriminantes	88
6	Metodologia Proposta	91
6.1	Parâmetros de Avaliação do Desempenho	91
6.2	Extração de Características	92
6.2.1	Características Globais	92
6.2.2	Características Segmentadas	92
6.3	Classificadores Neurais Supervisionados	93
6.3.1	Especificações dos Classificadores Neurais Utilizados	93
6.3.2	Classificador Global	94
6.3.3	Classificadores Segmentados	94
6.3.4	Combinação de Múltiplos Classificadores	95
7	Resultados - Dados Simulados	97
7.1	Características dos Sinais simulados	98
8	Resultados - Dados Experimentais	101
8.1	Validação com Sinais de Ráios Cósmicos	101
9	Conclusões	104
	Referências Bibliográficas	105
	A Aspectos Teóricos das Técnicas de Extração de Características	128
A.1	Mapas auto-organizáveis	128
A.1.1	Quantização vetorial por aprendizado	130
A.1.2	Classificação a partir do mapa de características	132
A.2	Análise de Componentes Principais	132
A.3	Técnicas de Pré-Processamento - Compactação	132
A.3.1	Análise de Componentes Principais	132
A.3.2	Redução de dimensão	134
A.4	Análise de Componentes Independentes	135
A.4.1	Princípios de estimação dos componentes independentes . . .	135

A.4.2	Pré-processamento dos sinais para ICA	139
A.4.3	Principais algoritmos para ICA	139
A.4.4	Extensões ao modelo básico de ICA	143
A.5	ICA não-linear	145
A.5.1	Algoritmo Taleb-Jutten para o modelo PNL	145
B	Conceitos fundamentais em classificação de sinais	147
B.1	Teste de Hipóteses	147
B.2	Critério de Bayes	148
B.3	Filtros Casados	149
B.4	Discriminante linear de Fisher	151
B.5	Classificadores Neurais	151
C	Algoritmos Genéticos	154
C.1	Algoritmo Genético como Método de Otimização	154
C.2	Estrutura de um Algoritmo Genético	155
C.2.1	Conceitos Principais	155
C.2.2	Escalonamento de Aptidão	157
C.2.3	Implementação de um Algoritmo Genético	159
C.2.4	O Algoritmo Genético Utilizado	159
D	Lista de Publicações	162
D.1	Publicações em periódicos	162
D.2	Capítulo de Livro	164
D.3	Artigos em Conferências	164
D.4	Resumos em Conferências	168

Listas de Figuras

2.1	Diagrama do Modelo Padrão, mostrando as partículas elementares incluindo o, ainda não confirmado, bóson de Higgs.	10
2.2	Mapa de localização dos detectores do LHC.	14
2.3	Diagrama esquemático do ATLAS.	16
2.4	Eixo de coordenadas do ATLAS.	16
2.5	Cortes (a) transversal e (b) axial do ATLAS.	18
2.6	Disposição em camadas dos calorímetros do ATLAS.	24
2.7	Granularidade e profundidade das camadas do calorímetro eletromagnético.	26
2.8	Energia perdida por elétrons antes do calorímetro.	27
2.9	Energia normalizada média depositada nas secções eletromagnética e hadrônica em função de η para elétrons.	28
2.10	Erro relativo do calorímetro na medição da energia de elétrons para diferentes valores de (a) energia e (b) η .	29
3.1	Diagrama em blocos do sistema de filtragem genérico para um experimento de HEP.	31
3.2	Diagrama em blocos do sistema de filtragem e aquisição de dados do ATLAS.	37
3.3	Diagrama em blocos do primeiro nível de filtragem.	41
3.4	Diagrama em blocos do segundo nível de filtragem.	44
4.1	Janela deslizante analisada pelo L1 no calorímetro eletromagnético.	48
4.2	Processo de identificação de elétrons no L2.	50
4.3	Distribuição de elétrons e jatos quanto à razão $\frac{E_{HAD}}{E_{EM}}$ para eventos com energia total (a) menor que 20 GeV e (b) maior que 60 GeV.	52

4.4	Distribuição de elétrons e jatos quanto à razão de forma ($R_{shape} = \frac{E_{3\times 7}}{E_{7\times 7}}$) para eventos com energia total (a) menor que 20 GeV e (b) maior que 60 GeV.	52
4.5	Fluxo de processamento do <i>Neural Ringer</i>	53
4.6	Diagrama do processo de construção dos anéis.	55
4.7	Sinais típicos medidos na camada EM2 (acima) e os respectivos sinais em anéis (abaixo), respectivamente para um elétron (esquerda) e um jato (direita).	56
4.8	Sinais em anéis para (a) elétron típico, (b) jato típico e (c) jato com perfil semelhante ao de elétrons.	57
4.9	Fluxo de processamento das extensões ao <i>Neural Ringer</i>	61
5.1	Diagrama do <i>cocktail party problem</i>	67
5.2	Sinais (a) fonte, (b) observados e (c) recuperados através da ICA.	68
5.3	Diagrama do modelo de mistura PNL.	71
5.4	Diagrama do modelo PNL-L.	72
5.5	Diagrama do modelo da Mono não-linearidade.	73
5.6	NLICA a partir de SOM.	73
5.7	Diagrama do algoritmo MISEP.	76
5.8	Diagrama do modelo da ICA local.	77
5.9	Modelos neurais para estimar (a) a primeira e (b) a k-ésima PCD.	85
5.10	Diagramas de (a) treinamento e (b) operação dos algoritmos de ICA/NLICA utilizando informação das classes na entrada.	85
5.11	Modelo da ICA que utiliza a informação das classes em conjunto com os componentes estimados na função custo.	86
5.12	Sistema de classificação baseado em modelos da ICA estimados para cada classe.	87
5.13	Modelo Pós Não-linear modificado.	88
5.14	Procedimento de treinamento para o modelo pós não-linear modificado.	89
6.1	Processo de extração de características globais.	92
6.2	Processo de extração de características segmentadas.	93
6.3	Decisão utilizando classificador global.	94

6.4	Decisão utilizando classificadores segmentados.	95
7.1	Distribuição em energia, η e ϕ das assinaturas de elétrons (esquerda) e jatos (direita) aprovadas pelos cortes de primeiro nível tipo E10 e E15i.	99
7.2	Sinais em anéis (média e desvio padrão) para elétrons (esquerda) e jatos (direita) que chegam ao L2 no corte E10.	100
7.3	Sinais em anéis (média e desvio padrão) para elétrons (esquerda) e jatos (direita) que chegam ao L2 no corte E15.	100
8.1	Visualização de um evento do LHC (7 GeV) onde foram produzidos um elétron e três jatos.	102
8.2	Visualização de um evento do LHC (7 GeV) onde foram produzidos dois múons a partir do decaimento de um bóson Z.	103
A.1	Diagrama de um mapa auto-organizável	128
A.2	Diagrama da classificação a partir do mapa de características	132
A.3	Compressão e recuperação do sinal \mathbf{x} utilizando a transformação por PCA.	134
A.4	Diagrama do algoritmo de Taleb-Jutten para o modelo PNL.	146
B.1	Esquemático do problema de classificação binário.	147
B.2	Diagrama em blocos de um filtro casado.	150
B.3	Exemplo de uma rede neural utilizada para separação dos sinais de entrada em 3 classes.	152
C.1	Exemplo de um cromossomo binário	156
C.2	(a) Recombinação em ponto único e (b) recombinação uniforme. . . .	158
C.3	Fluxo de um Algoritmo Genético	160

Listas de Tabelas

2.1	Intensidade relativa (em comparação com a interação forte) dos diversos tipos de interação.	10
2.2	Exemplos de valores da massa (em energia equivalente) de algumas partículas.	12
2.3	Principais objetos de interesse no ATLAS.	20
2.4	Região de cobertura em η , granularidade e número de canais de leitura das camadas dos calorímetros.	25
3.1	Principais características do sistema de <i>trigger</i> do ATLAS, onde Te e Ts são, respectivamente, as taxas de eventos na entrada e na saída e $Cr = Te/Ts$ é o coeficiente de redução de eventos.	39
3.2	Freqüência esperada para os principais canais de <i>trigger</i> no primeiro nível de filtragem do ATLAS ($L=10^{34}\text{cm}^{-2}s^{-1}$).	41
4.1	Tempo gasto nas diversas etapas necessárias para a execução <i>online</i> do <i>Neural Ringer</i>	60

Símbolos e Abreviaturas

Símbolos

c	Velocidade da Luz no Vácuo
E	Energia
eV	Elétron-Volt
$\mathcal{E}(.)$	Operador esperança
Ef_i	Eficiência de discriminação da classe i
E_T	Energia transversa
E_T^{miss}	Energia transversa faltante
$H(.)$	Entropia
$I(.)$	Informação mútua
$J(.)$	Negentropia
$kurt(.)$	Curtose
m	Massa
p_T	Momento transverso
$p_x(x)$	Função de densidade de probabilidade
η	Pseudo-rapidez (sistema de coordenadas do ATLAS)
θ	Ângulo polar (sistema de coordenadas do ATLAS)
$\lambda(.)$	Razão de semelhança
ϕ	Ângulo azimutal (sistema de coordenadas do ATLAS)
e^-/j	Elétron/Jato
\mathbf{s}	Vetor das fontes independentes
\mathbf{x}	Vetor dos sinais observados (medidos)
\mathbf{y}	Vetor dos componentes independentes estimados
\mathbf{z}	Vetor dos sinais branqueados

Abreviaturas

No caso de algumas abreviaturas internacionalmente conhecidas, optou-se por mantê-las em inglês.

ALICE	Detector do LHC
ATLAS	<i>A Toroidal LHC Aparattus</i>
BOOSTER	Experimento do laboratório Fermilab
CERN	Centro Europeu Para Pesquisa Nuclear
CHOOZ	Experimento de Física de Alta Energia
CMB	<i>Cosmic Microwave Background</i>
CMS	Detector do LHC
CP	<i>Charge parity</i>
CPLD	<i>Complex Programmable Logic Device</i>
DESY	<i>Deutsches Elektronen-Synchrotron</i>
EF	Filtro de Eventos (<i>Event Filter</i>)
EM	Eletromagnético
FastICA	Algoritmo de extração das componentes independentes
FPGA	<i>Field-Programmable Gate Array</i>
Fermilab	<i>Fermi National Accelerator Laboratory</i>
GTM	<i>Generative topographic mapping</i>
H1	Detector do acelerador HERA
HLT	<i>High Level Trigger</i>
HEP	<i>High-Energy Physics</i>
HERA	Experimento do laboratório DESY
HEGRA	Experimento de astrofísica de alta energia
IC	Componentes independentes
ICA	<i>Independent components analysis</i>
JADE	<i>Joint Approximate Diagonalization of Eigen-matrices</i>
LEAR	Experimento do CERN (desativado)
LEP	Experimento do CERN (desativado)

LHC	<i>Large Hadron Collider</i>
LVL1	Primeiro nível de filtragem do ATLAS
LVL2	Segundo nível de filtragem do ATLAS
LVQ	Quantização vetorial por aprendizado
MLP	Perceptron de múltiplas camadas
NLICA	<i>Non-linear independent component analysis</i>
NLPCA	<i>Non-linear principal component analysis</i>
PC	Computador pessoal
PCA	<i>Principal components analysis</i>
PNL	<i>Post Nonlinear</i>
PCD	<i>Principal Components for Discrimination</i>
OGLE	Telescópio instalado no Chile
QV	Quantização vetorial
RNA	Rede Neural Artificial
ROC	<i>Receiver operating characteristic</i>
RoI	Região de interesse (<i>Region of Interest</i>)
RPROP	<i>Resilient back-propagation</i>
SICA	<i>Segmented independent component analysis</i>
SPCA	<i>Segmented principal component analysis</i>
SM	Modelo Padrão de interação entre as partículas elementares (<i>Standard Model</i>)
SOM	Mapa auto-organizável
SP	Figura de mérito de desempenho de classificação
SUSY	Supersimetria
T2Calo	Algoritmo de discriminação de elétrons no L2 do ATLAS
TOTEM	Experimento do LHC
TT	Torre de <i>trigger</i>

Capítulo 1

Introdução

O processamento estatístico de sinais encontra aplicações nas diversas áreas do conhecimento, desde medicina e saúde pública até as bolsas de valores. Seu uso pode simplificar as tarefas de análise de dados e classificação, pois permite mapear o conjunto de sinais em um espaço onde sua estrutura fundamental está mais acessível.

Este trabalho descreve a aplicação das técnicas de processamento estatístico no sistema de filtragem (detecção) *online* de um detector de partículas elementares de altas energias. O objetivo é extrair características relevantes para guiar o processo de identificação das partículas.

1.1 Contexto

Com os constantes avanços tecnológicos dos sistemas eletrônicos de aquisição de informações, é crescente a necessidade de técnicas eficientes para o processamento *online* de sinais. As grandezas físicas são registradas por elementos sensores, que podem ser únicos (como na medição da velocidade de um motor), ou combinados aos milhares para obter o resultado final (como na captura de vídeo e imagem digitais).

Em aplicações onde a fina granularidade da informação é necessária para descrever adequadamente o processo físico em questão, e sinais com alta dimensão são assim gerados por sistemas de medição compostos por um número elevado de sensores, o custo computacional geralmente é alto. Em alguns casos, a informação disponível pode estar segmentada, pois foi produzida a partir de conjuntos de sensores com características distintas.

Se há a necessidade de uma resposta rápida, pode ser utilizada a combinação de técnicas de compactação de sinais e processamento distribuído. O cenário pode ficar ainda mais complicado quando o volume de dados é alto e o problema a ser resolvido apresenta elevado grau de complexidade.

O ambiente de aplicação deste trabalho é o sistema *online* de filtragem do ATLAS (*A Toroidal LHC Aparatus*) [1], maior detector de propósito geral do acelerador de partículas LHC (*Large Hadron Collider*) [2]. O LHC entrou em operação no final de 2008, logo depois, passou por reparos no sistema de resfriamento de um dos seus supercondutores, voltando a operar no final de 2009.

No LHC, os sinais de interesse são raros, estão imersos em um intenso ruído de fundo e a perda de um desses eventos compromete severamente o desempenho dos detectores. Neste caso, é necessária uma estratégia de filtragem capaz de remover, ou pelo menos atenuar a intensidade do ruído sem perder os eventos de interesse.

Combinado a isso, a taxa de ocorrência de eventos é extremamente elevada, fazendo com que o intervalo entre eventos consecutivos seja extremamente pequeno. Considerando ainda que os detectores são altamente segmentados e apresentam fina granularidade de células detectoras, a quantidade de informação produzida é enorme. Neste contexto, a seleção de eventos precisa ser realizada de modo *online* e sob severas restrições no tempo de processamento.

Técnicas que utilizam informações da estatística dos sinais, como Análise de Componentes Principais (PCA - *Principal Component Analysis*) [3], Análise de Componentes Independentes ICA (*Independent Component Analysis*) [4] e Redes Neurais Artificiais (RNA) [5], são frequentemente utilizadas na solução de problemas onde há a necessidade de processamento veloz, flexível e eficiente.

1.2 Motivação

Desde o final do século 19, quando foi descoberto o elétron, o estudo da física de partículas elementares de altas energias, ou simplesmente física de altas energias (HEP *High-Energy Physics*), teve um crescimento acentuado. Na década de 1950, com o uso dos aceleradores, foram descobertas centenas de novas partículas. A física de partículas de altas energias pretende encontrar os componentes fundamentais da

matéria e descrever suas formas de interação.

O LHC [2] é o maior e mais potente acelerador de partículas da atualidade e está em operação no CERN (Organização Européia para Pesquisa Nuclear) desde 2008. Ao operar na máxima capacidade, produzirá uma taxa de colisões que chegará a 40MHz. Entretanto, as assinaturas de interesse ocorrerão numa freqüência muito menor, o que faz do sistema de filtragem *online* um componente fundamental para os detectores.

O ATLAS [1] é um detector de propósito geral do LHC e está posicionado em um dos pontos de colisão. Entre os principais objetivos do ATLAS, pode-se destacar a busca pela partícula conhecida como *bóson de Higgs*, que segundo estudos teóricos seria responsável por interagir com as partículas fornecendo-lhes massa [6]. A partícula de Higgs ainda não foi verificada experimentalmente.

Parcela importante das informações necessárias para a caracterização dos eventos é obtida do sistema de calorímetros, que, no ATLAS, é subdividido em 7 camadas. Os calorímetros são medidores de energia compostos por um grande número de sensores (células). Ao interagirem com o material do calorímetro as partículas perdem energia (e consequentemente velocidade). As células dos calorímetros quantificam a energia perdida pelas partículas incidentes e a informação do perfil de deposição de energia é utilizada para a caracterização do tipo de partícula.

Os objetivos principais dos sistemas de filtragem *online* em experimentos de físicas de altas energias são maximizar a probabilidade de detecção (e consequente armazenamento) dos eventos de interesse, e minimizar a probabilidade de armazenar eventos não desejados (ruído de fundo ou falso-alarme). Em um ambiente como este, a alta dimensão dos dados, o intenso ruído de fundo e o curto tempo de resposta exigido são sérios entraves para o processamento e classificação *online* de eventos.

No ATLAS, o sistema *online* de filtragem (*trigger*) de eventos é composto por três níveis de seleção sequenciais. O ruído de fundo é gradualmente reduzido a cada nível de filtragem, esperando-se armazenar, em mídia permanente, uma taxa máxima de 200 Hz [7]. Considerando que a freqüência de colisões é 40 MHz, deve haver uma redução de 2×10^5 vezes.

No contexto dos diversos canais de interesse para a física no ATLAS, este trabalho se encontra focado na discriminação elétron/jato (e^-/j). Os elétrons podem estar

envolvidos em fenômenos como o decaimento do bóson de Higgs, supersimetria e a descoberta de novos bósons. Porém, em termos de calorimetria, alguns jatos podem apresentar um perfil de deposição de energia semelhante ao dos elétrons. Portanto, os jatos representam ruído de fundo no processo de identificação de elétrons. Apenas uma parcela dos candidatos a elétrons aceitos pelo primeiro nível são realmente elétrons; cabendo à filtragem de alto nível reduzir ainda mais o ruído de fundo, mantendo a maior parte das assinaturas de interesse.

Considerando a alta taxa de eventos e a intensidade do ruído de fundo produzidos pelo LHC, a busca por algoritmos de filtragem *online* mais eficientes é uma tarefa importante. A redução no número de eventos não relevantes (ruído de fundo) armazenados em mídia permanente significa maior eficiência na análise *offline* dos eventos de interesse e redução do espaço (mídia) necessário para armazenamento.

Neste trabalho estão sendo propostas alternativas para o algoritmo padrão de detecção de elétrons em uso atualmente no segundo nível de filtragem (L2) do detector ATLAS. Os algoritmos desenvolvidos apresentam maior eficiência de discriminação e tempo de processamento dentro das restrições do L2.

1.3 Trabalhos Anteriores

Considerando os desafios existentes no ambiente de filtragem *online* do ATLAS, no qual os sinais são adquiridos com fina segmentação, estão imersos em intenso ruído de fundo e as assinaturas de interesse são raras, técnicas avançadas de extração de características podem ser utilizadas para melhorar a eficiência de classificação.

No trabalho [8] foi inicialmente proposto o uso de um classificador neural supervisionado (arquitetura *Perceptron* de Múltiplas Camadas) para o canal elétron/jato do segundo nível de filtragem do detector ATLAS. Utilizando informação especialista a respeito do problema, os sinais medidos nos calorímetros são formatados em anéis concêntricos de deposição de energia. A formatação dos anéis preserva a informação discriminante do perfil de deposição de energia e compacta a informação (de 1000 células para 100 anéis).

Em [9] o sistema neural de identificação de elétrons (que ficou conhecido como

Neural Ringer) foi implementado no sistema (*software*) de filtragem do ATLAS. Numa comparação de desempenho entre o Neural Ringer e o discriminador oficial do ATLAS (T2Calo) foi mostrado que o *Ringer* apresenta desempenho superior e é capaz de operar dentro da janela de tempo permitida para o L2.

No trabalho [10] alguns métodos de compactação como Análise de Componentes Principais (PCA - *Principal Component Analysis*) [3] e Componentes Principais de Discriminação (PCD - *Principal Components of Discrimination*) [11] foram aplicados, em conjunto com o modelo linear da Análise de Componentes Independentes (ICA - *Independent Component Analysis*) [4], sobre os sinais em anéis como um pré-processamento para o classificador neural. A utilização destas técnicas de extração de características e compactação permitiu aumento do desempenho de classificação e redução do tempo necessário para tomada de decisão. Neste trabalho foi realizada também uma implementação otimizada do *Neural Ringer* no sistema *online* de filtragem do ATLAS operando como uma sub-rotina do T2Calo. Neste nova versão o custo computacional foi reduzido e o Ringer utiliza uma parte do processamento realizado pelo T2Calo.

1.4 Objetivos

Os calorímetros são projetados para serem detectores lineares, porém, diversas fontes de não-linearidades podem surgir numa implementação prática [12]. Neste caso, um método não-linear de extração de características talvez seja mais indicado para o problema.

O principal objetivo do presente trabalho é avaliar o desempenho obtido pelo discriminador *Neural Ringer* quando os sinais em anéis são pré-processados por métodos de extração de características baseados no modelo não-linear da análise de componentes independentes (NLICA - *Nonlinear Independent Component Analysis*) [13].

Neste contexto, foram utilizados diversos modelos e algoritmos de estimação da NLICA e seus resultados comparados em termos da eficiência de discriminação e do tempo de processamento. Foi proposta também a segmentação dos processos de extração de características e classificação, visando explorar adequadamente toda

segmentação e granularidade disponíveis no detector.

O modelo da NLICA foi originalmente definido para realizar a extração de características de modo não-supervisionado. Ou seja, não há como garantir que a transformação seja útil para o problema de classificação (no sentido de revelar características discriminantes). Neste trabalho foram propostas modificações no modelo tradicional da NLICA visando a estimativa de componentes com maior poder de discriminação entre as classes.

1.5 Metodologia

O projeto dos discriminadores de partículas foi realizado a partir de um conjunto de eventos simulados. Estas simulações, por técnicas de Monte Carlo [14], consideram todas as características físicas do detector e do acelerador. Foram utilizados, também, conjuntos de dados experimentais obtidos na fase inicial de operação do LHC. Sabe-se que os raios cósmicos representam ruído de fundo para o canal elétron/jato, pois produzem múons no detector. Assim, os algoritmos propostos foram também testados para uma base de dados composta de eventos de raios cósmicos, visando verificar a capacidade de rejeição para esse sinal. Uma outra análise utilizou eventos de colisão obtidos recentemente na fase inicial de operação do LHC.

Foram aplicados diversos algoritmos para estimar o modelo não-linear das componentes independentes (NLICA). Visando explorar adequadamente as características do detector, o modo de executar as tarefas de extração de características e classificação foi variado entre as abordagens segmentada (onde o processamento é feito a nível de cada camada do calorímetro) e não-segmentada (quando os sinais em anéis, gerados a partir de todas as camadas, são concatenados em um único vetor).

Entre os diversos modelos existentes para a estimação da NLICA, neste trabalho foram utilizados os modelos sem restrição estrutural (através dos mapas auto-organizáveis) e o modelo pós não-linear (que restrige os mapeamentos não-lineares possíveis a uma mistura linear seguida de funções não-lineares aplicadas a cada componente desta mistura). A ICA Local, que é um modelo diretamente ligado ao da NLICA também foi utilizada. Para cada modelo proposto foi realizado um estudo

comparativo de desempenho com o Neural Ringer quando este opera diretamente sobre os sinais em anéis.

Considerando o sistema de classificação neural, foi proposta a utilização de classificadores especialistas na informação de cada camada do calorímetro. Diversos modos de combinar a informação deste conjunto de classificadores foram testados. Desta forma verificou-se que há redundância de informação entre as diversas camadas do calorímetro. Pode-se, então, obter o mesmo desempenho de discriminação utilizando-se apenas uma parcela das informações disponíveis e, assim, contribuir para a redução do tempo de processamento.

1.6 Conteúdo do trabalho

No Capítulo 2 será apresentado o ambiente científico no qual o trabalho está sendo desenvolvido, contextualizando o detector de partículas ATLAS, o acelerador LHC e o CERN. Uma descrição dos sistemas de *trigger online* em experimentos de física de altas energias e, mais especificamente, no detector ATLAS serão apresentadas no Capítulo 3.

É descrito, no Capítulo 4, o processo de seleção de elétrons utilizando informações de calorimetria no detector ATLAS. No Capítulo 5, serão mostrados os fundamentos teóricos das técnicas de extração de características e teste de hipótese que serão utilizadas para a otimização do sistema de filtragem do ATLAS.

A metodologia empregada no desenvolvimento deste trabalho é descrita no Capítulo 6. Os conjuntos de dados utilizados para desenvolvimento e teste dos sistemas de classificação propostos serão apresentados no Capítulo 7.

Os resultados obtidos serão apresentadas no Capítulo 8. As conclusões e os futuros trabalhos são os tópicos abordados no Capítulo 9.

Nos Apêndices A e B são fornecidas as bases matemáticas para uma melhor compreensão, respectivamente, dos algoritmos de extração de características e classificação utilizados. Finalmente, no Apêndice C serão listadas as publicações produzidas com os resultados obtidos neste trabalho.

Capítulo 2

Física de Partículas e o detector ATLAS

Os experimentos de Física de Partículas Elementares têm como principais objetivos a confirmação de modelos desenvolvidos teoricamente e a identificação de novos fenômenos. O experimento LHC [2] é o maior acelerador de partículas já desenvolvido e quando operando em máxima capacidade, serão gerados aproximadamente 10^9 interações por segundo. Os detectores são responsáveis por selecionar, dentro de um conjunto enorme, os raros eventos de real interesse. O projeto e a montagem do detector ATLAS (*A Thoroidal LHC Apparatus*) [1] foram conduzidos por uma colaboração de 36 países, conhecida como *ATLAS Collaboration*, contando com pesquisadores de mais de 150 universidades e centros de pesquisa [15]. Sendo um dos detectores de propósito geral do experimento LHC, o ATLAS tem formato cilíndrico e foi projetado para cobrir um ângulo sólido próximo a 4π ao redor da região de colisão.

2.1 Panorama geral da física de partículas elementares

A noção de que a matéria é composta por um conjunto de constituintes elementares existe há mais de 2000 anos, desde o tempo dos filósofos gregos [16]. No decorrer do século 20, a compreensão dos componentes elementares da matéria forneceu à comunidade científica mundial informações importantes a respeito das leis fundamentais

da natureza. O estudo da física de partículas elementares teve início no final do século 19, quando foi descoberto o elétron, experimentando um crescimento acen-tuado na década de 1950, quando foram descobertas centenas de novas partículas. A partir de sua criação em 1954, o CERN (Centro Europeu para Pesquisa Nuclear) contribuiu significativamente nesse processo.

Inicialmente, pensava-se que a matéria era constituída de partículas subatômicas, chamadas elétrons, prótons e nêutrons. Mais tarde, descobriu-se que os prótons e nêutrons são compostos de quarks. Hoje sabe-se que toda a matéria no universo é composta de léptons e quarks. Existem ainda outras partículas elementares que são responsáveis por promover a interação entre léptons e quarks. Juntamente com as descobertas experimentais, estudos teóricos possibilitaram o desenvolvimento do Modelo Padrão (*SM-Standard Model*) [6] que descreve e prevê, de forma unificada, o comportamento das partículas e das forças de interação.

As características e propriedades dos processos nucleares dependem da ener-gia envolvida. A unidade de energia mais utilizada, neste contexto, é o elétron-volt (eV) e seus múltiplos (MeV, GeV ou TeV). O elétron-volt é definido como a energia necessária para aumentar o potencial elétrico de um elétron em 1 volt (em unidades do Sistema Internacional temos: $1\text{eV}=1,6\times10^{-19}\text{J}$) [17]. Os fenômenos produzidos quando a energia das partículas é menor que 20MeV são chamados de física a baixas energias. A faixa entre 20MeV e 400MeV corresponde à física a energias intermediárias, e finalmente, fenômenos com energia superior a 400MeV são estudados na física a altas energias [18].

Os experimentos realizados a partir do início da década de 1970 foram fundamen-tais para o desenvolvimento e teste do Modelo Padrão, mas também despertaram dúvidas a respeito de questões que não são completamente respondidas. O SM (ver Figura 2.1), divide as partículas elementares em quarks e léptons. Existem seis ti-pos de léptons (elétron, mûon, tau e três neutrino) e seis tipos de quarks (*up, down, charm, strange, top e bottom*).

Atualmente são conhecidas quatro formas de interação (ou acoplamento) entre as partículas elementares, são elas eletromagnética, gravitacional, fraca e forte. A força gravitacional é predominante entre corpos massivos separados por longas distâncias. Entre partículas, onde a massa é da ordem de 10^{-27}Kg , a interação gravitacional

	Fermions			Bósons
Quarks	u up	c charm	t top	y fóton
	d down	s strange	b bottom	Z
Leptons	V_e elétron neutrino	V_μ múon neutrino	V_τ tau neutrino	W
	e Elétron	μ Múon	τ tau	g glúon
Ainda não confirmado	Bóson de Higgs			

Figura 2.1: Diagrama do Modelo Padrão, mostrando as partículas elementares incluindo o, ainda não confirmado, bóson de Higgs.

tem intensidade muito baixa. Para distâncias maiores que 10^{-13}cm , a força eletromagnética domina, enquanto que para distâncias menores, as interações forte e fraca se destacam. A intensidade relativa (em comparação com a interação forte) dos quatro tipos de interação é mostrada na Tabela 2.1 [19].

Existem algumas teorias que descrevem como as interações elementares ocorrem. A eletrodinâmica quântica (QED - *Quantum Eletrodynamics*) considera que os processos elétricos e magnéticos acontecem a partir da interação fundamental entre dois elétrons, com a troca de um fóton. O fóton é a partícula mediadora da interação eletromagnética.

De modo semelhantes são definidos os outros tipos de interação. Para a interação

Tabela 2.1: Intensidade relativa (em comparação com a interação forte) dos diversos tipos de interação.

Interação	Intensidade Relativa
Forte	1
Eletromagnética	10^{-2}
Fraca	10^{-5}
Gravitacional	10^{-39}

forte, as partículas mediadoras são os glúons. Os bósons W e Z mediam a interação fraca e o gráviton a interação gravitacional. A teoria da cromodinâmica quântica (QCD - *Quantum Chromodynamics*) [20], por exemplo, descreve como os quarks e glúons interagem para formar os hadrons (prótons e neutrons são exemplos de hadrons). Uma diferença entre a QED e a QCD é que na última, os quarks e glúons não são observados como partículas livres (mas apenas na forma de hadrons), em oposição aos elétrons e fótons da QED. Mesmo em colisões de alta energia, quando quarks são produzidos, eles se afastam rapidamente uns dos outros, e antes que possam ser observados, se convertem em “jatos” de hadrons (ou jatos hadrônicos) [17].

Considerando as partículas e formas de interação já verificadas experimentalmente e uma formulação atualmente aceita pela física teórica (conhecida como *Gauge Invariance*), todas as partículas elementares teriam massa nula. Esta previsão é contrária aos resultados experimentais e somente pode ser corrigida assumindo-se que existe um outro tipo de interação. Esta interação foi prevista pelo cientista inglês Peter Higgs em 1964, tendo como partícula mediadora o bóson de Higgs, que é responsável por fornecer massa às partículas [17]. A existência do bóson de Higgs é a mais importante previsão do modelo padrão ainda não verificada experimentalmente e sua busca é de máxima importância para a física de partículas.

Nas últimas décadas, os experimentos com aceleradores de partículas tornaram gigantescos empreendimentos envolvendo milhares de físicos e engenheiros, com contribuição financeira e intelectual de dezenas de países.

As partículas são injetadas na máquina por dispositivos que produzem uma fonte de alta intensidade e baixa energia. Os aceleradores usam força eletromagnética para aumentar a energia de partículas estáveis e carregadas eletricamente. Quanto às características construtivas, os aceleradores podem ser divididos em alvo-fixo ou colisionadores de feixes. Nos aceleradores de alvo-fixo, as partículas são aceleradas até a máxima energia, quando o feixe é retirado da máquina e direcionado a um alvo estacionário. Nos colisionadores, feixes de partículas são acelerados, em sentidos opostos, e quando a energia desejada é atingida, os feixes são colocados em rota de colisão em alguns pontos específicos do percurso. Quanto ao percurso dos feixes, os aceleradores podem ser lineares ou circulares [17].

Imediatamente após a colisão, uma grande quantidade de partículas elementares

é produzida. Algumas delas são estáveis, porém outras têm curtíssimo tempo de vida. Os elétrons e prótons, por exemplo, têm vida média superior a 10^{23} anos, enquanto os mísions podem ter vida média da ordem de 10^{-6} segundos [18]. Para que todos os eventos sejam corretamente identificados, um complexo sistema de detecção precisa ser construído.

A massa (m) das partículas é usualmente expressa em função da energia equivalente ($m = E/c^2$, onde c é a velocidade da luz e E a energia). Os bósons W e Z, por exemplo, têm massa, respectivamente, de 80 e 91 GeV/c² (onde $1\text{GeV}/c^2=1,78\times10^{-27}\text{ kg}$). Na Tabela 2.2, são mostradas as massas de algumas partículas importantes. Ainda não é possível determinar a massa esperada para o bóson de Higgs, porém, a partir do conhecimento adquirido com os experimentos operados antes do LHC, sua massa deve ser maior que 115 GeV/c² [6].

O acelerador LHC (*Large Hadron Collider* - Grande Colisionador Hadrônico) [2, 21] em operação no CERN (Centro Europeu para Pesquisa Nuclear) [22], será capaz de atingir energia de até 14 TeV e permitir a visualização do bóson de Higgs, caso o Modelo Padrão esteja correto.

Tabela 2.2: Exemplos de valores da massa (em energia equivalente) de algumas partículas.

Partícula	Massa (GeV/c ²)
Elétron	0,000511
Próton	0,938
Bóson W	80
Bóson Z	91
Bóson de Higgs ou outras novas partículas	>115

Além da busca pelo bóson de Higgs, existem outras questões que precisam ser respondidas em física de partículas. É um desejo antigo dos físicos, desde Einstein, a unificação das teorias sobre as forças de interação entre as partículas, incluindo no Modelo Padrão a força gravitacional. O estudo da física de partículas também é fundamental para o entendimento da natureza e origem do universo. Pretende-se, por exemplo, descobrir informações sobre a composição da matéria escura [23], super-simetria [24] e violação de CP (do inglês *Charge Parity*) [25].

Entre os laboratórios que conduzem experimentos em física de partículas pode-se destacar: CERN, DESY [26], KEK [27], Fermilab [28], SLAC [29] e Brookhaven [30], localizados respectivamente na Suíça, Alemanha, Japão e os três últimos nos Estados Unidos.

O CERN, fundado em 1954, é atualmente um dos maiores centros de pesquisas em física de partículas do mundo. Localizado em Genebra, Suíça, funciona com base num complexo sistema de colaboração internacional, que envolve centenas de instituições de pesquisa em centenas de países [22].

2.2 O acelerador LHC

No CERN, foi projetado e construído o experimento LHC (*Large Hadron Collider* ou Grande Colisionador de Hádrons), que iniciou sua operação no final de 2008 [2, 31]. O LHC pode atingir níveis de energia de, aproximadamente, 14 TeV. O percurso do acelerador, localizado na fronteira franco-suíça, é aproximadamente circular, com 27 Km de comprimento, numa profundidade do solo que varia de 50 a 150 metros (ver Figura 2.2). Feixes de prótons são acelerados em sentidos opostos e direcionados para colisão nos centros dos detectores.

O LHC tem 6 detectores com propósitos diferentes: ATLAS [1], CMS [32], LHCb [33], LHCf [34], ALICE [35] e TOTEM [36]. O ATLAS e o LHCf estão localizados em Meyrin, Suíça, o CMS e o TOTEM em Cessy, França, o ALICE em St. Genis-Pouilly, França e o LHCb em Ferney-Voltaire, França (os experimentos TOTEM e LHCf não são mostrados na Figura 2.2 pois estão em locais próximos respectivamente ao CMS e ATLAS). O ATLAS (*A Toroidal LHC ApparatuS*) e o CMS (*Compact Muon Solenoid*) são detectores de propósito geral, enquanto os outros são dedicados a aplicações específicas, como o LHCb, que é dedicado a explorar informações sobre a física proveniente dos hádrons do tipo **b** produzida nas colisões do LHC.

Quando operando em máxima capacidade, o LHC irá produzir colisões de feixes de prótons a cada 25 ns e atingirá energia 7 vezes maior que o Tevatron, que é o acelerador de maior energia em operação atualmente, funcionando no Fermilab.

O número de colisões por centímetro quadrado produzidas por segundo é cha-

mado de luminosidade (L) [17]:

$$L = n \frac{N_1 N_2}{A} f \quad (2.1)$$

onde n é o número de feixes, N_1 e N_2 o número de partículas em cada feixe, A a área da seção transversal do feixe e f a frequência de colisão. Quanto maior a luminosidade do experimento, maior a quantidade de informação (partículas) gerada.

Quando operando numa luminosidade jamais alcançada por outros experimentos ($10^{34} cm^{-2}s^{-1}$), espera-se, em média, 25 choques próton-próton a cada cruzamento de feixe (colisão). Como a frequencia de colisões é de 40 MHZ, a taxa de interações será da ordem de 1 GHz. No LHC, fatores como a alta taxa de interações, as altas doses de radiação, a multiplicidade de partículas e largas faixas de energia a cobrir, em conjunto com a necessidade de medições precisas, definiram novos padrões para o projeto dos detectores.

A seguir, serão descritas, de modo geral, as principais características do detector ATLAS.

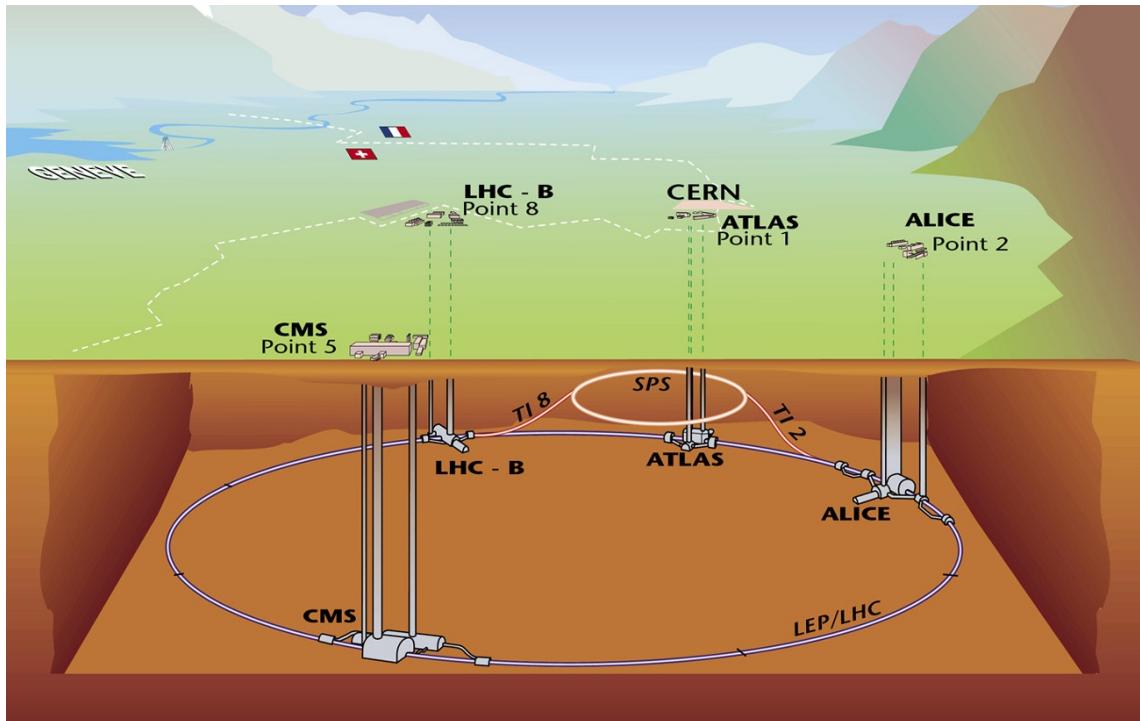


Figura 2.2: Mapa de localização dos detectores do LHC.

2.3 Características gerais do detector ATLAS

Os métodos de detecção em física têm como princípio básico promover interação entre as partículas em estudo e um material conhecido, produzindo informações sobre a natureza e as características da própria partícula. Os instrumentos que possibilitam a medição experimental destas quantidades físicas são chamados de detectores. Em particular, os detectores de energia são chamados de calorímetros. À medida que a energia envolvida aumenta, o sistema de detecção precisa ser mais sofisticado [18].

O ATLAS foi fruto do trabalho de uma grande colaboração, que envolveu milhares de físicos, engenheiros, técnicos e estudantes por um período de vinte anos de projeto, desenvolvimento, fabricação e instalação.

O detector tem 45 m de comprimento, 25 m de altura e pesa aproximadamente 7.000 toneladas, sendo dividido em subsistemas dispostos em camadas. Conforme ilustrado na Figura 2.3, os principais subsistemas são: detector de trajetórias (ou de traço), calorímetros eletromagnético e hadrônico e os detectores de múons.

A função dos detectores de trajetória é medir o momento das partículas electricamente carregadas, a partir da curvatura de sua trajetória, quando imersos no campo magnético do solenóide central [37]. Caminhando do eixo central para as extremidades, em sequência, estão os calorímetros, que medem a energia depositada pelas partículas [38]. Na interação com as células do calorímetro, são produzidos chuveiros de partículas secundárias [12]. Num último estágio está o detector de múons, sistema dedicado à detecção destas partículas, que são as únicas que não ficam contidas nos calorímetros [39].

O sistema xyz de coordenadas do ATLAS é único para todos os subsistemas. Conforme mostrado na Figura 2.4, a direção do feixe do LHC define o eixo z , e os eixos x e y formam um plano transverso ao feixe. A direção positiva do eixo x é definida apontando do ponto de interação para o centro do anel do LHC, e o eixo y positivo aponta para cima. O ângulo azimutal é obtido a partir de:

$$\phi = \arctg(x/y), \quad (2.2)$$

sendo que $\phi = 0$ corresponde ao eixo x positivo e ϕ aumenta no sentido horário.

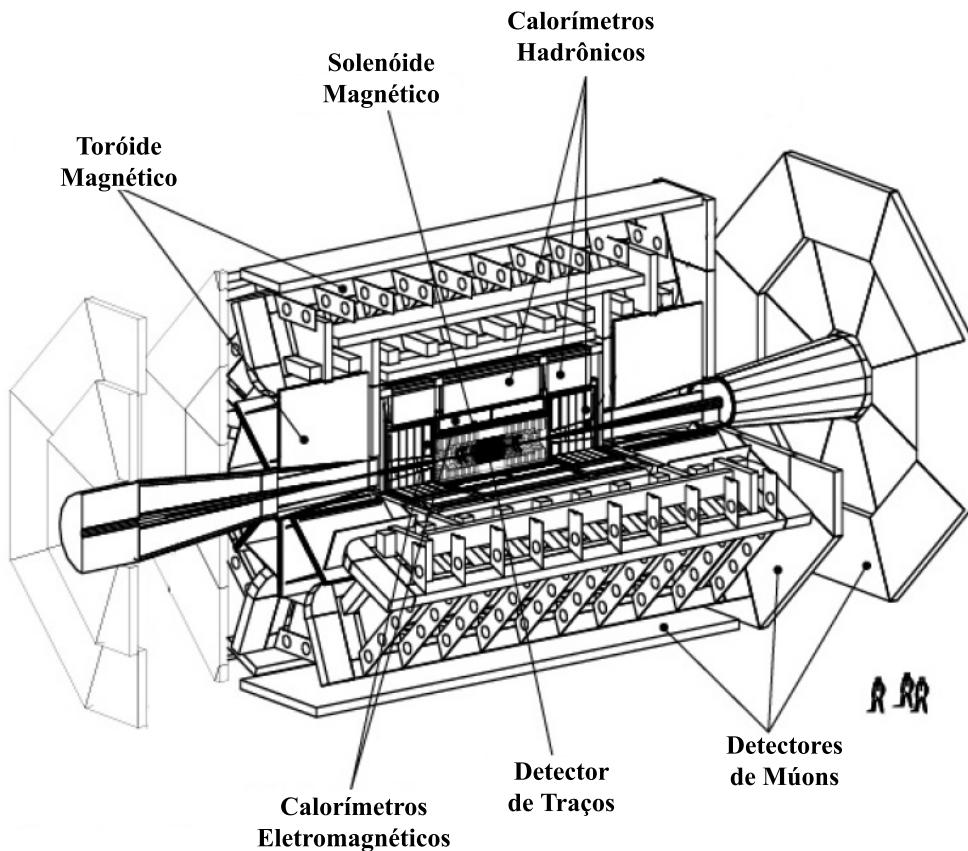


Figura 2.3: Diagrama esquemático do ATLAS.

O ângulo polar θ é medido a partir do eixo do feixe (eixo z positivo). O momento transverso p_T , a energia transversa E_T e a energia transversa perdida E_T^{miss} são definidas no plano xy . A pseudo-rapidez η é calculada a partir do ângulo θ de espllhamento em relação ao eixo z (ângulo de saída das partículas após a colisão) [7]:

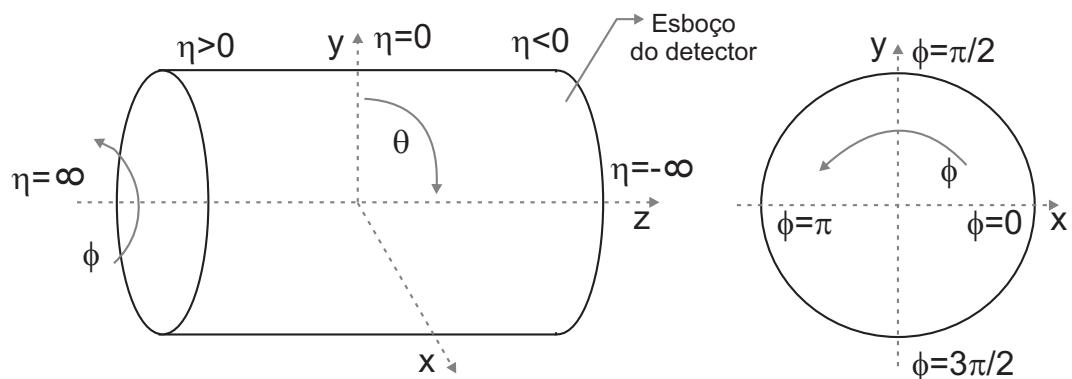


Figura 2.4: Eixo de coordenadas do ATLAS.

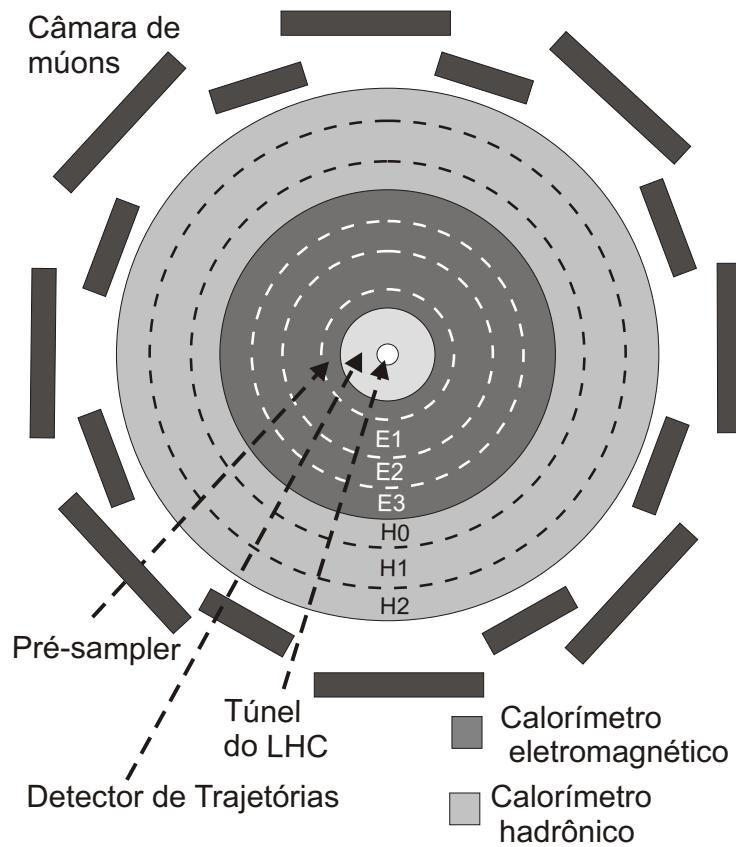
$$\eta = -\ln \operatorname{tg}(\theta/2). \quad (2.3)$$

A partir das definições das equações 2.2 e 2.3, define-se o eixo (η, ϕ) , onde o ângulo ϕ representa a rotação e η a direção de projeção das partículas após a colisão. Grandes valores da pseudo-rapidez indicam que a colisão das partículas não foi frontal, pois o ângulo de saída, após o choque, é pequeno, no limite quando $\theta \rightarrow 0$ então $\eta \rightarrow \infty$. Nesse tipo de colisão, como quase não houve choque, a produção de partículas elementares é pequena. O ATLAS foi projetado com baixa resolução para $\eta > 3$.

