REDUÇÃO DE DIMENSIONALIDADE NUM DETECTOR FINAMENTE SEGMENTADO UTILIZANDO A ANÁLISE DE COMPONENTES PRINCIPAIS NÃO-LINEARES

Danilo L. de Souza*, José Manoel Seixas*

*Laboratório de Processamento de Sinais, COPPE/POLI
Universidade Federal do Rio de Janeiro CP 68504, CEP 21941-972
Rio de Janeiro, RJ, Brasil

Emails: danilo@lps.ufrj.br, seixas@lps.ufrj.br

Abstract— ATLAS is one of the main detectors of the next generation collider experiment, LHC. The high amount of information produced per collision requires a complex triggering system, which will perform on-line selection of interesting events for the physics under investigation. This work concentrates on data compaction and classification for the calorimeter system in ATLAS. A monitoring system has been implemented, in the second-level trigger, for on-line monitoring of physics objects selected. Due to detector segmentation in several layers, compaction is performed using a segmented approach on regions of interest previously selected, which are classified by compact neural networks. High compaction rates has been achieved by using Nonlinear Principal Component Analysis. In the classification stage, a 96% detection efficiency for electrons has been achieved, for a false alarm probability below 4%.

Keywords— NLPCA, ATLAS, Neural Networks, Detectors.

Resumo— O ATLAS é um dos principais detectores presentes na próxima geração de colisionadores de partículas. A grande quantidade de informação produzida por colisão exige um complexo sistema de filtragem, que irá selecionar *on-line* os eventos mais interessantes para a física investigada. Este trabalho se concentra na compactação e classificação dos dados de calorimetria do ATLAS. Uma proposta de monitoramento *on-line* de sinais é apresentada para o segundo nível de filtragem. Devido a segmentação do detector em diversas camadas a compactação é realizada uma abordagem, também, segmentada das regiões de interesse previamente selecionadas, que são classificadas por redes neurais compactas. Altas taxas de compactação foram obtidas através da Análise de Componentes Principais Não-Lineares. Na etapa de classificação uma eficiência na detecção de elétrons de 96% foi alcançada, para uma probabilidade de falso alarme abaixo de 4%.

Palavras-chave— NLPCA, ATLAS, Redes Neurais, Detectores.

1 Introdução

O LHC (Large Hadron Collider), experimento do CERN (European Laboratory for Particle Physics), colisionará prótons com elevada energia, através de um colisionador circular com $26,7$km de circunferência. Cada colisão constitui um evento a ser analisados através de detectores acoplados ao LHC. O ATLAS (A Toroidal LHC ApparatuS) é um dos principais detectores do LHC (Collaboration, 1999a; Collaboration, 1999b).

Uma grande quantidade de dados é gerada a cada colisão no LHC. Por outro lado, a física de interesse ocorre com uma freqüência muito menor do que a taxa total de geração de eventos. Como cada evento carrega aproximadamente $1,5$ MByte de dados, e espera-se uma taxa de 40 MHz de eventos sendo gerados, o fluxo de dados será da ordem de 60 TBytes por segundo, impossibilitando o armazenamento completo desses eventos para análise *off-line*. Desta maneira, um sistema de filtragem *online* torna-se indispensável para o experimento. Este sistema é constituído de três níveis sequênciais de seleção de eventos, cada nível refina a decisão realizada pelo nível anterior, onde necessariamente aplica critérios adicionais de seleção. Todo esse processamento requer uma excelente eficiência para preservar a raridade da física de interesse, como por exemplo o bóson de Higgs (Collaboration, 1999a; Collaboration, 1999b).

As técnicas de análise de componentes, sejam principais (lineares ou não) ou independentes, são abordagens que vêm sendo usadas com sucesso na extração de características, compactação e visualização de dados em diferentes áreas de aplicação. O que é essencial para a aplicação dessas técnicas é que os dados sejam mutuamente correlacioandos e que haja, então, um certo grau de redundância das amostras.

A Análise de Componentes Principais (*Principal Component Analysis*, PCA), tem sido extensivamente utilizada na compactação de sinais e representação do mesmo a partir de componentes de maior energia. Estes componentes principais correspondem aos autovetores da matriz de correlação relativa ao conjunto de dados (Hyvärinen et al., 2001).

