

DISCRIMINADORES NEURAIIS DE PARTÍCULAS PARA UM DETECTOR SUBMETIDO A UMA ALTA TAXA DE EVENTOS

THIAGO CIODARO XAVIER*, ANDRÉ DOS ANJOS RABELLO†, JOSÉ MANOEL DE SEIXAS*

**Laboratório de Processamento de Sinais (LPS), Escola Politécnica, COPPE, UFRJ*

†*Universidade de Wisconsin*

Emails: xavier@lps.ufrj.br, Andre.Dos.Anjos@cern.ch, seixas@lps.ufrj.br

Abstract— This article presents the results of using neural networks for the optimization of the ATLAS online filtering system, one of the main detectors of the particle collider LHC (Large Hadron Collider). The Regions of Interests of the ATLAS energy measurer calorimeter are mapped in 100 rings of energy deposition, which feed a classifier neural network to classify them as electron or jet. For the signal pre-processing, it is used a relevance mapping and PCA (Principal Component Analysis) to compact the information, increasing the processing speed and, eventually, increasing the detection efficiency, with a decreasing of false alarm rate.

Keywords— CERN, ATLAS, nuclear physics, filter, classifier, neural networks.

Resumo— Esse artigo mostra os resultados da aplicação de redes neurais na otimização do sistema de filtragem online do detector ATLAS, um dos principais detectores posicionados ao redor do ponto de colisão do colisionador de partículas de última geração, o LHC (Large Hadron Collider). As Regiões de Interesse do calorímetro medidor de energia do ATLAS são mapeadas em 100 anéis de deposição de energia, os quais alimentam a rede neural classificadora, que a classifica como elétron ou jato. Como pré-processamento dos sinais, utiliza-se um mapeamento de relevância e PCA (Análise de Componentes Principais) para a compactação da informação, aumentando-se a velocidade de processamento e, eventualmente, aumentando-se a eficiência de detecção, com diminuição da taxa de falso alarme.

Palavras-chave— CERN, ATLAS, física nuclear, filtragem, classificação, redes neurais.

1 Introdução

O LHC (*Large Hadron Collider*) é um acelerador de partículas que está sendo construído no CERN (CERN, 2006) (Centro Europeu para a Pesquisa Nuclear) e colisionará partículas com elevada energia no centro de massa, através de um túnel circular com 26,7 km de circunferência dotado de poderosos ímãs para a aceleração das partículas (LHC, 2006). Cada colisão constitui um evento a ser analisado através de detectores posicionados em torno dos pontos de colisão. O ATLAS (ATLAS, 2006) (*A Toroidal LHC Apparatus*) é um dos detectores do LHC. Ele é composto por diferentes sistemas de detecção – o detector de traços, os calorímetros eletromagnético e hadrônico e a câmara de múons.

Os eventos de possível interesse do LHC são muito raros, demandando dias de operação a uma taxa de eventos elevada (em torno de 40 MHz). Como cada evento carrega aproximadamente 1 MB de informação, tem-se um fluxo de dados da ordem de 40 TB por segundo, impossibilitando o armazenamento completo desses eventos para análise *offline* (High Level Trigger, 2003). Desta maneira, um sistema de filtragem online torna-se indispensável para o experimento, composto por níveis, de forma a reduzir a taxa de eventos gradualmente (High Level Trigger, 2003).

Este trabalho se concentra na separação de elétrons e jatos (ruído), a assinatura de maior frequência dos possíveis eventos de interesse do LHC, utilizando somente as informações de calori-

metria do detector ATLAS. O discriminador neural proposto atuará no segundo nível de filtragem, com acesso a granularidade plena dos calorímetros e a todos os seus milhares de canais de leitura.

2 O Sistema de Calorimetria

O sistema de calorimetria do ATLAS é composto pelos calorímetros eletromagnético e hadrônico. O calorímetro eletromagnético, LArg (LArg Collaboration, 1996) (Liquid Argon Calorimeter), é dividido em 3 camadas com diferentes granularidades, sendo a segunda camada a mais profunda. Além disso, a seção e.m. é dividida em duas partes: o barril e a tampa. O barril e as duas tampas completam totalmente o espaço ao redor da colisão. O calorímetro e.m. tem, também, um pré-irradiador, que funciona praticamente como um calorímetro muito fino, com a função de recuperar a informação perdida no material morto da seção e.m. (fios, encapsamentos etc).