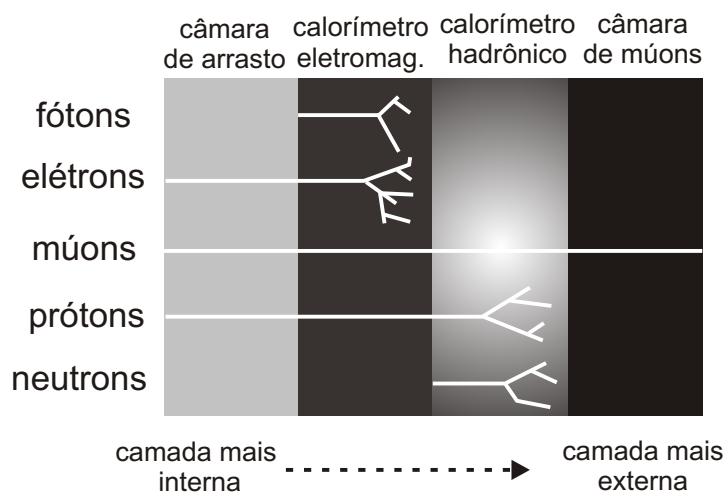
Após uma colisão, as partículas geradas interagem com o material do detector, perdendo energia e consequentemente velocidade. Na Figura 2.5(a) pode-se visualizar os subsistemas do ATLAS num corte transversal (paralelo ao plano xy). Percebe-se que a câmara de mísseis e o calorímetro hadrônico têm as maiores dimensões. Na figura 2.5(b) é mostrada a penetração e consequente visualização esperadas de algumas partículas nas camadas do detector. Não espera-se, por exemplo, deposição de energia de fótons ou elétrons além das camadas eletromagnéticas do calorímetro, pois estas partículas interagem intensamente com os materiais que compõem a seção eletromagnética, perdendo toda sua energia. As partículas hadrônicas (por exemplo: prótons e nêutrons) em geral interagem menos intensamente com o calorímetro eletromagnético e precisam das camadas hadrônicas para serem paradas. Os mísseis são partículas que perdem pouca energia nos calorímetros, necessitando de um sistema especial para serem detectados, o detector de mísseis.

O ATLAS foi projetado e construído considerando-se as condições experimentais produzidas pelas colisões do LHC e com o objetivo de obter informações importantes para responder às questões chave em física de partículas. Entre os principais critérios utilizados para o projeto do detector pode-se destacar [1]:

- Os elementos sensores e circuitos eletrônicos devem apresentar resposta rápida e resistência a altos níveis de radiação. A fina granularidade do detector é importante para reduzir a influência dos eventos sobrepostos.
- Excelente calorimetria eletromagnética para a identificação de elétrons e fótons, complementada por informações acuradas dos calorímetros hadrônicos, para medições de jatos hadrônicos e energia transversa E_T ;



(a)



(b)

Figura 2.5: Cortes (a) transversal e (b) axial do ATLAS.

- Eficiente sistema de identificação de trajetória para medição do momento;
- Alta precisão na identificação de múons;
- Alta aceitação em pseudo-rapidez (η) com cobertura quase total no ângulo azimutal (ϕ);
- Alta eficiência do sistema de filtragem (*trigger*), armazenando a maioria dos eventos físicos de interesse e reduzindo ao máximo o ruído de fundo (informação não relevante) produzido nas colisões do LHC.

Conforme mencionado no Capítulo 1, o estudo conduzido neste trabalho utiliza informações do sistema de calorímetros do ATLAS e propõe algoritmos para a otimização da detecção (*trigger*) de elétrons. Nas próximas seções, serão descritos o sistema de calorímetros do ATLAS e a importância da detecção de elétrons para o desempenho do experimento.

2.4 Principais objetos de interesse no ATLAS

Dentre os eventos gerados nas colisões do LHC, apenas uma pequena parte será útil para a caracterização dos processos da “nova física”. Com o LHC operando em alta luminosidade podem ocorrer até 10^9 interações por segundo, porém, os eventos de interesse são muito raros. Por exemplo a depender de sua massa, a taxa de ocorrência do bóson de Higgs pode variar de 0,01 a 0,1 Hz e eventos de supersimetria são esperados a 1 Hz [40]

Provar a existência do bóson de Higgs é um dos principais objetivos do LHC. Com o conhecimento adquirido até agora, não é possível determinar sua massa m_H , embora seu limite inferior ($m_H > 114 \text{ GeV}/c^2$) tenha sido determinado pelos resultados obtidos em outros aceleradores, como o LEP (*Large Electron Positron Collider*, acelerador que operou no CERN de 1989 a 2000) [41]. O limite superior esperado por estudos teóricos é $\sim 1 \text{ TeV}/c^2$ [42].

Considerando os diversos decaimentos possíveis para a partícula de Higgs, espera-se que o canal mais limpo para seu estudo aconteça se sua massa estiver aproximadamente na faixa $150 < m_H < 700 \text{ GeV}/c^2$ [1]. Neste caso, o Higgs pode apresentar

o decaimento em 2 bósons Z, com cada Z, por sua vez, decaindo¹ em dois léptons (elétrons ou múons):

$$H \rightarrow ZZ^{(*)} \rightarrow l^+l^-l^+l^- \quad (2.4)$$

Para este canal de busca do Higgs, é fundamental que os detectores tenham um sistema de filtragem capaz de identificar com alta eficiência elétrons e múons. Estas duas partículas, assim como os fótons, jatos e os taus, são importantes também para o melhor entendimento da supersimetria (SUSY - *supersimetry*) [6]. Os taus também podem levar aos modelos de Higgs estendidos [43]. Outro fenômeno importante é a energia transversa perdida (E_T^{miss}), que geralmente é atribuído a partículas que passaram pelo detector e não foram detectadas. A E_T^{miss} pode ser identificada a partir da violação dos princípios de conservação de energia e momento.

Um resumo com os principais objetos de interesse no ATLAS e suas aplicações na física é mostrado na Tabela 2.3.

Tabela 2.3: Principais objetos de interesse no ATLAS, extraída de [44].

Objeto	Física de interesse
Elétron	Higgs, SUSY, dimensões-extra, novos bósons e quark top
Fóton	Higgs, SUSY e dimensões-extra
Múon	Higgs, SUSY, extra-dimensões, novos bósons e quark top
Jato	SUSY e ressonâncias
Jato + E_T^{miss}	SUSY e lépton-quarks
Tau + E_T^{miss}	Modelo de Higgs estendido e SUSY

A identificação *online* destes objetos dentro de um universo enorme de informações é realizada pelo sistema de filtragem (*trigger*). Conforme mencionado, um dos objetos de interesse na filtragem do ATLAS é o elétron. Para a identificação de elétrons, a informação obtida no sistema de calorímetros é muito importante. Em termos da calorimetria, as assinaturas de elétrons podem ser confundidas com perfis de energia gerados por alguns jatos hadrônicos (espacialmente concentrados nas camadas eletromagnéticas e com pouca energia nas hadrônicas). Considerando que a produção de jatos será muito frequente nas colisões do LHC, estes formarão

¹No decaimento de partículas elementares, parte da massa da partícula é convertida em energia e o restante em massa de outras partículas

um intenso ruído de fundo para a identificação de elétrons, tornando a discriminação elétron/jato (e^-/j) importante para o desempenho do detector.

2.5 O sistema de calorimetria do ATLAS

2.5.1 Breve introdução à calorimetria

Calorímetros [12] representam uma importante classe de detectores para medição de energia e posição da partícula. Durante o processo de absorção, as partículas interagem com o material dos calorímetros gerando partículas secundárias que, por sua vez, interagem também gerando outras partículas e assim por diante. Este processo é chamado de cascata ou chuveiro de partículas. Os calorímetros podem apresentar resposta muito rápida, da ordem de nanosegundos [12], e, por isso, são utilizados intensamente pelo sistema de filtragem online (*trigger*). As prováveis classes de partículas são identificadas a partir das características esperadas para o seu perfil de deposição de energia.

Calorímetros são detectores de absorção total. O processo de medição utilizado é destrutivo e as partículas não estão disponíveis após a passagem pelos calorímetros (com exceção aos mísseis, que conseguem penetrar em grande quantidade de matéria e necessitam de um detector especial a câmara de mísseis). Quando partículas atravessam matéria, elas em geral interagem e perdem assim uma parte de sua energia. Neste processo o meio é excitado ou aquecido (daí o termo calorimetria). Na prática absorção “total” significa 99 a 99,9 % da energia (ou até um pouco menos). A depender do tipo de partícula e da energia envolvida, a partícula pode exceder os limites do calorímetro (vazar) e assim interferir em outros detectores (como a câmara de mísseis) [12].

Os calorímetros podem ser classificados em homogêneos ou amostradores. No calorímetro homogêneo todo o volume do detector é sensível às partículas e contribui para produção do sinal. No calorímetro amostrador, o material passivo interage com as partículas (absorvendo sua energia) e o material ativo produz o sinal [12].

A depender do tipo de partícula, a interação com o calorímetro ocorre de modo distinto. Partículas eletromagnéticas (EM), como elétrons e pósitrons, interagem com a matéria gerando um chuveiro de partículas menos energéticas e necessitam de

pequena quantidade de material para serem totalmente absorvidas. Os mísions, por outro lado, perdem sua energia muito lentamente necessitando de grande quantidade de matéria para a absorção total. As partículas hadrônicas interagem com a matéria através da força nuclear forte. O processo de interação é muito mais complexo que o EM e uma variedade muito grande de fenômenos pode ocorrer. Os hadrons podem, por exemplo, se comportar de modo semelhante a elétrons e mísions, ou se envolverem em uma interação nuclear e se transformarem em 15 novos hadrons. Uma parte da energia das interações hadrônicas não é visível (detectável) pelos calorímetros pois os hadrons neutros não ionizam o calorímetro e a energia é perdida em interações nucleares (não detectadas pelo calorímetro) [12].

Como as características dos chuveiros eletromagnéticos e hadrônicos são diferentes, na prática, são utilizados tipos de calorímetro específicos para estas classes de partículas [17]. O calorímetro eletromagnético é usualmente instalado internamente ao hadrônico. As partículas eletromagnéticas (ex: elétrons e fótons) apresentam perfil de deposição de energia que, em geral, é concentrado em torno do ponto de colisão. Tipicamente, as partículas eletromagnéticas são absorvidas completamente nos calorímetros eletromagnéticos. Os chuveiros hadrônicos apresentam formas variadas e iniciam sua interação com o calorímetro eletromagnético, mas, em geral, somente são completamente absorvidas nas camadas hadrônicas (mais externas).

Os calorímetros deveriam ser intrinsecamente lineares para a detecção de partículas EM. Por exemplo um par de elétrons de 10 GeV deveria gerar um sinal de mesma intensidade que um único elétron de 20 GeV. Na prática, desvios do comportamento linear (para partículas EM) podem ser observados na prática devido a fenômenos como [12]:

- Saturação das foto-multiplicadoras (PMT - *Photo-Multipliers*) - as PMT convertem luz dos calorímetros cintiladores em sinais elétricos, sua saturação implica em distorção não-linear dos sinais elétricos;
- Vazamento do chuveiro - com o aumento da energia algumas partículas podem extrapolar os limites do detector, havendo, neste caso, perda de parte da informação;
- Recombinação dos elétrons com íons do material ativo - se isso ocorrer a io-

nização não é detectada;

- Atenuação da luz - a luz emitida pelo material ativo (cintilante) pode ser atenuada antes de atingir as PMT.

Calorímetros homogêneos são intrinsecamente não-lineares para a detecção de hadrons e jatos. A fração EM de chuveiros hadrônicos depende da energia e varia bastante de evento para evento, tornando a resposta hadrônica não-constante em função da energia (tanto para calorímetros homogêneos como para os amostradores). A fração não EM (que produz interações nucleares) é, em geral, menor. Em qualquer calorímetro a resolução em energia para hadrons é pior que para elétrons de mesma energia. Isso se deve ao fato de ocorrerem flutuações na energia visível (detectável) aos calorímetros [12].

Diferente de outros tipos de detectores, a precisão dos calorímetros aumenta com a energia [12]:

$$\frac{\sigma_E}{E} \propto \frac{1}{\sqrt{E}} \quad (2.5)$$

onde E é a energia incidente por partícula e σ_E a variação (desvio padrão) esperado na medição. Outras fontes de flutuações de menor importância contribuem com fatores de outra ordem como ruído eletrônico: $\propto 1/E$ (domina em baixa energia, principalmente em calorímetros de Argônio Líquido - LAr); e vazamento lateral do chuveiro: $\propto 1/(E)^{1/4}$. Existem ainda flutuações que são independentes da energia. Uma característica interessante é que as flutuações podem não ser simétricas em torno do valor médio. Um estudo detalhado a respeito das flutuações encontradas em calorímetros pode ser encontrado em [12].

Para o calorímetro do ATLAS, foi calculada experimentalmente em [38] a resolução esperada . Os valores encontrados foram $0,1/\sqrt{E}$ e $0,4/\sqrt{E}$, respectivamente para os calorímetros eletromagnético (de argônio líquido) e hadrônico (de telhas cintilantes). A seguir, o sistema de calorímetros do ATLAS será descrito mais detalhadamente.

2.5.2 Características dos calorímetros do ATLAS

O sistema de calorímetros do detector ATLAS [1] é sub-dividido em 7 camadas [38], sendo 4 eletromagnéticas (PS, E1, E2, E3) e 3 hadrônicas (H0, H1 e H2), conforme

Figura 2.6. Cada camada apresenta diferente concentração de células detectoras por unidade de área (granularidade). O calorímetro eletromagnético (EM) é composto de finas folhas de chumbo separadas por dispositivos sensores de argônio líquido, cobrindo a região onde $|\eta| < 3,2$. As três camadas do calorímetro eletromagnético são divididas em barril (região central do detector, onde $|\eta| < 1,5$) e tampa (regiões mais externas onde $1,4 < |\eta| < 3,2$). Na região onde $|\eta| < 1,8$, imediatamente antes da primeira camada EM existe uma fina camada de argônio líquido chamada pré-amostrador (*presampler* ou PS). O pré-amostrador é importante para corrigir medições nas quais existe perda de energia no caminho até os calorímetros.

O calorímetro hadrônico envolve o eletromagnético. Na região onde $|\eta| < 1,7$, ele é composto de placas absorvedoras de aço separadas por telhas de material plástico cintilante. Quando as partículas atravessam as telhas elas emitem luz de intensidade proporcional à energia incidente [45]. O sinal luminoso é então convertido em elétrico através de placas foto-multiplicadoras. Esta parte do calorímetro hadrônico do ATLAS é conhecida como calorímetro de telhas (*Tile Calorimeter* ou simplesmente *TileCal*) e é dividida em barril ($|\eta| < 1,0$) e barril-extendedo ($0,8 < |\eta| < 1,7$) [46]. Para a tampa do calorímetro hadrônico ($|\eta| > 1,5$) utiliza-se a tecnologia do argônio líquido.

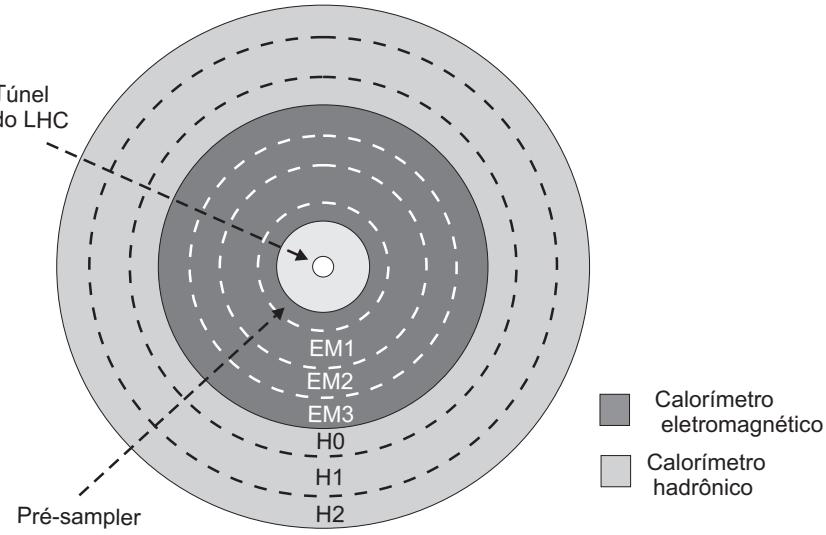


Figura 2.6: Disposição em camadas dos calorímetros do ATLAS.

A informação da energia depositada nas camadas do calorímetro, com fina segmentação, é muito importante para a caracterização das partículas. Após um evento ser aceito pelo sistema de filtragem, todas as informações relativas a este evento são

Tabela 2.4: Região de cobertura em η , granularidade e número de canais de leitura das camadas dos calorímetros.

Pre-amostrador	Barril	Tampa
Cobertura	$ \eta < 1,52$	$1,5 < \eta < 1,8$
Granularidade ($\Delta\eta \times \Delta\phi$)	$0,025 \times 0,1$	$0,025 \times 0,1$
Canais de Leitura	7808	1536 (ambos os lados)
Eletromagnético	Barril	Tampa
Cobertura	$ \eta < 1,475$	$1,375 < \eta < 3,2$
Granularidade ($\Delta\eta \times \Delta\phi$)		
Camada 1	$0,025/8 \times 0,1$	$0,025/8 \times 0,1$ a $0,1 \times 0,1$
Camada 2	$0,025 \times 0,025$	$0,025 \times 0,025$ a $0,1 \times 0,1$
Camada 3	$0,050 \times 0,025$	$0,05 \times 0,025$
Canais de Leitura	101760	62208 (ambos os lados)
Had. Telhas Cintilantes	Barril	Barril estendido
Cobertura	$ \eta < 1,0$	$0,8 < \eta < 1,7$
Granularidade ($\Delta\eta \times \Delta\phi$)		
Camadas 1, e 2	$0,1 \times 0,1$	$0,1 \times 0,1$
Camada 3	$0,2 \times 0,1$	$0,2 \times 0,1$
Canais de Leitura	5760	4092 (ambos os lados)
Had. Argônio Líquido	Tampa	
Cobertura	$1,5 < \eta < 3,2$	
Granularidade ($\Delta\eta \times \Delta\phi$)		
Camadas 1, 2 e 3	$0,1 \times 0,1$ a $0,2 \times 0,2$	
Canais de Leitura	5632 (ambos os lados)	

armazenadas em mídia permanente para posterior análise *off-line*. A granularidade, ou quantidade de células por unidade de área, varia entre as camadas do calorímetro. A Tabela 2.4 traz informações sobre a região de cobertura em η , granularidade e quantidade de canais de leitura (células sensoras) de cada camada do calorímetro.

Conforme ilustrado na Figura 2.7, percebe-se que a primeira camada eletromagnética apresenta mais fina segmentação em η possibilitando medição precisa do

ponto de colisão nessa coordenada. A segunda camada apresenta células detectoras quadradas e maior profundidade, absorvendo maior parcela da energia. A terceira camada, por sua vez, captura os detalhes do final do chuveiro eletromagnético.

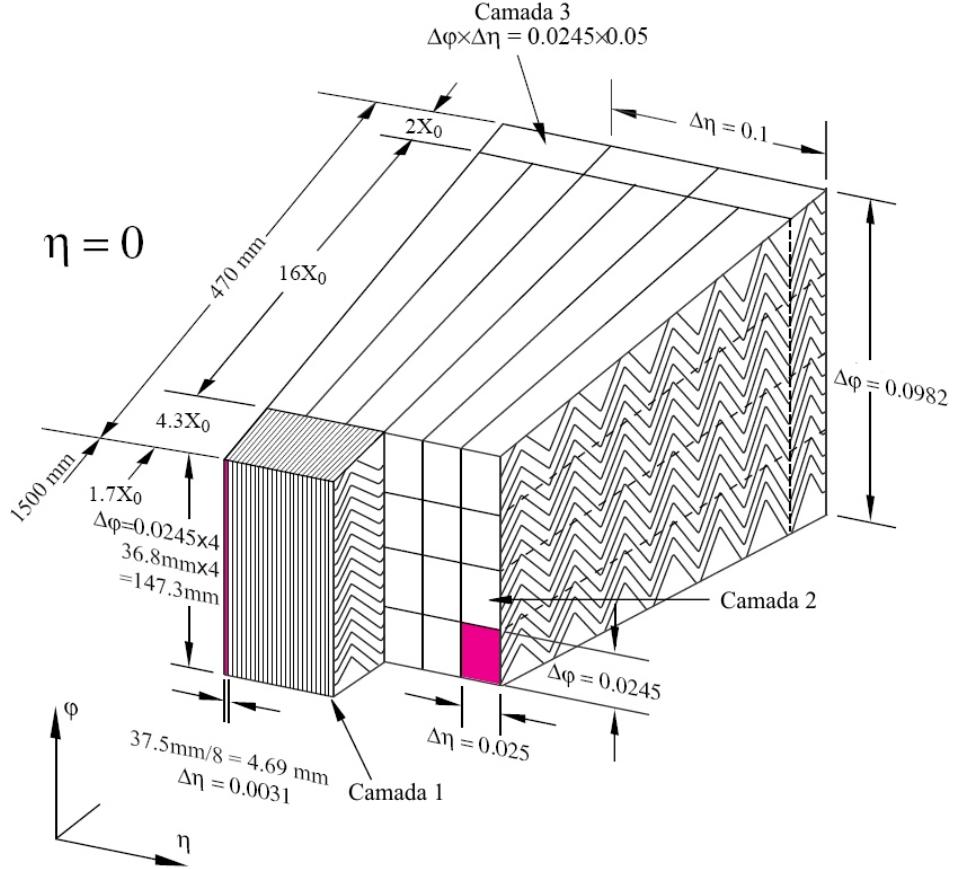


Figura 2.7: Granularidade e profundidade das camadas do calorímetro eletromagnético, extraído de [1].

Calorímetros são construídos utilizando estruturas modulares. Os cabos de transmissão de sinais e alimentação passam por espaços existentes entre os módulos (conhecidos como *cracks*). No ATLAS, na região onde $|\eta| \approx 1,5$ (interconexão entre barril e barril-extendedo) existe uma descontinuidade nos calorímetros para passagem de cabos de alimentação e dados do detector de trajetória e do barril do calorímetro eletromagnético [1]. Nesta faixa, há menor quantidade de células detectoras e, consequentemente, baixa resolução nas medições obtidas (o que pode representar um problema para o sistema de filtragem).

2.5.3 Desempenho esperado dos calorímetros

Devido a limitações construtivas, a precisão nas medições dos calorímetros do ATLAS varia com a energia e a posição de interação. Uma característica particular do ATLAS é a grande quantidade de material instalado entre o ponto de colisão e o sistema de calorímetros, fato que provoca medição incorreta da energia de elétrons, por exemplo, em 119 GeV 5% da energia de elétrons é perdida neste material. A Figura 2.8 ilustra este problema. Uma fina camada detectora (chamada Pré-Amostrador ou *Pre-Sampler* - PS) foi instalada imediatamente antes da primeira camada do calorímetro eletromagnético numa tentativa de estimar a energia perdida pelas partículas antes de chegarem ao calorímetro. Os sinais medidos no PS são ponderados por um fator α e somados aos sinais medidos nas outras camadas para compor a energia total do evento [12]:

$$E_{tot} = E_{calo} + \alpha E_{ps}. \quad (2.6)$$

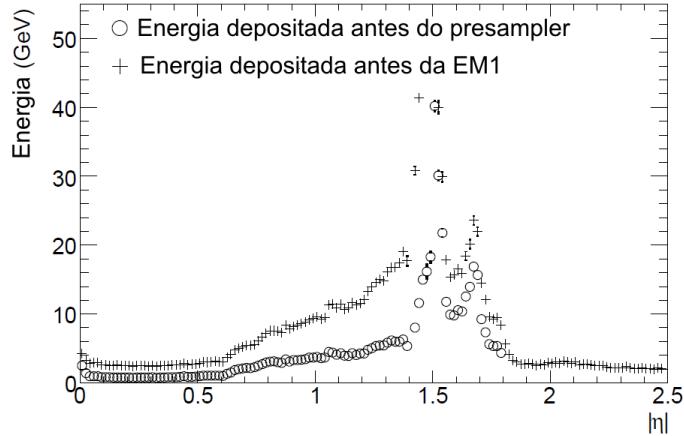


Figura 2.8: Energia perdida por elétrons antes do calorímetro, adaptado de [1].

Um outro fator que pode provocar medição incorreta nos calorímetros é o vazamento de energia de elétrons além da terceira camada eletromagnética. O calorímetro eletromagnético foi projetado para conter todo o chuveiro eletromagnético, porém, principalmente os eventos que insidem na região de transição entre o barril e o barril estendido, por encontrarem menor quantidade de material para interagirem, podem alcançar o calorímetro hadrônico. Essa região é conhecida como *crack*, fica localizada próximo a $\eta = 1.5$ e apresenta um número reduzido de células detectoras.

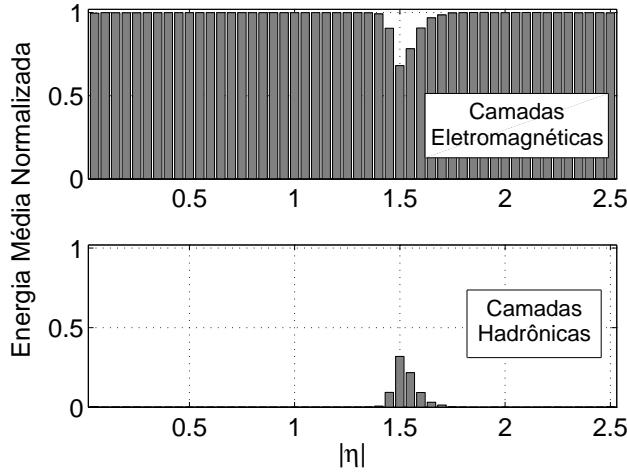


Figura 2.9: Energia normalizada média depositada nas secções eletromagnética e hadrônica em função de η para elétrons.

Conforme ilustrado na Figura 2.9, a energia de elétrons é quase totalmente concentrada no calorímetro eletromagnético, porém, quando a partícula interage próximo ao *crack*, uma parcela significativa de energia chega ao calorímetro hadrônico.

Numa tentativa de corrigir as medições de energia, fatores de ponderação dependentes de η são utilizados na estimativa da energia total do evento [1]:

$$E = s(\eta)[c(\eta) + w_0(\eta).E_{PS} + E_{EM1} + E_{EM2} + w_3(\eta).E_{EM3}] \quad (2.7)$$

sendo s um fator global de correção, c uma tendência (ou *bias*), w_0 corrige a energia perdida antes do calorímetro e w_3 corrige o vazamento longitudinal de energia, os termos E_{PS} , E_{EM1} , E_{EM2} e E_{EM3} representam a energia medida em cada camada do calorímetro eletromagnético (*presampler*, primeira, segunda e terceira camadas eletromagnéticas).

Os erros relativos ($\frac{\sigma_E}{E}$) esperados em função da energia e de η para assinaturas de elétrons são mostrados na Figura 2.10. Percebe-se que a precisão diminui para baixas energias e próximo ao *crack* ($\eta \approx 1.5$). Sempre que o erro de medição aumenta, a caracterização das partículas é prejudicada, provocando queda de desempenho dos algoritmos de filtragem de elétrons.

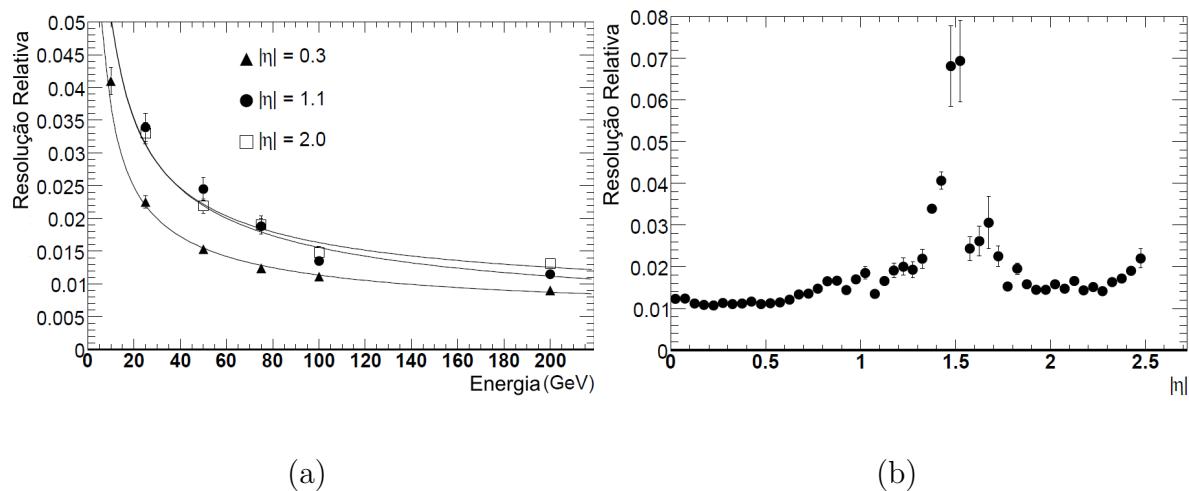


Figura 2.10: Erro relativo do calorímetro na medição da energia de elétrons para diferentes valores de (a) energia e (b) η , adaptado de [1].

Capítulo 3

Filtragem Online no ATLAS

Neste capítulo será apresentada uma breve introdução aos sistemas de filtragem online (*trigger*) em experimentos de física de altas energias (HEP - *High-Energy Physics*) e em seguida serão descritas as características gerais do sistema de filtragem e aquisição de dados (TDAQ - *Trigger and Data Aquisition*) do detector ATLAS.

3.1 Introdução aos sistemas de filtragem em HEP

A maioria dos fenômenos físicos que são estudadas atualmente em experimentos de física de altas energias são raros, pois a grande parte da informação produzida representa processos já conhecidos (identificados e estudados anteriormente em outros experimentos) [47]. Por exemplo, o LHC produzirá uma taxa de eventos da ordem de 10^9 Hz para alta luminosidade ($L = 10^{34}\text{cm}^{-2}\text{s}^{-1}$). Dependendo de sua massa, a taxa de produção esperada para o bóson de Higgs varia entre 10^{-1} e 10^{-2} Hz [40]. Neste caso a frequência de interesse é de 10^{10} a 10^{11} vezes menor que a taxa de eventos produzidos, o que significa dizer que todo o restante da informação produzida representa ruído de fundo para a identificação da física de interesse.

Conforme ilustrado na Figura 3.1, os sistemas de *trigger*, em geral, utilizam diferentes níveis hierárquicos de filtragem, onde os níveis mais baixos operam em janelas de tempo extremamente curtas e são responsáveis pela rejeição de eventos utilizando critérios mais simples e óbvios, enquanto que os níveis mais altos implementam análises mais complexas, pois dispõem de mais tempo para a tomada de decisão. Como os níveis são hierárquicos (sequenciais), uma vez que o evento

foi rejeitado em um dado nível ele não está mais disponível para análise nos níveis posteriores [47].

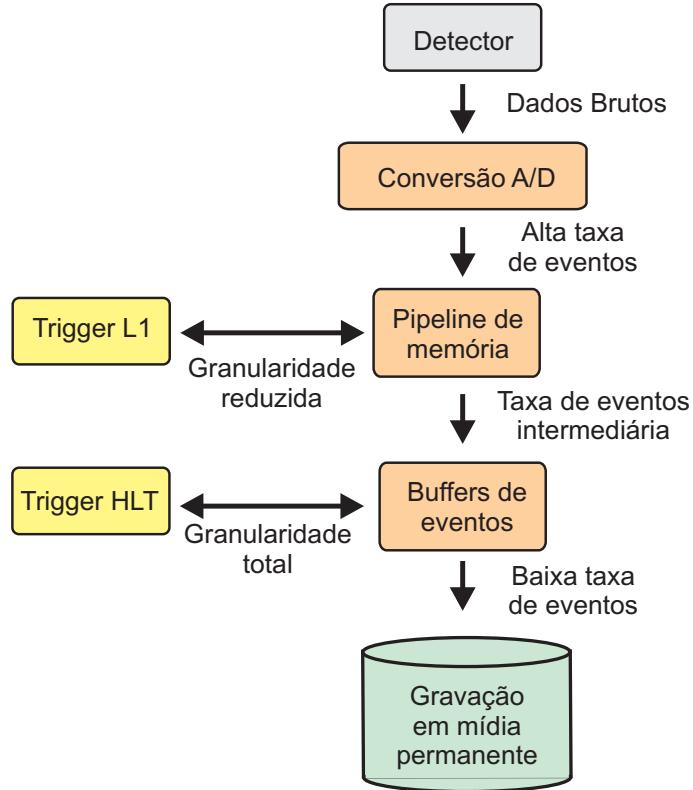


Figura 3.1: Diagrama em blocos do sistema de filtragem genérico para um experimento de HEP.

O primeiro nível de filtragem (*L1 - Level One*) tem disponível um tempo muito curto para tomada de decisão (da ordem de μs), sendo, tipicamente, implementado através de *hardware* dedicado [48], utilizando, por exemplo, FPGA (*Field Programmable Gate Arrays* [49]), DSP (*Digital Signal Processors* [50]) ou ASIC (*Application Specific Integrated Circuit* [51]). Os detectores ATLAS, CMS, CDF [52] e D0 [53], por exemplo, utilizam filtragem de primeiro nível em *hardware* dedicado.

A filtragem de alto nível (*HLT - high-level trigger*) pode ser implementada através de *software* e executado em paralelo por computadores pessoais (PCs). Este estágio dispõe de um maior tempo para produzir a decisão de aceitação ou rejeição e opera com uma menor taxa de eventos (uma vez que uma parcela do ruído de fundo já foi rejeitada pelo primeiro nível). As análises executadas no HLT envolvem operações mais complexas (comparadas ao L1) e podem até requerer a recomposição completa do evento [48]. O detector H1 do acelerador HERA é uma exceção, pois no segundo

nível de filtragem utiliza redes neurais artificiais implementadas em *hardware*, mais detalhes a respeito do sistema de detecção neural do H1 serão fornecidos na próxima seção.

As características de seleção de cada nível variam de acordo com o experimento em questão [47]. Em modernos experimentos de física de alta energia os detectores são divididos em sub-sistemas responsáveis pela detecção de classes específicas de assinaturas. Por exemplo os calorímetros podem ser utilizados para a identificação de candidatos a elétrons, fótons e jatos, já os mísseis necessitam de um detector específico (o sistema de mísseis). O sistema de filtragem utiliza informações destes sub-detectores para identificar as assinaturas da física de interesse.

Considerando os detectores ATLAS e CMS (os dois detectores de propósito geral do LHC), diferentes soluções foram utilizadas para resolver o problema de filtragem e aquisição de dados. No ATLAS, o nível 1 transmite para o nível 2 um sub-conjunto da informação total do detector (conhecido como região de interesse - RoI, do inglês *Region of Interest*) que contém as características necessárias para a identificação da assinatura em questão [54]. O CMS utiliza um sistema de filtragem de primeiro nível que envia toda a informação do evento ao HLT [55].

A utilização das RoI exige inteligência para selecionar adequadamente a informação necessária, mas a grande vantagem é que a taxa de transmissão de dados é reduzida, pois se um evento for rejeitado no nível 2 não há a necessidade de transmitir toda a informação referente ao mesmo. A solução utilizada no CMS exige maior largura de banda para transmissão de dados, porém não há a necessidade de inteligência para efetuar a seleção da RoI. No CMS o problema da largura de banda foi solucionado a partir da utilização de redes de transmissão de dados operando em paralelo e de modo relativamente independente [55].

3.1.1 Aplicações de redes neurais artificiais

As redes neurais artificiais (RNA) [5] de treinamento supervisionado vêm sendo utilizadas em experimentos de física de altas energias, desde o início da década de 1990, com objetivo de auxiliar na detecção *online* dos eventos e também para caracterização dos fenômenos físicos de interesse em análises *offline*. Mais detalhes a respeito das redes neurais artificiais (e sua utilização como classificador supervisio-

nado) podem ser encontrados no Apêndice B.

Em geral, nas aplicações de filtragem *online* são utilizados classificadores neurais na arquitetura Perceptron de Múltiplas Camadas com treinamento supervisionado [5]. A escolha de redes neurais para aplicações em sistemas de *trigger* é motivada pelo fato do problema ser essencialmente de reconhecimento de padrões (identificação das assinaturas de interesse). Uma outra característica particular é que as assinaturas não tem estrutura temporal, representando as medições instantâneas dos diversos sub-detectores. Finalmente, a possibilidade de realizar implementação paralela (seja em *hardware* ou *software*) possibilita a redução do tempo de processamento [56].

O primeiro sistema de filtragem *online* (*trigger*) baseado em redes neurais foi projetado e testado no detector D0 do FermiLab para a identificação da trajetória de mísulas [57, 58]. O *trigger* neural não foi implementado no experimento, mas operou em paralelo com o algoritmo oficial, obtendo uma resolução 40 vezes melhor. Uma rede neural tipo *Perceptron* de Múltiplas Camadas (alimentada adiante) foi utilizada nesta aplicação. Ainda para o detector D0, conforme detalhado no trabalho [59], foram realizados estudos a respeito da aplicação de uma rede neural para a discriminação entre elétrons (produzidos em decaimentos do tipo $Z \rightarrow ee$) e jatos (jatos duplos com alto momento transverso) a partir de dados simulados. Neste trabalho, sinais simulados do perfil de deposição de energia (dados brutos provenientes da região adjacente ao pico de energia) foram utilizados como entradas para a rede neural. O *trigger* neural obteve eficiência de 90 %, desempenho muito superior ao algoritmo padrão do experimento, que identificou corretamente apenas 75 % dos elétrons.

O primeiro sistema de *trigger* neural operando em um experimento de HEP [60] foi implementado no segundo nível de filtragem do detector H1 (um dos detectores do acelerador HERA, que operou no laboratório DESY na Alemanha) [61]. Conforme descrito em [62], um classificador neural, utilizando funções de ativação do tipo sigmoidal, foi treinado para cada canal de filtragem existente no experimento. No H1, o sistema de *trigger* neural foi capaz de produzir a decisão em $8 \mu\text{s}$ e operou em paralelo com um sistema de detecção tradicional (baseado em cortes lineares nos parâmetros de interesse), porém, em alguns momentos, o *trigger* neural operou

sozinho.

Visando minimizar o tempo de processamento, as redes neurais dos detectores D0 e H1 foram implementadas em *hardware* dedicado [63]. Outras aplicações de rede neural implementadas em *hardware* podem ser encontradas em [64, 65, 66].

No experimento HEGRA (*High-Energy Gamma Ray Astronomy*) [67], composto por um conjunto de detectores de raios cósmicos instalados a 2000 metros de altitude, na ilha La Palma, Espanha, uma rede neural foi utilizada para separar eventos de raios cósmicos carregados eletricamente (γ -induzidos) do ruído de fundo composto por chuveiros induzidos por hadrons (que é 100 vezes mais frequente que a assinatura de interesse). O classificador neural foi alimentado por medições de um conjunto de 221 cintiladores, que possibilitam a reconstrução da direção da partícula primária. Diversas topologias de classificadores neurais foram testadas e foi atingida uma rejeição de 92% do ruído de fundo para aceitação de 60% da física de interesse.

Um *trigger* de segundo nível baseado em redes neurais e informação de calorimetria foi proposto para o detector ATLAS no trabalho [8]. Informações do perfil de deposição de energia nos calorímetros foram utilizados para produzir a identificação de elétrons a partir de uma rede neural. Em [68], numa sequencia do trabalho anterior, foi desenvolvido para o segundo nível de filtragem do ATLAS o discriminador *Neural Ringer*, que utilizada informação especialista da física de interesse para pré-processamento dos dados brutos do calorímetro, produzindo-se anéis concêntricos de deposição de energia. A rede neural classificadora (arquitetura MLP) opera sobre os sinais em anéis. Foi obtida eficiência de discriminação superior ao discriminador padrão do ATLAS, que opera a partir de cortes lineares em parâmetros calculados do perfil de deposição de energia. Mais detalhes a respeito do *Neural Ringer* e suas extensões serão fornecidos na seção 4.3.

As redes neurais são aplicadas também na análise *offline* de eventos. Nas aplicações *offline* não existem restrições severas quanto ao tempo de processamento ou risco de “perder” eventos de interesse. Nesse caso trabalha-se com dados gravados em mídia permanente, com o objetivo de estimar, com precisão, as características de cada evento.

No detector ALEPH do experimento LEP, um classificador neural (arquitetura MLP) foi utilizado para separar os quarks em três classes distintas (quarks b, c

e leves). Após testes de desempenho, a topologia escolhida ($20 \times 20 \times 8 \times 3$) utilizou 20 variáveis de entrada, duas camadas escondidas e 3 neurônios na camada de saída (cada um correspondendo a uma das classes) [69]. Neste trabalho, foi obtida eficiência de classificação da ordem de 90 % para cada classe.

Redes neurais artificiais foram utilizadas também para a estimação da massa de top-quarks e eventos de raios cósmicos, respectivamente, no acelerador Tevatron do Fermilab [70] e no observatório Pierre Auger [71]. A massa de top-quarks é extremamente alta se comparado às demais partículas elementares (aproximadamente 40 vezes maior que o segundo quark mais pesado). O motivo para tal característica ainda não está completamente esclarecido na teoria, portanto a medição precisa da massa destes eventos (que são relativamente raros) é muito importante. Com o uso da rede neural em [70], foi obtida a medição mais precisa da massa de top-quarks (até a publicação do referido trabalho) para a assinatura utilizada. Em eventos de raios cósmicos, a estimativa da massa (que é realizada de modo indireto a partir de parâmetros da cascata desenvolvida na atmosfera) é importante na determinação da origem e da natureza dos raios cósmicos primários. No trabalho [71], eventos simulados com energia variando entre 10^{18} e 10^{19} eV foram utilizados para treinamento e teste da rede neural (arquitetura MLP). A rede foi alimentada a partir de parâmetros estimados dos eventos e obteve eficiência da ordem de 99%.

Um exemplo recente de aplicação de rede neural para análise *offline* pode ser encontrado em [72]. Neste trabalho, uma rede *Perceptron* de Múltiplas Camadas foi utilizada para realizar a reconstrução do ponto de interação de raios gama com um detector composto de cristais cintiladores. O treinamento da rede foi realizado de modo supervisionado a partir de eventos experimentais (para baixas energias) e simulados (por *Monte Carlo* para altas energias). Os resultados indicaram que a rede neural foi capaz de realizar a estimativa com alta precisão, superando métodos tradicionalmente utilizados (como os métodos do centroide [73] e da máxima verossimilhança [74]), atingindo precisão menor que 1 mm para baixas energias e aproximadamente igual a 2,1 mm para altas energias.

Mesmo com todos os exemplos de aplicações bem sucedidas de redes neurais, o seu uso está longe de ser uma unanimidade entre a comunidade de HEP. Uma característica particular do campo de aplicação é que há uma busca pelo entendimento de

novos fenômenos, que são representados nos dados simulados (que em geral são utilizados no treinamento dos classificadores) por modelos teóricos aproximados. Um experimento pode concluir que um modelo teórico existente está incompleto ou até mesmo errado, nesse caso o treinamento do sistema ficaria comprometido. O uso de redes neurais é justificado pela facilidade de operação em alta dimensão (inúmeras variáveis são utilizadas no processo de identificação da física de interesse), porém, a dependência dos modelos teóricos é mais difícil de ser verificada ou corrigida nos classificadores não-lineares (em comparação com os cortes lineares mais usualmente utilizados em HEP) [60]. Essa característica particular da HEP talvez tenha produzido uma resistência maior ao uso de redes neurais em comparação a outros campos da ciência. No caso da aplicação no detector H1, toda a colaboração do experimento precisou se convencer que as redes não eram uma caixa preta misteriosa e um esforço conjunto foi feito no sentido de entender como as redes funcionam e como podem ser ajustadas de modo ótimo para cada problema [63]. De um modo geral, o uso de redes neurais está consolidado e bem aceito como ferramenta importante na análise *offline*, porém, no trigger online a situação ainda é ambígua, com grupos a favor e outros contra [60].

3.2 O sistema de filtragem *online* do ATLAS

No sistema de filtragem *online* do ATLAS as estratégias de aceitação de eventos devem garantir que as informações de interesse não sejam perdidas, reduzindo ao máximo a quantidade de ruído de fundo (eventos não-relevantes) gravados em mídia permanente [75].

Considerando as condições de operação do LHC (40×10^6 colisões por segundo, frequencia de interações de 1 GHZ, eventos de interesse raros e imersos em intenso ruído de fundo) e sabendo que, a cada evento (colisão do LHC), no ATLAS são gerados aproximadamente 1,5 MByte de informação, é produzido no detector aproximadamente 60 TBytes de informação por segundo. Com a tecnologia disponível atualmente não é viável armazenar essa quantidade de informação. Mesmo que as leituras de todos os eventos fossem acumuladas, o processo de filtragem *offline* sobre toda essa massa de dados não seria possível. Assim, é necessário um sistema

eficiente de filtragem *online*.

O sistema de filtragem do ATLAS (usualmente chamado de sistema de *trigger*) acessa informação dos três principais sub-detectores, o detector de trajetórias, os calorímetros e a câmara de múons. O *trigger* online opera em 3 níveis sequenciais de seleção de eventos: o nível 1 (L1), o nível 2 (L2) e o filtro de eventos (EF - *event filter*), sendo que os dois últimos juntos compõem a filtragem de alto nível (HLT - *High Level Trigger*) [44]. Cada nível é responsável pela rejeição de uma parcela das assinaturas não relevantes (ruído de fundo), refinando a decisão do nível anterior. A Figura 3.2 ilustra as principais características dos três níveis de filtragem de eventos no ATLAS.

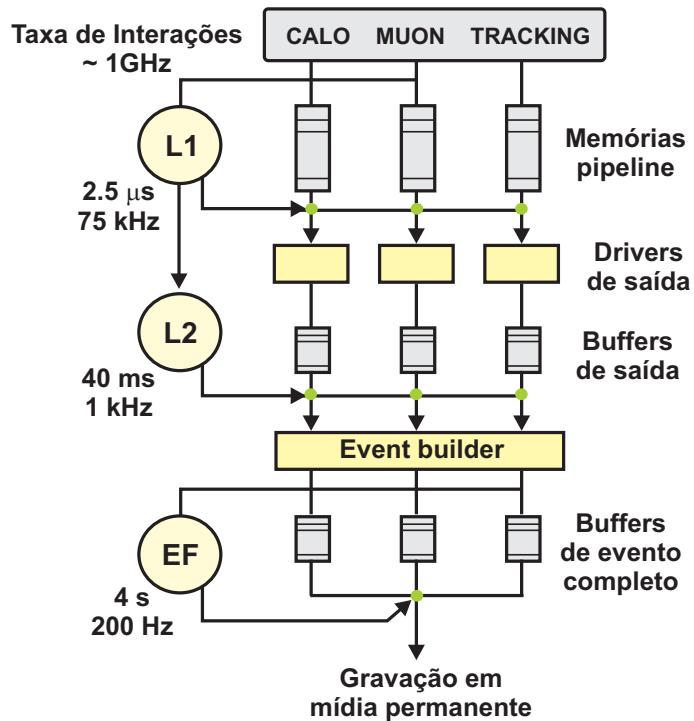


Figura 3.2: Diagrama em blocos do sistema de filtragem e aquisição de dados do ATLAS, extraído de [45].