A Análise de Componentes Principais Não-Lineares (*Nonlinear Principal Component Analysis*, NLPCA) é utilizada para identificar e remover correlações assim como a PCA. Enquanto esta identifica somente correlação linear entre as variáveis, a NLPCA abrange tanto a correlação linear quanto a correlação não-linear, sem restrição quanto ao tipo de não-linearidade presente (Hyvärinen et al., 2001).

Este trabalho tem como objetivo realizar uma redução da alta dimensionalidade dos dados entrada do sistema de calorimetria (medição de energia) do ATLAS através de NLPCA, usando a informação assim compactada para alimentar uma rede neural discriminadora que objetiva identificar a rara física de interesse que se encontra imersa no imenso ruído de fundo da experiência.

2 ATLAS

O ATLAS é composto por diferentes sistemas de detecção:

- Câmera de Arrasto - primeiro detector, localizado na região mais interna e altamente segmentado. A informação oferecida é uma imagem das trajetórias descritas pelas partículas.
- Calorímetros Hadrônico e Eletromagnético - são capazes de medir a energia depositada pela partícula incidente. Permite a identificação de partículas.
- Câmara de Múons - ´e a parte mais afastada do centro e tem como objetivo a detecção de múons.

O sistema de filtragem do ATLAS é concebido em três níveis (figura 1), cada um com suas características definidas. Parte-se de uma taxa de ocorrência de eventos de 40MHz que, com a multiplicidade de eventos por colisão em alta luminosidade pode chegar a 1GHz, para uma taxa de gravação de eventos em mídia permanente (*off-line*) de aproximadamente 100Hz.

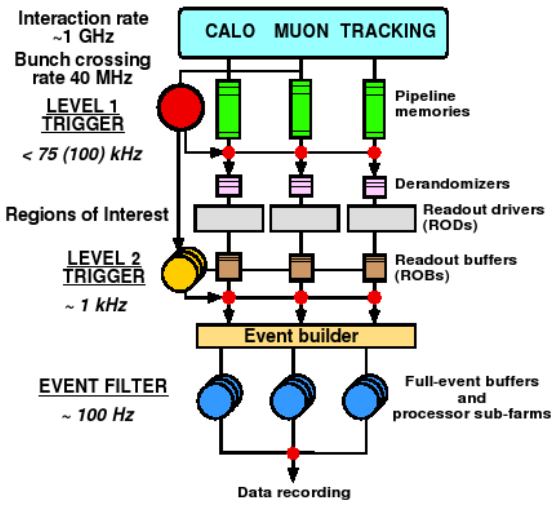


Figura 1: Sistema de filtragem do ATLAS. Extraído de (Collaboration, 1999a).

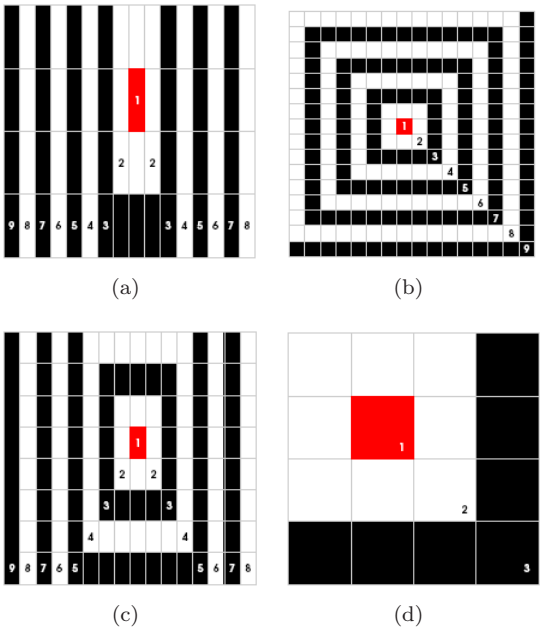
O primeiro nível de *trigger* do ATLAS realiza a seleção inicial dos eventos que é, baseada na informação obtida com granularidade menos fina de um sub-conjunto dos seus detectores (calorimetria e detecção rápida de múons). Este nível, então, descarta eventos com características bem distintas dos canais de interesse desejados. A taxa máxima de saída do primeiro nível de filtragem está limitada em 75kHz (podendo ser expandida para 100kHz) (Collaboration, 1999a).