Os calorímetros hadrônicos do ATLAS (TileCal Collaboration, 1996) são formados pelo TileCal, calorímetro de telhas, e a tampa hadrônica. O TileCal é posicionado radialmente depois do LArg, abrangendo as porções do barril e sua extensão. Assim como a seção e.m., a seção hadrônica também possui granularidade variável.

Os dois calorímetros são dispostos ao redor do ponto de impacto e, por isso, têm o formato cilíndrico. O sistema de coordenadas do ATLAS, então, acompanha a direção do feixe de partículas

provenientes da colisão. As coordenadas empregadas são η , ϕ e z , representando, respectivamente, a direção das partículas após a colisão (também chamada de pseudo-rapidez) e a rotação em torno do eixo de colisão z .

3 O Sistema de Filtragem do ATLAS

O sistema de filtragem do ATLAS deve ser capaz de separar, de forma eficiente e rápida, física nova da massa de dados já conhecida, que funciona como ruído de fundo do experimento. Para isto, foi concebido o Sistema de Trigger (High Level Trigger, 2003), dividido em três níveis (Figura 1). É importante que se mantenha o tempo de latência (tempo para formar e distribuir a decisão do filtro) no valor mais baixo possível para cada nível de operação de modo a se adequar à elevada taxa de eventos do LHC. Durante esse tempo, a informação de todos os canais do detector precisa ser retida em memórias do tipo pipeline (High Level Trigger, 2003).

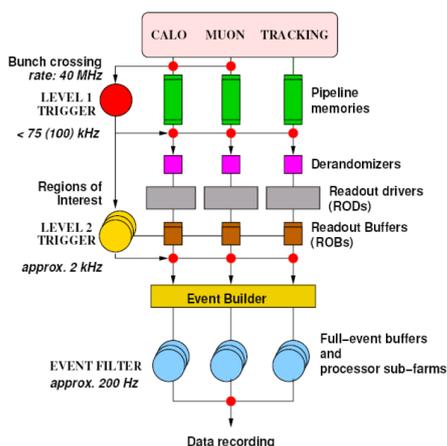


Figura 1: Sistema de Trigger

O primeiro nível de Trigger realiza a seleção inicial, possui um tempo de latência de $2\mu s$ e a maior taxa de filtragem de eventos. Assim, seu processamento é feito todo em hardware de alta velocidade, basicamente com FPGAs (Field Programmable Gate Array), e ele utiliza uma informação de calorimetria com granularidade menos fina, compondo uma soma de células. Este nível é responsável, também, por selecionar as regiões do calorímetro onde houve deposição significativa de energia, as chamadas Regiões de Interesse (RoIs), e passá-las ao segundo nível.

O segundo nível irá, então, processar somente os eventos selecionados pelo primeiro nível, já separados em RoIs. Os eventos rejeitados pelo primeiro nível são definitivamente descartados. Neste segundo nível, utiliza-se a granularidade plena do ATLAS, de modo a refinar a decisão do primeiro nível. Como a taxa de eventos é menor no segundo nível (100kHz), seu tempo de latência

é de 10ms, sendo sua implementação totalmente em software. Para ilustrar, na Figura 2, são mostradas RoIs típicas da segunda camada da seção e.m., para um elétron e um jato que foram aceitos pelo primeiro nível.

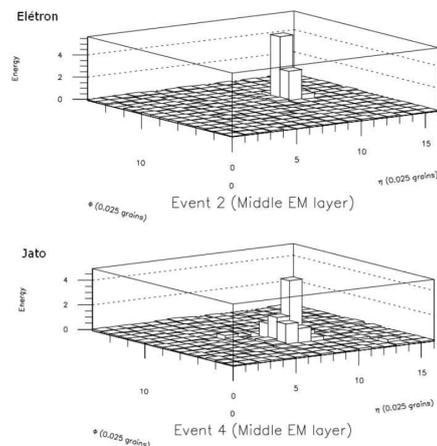


Figura 2: Duas RoIs típicas da segunda camada e.m..

O terceiro nível possui uma latência maior (aproximadamente 1s), esperando-se que a taxa de gravação de eventos em mídias permanentes não passe de 100Hz. A seleção final é, então, feita offline.