O primeiro nível tem sérias restrições quanto ao tempo de processamento (latência máxima de $2.5\mu s$), recebendo a plena taxa de eventos do LHC como entrada. Esse nível é implementado em *hardware* dedicado, e para produzir a decisão rapidamente, usa apenas parte da resolução disponível ao detector. A combinação da baixa resolução com critérios de seleção simplificados resulta numa aceitação de parcela considerável do ruído de fundo pelo primeiro nível, que deve ser gradual-

mente reduzida pelos níveis subsequentes. Até a decisão do primeiro nível quanto à aceitação ou rejeição do evento, toda a informação do detector relacionada a esse evento é mantida em um *pipeline* de memórias. O L1 entrega ao segundo nível a localização das áreas onde possivelmente aconteceram eventos de interesse, regiões estas conhecidas como RoI (*Regions of Interest*).

A seleção de eventos no L2 é feita através de *software* especializado, rodando em um conjunto de ≈ 700 PCs (Computadores Pessoais) dedicados operando em ambiente de processamento distribuído [54]. Neste nível, o tempo máximo para tomada de decisão é 40 ms e estão disponíveis informações (que podem ser descarregadas através pelos *drivers* de saída - ROD *Read-Out Drivers*) do detector de trajetórias, das câmaras de mísseis, assim como a total resolução dos calorímetros. Dispondo de um tempo maior e da total resolução do detector, o L2 utiliza critérios de seleção mais refinados (em comparação com o L1), reduzindo consideravelmente o ruído de fundo, sem perder muitas assinaturas de interesse.

O filtro de eventos é o último estágio do sistema de filtragem, recebendo uma taxa de eventos mais baixa, tendo latência de alguns segundos para tomada de decisão. O EF tem acesso a todo o evento (não somente à RoI como no L2) e utiliza técnicas semelhantes às da reconstrução *offline* operando num conjunto de ≈ 2000 PCs [54]. Os eventos aprovados nesse nível são armazenados em mídia permanente para posterior análise *offline*.

A estratégia de processamento sequencial permite que os eventos sejam rejeitados na primeira etapa possível, minimizando a necessidade de acesso a informações, e facilitando o ajuste e eventual modificação das estratégias de extração de características. Na Tabela 3.1 é apresentado um resumo das principais características dos 3 níveis de filtragem do ATLAS.

Um problema que afetará os algoritmos de extração de características é o efeito de empilhamento (do inglês *pile-up*), que ocorre quando há uma sobreposição de assinaturas em regiões do detector [12], ou seja, uma assinatura que ainda se desenvolve tem seu padrão de deposição de energia distorcido por uma nova que chega e se sobrepõe, gerando um ruído de fundo que pode atingir grande intensidade.

Para o projeto e teste dos métodos de extração de características, foram usados conhecimentos prévios adquiridos em outras experiências com aceleradores de

Tabela 3.1: Principais características do sistema de *trigger* do ATLAS, onde Te e Ts são, respectivamente, as taxas de eventos na entrada e na saída e $Cr = Te/Ts$ é o coeficiente de redução de eventos.

Nível	Te (Hz)	Ts (Hz)	Cr	Latência (s)	Implementação
L1	40×10^6	$75 \times (100)10^3$	533 (400)	$2,5 \times 10^{-6}$	<i>Hardware</i>
L2	$75(100) \times 10^3$	$3,5 \times 10^3$	21 (29)	40×10^{-3}	<i>Software</i>
EF	$3,5 \times 10^3$	≈ 200	17	≈ 4	<i>Software</i>

partículas e eventos simulados através de técnicas de Monte Carlo [14]. As simulações utilizam modelos estocásticos que descrevem as interações, levando em conta as características físicas do acelerador e do detector, assim como os efeitos de cada nível de filtragem. Para o ATLAS, foram utilizados geradores de eventos para colisões próton-próton como HERWIG, ISAJET, GEANT e PYTHIA, descritos em [45] e [76]. Os algoritmos de classificação e extração de características foram projetados para os dados simulados e posteriormente adaptados para a realidade de operação quando do início da aquisição de dados.

3.2.1 Primeiro nível de filtragem

O primeiro nível de filtragem (L1) é responsável por reduzir a taxa de eventos de aproximadamente 40 MHz para 75 kHz (a taxa de saída do L1 pode ser aumentada até 100 kHz, a depender das condições de operação do detector), implicando numa redução da ordem de 500 vezes. A decisão do primeiro nível deve ser tomada até $2,5\mu s$ após o cruzamento de feixes (colisão) ao qual o evento está associado.

O L1 identifica as assinaturas básicas de interesse e, para tornar mais rápida a tomada de decisão, a granularidade dos subsistemas do detector é menos fina [75]. Por exemplo, considerando o sistema de calorímetros, que possui mais de 100.000 células detectoras, o L1 utiliza apenas a informação de 7000 “torres” de soma analógicas (as torres são obtidas somando-se a energia de células dentro de regiões de $0,1 \times 0,1$ em $\Delta\eta \times \Delta\phi$) [77].

A tomada de decisão quanto à aceitação ou rejeição de eventos no nível 1 é feita pelo processador central de *trigger* (*central trigger processor* - CTP), que combina as informações disponíveis nos calorímetros, para a detecção de partículas eletro-

magnéticas e hadrônicas, e nos sub-detectores RPC (*Resistive Plate Chamber*) e TGC (*Thin Gap Chambers*) para a detecção de múons, conforme mostrado na Figura 3.3. O filtro dos calorímetros é dividido em três sub-sistemas (Pré-processador, Processador de Regiões e Processador de Soma de Energia/Jato). O pré-processador digitaliza os sinais medidos e envia as informações para os processadores de regiões (CP - *Cluster Processor*) e de soma de energia / jatos (JEP - *Jet/Energy Sum Processor*). O CP é responsável pela identificação dos candidatos a elétrons, fótons e τ -leptons e o JEP identifica candidatos a jatos e produz a soma global de energia do evento. Considerando o filtro de múons, cada um dos sub-detectores é responsável pela identificação dos candidatos a múons em uma região do detector, o RPC opera no barril e o TGC na tampa. Se a assinatura analisada satisfaz um critério de aceitação de algum dos sub-detectores (calorímetros, RPC ou TGC), então ela é aceita pelo L1 e enviada para uma análise mais criteriosa na filtragem de alto-nível.

Até a tomada de decisão de nível 1, toda a informação do evento é armazenada em memórias tipo *pipeline*. Quando um evento é aceito, as informações referentes a este são descarregadas para uso pelo nível 2 de filtragem. As informações dos eventos rejeitados são descartadas (não estando mais acessível para a filtragem de alto-nível).

O L1 é responsável por fornecer ao HLT informações sobre a posição (no plano η, ϕ) onde os eventos aceitos ocorreram, sinalizando assim as regiões de interesse (RoI) para análise no L2. O nível 1 também fornece outras informações importantes, como o critério utilizado para aceitação do evento e a identificação da colisão (cruzamento de feixe) ao qual o evento está associado. O sistema de filtragem do L1 é implementado utilizando hardware dedicado (através de FPGAs) conforme detalhado em [78].

Na Tabela 3.2 são mostradas as freqüências dos principais canais de filtragem esperadas na saída do L1 quando o LHC estiver operando em alta luminosidade ($L=10^{34}\text{cm}^{-2}s^{-1}$). Os principais objetos de *trigger* a serem identificados são candidatos a múons, elétrons/fótons (regiões eletromagnéticas), taus, jatos, e E_T^{miss} . Pode-se perceber que a taxa total desta simulação é da ordem de 40 kHz, aproximadamente 2 vezes menor que a freqüência de saída esperada para o nível 1 na operação do LHC [75].

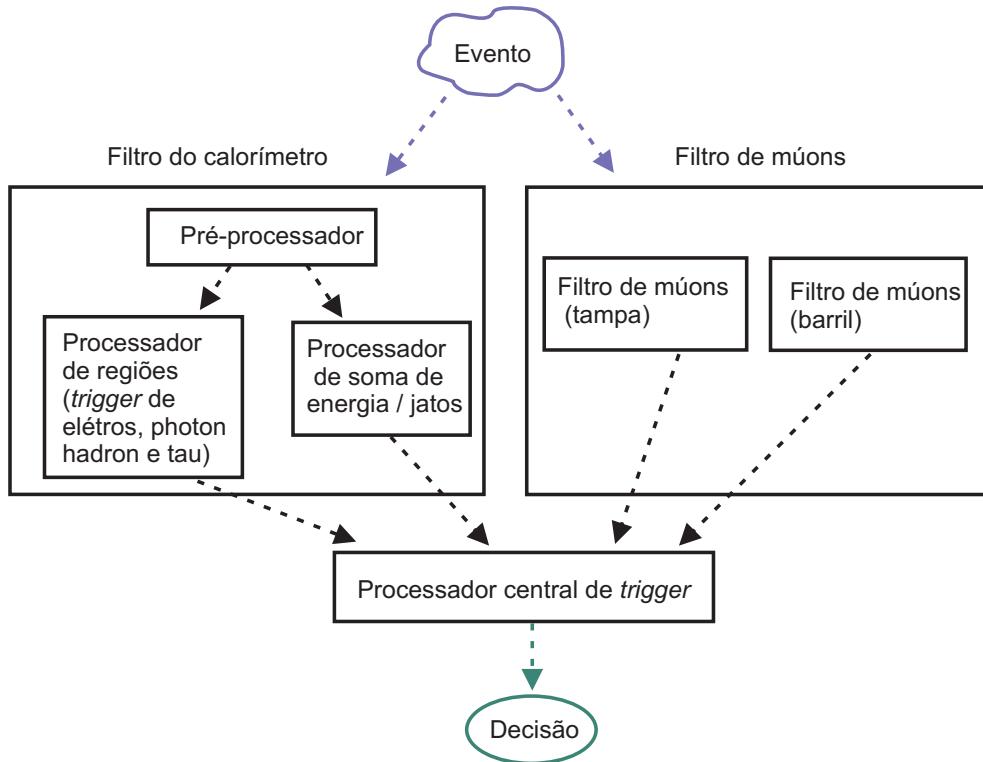


Figura 3.3: Diagrama em blocos do primeiro nível de filtragem.

Tabela 3.2: Freqüência esperada para os principais canais de *trigger* no primeiro nível de filtragem do ATLAS ($L=10^{34}\text{cm}^{-2}\text{s}^{-1}$) , extraída de [75].

Canal	Freqüência (kHz)
Um múon	4
Par de múons	1
Região eletromagnética	22
Par de regiões eletromagnéticas	5
Um jato	0,2
Três jatos	0,2
Quatro jatos	0,2
Jato e E_T^{miss}	0,5
Tau e E_T^{miss}	1
Múon e região eletromagnética	0,4
Outras condições	5
Total	≈ 40

3.2.2 Filtragem de alto nível

O sistema de filtragem de nível 2 (L2) e o filtro de eventos (EF) são responsáveis pela filtragem de alto nível (HLT - *high-level trigger*) do ATLAS. O nível 2 deve reduzir a taxa de eventos de 75 kHz (podendo chegar até 100 kHz) para \approx 3,5 kHz, tendo um tempo de latência de aproximadamente 10 ms para tomar a decisão [76]. O EF precisa diminuir a taxa de eventos de 3,5 kHz para 200 Hz. Os eventos que forem aceitos pelos três níveis de filtragem serão armazenados em mídia permanente para futura análise *offline*. O tempo para a tomada de decisão no filtro de eventos é de alguns segundos. O HLT é implementado em *software* e opera em um conjunto de PCs (\approx 2800) operando em paralelo [54]. Considerando os efeitos conjuntos do segundo nível e do filtro de eventos, a filtragem de alto nível deve reduzir em aproximadamente 500 vezes a taxa de eventos.

O segundo nível opera guiado pelas informações da RoI fornecidas pelo L1. Conforme mostrado na Figura 3.4, após um evento ser aceito pelo L1, as informações (fragmentos) da RoI geradas por diferentes detectores são descarregadas no construtor de RoI (RoIB - *RoI Builder*). Com o uso das RoIs, apenas \approx 2% da informação do detector é necessária para produzir a decisão do L2, reduzindo consideravelmente a taxa de transmissão na rede de dados [54]. O RoIB agrupa os fragmentos e transmite o registro produzido para um supervisor do segundo nível (L2SV - *Level 2 supervisor*), que atribuirá a RoI recebida a uma unidade de processamento do L2 (L2PU - *Level 2 processing units*). A L2PU executa os algoritmos de seleção do segundo nível, que utilizam plena granularidade dos detectores e produzem a decisão (aceite ou rejeição) do L2. Enquanto a decisão do segundo nível é aguardada, todas as informações dos eventos aceitos pelo L1 são armazenadas nos ROBs (*Read-Out Buffers*). Quando a decisão do L2 é produzida, o L2PU informa ao L2SV, que em caso de aceite, envia as informações do evento completo (que estavam armazenadas temporariamente nos ROBs), através da rede do L2, para o Filtro de Eventos (EF). Quando o evento é rejeitado, as informações referentes a ele são descartadas.

As principais características desejadas para os algoritmos de filtragem no L2 são listadas a seguir [76]:

- alta eficiência ($> 95\%$) por RoI selecionada no L1;

- eficiência uniforme em η (o que é difícil pois o detector apresenta discontinuidades) e eficiência uniforme ou crescente com E_T ;
- redução do ruído de fundo minimizando a taxa de eventos classificados de forma incorreta (falso alarme);
- robustez em relação à luminosidade, ruído de medição, imperfeições de alinhamento e calibração.

Quando um evento é aceito pelo L2, o construtor de eventos (EB - *Event Builder*) coleta nos ROBs toda a informação do evento e a disponibiliza ao Filtro de Eventos (EF - *Event Filter*) para a última etapa da filtragem *online*. O evento completo é armazenado nas Fazendas de Entrada do Filtro de Eventos (*EF Sub-Farm Input -SFI*) [40]. O EF reduz a taxa de eventos a 200 Hz tendo disponível até 4 segundos para a tomada de decisão. Os algoritmos do EF analisam todo o evento (não se restringem apenas à RoI como no L2) e operam de modo semelhante à análise *offline* [54]. Quando o evento é aceito pelo EF, todas as informações referentes a ele são armazenadas (gravadas) em mídia permanente.

O HLT foi desenvolvido utilizando, sempre que possível, tecnologias padronizadas (disponíveis comercialmente) [79]. Todos os processadores utilizados são de uso geral (semelhantes aos utilizados em computadores pessoais) e praticamente toda a comunicação é feita através de redes *Gigabit Ethernet* [80]. As aplicações estão sendo desenvolvidas utilizando C++ [81]. Estas escolhas foram feitas considerando-se padronização, velocidade, confiabilidade e facilidade na reposição de equipamentos.

Mais informações sobre os componentes dos sistemas de filtragem podem ser encontradas em [75], [76], [45] e [43].

Desafios do HLT

Conforme mencionado anteriormente, a filtragem de alto nível pode utilizar toda a granularidade e precisão dos subdetectores do ATLAS. Então, percebe-se que, os algoritmos do HLT operam num espaço de decisão multidimensional, com restrições no tempo de processamento e rígidos padrões de eficiência. Deve-se notar também que, as características do sistema de filtragem podem mudar com o melhor conhecimento do detector, após os testes e o início de operação. Percebe-se, então, que

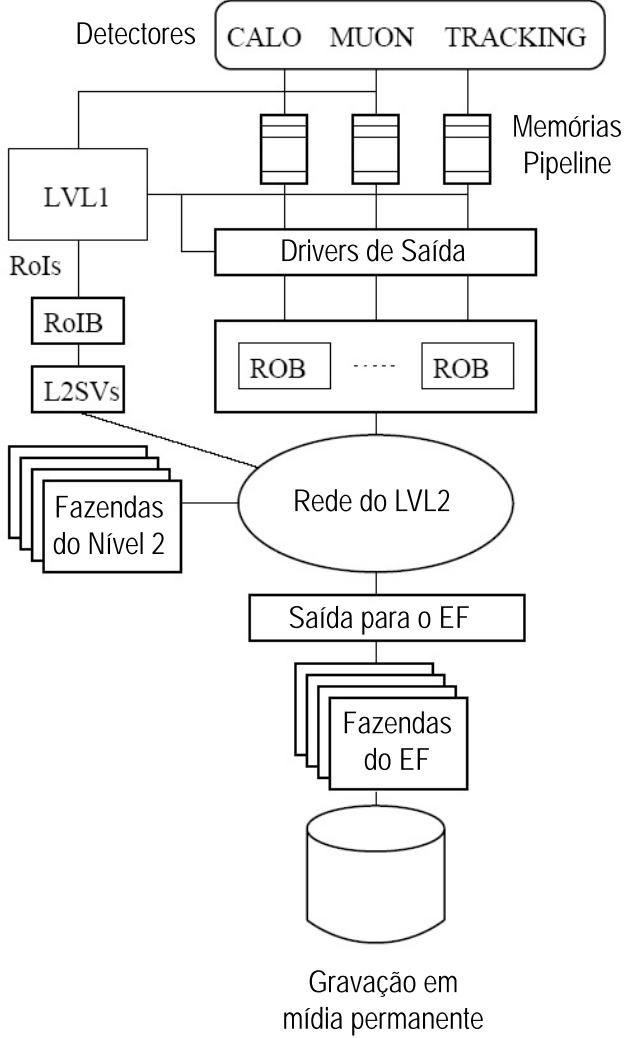


Figura 3.4: Diagrama em blocos do segundo nível de filtragem.

o sistema de *trigger* de alto nível demanda esforços no sentido de propor, testar e comparar diferentes estratégias de seleção.

Embora existam algoritmos desenvolvidos pela colaboração do ATLAS para a filtragem de alto nível (nos diversos canais de interesse), pesquisas continuam sendo conduzidas com o objetivo de propor rotinas de filtragem alternativas que fornecem maior eficiência de discriminação da física de interesse e, ao mesmo tempo, maior rejeição do ruído de fundo.

Neste contexto, se encaixa este trabalho, que visa propor uma otimização ao Neural Ringer [9, 10] que é um discriminador alternativo para o canal elétron/jato e apresenta desempenho superior ao T2Calo (discriminador oficial do ATLAS) e tempo de processamento dentro da janela permitida ao L2 (≈ 40 ms).

No próximo capítulo será descrito o processo de identificação de elétrons a partir de informações de calorimetria. Considerando o L2 do ATLAS, serão descritos os discriminadores T2Calo e *Neural Ringer*.

3.2.3 Plataforma de *Software* do Sistema de Filtragem

A colaboração do ATLAS desenvolveu um conjunto de softwares e ferramentas de controle (conhecidos como Athena) [82] que permitem aos membros da colaboração, independente de sua localização geográfica, acesso e análise dos dados gerados no detector.

O Athena, em sua estrutura modular, dispõe de rotinas para:

- Simulação (incorporando algoritmos como o Pythia, que simula a colisão dos feixes de prótons e o Geant, que simula a interação das partículas com a matéria e o comportamento dos detectores);
- Filtragem (*Trigger*);
- Reconstrução do evento;
- Análise da física.

O núcleo (*kernel*) do Athena é baseado no projeto Gaudi [83], desenvolvido originalmente para o detector LHCb e posteriormente adaptado para as necessidades do ATLAS. O Athena, na verdade, define uma estrutura (*framework*) de controle comum para todos os aplicativos e análises necessárias para a colaboração do ATLAS.

Entre os principais benefícios do uso de uma estrutura comum pode-se mencionar:

- Os desenvolvedores compartilham a mesma estrutura, onde podem inserir os seus próprios códigos, de acordo com sua necessidade específica;
- A comunicação e compatibilidade entre os diversos componentes é mais facilmente garantida;
- Facilidade para re-uso de código, poupando tempo no desenvolvimento de rotinas mais complexas.

O Athena está sendo utilizado no desenvolvimento e teste de algoritmos tanto para seleção de eventos (filtragem *online*), como para reconstrução e análise da física (modo *offline*).

Capítulo 4

Detecção de Elétrons a partir de Informações de Calorimetria no ATLAS

Neste capítulo serão descritos os principais aspectos do processo de filtragem *online* de elétrons no detector ATLAS. O conhecimento adequado da energia de elétrons em uma larga faixa (de alguns GeV a poucos TeV) é necessário para a busca pela partícula de Higgs no decaimento $H \rightarrow ZZ^* \rightarrow 4e$ (conforme previsto no Modelo Padrão) e para medições precisas de fenômenos não descritos pelo Modelo Padrão. A detecção de elétrons é também muito importante para os processos de calibração e alinhamento do detector durante a fase inicial de operação. Com a análise de eventos conhecidos (como por exemplo o decaimento $Z \rightarrow ee$) pode-se gerar informações valiosas para o melhor conhecimento das características do detector [84]. A discriminação destas partículas é baseada, principalmente, nas informações do sistema de calorímetros. Considerando informações do perfil de deposição de energia medido nos calorímetros, os elétrons podem ser confundidos com jatos hadrônicos (que neste caso representam ruído de fundo para o experimento). No LHC, a geração de jatos será bastante intensa (por exemplo, para energia de 40 GeV a relação de produção elétron/jato é da ordem de 10^{-5} [1]) o que tornará difícil a identificação da física de interesse. Tradicionalmente, a identificação de elétrons é realizada analisando-se o formato transversal do chuveiro e o vazamento para as camadas hadrônicas [85]. Estes parâmetros variam com a energia da partícula e nem sempre são capazes de

produzir a eficiência de discriminação desejada.

4.1 Filtragem de elétrons no L1

Considerando as características desejadas para o primeiro nível de filtragem (L1), onde a taxa de entrada é da ordem de 10^6 eventos por segundo e uma taxa de redução da ordem de 500 vezes é exigida em no máximo $25\mu\text{s}$, para tornar mais rápido o processo de decisão, este nível opera com granularidade reduzida no sistema de calorímetros. No L1, as células dos calorímetros são somadas (através de somadores analógicos) para produzir os sinais conhecidos como torres de *trigger* (TT). O tamanho das torres de *trigger* é diferente para as camadas eletromagnéticas e hadrônicas. No calorímetro eletromagnético, cada TT cobre uma área de $0,1 \times 0,1$ no plano $\eta \times \phi$ (ver Figura 4.1), já no calorímetro hadrônico, cada TT representa uma região de $0,2 \times 0,2$ no plano $\eta \times \phi$.

A filtragem de elétrons no L1 é baseada em cortes lineares nos parâmetros do perfil de deposição de energia medido nos calorímetros. A definição dos critérios de seleção leva em conta o conhecimento das características típicas do perfil de objetos eletromagnéticos (elétrons e fótons), que geralmente apresentam:

- concentração ao redor do ponto de colisão (centro da região de interesse - RoI);
- alta concentração de energia na seção eletromagnética;
- baixa concentração de energia na seção hadrônica.

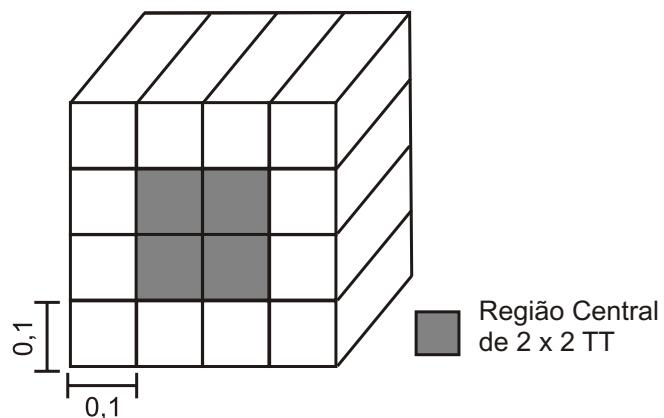


Figura 4.1: Janela deslizante analisada pelo L1 no calorímetro eletromagnético.

A partir da análise de uma janela deslizante, cobrindo uma região de 4×4 TT (16 no total), alguns critérios podem ser utilizados pelo L1 (a depender das características desejadas para a física de interesse pode-se variar os patamares de corte) para definir um possível candidato a elétron:

1. Os valores de energia das TTs da região central (de 2×2 TT) são somados dois a dois e utiliza-se o maior valor encontrado, que deve exceder um patamar de energia eletromagnética para o evento ser aceito neste critério;
2. O nível de energia na periferia (fora da região central de 2×2 TT) é calculado para verificar o isolamento em energia do objeto em questão. Para o evento ser aceito a energia periférica deve ser menor que o patamar estabelecido;
3. A energia depositada nas camadas hadrônicas é somada para verificar o vazamento de energia fora do calorímetro eletromagnético (isolamento hadrônico). Para o evento ser aceito a energia hadrônica deve ser menor que o patamar estabelecido.

Os patamares de seleção do nível 1 podem ser ajustados e os critérios combinados, a depender das características da física de interesse que se deseja estudar num dado momento de operação do detector.

Neste trabalho, foram utilizados sinais simulados do canal elétron/jato aprovados em dois cortes de primeiro nível diferentes. No primeiro conjunto, foi realizado apenas um corte em energia num patamar igual a 10 GeV, esse corte é conhecido como E10 (ou E10-*loose*, para indicar que não houve isolamento). No segundo conjunto, o patamar de corte em energia utilizado foi 15 GeV, e foram realizados também cortes por isolamento (periférico e hadrônico), produzindo os sinais do conjunto E15i (onde o sufixo i indica a utilização de cortes por isolamento). Mais detalhes a respeito das características dos conjuntos simulados E10 e E15i serão fornecidos no Capítulo 7.

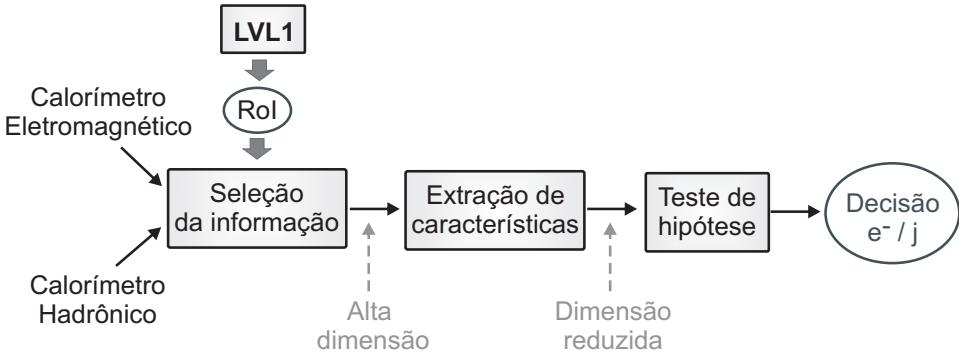


Figura 4.2: Processo de identificação de elétrons no L2.

4.2 Filtragem de elétrons no L2 - Algoritmo T2Calo

Conforme mencionado anteriormente, o segundo nível de filtragem (L2) tem aproximadamente 40 ms para produzir a decisão (aceitação/rejeição) das assinaturas aprovadas pelo L1. O L2 opera apenas sobre a região de interesse (RoI) indicada pelo L1, que no caso da identificação de elétrons envolve os calorímetros e, em alguns casos, o detector de trajetória (*tracking system*).

O processo de discriminação de elétrons no L2 baseado em informações de calorimetria, conforme ilustrado na Figura 4.2, pode ser dividido em duas etapas distintas. Inicialmente é executada a extração de características (onde parâmetros discriminantes são estimados a partir dos dados brutos medidos nos calorímetros) e, em seguida, o teste de hipóteses (onde a discriminação propriamente dita é realizada a partir das características estimadas).

O **T2Calo** [76] é o algoritmo padrão no L2 do ATLAS para a identificação de objetos eletromagnéticos (elétrons ou fôtons) a partir da informação dos calorímetros. O T2Calo utiliza parâmetros que estimam a forma dos chuveiros de deposição de energia, operando de modo semelhante ao algoritmo de filtragem do L1, porém agora, toda a granularidade dos calorímetros está disponível.

O primeiro passo é refinar a posição do centro da RoI fornecida pelo L1, encontrando a célula de maior energia (célula mais quente) na segunda camada do calorímetro eletromagnético (η_1, ϕ_1). Essa posição será posteriormente refinada pelo cálculo da posição da média ponderada da energia em uma janela de 3×7 células

em (η, ϕ) .

Para a seleção dos objetos eletromagnéticos, o T2Calo estima os parâmetros descritos a seguir:

- Razão de Forma: $R_{shape} = E_{3 \times 7}/E_{7 \times 7}$, onde $E_{n \times m}$ é a energia depositada numa janela de $n \times m$ células em torno de (η_1, ϕ_1) na segunda camada eletromagnética;
- Razão de Energia: $E_{ratio} = (E_{1st} - E_{2nd})/(E_{1st} + E_{2nd})$, onde E_{1st} e E_{2nd} são o máximo e o segundo maior valor de energia encontrados na primeira camada eletromagnética numa janela de $\Delta\eta \times \Delta\phi = 0,125 \times 0,2$ em torno do centro da RoI (η_1, ϕ_1) ;
- A energia eletromagnética total: E_{EM} é calculada a partir da soma da energia concentrada em janelas de 3×7 células em torno de (η_1, ϕ_1) nas três camadas eletromagnéticas;
- A energia hadrônica total: E_{HAD} é calculada a partir da soma da energia concentrada em janelas de $\Delta\eta \times \Delta\phi = 0,2 \times 0,2$ em torno do centro da RoI nas três camadas hadrônicas.

Considerando que o perfil de deposição de energia dos elétrons é, em geral, mais concentrado ao redor do ponto de máximo e contido na seção eletromagnética do calorímetro, o T2Calo opera através de cortes lineares nos parâmetros estimados acima. Por exemplo, analisando os parâmetros R_{shape} , E_{ratio} e E_{EM} , os eventos são aprovados caso os valores calculados sejam maiores que um patamar de corte pré-estabelecido. Para E_{HAD} acontece o inverso, e um candidato a elétron é selecionado se o valor calculado para este parâmetro for menor que o patamar.

Embora a concentração em torno do centro da RoI e o vazamento hadrônico sejam características bastante úteis na identificação de elétrons, uma análise um pouco mais detalhada revela que, pelo fato dos jatos apresentarem uma grande flutuação em suas características, o uso apenas destes parâmetros pode não produzir o desempenho de discriminação desejado. Na Figura 4.3, considerando a razão $\frac{E_{HAD}}{E_{EM}}$ (num conjunto de eventos simulados para o L2 do ATLAS), percebe-se que há uma razoável superposição entre elétrons e jatos, essa limitação aumenta em baixa

energia. Comportamento semelhante é verificado na Figura 4.4, onde elétrons e jatos são comparados a partir da razão de forma ($R_{shape} = \frac{E_{3 \times 7}}{E_{7 \times 7}}$).

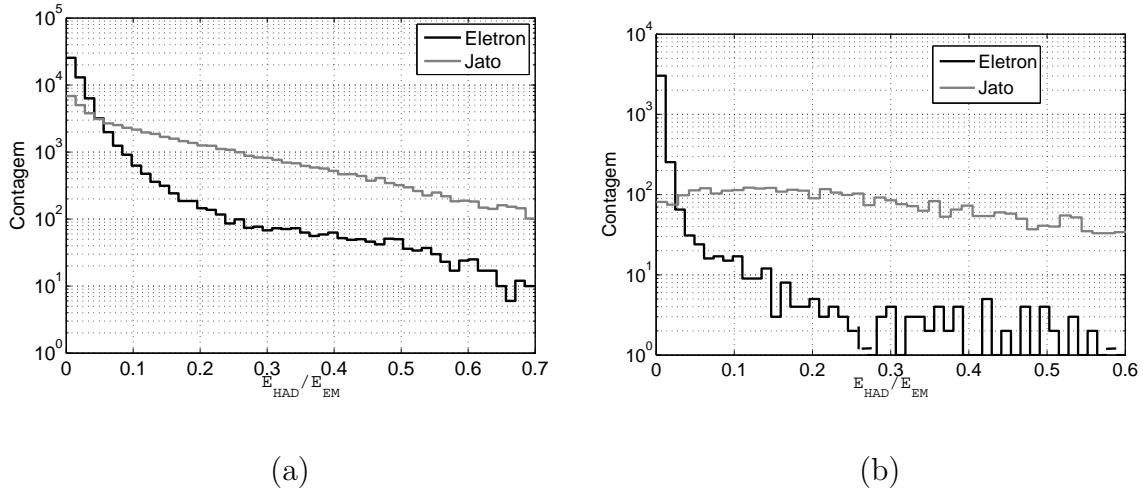


Figura 4.3: Distribuição de elétrons e jatos quanto à razão $\frac{E_{HAD}}{E_{EM}}$ para eventos com energia total (a) menor que 20 GeV e (b) maior que 60 GeV.

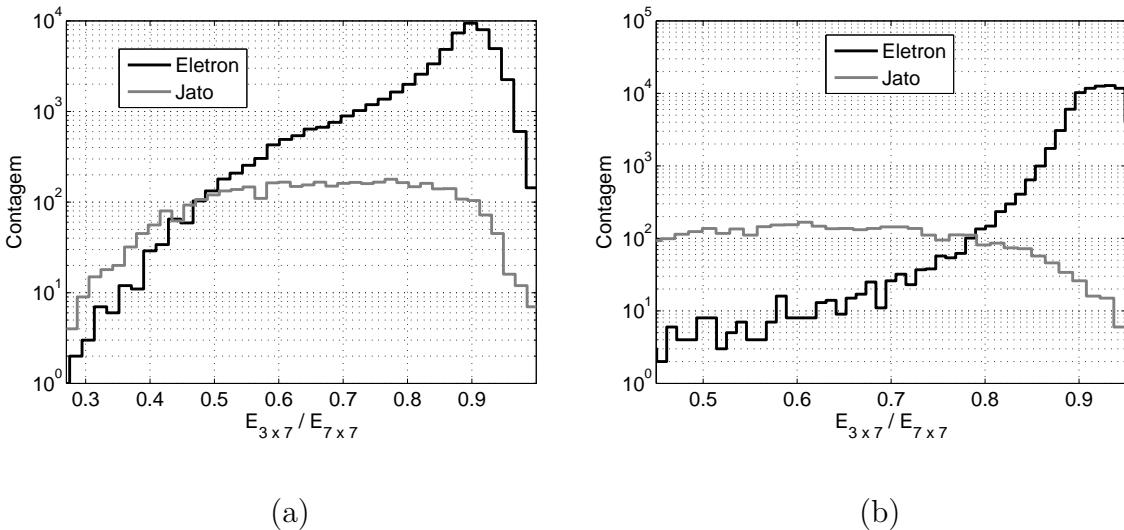


Figura 4.4: Distribuição de elétrons e jatos quanto à razão de forma ($R_{shape} = \frac{E_{3 \times 7}}{E_{7 \times 7}}$) para eventos com energia total (a) menor que 20 GeV e (b) maior que 60 GeV.

Num ambiente complexo, como o que é gerado a cada colisão do LHC, onde o nível de ruído de fundo é extremamente alto e os eventos de interesse são raros, a busca por algoritmos de filtragem mais eficientes é muito importante. Aumentar o desempenho de discriminação significa menor quantidade de eventos não relevan-

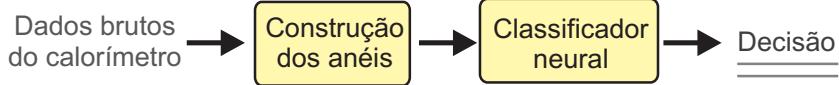


Figura 4.5: Fluxo de processamento do *Neural Ringer*.

tes armazenados mídia permanente, economizando recursos e facilitando a análise *offline*.

A seguir será descrito o *Neural Ringer*, que é um discriminador alternativo para o canal elétron/jato no L2 do ATLAS. Utilizando informação da física de interesse para o processo de extração de características e um classificador neural supervisionado para o teste de hipóteses, o *Neural Ringer* apresenta eficiência superior ao T2Calo e tempo de execução dentro da janela permitida para o L2.

4.3 Neural Ringer - Alternativa para Filtragem de Elétrons no L2

Um discriminador alternativo foi proposto inicialmente em [8] para o canal elétron/jato no segundo nível de filtragem do ATLAS. Conforme ilustrado na Figura 4.5, o processo de extração de características compreende um arranjo topológico dos sinais medidos no calorímetro (com a construção de anéis concêntricos a partir do perfil de deposição de energia), e o teste de hipóteses é realizado por um classificador neural supervisionado (arquitetura *Perceptron* de Múltiplas Camadas) [5]. Este discriminador ficou conhecido como *Neural Ringer*.

4.3.1 Extração de Características - Anelamento

Os sinais do perfil de deposição de energia utilizados para discriminação elétron/jato são medidos nas sete camadas dos calorímetros do ATLAS. Considerando a granularidade de cada camada, uma Região de Interesse (RoI) típica (de tamanho $0,4 \times 0,4$ em $\eta \times \phi$) é descrita por aproximadamente 1000 células [9].

Visando compactar os sinais dos calorímetros e, ao mesmo tempo, manter a interpretação física do perfil de deposição de energia, as células sensoras de cada camada são agrupadas em anéis concêntricos de deposição de energia. Conforme

ilustrado na Figura 4.6, o procedimento de geração dos sinais em anéis é executado em cada camada do calorímetro e envolve os passos a seguir:

1. Seleção das regiões de interesse ($0,4 \times 0,4$ em $\eta \times \phi$ ao redor da célula mais energética);
2. A célula mais energética da camada é considerada como o primeiro anel (e consequentemente o “centro” da ROI);
3. As células ao redor do primeiro anel (imediatamente adjacentes) definem o segundo anel e assim sucessivamente (até cobrir toda a janela da ROI);
4. A energia amostrada pelas células pertencentes a um dado anel são somadas produzindo o sinal de energia em anéis;
5. Este processo é repetido para todas as sete camadas do calorímetro;
6. Os sinais em anéis formados a partir das diferentes camadas são concatenados em um único vetor e normalizados.

É interessante notar que, devido à diferença de granularidade entre as camadas do calorímetro, um número diferente de anéis é gerado para cada uma delas. A tabela no final da Figura 4.6 mostra a distribuição dos anéis por camada. No processo de formatação é considerado um número fixo de anéis por camada, porém, dependendo da posição do centro da ROI, esta pode não estar completa (as células sensoras que compõem os anéis podem não existir caso a partícula interaja próximo a uma extremidade ou descontinuidade do calorímetro), neste caso, assume-se igual a zero as leituras das células que faltam. Outra característica a ser observada é que, em algumas camadas, as células não são quadradas, neste caso, falar em anéis é uma abstração (pois ao final do processo descrito acima podem não existir realmente anéis de células sensoras).

São mostrados na Figura 4.7, considerando a segunda camada eletromagnética (E2), os sinais medidos para um elétron e um jato típicos e os respectivos sinais em anéis gerados para esta camada. A Figura 4.8 mostra sinais em anéis (após concatenação de todas as camadas e normalização), respectivamente, para um elétron

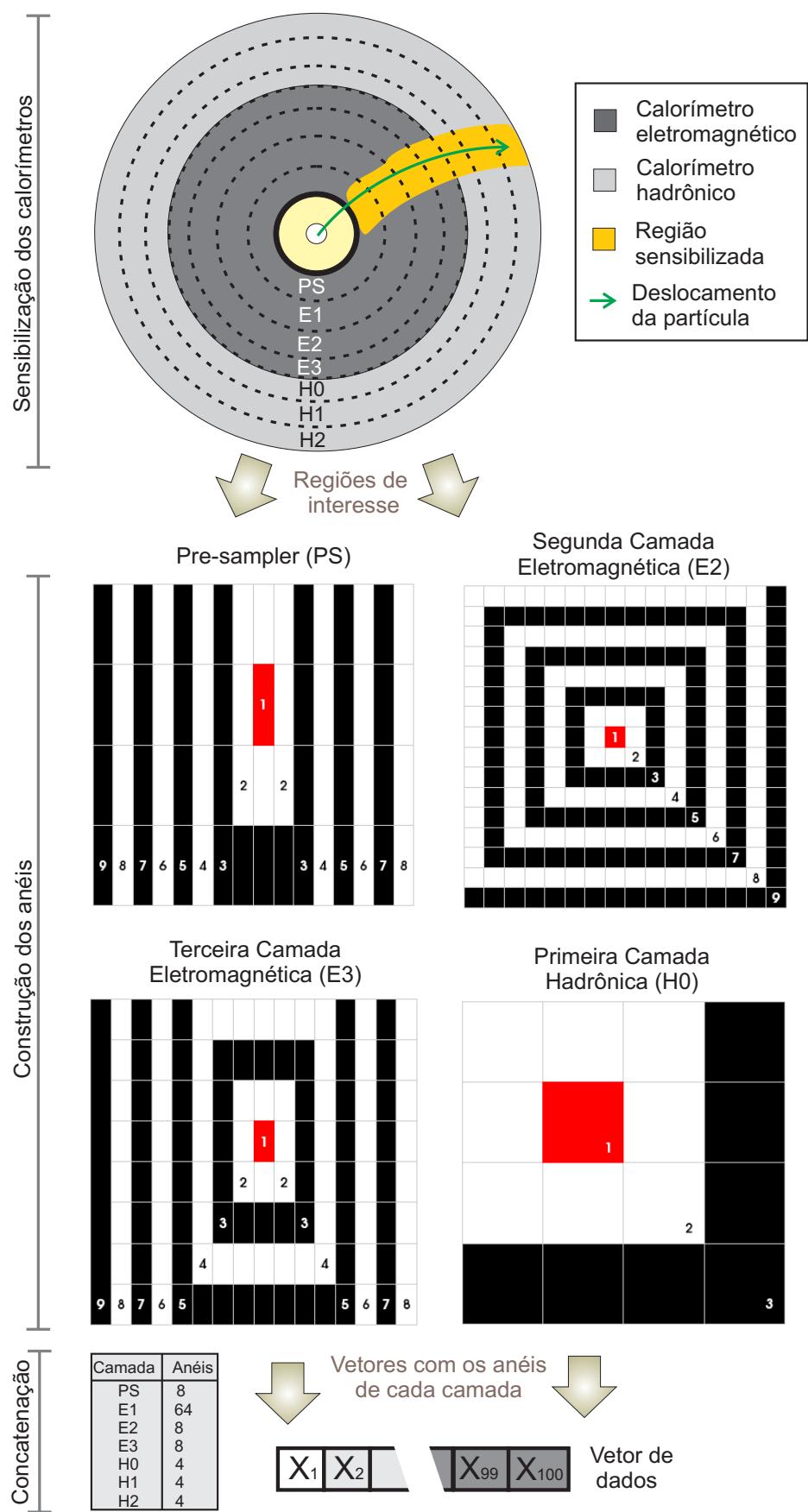


Figura 4.6: Diagrama do processo de construção dos anéis.

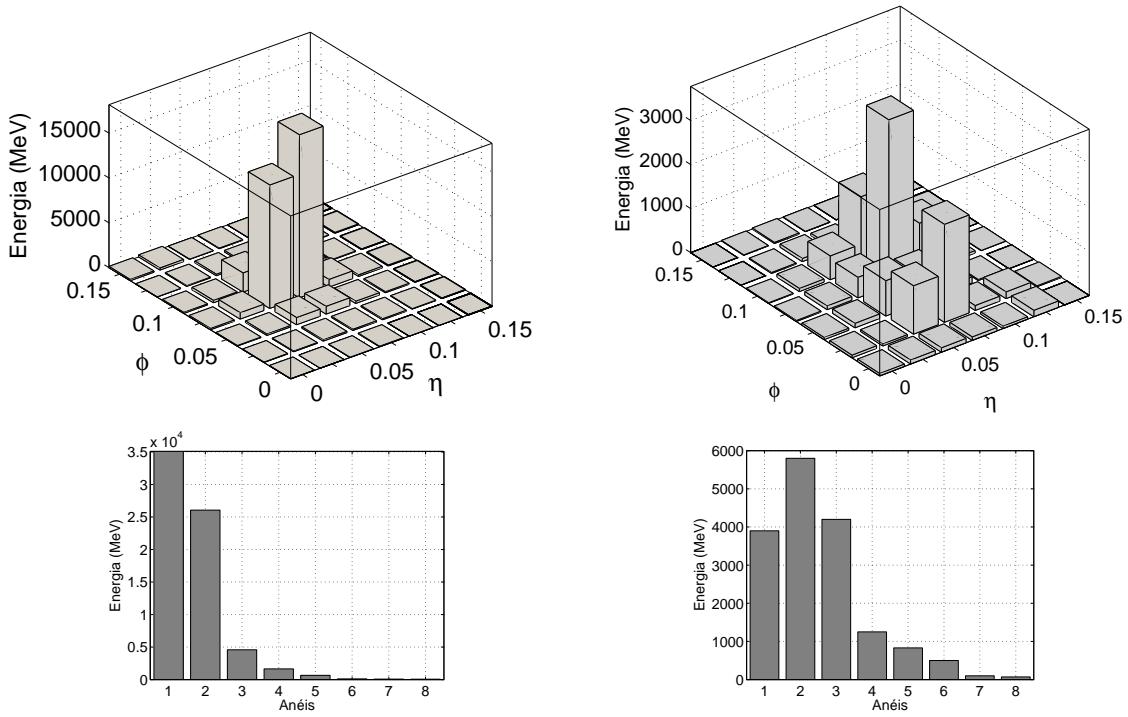


Figura 4.7: Sinais típicos medidos na camada EM2 (acima) e os respectivos sinais em anéis (abaixo), respectivamente para um elétron (esquerda) e um jato (direita).

típico, um jato típico e um jato com perfil semelhante ao de elétrons. Percebe-se que, o perfil de deposição medido para elétrons típicos apresenta pouco espalhamento (é contido em uma pequena região) e é concentrado nas camadas eletromagnéticas. Os jatos apresentam uma maior variação em suas características físicas. Essas partículas tipicamente depositam uma considerável quantidade de energia nas camadas hadrônicas e possuem perfil de deposição com pouca concentração espacial, ou seja, mais espalhados (ver Figura 4.8-b). Entretanto, alguns jatos (como o mostrado na Figura 4.8-c) podem apresentar perfil de deposição de energia semelhante ao de elétrons. Esse último tipo de assinatura hadrônica representa a maioria dos jatos aprovados como elétrons pelo L1, compondo um ruído de fundo de difícil identificação.

4.3.2 Normalização

Antes da utilização nos sistemas de classificação os sinais em anéis precisam ser normalizados. No trabalho [10], foram testadas diferentes formas de normalização

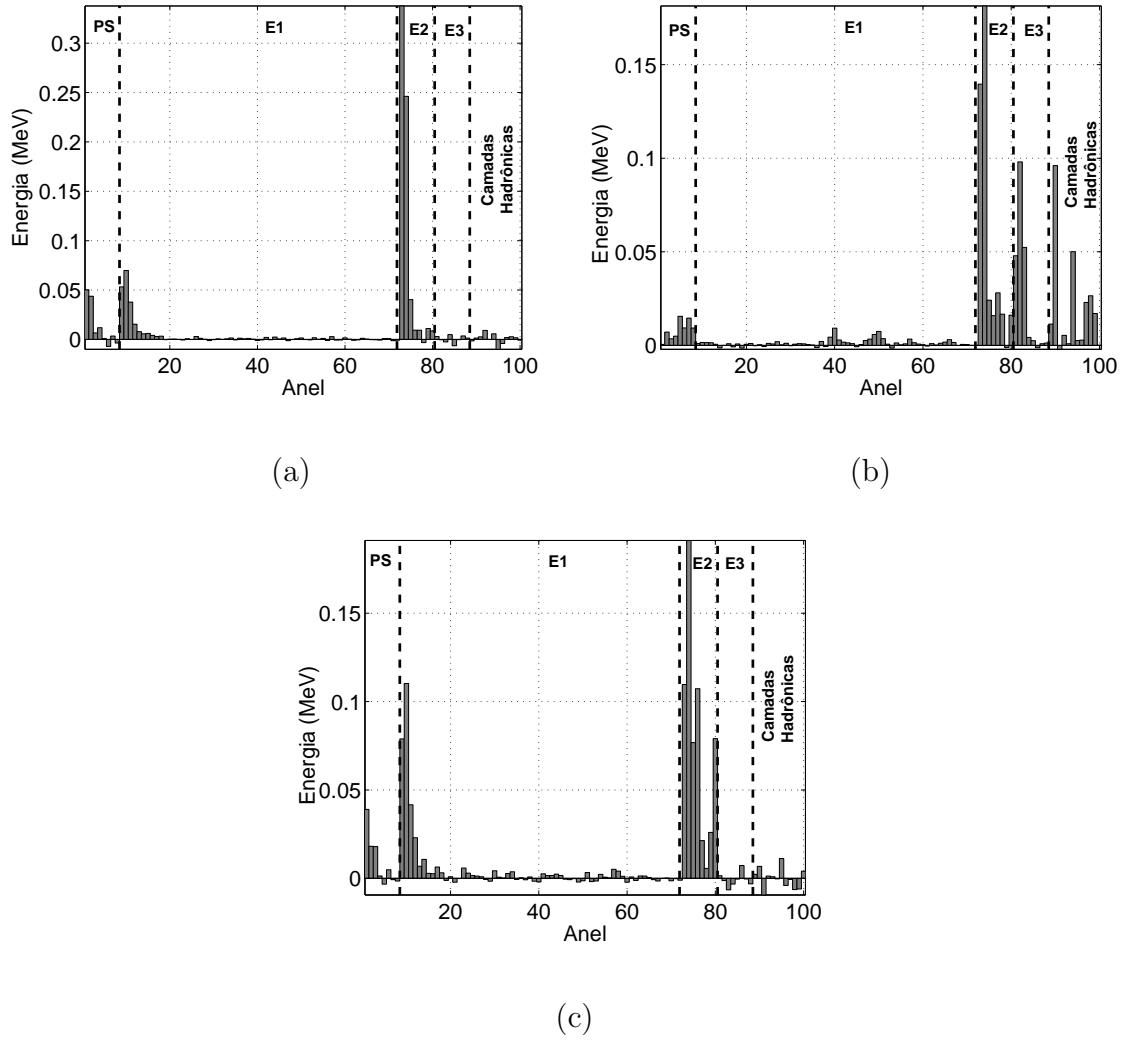


Figura 4.8: Sinais em anéis para (a) elétron típico, (b) jato típico e (c) jato com perfil semelhante ao de elétrons.

dos sinais em anéis como:

- **Energia total:** dividindo-se a energia de cada anel (Ea_i) pela energia total do evento:
$$E_{Ni} = \frac{Ea_i}{\sum_{i=1}^{100} Ea_i}. \quad (4.1)$$
- **Energia da camada:** dividindo-se a energia de cada anel pela soma da energia da camada a qual ele pertence;
- **Energia da seção do calorímetro:** dividindo-se a energia de cada anel pela soma da energia na seção do calorímetro (eletromagnética ou hadrônica) que contém o referido anel.