De forma a reduzir o volume de tráfego de dados para o segundo nível, e, conseqüentemente, aumentar a banda passante entre estes dois níveis, o primeiro nível já marca as regiões de interesse (Region of Interest - RoIs), que correspondem às regiões do detector onde houve efetivamente a incidência de algum evento, de tal maneira que o segundo nível observará somente estas regiões, e não toda a área do detector numa tomada de decisão.

O segundo nível de *trigger* usa as RoIs marcadas para refinar a decisão, validando a etiquetagem (resultado da classificação do primeiro nível) de cada RoI através da utilização da informação completa (granularidade fina). Para o segundo nível, então, o tempo de latência é de aproximadamente 10ms, onde a taxa de eventos que devem trafegar deste para o terceiro nível é da ordem de 1kHz (Group, 2003).

O processamento das células do calorímetro com toda a granularidade disponível é inviável, pois a dimensionalidade do espaço de entrada torna proibitivo qualquer método direto de disriminação, considerando-se o tempo necessário de processamento para a operação *on-line*. Uma abordagem adotada para contornar esta dificuldade foi a utilização da formatação dos dados de entrada em soma de anéis concêntricos (dos Anjos et al., 2006). A formatação em anéis aproveita-se do isotropismo na deposição de energia por parte dos objetos que interagem com o calorímetro. A tendência é que as partículas desenvolvam suas cascastas no sentido externo dos calorímetros, com raio crescente, de forma semelhante a um cone. Os anéis de cada camada são extraídos após a identificação da célula com maior deposição de energia. A partir deste primeiro anel procura-se, então, as células vizinhas (concêntricas a este ponto) que a circundam, formando o segundo anel. Procegue-se este algoritmo, buscando o anel seguinte, até que se esgotem todas as células na camada. A figura 2 ilustra a formação dos aneis de algumas camadas dos calorímetros do ATLAS. Além disso, os dados em anéis, traçados a partir da célula de maior energia por camada do calorímetro, são normalizados de modo a evidenciar a participação das células menos energéticas no processamento de discriminação baseado na calorimetria (dos Anjos et al., 2006).

Já que os calorímetros são divididos em camadas com granularidade variante (figura 2), é natural propor que cada camada tenha seu conjunto de anéis. Assim, pode-se explorar eficientemente a informação detalhada oferecida pelos caloríme-



(a) (b) (c) (d)

Figura 2: Formatação em anéis: (a) *Pre-Sampler*; (b) *EM middle layer*; (c) *EM back layer*; (d) *HAD layer* 0.

Tabela 1: Anéis distribuídos pelas camadas do calorímetro do ATLAS.

Camada	# de Anéis
Pre-sampler	8
EM front layer	64
EM middle layer	8
EM back layer	8
HAD layer 0	4
HAD layer 1	4
HAD layer 2	4

tros do ATLAS (dos Anjos et al., 2006). Das 7 camadas do calorímetro, obtém-se o total de 100 anéis, conforme a tabela 1.

3 Análise de Componentes Principais

É vasta a literatura que mostra a PCA como uma solução ótima para diferentes problemas de representação da informação (Karhunen and Joutsensalo, 1995; Hyvärinen et al., 2001). Isso inclui:

- Maximização das variâncias linearmente transformadas.
- Minimização do erro médio quadrático de reconstrução, quando os dados de entrada são aproximados usando um sub-espaço linear dimensionalmente menor.
- Descorrelação das saídas.
- Minimização da representação da entropia.

Dado um conjunto de dados multivariados o objetivo é encontrar um conjunto menor de variáveis, com menos redundância, e que forneça uma representação tão boa quanto possível. A PCA usa uma base de autovetores para descorrelacionar linearmente a informação contida na amostra de dados estudada. Usando somente a correlação linear, a PCA baseia-se em estatística de segunda ordem, contudo, problemas reais podem envolver relações mais complexas (Hyvärinen et al., 2001).