4 Algoritmo de Extração de Características

O algoritmo de referência no segundo nível, com base em calorimetria, é o algoritmo T2Calo (ATLAS HLT/DAQ/DCS Group, 2000). Observando as RoIs de elétrons e jatos típicos do segundo nível de trigger (Figura 2), nota-se que as sutis diferenças entre os dois padrões de deposição de energia se encontram no espalhamento da distribuição de energia e nas relações dos picos energéticos. São essas características que o T2Calo procura extrair através de suas variáveis discriminantes.

As variáveis do T2Calo são calculadas através de janelas no plano η e ϕ , ao redor da célula mais energética, do centro do feixe. Posteriormente, elas são apresentadas a um Algoritmo de Hipótese que, através de cortes lineares de energia baseados em informação especializada, irá marcar o evento como elétron ou jato.

5 Anelador Neural - Neuralringer

Assim como o Algoritmo do T2Calo extrai as características de deposição de energia das RoIs através de janelas (cluster) de células, numa área em η e ϕ , a proposta do neuralringer é fazer o mapeamento das RoIs através de anéis e utilizar

redes neurais artificiais como Algoritmo de Hipótese (Anjos, 2003).

O mapeamento em anéis é feito de forma abstrata, uma vez que a granularidade de certas camadas do calorímetro impede a visualização de anéis. A Figura 3 mostra um conjunto de anéis para algumas camadas do calorímetro. Uma vez encontrada a célula mais energética da camada, os anéis são criados de forma concêntrica. Esse tipo de mapeamento também tenta extrair as características de espalhamento e relação de energia entre as camadas do calorímetro. Eventualmente, podem ser criados anéis incompletos.

A normalização utilizada foi a do tipo *seqüencial*. Para cada camada, calcula-se a energia total depositada e divide-se o valor da energia do primeiro anel por essa energia. A energia do segundo anel é dividida pelo valor desta energia do primeiro anel, e assim sucessivamente até que o número de anéis se esgote.

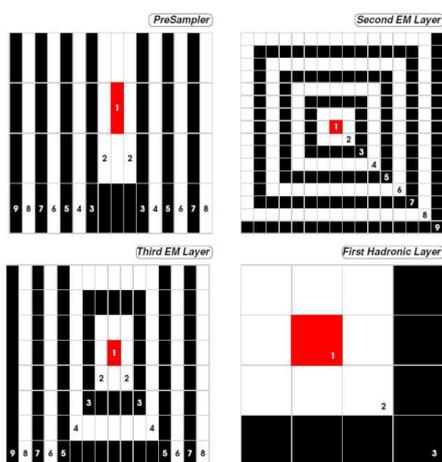


Figura 3: Conjunto de anéis para algumas camadas do calorímetro.

Um problema ocorrerá: os valores energéticos nos últimos anéis de cada camada, aqueles nos quais a relação sinal ruído é a menor, terão o menor fator de normalização e, portanto, a maior amplificação. Isto é prejudicial ao desempenho de qualquer discriminador, uma vez que beneficiará canais ruidosos. A solução é limitar a variação do fator de normalização por camada. A Tabela 1 exemplifica a execução do algoritmo.

Anel	Normalização	Soma em Anel
1	$E(\text{total da camada})$	E_1
2	$E - E_1$	E_2
...
N	$E - E_1 - \dots - E_{N-1}$	E_N

Tabela 1: Normalização Sequencial.

Como pode-se especificar o tamanho em η e ϕ dos anéis de cada camada, de forma a se adaptar à diferente granularidade de cada uma, a Tabela 2

mostra a quantidade de anéis utilizada para cada camada do calorímetro, totalizando 100 anéis.

Camadas Eletromagnéticas			Camadas Hadrônicas			
Pre-sampler	1 ^a	2 ^a	3 ^a	1 ^a	2 ^a	3 ^a
8	64	8	8	4	4	4

Tabela 2: Quantidade de anéis por camada do calorímetro.