- **Sequencial:** esse procedimento de normalização visa amplificar as diferenças no perfil lateral dos chuveiros (realçando a energia distante do centro da ROI, que em alguns casos, pode ser útil para a discriminação). A energia do anel i da camada c (Ea_{ic}) é normalizada através de:

$$E_{Ni}^{\{c\}} = \frac{Ea_i^{\{c\}}}{E^{\{c\}} - \sum_{j=1}^{i-1} Ea_j^{\{c\}}} \quad (4.2)$$

onde $E^{\{c\}}$ é a energia total da camada c.

Após um estudo comparativo, concluiu-se que a normalização por energia total atendia um compromisso entre rapidez de execução e eficiência de discriminação, sendo assim, este procedimento foi adotado em todos os resultados simulados a serem mostrados nessa tese. A normalização sequencial produz eficiência semelhante à normalização por energia total e, embora tenha custo computacional um pouco mais elevado, é o procedimento em operação no detector atualmente. Portanto, nos conjuntos de dados experimentais (medidos no detector) este é o procedimento de normalização utilizado.

4.3.3 Teste de Hipóteses - Classificador Neural

Para o processo de discriminação propriamente dito (teste de hipóteses), o Neural Ringer utiliza um classificador neural supervisionado na arquitetura *Perceptron* de Múltiplas Camadas (MLP *Multi-Layer Perceptron*) [5]. Os classificadores neurais são capazes de produzir superfícies não-lineares de separação, contribuindo, em geral, para aumento na eficiência de discriminação quando comparados aos discriminadores lineares. Devido a sua estrutura paralelamente distribuída, uma rede neural é capaz de responder rapidamente quando um padrão de entrada é aplicado. Considerando as características expostas (alta eficiência e processamento veloz), os classificadores neurais se tornam uma opção interessante ao ambiente do segundo nível de filtragem *online* do ATLAS. Mais detalhes a respeito dos classificadores neurais na arquitetura MLP podem ser encontrados no Apêndice B.

4.3.4 Tempo de Execução

Um aspecto crucial nos algoritmos para o sistema de filtragem *online* é o tempo de execução. No caso do *Neural Ringer*, que é um algoritmo para o segundo nível de filtragem, a janela de tempo máxima permitida é de 40 ms. Uma análise detalhada do tempo de processamento requerido pelo *Neural Ringer* foi realizada no trabalho [10]. O algoritmo foi incorporado como uma sub-rotina do T2Calo, aproveitando parte do processamento realizado por este último algoritmo.

As diversas etapas necessárias para a execução do *Neural Ringer*, e seus respectivos tempos de execução no ambiente do L2, são listadas a seguir:

1. Inicialmente, são selecionadas as informações necessárias, o que é feito na etapa de seleção de região (*Region Selector*) do T2Calo, que solicita os fragmentos da RoI desejada. Essa etapa, por envolver um intenso fluxo de informação entre o pipeline de memória e o segundo nível de filtragem, corresponde a 40.7 % do tempo total (em média $0,4927 \pm 0,0787$ ms).
2. A seguir, o T2Calo realiza um pré-processamento nos sinais, onde são calculadas as suas 4 variáveis de decisão e os valores de E_T , η e ϕ com uma maior precisão que no L1. Este passo de processamento equivale a 11.6 % do tempo total de execução ($0,1408 \pm 0,0148$ ms). A partir daqui é iniciada a sequência de rotinas específicas do *Neural Ringer*.
3. Apenas parte da informação necessária para o *Neural Ringer* é utilizada pelo T2Calo, então, o restante precisa ser solicitado neste momento pelo *Region Selector* do *Neural Ringer*. Um novo fluxo de informação é estabelecido entre o L2 e as memórias de armazenamento temporário, consumindo cerca de 36 % do tempo total ($0,4375 \pm 0,0996$ ms).
4. Na etapa de pré-processamento do *Neural Ringer*, as informações da RoI são utilizadas na construção dos anéis para cada camada do calorímetro (conforme procedimento descrito na Seção 4.3.1). O tempo gasto nesta etapa corresponde a aproximadamente 8% do total ($0,0986 \pm 0,0165$ ms). Finalmente, a classificação dos sinais em anéis pela rede neural requer aproximadamente 3% do tempo total ($0,0387 \pm 0,0018$ ms).

Contabilizando todas as etapas de processamento chega-se a um tempo total de $1,2352 \pm 0,1288$ ms. Somando-se a isso a análise do perfil da trajetória da partícula medido no detector de traço (a identificação de elétrons no L2 pode envolver a análise de sinais de dois sub-detectores, os calorímetros e os detectores de traço), chega-se a $8,04 \pm 2,04$ ms, valor que ainda está muito abaixo do tempo máximo exigido para o nível 2, que é de 40 ms. Na Tabela 4.1, são mostradas as diversas etapas necessárias para a execução do *Neural Ringer*, juntamente com o tempo requerido por cada uma delas (em valores absolutos e também percentuais).

Tabela 4.1: Tempo gasto nas diversas etapas necessárias para a execução *online* do *Neural Ringer*.

Etapas	Tempo (ms)	% do Total
<i>Region Selector - T2Calo</i>	$0,4927 \pm 0,0787$	40,7
Pré-processamento - T2Calo	$0,1408 \pm 0,0148$	11,6
<i>Region Selector - Ringer</i>	$0,4375 \pm 0,0996$	36,1
Pré-processamento (anelamento) - <i>Ringer</i>	$0,0986 \pm 0,0165$	8,2
Normalização	$0,0026 \pm 0,0015$	0,2
Classificação Neural	$0,0387 \pm 0,0018$	3,2
Total	$1,2109 \pm 0,1288$	100

Embora o *Neural Ringer* tenha, em comparação ao T2Calo, produzido um aumento no tempo de execução da ordem de 91% (de $0,6469 \pm 0,0802$ ms para $1,2352 \pm 0,1288$ ms, ou seja um aumento de 0,5883 ms), isso é compensado por uma eficiência de discriminação muito superior do *Ringer*, conforme será mostrado no Capítulo 8.

O excesso de tempo do *Ringer* em relação ao T2Calo é compensado no processamento do Filtro de Eventos, pois, como uma quantidade significativamente menor de ruído de fundo é enviado para o último nível, um menor tempo será gasto nesta etapa.

4.4 Extensões ao Neural Ringer

Com eficiência mais alta que o algoritmo padrão (T2Calo) e tempo de processamento dentro da janela aceita para o L2, o discriminador *Neural Ringer* mostrou-se uma opção bastante interessante para o problema da filtragem *online* de elétrons



Figura 4.9: Fluxo de processamento das extensões ao *Neural Ringer*.

no ATLAS. Em sistemas neurais de classificação, sabe-se que o uso de uma etapa de pré-processamento adequada pode ser decisivo para o aumento da eficiência de discriminação [86]. Neste contexto, trabalhos vêm sendo desenvolvidos com o objetivo de melhorar a eficiência do Neural Ringer com a adição de uma etapa de pré-processamento ao classificador neural (ver Figura 4.9), na qual, os sinais em anéis são mapeados em um conjunto (em geral mais compacto) de características discriminantes. A seguir, serão descritos a importância da etapa de pré-processamento em classificadores neurais (em um caso genérico) e os trabalhos desenvolvidos neste sentido no contexto do projeto *Neural Ringer*.

4.4.1 Importância do pré-processamento para classificadores neurais

Embora as redes neurais possam, essencialmente, realizar quaisquer mapeamentos não-lineares, pode-se observar que, o uso direto dos dados brutos como entradas geralmente produz um desempenho pior do que quando algum pré-processamento é aplicado [86].

Da teoria da informação, considerando duas variáveis \mathbf{x} e c (respectivamente o conjunto de características e os rótulos de classes), para qualquer transformação determinística $T(\cdot)$, a informação mútua (I) entre $T(\mathbf{x})$ e c tem como limite superior a informação mútua entre \mathbf{x} e c [87]:

$$I(T(\mathbf{x}); c) \leq I(\mathbf{x}; c). \quad (4.3)$$

Ou seja, nenhuma transformação é capaz de acrescentar informação, ao conjunto de características \mathbf{x} , a respeito das classes c .

Considerando o resultado da Equação 4.3, o que justifica o melhor desempenho dos classificadores após a etapa de pré-processamento, uma vez que ela não é capaz de adicionar informação a respeito do problema?

A resposta pode ser encontrada considerando-se que, os sistemas neurais de classificação são obtidos a partir de um processo iterativo de ajuste de pesos. Então, um pré-processamento eficiente, ou seja, capaz de revelar características discriminantes que estavam inicialmente ocultas nos dados brutos, pode se tornar decisivo para a obtenção de um classificador com melhor desempenho.

É importante notar que, todas as informações utilizadas para a discriminação estão presentes nos dados brutos. O pré-processamento, que neste contexto é conhecido como extração de características, é responsável apenas por uma transformação que torna as características discriminantes mais evidentes.

Um outro aspecto a ser analisado é a chamada “maldição da dimensionalidade” (do inglês *curse of dimensionality*) [86]. Sabe-se que, a utilização de um número elevado de entradas para o sistema classificador acaba dificultando o processo de treinamento. Isso ocorre pois, quanto maior a dimensão dos dados de entrada, maior a complexidade da rede neural (e consequentemente maior o número de parâmetros a serem ajustados), requerendo também maior informação estatística (exemplos de treinamento) para o treino adequado do classificador.

A solução adotada, na maioria dos casos, é pré-processar os sinais com uma transformação que reduza a dimensionalidade do problema. Pois, entre os atributos disponíveis, alguns podem apresentar informações redundantes ou até mesmo irrelevantes para o processo de classificação, dificultando o processo de treinamento dos classificadores. Porém, é preciso escolher adequadamente a informação a ser descartada para minimizar a chance da perda de características relevantes para a discriminação das classes.

Resumindo, a transformação ótima deve, entre outras coisas, mapear os atributos disponíveis em um número reduzido de características, eliminar a redundância e manter toda a informação discriminante para o problema.

4.4.2 Pré-processamento linear

Voltando para o contexto do discriminador *Neural Ringer*, embora a construção dos anéis já seja responsável por uma considerável redução de dimensão (por um fator de 10 vezes), no trabalho [88] foi realizado um estudo detalhado a respeito da utilização da Análise de Componentes Principais (PCA - *Principal Component Analysis*) [3]

para compactação. A PCA foi aplicada aos sinais em anéis e o classificador neural treinado a partir dos componentes principais estimados. Com a PCA, foi alcançado um fator de redução de aproximadamente 4 vezes, mantendo desempenho semelhante ao Neural Ringer tradicional. Visando explorar toda a granularidade e segmentação disponível aos calorímetros do ATLAS, foi proposta a realização do processo de compactação de modo segmentado (a nível de cada camada do calorímetro).

No trabalho [10], foi utilizada uma técnica de compactação mais adequada a problemas de classificação, as Componentes Principais de Discriminação (PCD - *Principal Components for Discrimination*) [11]. Foi realizado também um estudo referente à aplicação da Análise de Componentes Independentes (ICA *Independent Component Analysis*) [4] como uma etapa adicional de pré-processamento, aparecendo logo após a compactação. Os resultados mostraram que a combinação de ICA e PCD foi capaz de melhorar o desempenho em relação ao Neural Ringer tradicional, sem modificar significativamente o tempo de processamento necessário para a tomada de decisão.

4.4.3 Pré-processamento não-linear

Conforme descrito no Capítulo 2, os calorímetros, embora projetados para serem detectores lineares na identificação de elétrons, na prática, seu comportamento pode apresentar características não-lineares, devido a fenômenos como saturação de sensores, atenuação da luz (em calorímetros cintilantes), etc. Então, é possível que uma técnica não-linear de extração de características seja mais adequada ao problema.

Os estudos desenvolvidos no contexto desta tese, tiveram o objetivo de investigar a aplicação da versão não-linear da análise de componentes independentes (NLICA - *Nonlinear Independent Component Analysis*) [13, 89] para extração de características, como uma etapa de pré-processamento ao *Neural Ringer*.

Considerando que existem diversos algoritmos para a NLICA, neste trabalho foram realizados estudos com alguns destes métodos. Os diferentes modelos foram comparados, considerando eficiência de discriminação e tempo de processamento necessário para produzir a decisão. No próximo capítulo serão mostrados os fundamentos teóricos da análise de componentes independentes nos seus modelos linear e não-linear, bem como suas aplicações em problemas de extração de características.

Capítulo 5

Análise de Componentes Independentes

Neste capítulo, serão mostradas, de modo resumido, as principais características dos modelos linear e não-linear da análise de componentes independentes (respectivamente ICA e NLICA). Também serão brevemente apresentados os algoritmos utilizados para estimação da NLICA no contexto deste trabalho. A seguir, será realizada uma revisão bibliográfica das aplicações da análise de componentes independentes em problemas de extração de características, dando enfoque especial a Física de Altas Energias (HEP - *High-Energy Physics*). Em complementação a este Capítulo, no Apêndice A, são apresentados, de modo mais detalhado (com um maior formalismo matemático), alguns dos algoritmos mais utilizados na literatura para ICA/NLICA, juntamente com os diversos parâmetros empregados nesta rotinas para estimar o grau de independência entre os sinais.

5.1 Modelo linear da ICA

Em muitos problemas de processamento de sinais multi-dimensionais¹ deseja-se encontrar uma transformação que, de algum modo, torne a estrutura essencial dos dados mais acessível [4]. Entre as técnicas lineares que buscam, através de pre-

¹Sinais multi-dimensionais são, em geral, produzidos por sistemas de medição com múltiplos sensores, mas também podem surgir a partir da aplicação de transformações (como a transformada de Fourier ou Wavelet) a sinais uni-dimensionais.

missas distintas, por uma nova representação para os sinais multi-dimensionais, pode-se mencionar a Análise de Componentes Principais (PCA - *Principal Component Analysis*) [3], a Análise de Fatores (*Factor Analysis*) [90] e a Análise de Componentes Independentes (ICA - *Independent Component Analysis*) [4].

Entre as técnicas listadas acima, a análise de componentes independentes busca por uma transformação onde os componentes na saída são estatisticamente independentes. A ICA vêm sendo aplicada na solução de diversos problemas na área de processamento de sinais como cancelamento de ruído [91], sonar passivo [92], telecomunicações [93], reconhecimento facial [94] e biomédica [95].

- **Independência Estatística:** Considerando duas variáveis aleatórias (VAs) y_1 e y_2 , se elas são independentes, então o conhecimento de uma não traz nenhuma informação a respeito da outra. Matematicamente, y_1 e y_2 são independentes estatisticamente se e somente se [96]:

$$p_{y_1,y_2}(y_1, y_2) = p_{y_1}(y_1)p_{y_2}(y_2), \quad (5.1)$$

onde $p_{y_1,y_2}(y_1, y_2)$, $p_{y_1}(y_1)$ e $p_{y_2}(y_2)$ são respectivamente as funções de densidade de probabilidade (pdf - *probability density function*) conjunta e marginais de y_1 e y_2 [96].

Pode-se obter uma expressão equivalente à equação (5.1) se, para todas as funções $g(y_1)$ e $h(y_2)$ absolutamente integráveis em y_1 e y_2 , vale a igualdade:

$$\mathcal{E}\{g(y_1)h(y_2)\} = \mathcal{E}\{g(y_1)\}\mathcal{E}\{h(y_2)\} \quad (5.2)$$

Considerando que a estimativa das funções de densidade de probabilidade (necessárias na equação (5.1)) é um problema de difícil solução, pois, em geral os componentes independentes são desconhecidos, uma vantagem da equação (5.2) é que as pdfs não são necessárias. A definição de independência pode ser facilmente estendida para mais de duas variáveis aleatórias. Percebe-se que a independência é um princípio mais restritivo que a descorrelação (quando $g(x) = x$ e $h(y) = y$). O conceito de independência envolve o conhecimento de toda a estatística dos dados, sendo assim muito mais abrangente

que a descorrelação (utilizada pela PCA), que somente utiliza estatística de segunda ordem (variância).

Na ICA, considera-se que um sinal multi-dimensional $\mathbf{x}(t) = [x_1(t), \dots, x_N(t)]^T$ observado (ou medido) é gerado a partir da combinação linear das fontes independentes $\mathbf{s}(t) = [s_1(t), \dots, s_N(t)]^T$. Na forma matricial e suprimindo-se o índice temporal t, pode-se escrever [97]:

$$\begin{bmatrix} x_1(t) \\ \vdots \\ x_N(t) \end{bmatrix} = \underbrace{\begin{bmatrix} a_{11} & \dots & a_{1N} \\ \vdots & \ddots & \vdots \\ a_{N1} & \dots & a_{NN} \end{bmatrix}}_{\mathbf{A}} \times \begin{bmatrix} s_1(t) \\ \vdots \\ s_N(t) \end{bmatrix}, \quad (5.3)$$

ou na forma matricial (omitindo o índice temporal):

$$\mathbf{x} = \mathbf{As}, \quad (5.4)$$

onde \mathbf{A} é a matriz de mistura.

O objetivo final da ICA é encontrar uma aproximação \mathbf{y} das fontes independentes, utilizando apenas os sinais observados \mathbf{x} . O vetor \mathbf{y} é definido por:

$$\mathbf{y} = \mathbf{Wx}, \quad (5.5)$$

sendo \mathbf{W} a matriz de separação. Se $\mathbf{W} = \mathbf{A}^{-1} \rightarrow \mathbf{y} = \mathbf{s}$, então o problema foi completamente solucionado.

Um problema clássico que pode ser solucionado usando-se ICA é conhecido como *cocktail-party problem* [4], e está formulado de forma simplificada, omitindo atrasos temporais e outros fenômenos físicos, como a existência de múltiplas reflexões, nas equações (5.6) e (5.7) (ver Figura 5.1). Considerando que numa sala existem duas pessoas falando simultaneamente e dois microfones em diferentes posições, os sinais gravados $x_1(t)$ e $x_2(t)$ são uma soma ponderada das fontes $s_1(t)$ e $s_2(t)$:

$$x_1(t) = a_{11}s_1(t) + a_{12}s_2(t) \quad (5.6)$$

$$x_2(t) = a_{21}s_1(t) + a_{22}s_2(t); \quad (5.7)$$

os coeficientes a_{ij} dependem das distâncias dos microfones às pessoas, e são os elementos da matriz de mistura \mathbf{A} , onde:

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}. \quad (5.8)$$

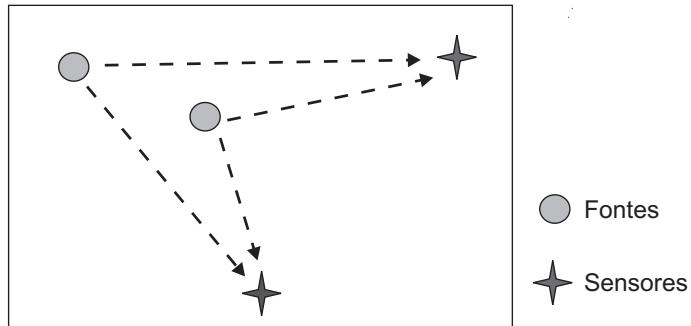


Figura 5.1: Diagrama do *cocktail party problem*.

Em um exemplo de aplicação da ICA, a Figura 5.2-a mostra as fontes $s_1(t)$ e $s_2(t)$, que foram misturadas linearmente, gerando os sinais $x_1(t)$ e $x_2(t)$ da Figura 5.2-b. Após a aplicação de um algoritmo para extração das componentes independentes (FastICA [98]), foram obtidas as curvas da Figura 5.2-c. Percebe-se que os sinais recuperados são cópias dos originais, a menos de fatores multiplicativos. Esta é uma das limitações inerentes ao modelo da ICA, não há como garantir o fator de escala (que pode ser positivo ou negativo) ou a ordem de extração dos componentes.

As técnicas de ICA foram desenvolvidas inicialmente para solucionar problemas de separação cega de sinais (BSS - *Blind Signal Separation*) semelhantes ao *cocktail-party problem*, porém, mais recentemente, surgiram outras aplicações interessantes, como extração de características, separação de fontes em telecomunicações e redução de ruído em imagens [4, 98]. Atualmente, ICA é aplicada com sucesso tanto para separação de sinais como para extração de características.

5.2 ICA não-linear

Em muitos problemas práticos, o modelo básico da ICA, onde os sinais observados são considerados combinações lineares e instantâneas das fontes, não representa

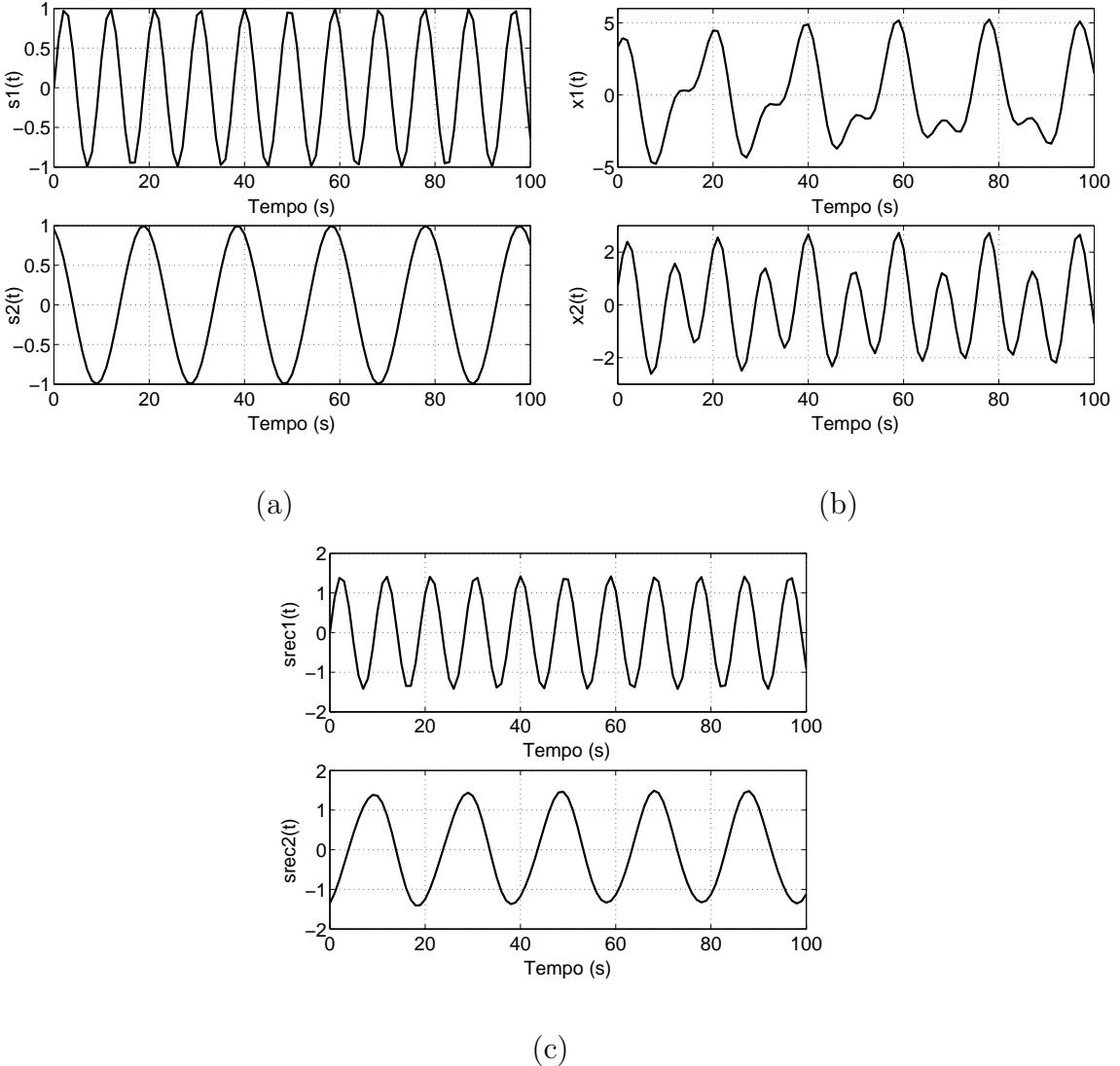


Figura 5.2: Sinais (a) fonte, (b) observados e (c) recuperados através da ICA.

corretamente o cenário real.

A equação (5.9) apresenta um modelo geral para as misturas não-lineares:

$$\mathbf{x} = \mathbf{F}(\mathbf{s}), \quad (5.9)$$

onde $\mathbf{F}(\cdot)$ é um mapeamento não-linear de $\mathbb{R}^N \rightarrow \mathbb{R}^N$, \mathbf{x} e \mathbf{s} são respectivamente os sinais observados e as fontes. A ICA não-linear consiste em encontrar o mapeamento $\mathbf{G}(\cdot): \mathbb{R}^N \rightarrow \mathbb{R}^N$ tal que os componentes de \mathbf{y} sejam estatisticamente independentes [89]:

$$\mathbf{y} = \mathbf{G}(\mathbf{x}). \quad (5.10)$$

Uma característica da NLICA é que o problema apresenta múltiplas soluções [99].

Se \mathbf{y}_1 e \mathbf{y}_2 são variáveis aleatórias independentes, é fácil provar que $f(\mathbf{y}_1)$ e $g(\mathbf{y}_2)$ (onde $f(\cdot)$ e $g(\cdot)$ são funções diferenciáveis) são também independentes [100]. Fica claro que, numa dada aplicação, sem o uso de restrições adequadas, infinitos mapeamentos inversos \mathbf{G} satisfazem a condição de independência entre os sinais estimados y_i , $i=1,\dots,N$. Se o objetivo do problema for realizar a separação das fontes de modo não-supervisionado ou cego (do inglês *Blind Signal Separation - BSS*), neste caso, deseja-se que os componentes y_i sejam aproximações das fontes independentes (\mathbf{s}) que produziram os sinais observados (\mathbf{x}). Então, informações a respeito do modelo de mistura ou das fontes devem ser conhecidas a priori. A NLICA vem sendo aplicada com sucesso em problemas como processamento de sinais de fala [101, 102], processamento de imagens [103, 104], predição de séries de ações em bolsas de valores [105] e processamento de sinais de um *array* de sensores [106, 107].

Em geral, o número de parâmetros a serem estimados num modelo de ICA não-linear é maior do que no caso linear. Os algoritmos de NLICA, se comparados aos de ICA, apresentam maior complexidade computacional e convergência mais lenta [99]. Considerando estas limitações, nas aplicações de NLICA deve-se verificar se existem restrições quanto ao aumento do custo computacional no processo de estimação dos componentes. Em problemas de separação cega de fontes, o algoritmo a ser utilizado deve ser escolhido utilizando informações a respeito do modelo de mistura.

Entre os algoritmos de NLICA propostos na literatura, uma classe de métodos impõe restrições estruturais ao modelo de mistura. Neste caso, pode-se garantir que os componentes estimados são aproximadamente iguais às fontes (a menos das indeterminações de fator multiplicativo e ordem de estimação dos sinais, assim como ocorre no modelo linear). Existem também alguns métodos (ou algoritmos) que não impõem restrições ao modelo de mistura, neste caso, o mapeamento não-linear dos dados observados em componentes independentes não garante a separação das fontes. Outro método diretamente relacionado com a NLICA, chamado de ICA Local, propõe uma etapa de agrupamento dos sinais em conjuntos de características semelhantes, que deve ser realizada antes da ICA. O agrupamento produz um mapeamento não-linear dos dados e a ICA (linear) estima os componentes independentes. Mais informações a respeito dos diversos algoritmos e modelos de NLICA serão fornecidas nas próximas seções.

5.2.1 Unicidade da Solução em NLICA

Conforme mencionado anteriormente, no caso não-linear, a independência estatística não é suficiente para garantir a separação das fontes. Se duas variáveis aleatórias y_1 e y_2 são independentes, então $p_{y_1,y_2}(y_1, y_2) = p_{y_1}(y_1)p_{y_2}(y_2)$. Para funções diferenciáveis f e g , pode-se provar que [108]:

$$p_{f(y_1),g(y_2)}(y_1, y_2) = p_{f(y_1)}(y_1)p_{g(y_2)}(y_2), \quad (5.11)$$

e então as variáveis $f(y_1)$ e $g(y_2)$ são também independentes. Esta indeterminação, diferente do fator de escala e da ordem de estimativação dos componentes (que são inerentes a ICA linear), não são aceitáveis em um problema de separação de fontes.

Estudos teóricos indicaram que a unicidade da solução da NLICA pode ser conseguida se o problema apresentar pelo menos uma das características a seguir [100]:

- O número de componentes é igual a dois. Deste modo, os sinais podem ser considerados como uma variável complexa.
- As pdf dos componentes independentes são limitadas a valores conhecidos.
- O mapeamento \mathbf{F} preserva o zero ($\mathbf{F}(0) = 0$) e é unívoco.
- O modelo de mistura é conhecido a priori e utilizado como informação para o algoritmo de estimativação das componentes independentes.

5.2.2 Modelos com restrições estruturais

Um caso especial da ICA não-linear são os métodos de estimativação dos componentes independentes que utilizam informações a respeito do modelo não-linear que gerou os dados observados. Estas informações se configuram em restrições estruturais ao mapeamento inverso (que é estimado pelo algoritmo). Entre os modelos com restrições estruturais pode-se destacar o de misturas **pós não-lineares** (PNL), que será descrito a seguir.

Misturas Pós Não-Lineares

O modelo de misturas pós não-lineares [108] é um dos mais utilizados na literatura, com aplicações em processamento de fala [109, 110], separação de sinais de áudio [111, 112] e processamento de imagens [113].

No modelo PNL, considera-se que inicialmente ocorre uma combinação linear das fontes (como no modelo básico de ICA), e as funções não-lineares f_i são aplicadas antes da observação dos sinais x_i :

$$x_i = f_i \left(\sum_{j=1}^n a_{ij} s_j \right). \quad (5.12)$$

É importante notar que as não-linearidades são aplicadas individualmente a cada componente da mistura linear (não são permitidas não-linearidades cruzadas). A Figura 5.3 ilustra o modelo de misturas PNL.

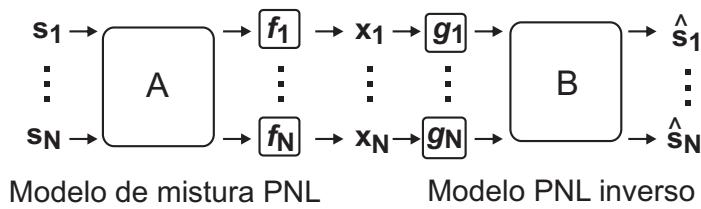


Figura 5.3: Diagrama do modelo de mistura PNL.

A consideração de que as misturas são pós não-lineares permite uma grande simplificação do problema, e as indeterminações existentes se tornam semelhantes às do caso linear. A modelagem através da equação (5.12) satisfaz grande parte dos fenômenos não-lineares, como, por exemplo, a modelagem da distorção de sensores num meio de propagação linear.

Diversos algoritmos foram propostos na literatura para estimação do modelo PNL. Um dos primeiros trabalhos [108], utiliza redes neurais para estimar cada função não-linear g_i e o ajuste dos pesos é feito a partir da minimização da informação mútua usando o método do gradiente (mais detalhes a respeito deste algoritmo podem ser encontrados no Apêndice A). Na estimação do modelo PNL usando um algoritmo do tipo gradiente descendente (que realiza uma busca local), o grande número de parâmetros e as características não-lineares do problema contribuem para a convergência em mínimos locais. Visando minimizar esta limitação, foram propostos algoritmos de estimação do modelo PNL usando métodos de busca global como algoritmos genéticos [114, 115], reconhecimento simulado (*Simulated Annealing*) e aprendizado competitivo (*Competitive Learning*) [109].

Outros modelos de misturas com restrições estruturais

Alguns modelos com restrições estruturais diferentes do PNL foram propostos na literatura. No trabalho [112], o modelo de mistura é definido por:

$$\mathbf{x} = \mathbf{A}_2 f(\mathbf{A}_1 \mathbf{s}), \quad (5.13)$$

sendo \mathbf{A}_1 e \mathbf{A}_2 matrizes quadradas e $f = [f_1, f_2, \dots, f_N]^T$ um mapeamento com funções não-lineares aplicadas a cada componente (assim como o modelo PNL, este também não permite não-linearidades aplicadas a mais de um componente). O modelo definido na Equação 5.13 e ilustrado também na Figura 5.4 é chamado Pós Não-linear Linear (PNL-L). O bloco linear \mathbf{A}_2 é executado após a aplicação das funções não-lineares, produzindo um modelo mais geral que o PNL. Nos trabalhos [116, 117] são propostos algoritmos baseados em redes neurais para a estimativa do modelo PNL-L.

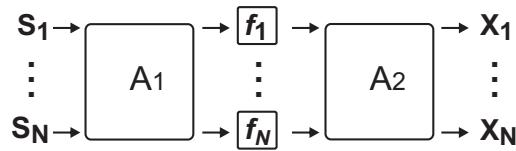


Figura 5.4: Diagrama do modelo PNL-L.

Em [117], um modelo estrutural chamado mono não-linearidade (ver Figura 5.5) foi proposto para o problema da NLICA. Neste modelo os sinais observados são gerados a partir de:

$$\mathbf{x} = f^{-1}(\mathbf{A}f(\mathbf{s})). \quad (5.14)$$

Este modelo (chamado de mistura de mono não-linearidade e ilustrado na Figura 5.5) é dito mais geral que o PNL, pois as funções não-lineares podem ser aplicadas a mais de um componente (as funções f_i podem ser funções não-lineares de mais de uma variável). A análise deste modelo a partir da teoria da análise funcional (*functional analysis*) [118], mostra que esta arquitetura pode representar qualquer mistura com duas camadas de não-linearidades [117].

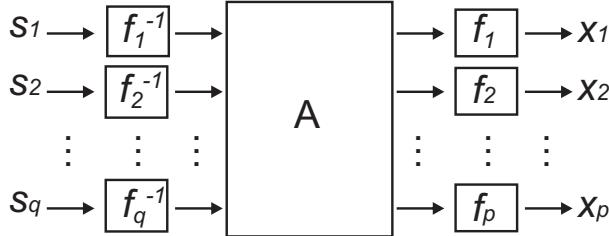


Figura 5.5: Diagrama do modelo da Mono não-linearidade.

5.2.3 Algoritmos sem restrições estruturais

Se nenhuma restrição ao modelo de mistura é imposta, não há garantia que os componentes independentes estimados estejam relacionados com as fontes (ver Seção 5.2.1). Entre os métodos de NLICA sem restrições estruturais, pode-se destacar o uso de mapas auto-organizáveis e os métodos que utilizam inferência Bayesiana.

NLICA a partir de Mapas Auto-Organizáveis

Uma das primeiras tentativas bem sucedidas de realizar NLICA utilizou mapas auto-organizáveis [119]. Pode-se provar que as coordenadas y_1 e y_2 do neurônio vencedor no mapa (ver Figura 5.6) são independentes e aproximadamente uniformemente distribuídas [119]. Para estimar a NLICA, o SOM é treinado usando como entradas os sinais observados, e as coordenadas do neurônio vencedor correspondem a uma aproximação dos componentes independentes.

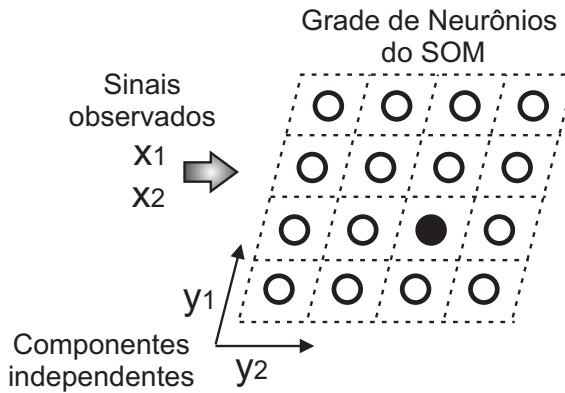


Figura 5.6: NLICA a partir de SOM.

Entre as desvantagens do método pode-se destacar [99]:

- O mapeamento é discreto (existe um número limitado de neurônios no mapa),

então, algum tipo de regularização é necessária para produzir componentes contínuos. Esse problema pode ser minimizado aumentando-se o número de neurônios do mapa.

- Os componentes a serem estimados devem ter pdf sub-gaussiana (quanto mais próxima da distribuição uniforme, melhor).
- O custo computacional para treinamento dos mapas aumenta rapidamente com o número de componentes independentes a serem estimados.

Para avaliar o custo computacional, o número de parâmetros N_p do SOM pode ser estimado pela expressão:

$$N_p = N \times (Q_L)^N \quad (5.15)$$

onde N é o número de componentes a serem estimados (que é considerado igual ao número de sinais observados) e Q_L é o número de níveis de quantização desejado [89].

Diversas aplicações do SOM para estimação da NLICA podem ser encontradas na literatura, entre elas pode-se citar os trabalhos [113] e [103], onde os objetivos eram, respectivamente, separação e remoção de ruído em imagens.

Com uma formulação alternativa aos SOM, o Mapeamento Topográfico Gênerativo (GTM - *Generative Topographic Mapping*) foi introduzido em [120], e apresenta princípios estatísticos mais fundamentados que o mapa SOM. O método GTM básico tem poucas vantagens práticas em relação aos Mapas Auto-Organizáveis, pois aqui as componentes independentes também são assumidas como processos uniformemente distribuídos e o espaço de características é formado a partir de uma grade retangular discreta m-dimensional. Porém, devido à sua formulação matemática mais fundamentada, o GTM pode ser estendido para variáveis não uniformes. O trabalho [121] propõe uma modificação à formulação básica onde são introduzidos coeficientes de ponderação que permitem a estimação de componentes independentes com qualquer tipo de distribuição. Os componentes são modelados como misturas de sinais Gaussianos e os parâmetros são estimados usando o algoritmo *Expectation Maximization* [98]. O treinamento do GMT envolve dois passos, a avaliação da probabilidade a *posteriori* e a adaptação dos parâmetros do modelo.

NLICA a partir de Inferência Bayesiana

Nos métodos baseados em inferência Bayesiana, considera-se que os sinais observados são gerados a partir de [122]:

$$\mathbf{x} = f(\mathbf{s}) + \mathbf{n} \quad (5.16)$$

onde \mathbf{n} é definido como ruído Gaussiano independente dos componentes a serem estimados.

Os componentes independentes são modelados como misturas de sinais de distribuição Gaussiana. Pode-se provar que, dado um número suficiente de Gaussianas, virtualmente qualquer distribuição de probabilidade pode ser modelada com uma certa precisão [122]. Uma variação deste método foi aplicada em [123] para o modelo linear da ICA. Em grande parte dos algoritmos Bayesianos para NLICA, redes neurais tipo MLP de duas camadas são treinadas para aproximar o mapeamento não linear [99]:

$$f(\mathbf{s}) = \mathbf{B}\Phi(\mathbf{As} + \mathbf{a}) + \mathbf{b} \quad (5.17)$$

Em um método de estimação Bayesiano, probabilidades *a posteriori* são associadas a cada modelo não-linear que, possivelmente, teria gerado os dados observados. Verificar uma quantidade tão grande de modelos não é possível na prática; então, os métodos Bayesianos para NLICA utilizam uma técnica chamada de “aprendizagem amostral” (EL - *ensemble learning*) [124]. Na EL, somente o conjunto mais provável de modelos é testado utilizando uma aproximação paramétrica que é ajustada à probabilidade *a posteriori* [125].

Métodos Bayesianos de NLICA foram propostos em [126] e [127]. No trabalho [128] foram realizados testes experimentais para comparar o desempenho dos modelos Bayesiano e PNL na estimação dos componentes independentes, as principais conclusões foram:

- os algoritmos PNL apresentam desempenho superior quando as misturas seguem o modelo PNL clássico (não-linearidades inversíveis e mesmo número de componentes independentes e sinais observados);
- o desempenho de ambos os métodos pode ser melhorada a partir da exploração da informação de mais misturas que componentes independentes;

- a principal vantagem do método Bayesiano é que mapeamentos mais genéricos podem ser produzidos (uma vez que não há restrições estruturais). Estes métodos geralmente apresentam maior custo computacional e necessitam de várias inicializações para obter uma solução ótima (podem apresentar problemas com mínimos locais da função custo).

No trabalho [107] um algoritmo Bayesiano de NLICA foi utilizado com sucesso para a separação de sinais medidos em um conjunto de sensores químicos.

O algoritmo MISEP

O algoritmo MISEP [129] utiliza a minimização da Informação Mútua (ver Apêndice A) como estratégia para busca pelos componentes independentes, e é considerado como uma extensão do método INFOMAX [4], podendo ser utilizado para estimar tanto o modelo linear quanto o não-linear da ICA. Na Figura 5.7 pode-se observar um diagrama do MISEP (para duas entradas e duas saídas), onde x_i e y_i são respetivamente os sinais observados e os componentes independentes estimados, o bloco $\mathbf{G}(\cdot)$, no caso linear, aproxima a matriz de separação \mathbf{W} , e para a NLICA, deve fornecer uma aproximação do mapeamento não-linear inverso. As funções não-lineares ψ_i e as variáveis de saída z_i são utilizadas apenas no processo de treinamento. Após a convergência do algoritmo, as não-linearidades devem ser aproximações da função de probabilidade cumulativa (cdf - *cumulative probability function*) dos componentes independentes.

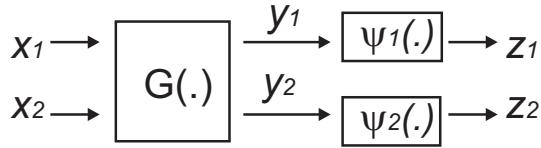


Figura 5.7: Diagrama do algoritmo MISEP.

Para a aplicação em NLICA, o bloco \mathbf{G} é estimado por uma rede neural (que pode utilizar tanto a arquitetura *perceptron* de múltiplas camadas - MLP como rede de funções de base radial - RBF). Como o objetivo é estimar a cdf, as saídas z_i são restritas ao intervalo $[0,1]$ e as ψ_i são limitadas a funções estritamente crescentes. Para estimação iterativa de cada função ψ_i é utilizada uma redes neurais MLP com uma camada oculta (de neurônios sigmoidais) e uma camada de saída (linear). Estas

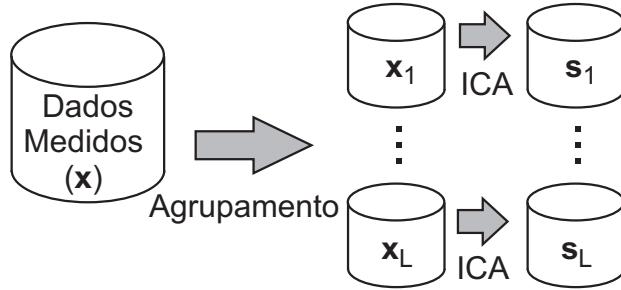


Figura 5.8: Diagrama do modelo da ICA local.

redes tem uma entrada (y_i) e uma saída (z_i). O treinamento do modelo MISEP é feito a partir da maximização da entropia das saídas z_i , o que acaba produzindo a minimização da informação mútua dos componentes y_i , mais detalhes podem ser encontrados em [129]

O MISEP foi aplicado em processamento de sinais de áudio [129] e separação de imagens [130]. Foram propostas também, modificações ao algoritmo MISEP visando otimizar a estimação dos componentes independentes quando as misturas seguem os modelos pós não-linear (PNL) [131] e pós não-linear-linear (PNL-L) [132].

5.2.4 ICA Local

Se o modelo da ICA for utilizado para extração de características (ao invés de separação de fontes), uma melhor descrição dos dados pode ser obtida se forem exploradas características locais. Considerando um conjunto de sinais multi-dimensionais com estatística variável, o modelo da ICA linear pode não ser capaz de revelar a estrutura fundamental dos dados. Neste caso, é mais razoável realizar a extração de características (estimação das componentes independentes) a partir de k subconjuntos dos dados (ver Figura 5.8). Os sinais pertencentes ao k -ésimo subconjunto apresentam características semelhantes. Este procedimento leva ao modelo da ICA local.

Conforme proposto em [133], um conjunto de dados de alta dimensão pode ser separado em sub-conjuntos, através de algum algoritmo de agrupamento como o *k-means* [134] ou SOM [135], e componentes independentes lineares são então estimadas de cada subconjunto. Caso não exista informação a priori a respeito do número de agrupamentos, metodologias foram propostas nos trabalhos [136, 137]

para sua estimação.

Na ICA Local, o agrupamento é responsável por uma representação não-linear dos dados, enquanto modelos de ICA linear aplicados a cada sub-conjunto (*cluster*) descrevem as características locais dos dados. A ICA local pode ser considerada como um compromisso entre os modelos linear e não-linear da ICA [99]. O objetivo é obter uma melhor representação dos dados (se comparado com o modelo linear da ICA), evitando os problemas computacionais do modelo não-linear [138]. Em diferentes abordagens, os agrupamentos podem ser montados com superposição, usando por exemplo fronteiras *fuzzy* [139, 140], ou sem superposição [138, 141].

Nos trabalhos [142, 143] a ICA Local foi aplicada para a estimação da informação mútua. A informação mútua [87] é uma importante ferramenta em diversas aplicações de processamento de sinais, especialmente na seleção de características. O modelo da ICA Local foi utilizado também em [?] para a remoção de ruído. O modelo proposto consiste em realizar k agrupamentos de versões atrasadas no tempo dos sinais medidos e estimar a ICA linear para cada um destes conjuntos. Através da abordagem proposta, foi alcançado um aumento da relação sinal/ruído da ordem de 10 dB.

5.3 Aplicações de ICA e NLICA para extração de características

Nessa Seção serão descritas, de modo resumido, algumas aplicações da análise de componentes independentes, nos seus modelos linear e não-linear (respectivamente ICA e NLICA), para a extração de características.