Em muitos casos as variáveis que compõem o vetor de informação de entrada apresentam correlação não-linear, de tal modo que uma análise de componentes principais não-linear pode ser mais eficiente do que uma técnica linear padrão. A NLPCA é análoga à PCA, do ponto de vista da minimização do erro quadrático, acrescida de mapeamentos não-lineares entre os espaços de dimensão original $\mathbb{R}^l$ e de dimensão reduzida $\mathbb{R}^f$. Dessa forma, se houver relação não-linear entre as variáveis, a NLPCA descreverá os dados com maior acurácia e/ou com um número menor de componentes que a PCA, a menos que as restrições estatísticas não permitam a formulação de funções mais complexas (Kramer, 1991).

Outra motivação para usar não-linearidades é que elas introduzem de uma maneira implícita estatística de ordem elevada na computação. Estatística de ordem elevada ou definida por cumulantes iguais ou maiores que segundos momentos, são necessários para uma boa caracterização de dados não gaussianos, o que compreende muitos casos reais (Karhunen and Joutsensalo, 1995; Kramer, 1991).

Em NLPCA, o mapeamento no espaço de componentes é generalizado para permitir não-linearidades arbitrárias

$$\mathbf{y} = \mathbf{g}(\mathbf{x}) \tag{1}$$

onde $\mathbf{g}$ é um vetor de funções não-lineares, isto é, $\mathbf{g} = \{g_1, g_2, \ldots, g_f\}$, sendo g_i a i-ésima componente não-linear e $\mathbf{y}_i$ está representando o i-ésimo elemento de $\mathbf{y}$.

A transformação inversa, restaurando a dimensão original dos dados é implementada por um segundo vetor de funções não-lineares, $\mathbf{h} = \{h_1, h_2, \ldots, h_l\}$.

$$\hat{\mathbf{x}} = \mathbf{h}(\mathbf{y}) \tag{2}$$

A perda de informação, ou erro, é medida por

$$\mathbf{E} = \mathbf{x} - \hat{\mathbf{x}} \tag{3}$$

e as funções $\mathbf{g}$ e $\mathbf{h}$ são selecionadas para minimizar $\|\mathbf{E}\|$, como na PCA.

A abordagem proposta por Kramer (1991) utiliza uma rede neural para obtenção de componentes principais não-lineares. As funções $\mathbf{g}$ e $\mathbf{h}$ são modeladas por meio de redes neurais de três camadas. A rede $\mathbf{g}$ contém uma camada escondida denominada camada de codificação com M_1 nós com função de ativação sigmoidal (geralmente

$M_1 > l$). A rede **h** contém uma camada escondida denominada camada de decodificação com M_2 nós com função de ativação sigmoidal (geralmente $M_2 > l$). É comum $M_1 = M_2$.

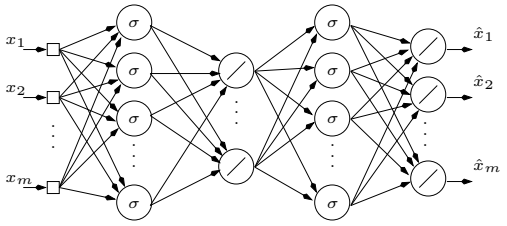


Figura 3: Rede neural autoassociativa.

As redes para gerar **g** e **h** requerem treinamento supervisionado. No entanto, ao treinar a rede representando **g**, a saída **Y** desejada é desconhecida. Para a rede representando **h**, as saídas desejadas são conhecidas (a entrada **X** da rede **g**) mas as entradas correspondentes, **Y**, não são. Como a saída da rede **g** e a entrada da rede **h** são, ambas, **Y**, as duas redes podem ser combinadas formando, agora, uma única rede de 5 camadas (figura 3).

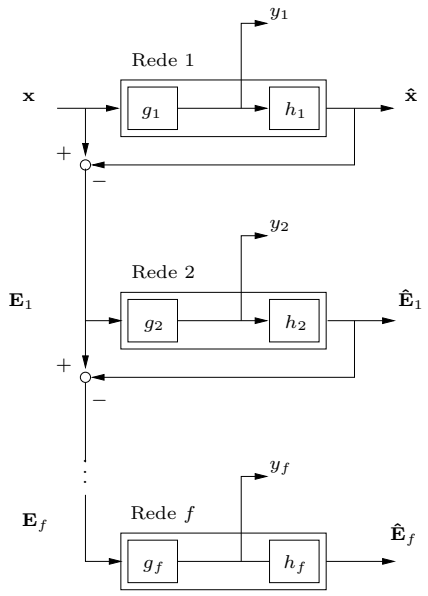


Figura 4: f redes autoassociativas.