Os dados de simulação que foram utilizados nesse artigo formam um conjunto contendo, aproximadamente, 2400 RoIs de elétrons e 800 RoIs de jatos. Os conjuntos de RoIs foi dividido em 2 grupos, treino e teste, sendo o grupo de teste também utilizado como grupo de validação do treinamento. Para lidar com a elevada diferença no número de eventos de elétrons e jatos, este último foi replicado no conjunto de treino, evitando-se uma eventual tendência no treinamento. O algoritmo de treinamento supervisionado utilizado para o treinamento da rede foi o Resilient Backpropagation (Riedmiller and Braun, 1993).

A figura de mérito escolhida para a discriminação entre elétrons e jatos foi o produto SP, que relaciona as probabilidades de detecção de elétrons e jatos de acordo com a fórmula:

$$SP = (P_e + P_j) * (P_e * P_j)$$

Foram treinadas redes com diversas topologias (1 a 15 neurônios na camada escondida) e com 5 inicializações para cada tentativa de treinamento.

Será utilizada uma função tangente hiperbólica como a função de ativação dos neurônios, seja na camada escondida, seja na camada de saída. As variáveis de entrada são esferizadas (média nula e variância unitária) de acordo com o conjunto de treinamento.

6 Compactação da informação

Afim de diminuir a dimensão do espaço de entrada da rede neural, de forma a aumentar a velocidade de processamento e, eventualmente, aumentar também a eficiência do classificador, aplicou-se um mapeamento de relevância e a Análise de Componentes Principais.

6.1 Relevância

O estudo de relevância da entrada mostra quais são as variáveis mais importantes à discriminação. As outras podem, então, ser descartadas do processamento. Para avaliar a relevância de cada componente, podemos proceder de duas formas: pela variação média quadrática da resposta da rede ou pela variação do SP. Assim, temos:

$$R_{MSE_{x_i}} = \frac{1}{N} \sum_{j=1}^N [O(\vec{x}_j) - O(\vec{x}_j |_{x_{j,i}=\bar{x}_i})]^2$$

$$R_{SP_{x_i}} = SP(\vec{x}_j) - SP(\vec{x}_j |_{x_{j,i}=\bar{x}_i})$$

Onde R_{x_i} é a relevância da i -ésima componente, $O(\vec{x})$ é a resposta da rede neural ao estímulo \vec{x} e \bar{x}_i representa o valor médio da componente \vec{x} . As variáveis entre uma faixa de valores entre a variável mais relevante e uma ordem de grandeza abaixo serão ditas relevantes.

A Figura 4 mostra a relevância dos 100 anéis calculadas por MSE e SP. A linha vertical destacada em cada figura é o patamar de relevância, onde anéis com relevância maior que este valor são considerados relevantes. Já as linhas horizontais destacadas marcam a passagem de uma camada do calorímetro para a outra, de forma crescente (pre-sampler e.m., 1ª camada e.m....até a 3ª camada hadrônica). Assim, o número de anéis selecionados seria: MSE treino, 15 anéis; MSE teste, 19 anéis; SP treino, 23 anéis; SP teste, 35 anéis. Pode-se notar, pelo estudo de relevância por SP, que alguns anéis confundem a rede (anéis cuja relevância é negativa), uma vez que o produto SP aumenta quando eles são substituídos pela sua média. Um estudo futuro é a retirada de somente esses anéis.

6.2 PCA - Análise de Componentes Principais

A análise de componentes principais é uma técnica estatística que tenta decorrelacionar a informação contida numa base de dados através de uma mudança de base (JOLLIFFE, I.T., 2002). As novas componentes que formam esta base são chamadas de componentes principais e são ordenadas de acordo com sua energia (variância). Dependendo do valor de sua energia, podemos simplesmente descartar uma componente minoritária, acarretando uma perda controlada da informação.

Nesta aplicação, pode-se utilizar uma extração de componentes principais segmentada. Como o calorímetro é dividido em camadas de deposição de energia, as componentes principais podem, então, ser extraídas por camadas. A seleção de componentes pode se dar pela curva de energia ou pela relevância de uma componente para o discriminador neural.