No trabalho [144], ICA foi utilizada como pré-processamento para problemas de classificação em nove bases de dados diferentes (obtidas no repositório de bases de dados para aprendizado de máquina da Universidade da Califórnia - Irvine, CA, Estados Unidos [145]). Entre os problemas testados, estão a classificação de vinhos a partir de características físicas e químicas, a identificação da existência de câncer em amostras de tecido da mama, a identificação isolada de vogais independente do locutor e previsão de sobrevida de pacientes que sofreram ataque cardíaco a partir do resultado do eletrocardiograma. A transformação da ICA foi estimada através

do algoritmo JADE [146]. Utilizando-se classificadores neurais (MLP), a eficiência foi comparada para sinais sem pré-processamento, sinais branqueados e sinais após ICA. Em alguns casos (como na identificação de vogais) o uso da ICA produziu uma redução do erro de identificação (24,13% sem pré-processamento, 21,05% após o branqueamento e 20,77% após a ICA). Em outros casos, porém, a aplicação da ICA tornou mais difícil o problema de classificação (como no caso da identificação do cancer de mama, onde, sem pré-processamento, o erro foi de 2,55% e após a ICA aumentou para 2,63%). Analisando-se todos os resultados conclui-se que, nem sempre a aplicação da ICA contribui para um aumento na eficiência, esse fato é intensificado em problemas onde o modelo da mistura linear não se aplica (pois possivelmente existem não-linearidades envolvidas). O uso da ICA parece tornar mais suave a curva de erro de treinamento das redes neurais, contribuindo para a diminuição da quantidade de mínimos locais e, consequentemente, da probabilidade do treinamento ficar estacionado num desses mínimos.

Em [147], ICA foi novamente utilizada para detecção do cancer de mama a partir de imagens digitalizadas de mamografias. Nesse trabalho, os componentes independentes foram estimados através do algoritmo FastICA [148] e classificadores neurais (MLP) foram utilizados para produzir a decisão. As amostras disponíveis pertenciam a três classes distintas (normais, com alterações benignas e com alterações malignas). A ICA foi estimada a partir de pequenas janelas nas mamografias onde as classes de interesse eram mais facilmente identificadas. Foram obtidas eficiências de identificação da ordem de 99,9% para as amostras normais, 86,8% e 91,1% respectivamente para amostras com alterações benignas e malignas.

Microarranjos de DNA foram pré-processados por ICA em [149] para a classificação através de máquinas de vetor de suporte (SVM - *Support Vector Machines*) [5]. Os microarranjos de DNA são fragmentos genômicos que representam segmentos gênicos em particular. Nesse trabalho, o algoritmo FastICA foi utilizado para extrair características dos microarranjos (de quatro bases de dados distintas) com o objetivo de identificar a presença de diferentes tipos de tumores (de colo de útero, leucemia, de fígado e do sistema nervoso). As eficiências de identificação obtidas para os quatro tipos foram, respectivamente, 90%, 100%, 74% e 76%. Uma outra aplicação de ICA no mesmo problema pode ser encontrada em [150].

Ainda na área biomédica, no trabalho [151], a ICA foi utilizada como pré-processamento para um mapeamento não-supervisionado de características oculares, com o objetivo de identificar a presença de glaucoma. A partir de padrões de um exame conhecido como *standard automated perimetry* (SAP), aplicou-se a ICA e o agrupamento (não-supervisionado) foi realizado sobre os componentes independentes estimados. Através dessa abordagem, 98,4% das assinaturas de olhos com padrão óptico normal foram corretamente classificadas e, considerando-se os olhos com glaucoma, o acerto foi de 68,6%.

A análise de componentes independentes foi utilizada em [152] para a análise de sinais de espectrometria eletrônica de perda de energia (EELS - *Electron Energy Loss Spectroscopy*). A EELS [153] pode ser empregada para medições precisas de espessura (com resolução da ordem de 0.1 nm), pressão e análise de composição química. A ICA, através do algoritmo SOBI [154], foi utilizada como ferramenta complementar de análise dos espectros eletrônicos produzidos. O uso da ICA possibilitou a análise simultânea de dois espectros misturados e eliminou escolhas subjetivas durante a análise (que sem o uso da ICA precisam ser feitas pelo usuário).

O modelo não-linear da ICA também é utilizado em problemas de extração de características, por exemplo, o trabalho [155] ilustra a aplicação da NLICA num problema de classificação de sinais de eletroencefalograma (EEG). O objetivo é a separação das diferentes atividades cerebrais independentes, porém como não há garantia que o processo de combinação é linear, utilizou-se o modelo da NLICA numa tentativa de modelar dinâmicas cerebrais não-lineares. O modelo pós não-linear (PNL) foi empregado para estimar os componentes independentes. A informação mútua foi utilizada como medida da independência e um algoritmo genético [156] buscou sua minimização. Múltiplos classificadores lineares (cada um treinado com um dos componentes estimados) foram utilizados para identificar os sinais provenientes do movimento da mão. Uma combinação das saídas dos múltiplos classificadores foi utilizada para produzir a decisão final. A eficiência de identificação a partir dos sinais medidos (sem pré-processamento) foi de 73,84%, aumentando para 74,61% e 77,95% quando utilizados, respectivamente, pré-processamento por ICA e NLICA.

A NLICA foi utilizada no trabalho [157] visando à extração de características de séries temporais de ações para a previsão do índice diário de uma bolsa de valores.

Para formar o vetor N-dimensional de entrada para a NLICA, foram utilizados a série com os valores de fechamento diário da bolsa e N-1 versões atrasadas desta série. O algoritmo MISEP [129] foi utilizado para estimar os componentes independentes. Um modelo de regressão por vetor de suporte (*Support Vector Regression*) foi utilizado para prever o comportamento da bolsa. Os resultados obtidos com a NLICA foram comparados com pré-processamento por ICA e PCA e as eficiências obtidas foram, respectivamente, 80%, 75% e 79%. Também neste exemplo, o modelo da NLICA mostrou-se mais eficiente para evidenciar as características discriminantes do problema em questão.

Em Física de Altas Energias (HEP - *High-Energy Physics*) também são encontradas algumas aplicações de PCA e ICA para extração de características. Alguns desses trabalhos serão descritos brevemente na próxima seção. Para a NLICA, talvez por ser uma técnica ainda menos difundida (em comparação com PCA e ICA), não foram encontradas aplicações na área de HEP.

5.4 Aplicações em Física de Altas Energias e Áreas Correlatas

A partir do final da década de 1990, os métodos de aprendizado estatístico multi-variável como PCA, ICA e SOM vêm sendo aplicados com sucesso em problemas na área de física de alta energia.

Um dos primeiros trabalhos neste tópico [158], foi publicado em 1998 e utiliza mapas auto-organizáveis (SOM) para a classificação de eventos de raios gama em astronomia de alta energia. Em [159], mapas SOM foram aplicados para a separação de bósons W do ruído de fundo composto por jatos hadrônicos. No trabalho [160], o ruído de fundo gerado na aceleração do feixe de partículas foi rejeitado a partir de mapas auto-organizáveis modificados. Redes SOM também foram utilizadas com sucesso para análise, classificação e monitoramento de sinais do telescópio OGLE (no Chile) [161] e para a identificação de prováveis assinaturas de bósons de Higgs [162].

A análise de componentes principais (PCA) é um técnica de descorrelação e compactação bastante utilizada em diversas áreas do conhecimento. Em física de alta energia, PCA foi aplicada para a seleção de variáveis de entrada de um discri-

minador neural no trabalho [163]. Em [164], são apresentadas diversas aplicações em HEP onde é utilizada a PCA para extração de características e compactação. No trabalho [165], sinais ópticos de nebulosas planetárias são processados por PCA com o objetivo de extrair informações a respeito de suas características morfológicas. Numa outra aplicação em astrofísica, PCA é utilizada, em conjunto com ICA, para a remoção do ruído de fundo e de outras fontes de interferência, permitindo melhor visualização de dados de ventos e tempestades solares [166]. O trabalho [88] utiliza a PCA, de forma segmentada, para compactação de sinais de calorimetria do ATLAS, em seguida classificadores neurais realizam a decisão elétron/jato, conseguindo boa eficiência de classificação.

A análise de componentes independentes (ICA) tem aplicação mais recente em HEP, sendo que um dos primeiros trabalhos foi publicado em 2005 e descreve a eliminação de ruído na análise de sinais do feixe de partículas do experimento BOOSTER do Fermilab [167]. Neste trabalho também é realizada uma comparação com um sistema semelhante baseado em PCA e a ICA apresenta resultados melhores. No trabalho [168], ICA é utilizada para análise de dados multi-variados em experimentos de física atômica e nuclear. A aplicação de ICA proporcionou redução do ruído de fundo, permitindo melhor visualização do sinal de interesse. ICA também foi aplicado com sucesso para separação de sinais em astrofísica de alta energia conforme detalhado a seguir. Em [169], ICA foi aplicada para a separação de imagens de fontes sobrepostas adquiridas pelo satélite Planck da Agência Espacial Européia; no trabalho [170] utiliza-se a análise de componentes independentes, em substituição aos filtros casados, para a decomposição de sinais astrofísicos simulados compostos pela combinação de moléculas elementares em estado congelado. Ainda na área de astrofísica, nos trabalhos [171, 172] ICA foi aplicado para a caracterização da radiação cósmica de fundo em microondas (CMB - *Cosmic Microwave Background*). A CMB é uma forma de energia eletromagnética que preenche todo o universo e foi inicialmente observada em 1965. A CMB é visualizada apenas por rádio-telescópios.

A partir destes exemplos, percebe-se que, apesar da aplicação mais recente em física de alta energia e áreas correlatas, diversos problemas de extração de características, remoção de ruído, agrupamento não-supervisionado (*clustering*) e visualização vêm sendo resolvidos com a aplicação das técnicas estatísticas de processa-

mento não-supervisionado de sinais.

5.5 Utilizando Informação das Classes na Estimação dos Componentes Independentes

Conforme visto anteriormente, a ICA surgiu como uma alternativa para a extração de características, pois é capaz de transformar os atributos em um conjunto de componentes estatisticamente independentes, eliminando a redundância entre eles. Porém, quando há a necessidade de redução de dimensão (compactação), no contexto da ICA, o método mais utilizado é a PCA (Análise de Componentes Principais, ver mais detalhes no Apêndice A). A PCA não é adequada a um problema de extração de características, pois um componente de baixa energia (que seria eliminado após a compactação) pode apresentar grande importância para a discriminação das classes.

Um outro problema é que o modelo da ICA/NLICA não foi originalmente desenvolvido para extração de características discriminantes e, conforme mostrado em [144], não há como garantir que a estimação dos componentes independentes seja útil neste contexto. Os benefícios alcançados pela aplicação da ICA como pré-processamento dependem de características da base de dados utilizada.

Visando minimizar as limitações da ICA indicadas acima, foram utilizadas neste trabalho algumas modificações para incluir informação das classes no procedimento de estimação dos componentes independentes.

Um modo alternativo à PCA para realizar a compactação, incluindo a informação das classes, é através do método conhecido como Componentes Principais de Discriminação (PCD - *Principal Components for Discrimination*) [11]. Essa abordagem será descrita mais detalhadamente na Seção 5.5.1.

A informação das classes pode ainda ser utilizada no contexto da ICA de diferentes modos. Um procedimento simples, proposto em [173], é a inclusão dos rótulos de classe como atributos de entrada para os algoritmos de estimação dos componentes independentes (ver Seção 5.5.2). Uma outra abordagem possível é realizar a estimação dos modelos da ICA separadamente para amostras de cada classe . Alternativamente, um método para estimação de componentes independentes e discriminantes foi desenvolvido no contexto desta tese e será mostrado na Seção 5.5.4.

5.5.1 Componentes Principais de Discriminação

Considerando um problema de classificação de padrões, o uso da PCA para compactação pode ser prejudicial, pois, os componentes menos energéticos (que são eliminados após a PCA) podem carregar informações discriminantes. Neste caso, pode-se utilizar técnicas de compactação mais adequadas. As componentes principais de discriminação (PCD - *Principal Discriminating Components*) [11, 174] são obtidas a partir da projeção dos sinais de entrada em um conjunto compacto que carrega toda a informação importante para discriminação entre as classes.

Conforme proposto em [11], para um problema de classificação, o objetivo da PCD é obter uma projeção linear dos sinais de entrada $\mathbf{x} = [x_1, \dots, x_N]^T$ nos componentes $\mathbf{z} = [z_1, \dots, z_K]^T$ (com $K < N$) que maximizam a discriminação entre as classes (ou seja, z_i são os componentes principais de discriminação).

Considerando um problema de discriminação onde existem apenas duas classes possíveis, os PCD podem ser estimados a partir de uma rede neural (de arquitetura MLP - *Multi-Layer Perceptron*) [5] com uma camada oculta e um neurônio de saída, treinada para obter máxima discriminação entre as classes. Conforme indicado na Figura 5.9-a, uma rede neural com um neurônio na camada oculta é capaz de estimar o primeiro PCD, que é obtido a partir da projeção das entradas na direção dos pesos sinápticos do neurônio oculto:

$$z_1 = [b_{1,1}, b_{1,2}, \dots, b_{1,N}] \times [x_1, \dots, x_N]^T + b_{0,1}, \quad (5.18)$$

onde $b_{0,1}$ é o *bias* do neurônio. Adicionando-se mais neurônios ocultos consegue-se estimar os demais PCD (conforme ilustrado na Figura 5.9-b). No processo de estimação, estatística de ordem elevada é acessada a partir da utilização de funções de ativação não-lineares. O treinamento da rede neural pode ser feito com o congelamento dos pesos da camada de entrada correspondentes aos componentes já estimados, ou seja, na estimativa do l -ésimo componente, os pesos $b_{i,j}$, com $i < l$ e $j = 1, \dots, N$ não são ajustados. Os demais pesos da rede são ajustados a cada novo componente estimado.

Outros modelos que, de modo semelhante aos PCD, utilizam redes neurais para extrair características discriminantes de um conjunto de sinais foram propostos na literatura em [175, 176, 177].

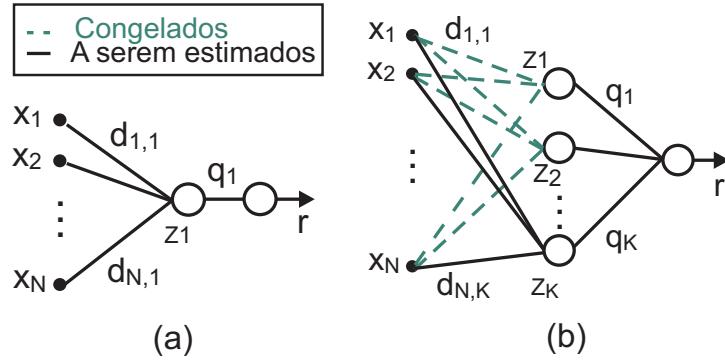


Figura 5.9: Modelos neurais para estimar (a) a primeira e (b) a k -ésima PCD.

5.5.2 Utilizando os rótulos de classe como sinais de entrada para os algoritmos de ICA

No trabalho [173] foi proposta a utilização dos rótulos de classes como entrada para os algoritmos de estimação dos componentes independentes. Conforme mostrado na Figura 5.10, num problema de classificação binária (com apenas duas classes), para cada exemplo de entrada é associado um novo atributo c com valor igual a 1 (para a classe 1) e -1 (para a classe 2). O bloco G pode ser utilizado para estimar os modelos linear ou não-linear da ICA, a depender do algoritmo de treinamento utilizado.

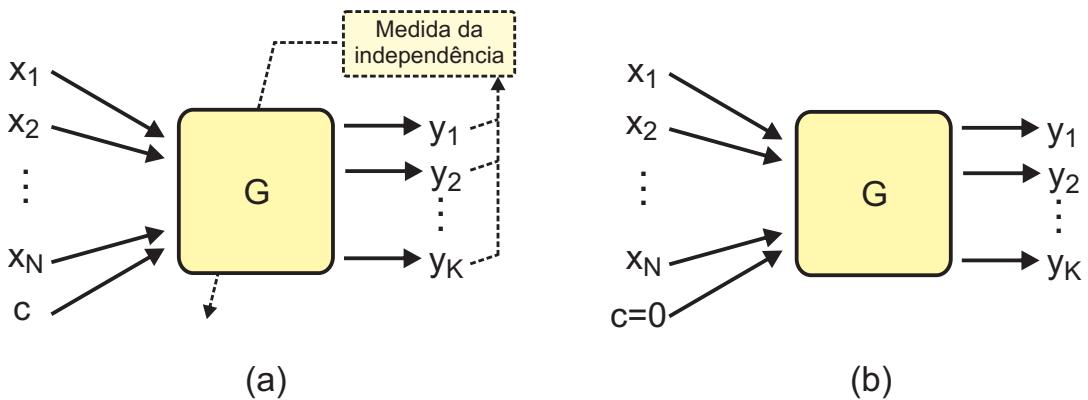


Figura 5.10: Diagramas de (a) treinamento e (b) operação dos algoritmos de ICA/NLICA utilizando informação das classes na entrada.

O parâmetro c é adicionado ao vetor de atributos original $\mathbf{x} = [x_1, x_2, \dots, x_N]$, gerando $\mathbf{x}_C = [x_1, x_2, \dots, x_N, c]$. Para treinamento dos algoritmos de ICA utiliza-se como entrada o vetor \mathbf{x}_C .

Como num cenário prático de operação do sistema classificador os rótulos de

classe não estarão disponíveis, a informação das classes deve ser removida dos componentes estimados. Isso pode ser feito removendo-se as conexões da entrada c ao modelo, ou substituindo c por um vetor de zeros.

Em um trabalho semelhante [178], o mesmo procedimento de treinamento foi adotado, porém o modelo da ICA utilizado é linear e quadrado (mesmo número de entradas e saídas, não havendo portanto redução de dimensão). Após a estimação dos componentes, aqueles que tem pesos de baixa amplitude são eliminados, produzindo um conjunto mais compacto de características discriminantes.

Um outro modo de utilizar a informação das classes no processo de estimação dos componentes independentes foi proposto em [179]. Neste caso, conforme ilustrado na Figura 5.11, os componentes estimados na saída devem ser independentes e apresentar máxima informação mútua com os rótulos de classe c . A função custo a ser maximizada (este trabalho restringe-se ao modelo linear da ICA) é definida como:

$$L(\mathbf{W}) = -\log |\det \mathbf{W}| + \sum_{i=1}^d H(z_i) - \sum_{i=1}^d I(z_i, c) \quad (5.19)$$

onde \mathbf{W} é a matriz da ICA, z_i são os componentes independentes estimados e c o vetor de rótulos de classe.

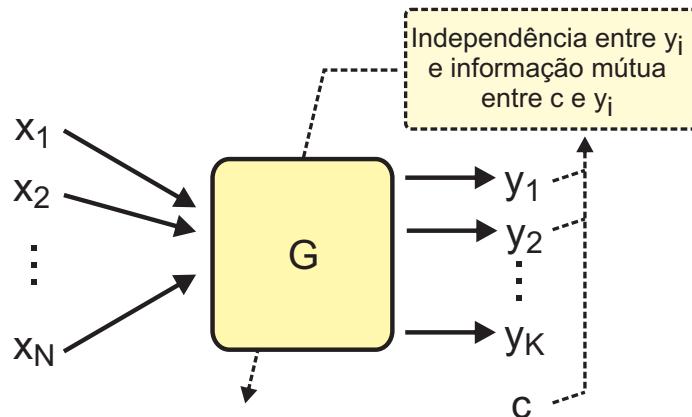


Figura 5.11: Modelo da ICA que utiliza a informação das classes em conjunto com os componentes estimados na função custo.

5.5.3 Componentes Independentes para Cada Classe

Um outro modo de utilizar informação supervisionada é realizar a estimação de modelos G_m da ICA utilizando apenas amostras da classe m . Este procedimento foi proposto em [180], e está ilustrado na Figura 5.12. Pode-se observar que as saídas de cada bloco G_m são utilizadas como entradas para um classificador especialista na identificação da classe m . Ao final do processamento, um combinador utiliza as informações dos m classificadores para atribuir o rótulo de classe (produzindo a decisão final).

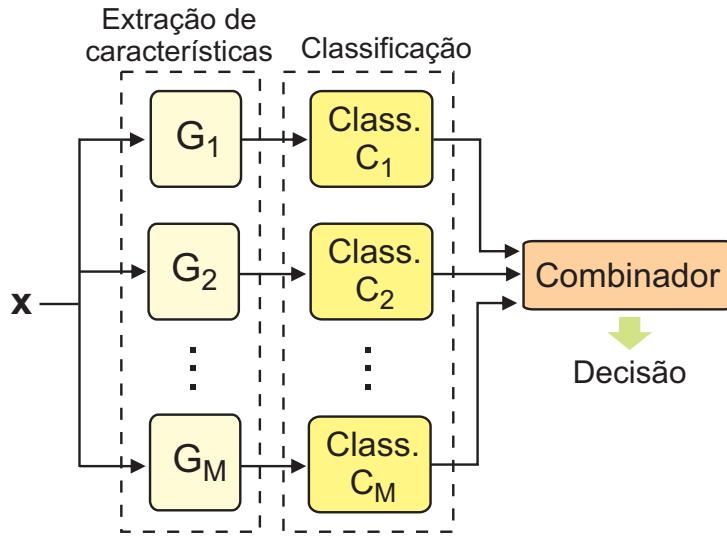


Figura 5.12: Sistema de classificação baseado em modelos da ICA estimados para cada classe.

No trabalho [180], este modelo foi capaz de produzir alta eficiência no diagnóstico acústico de equipamentos industriais. A grande vantagem desta abordagem é que pode-se utilizar algoritmos tradicionais de ICA/NLICA, sem a necessidade de modificações (o que não ocorre para os métodos descritos na seção 5.5.2). No trabalho descrito acima foi utilizado o algoritmo FastICA para estimação dos componentes independentes.

5.5.4 Proposta de algoritmo para estimação de componentes independentes e discriminantes

No desenvolvimento desta tese, foi proposto um método alternativo para estimação de componentes independentes e discriminantes. No trabalho [181], foi derivado um algoritmo de treinamento para um modelo pós não-linear (PNL) modificado. Este método será descrito a seguir e, em sequencia, será mostrado como ele pode ser estendido para um modelo geral (sem restrições estruturais) da NLICA.

Conforme ilustrado na Figura 5.13, um bloco de compactação foi adicionado ao modelo PNL, com o objetivo de transformar o conjunto de N atributos em K componentes (com $N < K$). Assim, a arquitetura proposta é adequada para o caso sobre-determinado (quando existem mais sinais observados do que fontes).

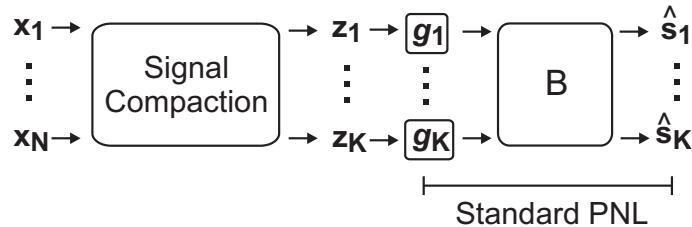


Figura 5.13: Modelo Pós Não-linear modificado.

O modelo PNL modificado pode ser estimado a partir de dois procedimentos distintos. Uma abordagem possível é executar a estimação do bloco de compactação de modo independente do modelo PNL. Neste caso, a compactação se configura numa etapa de pré-processamento para a NLICA e pode ser executada, por exemplo através da transformação PCD.

De modo alternativo, o modelo PNL modificado pode ser estimado através do procedimento mostrado na Figura 5.14, que combina informação de duas funções custo diferentes $c_1(\hat{s})$, que mede a independência estatística entre os componentes estimados \hat{s} , e $c_2(y)$, que avalia a eficiência de discriminação produzida a partir de um discriminante linear (DL) [134], onde y é a saída do classificador.

Considerando um bloco de compactação linear, os componentes independentes estimados são descritos por:

$$\hat{s}_i = \sum_{j=1}^K b_{ij} g_j(z_j) \quad i = 1, \dots, K \quad (5.20)$$

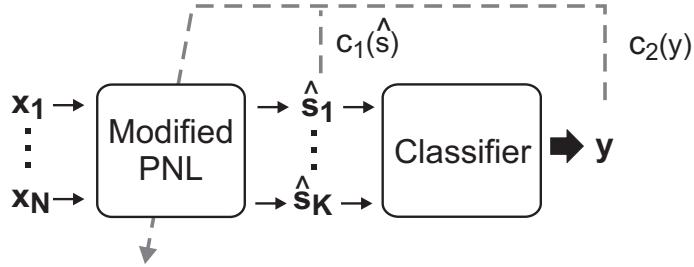


Figura 5.14: Procedimento de treinamento para o modelo pós não-linear modificado.

onde $\mathbf{z} = \mathbf{Dx}$, \mathbf{D} é uma matriz retangular ($K \times N$) de compactação e os b_{ij} são elementos da matriz quadrada \mathbf{B} ($K \times K$).

A estimativa das não-linearidades $g_i(\cdot)$ é feita de modo semelhante ao proposto no trabalho [108]. Cada função é aproximada por:

$$g_i(z_i) = \sum_{h=1}^{N_H} \beta_{hi} \tanh(\omega_{hi} z_i - \eta_{hi}) \quad i = 1, \dots, K \quad (5.21)$$

onde β_{hi} , ω_{hi} e η_{hi} são parâmetros a serem determinados. Um algoritmo genético [182] foi utilizado para estimar o conjunto de parâmetros \mathbf{D} , \mathbf{B} , β_{hi} , ω_{hi} e η_{hi} que maximiza a função custo definida por:

$$c(\hat{\mathbf{s}}, \mathbf{y}) = \frac{\alpha_1}{c_1(\hat{\mathbf{s}}) + \alpha_3} + \alpha_2 c_2(\mathbf{y}) \quad (5.22)$$

sendo α_1 , α_2 e α_3 constantes a serem previamente escolhidas. É importante observar que o propósito de α_3 é limitar o primeiro termo da Equação 5.22 quando $c_1(\hat{\mathbf{s}}) \rightarrow 0$. Valores adequados para as três constantes serão indicados a seguir. O número de componentes independentes a serem estimados (K) precisa ser escolhido a priori. Na prática, se K for desconhecido, pode-se utilizar um procedimento semelhante ao descrito na Seção 5.5.1 para a escolha do número de componentes principais de discriminação.

A função custo que avalia a independência estatística ($c_1(\hat{\mathbf{s}})$) utiliza uma medida do cumulante de quarta ordem, semelhante à proposta no algoritmo JADE para a ICA [146]:

$$c_1(\hat{\mathbf{s}}) = \sum_{\substack{i,j=1 \\ i < j}}^K \sum_{l,m=1}^K \text{cum}\{s_i, s_j, s_l, s_m\}^2 \quad (5.23)$$

sendo $\text{cum}\{s_i, s_j, s_l, s_m\}$ o cumulante de quarta ordem [4]:

$$\begin{aligned} \text{cum}\{s_i, s_j, s_l, s_m\} = & E\{s_i, s_j, s_l, s_m\} - E\{s_i, s_j\}E\{s_l, s_m\} \\ & - E\{s_i, s_l\}E\{s_j, s_m\} - E\{s_i, s_m\}E\{s_j, s_l\} \end{aligned} \quad (5.24)$$

Um modo para calcular a medida da independência baseada no cumulante de quarta ordem foi proposta em [183], e foi utilizada neste trabalho para estimar $c_1(\hat{\mathbf{s}})$. É interessante notar que $c_1(\hat{\mathbf{s}})$ é uma sempre não-negativa e zero para sinais independentes, então, maximizar a independência entre os componentes $\hat{\mathbf{s}}$ implica em minimizar $c_1(\hat{\mathbf{s}})$.

A função custo que avalia a eficiência de discriminação é o índice soma-produto (SP) normalizado, que para um problema de classificação em M classes é descrito por:

$$c_2(\mathbf{y}) = \frac{\sum_{i=1}^M Ef_i}{M} \times \sqrt[M]{\prod_{i=1}^M Ef_i} \quad (5.25)$$

onde Ef_i é a eficiência de discriminação obtida para a classe i . A função definida na Equação 5.25 varia no intervalo $[0,1]$ e alcança o máximo quando $Ef_i = 1$ para $i = 1, \dots, M$ (eficiência total). Uma característica de $c_2(\mathbf{y})$ é sua sensibilidade a degradação da eficiência de qualquer classe.

Considerando que $c_1(\hat{\mathbf{s}}) \geq 0$ e $0 \leq c_2(\mathbf{y}) \leq 1$, as constantes α_i , na Equação 5.22 são escolhidas para produzir $0 \leq c(\hat{\mathbf{s}}, \mathbf{y}) \leq 1$. Usando por exemplo, $\alpha_1 = \alpha_3/2$, $\alpha_2 = 0.5$ e $\alpha_3 = 0.001$, o mesmo fator de ponderação é dado para os termos de ambos, c_1 e c_2 .

A rotina utilizada para a otimização de $c(\hat{\mathbf{s}}, \mathbf{y})$ é um algoritmo genético simples, conforme mostrado em [156], ao qual foram adicionados, elitismo, *crossover* uniforme e genocídio periódico. Mais detalhes a respeito são mostrados no Apêndice C.

Capítulo 6

Metodologia Proposta

Os sistemas de filtragem propostos neste trabalho são divididos em duas etapas distintas. Inicialmente, os sinais medidos são pré-processados (utilizando modelos de ICA/NLICA) para que suas características discriminantes se tornem mais acessíveis. As características extraídas são utilizadas como entrada para os classificadores propriamente ditos (classificadores neurais supervisionados na arquitetura MLP). Nesta Seção, serão descritos os sistemas de filtragem e o procedimento de treinamento utilizados.

Os métodos de extração de características utilizados neste trabalho operam sobre os sinais dos calorímetros do ATLAS formatados em anéis (conforme descrito no Capítulo 4). O propósito é aumentar a eficiência de discriminação de elétrons do *Neural Ringer* (discriminador de elétrons alternativo do segundo nível de filtragem online do ATLAS).

Considerando que cada assinatura é descrita por 100 anéis, gerados em sete camadas do calorímetro, e que, cada camada apresenta diferentes características físicas (quantidade e tipo de sensores)

A estimação das características discriminantes dos anéis pode ser realizada de duas formas distintas conforme descrito a seguir.

6.1 Parâmetros de Avaliação do Desempenho

Para avaliar o desempenho de discriminação são utilizados a curva ROC (Receiver Operating Characteristic) [184], que mostra como as probabilidades de detecção e

falso alarme (respectivamente PD e PF) variam com o patamar de decisão, e o índice SP [68], que é definido por:

$$SP = \frac{(Ef_e + Ef_j) \times \sqrt{(Ef_e \times Ef_j)}}{2} \quad (6.1)$$

onde Ef_e e Ef_j são as eficiências obtidas, respectivamente, para elétrons e jatos.

(SP, ROC, definir e comentar)

6.2 Extração de Características

...

6.2.1 Características Globais

Os sinais em anéis produzidos em todas as camadas do calorímetro podem ser juntados em um único vetor de características (de 100 componentes). Este vetor é então utilizado como entrada para os algoritmos de aprendizado estatístico, este procedimento é ilustrado na Figura 6.1

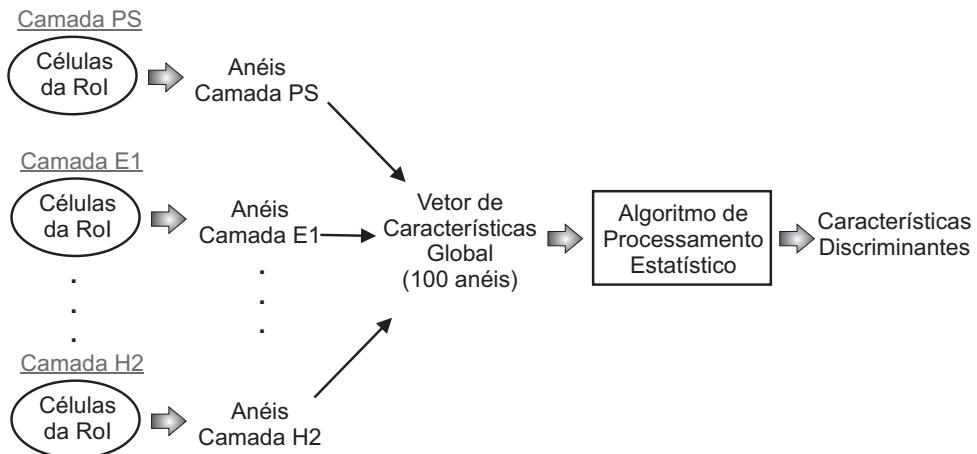


Figura 6.1: Processo de extração de características globais.

6.2.2 Características Segmentadas

Alternativamente, as características discriminantes podem ser estimadas separadamente para os anéis produzidos em cada camada do calorímetro. Este procedimento, em geral, produz resultados de mais fácil interpretação física pois sabe-se que cada

camada do calorímetro possui características distintas como o tipo e a granularidade das células detectoras. A Figura 6.2 ilustra este procedimento.

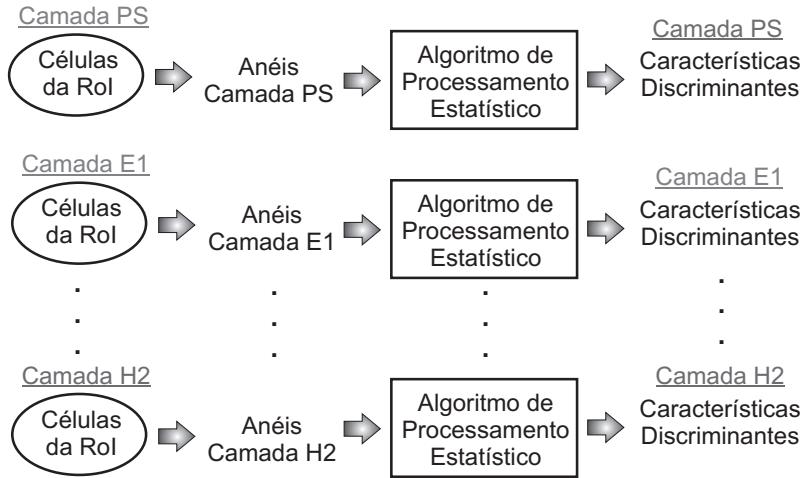


Figura 6.2: Processo de extração de características segmentadas.

6.3 Classificadores Neurais Supervisionados

Neste trabalho foram utilizados classificadores neurais tipo perceptrons de múltiplas camadas (MLP). Mais detalhes sobre classificação de sinais e a implementação de classificadores neurais supervisionados podem ser encontrados no Apêndice B. Os classificadores, assim como os algoritmos de extração de características, podem ser aplicados de modo global e de modo segmentado.

6.3.1 Especificações dos Classificadores Neurais Utilizados

Neste trabalho, foram utilizados classificadores neurais supervisionados na arquitetura Perceptron de Múltiplas camadas (MLP - *Multi-Layer Perceptron*) [5]. Foram utilizadas redes com uma camada oculta e um neurônio na camada de saída. Todos os neurônios utilizam funções de ativação tipo tangente hiperbólica. Em trabalhos anteriores [9, 10], bons resultados de classificação foram obtidos com redes de 10 neurônios na camada oculta, desta forma, decidiu-se utilizar o mesmo número de neurônios para esta camada.

Para o treinamento dos classificadores segmentados foi utilizado o algoritmo RPROP (*Resilient Back-propagation*) [185].

O treinamento dos discriminadores foi realizado a partir de um processo de validação cruzada. Os sinais disponíveis para cada classe foram, inicialmente, divididos de modo aleatório em 12 grupos (com aproximadamente a mesma quantidade de eventos). A seguir, para cada processo de treinamento, sorteava-se 4 grupos distintos (em cada classe) para compor os conjuntos de treino, validação e teste. O número de neurônios na camada escondida dos classificadores neurais foi escolhido, a partir de testes exaustivos, visando maximizar a eficiência de discriminação.

Para o treinamento dos classificadores segmentados foi utilizado o algoritmo RPROP (*Resilient Back-propagation*) [185]. Foram utilizadas redes MLP com uma camada escondida, um neurônio de saída e função de ativação tangente hiperbólica. O número de neurônios da camada escondida foi obtido após testes exaustivos com várias configurações.

Explicar a validação cruzada e o modo de parada do algoritmo (falar do RPROP)
Modo de treinamento (validação cruzada ... ver torres)

6.3.2 Classificador Global

O classificador global opera sobre um conjunto de características discriminantes globais, produzindo a decisão elétron/jato. A Figura 6.3 ilustra a operação do classificador global.

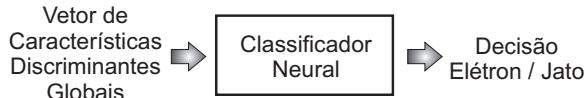


Figura 6.3: Decisão utilizando classificador global.

6.3.3 Classificadores Segmentados

Considerando o sistema de calorímetros do detector ATLAS, a informação disponível está segmentada em diversas camadas com características físicas e granularidade diferentes. Para melhor explorar as características do detector, está sendo proposto o uso de um conjunto de classificadores neurais, cada um especialista na informação de uma camada do calorímetro (ver Figura 6.4).

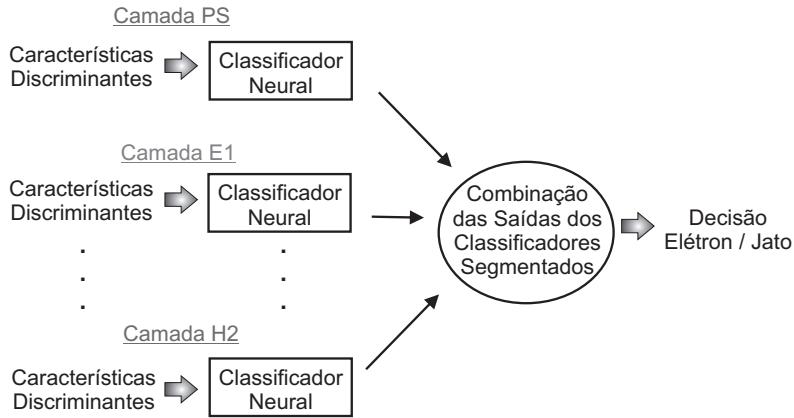


Figura 6.4: Decisão utilizando classificadores segmentados.

Um problema que surge na utilização de múltiplos classificadores segmentados é como combinar suas saídas para produzir da decisão final. Neste contexto, pode-se utilizar as informações de dois modos, através dos rótulos de classe atribuídos por cada classificador ou utilizando-se a saída contínua (que no caso dos classificadores utilizados varia de -1 a 1).

6.3.4 Combinação de Múltiplos Classificadores

Quando estão disponíveis informações de múltiplos classificadores, surge o problema de como combiná-las de forma ótima. A depender do tipo de saída escolhida para os classificadores, sendo variáveis contínuas (com excursão de -1 a 1) ou variáveis discretas (rótulos de classe), a combinação pode ser realizada através de estratégias distintas [186].

Considerando K classificadores com saídas contínuas u_k , uma forma usualmente utilizada para combinação é a média das saídas:

$$\mu(\mathbf{x}) = \sum_{k=1}^K u_k(\mathbf{x}) \quad (6.2)$$

Considerando que os múltiplos classificadores apresentam eficiência diferente, pode-se dar aos mais eficientes maior poder de decisão com o uso de fatores de ponderação α_k :

$$\mu(\mathbf{x}) = \sum_{k=1}^K \alpha_k u_k(\mathbf{x}) \quad (6.3)$$

Outra forma para a combinação de classificadores de saídas contínuas u_k é o

cálculo da média geométrica:

$$\mu(\mathbf{x}) = \sqrt[K]{\prod_{k=1}^K u_k(\mathbf{x})} \quad (6.4)$$

Alternativamente, considerando que a saída dos múltiplos classificadores é o rótulo de classe associado ao vetor de entrada \mathbf{x} , um método muito utilizado para combinação das informações é a votação da maioria [186]. Neste caso, também podem ser utilizados fatores de ponderação, caso as eficiências dos classificadores sejam diferentes. Deste modo, o voto de um classificador mais eficiente tem mais influência na decisão final.

Capítulo 7

Resultados - Dados Simulados

Nesta seção serão descritas as principais características das bases de dados utilizadas.

Os sinais podem ser divididos em duas classes, simulados e experimentais:

- **Sinais Simulados:** Em física de altas energias, o projeto dos detectores e sistemas de filtragem é realizado com o auxílio de de simuladores de colisões. Os simuladores utilizam informações a respeito da física (obtidas em experimentos anteriores e em modelos teóricos) e são capazes de reproduzir as características esperadas nas colisões de um acelerador (que ainda não está operacional). Os sinais simulados foram obtidos a partir de simuladores para colisões próton-próton [45] que utilizam a técnica de Monte Carlo [14]. As simulações consideram as características esperadas nas colisões do LHC e as especificações do ATLAS. Assim como na operação real do detector, é possível aplicar diferentes cortes (formas de seleção) de nível 1 nos dados simulados. Nos dados simulados existe a informação do tipo de partícula correspondente a cada assinatura, permitindo o treinamento supervisionado dos algoritmos de filtragem.
- **Sinais Experimentais:** São sinais que foram adquiridos no detector ATLAS, na fase de testes e nas primeiras colisões do LHC, realizadas no final de 2009 e durante o ano de 2010. Nos dados experimentais não está disponível a informação do tipo de partícula correspondente a cada assinatura (exceto no caso dos sinais produzidos por raios cósmicos). Nesse caso, outras estratégias precisam ser utilizadas para avaliar o desempenho dos algoritmos. É comum

fazer uma comparação com a identificação obtida no processo de reconstrução *offline*.

7.1 Características dos Sinais simulados

Neste trabalho foram utilizados dois conjuntos distintos de dados simulados. Ao primeiro conjunto, foi aplicado um corte de nível 1 do tipo E10 *loose* e ao segundo um corte tipo E15i.

A seleção de nível 1 a partir do Corte E10 *loose* realiza apenas um corte linear na energia total do evento, deixando passar somente eventos com energia da ordem de 10 GeV ou maior (os eventos de baixa energia são eliminados). Este corte é mais utilizado para teste do sistema de filtragem no período inicial de operação do detector, pois é pouco provável a identificação de assinaturas relevantes, uma vez que, uma grande quantidade de eventos de falso-alarme é aprovada pelo nível 1.

Na filtragem E15i, a energia mínima agora é da ordem de 15 GeV e, adicionalmente, foram efetuados cortes considerando outras características do perfil de deposição de energia como

- corte de vazamento para as camadas hadrônicas - a energia nas camadas hadrônicas é calculada e se ela for maior que um valor limite o evento é rejeitado pelo L1;
- corte de isolamento em energia - efetuado a partir da relação entre a energia no centro ($\sum E_C$) e na periferia ($\sum E_P$) do perfil de deposição de energia da segunda camada eletromagnética, se $\frac{\sum E_P}{\sum E_C}$ for maior que um limiar pré-estabelecido o evento é rejeitado pelo L1.

Este corte de nível 1 se aproxima da característica esperada para a fase operação do detector.

... Colocar tabela mostrando as quantidades de sinais de cada classe antes e depois do L1 (fatores de rejeição) ...

Na Figura ?? são mostradas as distribuições em energia, η (pseudorapidez) e ϕ (ângulo azimutal) para elétrons e jatos dos conjuntos aos quais foram aplicados os cortes de L1 tipo E10 e E15i. Percebe-se que...

Nas Figuras 7.2 e 7.3 são mostrados os eventos médios (média e desvio padrão de cada anel) para elétrons e jatos após os cortes E10 *loose* e E15i. Pode-se observar que os perfis dos elétrons são semelhantes nos dois casos. Por outro lado, os jatos E15i apresentam uma menor energia hadrônica e menor espalhamento (o que já era esperado, uma vez que foram efetuados cortes considerando estes parâmetros) do que os jatos E10, sendo mais parecidos com os elétrons e portanto representando um ruído de fundo de mais difícil identificação.

... substituir os histogramas e10 e e15i por histogramas combinados tipo root.

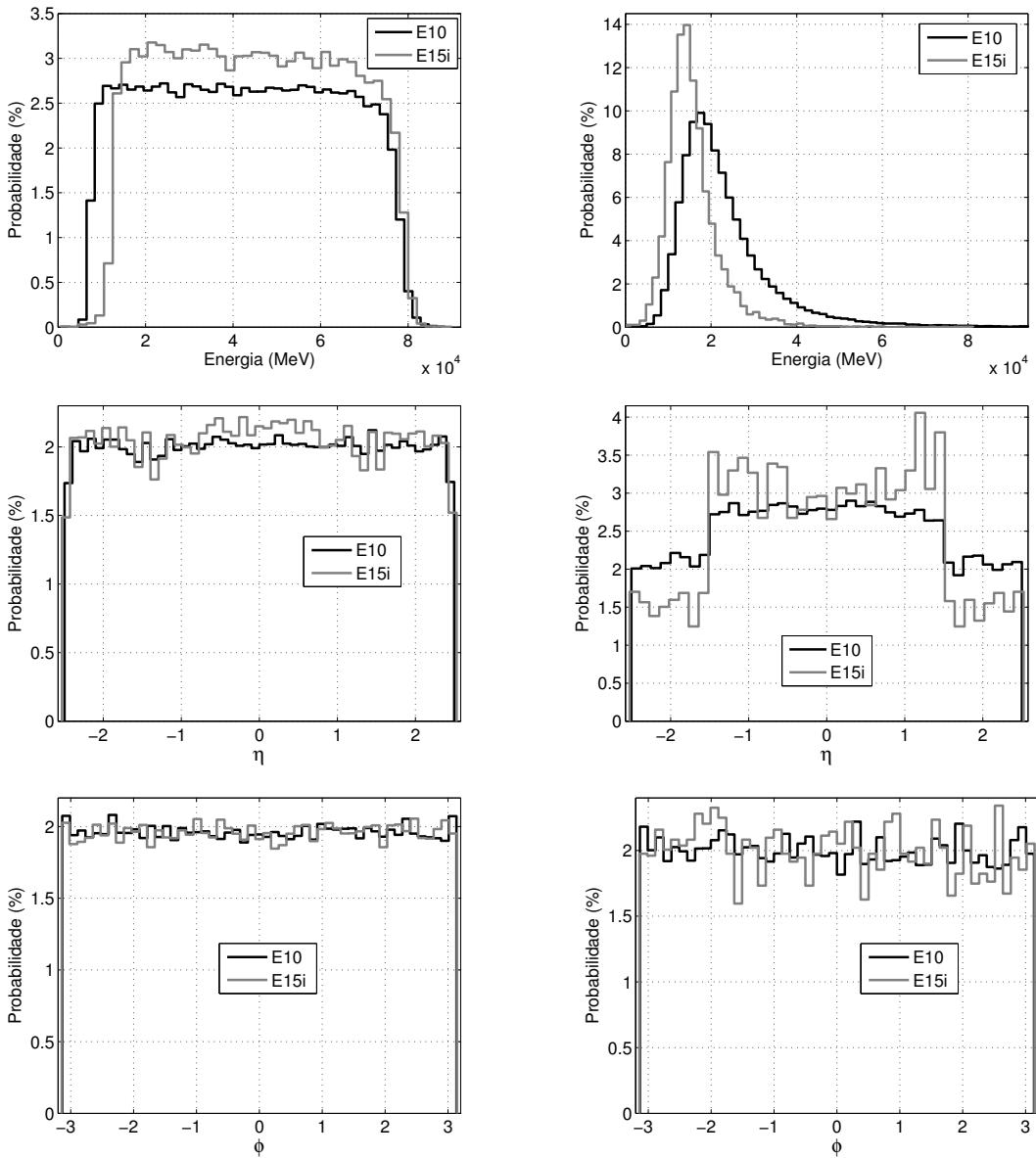


Figura 7.1: Distribuição em energia, η e ϕ das assinaturas de elétrons (esquerda) e jatos (direita) aprovadas pelos cortes de primeiro nível tipo E10 e E15i.

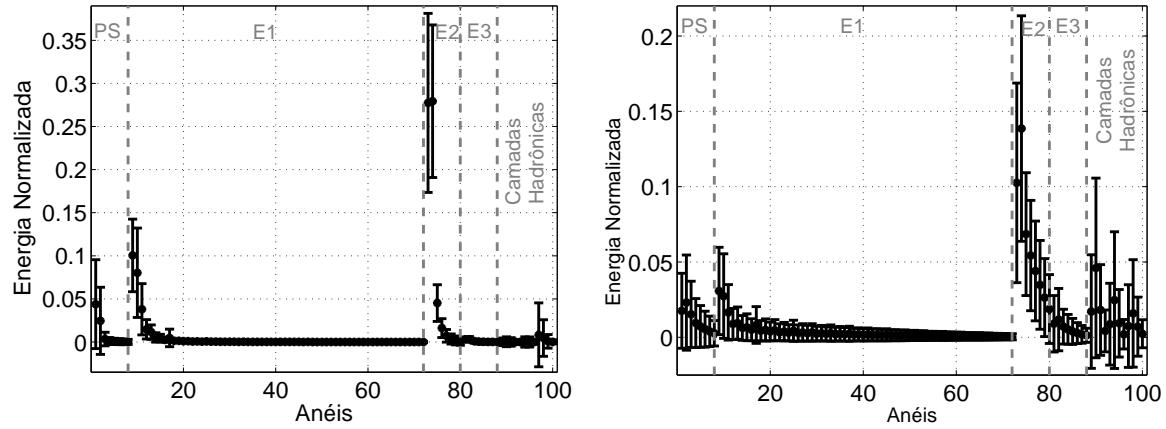


Figura 7.2: Sinais em anéis (média e desvio padrão) para elétrons (esquerda) e jatos (direita) que chegam ao L2 no corte E10.