A rede combinada contém três camadas escondidas: a camada de codificação, envolvida em representar **g**, a camada intermediária, cujas saídas representam **Y**, além da camada de decodificação, envolvida em modelar **h**. A camada intermediária é também chamada de camada gargalo, pois é a camada com o menor número de nós. As camadas de entrada e saída representam **X** e $\hat{\mathbf{X}}$ respectivamente. Os nós das camada de codificação e de decodificação são não-lineares enquanto que os nós das camadas gargalo e de saída são tanto lineares quanto não-lineares (desde que a função de ativação não-linear usada na saída abranja a faixa dinâmica do sinal de entrada) (Kramer, 1991).

A figura 4 mostra um esquema de f redes autoassociativas utilizadas na extração de f componentes principais.

4 Base de Dados e Figura de Mérito

A base de dados disponível conta com 30090 observações, sendo 22581 elétrons e 7509 jatos. Os dados foram divididos em dois grupos: treinamento e teste. Para o treinamento foram escolhidos aleatoriamente 11298 elétrons e 3774 jatos, restando 11283 elétrons e 3735 jatos para a etapa de teste.

A comparação dos resultados de classificação se dará através da curva de eficiência de detecção de elétrons versus falso alarme (jatos rotulados como elétrons), conhecida como ROC (*Receiver Operating Characteristic*) e do produto SP. Considerando-se o conjunto de teste, o produto SP é definido como o produto da soma das eficiências pelo produto destas, ou seja:

$$SP = (\mathrm{Ef}_e + \mathrm{Ef}_j) \cdot (\mathrm{Ef}_e \cdot \mathrm{Ef}_j) \qquad (4)$$

onde Ef_e é a eficiência de detecção de elétrons e Ef_j é a eficiência de detecção de jatos. O produto SP restringe a existência de um desiquilíbrio entre as eficiências das duas classes, uma vez que uma eficiência reduzida em uma das classes implica em valores baixos para o produto SP (dos Anjos et al., 2006). O seu valor máximo é 2, para uma classificação binária.

Os valores apresentados por Seixas and Júnior (2006) serão utilizados como figuras de mérito, já que representam o estado da arte. Uma partição binária dos dados de teste foi realizada através de um classificador neural com topologia: $100 - 5 - 1$, ou seja operando sobre os dados dos 100 anéis. Um valor final de produto SP igual a $1,80$ para uma eficiência de $97,28\%$ e falso alarme de $3,96\%$ foi obtido. Numa aplicação da análise de componentes principais lineares nos dados das diferentes camadas (com retenção de 95% da energia, tabela 2) o desempenho final do classificador (melhor topologia: $85 - 15 - 1$) em função do produto SP ficou em $1,83$. A eficiência na detecção de elétrons alcançou $97,05\%$, para um falso alarme de $2,78\%$.

Tabela 2: MSE de reconstrução em função do número de componentes principais retidas por camada de cada calorímetro.

Camada	Comp.	MSE
Pre-sampler	8	0
EM front layer	52	0,00042
EM middle layer	6	0,003
EM back layer	7	0,0336
HAD layer 0	4	0
HAD layer 1	4	0
HAD layer 2	4	0

5 Compactação Segmentada

A redução de dimensionalidade dos dados de entrada, na abordagem segmentada, considera a divisão entre as diferentes camadas dos calorímetros como diferentes processos aleatórios.

A figura 5 mostra o MSE de reconstrução em fução do número de neurônios na segunda e quarta camadas da rede autoassociativa, para 5 diferentes inicializões do algoritmo, na extração das componentes principais não-lineares da camada *Pre-Sampler* do calorímetro. O erro de reconstrução não parece ser uma função apenas do número de neurônios alocados nessas duas camadas, tendo em vista que o menor MSE alcançado foi obtido com um número intermediário entre o 2 e 9 neurônios.