A Figura 5 mostra a curva de carga para as componentes extraídas de forma não segmentada. O eixo vertical mostra o percentual de energia acumulado conforme a energia de cada componente é somada. Pode-se ver que com 40 componentes temos 95% de energia acumulada. Já a Figura 6 mostra a relevância por SP das componentes principais (a relevância por MSE mostrou resultado

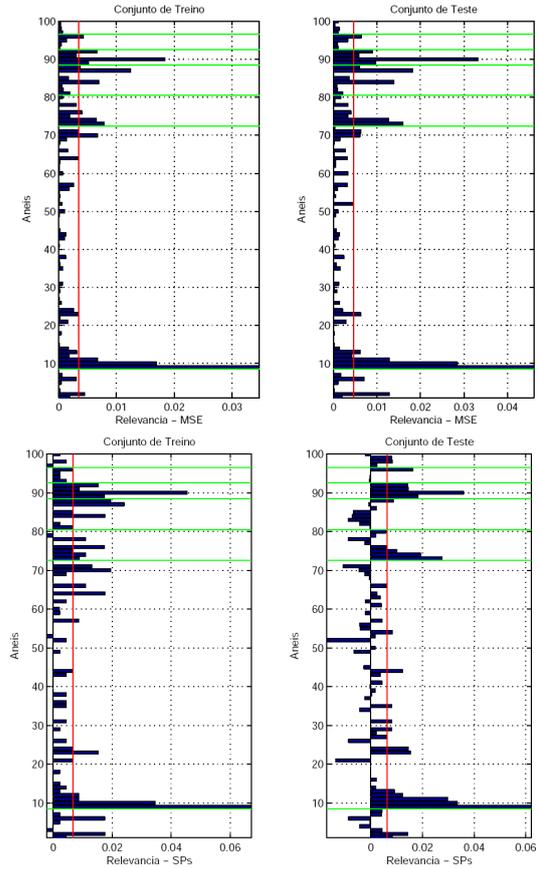


Figura 4: Relevância para os 100 anéis.

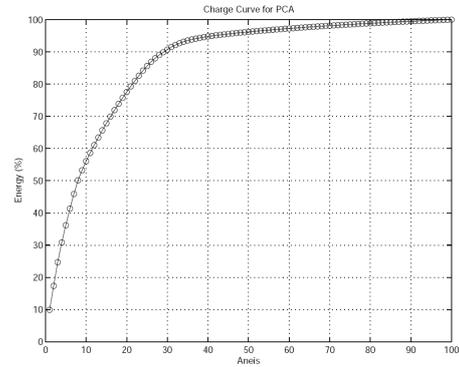


Figura 5: Curva de carga para os 100 anéis, de forma não segmentada.

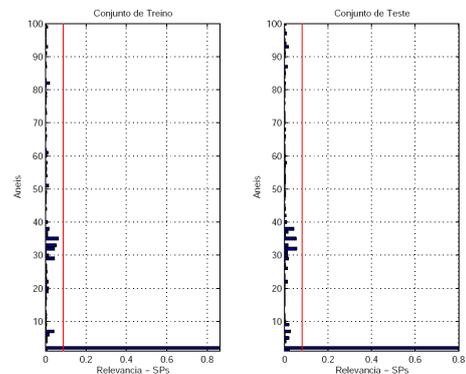


Figura 6: Relevância por SP para PCA extraída de forma não segmentada.

semelhante). Note-se que somente uma componente é considerada relevante.

Se olharmos para a distribuição dessa componente mais relevante (Figura 7), podemos ver que ela é bastante seletiva. A Figura 7 mostra também o auto-vetor relacionado à segunda componente principal, a mais relevante. Pode-se ver que ela possui forte representação nos anéis das duas seções do calorímetro.

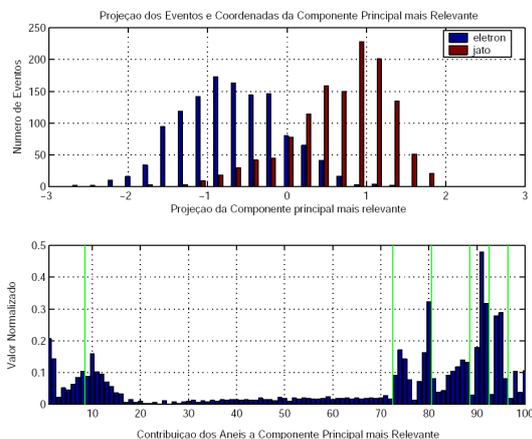


Figura 7: Distribuição e Coordenadas da componentes principal mais relevante.