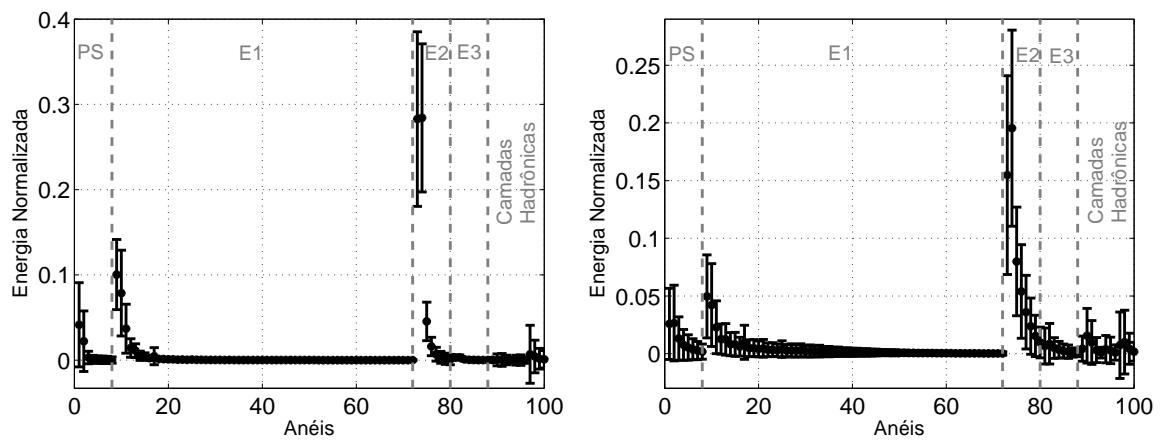


Figura 7.3: Sinais em anéis (média e desvio padrão) para elétrons (esquerda) e jatos (direita) que chegam ao L2 no corte E15.

Capítulo 8

Resultados - Dados Experimentais

Neste capítulo serão apresentados os resultados obtidos através da aplicação da metodologia proposta para sinais medidos no detector ATLAS. Os sinais utilizados são divididos em dois tipos. Inicialmente, será verificada a capacidade dos algoritmos propostos para identificar (e consequentemente eliminar) os sinais induzidos no detector por raios cósmicos. A seguir, serão utilizados dados medidos no detector na fase inicial de operação do LHC, onde a energia total dos feixes está limitada a 7 TeV (funcionando em plena capacidade, o LHC atingirá energias da ordem de 14 TeV).

Na Figura 8.1 pode-se visualizar um evento (colisão) do LHC onde foram gerados um elétron e três jatos. O elétron está marcado em vermelho e aponta para baixo. Neste diagrama, o calorímetro eletromagnético e o hadrônico estão marcados respectivamente de verde e ...??... . Percebe-se que o elétron não sensibiliza a seção eletromagnética. Um outro evento do LHC visualizado pelo ATLAS é mostrado na Figura 8.2, neste caso, pode-se observar dois múons, que possivelmente foram gerados do decaimento de um bóson Z.

8.1 Validação com Sinais de Ráios Cósmicos

Os raios cósmicos [187] são partículas originadas no espaço, que se deslocam com velocidade próxima a da luz e, ao penetrarem na atmosfera terrestre, interagem com os átomos que a constituem produzindo uma "cascata" de novas partículas menos energéticas (conhecidas como raios cósmicos secundários). O poder de penetração

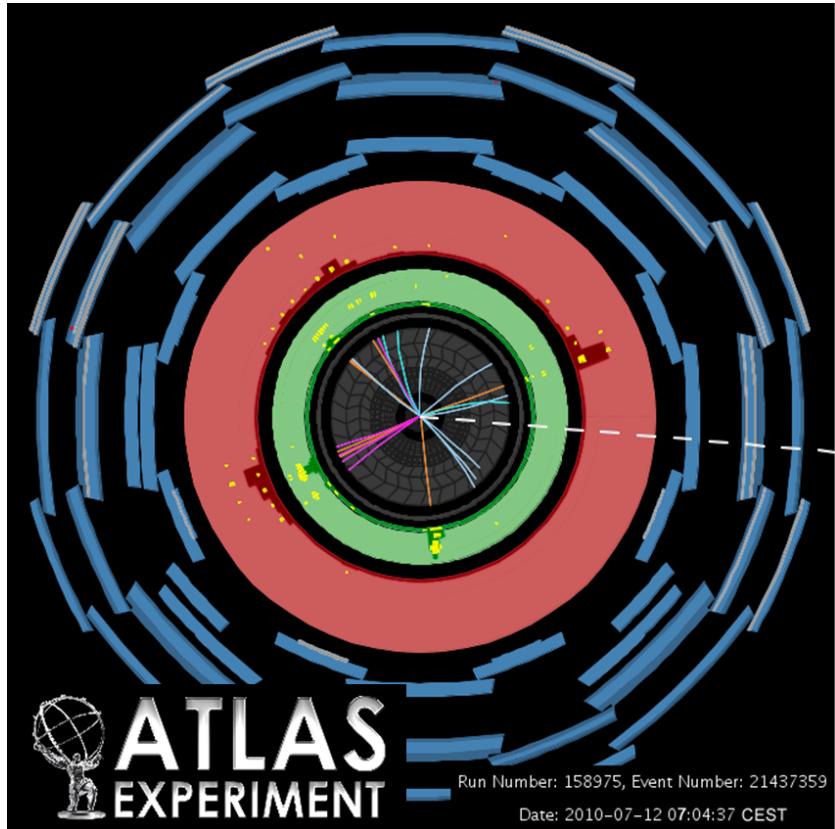


Figura 8.1: Visualização de um evento do LHC (energia total dos feixes igual a 7 GeV) onde foram produzidos um elétron (marcado em vermelho, apontando para baixo) e três jatos, extraído de [15].

dos raios cósmicos é alto, podendo atingir o detector ATLAS (instalado a uma profundidade de aproximadamente 100 m) e interagir com o material dos calorímetros. A energia destas partículas também pode ser alta, atingindo até 10^{20} eV.

Os sinais produzidos no detector pelos raios cósmicos constituem uma fonte de ruído de fundo para o experimento, que deve ser eliminada (ou pelo menos atenuada) pelo sistema de filtragem. Em momentos quando o LHC está desligado (não havendo portanto outra fonte de sinal para os calorímetros), foram coletadas diversas assinaturas originadas por raios cósmicos que serão utilizadas para verificar a robustez do sistema de filtragem a essa fonte de ruído.

Na fase inicial de operação do LHC, sua energia máxima está limitada em 7 TeV (apenas metade da energia para qual o colisionador foi projetado que é de aproximadamente 14 TeV, porém um valor jamais alcançada por outro experimento). Nos dados simulados, utilizados para projeto dos discriminadores, a faixa de energia

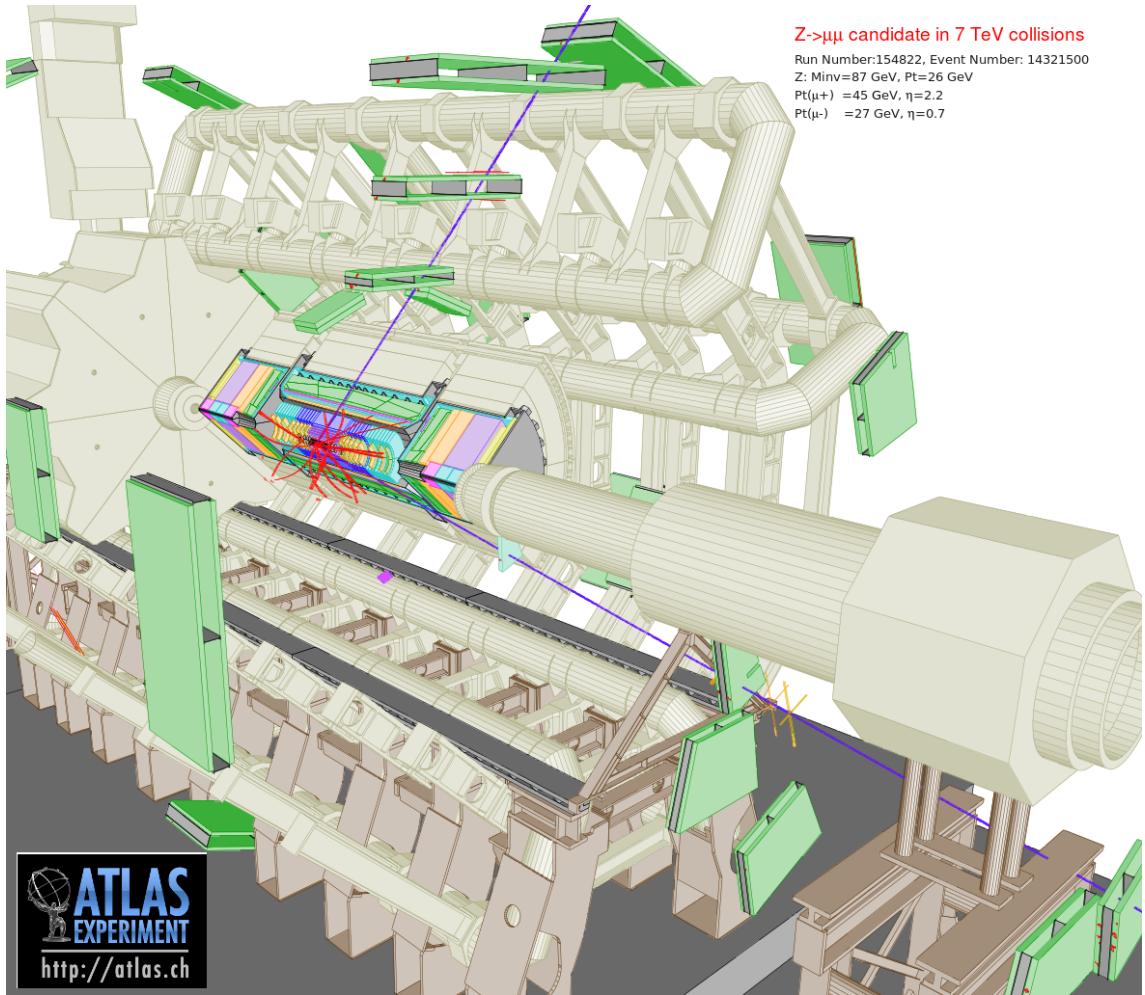


Figura 8.2: Visualização de um evento do LHC (energia total dos feixes igual a 7 GeV) onde foram produzidos dois múons a partir do decaimento de um bóson Z, extraído de [15].

dos feixes era de 14 TeV, desta forma, os resultados obtidos com a aplicação direta dos discriminadores propostos aos dados experimentais disponíveis estão seriamente comprometidos, uma vez que as características das partículas dependem fortemente da energia envolvida.

Nos sinais experimentais não há a caracterização prévia do tipo de partícula (como havia nos sinais simulados). Neste caso, foram utilizadas informações obtidas da reconstrução *offline*, que indicam a probabilidade de um certo evento ser ou não um elétron. Entre os parâmetros calculados pode-se destacar: ...

(, gráficos eta, phi e energia, qual o corte de nível 1 ??)

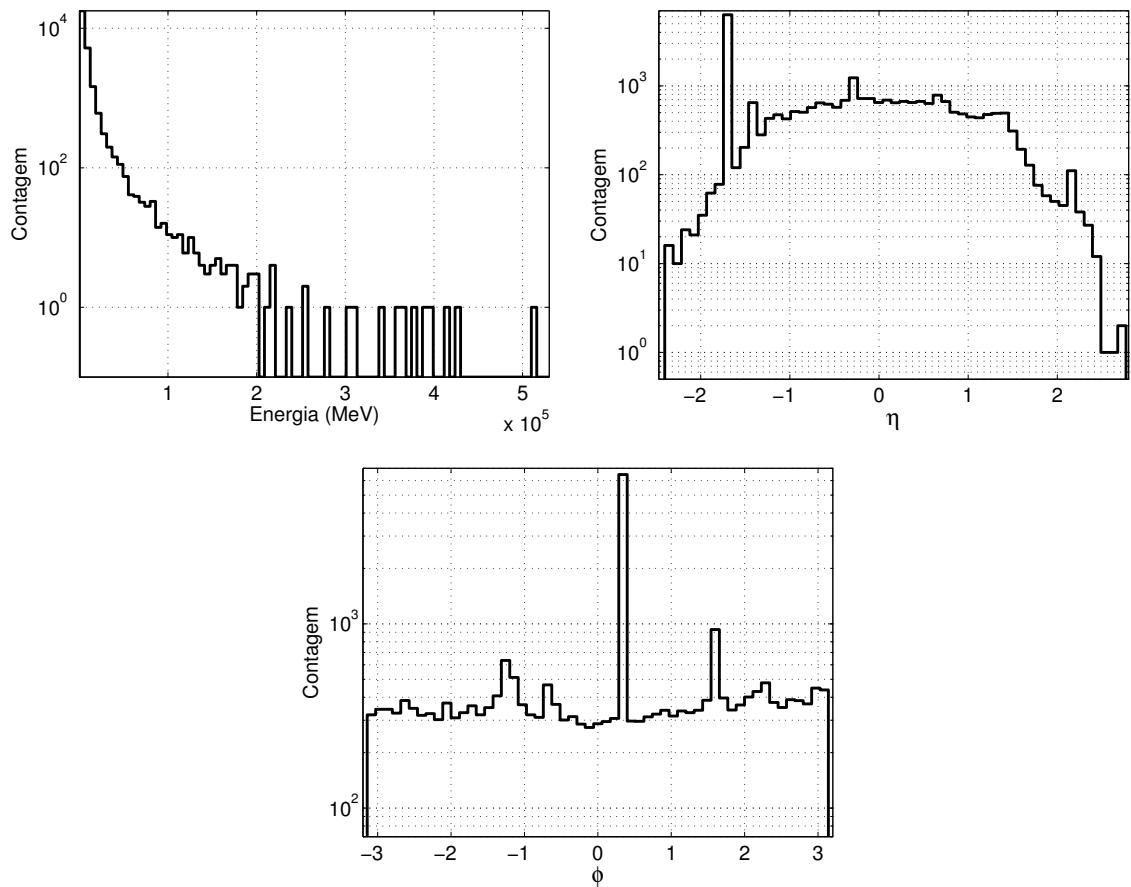


Figura 8.3: Sinais em anéis (média e desvio padrão) para elétrons (esquerda) e jatos (direita) que chegam ao L2 no corte E15.

Capítulo 9

Conclusões

Aqui vão as conclusões ...

Referências Bibliográficas

- [1] ATLAS COLABORATION, “ATLAS Experiment at CERN Large Hadron Collider”, *Journal of Instrumentation*, v. 3, n. S08003, pp. 1–407, 2008.
- [2] EVANS, L., BRYANT, P., “LHC Machine”, *Journal of Instrumentation*, v. 3, n. S08001, pp. 1–158, 2008.
- [3] JOLLIFFE, I. T., *Principal Component Analysis*. 2nd ed. Springer: Nova York, Estados Unidos, 2002.
- [4] HYVARINEN, A., KARHUNEN, J., OJA, E., *Independent Component Analysis*. Wiley: Nova York, Estados Unidos, 2001.
- [5] HAYKIN, S., *Neural Networks and Learning Machines*. 3rd ed. Prentice Hall: Nova Jersey, Estados Unidos, 2008.
- [6] COMMITTEE ON ELEMENTARY PARTICLE PHYSICS IN THE 21ST CENTURY, N. R. C., *Revealing the Hidden Nature of Space and Time: Charting the Course for Elementary Particle Physics*. National Academic Press: Washington-DC, Estados Unidos, 2006.
- [7] ATLAS-COLLABORATION, *Letter of Intents for a General Purpose pp Experiment at the Large Hadron Collider at CERN*, Tech. rep., CERN, Genebra, Suíça, Outubro 1992.
- [8] SEIXAS, J. M., L. CALÔBA, M. S., “Neural Second-Level trigger system based on calorimetry”, *Computer Physics Communications*, v. 95, n. 2-3, pp. 143–157, Junho 1996.
- [9] DOS ANJOS, A. R., *Sistema de filtragem online aplicada a um ambiente com alta taxa de eventos*, Ph.D. Thesis, COPPE/UFRJ, Dezembro 2006.

- [10] TORRES, R. C., *Sistema Online de Filtragem em um Ambiente com Alta Taxa de Eventos e Fina Granularidade*, Ph.D. Thesis, COPPE/UFRJ, Março 2010.
- [11] CALOBA, L., SEIXAS, J., PEREIRA, F., “Neural Discriminating Analysis for a Second-Level Trigger System”. In: *Proceedings of the International Conference on Computing in High Energy Physics (CHEP95)*, Rio de Janeiro, Brasil, Setembro 1995.
- [12] WIGMANS, R., *Calorimetry: Energy Measurement in Particle Physics*. Clarendon Press: Gloucestershire, Reino Unido, 2000.
- [13] ALMEIDA, L. B., *Nonlinear Source Separation*. Morgan and Claypool, 2006.
- [14] ROBERT, C., CASELLA, G., *Monte Carlo Statistical Methods*. Springer-Verlag: Nova York, Estados Unidos, 2004.
- [15] ATLAS-COLLABORATION, “Experiment Overview”, Disponível em :<<http://www.atlas.ch/>>. Acesso em: 27 de agosto de 2010.
- [16] FERNOW, R. C., *Introduction to Experimental Particle Physics*. Cambridge University Press: Cambridge-Reino Unido, 1986.
- [17] MARTIN, B. R., *Nuclear and Particle Physics, An Introduction*. Wiley: West Sussex, Reino Unido, 2006.
- [18] CHUNG, K. C., *Introdução à Física Nuclear*. EdUERJ: Rio de Janeiro, 2001.
- [19] VELTMAN, M. J. G., *Facts and Mysteries in Elementary Particle Physics*. World Scientific: Nova Jersey, Estados Unidos, 2003.
- [20] ELLIS, R. K., STIRLING, W. J., WEBBER, B. R., *QCD and Collider Physics*. Cambridge Monographs on Particle Physics, Nuclear Physics and Cosmology, Cambridge University Press: Cambridge, Reino Unido, 2003.
- [21] PRICE, M., “The LHC project”, *Nuclear Instruments and Methods in Physics Research A*, v. 478, pp. 46–61, 2002.

- [22] CERN-COLABORATION, “CERN”, Disponível em :<<http://www.cern.ch/>>. Acesso em: 20 de março de 2007.
- [23] BERTONE, G., HOOPER, D., SILK, J., “Particle dark matter, evidence, candidates and constraints”, *Physics Reports*, v. 405, pp. 279–390, 2005.
- [24] LYKKEN, J. D., “Introduction to Supersymmetry”, Lectures at the Theoretical Advanced Study Institute (TASI 96):Fields, Strings, and Duality, Junho 1996.
- [25] BIGI, I., SANDA, A., *CP violation*. Cambridge University Press: Cambridge, Reino Unido, 1999.
- [26] DESY, “Deutsches Elektronen-SynchrotronDie”, Disponível em :<<http://www.desy.de/>>. Acesso em: 02 de julho de 2009.
- [27] KEK, “KEK High-Energy Accelerator”, Disponível em: <<http://www.kek.jp/>>. Acesso em: 02 de julho de 2009.
- [28] FERMILAB, “Fermi National Accelerator Laboratory”, Disponível em :<<http://www.fnal.gov/>>. Acesso em: 02 de julho de 2009.
- [29] SLAC, “SLAC Laboratory”, Disponível em: <<http://www.slac.stanford.edu/>>. Acesso em: 02 de julho de 2009.
- [30] BROOKHAVEN, “Brookhaven National Laboratory”, Disponível em: <<http://www.bnl.gov/>>. Acesso em: 02 de julho de 2009, 2009.
- [31] LHC-COLLABORATION, “Large Hadron Collider”, Disponível em: <<http://lhcb.web.cern.ch/lhc/>>. Acesso em: 28 de março de 2007.
- [32] CMS COLABORATION, “CMS Experiment at the CERN LHC”, *Journal of Instrumentation*, v. 3, n. S08004, pp. 1–334, 2008.
- [33] LHCb COLABORATION, “LHCb Detector at the LHC”, *Journal of Instrumentation*, v. 3, n. S08005, pp. 1–205, 2008.
- [34] LHCf COLABORATION, “LHCf Experiment at the CERN Large Hadron Collider”, *Journal of Instrumentation*, v. 3, n. S08006, pp. 1–36, 2008.

- [35] ALICE COLABORATION, “ALICE Experiment at the CERN LHC”, *Journal of Instrumentation*, v. 3, n. S08002, pp. 1–245, 2008.
- [36] TOTEM COLABORATION, “TOTEM Experiment at the CERN Large Hadron Collider”, *Journal of Instrumentation*, v. 3, n. S08007, pp. 1–107, 2008.
- [37] ROSSI, L., “The ATLAS tracking and vertexing detector”, *Nuclear Instruments and Methods in Physics Research Section A*, v. 580, n. 2, 2007.
- [38] PUZO, P., “ATLAS Calorimetry”, *Nuclear Instruments and Methods in Physics Research A*, v. 494, n. 1, pp. 340–345, 2002.
- [39] PONTECORVO, L., “The ATLAS muon spectrometer”, *The European Physical Journal C - Particles and Fields*, v. 34, n. 1, 2004.
- [40] PADILLA, C., “The ATLAS Trigger System”. In: *Proceedings of the 16th IEEE-NPSS Real Time Conference*, v. 1, pp. 326–333, Pequim, China, 2009.
- [41] DREES, J., “Review of final LEP results or a tribute to LEP”. In: *Proceedings of the International Symposium on Lepton and Photon Interactions at High Energies*, pp. 349–373, Roma, Itália, Julho 2001.
- [42] KANE, G., PIERCE, A., (eds), *Perspectives on LHC Physics*. World Scientific: Singapura, 2008.
- [43] ATLAS-COLLABORATION, *ATLAS Detector and Physics Performance Technical Design Report, Volume 2*, Tech. rep., CERN, Genebra, Suiça, Maio 1999.
- [44] HAUSER, R., “The ATLAS High Level Trigger System”. In: *Proceedings of the U.S. Atlas Computing Meeting*, v. 1, pp. 1–5, Upton, EUA, 2003.
- [45] ATLAS-COLLABORATION, *ATLAS Detector and Physics Performance Technical Design Report, Volume 1*, Tech. rep., CERN, Genebra, Suiça, Maio 1999.

- [46]ADRAGNA, ET AL., P., “The ATLAS Hadronic Tile Calorimeter: From Construction Toward Physics”, *IEEE Transactions on Nuclear Science*, v. 53, n. 5, pp. 1275–1281, Junho 2006.
- [47]WATTS, G., “Review of Triggering”. In: *Proceedings of the IEEE Nuclear Science Symposium and Medical Imaging Conference*, v. 1, pp. 282–287, Portland, EUA, Outubro 2003.
- [48]VOLKER LINDENSTRUTH, I. K., “Overview of trigger systems”, *Nuclear Instruments and Methods in Physics Research A*, v. 535, n. 1, pp. 48–56, 2004.
- [49]DA COSTA, C., *Projeto de Circuitos Digitais com FPGA*. Érica, 2009.
- [50]DINIZ, P. S. R., DA SILVA, E. A. B., NETO, S. L., *Processamento Digital de Sinais*. Bookman, 2004.
- [51]SMITH, M. J. S., *Application Specific Integrated Circuits*. Addison-Wesley Professional: Nova York, Estados Unidos, 1997.
- [52]CDF-COLLABORATION, “The CDF Experiment”, Disponível em: <<http://www-cdf.fnal.gov/>>. Acesso em: 21 de abril de 2010.
- [53]D0-COLLABORATION, “The D0 Experiment”, Disponível em: <<http://www-cdf.fnal.gov/>>. Acesso em: 21 de abril de 2010.
- [54]RIU, I., ABOLINS, M., ET. AL, P. A., “Integration of the Trigger and Data Acquisition Systems in ATLAS”, *IEEE Transactions on Nuclear Science*, v. 55, n. 1, pp. 106–112, 2008.
- [55]SEEZ, C., “The CMS trigger system”, *The European Physical Journal C - Particles and Fields*, v. 34, n. 1, pp. 151–159, 2004.
- [56]KIESLING, C., JANAUSCHEK, L., PLACAKYTE, R., et al., “The H1 neural network trigger”. In: *Proceedings of the Intelligent Sensors, Sensor Networks and Information Processing Conference*, v. 1, pp. 319–324, Melbourne, Austrália, 2004.

- [57] LINDSEY, C., DENBY, B., HAGGERTY, H., et al., “Real Time Track Finding with a VLSI Neural Network”, *Nuclear Instruments and Methods in Physics Research A*, v. A, n. 317, pp. 346–356, 1992.
- [58] LINDSEY, C. S., DENBY, B., HAGGERTY, H., “Drift chamber tracking with neural networks”, *IEEE Transactions on Nuclear Science*, v. 40, n. 4, pp. 607–614, 2002.
- [59] CUTTS, D., HOFTUN, J. S., SORNBORGER, A., “The use of neural networks in the D0 data aquisition system”, *IEEE Transaction on Nuclear Science*, v. 36, n. 5, pp. 1490–1493, 1989.
- [60] DENBY, B., “Neural Networks in high-energy physics, a ten year perspective”, *Computer Physics Communications*, v. 1, n. 119, pp. 219–231, 1999.
- [61] NGUYET, T. T., “Searches for new physics by the H1 experiment at HERA”, *Acta Physica Polonica B Proceedings Supplement*, v. 1, n. 2, pp. 407–409, 2008.
- [62] ZIMMERMANN, J., KIESLING, C., “Neural networks for the H1 experiment”. In: *Proceedings of the IEEE Nuclear Science Symposium*, v. 3, pp. 1869–1872, Roma, Itália, 2004.
- [63] KOLANOSKI, H., “Application of artificial neural networks in particle physics”, *Nuclear Instruments and Methods in Physics Research A*, v. 1, n. 367, pp. 14–20, 1995.
- [64] SPENCER, J. E., “Real-Time Applications of Neural Nets”, *IEEE Transactions on Nuclear Science*, v. 36, n. 5, pp. 1485–1489, 1989.
- [65] DENBY, B., GARDA, P., GRANADO, B., et al., “Fast triggering in high-energy physics experiments using hardware neural networks”, *IEEE Transactions on Neural Networks*, v. 14, n. 5, pp. 1010–1027, 2003.
- [66] WON, E., “A hardware implementation of artificial neural networks using field programmable gate arrays”, *Nuclear Instruments and Metohds in Physics Research A*, v. 581, n. 1, pp. 816–820, 2007.

- [67] WESTERHOFF, S., MEYER, H., “Neural nets as a tool for separating gamma and hadron induced Air Showers”. In: *Proceedings of the AIHENP*, v. 1, n. 1, pp. 1–8, Pisa, Itália, 1995.
- [68] DOS ANJOS, A., TORRES, R., SEIXAS, J., “Neural triggering system operating on high resolution calorimetry information”, *Nuclear Instruments and Methods in Physics Research A*, v. 559, n. 1, pp. 134–138, 2006.
- [69] PRORIOL, J., “Selection of variables for neural network analysis. Comparison of several methods with high energy physics analysis”, *Nuclear Instruments and Methods in Research A*, v. 361, n. 3, pp. 581–585, 1995.
- [70] WHITESON, S., WHITESON, D., “Machine learning for event selection in high energy physics”, *Engineering Applications of Artificial Intelligence*, v. 22, n. 8, pp. 1203–1217, 2009.
- [71] RIGGI, S., CARUSO, R., INSOLIA, A., et al., “A neural network approach to event-by-event cosmic ray primary mass identification”, *Proceedings of Science*, v. ACAT07, n. 035, pp. 1–13, 2007.
- [72] MARONE, A., ANS F. CADINI, C. F., ZIO, E., et al., “Employing Neural Networks to Determine the Position of Interaction of Medium-High Energy Gamma Rays”. In: *Proceedings of the IEEE Nuclear Science Symposium*, v. 1, pp. 645–649, Orlando, Estados Unidos, 2009.
- [73] CHOULAKIAN, V., “The optimality of the centroid method”, *Psychometrika Journal*, v. 68, n. 3, pp. 473–475, 2006.
- [74] ALDRICH, J., “R. A. Fisher and the making of maximum likelihood 1912–1922”, *Statistical Science*, v. 12, n. 3, pp. 162–176, 1997.
- [75] ATLAS-COLLABORATION, *First-Level Trigger Technical Design Report*, Tech. rep., CERN, Genebra, Suiça, Junho 1998.
- [76] ATLAS-COLLABORATION, *ATLAS High-Level Trigger, Data Acquisition and Controls Technical Design Report*, Tech. rep., CERN, Genebra, Suiça, Maio 2003.

- [77] CERQUEIRA, A. S., CALÔBA, L. P., SEIXAS, J. M., “Analog System for Building the First-Level Triggering Signal Provided By the Hadronic Calorimeter of ATLAS Detector”, *Nuclear Instruments and Methods in Physics Research - A*, v. 570, n. 1, pp. 117–125, 2007.
- [78] GARVEY, J., HILLIER, S., MAHOUT, G., et al., “Use of an FPGA to identify electromagnetic clusters and isolated hadrons in the ATLAS level-1 calorimeter trigger”, *Nuclear Instruments and Methods in Physics Research A*, v. 512, n. 1, pp. 400–432, 2003.
- [79] DOS ANJOS, A., ABOLINS, M., ARMSTRONG, S., “The Second Level Trigger of the ATLAS Experiment at CERN LHC”, *IEEE Transactions on Nuclear Science*, v. 51, n. 3, pp. 909–914, 2004.
- [80] LANE, W. G., CUNNINGHAM, D., *Gigabit Ethernet Networking*. Sams: Nova York, Estados Unidos, 1999.
- [81] SHILDT, H., *C++: The Complete Reference*. McGraw-Hill: Nova York, Estados Unidos, 2003.
- [82] ATLAS-COLLABORATION, *Athena Developer Guide*, Tech. rep., CERN, Genebra, Suiça, Agosto 2001.
- [83] CORTI, G., CATTANEO, M., CHARPENTIER, M., “Software for the LHCb Experiment”, *IEEE Transactions on Nuclear Science*, v. 53, n. 3, pp. 1323–1328, 2006.
- [84] SCHIAVI, C., “Implementation and Performance of the High Level Trigger Electron and Photon Selection for the ATLAS Experiment at the LHC”, *IEEE Transactions on Nuclear Science*, v. 53, n. 5, pp. 1424–1429, 2006.
- [85] MELLO, A. G., DOS ANJOS, A., ARMSTRONG, S., et al., “Overview of the High-Level Trigger Electron and Photon Selection for the ATLAS Experiment at the LHC”, *IEEE Transactions on Nuclear Science*, v. 53, n. 5, pp. 2839–2843, 2006.
- [86] BISHOP, C. M., *Neural Networks for Pattern Recognition*. Clarendon Press: Gloucestershire, Reino Unido, 1995.

- [87] COVER, T. M., THOMAS, J. A., *Elements of Information Theory*. Wiley: Nova Jersey, Estados Unidos, 2006.
- [88] LIMA JR., H., SEIXAS, J., “A segmented principal component analysis applied to calorimetry information at ATLAS”, *Nuclear Instruments and Methods in Physics Research A*, v. 559, n. 1, pp. 129–133, Abril 2006.
- [89] SIMAS FILHO, E., SEIXAS, J. M., CALOBA, L. P., “Nonlinear Independent Component Analysis, Theoretical Review and Applications”, *Learning and Nonlinear Models*, v. 5, n. 2, pp. 99–120, 2007.
- [90] HARMAN, H. H., *Modern Factor Analysis*. 2nd ed. University of Chicago Press: Chicago, Estados Unidos, 1967.
- [91] PARK, H. M., OH, S. H., LEE, S. Y., “Adaptive noise cancelling based on independent component analysis”, *Electronics Letters*, v. 38, n. 15, pp. 832–833, 2002.
- [92] MOURA, N. N., FILHO, E. F. S., SEIXAS, J. M., “Advances in Sonar Technology”, chap. Independent Component Analysis for Passive Sonar Signal Processing, pp. 91–110, In-Tech, 2009.
- [93] SARPERI, L., ZHU, X., NANDI, A. K., “Blind OFDM receiver based on independent component analysis for MIMO systems”, *IEEE Transactions on Wireless Communications*, v. 6, n. 11, pp. 4079–4089, 2007.
- [94] KWAK, K. C., PEDRYCZ, W., “Face recognition using an enhanced independent component analysis approach”, *IEEE Transactions on Neural Networks*, v. 18, n. 2, pp. 530–541, 2007.
- [95] ESCUDERO, J., HORNERO, R., ABASOLO, D., et al., “Artifact removal in magneto-encephalogram background activity with independent component analysis”, *IEEE Transactions on Biomedical Engineering*, v. 54, n. 11, pp. 1965–1973, 2007.
- [96] PAPOULIS, A., *Probability, Random Variables, and Stochastic Processes*. 3rd ed. McGraw-Hill: Nova York, Estados Unidos, 1991.

- [97] CICHOCKI, A., AMARI, S., *Adaptive Blind Signal and Image Processing*. John Wiley and Sons: West Sussex, Reino Unido, 2002.
- [98] HYVÄRINEN, A., OJA, E., “Independent Component Analysis: Algorithms and Applications”, *Neural Networks*, v. 13, n. 4-5, pp. 411–430, 2000.
- [99] JUTTEN, C., KARHUNEN, J., “Advances in Nonlinear Blind Source Separation”. In: *Proceedings of the 4th Int. Symp. on Independent Component Analysis and Blind Signal Separation*, v. 1, pp. 245–256, Nara, Japão, 2003.
- [100] HYVÄRINEN, A., PAJUNEN, P., “Nonlinear Independent Component Analysis: Existence and Uniqueness results”, *Neural Networks*, v. 12, n. 3, pp. 429–439, 1999.
- [101] ROJAS, F., PUNTONET, C. G., ROJAS, I., “Independent component analysis evolution based method for nonlinear speech processing”, *Artificial Neural Nets Problem Solving Methods, PT II*, v. 2687, pp. 679–686, 2003.
- [102] MIYABE, S., JUANG, B.-H., SARUWATARI, H., et al., “Kernel-based nonlinear independent component analysis for underdetermined blind source separation”. In: *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, v. 1, pp. 1641–1644, Taipei, Taiwan, 2009.
- [103] HARITOPOULOS, M., YIN, H., ALLINSON, N. M., “Image denoising using self-organizing map-based nonlinear independent component analysis”, *Neural Networks*, v. 15, n. 8-9, pp. 1085–1098, 2002.
- [104] ALMEIDA, M. S., ALMEIDA, L. B., “Wavelet-based separation of nonlinear show-through and bleed-through image mixtures”, *Neurocomputing*, v. 72, n. 1-3, pp. 57–70, 2008.
- [105] LU, C.-J., WU, J.-Y., FAN, C.-R., et al., “Forecasting stock price using Nonlinear independent component analysis and support vector regression”. In: *Proceedings of the IEEE International Conference on Industrial Engi-*

neering and Engineering Management, v. 1, pp. 2370–2374, Hong Kong, 2009.

- [106] DUARTE, L. T., JUTTEN, C., “A Mutual Information Minimization Approach for a Class of Nonlinear Recurrent Separating Systems”. In: *Proceedings of the International Workshop on Machine Learning for Signal Processing*, v. 1, pp. 122–127, Thessaloniki, Greece, 2007.
- [107] DUARTE, L. T., JUTTEN, C., MOUSSAOUI, S., “Ion Selective Electrode Array Based on a Bayesian Nonlinear Source Separation Method”. In: *Independent Component Analysis And Signal Separation, 8th International Conference, Lecture Notes In Computer Science*, pp. 662–669, Springer: Paraty, Brasil, Março 2009.
- [108] TALEB, A., JUTTEN, C., “Source Separation in post-nonlinear mixtures”, *IEEE Transactions on Signal Processing*, v. 47, n. 10, pp. 2807–2820, 1999.
- [109] ROJAS, F., PUNTONET, C. G., RODRÍGUEZ-ÁLVAREZ, M., et al., “Blind Source Separation in Post-Nonlinear Mixtures Using Competitive Learning, Simulated Annealing, and a Genetic Algorithm”, *IEEE Transactions on Systems, Man and Cybernetics-Part C*, v. 34, n. 4, pp. 407–416, 2004.
- [110] NGUYEN, T. V., PATRA, J. C., DAS, A., “A Geometric Approach to Post-Nonlinear Mixture in Blind Source Separation”. In: *Proceedings of the International Conference on Communications Systems*, pp. 260–264, Cin-gapura, Setembro 2004.
- [111] WEI, C., KHOR, L., WOO, W., et al., “A Novel Iterative Conditional Maximization Method for Post-Nonlinear Underdetermined Blind Source Separation”. In: *Proceedings of the International Conference on Digital Signal Processing*, pp. 551–554, Cardiff, Wales, Inglaterra, Julho 2007.
- [112] SOLAZZI, M., UNCINI, A., “Spline neural networks for blind separation of post-nonlinear-linear mixtures”, *IEEE Transactions on Circuits and Systems - I: Regular Papers*, v. 51, n. 4, pp. 817–829, 2004.

- [113] WANG, J., WANG, M., FANG, Y., “Post-nonlinear blind image separation algorithm with SOM initialization method”. In: *Proceedings of the IET Conference on Wireless, Mobile and Sensor Networks*, v. 1, pp. 978–982, Xangai, China, 2007.
- [114] KAI, S., QI, W., MINGLI, D., “Approach to nonlinear blind source separation based on niche genetic algorithm”. In: *Proceedings of the Sixth International Conference on Intelligent Systems Design and Applications*, v. 1, pp. 441–445, Jinan, China, 2006.
- [115] TAN, Y., WANG, J., “Nonlinear Blind Source Separation using Higher Order Statistics and a Genetic Algorithm”, *IEEE Transactions on Evolutionary Computation*, v. 5, n. 6, pp. 600–612, 2001.
- [116] WOO, W. L., DLAY, S. S., “Nonlinear blind source separation using a hybrid RBF-FMLP network”, *IEE Proceedings on Vision, Image and Signal Processing*, v. 8, n. 2, pp. 173–183, 2005.
- [117] GAO, P., WOO, W., DLAY, S. S., “Neural network approaches to nonlinear blind source separation”. In: *Proceedings of the Eighth International Symposium on Signal Processing and Its Applications*, v. 1, pp. 78–81, Sydney, Australia, 2005.
- [118] GRIFFEL, D. H., *Applied Functional Analysis*. Dover: West Sussex, Reino Unido, 1985.
- [119] PAJUNEN, P., HYVARINEN, A., KARHUNEN, J., “Nonlinear blind source separation by self-organizing maps”. In: *Proceedings of the International Conference on Neural Information Processing*, v. 1, pp. 1207–1210, Hong Kong, Setembro 1996.
- [120] BISHOP, C. M., SVENSEN, M., WILLIANS, C. K. I., “GTM: The generative topographic mapping”, *Neural Computation*, v. 10, pp. 215–234, 1998.
- [121] PAJUNEN, P., KARHUNEN, J., “A maximum likelihood approach to nonlinear blind source separation”. In: *Proceedings of the International Conference on Neural Information Processing*, v. 1, pp. 1207–1210, Hong Kong, Setembro 1996.

- rence on Artificial Neural Networks*, v. 1, pp. 541–546, Lausanne, Suiça, 1997.
- [122] LAPPALAINEN, H., HONKELA, A., “Advances in Independent Component Analysis”, chap. Bayesian nonlinear independent component analysis by multi-layer perceptrons, pp. 93–121, Springer, 2000.
- [123] LAPPALAINEN, H., “Ensemble learning for independent component analysis”. In: *Proceedings of the International Workshop on Independent Component Analysis and Blind Signal Separation*, v. 1, pp. 7–12, Aussois, France, 1999.
- [124] LAPPALAINEN, H., MISKIN, J. W., “Advances in Independent Component Analysis”, chap. Ensemble learning, pp. 75–92, Springer, 2000.
- [125] VALPOLA, H., “Nonlinear independent component analysis using ensemble learning: Theory”. In: *Proceedings of the International Workshop on Independent Component Analysis and Blind Signal Separation*, v. 1, pp. 351–356, Helsinki, Finlândia, 2000.
- [126] HONKELA, A., “Approximating nonlinear transformations of probability distributions for nonlinear independent component analysis”. In: *Proceedings of the IEEE International Joint Conference on Neural Networks*, v. 3, pp. 2169–2174, Budapest, Hungria, 2004.
- [127] HONKELA, A., VALPOLA, H., ILIN, A., et al., “Blind separation of nonlinear mixtures by variational bayesian learning”, *Digital Signal Processing*, v. 17, n. 5, pp. 914–934, 2007.
- [128] ILIN, A., ACHARD, S., JUTTEN, C., “Bayesian versus constrained structure approaches for source separation in post-nonlinear mixtures”. In: *Proceedings of the IEEE International Joint Conference on Neural Networks*, v. 3, pp. 2181–2186, Budapest, Hungria, Julho 2004.
- [129] ALMEIDA, L. B., “Linear and Nonlinear ICA based on Mutual Information”, *The Journal of Machine Learning Research*, v. 4, n. 7, pp. 1297–1318, 2004.

- [130] ALMEIDA, L. B., HYVÄRINEN, A., “Separating a real-life nonlinear image mixture”, *Journal of Machine Learning Research*, v. 6, pp. 1199–1229, 2005.
- [131] ZHENG, C.-H., HUANG, D.-S., LI, K., et al., “MISEP Method for Post-nonlinear Blind Source Separation”, *Neural Computation*, v. 19, n. 9, pp. 2557–2578, 2007.
- [132] SUN, Z.-L., “An extension of MISEP for post-nonlinear-linear mixture separation”, *IEEE Transactions on Circuits and Systems*, v. 56, n. 8, pp. 654–658, 2009.
- [133] KARHUNEN, J., MALAROIU, S., “Local independent component analysis using clustering”. In: *Proceedings of the International Workshop on Independent Component Analysis and Blind Signal Separation*, v. 1, pp. 1–6, Aussois, France, 1999.
- [134] DUDA, R. O., HART, P. E., STORK, D. G., *Pattern Classification*. 2nd ed. Wiley: Nova York, Estados Unidos, 2000.
- [135] KOHONEN, T., OJA, E., SIMULA, O., et al., “Engineering Applications of the Self-Organizing Map”, *Proceedings of the IEEE*, v. 84, n. 10, pp. 1358–1384, Outubro 1996.
- [136] KOTHARI, R., PITTS, D., “On finding the number of clusters”, *Pattern Recognition Letters*, v. 20, n. 1, pp. 405–416, 1999.
- [137] TAKASU, A., “Lecture Notes In Computer Science”, v. 1532, chap. On the Number of Clusters in Cluster Analysis, pp. 62–63, Springer, 1998.
- [138] KARHUNEN, J., MALAROIU, S., ILMONIEMI, M., “Local linear independent component analysis based on clustering”, *International Journal of Neural Systems*, v. 10, pp. 439–451, 2000.
- [139] HONDA, K., ICHIHASHI, H., OHUE, M., et al., “Extraction of local independent components using fuzzy clustering”. In: *Proceedings of 6th International Conference on Neural Networks and Soft Computing*, v. 1, pp. 837–842, Zakopane, Polônia, 2000.

- [140] MAENAKA, T., HONDA, K., ICHIHASHI, H., “Local independent component analysis with fuzzy clustering and regression-principal component analysis”. In: *Proceedings of IEEE International Conference on Fuzzy Systems*, v. 1, pp. 857–862, Vancouver, Canada, 2006.
- [141] PALMIERI, F., BUDILLON, A., “Advances in Independent Component Analysis”, chap. Multi-class independent component analysis for rank deficient distributions, pp. 145–160, Springer-Verlag, 2000.
- [142] LAN, T., ERDOGMUS, D., “Local Linear ICA for Mutual Information Estimation in Feature Selection”. In: *Proceedings of the IEEE Workshop on Machine Learning for Signal Processing*, v. 1, n. 1, pp. 3–8, Connecticut, Estados Unidos, Setembro 2005.
- [143] LAN, T., HUANG, Y., ERDOGMUS, D., “Independent Component Analysis and Blind Signal Separation”, v. 3889, chap. A Comparison of Linear ICA and Local Linear ICA for Mutual Information Based Feature Ranking, pp. 823–830, *Lecture Notes in Computer Science*, Springer, 2006.
- [144] SANCHEZ-POBLADOR, V., MONTE-MORENO, E., SOLÉ-CASALS, J., “Independent Component Analysis and Blind Signal Separation”, v. 3195, chap. ICA as a Preprocessing Technique for Classification, pp. 1165–1172, *Lecture Notes in Computer Science*, Springer, 2004.
- [145]
- [146] CARDOSO, J.-F., SOULOUMIAC, A., “Blind Beamforming for Non-gaussian Signals”, *IEE Proceedings-F*, v. 140, n. 6, pp. 362–370, Novembro 1993.
- [147] CABALLERO, R. G., ORELLANA, C. J. G., MACÍAS, M. M., et al., “Independent Component Analysis Applied to Breast Cancer Detection on Digitized Mammograms”, *International Congress Series*, v. 1281, n. 1, pp. 1052–1057, 2005.
- [148] HYVARINEN, A., “Fast and Robust Fixed-Point Algorithms for Independent Component Analysis”, *IEEE Transactions on Neural Networks*, v. 10, n. 3, pp. 626–634, 1999.

- [149] ZHENG, C. H., HUANG, D. S., SHANG, L., “Feature Selection in Independent Component Subspace for Microarray Data Classification”, *Neurocomputing*, v. 69, n. 1, pp. 2407–2410, 2006.
- [150] APOLLONI, B., BASSIS, S., BREGA, A., “Feature Selection via Boolean Independent Component Analysis”, *Information Sciences*, v. 179, pp. 3815–3831, 2009.
- [151] GOLDBAUM, M. H., “Unsupervised Learning with Independent Component Analysis can Identify Patterns of Glaucomatous Visual Field Defects”, *Transactions of the American Ophthalmology Society*, v. 103, n. 1, pp. 270–280, 2005.
- [152] BONNET, N., NUZILLARD, D., “Independent Component Analysis, a New Possibility for Analysing Series of Electron Energy Loss Spectra”, *Ultra-microscopy*, v. 102, n. 1, pp. 327–337, 2005.
- [153] EGERTON, R. F., *Electron Energy Loss Spectroscopy in the Electron Microscope*. Plenum: Nova York, Estados Unidos, 1996.
- [154] BELOUCHRANI, A., MERAIN, K. A., CARDOSO, J. F., et al., “A Blind Source Separation Technique based on Second Order Statistics”, *IEEE Transactions on Signal Processing*, v. 45, n. 2, pp. 434–444, 1997.
- [155] OVEISI, F., “EEG Signal Classification Using Nonlinear Independent Component Analysis”. In: *Proceedings of the International Conference on Acoustics, Speech and Signal Processing*, v. 1, pp. 361–364, Taipei, Taiwan, 2009.
- [156] GOLDBERG, D. E., *Genetic Algorithms in Search, Optimization and Machine Learning*. Addison-Wesley: Boston, Estados Unidos, 1989.
- [157] LU, C. J., FAN, C. R., CHIU, C. C., “Forecasting Stock Price Using Nonlinear Independent Component Analysis and Support Vector Machine”. In: *Proceedings of the International Conference on Industrial Engineering and Engineering Management*, v. 1, pp. 2370–2374, Hong Kong, 2009.