Já para as componentes principais extraídas de forma segmentada, as curvas de carga não mostraram uma compactação muito elevada. Das 100 componentes, a seleção por energia, levando em conta entre 90% e 95% de energia acumulada para cada camada, mantém 71 componentes. Entretanto, a relevância por MSE e SP dessas componentes principais (Figura 8) mostrou que apenas 10 a 15 componentes podem ser mantidas. Note-se que a maior parte das componentes mais energética, para cada camada, foram consideradas relevantes.

7 Resultados Neurais

Feitas todas as análises quanto à compactação, diversas redes neurais foram treinadas. As redes são comparadas através da curva ROC (Receiver Operating Characteristics) de cada uma e pelo produto SP alcançado. Conforme andamos na curva ROC, o produto SP é variado. Figura 9 mostra as curvas ROC das redes treinadas cortando-se o número de anéis por relevância. Pode-se ver que se fixarmos um nível de 10% de falso alarme, todas as redes mostram resultados semelhantes, com 95% de eficiência de detecção. Porém, se fixarmos em 5% o falso alarme, os cortes por relevância MSE apresentam uma eficiência de detecção levemente maior àqueles adotados por relevância SP, mas a performance é cerca de 10 pontos percentuais melhor do que alcançado com o uso da informação completa (100 anéis).

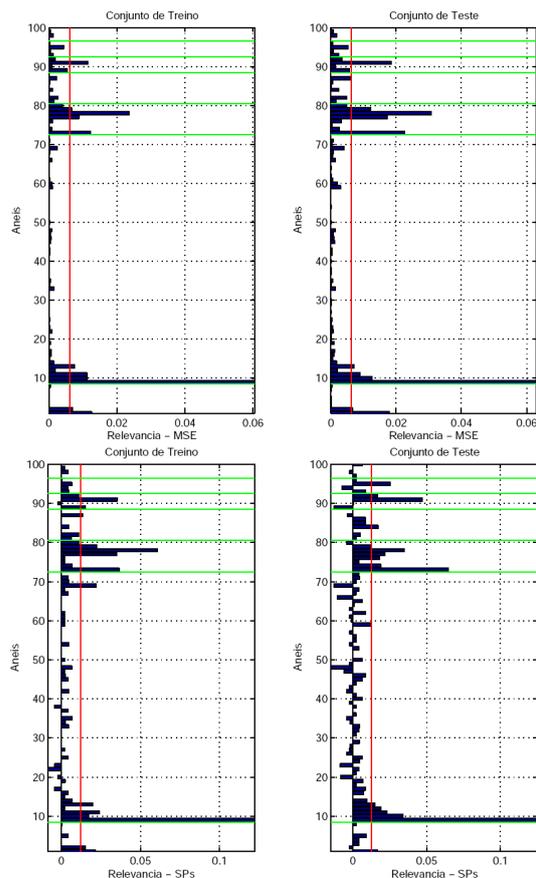


Figura 8: Relevância das componentes principais extraídas de forma segmentada.

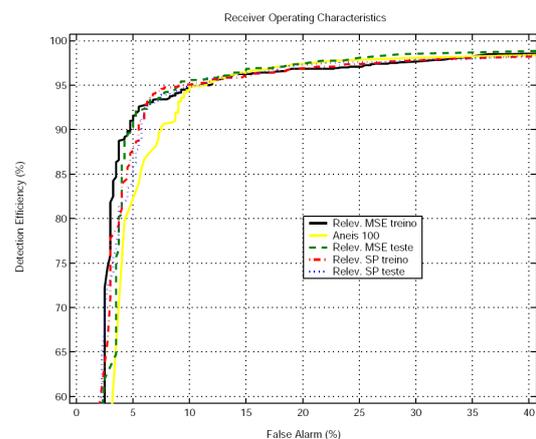


Figura 9: Curva ROC dos discriminadores com anéis cortados por relevância.

Para as análises com PCA segmentada e não segmentada, as respectivas ROC se encontram na Figura 10. As redes adotando cortes por relevância do conjunto de treino e teste, por MSE e Sp, tiveram desempenho similares e estão representadas somente pela ROC referente ao conjunto de treino. Pode-se ver que a componente mais relevante da PCA não segmentada consegue resultados inferiores, porém significativos por se tratar de apenas uma componente. Se fixarmos também um nível de falso alarme em 5%, o discriminador com corte por PCA segmentada e relevância por SP consegue uma eficiência de detecção entre 2,5 e 15 pontos percentuais melhor que os outros métodos. Já se fixarmos em 10% de falso alarme, os discriminadores têm praticamente a mesma eficiência de detecção.