- [158] LANG, M. J., “Application of Kohonen network classifier in TeV gamma-ray astronomy”, *Journal of Physics G: Nuclear Particle Physics*, v. 1, n. 24, pp. 2279–2287, 1998.
- [159] BECKS, K. H., DRESS, J., FLAGMEYER, U., et al., “Separation of hadronic W-decays from QCD-background with self-organizing maps”, *Nuclear Instruments and Methods in Physics Research A*, v. 426, n. 2, pp. 599–604, 1999.
- [160] LANGE, J. S., “Transputer self-organizing map algorithm for beam background rejection at the BELLE silicon vertex detector”, *Nuclear Instruments and Methods in Physics Research A*, v. 420, n. 1, pp. 288–309, 1999.
- [161] WYZYKOWSKI, L., BELOKUROV, V., “Self-organizing maps, an application to the OGLE data and the Gaia Science Alerts”. In: *Proceedings of the International Conference on Classification and discovery in large astronomical surveys, AIP Conference Proceedings*, v. 1083, n. 1, pp. 201–206, Rinberg Castle, Alemania, 2008.
- [162] HEIKKINEN, A., “Separation of Higgs boson signal from Drell-Yan background with self-organizing maps”, *Proceedings of Science*, v. ACAT07, n. 065, pp. 1–8, 2007.
- [163] PRORIOL, J., “Selection of variables for neural network analysis. Comparison of several methods with high-energy physics data”, *Nuclear Instruments and Methods in Physics Research A*, v. 361, n. 1, pp. 581–585, 1995.
- [164] WOLTER, M., “Multivariate Analysis Methods in Physics”, *Physics of Particles and Nuclei*, v. 38, n. 2, pp. 255–268, 2007.
- [165] AKRAS, S., BOUMIS, P., “A principal component analysis approach to the morphology of planetary nebulae”, *Astrophysics Journal*, v. 1, n. 1, pp. 1–8, 2007.

- [166] CADAVID, A. C., LAWRENCE, J. K., RUZMAIKIN, A., “Principal Components and Independent Component Analysis of Solar and Space Data”, *Solar Physics Journal*, v. 248, n. 2, pp. 247–261, 2008.
- [167] HUANG, X., LEE, S. Y., PREBYS, E., et al., “Application of independent component analysis to Fermilab BOOSTER”, *Physical Review Special Topics, Accelerators and Beams*, v. 8, n. 6, pp. 1–14, 2005.
- [168] FERNANDEZ, C., NAYERI, S., “Independent Component Analysis Applications in Physics”. In: *Proceedings of the Int. Joint Conference on Neural Networks*, v. 1, n. 1, pp. 2213–2216, Montreal, Canada, 2005.
- [169] COSTAGLIA, M., KURUOGLU, E. E., AHMED, A., “Astrophysical Source Separation using Particle Filters”, *Lecture Notes in Computer Science - Independent Component Analysis and Blind Signal Separation*, v. 3195, pp. 930–937, 2004.
- [170] IGUAL, J., LINHARES, R., SALAZAR, A., “Source Separation of Astrophysical Ice Mixtures”, *Lecture Notes in Computer Science - Independent Component Analysis and Blind Signal Separation*, v. 3889, pp. 368–375, 2006.
- [171] MOUDDEN, Y., ABRIAL, P., MELIN, J. B., et al., “Independent Component Separation from Incomplete Spherical Data using Wavelets. Application to CMB Data Analysis”. In: *Proceedings of the International Conference on Physics in Signal and Image Processing*, v. 1, pp. 1–6, Toulouse, France, 2005.
- [172] VIOO, R., ANDREANI, P., “A Modified ICA Approach for Signal Separation in CMB maps”, *Astronomy and Astrophysics Journal*, v. 802, pp. 1–12, Fevereiro 2008.
- [173] NOJUN KWAK, M., CHOI, C.-H., “Feature Extraction Based on ICA for Binary Classification Problems”, *IEEE Transactions on Knowledge and Data Engineering*, v. 15, n. 6, pp. 1374–1388, 2003.

- [174] CANER, E. S., SEIXAS, J., “Neural Discriminating Analysis on Pre-processed Data”. In: *Proceedings of the 6th IEEE International Conference on Electronics, Circuits and Systems (ICECS)*, pp. 415–418, Pafos, Cyprus, Setembro 1999.
- [175] LU, H., SETIONO, R., LIU, H., “Effective Data Mining Using Neural Networks”, *IEEE Transactions on Knowledge and Data Engineering*, v. 8, n. 6, pp. 957–961, 1996.
- [176] STEPPE, J. M., ANS S. K. ROGERS, K. W. B. J., “Integrated Feature Architecture Selection”, *IEEE Transactions on Neural Networks*, v. 7, n. 4, pp. 1007–1014, 1996.
- [177] MCGARRY, K. J., WERMTER, S., MACINTYRE, J., “Knowledge Extraction from Radial Basis Functions Networks and Multi-Layer Perceptrons”. In: *Proceedings of the International Joint Conference on Neural Networks*, v. 4, pp. 2494–2497, Washington-DC, Estados Unidos, 1999.
- [178] KWAK, N., CHOI, C. H., CHOI, J. Y., “Lecture Notes In Computer Science (ICANN)”, v. 2130, chap. Feature Extraction Using ICA, pp. 568–573, Springer, 2001.
- [179] AKAHO, S., “Conditionally Independent Component Analysis for Supervised Feature Extraction”, *Neurocomputing*, v. 49, n. 1, pp. 139–150, 2002.
- [180] KOTANI, M., OZAWA, S., “Feature Extraction Using Independent Components of Each Category”, *Neural Processing Letters*, v. 22, n. 1, pp. 113–124, 2005.
- [181] SIMAS FILHO, E. F., SEIXAS, J. M., CALOBA, L. P., “Modified Post-Nonlinear ICA Model for Online Neural Discrimination”, *Neurocomputing*, v. 0, n. 0, pp. 0, 2010.
- [182] HAUPT, R. L., HAUPT, S. E., *Practical Genetic Algorithms*. Wiley-Interscience: Nova York, Estados Unidos, 2004.

- [183] GEORGIEVA, P., RALESCU, A., RALESCU, D., “Cross-cumulants measure for independence”, *Journal of Statistical Planning and Inference*, v. 137, n. 1, pp. 1085–1098, 2006.
- [184] TREES, H. L. V., *Detection, Estimation, and Modulation Theory, Part I*. Wiley: Nova York, Estados Unidos, 2001.
- [185] RIEDMILLER, M., BRAUN, H., “A direct adaptive method for faster back-propagation learning, the RPROP algorithm”, *Proc. of Int. Conf. on Neural Networks*, pp. 586–591, 1993.
- [186] KUNCHEVA, L., *Combining Pattern Classifiers, Methods and Algorithms*. Wiley: Nova York, Estados Unidos, 2004.
- [187] SCHLICKEISER, R., *Cosmic Ray Astrophysics*. Springer: Nova York, Estados Unidos, 2002.
- [188] KOHONEN, T., *Self Organizing Maps*. 3rd ed. Springer: Berlin, Alemanha, 2001.
- [189] BASE, A. M., GRUBER, P., THEISA, F., et al., “Blind source separation based on self-organizing neural network”, *Engineering Applications of Artificial Intelligence*, v. 19, n. 3, pp. 305–311, 2006.
- [190] GERSHO, A., “On the Structure of Vector Quantizers”, *IEEE Transactions on Information Theory*, v. 28, n. 2, pp. 157–166, Março 1982.
- [191] GRAY, R. M., “Vector Quantization”, *IEEE ASSP Magazine*, v. Part 1, pp. 4–29, Abril 1984.
- [192] KOHONEN, T., “Improved versions of learning vector quantization”. In: *Proceedings of the International Joint Conference on Neural Networks*, v. 1, pp. 545–550, San Diego, Estados Unidos, 1990.
- [193] BARAS, J. S., DEY, S., “Combined Compression and Classification with Learning Vector Quantization”, *IEEE Transactions on Information Theory*, v. 45, n. 6, pp. 1911–1920, 1999.

- [194] FREEMAN, J. A., SKAPURA, D. M., *Neural Networks Algorithms, Applications, and Programming Techniques*. Addison Wesley, 1991.
- [195] OJA, E., “Neural Networks, Principal Components, and Subspaces”, *International Journal of Neural Systems*, v. 1, n. 1, pp. 61–68, 1989.
- [196] DE OLIVEIRA E SOUZA FILHO, J. B., *Classificação Neural de Sinais Passivos*, Ph.D. Thesis, COPPE/UFRJ, Julho 2007.
- [197] PEEBLES JR., P. Z., *Probability, Random Variables and Random Signal Principles*. Mc Graw Hill: Nova York, Estados Unidos, 2001.
- [198] SPIEGEL, M. R., SCHILLER, J. J., SRINIVASAN, R. A., *Probability and Statistics*. 2nd ed. McGraw-Hill: Nova York, Estados Unidos, 2000.
- [199] SHANNON, C. E., “A Mathematical Theory of Communication”, *The Bell System Technical Journal*, v. 27, n. 6, pp. 379–423, Julho 1948.
- [200] HYVARINEN, A., “New approximations of differential entropy for independent component analysis and projection pursuit”, *Advances in Neural Information Signal Processing*, v. 10, pp. 273–279, 1998.
- [201] CICHOCKI, A., UNBEHAUEN, R., “Robust Neural Networks with On-Line Learning for Blind Identification and Blind Separation of Sources”, *IEEE Transactions on Circuits and Systems-I: Fundamental Theory and Applications*, v. 43, n. 11, Novembro 1996.
- [202] LUENBERGER, D. G., *Linear and Nonlinear Programming*. Addison-Wesley: Boston, Estados Unidos, 1984.
- [203] MICHAL, A. D., *Matrix and Tensor Calculus*. 1st ed. Dover: West Sussex, Reino Unido, 2008.
- [204] CARDOSO, J. F., “Source separation using higher order moments”. In: *Proceedings of the IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, v. 4, pp. 2109–2112, Glasgow, Reino Unido, 1989.
- [205] AKUZAWA, T., MURATA, N., “Multiplicative Nonholonomic Newton-like Algorithm”, *Chaos, Solitons and Fractals*, v. 12, n. 4, pp. 785–793, 2001.

- [206] AKUZAWA, T., “Extended Quasi-Newton Method for the ICA”. In: *Proceedings of the Int. Workshop Independent Component Anal. Blind Signal Separation*, pp. 521–525, Helsinki, Finlândia, 2000.
- [207] CUENCA, W. M., LEVY, A. F., SEIXAS, J. M., et al., “Análise de Componentes Independentes para Extração Cega de Ruídos Gaussianos dos Sinais de Descargas Parciais em Equipamentos de Alta Tensão”. In: *Anais do Congresso Brasileiro de Redes Neurais*, v. 1, pp. 133–138, São Paulo, 2003.
- [208] CICHOCKI, A., DOUGLAS, S., AMARI, S., “Robust techniques for independent component analysis with noisy data”, *Neurocomputing*, v. 22, n. 1-3, pp. 113–129, 1998.
- [209] HYVARINEN, A., “Independent Component Analysis in the Presence of Gaussian Noise by Maximizing Joint Likelihood”, *Neurocomputing*, v. 22, n. 1-3, pp. 49–67, 1998.
- [210] HYVARINEN, A., “Sparse Code Shrinkage: Denoising of Nongaussian Data by Maximum Likelihood Estimation”, *Neural Computation*, v. 11, n. 7, pp. 1739–1768, 1999.
- [211] PARASCHIV-IONESCU, A., JUTTEN, C., AMINIAN, K., et al., “Source separation in strong noisy mixtures a study of wavelet de-noising pre-processing”. In: *IEEE International Conference on Acoustics, Speech, and Signal Processing*, v. 2, pp. 1681–1684, Orlando, Estados Unidos, 2002.
- [212] TAN, Y., WANG, J., “Nonlinear blind separation using an RBF network model”. In: *Proceedings of the IEEE International Symposium on Circuits and Systems*, v. 3, pp. 634–637, Genebra, Suiça, 2000.
- [213] SHANMUGAN, K. S., BREIPOHL, A. M., *Random Signals, Detection, Estimation and Data Analysis*. Wiley: Nova York, Estados Unidos, 1988.
- [214] FISHER, R. A., “The Use of Multiple Measurements in Taxonomic Problems”, *Annals of Eugenics*, v. 7, pp. 179–188, 1936.

- [215] SIMAS FILHO, E. F., ALMEIDA, L. A. L., LIMA, A. C. C., “Vibration Monitoring of On-Load Tap Changers Using a Genetic Algorithm”. In: *Proceedings of the IEEE Instrumentation and Measurement Technology Conference*, v. 3, pp. 2288–2293, Ottawa, Canada, 2005.
- [216] SIMAS FILHO, E. F., ALMEIDA, L. A. L., “Filtragem inversa de medições de histerese térmica utilizando Redes Neurais”. In: *Anais do XVI Congresso Brasileiro de Automática*, v. 1, pp. 1–6, Salvador, Brasil, Outubro 2006.
- [217] ZOPPOLI, R., PARISINI, T., SANGUINETI, M., “Neural Aproximators for Function Optimization”, *Proceedings of the 35th conference on decision and control*, v. 3, pp. 3290–3293, 1996.
- [218] ICHIKAWA, Y., SAWA, T., “Neural Networks Application for Direct Feedback Controllers”, *IEEE Transactions on Neural Networks*, v. 3, n. 2, pp. 224–231, 1992.
- [219] WASSERMAN, P., *Neural Computing, Theory and Practice*. Van Nostrand Reinhold: Nova York, Estados Unidos, 1989.
- [220] WHITLEY, D., “An Overview of Evolutionary Algorithms: Practical Issues and Common Pitfalls”, *Colorado Advanced Software Institute Press*, 2002.
- [221] HOLLAND, J., “Adaptation in Natural and Artificial Systems”, *University of Michigan Press*, 1975.
- [222] MITCHEL, M., *An Introduction to Genetic Algorithms*. The MIT Press, 1996.
- [223] ACKLEY, D., “A Connectionist Machine for Genetic Hillclibimg”, *Kluwer Academic Publishers*, 1987.
- [224] SYSWERDA, G., “Uniform Crossover in Genetic Algorithms”, *Proceedings of the 3rd International Conference on Genetic Algorithms*, pp. 2–9, 1989.
- [225] TANOMARU, J., “Motivação, Fundamentos e Aplicações de Algoritmos Genéticos”, *II Congresso Brasileiro de Redes Neurais, III Escola de Redes Neurais*, 1995.

Apêndice A

Aspectos Teóricos das Técnicas de Extração de Características

Neste apêndice serão fornecidos os detalhes da teoria envolvida nos diversos métodos de extração de características utilizados neste trabalho.

A.1 Mapas auto-organizáveis

O mapa auto-organizável (SOM - *Self Organizing Map*) é uma rede neural com treinamento não-supervisionado, baseado na aprendizagem competitiva, que é capaz de realizar uma organização topológica das entradas. O SOM foi proposto por Teuvo Kohonen em 1982 [188], sendo capaz de realizar um mapeamento não-linear dos sinais de um espaço de entrada contínuo de dimensão k para um espaço de características discreto que, em geral, é bidimensional. Cada neurônio da grade está diretamente conectado a todos os nós de entrada. Na Figura A.1 pode-se visualizar o diagrama de um mapa auto-organizável bi-dimensional.

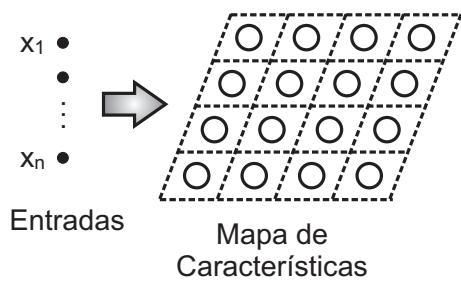


Figura A.1: Diagrama de um mapa auto-organizável

O mapa auto-organizável compacta a informação e preserva relações topológicas ou métricas do conjunto de sinais. Os SOM estão ligados à ICA por conseguirem extrair informações ocultas dos sinais de forma não supervisionada [189]. Uma aproximação das componentes independentes não-lineares pode ser obtida utilizando mapas auto-organizáveis [4].

Três processos estão envolvidos na formação do mapa auto-organizado: a **competição**, onde, para cada vetor de entrada, há apenas um neurônio vencedor; a **cooperação**, quando o neurônio vencedor determina uma vizinhança topológica de neurônios excitados; e a **adaptação**, que procede ao ajuste dos pesos sinápticos para reforçar a resposta do neurônio vencedor, e de seus vizinhos, ao padrão de entrada.

A atualização do vetor de pesos w_j do neurônio j é feita através da equação:

$$w_j(n+1) = w_j(n) + \eta(n)h_{ij}(n)(x(n) - w_j(n)), \quad (\text{A.1})$$

sendo $\eta(n)$ a taxa de aprendizagem, um tipo de função de vizinhança $h_{ij}(n)$ usualmente utilizada é definida por:

$$h_{ij}(n) = \exp(-d_{ij}^2/2\sigma^2(n)) \quad (\text{A.2})$$

onde d_{ij} é a distância do neurônio j para o neurônio vencedor i e $\sigma(n)$ é a largura da função vizinhança na n-ésima iteração.

O mapa de características possui algumas propriedades, listadas a seguir [5]:

1. é formado pelo conjunto de vetores de pesos sinápticos w_i no espaço de saída discreto e fornece uma boa aproximação para o espaço de entrada;
2. é ordenado de modo topológico. Padrões de entrada semelhantes são mapeados para regiões adjacentes no mapa de características;
3. regiões do espaço de entrada que possuem alta probabilidade de ocorrência são mapeadas para domínios maiores do espaço de saída;

No mapa de características, o neurônio que apresentar maior saída é considerado o vencedor, ou seja a saída do SOM é do tipo “vencedor leva tudo” (WTA - *winner*

takes all). O neurônio ativado é escolhido a partir de sua semelhança com a entrada apresentada. É comum a utilização da distância euclidiana como métrica da proximidade entre dois vetores; nesse caso, o neurônio vencedor é aquele que minimiza $i(\mathbf{x}) = \|\mathbf{x}(n) - \mathbf{w}_j\|$.

Uma outra forma de operar um mapa auto-organizável é utilizar as projeções dos sinais de entrada no mapa de características, ou seja as saídas u_j de cada neurônio j que pode ser calculada por:

$$u_j = \mathbf{x}^T \mathbf{w}_j \quad (\text{A.3})$$

O vetor $\mathbf{u} = [u_1, \dots, u_K]^T$ pode ser considerado como a projeção de \mathbf{x} no mapa de características.

Os mapas auto-organizáveis pertencem à classe de algoritmos de codificação vetorial, sendo capazes de encontrar, de forma otimizada, um número fixo de vetores ou palavras de código que melhor representem o conjunto de sinais.

A.1.1 Quantização vetorial por aprendizado

A quantização vetorial (VQ - *Vector Quantization*) é uma técnica de codificação onde um espaço de entrada é mapeado em um grupo finito de vetores representativos (*codebook*) [190]. A codificação é definida como um particionamento do espaço de entrada em um número finito de regiões. O quantizador realiza um mapeamento do espaço \mathbb{R}^k , em um subconjunto finito Y de \mathbb{R}^k :

$$Q : \mathbb{R}^k \rightarrow \mathbf{Y} \quad (\text{A.4})$$

sendo $\mathbf{Y} = \{y_1, y_2, \dots, y_k\}$ o livro de código (*codebook*). Para cada palavra de código y_i existe uma partição R_i do espaço de entrada que satisfaz:

$$R_i = Q^{-1}(\mathbf{y}_i) = \{\mathbf{x} \in \mathbb{R}^k : Q(\mathbf{x}) = \mathbf{y}_i\} \quad (\text{A.5})$$

$$\bigcup_{i=1}^N R_i = \mathbb{R}^k, \quad R_i \cap R_j = \emptyset, \quad i \neq j \quad (\text{A.6})$$

Quando um quantizador vetorial possui mínima distorção é denominado **quantizador de Voronoi**. Neste caso, diz-se que o espaço de entrada está particionado de

acordo com a regra do vizinho mais próximo, e as partições criadas são chamadas de células de Voronoi [191]. Usando-se a distância euclidiana como parâmetro de distorção, o quantizador Q^* é dito ótimo se, para qualquer outro quantizador Q , com o mesmo número de pontos, a condição abaixo é satisfeita:

$$E\|\mathbf{x} - Q^*(\mathbf{x})\|^2 \leq E\|\mathbf{x} - Q(\mathbf{x})\|^2 \quad (\text{A.7})$$

As palavras de código ou os vetores de Voronoi podem ser calculados de modo aproximado pelo algoritmo SOM. O *codebook* é formado a partir dos pesos sinápticos dos neurônios do mapa. As células de Voronoi são compostas pelos pontos do espaço de entrada que estão mais próximos do vetor de código correspondente.

Em um problema de classificação, pode-se empregar a quantização vetorial por aprendizado (*Learning Vector Quantization*) [192], que utiliza informações sobre as classes para mover ligeiramente os vetores de Voronoi, visando a uma melhora no desempenho de decisão do classificador.

Na sua forma básica, o algoritmo LVQ escolhe aleatoriamente um vetor de entrada \mathbf{x} ; quando seu rótulo de classe $\mathcal{C}_{\mathbf{x}_i}$ e o de um vetor de Voronoi \mathbf{w}_c concordam, então, \mathbf{w}_c é movido na direção de \mathbf{x} :

$$\mathcal{C}_{\mathbf{w}_c} = \mathcal{C}_{\mathbf{x}_i} \rightarrow \mathbf{w}_c(n+1) = \mathbf{w}_c(n) + \alpha[\mathbf{x} - \mathbf{w}_c(n)] \quad (\text{A.8})$$

onde α é a taxa de aprendizagem ($0 < \alpha < 1$). Em caso contrário, \mathbf{w} é afastado de \mathbf{x} :

$$\mathcal{C}_{\mathbf{w}_c} \neq \mathcal{C}_{\mathbf{x}_i} \rightarrow \mathbf{w}_c(n+1) = \mathbf{w}_c(n) - \alpha[\mathbf{x} - \mathbf{w}_c(n)] \quad (\text{A.9})$$

Conforme proposto em [192], podem ser implementadas algumas modificações na forma básica do algoritmo de LVQ, visando a melhorar o desempenho do método. Chega-se, então, aos algoritmos LVQ-2 e LVQ-2.1, que ajustam dois vetores de código simultaneamente.

Alguns exemplos da aplicação da quantização vetorial por aprendizado para compressão de sinais e classificação podem ser encontrados em [192] e [193].

A.1.2 Classificação a partir do mapa de características

Considerando um problema de classificação, o mapeamento auto-organizável consegue transformar o conjunto de sinais, revelando características ocultas. A nova organização do conjunto de entrada pode ser utilizada para guiar o processo de discriminação. Em [194] é proposta uma estratégia de classificação a partir do mapa de características onde uma rede neural MLP é conectada às saídas do SOM (ver Figura A.2). A MLP é treinada com supervisão usando informações a respeito das classes de sinais.

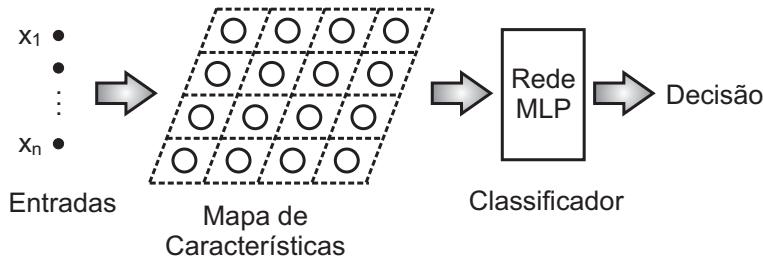


Figura A.2: Diagrama da classificação a partir do mapa de características

A.2 Análise de Componentes Principais

A.3 Técnicas de Pré-Processamento - Compactingação

No processamento de sinais multi-dimensionais, é comum a utilização de técnicas de processamento de sinais que visam a redução da dimensionalidade do problema. O objetivo é projetar os sinais N-dimensionais observados em

A.3.1 Análise de Componentes Principais

A análise de componentes principais (PCA - *Principal Component Analysis*) é uma técnica estatística de processamento de sinais diretamente ligada à transformação de *Karhunen-Loëve* [3]. O objetivo da PCA é encontrar uma transformação linear onde os sinais projetados sejam não-correlacionados e grande parcela da energia

(variância) esteja concentrada num pequeno número de componentes. Para isso, são exploradas informações da estatística de segunda ordem.

A análise de componentes principais é bastante usada para compactação de informação. Como a PCA projeta os sinais em componentes ordenados por energia, uma métrica geralmente utilizada para reduzir a dimensão dos dados consiste na seleção apenas dos componentes de maior energia, de modo que o sinal recuperado a partir da informação compactada tenha pequeno erro médio quadrático se comparado ao original. A seguir serão desenvolvidos, de forma resumida, os fundamentos matemáticos da PCA.

Considerando-se um vetor $\mathbf{x} = [x_1, \dots, x_N]^T$ aleatório com N elementos, assume-se que ele tenha média zero:

$$\mathcal{E}\{\mathbf{x}\} = 0 \quad (\text{A.10})$$

onde $\mathcal{E}\{\cdot\}$ é o operador esperança. Se \mathbf{x} tem média não nula faz-se $\mathbf{x} \leftarrow \mathbf{x} - \mathcal{E}\{\mathbf{x}\}$.

A projeção z_i de \mathbf{x} na direção de \mathbf{v}_i pode ser expressa por:

$$z_i = \mathbf{v}_i^T \mathbf{x} = \sum_{k=1}^N v_{ki} x_k \quad (\text{A.11})$$

Na transformação por PCA, os componentes z_i ($i = 1, \dots, N$) devem ser ortogonais e ordenados (de modo decrescente) pela variância das projeções, sendo, então, z_1 a projeção de máxima variância. Para tornar a variância independente da norma de \mathbf{v}_i , faz-se:

$$\mathbf{v}_i \leftarrow \frac{\mathbf{v}_i}{\|\mathbf{v}_i\|} \quad (\text{A.12})$$

Fazendo-se com que $\|\mathbf{v}_i\| = 1$, torna-se a variância função apenas da direção das projeções.

Como $\mathcal{E}\{\mathbf{x}\} = 0$, então $\mathcal{E}\{z_i\} = 0$, logo a variância da projeção z_i é calculada por $\mathcal{E}\{z_i^2\}$. Seguindo a definição da PCA, z_1 tem máxima variância; logo, \mathbf{v}_1 pode

ser encontrado pela maximização de [4]:

$$J_1^{PCA}(\mathbf{v}_1) = \mathcal{E}\{z_i^2\} = \mathcal{E}\{(\mathbf{v}_1^T \mathbf{x})^2\} = \mathbf{v}_1^T \mathcal{E}\{\mathbf{x}\mathbf{x}^T\} \mathbf{v}_1 = \mathbf{v}_1^T \mathbf{C}_x \mathbf{v}_1, \quad (\text{A.13})$$

onde \mathbf{C}_x é a matriz de covariância de \mathbf{x} .

A solução para o problema de maximização da equação (A.13) pode ser encontrada na álgebra linear, em função dos autovetores $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_N$ da matriz \mathbf{C}_x . A ordem dos autovetores é tal que os autovalores associados satisfazem $d_1 > d_2 > \dots > d_N$. Desta forma, tem-se:

$$\mathbf{v}_i = \mathbf{e}_i, \quad 1 \leq i \leq N \quad (\text{A.14})$$

Percebe-se que a PCA de \mathbf{x} e a decomposição por autovalores da matriz \mathbf{C}_x (de dimensão $N \times N$) são equivalentes. Limitações computacionais na extração das componentes principais utilizando as equações (A.11) e (A.14) aparecem quando a dimensão N do vetor \mathbf{x} aumenta, pois o processo de obtenção dos autovetores se torna proibitivamente lento. Nesse caso, uma solução é utilizar métodos iterativos de extração das componentes principais, através de redes neurais [195, 196].

A.3.2 Redução de dimensão

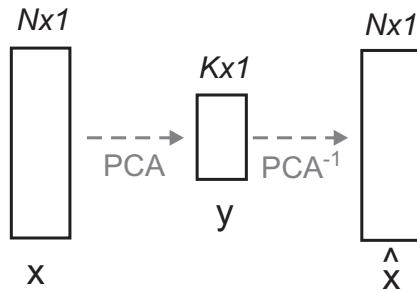


Figura A.3: Compressão e recuperação do sinal \mathbf{x} utilizando a transformação por PCA.

A principal aplicação da PCA é a compactação da informação. A redução de dimensão é obtida utilizando-se para a reconstrução do sinal original \mathbf{x} um número K de componentes principais sendo $K < N$. Na Figura A.3 é ilustrado o processo de redução de dimensão utilizando análise de componentes principais. Em geral, o

número de componentes é escolhido visando a preservar uma parcela V_e da energia total, de modo que $\hat{\mathbf{x}} \approx \mathbf{x}$. A variância explicada V_e de um conjunto de componentes pode ser calculada usando-se:

$$V_e(K) = \frac{\sum_{i=1}^K d_i}{\sum_{i=1}^N d_i}, \quad (\text{A.15})$$

sendo d_i o autovalor da matriz \mathbf{C}_x de covariância do processo correspondente à componente i .

A transformação por PCA é ótima no sentido de representação do sinal nas primeiras componentes, mas não há garantia de que a compactação facilite o processo de classificação. Quando as direções de maior variância coincidem com as de melhor discriminação das classes, então a PCA é também útil para o reconhecimento de padrões, em caso contrário, a redução de dimensão pode dificultar a separação. Entretanto, em problemas de classificação onde a dimensão da entrada é excessivamente grande o pré-processamento por PCA reduz o custo computacional e consequentemente o tempo de processamento.

A.4 Análise de Componentes Independentes

A.4.1 Princípios de estimativa dos componentes independentes

No modelo básico da ICA (ver equação (5.4)) assume-se que a matriz \mathbf{A} é quadrada e não são considerados os atrasos temporais nem a existência de ruído aditivo. O princípio básico para a extração das componentes independentes é obtido do teorema do limite central. Como a soma de duas variáveis aleatórias independentes é sempre mais próxima de uma distribuição normal do que as variáveis originais, os sinais misturados x_i , que são gerados a partir do somatório ponderado das fontes s_i , têm distribuições de probabilidade mais semelhantes à Gaussiana quando comparadas aos sinais originais. As fontes podem ser obtidas então pela maximização da não-Gaussianidade.

Maximização da não-Gaussianidade

A **curtose** é o cumulante de quarta ordem, e para uma variável y de média zero e variância unitária é definida por [197]:

$$kurt(y) = \mathcal{E}\{y^4\} - 3(\mathcal{E}\{y^2\})^2. \quad (\text{A.16})$$

Variando no intervalo $[-2, \infty)$, a curtose é igual a zero para uma variável Gaussiana, os valores negativos indicam sub-Gaussianidade e os positivos super-Gaussianidade.

A curtose é um parâmetro estatístico facilmente calculado a partir das realizações da variável aleatória, porém seu valor pode ser bastante influenciado por um pequeno conjunto de pontos na cauda da distribuição [198], sendo, nesse caso, pouco robusta para a estimativa da não-Gaussianidade. Conhecidos como intrusos (ou *outliers*) esses pontos podem realmente pertencer à variável aleatória, ou ter sido artificialmente introduzidos por algum fenômeno desconhecido, como erro de medida ou de digitação.

Uma estimativa alternativa da não-Gaussianidade pode ser obtida a partir da **negentropia**, que é calculada por [87]:

$$J(y) = H(y_{gauss}) - H(y), \quad (\text{A.17})$$

onde $H(\cdot)$ é a entropia, e y_{gauss} é uma variável aleatória Gaussiana com a mesma média e variância de y . A entropia é um dos conceitos básicos da teoria da informação e pode ser interpretada como o grau de informação contido em uma variável. Para uma variável aleatória discreta a entropia é definida como [199]:

$$H(Y) = - \sum_i P(Y = a_i) \log P(Y = a_i), \quad (\text{A.18})$$

onde os a_i são os possíveis valores da variável Y , e $P(Y = a_i)$ é a probabilidade de Y ser igual a a_i .

Um resultado importante obtido a partir da teoria da informação é que uma variável Gaussiana tem a máxima entropia entre todas as variáveis de mesma

variância. Considerando a equação (A.17), a negentropia é sempre não negativa e zero quando a variável é Gaussiana, servindo como uma medição da não-Gaussianidade. O grande problema no cálculo de $J(.)$ é a necessidade de se estimar as probabilidades da equação (A.18). Para evitar esse cálculo, utilizam-se aproximações da negentropia. Conforme descrito em [4], existem duas aproximações mais utilizadas para a negentropia, uma faz uso de cumulantes de ordem superior:

$$J(Y) \approx \frac{1}{12}E\{Y^3\}^2 + \frac{1}{48}kurt(Y)^2, \quad (\text{A.19})$$

e outra utiliza funções não-polinomiais [200]:

$$J(Y) \approx [k_1(E\{G_1(Y)\})^2 + k_2(E\{G_2(Y)\} - E\{G_2(\nu)\})^2], \quad (\text{A.20})$$

onde ν é uma variável aleatória Gaussiana de média zero e variância unitária. As funções não-lineares recomendadas em [200] são $G_1(y) = y \exp(-y^2/2)$ e $G_2(y) = |y|$ ou $G_2(y) = \exp(-y^2/2)$.

O uso de cumulantes traz de volta o problema da pouca robustez a *outliers*. É mostrado em [200] que o uso das funções não-polinomiais leva ao método da máxima entropia [4].

Minimização da informação mútua

Um outro método de estimação de ICA, também derivado da teoria da informação, é obtido pela minimização da informação mútua. A informação mútua $I(.)$ entre m variáveis aleatórias escalares y_i é definida como [98]:

$$I(y_1, y_2, \dots, y_m) = \sum_{i=1}^m H(y_i) - H(\mathbf{y}) \quad (\text{A.21})$$

A entropia $H(y_i)$ pode ser interpretada como o comprimento de código (ou a quantidade de informação) necessário para representar a variável y_i . Conforme a equação (A.21), a informação mútua é a diferença entre o somatório das entropias de cada uma das m variáveis y_i e a entropia do vetor aleatório $\mathbf{y} = [y_1, y_2, \dots, y_m]$. Pode-se provar que a codificação mais eficiente é obtida quando se utiliza o conjunto de variáveis \mathbf{y} . Utilizar as variáveis isoladamente sempre gera um maior código, menos quando as y_i são independentes, pois desta forma uma variável não carrega informação sobre as demais, sendo a informação mútua igual a zero. Desta

forma, $I(y_1, y_2, \dots, y_m)$ pode ser utilizada como uma medida da dependência entre as variáveis. A matriz \mathbf{W} de transformação inversa da ICA, conforme equação ??, pode ser estimada através da minimização da informação mútua dos sinais s_i recuperados.

ICA através da descorrelação não linear

A igualdade da equação:

$$\mathcal{E}\{g(x)h(y)\} = \mathcal{E}\{g(x)\}\mathcal{E}\{h(y)\} \quad (\text{A.22})$$

repetida aqui para comodidade do leitor, garante que as variáveis x e y são independentes quando todas funções $g(\cdot)$ e $h(\cdot)$, integráveis em x e y são descorrelacionadas. Portanto, a extração das ICs pode ser obtida testando-se a correlação entre todas as funções não-lineares $g(\cdot)$ e $h(\cdot)$.

Existem alguns algoritmos propostos na literatura para o problema da descorrelação não-linear, como o *Hérault-Jutten* [4] e o *Chichocki-Unbehauen* [201], mas como não é possível testar a descorrelação entre todas as funções não-lineares, escolhem-se $f(\cdot)$ e $g(\cdot)$ visando-se a obter boas aproximações das componentes independentes. O algoritmo *Hérault-Jutten*, por exemplo, aconselha o uso de $f(y) = y^3$ e $g(y) = \text{arctg}(y)$, já o *Chichocki-Unbehauen* sugere uma função polinomial e a tangente hiperbólica.

A PCA não-linear (NLPCA - *Non-linear Principal Component Analysis*) pode ser vista como uma extensão não linear da PCA, e é capaz de encontrar projeções descorrelacionadas não-linearmente. Enquanto o objetivo da PCA é minimizar o erro médio quadrático de reconstrução do sinal projetando as componentes numa base ortonormal, a NLPCA pode ser definida de modo simples através da função-objetivo a ser minimizada:

$$J(\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_n) = \mathcal{E}\{\|\mathbf{x} - \sum_{i=1}^n g_i(\mathbf{w}_i^T \mathbf{x})\mathbf{w}_i\|^2\}, \quad (\text{A.23})$$

onde $g_1(\cdot), g_2(\cdot), \dots, g_n(\cdot)$ é um conjunto de funções escalares e não-lineares, e os vetores \mathbf{w}_i formam a base do sub-espacó onde serão projetadas as entradas \mathbf{x} . Quando o mínimo de $J(\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_n)$ for encontrado, o produto $\mathbf{w}_i^T \mathbf{x}$ dará as componentes

principais não-lineares. Se $g_i(y) = y$ para todo i , então equação (A.23) se reduz à função objetivo da PCA. Quando os sinais satisfazem ao modelo da ICA, mostrado na equação (5.4), a NLPCA obtém uma aproximação das componentes independentes.

A.4.2 Pré-processamento dos sinais para ICA

Em geral, os algoritmos de extração das componentes independentes têm seu trabalho simplificado quando os sinais são centralizados, ou seja, têm sua média removida fazendo-se:

$$\mathbf{x} \leftarrow \mathbf{x} - \mathcal{E}\{\mathbf{x}\} \quad (\text{A.24})$$

Outra transformação importante é o branqueamento. Um vetor $\mathbf{z} = (z_1, z_2, \dots, z_n)^T$ é dito branco quando os elementos z_i são descorrelacionados e têm variância unitária. O branqueamento pode ser realizado por uma transformação linear:

$$\mathbf{z} = \mathbf{V}\mathbf{x} \quad (\text{A.25})$$

O branqueamento, que é apenas a descorrelação seguida de uma normalização, pode ser realizado por uma transformação através de PCA. Com as variáveis branqueadas a extração da ICA é facilitada, pois os sinais já estão descorrelacionados.

Em problemas com vetores de entrada de alta dimensão, é importante a compactação da informação através de PCA ou Análise de Relevância para facilitar o processo de extração das componentes independentes.

A.4.3 Principais algoritmos para ICA

Diversos algoritmos vêm sendo propostos para a extração das componentes independentes. Essas rotinas diferem basicamente no princípio teórico no qual fundamentam a obtenção das componentes independentes (não-Gaussianidade, informação mútua, descorrelação não-linear, etc) e na forma fazem a otimização da função objetivo escolhida. Os principais parâmetros para avaliação de desempenho são o tempo de processamento (complexidade computacional) e a precisão na extração das componentes.

Um estudo comparativo entre diversos métodos de estimação das componentes independentes foi realizado em [4]. O algoritmo **FastICA**, descrito com detalhes em [4] e [98], é o que apresenta menor custo computacional. Algoritmos que realizam descorrelação não linear e NLPCA têm desempenho semelhante ao FastICA em termos da precisão na obtenção da matriz \mathbf{W} , porém exigem maior esforço de computação. O algoritmo **JADE** (*Joint Approximate Diagonalization of Eigenmatrices*) proposto em [146] também é muito utilizado em ICA, mostrando bons resultados.

Algoritmo FastICA

Considerando as aproximações da negentropia mostradas nas Equações (A.19) e (A.20), e o fato de que a minimização da negentropia leva à independência estatística, no trabalho [148] foram propostos algoritmos de ponto fixo para ICA (chamados FastICA), que utilizam iterações semelhantes às de Newton [202]. Entre as vantagens deste algoritmo pode-se citar simplicidade computacional, baixa utilização de memória e boas características de convergência [98].

A partir de algumas manipulações da equação (A.20), o algoritmo FastICA para estimativa de uma componente independente é formulado a seguir para sinais pré-branqueados:

1. Escolha um vetor de pesos inicial \mathbf{w} de modo aleatório;
2. Faça $\mathbf{w}^+ = E\{\mathbf{x}g(\mathbf{w}^T \mathbf{x})\} - E\{g'(\mathbf{w}^T \mathbf{x})\}\mathbf{w}$;
3. $\mathbf{w} = \mathbf{w}^+ / \|\mathbf{w}^+\|$;
4. Se o algoritmo não tiver收敛ido voltar para o passo 2.

Os autores sugerem o uso de uma das funções $g(\cdot)$ a seguir:

$$g_1(x) = \tgh(a_1 x), \quad (\text{A.26})$$

$$g_2(x) = x \exp(-a_2 u^2 / 2), \quad (\text{A.27})$$

$$g_3(x) = x^3, \quad (\text{A.28})$$

onde $1 \leq a_1 \leq 2$ e $a_2 \approx 1$. A escolha da função não-linear pode ser guiada pelas características a seguir [148]: a função $g_1(\cdot)$ é indicada quando não há informação a

respeito da estatística das componentes independentes, pois o algoritmo apresenta resultados satisfatórios para qualquer tipo de distribuição; o uso de $g_2(\cdot)$ é indicado quando as componentes independentes são super-Gaussianas e o $g_3(\cdot)$ deve ser utilizada para estimar componentes sub-Gaussianas.

Para estimar mais de uma componente independente pode-se utilizar métodos de ortogonalização deflacionária como o de Gram-Schmidt [4].

Algoritmo JADE

No algoritmo JADE (*Joint Approximate Diagonalization of Eigenmatrices*), as informações estatísticas de segunda e quarta ordem são utilizadas a partir de uma abordagem tensorial. Tensores [203] são generalizações de alta-dimensão das matrizes. O tensor cumulante de quarta ordem \mathbf{T}_4 é uma “matriz” de quatro dimensões onde cada elemento é definido por $q_{ijkl} = \text{cum}(x_i, x_j, x_k, x_l)$, os índices i,j,k e l variaram de 1 até N (onde N é o número de sinais) e $\text{cum}(x_i, x_j, x_k, x_l)$ é o cumulante de quarta ordem:

$$\begin{aligned} \text{cum}(x_i, x_j, x_k, x_l) &= E\{x_i, x_j, x_k, x_l\} - E\{x_i, x_j\}E\{x_k, x_l\} \\ &\quad - E\{x_i, x_k\}E\{x_j, x_l\} - E\{x_k, x_j\}E\{x_i, x_l\} \end{aligned} \quad (\text{A.29})$$

Sabe-se que a diagonalização da matriz de correlação (\mathbf{C}_y) produz a descorrelação entre os componentes de \mathbf{y} [4]. Para sinais independentes, apenas quando $i=k=j=l$ os cumulantes de quarta-ordem são diferentes de zero. Considerando isso, os métodos Tensoriais de ICA propõe a diagonalização de \mathbf{T}_4 para alcançar a independência estatística [146].

Embora teoricamente simples, a utilização de métodos tensoriais de ICA exigem uma grande quantidade de recursos computacionais para a decomposição em auto-valores de matrizes de quarta-ordem. O algoritmo JADE propõe um método aproximado para a diagonalização de \mathbf{T}_4 , se tornando mais leve computacionalmente.

Considerando que os dados satisfazem o modelo da ICA para dados pré-branqueados, pode-se escrever:

$$\mathbf{z} = \mathbf{V}\mathbf{A}\mathbf{s} = \mathbf{W}^T\mathbf{s} \quad (\text{A.30})$$

onde $\mathbf{x} = \mathbf{As}$ são os sinais observados, \mathbf{V} é a matriz de branqueamento e $\mathbf{W}^T = VA$ é a matriz de misturas branqueada. Neste caso, pode-se provar (ver [4] que o tensor

cumulante de \mathbf{z} tem uma estrutura especial e suas auto-matrices são descritas por:

$$\mathbf{M} = \mathbf{w}_m \mathbf{w}_m^T \quad (\text{A.31})$$

onde $m=1,\dots,N$ e w_n são as colunas da matriz \mathbf{W}^T

O algoritmo JADE utiliza a transformação linear F_{ij} da matriz \mathbf{M} definida por:

$$F_{i,j}(\mathbf{M}) = \sum m_{kl} \text{cum}(x_i, x_j, x_k, x_l) \quad (\text{A.32})$$

onde m_{kl} é um elemento da matriz \mathbf{M} .

A decomposição em autovalores é vista como um processo de diagonalização, então busca-se a matriz \mathbf{W} que diagonaliza $F(\mathbf{M})$ para qualquer \mathbf{M} (ou seja, $Q = \mathbf{W}F(\mathbf{M}_i)\mathbf{W}^T$ é uma matriz diagonal).

A função custo do método JADE busca a diagonalização de \mathbf{Q} pela maximização da soma dos elementos de sua diagonal. As matrizes \mathbf{M}_i utilizadas são as automatrices do tensor cumulante dos dados, pois assim tem-se um conjunto de N matrizes que contém toda a informação relevante a respeito dos cumulantes.

Os métodos tensoriais [204, 146] foram, provavelmente, a primeira classe de algoritmos capazes de executar a ICA de modo realmente eficiente [4]. Atualmente, estes métodos são mais utilizados para sinais de baixa dimensão, pois o custo computacional aumenta rapidamente com o número de componentes a serem estimados.

Algoritmo Multiplicativo com Iteração de Newton

Um algoritmo multiplicativo para ICA foi proposto por Akuzawa e Murata em [205]. Usando a curtose como função para avaliar a independência, esse algoritmo utiliza estratégia de otimização de segunda ordem, através do método de Newton [202], para estimar as componentes independentes.

O algoritmo proposto por Akuzawa não requer pré-branqueamento, operando diretamente sobre os dados medidos. Resultados experimentais obtidos em [206, 207] indicam que o algoritmo de Akuzawa apresenta melhor desempenho que FastICA e JADE quando os sinais estão contaminados por ruído Gaussiano.

Considerando a transformação linear $\mathbf{Y} = \mathbf{C}\mathbf{X}$, o objetivo do algoritmo de Akuzawa é encontrar a matriz \mathbf{C} que maximiza a independência entre as componentes de \mathbf{y} . Os passos a seguir são executados durante as iterações:

1. Escolher C_0 (a matriz de separação inicial) e a matriz Δ_0 ($N \times N$);
2. Calcular a iteração $C_t = \exp(\Delta_t - 1)C_{t-1}$;
3. Avalie a função custo em C_t usando uma expansão de segunda ordem em torno de C_{t-1} ;
4. Δ_t é escolhido como o ponto de sela da função custo;
5. Retorne para o passo 2 até convergir.

Mais detalhes a respeito da execução do passo 4 podem ser encontradas em [205]. Modificações no método de Akuzawa foram propostos em [206] com o objetivo de reduzir o custo computacional pela substituição das iterações de Newton pelo método quasi-Newton [202].

Comparado com FastICA e JADE, o algoritmo multiplicativo de Akuzawa é mais lento (mesmo em sua versão modificada que utiliza o método de otimização quasi-Newton), suas vantagens aparecem quando o nível de ruído aumenta, neste caso o método de Akuzawa apresenta melhores resultados.

A.4.4 Extensões ao modelo básico de ICA

A análise de componentes independentes em sua formulação básica mostrada nas equações (5.4) e (??) é empregada com sucesso em uma grande variedade de aplicações. O modelo básico de ICA não considera que os sinais podem estar contaminados por ruído aditivo, ou que o sinal misturado seja gerado de forma não-linear. No ambiente do calorímetro do ATLAS, o ruído estará presente nas células absorvedoras de energia, e por características de construção do sistema, transformações não-lineares dos sinais poderão estar presentes.

ICA para sinais ruidosos

Sabe-se que os sinais reais no ambiente do calorímetro do ATLAS estarão contaminados por ruído. A formulação básica da ICA, mostrada na equação (5.4), não considera a presença de ruído. Um modelo mais realista permite a obtenção do modelo da ICA para sinais ruidosos (*Noisy ICA*) [4]:

$$\mathbf{x} = \mathbf{A}\mathbf{s} + \mathbf{n}, \quad (\text{A.33})$$

onde $\mathbf{n} = (n_1, n_2, \dots)$ é o vetor de ruído. Em geral, considera-se que o ruído é gaussiano e independente das componentes independentes. Assume-se também que a matriz de covariância do ruído $\Sigma = \sigma^2 I$ é conhecida.

A equação (A.33) pode ser associada a ruído de sensores, uma vez que o ruído n_i é adicionado separadamente a cada fonte x_i . O ruído de fonte, que é adicionado às componentes independentes s_i é representado pela equação a seguir:

$$\mathbf{x} = \mathbf{A}(\mathbf{s} + \mathbf{n}) \quad (\text{A.34})$$

Um caso especial que pode simplificar bastante o problema de estimação das componentes independentes em ambiente ruidoso acontece quando existem poucas fontes de ruído. Se o número total de componentes de ruído é menor que o de componentes independentes, o modelo básico da ICA pode ser adaptado. Define-se o vetor $\tilde{\mathbf{s}} = (s_1, \dots, s_k, n_1, \dots, n_l)^T$, onde s_i ($i = 1, \dots, k$) são as componentes independentes e n_j ($j = 1, \dots, l$) são as fontes de ruído. Assumindo-se que o número de misturas é $k + l$, então o modelo básico da ICA pode ser aplicado a $x = A\tilde{s}$ e usando-se um algoritmo deflacionário consegue-se estimar as k componentes menos gaussianas, que são as próprias ICs.

Na maioria das vezes, deseja-se considerar que o ruído foi adicionado a cada uma das misturas; assim, $k + l$ é maior que o número de misturas, e o modelo básico da ICA não pode ser aplicado ao vetor \tilde{s} .

O processo de estimação das componentes independentes já é bastante difícil na formulação básica, e quando o ruído é considerado o cenário piora consideravelmente. Nos modelos das equações (A.33) e (A.34) pode-se verificar que as ICs não são obtidas apenas com a inversão da matriz de mistura \mathbf{W} :

$$\mathbf{Wx} = \mathbf{s} + \mathbf{Wn} \quad (\text{A.35})$$

A solução do problema da equação (A.35) envolve duas etapas: o uso de métodos de otimização numérica para obter uma aproximação do ruído e, a partir de então, a estimação das componentes independentes. Em uma abordagem mais simples e que parece ser a mais promissora para o *noisy ICA*, são utilizadas técnicas de remoção de tendência para reduzir os efeitos do ruído, adaptando-se os métodos básicos de ICA para o caso ruidoso [208].

O algoritmo FastICA pode ser adaptado para sinais contaminados por ruído usando medidas não tendenciosas da Gaussianidade [209]:

$$\mathbf{w}^* = E\{\tilde{\mathbf{x}}g(\mathbf{w}^T)\tilde{\mathbf{x}}\} - (I + \tilde{\Sigma})\mathbf{w}E\{g'(\mathbf{w}^T\tilde{\mathbf{x}})\}, \quad (\text{A.36})$$

sendo que a norma do novo valor \mathbf{w}^* é tornada unitária após cada iteração, e $\tilde{\Sigma}$ é dado por:

$$\tilde{\Sigma} = E\{\tilde{\mathbf{n}}\tilde{\mathbf{n}}^T\} = (\mathbf{C} - \Sigma)^{-1/2}\Sigma(\mathbf{C} - \Sigma)^{-1/2} \quad (\text{A.37})$$

onde $\mathbf{C} = E\{\mathbf{x}\mathbf{x}^T\}$ é a matriz de covariância dos sinais ruidosos observados. A função $g(.)$ pode ser escolhida entre as abaixo:

Outras técnicas como o “encolhimento de código esparso” (*Sparse Code Shrinkage*) [210] também são utilizadas para a extração das componentes independentes para sinais ruidosos. Em casos onde o nível de ruído é muito alto, pode-se tentar o uso de técnicas de processamento de sinais como *wavelets* ou filtragem adaptativa [211].

A.5 ICA não-linear

Conforme mostrado no Capítulo ??, o modelo da ICA não-linear (NLICA) apresenta uma formulação mais geral que o linear. A seguir será mostrado o desenvolvimento teórico de um algoritmo para a estimação das componentes independentes no modelo pós não-linear.

A.5.1 Algoritmo Taleb-Jutten para o modelo PNL

Um dos primeiros algoritmos para o modelo pós não-linear da ICA foi proposto por Taleb e Jutten no trabalho [108]. Este algoritmo é robusto a variações na distribuição de probabilidade das fontes, pois executa estimativa iterativa da estatística das componentes independentes estimadas através do cálculo da função escore:

$$\psi = p'_{Y_i}(u)/p_{Y_i}(u), \quad (\text{A.38})$$

conforme Figura A.4.

Cada função não-linear g_k ($k=1,\dots,N$) é modelado por redes MLP com um neurônio linear na saída:

$$g_k(u) = \sum_{h=1}^{N_H} \xi^h \sigma(\omega^h u - \eta^h), \quad (\text{A.39})$$

onde N_H é o número de neurônios ocultos. A divergência de Kullback-Lieber é utilizada para encontrar as regras de aprendizado para a estimativa das funções não-lineares [108].

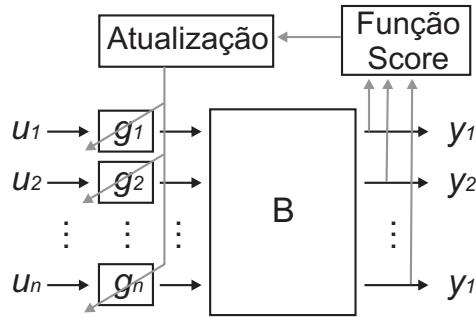


Figura A.4: Diagrama do algoritmo de Taleb-Jutten para o modelo PNL.

Como existem vários parâmetros a serem ajustados no modelo inverso proposto e a otimização envolve funções não-lineares, o algoritmo pode apresentar problemas de convergência para mínimos locais [99]. Diferentes procedimentos foram propostos na literatura para melhorar a eficiência de estimativa em modelos PNL. Em [115, 114] um algoritmo genético [156] foi utilizado para executar uma busca global, evitando o problema dos mínimos locais. O problema com esta abordagem é o aumento do custo computacional.

Redes neurais com arquiteturas alternativas também foram aplicadas com sucesso na separação de misturas PNL. Por exemplo, em [212] funções de base radial (RBF - *Radial Basis Function*). Em um outro trabalho [112], a separação foi realizada por redes neurais com funções de ativação do tipo *spline*.

Apêndice B

Conceitos fundamentais em classificação de sinais

A seguir serão mostrados os fundamentos teóricos de algumas técnicas de classificação de padrões, iniciando-se com uma visão geral do problema de decisão binária. Serão apresentadas técnicas lineares como Filtros Casados e Análise de Discriminantes, e não-lineares, como Redes Neurais.

B.1 Teste de Hipóteses

Considerando-se inicialmente a discriminação entre duas hipóteses H_1 e H_0 , o problema de classificação pode ser resumido pelo esquema da Figura B.1. A fonte gera as saídas, que após passarem por um meio probabilístico, precisam ser detectadas a partir das observações do processo. As regras de decisão, que formam o sistema classificador, são projetadas para maximizar a probabilidade de detecção correta.

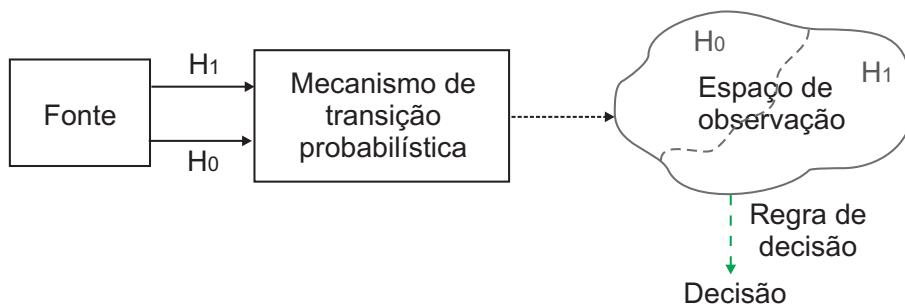


Figura B.1: Esquemático do problema de classificação binário.

No caso da decisão binária, cada vez que uma observação é efetuada 4 situações podem ocorrer:

decidir pela hipótese H_1 , sendo H_0 verdadeira;

decidir pela hipótese H_0 , sendo H_1 verdadeira;

decidir pela hipótese H_1 , sendo H_1 verdadeira;

decidir pela hipótese H_0 , sendo H_0 verdadeira.

As duas primeiras são erros de decisão e as duas últimas classificações corretas. Cada uma das hipóteses é associada a uma saída da fonte, que é mapeada em uma região do espaço de observação. Considerando um espaço de observação de dimensão N finita, um ponto neste espaço pode ser representado por um vetor:

$$\mathbf{r} = [r_1, r_2, \dots, r_N]. \quad (\text{B.1})$$

O mecanismo de transição probabilística gera pontos de acordo com as densidades de probabilidade condicionais $P_{\mathbf{r}/H_0}(\mathbf{R}/H_0)$ e $P_{\mathbf{r}/H_1}(\mathbf{R}/H_1)$. Quando essas probabilidades são conhecidas ou podem ser estimadas de alguma forma, o projeto do sistema classificador pode ser simplificado. Os critérios de *Bayes* e *Neyman-Pearson* são procedimentos clássicos utilizados para a escolha da regra de decisão.

B.2 Critério de Bayes

O critério de *Bayes* necessita do conhecimento das probabilidades a priori P_1 e P_0 de a fonte produzir H_1 ou H_0 , das probabilidades condicionais $P_{\mathbf{r}/H_0}(\mathbf{R}/H_0)$ e $P_{\mathbf{r}/H_1}(\mathbf{R}/H_1)$ e dos custos C_{ij} associados à escolha da hipótese i sendo j a verdadeira. O risco é, então, definido como [184]:

$$\begin{aligned} \mathfrak{R} = & C_{00}P_0 \int_{Z_0} P_{\mathbf{r}/H_0}(\mathbf{R}/H_0) d\mathbf{R} \\ & + C_{10}P_0 \int_{Z_1} P_{\mathbf{r}/H_0}(\mathbf{R}/H_0) d\mathbf{R} \\ & + C_{11}P_1 \int_{Z_1} P_{\mathbf{r}/H_1}(\mathbf{R}/H_1) d\mathbf{R} \\ & + C_{01}P_1 \int_{Z_0} P_{\mathbf{r}/H_1}(\mathbf{R}/H_1) d\mathbf{R} \end{aligned}, \quad (\text{B.2})$$

onde os elementos do espaço de observação que pertencem às partições Z_0 e Z_1 são associados, respectivamente, a H_0 e H_1 . As variáveis C_{ij} representam o custo da

escolha da hipótese i quando a hipótese verdadeira é a j . Em geral assume-se que o custo de uma decisão errada (C_{ij} sendo $i \neq j$) é maior do que o de um acerto (C_{ij} sendo $i = j$).

Minimizando o risco \mathfrak{R} da equação (B.2) chega-se a [184]:

$$\frac{P_{\mathbf{r}/H_1}(\mathbf{R}/H_1)}{P_{\mathbf{r}/H_0}(\mathbf{R}/H_0)} \gtrless_{H_0}^{H_1} \frac{P_0(C_{10} - C_{00})}{P_1(C_{01} - C_{11})}; \quad (\text{B.3})$$

a expressão à esquerda é chamada razão de semelhança ($\Lambda(\mathbf{R})$) e a fração à direita é o valor limiar (patamar) do teste (κ). Com isso, a equação (B.3) se reduz a:

$$\Lambda(\mathbf{R}) \gtrless_{H_0}^{H_1} \kappa; \quad (\text{B.4})$$

então, se a razão de semelhança é maior que o patamar, decide-se por H_1 , caso contrário, escolhe-se H_0 .

Quando os custos não são conhecidos, pode-se adotar o critério *minimax*, que minimiza o risco máximo; após algumas considerações chega-se a:

$$\begin{aligned} C_{00} &= C_{11} = 0 \\ C_{01}P_M &= C_{10}P_F \end{aligned}, \quad (\text{B.5})$$

onde $P_F = \int_{Z_1} P_{\mathbf{r}/H_0}(\mathbf{R}/H_0)d\mathbf{R}$ é a probabilidade de falso-alarme (terminologia usada em sistemas de radar, indicando que decidiu-se pela presença do alvo H_1 estando o mesmo ausente) e $P_M = \int_{Z_0} P_{\mathbf{r}/H_1}(\mathbf{R}/H_1)d\mathbf{R}$ é a probabilidade de perda do alvo.

O teste de *Neyman-Pearson* [213] é utilizado quando não se tem informações sobre os custos ou as probabilidades a priori. Escolhe-se um valor limite para a probabilidade de falso-alarme e procura-se minimizar a probabilidade de perda do alvo para o valor escolhido. Como o critério utiliza P_F e P_M é preciso conhecer as probabilidades condicionais $P_{\mathbf{r}/H_0}(\mathbf{R}/H_0)$ e $P_{\mathbf{r}/H_1}(\mathbf{R}/H_1)$.

B.3 Filtros Casados

Quando o processo de detecção é realizado em ambiente ruidoso o desempenho dos classificadores decaiu bastante. Os filtros casados são utilizados em processos de

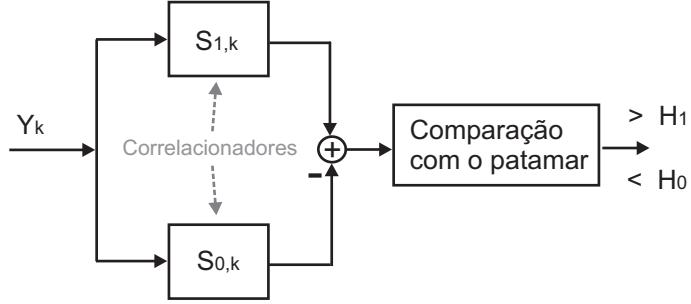


Figura B.2: Diagrama em blocos de um filtro casado.

detecção onde as formas de onda dos sinais corrompidos por ruído são conhecidas. Os filtros casados maximizam a razão sinal-ruído em ambiente de ruído branco. Um processo estocástico é dito branco quando sua densidade espectral de potência $S(\omega)$ é constante em todas as freqüências [197]:

$$S(\omega) = N_0/2 \quad (\text{B.6})$$

A autocorrelação do processo branco, calculada a partir da transformada de Fourier inversa de $S(\omega)$ é $R_{xx} = N_0/2\delta(t)$, ou seja, o ruído branco tem correlação nula com qualquer cópia deslocada dele mesmo.

No caso de sinais digitais, conforme mostrado na Figura B.2, o filtro casado é implementado através do cálculo da correlação do sinal recebido $y[k]$ com as formas de ondas esperadas $s_1[k]$ e $s_0[k]$. As saídas dos correlacionadores são subtraídas e comparadas com um patamar de decisão; se o valor for maior que o patamar decide-se por H_1 , caso contrário por H_0 .

Se o ruído de fundo não for branco, o desempenho dos filtros cai bastante. Esse problema é minimizado se o ruído é conhecido. Pode-se, então, projetar um filtro branqueador. A desvantagem do branqueador é que ele atua também sobre o sinal, alterando suas características estatísticas.

Embora muito utilizadas, por sua formulação matemática relativamente simples e bom desempenho em diversas aplicações, as técnicas mostradas até aqui apresentam as desvantagens de necessitarem conhecimento prévio a respeito das distribuições de probabilidade, custos ou forma de onda dos sinais de entrada. Em muitos casos práticos essas informações não estão disponíveis, sendo necessária a utilização de outros métodos de classificação. Além disso, o desempenho dessas técnicas piora

bastante em ambiente de ruído colorido e também quando as classes estão muito sobrepostas.

B.4 Discriminante linear de Fisher

A análise de discriminantes busca a direção \mathbf{w} onde as projeções \mathbf{y} dos sinais de entrada \mathbf{x} sejam maximamente separáveis. A análise por discriminante de Fisher (FDA - *Fisher Discriminant Analysis*) busca a direção ótima de discriminação utilizando 2 parâmetros: a distância inter-classes, e a distância intra-classes [214].

Numa formulação matricial o objetivo é encontrar a direção \mathbf{w}_0 que maximiza a expressão:

$$J(\mathbf{w}) = \frac{\mathbf{w}^T \mathbf{S_B} \mathbf{w}}{\mathbf{w}^T \mathbf{S_w} \mathbf{w}} \quad (\text{B.7})$$

onde \mathbf{m}_i é a média da classe i , $\mathbf{S_B} = (\mathbf{m}_1 - \mathbf{m}_2)(\mathbf{m}_1 - \mathbf{m}_2)^T$ é a matriz de separação inter-classes e $\mathbf{S_w} = \mathbf{S}_1 + \mathbf{S}_2$ é a matriz de separação intra-classes, sendo:

$$\mathbf{S}_i = \sum_{\mathbf{x} \in \mathcal{D}_i} (\mathbf{x} - \mathbf{m}_i)(\mathbf{x} - \mathbf{m}_i)^T \quad (\text{B.8})$$

Pode-se provar que a direção ótima que maximiza (B.7) é dada por [134]:

$$\mathbf{w} = \mathbf{S_w}^{-1}(\mathbf{m}_1 - \mathbf{m}_2) \quad (\text{B.9})$$

O discriminante de Fisher é capaz de encontrar a transformação linear ótima dos sinais de entrada de modo que os sinais projetados $\mathbf{y} = \mathbf{w}^T \mathbf{x}$ tenham máxima separação. Pode-se realizar a análise por discriminante de Fisher de modo analítico usando-se as equações (B.7), (B.8) e (B.9), ou de modo iterativo a partir de uma rede neural de uma camada.

B.5 Classificadores Neurais

As redes neurais artificiais (RNA) [5] são modelos matemáticos que emulam algumas características do cérebro humano, sendo capazes de adquirir conhecimento (aprender) e generalizar (responder corretamente a estímulos novos). Devido ao poder computacional, obtido de sua estrutura paralelamente distribuída, e às habilidades

de aprender e generalizar, as RNAs vêm sendo utilizadas em diversas aplicações como reconhecimento de padrões e classificação [215], processamento de sinais [216], aproximação de funções [217], controle e identificação de sistemas [218]. A seguir serão descritos os fundamentos matemáticos das redes neurais e sua aplicação como classificadores. Maior ênfase será dada às redes alimentadas adiante (*feedforward neural networks*) com treinamento supervisionado. Os livros [5] e [219] fornecem textos mais abrangentes sobre assunto.

Uma diferença fundamental entre os classificadores neurais e os métodos clássicos como os filtros casados é que nestes últimos é necessário formular um modelo matemático a partir dos sinais. Na abordagem neural, o classificador trabalha diretamente no conjunto de dados, ficando o modelo matemático implícito nos valores dos pesos sinápticos obtidos após o treinamento. Os classificadores neurais têm obtido resultados bastante animadores em problemas para os quais métodos clássicos não apresentam bom desempenho.

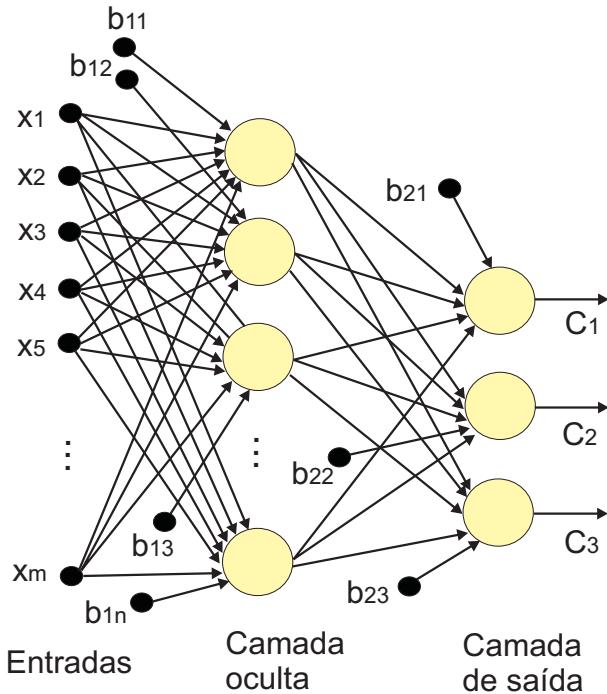


Figura B.3: Exemplo de uma rede neural utilizada para separação dos sinais de entrada em 3 classes.

Para o projeto do classificador neural dividem-se os pares entrada saída disponíveis nos conjuntos de treino, teste e validação. A divisão é importante para garantir que a rede treinada consiga generalizar bem o conhecimento adquirido. Apesar disso, é comum que a rede seja treinada com todos os dados disponíveis, e que o resultado seja validado com um conjunto de teste.

nas o conjunto de treino é considerado para o ajuste dos pesos sinápticos. O conjunto de validação é utilizado na parada do treinamento em caso de sobre-aprendizado. O sobre-aprendizado acontece quando a rede se ajusta demais ao conjunto de treino, perdendo capacidade de generalização. O resultado esperado do classificador é avaliado através do conjunto de teste, que não foi usado no ajuste dos pesos nem na parada do algoritmo.

Para decisão entre duas classes, pode-se usar na camada de saída um neurônio tipo tangente hiperbólica, associando-se $y = +1$ para uma classe e $y = -1$ para a outra. Para decisão entre N classes ($N \geq 3$), pode-se fazer com que a camada de saída tenha N neurônios. Associa-se um neurônio para cada classe. Quando a classe k estiver presente, a saída desejada no neurônio k associado deve ser $+1$ e nos demais -1 . Na fase de operação decide-se pela classe do neurônio com a maior saída. A rede da Figura B.3 é um exemplo de arquitetura que pode ser usada para a separação entre três classes a partir de um espaço de entrada de dimensão m .

Considerando uma rede de duas camadas (ver Figura B.3), a camada oculta é responsável por um mapeamento não-linear das entradas. Este mapeamento busca a máxima separação das classes diferentes, facilitando o processo de classificação realizado pela camada de saída.

Apêndice C

Algoritmos Genéticos

C.1 Algoritmo Genético como Método de Otimização

Os Algoritmos Genéticos (AG) pertencem à classe dos Algoritmos Evolutivos, que são ferramentas de busca e otimização bastante difundidas, utilizando operadores que simulam a evolução dos seres vivos para buscar a solução de problemas complexos [220].

Comparado com outros métodos de busca e otimização os AG's se destacam quando o problema tratado tem características altamente não lineares e o ambiente de busca é multi-modal (existem diversos máximos ou mínimos locais na função custo). Por ser um método de busca global, o AG tem menor probabilidade de convergir para soluções sub-ótimas (máximos/mínimos locais).

Outra vantagem é a facilidade de implementação, pois, em geral, os métodos numéricos precisam que a função objetivo $f(x)$ seja conhecida e diferenciável, pois se baseiam no cálculo das derivadas parciais. Já o AG utiliza apenas a função objetivo.

Idealizados inicialmente por John Holland e seus alunos nos anos 70 [221], na Universidade de Michigan, os Algoritmos Genéticos são métodos de busca baseados na teoria da evolução. O objetivo inicial de Holland era utilizar os mecanismos de evolução, adaptação e seleção natural dos seres vivos como modelos para sistemas computacionais.

Na natureza os indivíduos apresentam qualidades particulares que podem torná-los mais ou menos aptos para sobreviver às adversidades do meio ambiente. A seleção natural e os operadores genéticos, como a reprodução e a mutação, podem ocasionar o surgimento de exemplares com características mais favoráveis, melhorando as chances de sobrevivência da população.

Nos problemas computacionais, traçando um paralelo com a natureza, as possíveis soluções são chamadas cromossomos ou indivíduos, cada um possuindo características que o fazem melhor ou pior candidato à resolução do problema. Os melhores são selecionados para reproduzirem-se gerando uma nova geração. Espera-se que a cada iteração apareçam indivíduos mais aptos. A mutação, por sua vez, introduz aleatoriamente informações não existentes no conjunto original, contribuindo para a manutenção da diversidade populacional (explorando outras regiões do espaço de busca). Algumas características interessantes dos AG's esão listadas a seguir:

1. trabalham com os parâmetros codificados;
2. realizam a busca a partir de uma população de soluções candidatas (processamento em paralelo);
3. utilizam informações da função objetivo, e não de suas derivadas;
4. usam regras probabilísticas e não determinísticas.

C.2 Estrutura de um Algoritmo Genético

Não existe uma definição rigorosa para os Algoritmos Genéticos, entretanto, a maioria dos métodos assim denominados tem em comum as seguintes características: população composta de cromossomos, seleção de acordo com a função de adequabilidade, recombinação e mutação para a produzir uma nova geração [222]. A seguir tem-se uma breve descrição dos conceitos mais importantes na teoria dos Algoritmos Genéticos.

C.2.1 Conceitos Principais

- **Cromossomo** - Em geral, na teoria dos A.G.'s cada indivíduo é constituído de um único cromossomo, onde cada elemento equivale a um gene. Pode-se

utilizar codificação binária simples, códigos de gray, e até codificação decimal. A codificação binária é a mais comum, nela cada gene pode assumir apenas dois valores 0 ou 1. Os cromossomos podem ser vistos como um ponto do espaço de busca das prováveis soluções do problema.

- **Codificação** - As variáveis a serem otimizadas precisam ser codificadas nos cromossomos. A quantidade de bits destinada a cada variável deve ser escolhida de forma a minimizar o esforço computacional, mas sempre levando em conta a precisão desejada. Na figura C.1 tem-se a representação de um cromossomo binário, onde cada parâmetro é codificado por 4 bits distintos.

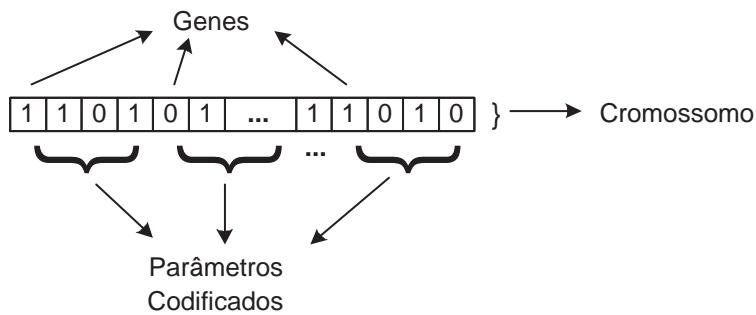


Figura C.1: Exemplo de um cromossomo binário

- **População** - Conjunto de cromossomos gerado inicialmente de forma aleatória, a população é atualizada a cada nova iteração do algoritmo, espera-se que a cada geração apareçam indivíduos mais aptos para a solução do problema. Em geral, usam-se populações de tamanho fixo.
- **Função Aptidão (*Fitness Function*)** - É na Função Aptidão ($fit(x)$) que são codificadas as informações sobre o sistema a ser otimizado. A cada geração são calculados os valores da Função Aptidão de todos os indivíduos. Esta informação será usada como base do processo de seleção.
- **Seleção** - Simula o processo de seleção natural, os indivíduos mais aptos, com maior valor da função aptidão, têm mais chances de serem escolhidos para o processo de recombinação.
- **Recombinação (*Crossover*)** - Promove a troca de informação genética. Os indivíduos são selecionados dois a dois, simulando um processo de reprodução

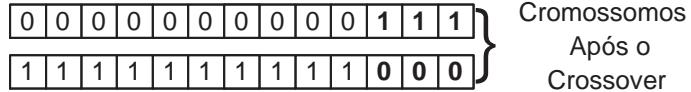
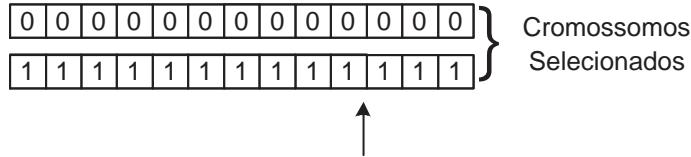
sexuada, e parte do material genético (bits) é trocado. Existem vários modos de proceder esta troca de bits, cada um dá origem a um tipo diferente de recombinação, a seguir temos a descrição dos tipos mais comuns de *crossover* em cromossomos binários:

- *Single-Point Crossover* - O modo mais simples e comum de recombinação, é escolhido aleatoriamente um ponto no cromossomo e são trocados os bits após este ponto.
- *Multi-Point Crossover* - Semelhante à recombinação de ponto único, pontos aleatórios são escolhidos, e trocados os bits dos segmentos que ficam entre estes pontos.
- *Uniform Crossover* - Alguns pontos são escolhidos de modo randômico e trocados os bits destes pontos. É gerado um cromossomo aleatório, as posições onde o valor do bit é igual a 1 serão trocadas. O *uniform crossover* foi proposto inicialmente em [223] e [224]. Na figura C.2 são ilustrados os processos de recombinação de ponto único e uniforme.
- **Mutação** - A mutação é responsável pela manutenção da diversidade genética, inserindo, de modo aleatório, novas informações nos indivíduos. No caso de codificação binária, o bit selecionado tem o seu valor invertido.
- **Fenótipo** - Em analogia com o termo biológico, o fenótipo de um indivíduo pode ser expresso pelo valor da função aptidão, ou pelos parâmetros decodificados.

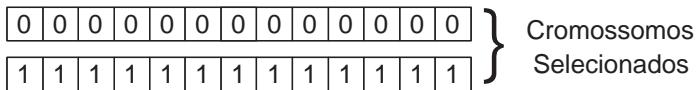
A recombinação e a mutação são controlados por parâmetros específicos, p_{rec} e p_{mut} , respectivamente. Estes parâmetros estabelecem a taxa de ocorrência dos processos. Exemplo: para $p_{rec} = 0.7$, a cada geração, 70% dos indivíduos são selecionados para participarem da recombinação, com $p_{mut} = 0.001$, a cada geração 0,1% dos bits sofrerão mutação.

C.2.2 Escalonamento de Aptidão

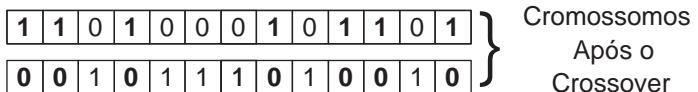
É importante salientar que problemas com a pressão seletiva (rigor com o qual é realizado o processo de seleção) do algoritmo podem levar à demora na convergência



(a)



1 1 0 1 0 0 0 1 0 1 1 0 1 → Padrão do Crossover



(b)

Figura C.2: (a) Recombinação em ponto único e (b) recombinação uniforme.

ou à convergência prematura. A última se caracteriza quando o AG concentra a busca em uma pequena região do domínio, e acaba apontando para um mínimo local. E a convergência lenta, acontece se o algoritmo, embora próximo do mínimo global, não consegue atender às condições de parada. Estes problemas podem ser minimizados com a adoção de um procedimento simples chamado escalonamento de aptidão (*fitness scaling*). No início do processo de busca é interessante que a pressão seletiva seja pequena, permitindo que o AG explore a maior porção possível do espaço de domínio. Após muitas gerações, a população já se encontra em um estágio mais avançado, sendo conveniente aumentar a pressão seletiva, uma vez que os indivíduos tem valores de *fitness* muito próximos, dificultando o destaque dos melhores candidatos. O escalonamento linear foi proposto por Goldberg em [156] e o valor da função aptidão escalonada (fit_e) é dada por:

$$fit_e = a fit + b \quad (C.1)$$

os coeficientes a e b são calculados pela rotina `fitescala.m`, descrita no apêndice B, utilizando valores entre 1,2 e 2 para o fator multiplicador Cm .

C.2.3 Implementação de um Algoritmo Genético

Um AG com codificação binária pode ser facilmente implementado utilizando-se *strings* de bits para designar os cromossomos. Os demais operadores podem ser implementados com operações de manipulação de bits. A maior parte do processamento computacional fica no cálculo da função aptidão, nesta etapa os cromossomos devem ser decodificados de valores binários em reais para obtenção dos parâmetros (x_i), e com estes calcula-se os valores da função aptidão ($fit(x)$) para cada indivíduo.

Para que a busca com um AG se torne eficiente, é preciso escolher corretamente os parâmetros do algoritmo. O tamanho da população, o número de bits destinado a cada variável e as taxas de recombinação e mutação exercem papel fundamental na velocidade de convergência e na globalidade da busca efetuada.

O uso de populações pequenas não permite o mapeamento de todas as regiões do espaço de domínio, já populações muito grandes tornam o algoritmo muito lento. Baixas taxas de mutação podem fazer a busca perder a generalidade, ou seja, ficar concentrada em pequenas partes do espaço de busca, valores de p_{mut} muito altos podem gerar uma busca aleatória. Na escolha do p_{rec} deve-se permitir que uma parte da população permaneça inalterada, sob pena de gerar uma nova geração com indivíduos inferiores aos da anterior. No algoritmo desenvolvido, após uma série de testes práticos, onde diferentes valores destes parâmetros foram utilizados, decidiu-se pela utilização de: $p_{mut} = 5\%$, $p_{rec} = 80\%$ e uma população de 150 indivíduos.

C.2.4 O Algoritmo Genético Utilizado

O algoritmo Genético usado é semelhante ao proposto por Goldberg em [156], com modificações no processo de recombinação e a implementação de estratégias elitistas. Diferente do *single-point crossover*, proposto por Goldberg, onde é escolhido um ponto e trocados os bits que estão depois deste ponto, foi utilizada a recombinação ou *crossover uniforme*, que é feita escolhendo-se aleatoriamente o número e a posição dos bits a serem trocados. O elitismo foi implementado garantindo-se

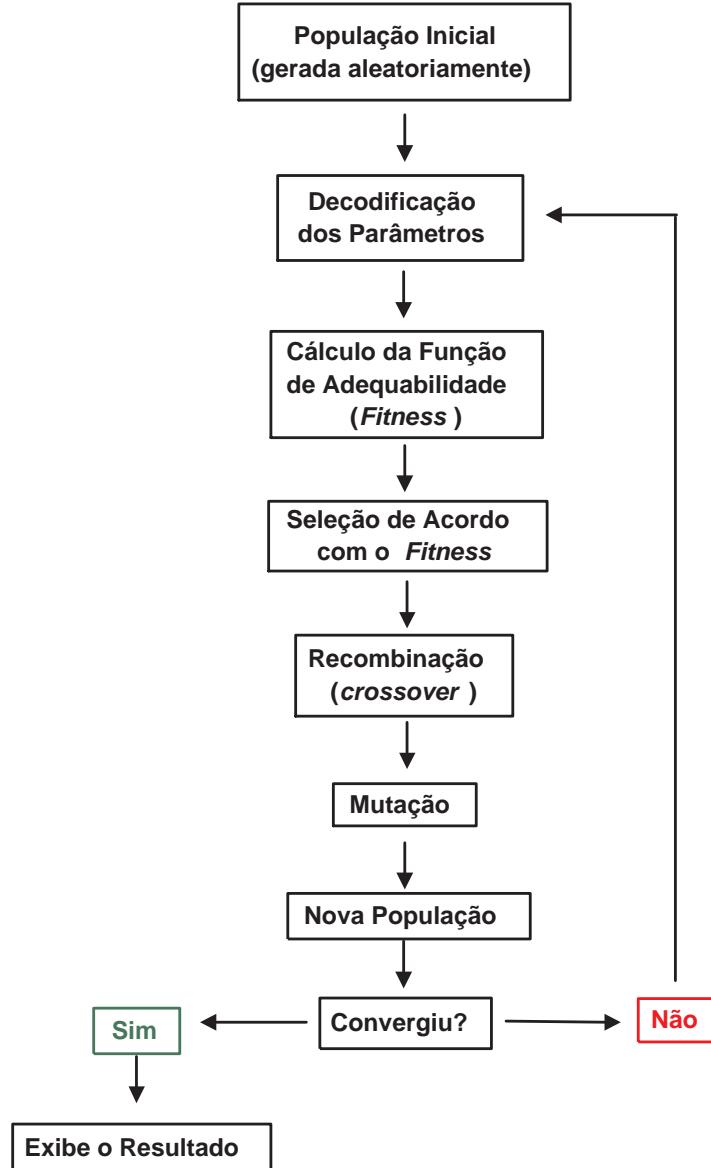


Figura C.3: Fluxo de um Algoritmo Genético

que cópias dos melhores indivíduos de cada geração estejam presentes na geração seguinte. Utilizou-se também, o escalonamento da aptidão, proposto em [156], este procedimento ameniza o problema da convergência prematura do algoritmo, regulando a pressão seletiva.

Escolheu-se como função objetivo a expressão do erro quadrático médio da equação ???. Como o AG foi desenvolvido para selecionar os indivíduos com maiores valores da função *fitness* ($fit(\mathbf{x})$), utilizou-se:

$$fit(\mathbf{x}) = \frac{1}{(f(\mathbf{x}, t) - f_e(\mathbf{x}, t))^2 + \zeta} \quad (C.2)$$

onde a constante ζ impede que $fit(\mathbf{x})$ assuma valores muito elevados à medida que o erro diminui.

Para uma boa amostragem do espaço de busca, é interessante usar populações grandes, mas, isso eleva muito o tempo de processamento. Na impossibilidade de fazer uso de um grande número de indivíduos, precisou-se implementar soluções alternativas. Na tentativa de corrigir problemas do algoritmo genético empregou-se alguns procedimentos:

Elitismo garante que os melhores indivíduos estejam presentes na geração seguinte, sem o risco de serem perdidos nos processos de *crossover* e mutação.

Recombinação uniforme para populações pequenas, age de modo mais eficiente que a de ponto único, facilitando a exploração das diversas regiões do espaço de busca.

Genocídio periódico Também visando evitar uma amostragem incompleta e a rápida perda de diversidade em populações pequenas, utilizou-se este procedimento de fácil implementação [225], que consiste em, a um número conveniente de gerações, aumentar o número de indivíduos de N para $\kappa \times N$, sendo κ um inteiro positivo, e depois, através de uma seleção rigorosa eliminar os $(\kappa - 1)N$ de menor aptidão.

Apêndice D

Listas de Publicações

D.1 Publicações em periódicos

1. *Eduardo F. Simas Filho, José M. de Seixas e Luiz P. Calôba. Modified Post-Nonlinear ICA for Online Neural Discrimination, Neurocomputing*, 2010 (aceito para publicação).

Resumo:

Neste trabalho foi proposta uma modificação no modelo pós não-linear da ICA (*Independent Component Analysis*) que permite a estimativa de um conjunto de componentes que são ao mesmo tempo independentes e discriminantes. O algoritmo proposto foi aplicado para o problema da identificação de elétrons no segundo nível de filtragem do detector ATLAS. O método proposto produziu melhor desempenho se comparado ao modelo pós não-linear tradicional.

2. *Eduardo F. Simas Filho, José M. de Seixas e Luiz P. Calôba. Optimized Calorimeter Signal Compaction for an Independent Component based ATLAS Electron/Jet Second-Level Trigger. Proceedings of Science* (nova denominação do *Journal of High Energy Physics/Online*), v. ACAT08, p. 1-10, 2009.

Resumo:

Neste trabalho foram estudados alguns métodos de compactação de sinais para serem aplicados como pré-processamento à análise de componentes

independentes. Entre os métodos testados pode-se destacar NLPCA (Análise de Componentes Principais Não-Lineares - *Nonlinear Principal Component Analysis*), PCA (Análise de Componentes Principais) e PCD (Componentes Principais de Discriminação - *Principal Components of Discrimination*). O processo de extração de características foi realizado de modo segmentado (a nível de cada camada do calorímetro).

3. *Eduardo F. Simas Filho, José M. de Seixas e Luiz P. Calôba. High-Energy Particles Online Discriminators Based on Nonlinear Independent Components.* *Lecture Notes in Computer Science*, v. 5441, p. 718-725, 2009.

Resumo:

Um estudo comparativo entre três modelos distintos de estimação das componentes independentes não-lineares para o problema da filtragem *online* de elétrons no ATLAS foi conduzido neste trabalho. A NLICA foi estimada a partir: de um modelo sem restrições estruturais (utilizando mapas auto-organizáveis), de um modelo com restrições estruturais (utilizando um algoritmo para o modelo pós não-linear) e da ICA Local (que é uma abordagem diretamente ligada ao problema da ICA não-linear). Os resultados obtidos foram comparados em termos do desempenho de discriminação.

4. *Eduardo F. Simas Filho, José M. de Seixas e Luiz P. Calôba. Self-organized mapping of calorimetry information for high efficient online electron/jet identification in ATLAS.* *Proceedings of Science* (nova denominação do *Journal of High Energy Physics/Online*), v. ACAT07, p. 1-13, 2007.

Resumo:

Um estudo da aplicação de mapas auto-organizáveis (SOM - *Self-organizing Maps*) para a extração de características na filtragem *online* de elétrons do ATLAS foi conduzido neste trabalho. Os SOM foram aplicados de modo não-segmentado (considerando como entrada um vetor composto pelos sinais em anéis gerados a partir de todas as camadas dos calorímetros).

5. *Eduardo F. Simas Filho, José M. de Seixas. Nonlinear Independent Com-*

ponent Analysis: Theoretical Review and Applications. *Learning and Nonlinear Models*, v. 5, p. 99-120, 2007.

Resumo:

Um tutorial a respeito da teoria e dos diversos algoritmos para estimação do modelo não-linear da análise de componentes independentes (em suas diversas variações) foi apresentado neste trabalho. Foram realizados testes de desempenho com diversos algoritmos utilizando uma base de dados onde havia controle sobre o processo da mistura dos sinais (consequentemente os sinais a serem estimados eram conhecidos). Uma aplicação da NLICA para extração de características no segundo nível de filtragem do detector ATLAS também foi mostrada.

D.2 Capítulo de Livro

1. *Rodrigo C. Torres, Eduardo F. Simas Filho, Danilo E. F. Lima e J. M. Seixas. Segmented Online Neural Filtering based on Independent Components of Pre-Processed Information. Signal Processing.* Vienna: In-Tech, 2010, v.1 , p. 337-358, (aceito para publicação).

Resumo:

Neste trabalho foi realizado um estudo comparativo de diversas abordagens baseadas em análise de componentes independentes (ICA - *Independent Component Analysis*) para extração de características na filtragem online do ATLAS. Os componentes independentes foram estimados de ambos os modos, segmentado e não-segmentado. Para a classificação foram utilizados classificadores neurais e classificadores lineares (discriminante linear de Fisher).

D.3 Artigos em Conferências

1. *Rodrigo C. Torres, Danilo E. F. Lima, Eduardo F. Simas Filho e J. M. Seixas. Neural Online Filtering Based on Preprocessed Calorimeter Data.* In: IEEE Nuclear Science Symposium and Medical Imaging Conference, p.

530-536, Orlando, 2009.

Resumo:

Um novo conjunto de sinais simulados, com características mais próximas da operação real esperada para o detector, foi utilizado neste trabalho. Um estudo detalhado sobre diferentes métodos de compactação como Análise de Componentes Principais (PCA - *Principal Component Analysis*) e Componentes Principais de Discriminação (PCD - *Principal Components of Discrimination*) foi conduzido.

2. *Eduardo F. Simas Filho, José M. de Seixas e Luiz P. Calôba. Combinação de Classificadores Neurais Segmentados com Pré-processamento por Análise de Componentes Independentes para um Sistema Online de Filtragem.* In: Congresso Brasileiro de Redes Neurais e Inteligência Computacional, p. 1-5, Ouro Preto-MG, 2009.

Resumo:

Neste trabalho foi proposto o uso de classificadores neurais especialistas nas informações de cada camada do calorímetro. Alguns métodos para combinar as saídas dos classificadores segmentados foram testados. Com o uso dos classificadores especialistas foi possível aumentar a eficiência de discriminação e ainda identificar que existe redundância na informação disponível nas diversas camadas dos calorímetros. Neste caso, foi mostrado que, mesmo sem utilizar os sinais de algumas camadas o desempenho de discriminação se mantém quase inalterado. Como o processo de formação dos sinais em anéis é responsável pela maior parte do tempo de processamento, quanto menor a quantidade de informação necessária mais rápida a decisão é tomada.

3. *Eduardo F. Simas Filho, José M. de Seixas e Luiz P. Calôba. Local Independent Component Analysis Applied to Highly Segmented Detectors.* In: IEEE International Symposium on Circuits and Systems (ISCAS08), v. 1, p. 3005-3008, Seattle, 2008.

Resumo:

Neste trabalho foi proposta a utilização do modelo da ICA Local para

a extração de características na filtragem online de segundo nível do ATLAS. Neste modelo um algoritmo não-supervisionado de agrupamento é utilizado para dividir o conjunto de sinais disponíveis em grupos de características semelhantes. A ICA é aplicada separadamente para os sinais de cada conjunto.

4. *Eduardo F. Simas Filho, José M. de Seixas e Luiz P. Calôba. Segmented Self-Organized Feature Extraction for Online Filtering in a High Event Rate Detector.* In: European Signal Processing Conference, p. 1-5, Lausanne, Suíça, 2008.

Resumo:

Mapas auto-organizáveis foram aplicados de modo segmentado (a nível de cada camada do calorímetro) para extração de características sobre os sinais em anéis. Um estudo para busca do tamanho ótimo do mapa foi conduzido considerando o erro médio de representação obtido. Foi mostrado que com a abordagem segmentada obteve-se melhor desempenho de discriminação do que no modo não-segmentado.

5. *Eduardo F. Simas Filho, José M. de Seixas e Luiz P. Calôba. Online Neural Filtering Operating Over Segmented Discriminating Components.* In: IEEE International Conference on Electronics, Circuits and Systems, v. 1. p. 530-533, Malta, 2008.

Resumo:

Foi proposta a estimativa dos componentes principais de discriminação (PCD - *Principal Discriminating Components*) de modo segmentado (a nível de cada camada do calorímetro), aproveitando, desta forma, toda a segmentação e granularidade disponíveis ao detector. As PCDs (obtidas em cada camada) foram concatenadas e utilizadas para treinar um classificador neural.

6. *Eduardo F. Simas Filho, José M. de Seixas e Luiz P. Calôba. Segmented Overdetermined Nonlinear Independent Component Analysis for Online Neural Filtering.* In: *Simpósio Brasileiro de Redes Neurais*, p. 159-164, Salvador, 2008.

Resumo:

Uma modificação no modelo pós não-linear da ICA foi proposta neste trabalho visando à aplicação em misturas sobre-determinadas (onde o número de sinais observados é maior que o número de componentes independentes a serem estimadas). O modelo proposto parece ser mais adequado para o problema da extração de características no ATLAS pois espera-se que o número de componentes independentes seja menor que o de anéis (100).

7. *Eduardo F. Simas Filho, José M. de Seixas e Luiz P. Calôba. Análise Não-Linear de Componentes Independentes Segmentadas para Filtragem Online num Detector com Alta Taxa de Eventos.* In: *Congresso Brasileiro de Automática*, p.1-6, Juiz de Fora, 2008.

Resumo:

Neste trabalho o modelo pós não-linear da NLICA (*Nonlinear Independent Component Analysis*) foi utilizado para extração de características de modo segmentado (a nível de cada camada do calorímetro). Foi utilizado um algoritmo que não permite a compactação dos sinais, então, após a estimativa das componentes foi calculada a relevância e as menos relevantes foram descartadas. Uma limitação deste procedimento é que todas (100) componentes independentes não-lineares precisam ser calculadas e muitas delas não são utilizadas para a discriminação, neste caso o esforço computacional (e consequentemente o tempo necessário para tomada de decisão) aumenta consideravelmente.

8. *Eduardo F. Simas Filho, José M. de Seixas e Luiz P. Calôba. Segmented Independent Component Analysis for online filtering using highly segmented detectors.* In: *International Conference on Intelligent Systems Design and Applications*, Rio de Janeiro, v. 1. p. 659-664, 2007.

Resumo:

A análise de componentes independentes (ICA - *Independent Component Analysis*) foi aplicada de modo segmentado (a nível de cada camada do calorímetro) para extração de características relevantes na discriminação de elétrons no segundo nível de filtragem do ATLAS. Para compactação foi

utilizada a análise de componentes principais (PCA - *Principal Component Analysis*)

9. *Eduardo F. Simas Filho, José M. de Seixas e Luiz P. Calôba. Análise de Componentes Independentes para Filtragem Online num Ambiente de Alta Taxa de Eventos e Informação Segmentada.* In: *VIII Congresso Brasileiro de Redes Neurais*, Florianópolis, p. 1-6, 2007.

Resumo:

A combinação PCA + ICA foi utilizada para compactação e extração de características no segundo nível de filtragem do ATLAS. Neste trabalho foi utilizada uma abordagem não-segmentada para extração de características, ou seja, os anéis produzidos em todas as camadas do calorímetro foram considerados como um único vetor de dados. Foi realizada também uma análise da relevância das componentes independentes estimadas, visando eliminar as componentes não-relevantes para a discriminação.

10. *Eduardo F. Simas Filho, Luiz P. Calôba e José M. de Seixas. Filtragem ótima para o Trigger de segundo nível do ATLAS baseado em calorimetria.* In: *XXVII Encontro Nacional de Física de Partículas e Campos*, Águas de Lindóia, p. 1-4, 2006.

Resumo:

Neste trabalho foram apresentados resultados iniciais da pesquisa onde os sinais em anéis foram identificados através de filtros casados. Visando a compactação dos sinais, testes foram conduzidos utilizando-se apenas os anéis mais energéticos de cada camada.

D.4 Resumos em Conferências

1. *Eduardo F. Simas Filho, José M. de Seixas e Luiz P. Calôba. ATLAS Neural Second-Level Trigger based on Nonlinear Independent Components of Segmented Calorimeter Information.* In: *Experimental High-Energy Physics and Associated Technologies Workshop*, Rio de Janeiro, 2008.

Resumo:

Um estudo envolvendo diversos métodos de estimação das componentes independentes no modelo não-linear foi conduzido neste trabalho.