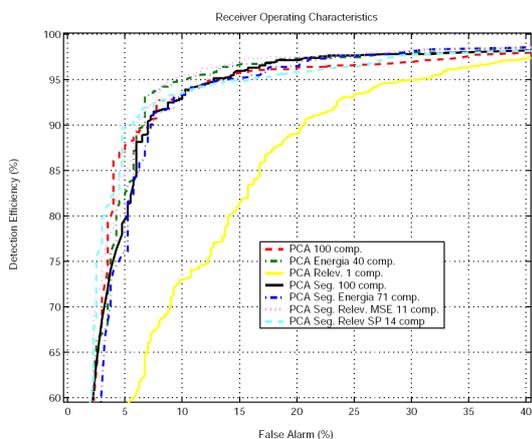


Figura 10: Curva ROC dos discriminadores com anéis cortados por PCA.

Assim, considerando o nível de 5% de falso alarme fixo e levando em conta o número de variáveis de entrada de cada análise, a Figura 11 mostra as curvas ROC para os melhores discriminadores, nas abordagens por relevância e por PCA. Já a Tabela 3 mostra o SP para um nível fixo de 5% de falso alarme.

	Anel 100	PCA Seg. Relev. SP	Relev. MSE treino	Relev. MSE teste
SP	1,410	1,577	1,623	1,612

Tabela 3: SP para os melhores discriminadores com falso alarme fixo em 5%.

8 Agradecimentos

Agradecemos ao apoio do CNPq e FAPERJ (Brasil), e CERN (Suíça) ao projeto. Também agradecemos aos nossos colegas da colaboração Trigger/DAQ por disponibilizar o conjunto de dados e pelas discussões frutíferas ao longo do desenvolvimento deste trabalho.

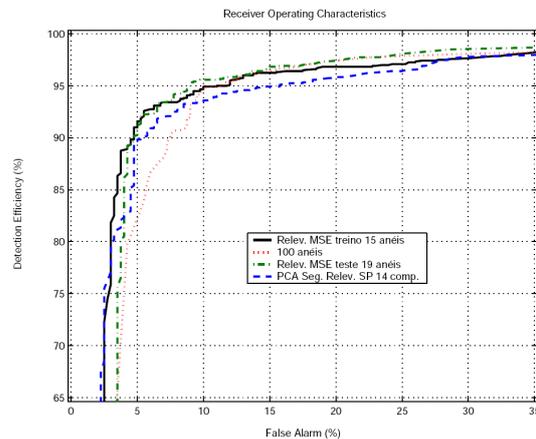


Figura 11: Curva ROC dos melhores discriminadores para as abordagens de cortes por PCA e relevância.

Referências

- Anjos, A.R.; Seixas, J. d. (2003). Neural particle discrimination for triggering interesting physics channels with calorimetry data., *Nuclear Instruments And Methods In Physics Research* pp. 713–715.
- ATLAS (2006). A toroidal lhc apparatus, <http://www.cern.ch/atlas>.
- ATLAS HLT/DAQ/DCS Group (2000). Atlas daq, ef, lvl2 and dcs technical progress report, *Technical report*, CERN.
- CERN (2006). European organization for nuclear research, <http://www.cern.ch>.
- High Level Trigger (2003). Atlas high level trigger, data acquisition and controls technical design report, *Manual*, CERN.
- JOLLIFFE, I.T. (2002). *Principal Component Analysis*, Springer-Verlag.
- LArg Collaboration (1996). Liquid argon calorimeter technical design report, *Manual*, CERN.
- LHC (2006). Large hadron collider, <http://lhc.web.cern.ch/lhc>.
- Riedmiller, M. and Braun, H. (1993). A direct adaptive method for faster backpropagation learning: The rprop algorithm, *Proceedings of the IEEE International Conference on Neural Networks*.
- TileCal Collaboration (1996). Tile calorimeter technical design report, *Technical report*, CERN.