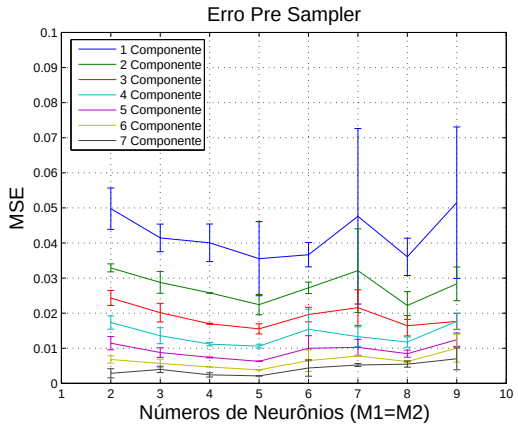


Figura 5: MSE de reconstrução em função de M_1 e M_2 para a camada *Pre-Sampler*.

Para definir a topologia da rede extratora usou-se da parcimônia na escolha dos parâmetros para cada um dos algoritmos. Os parâmetros M_1 e M_2 foram mantidos constantes ($M_1 = M_2$). Após exaustivos testes chegou-se a configuração do classificador com topologia igual a $33-5-1$. O MSE de reconstrução para cada uma das camadas é mostrado na tabela 3. Os dados projetados na dimensão reduzida, foram extraídos das redes que apresentaram os menores valores finais de MSE após as 5 inicializações aleatórias.

Tabela 3: MSE de reconstrução em função do número de componentes principais não-lineares.

Camada	$M_1 = M_2$	f	MSE
Pre-Sampler	5	6	0,0049
EM front layer	17	15	0,0054
EM middle layer	7	3	0,0093
EM back layer	6	3	0,0116
HAD layer 0	4	2	0,0063
HAD layer 1	4	2	0,0061
HAD layer 2	4	2	0,0039

A classificação para o conjunto de teste, na nova dimensão reduzida, apresentou um valor SP máximo igual a $1,782$, isso representa uma eficiência de $96,22\%$ para elétrons contra apenas $3,75\%$ de falso alarme. A figura 6 mostra a curva ROC e o ponto de produto SP máximo. Mesmo o SP máximo apresentando um valor ligeiramente inferior aos das figuras de mérito, a eficiência de detecção diminuiu menos de 1% em relação a cada uma, e a taxa de falso positivo ficou abaixo da apresentada pela figura de mérito que operava sobre os 100 anéis e menos de 1% superior a apresentada pela figura de mérito em que foi aplicada a análise de componentes principais lineares, tudo isso com uma redução de dimensionalidade de 67% e 38% em relação a cada uma das figuras de mérito, respectivamente. Com relação ao classificador operando sobre as componentes principais lineares, nota-se que a NLPCA produziu uma acentuada redução de dimensionalidade, o que favorece a operação *on-line* desejada, considerando-se a latência máxima de 10ms do segundo nível de filtragem.

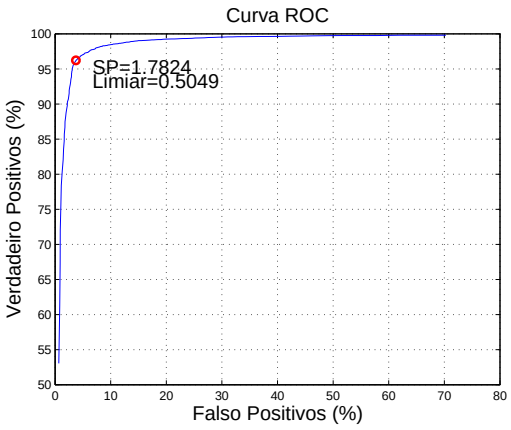


Figura 6: Curva ROC e Produto SP da abordagem NLPCA segmentada.

Os resultados apresentados são positivos, principalmente se levarmos em consideração que o classificador da figura de mérito que operava sobre as componentes principais lineares possui uma topologia maior, mais neurônios na camada escondida.

6 Análise de Relevância

Esta análise pode ser vista como um ajuste fino sobre o espaço amostral para a discriminação. Ela exprime, de forma quantitativa, a importância de cada variável de entrada da rede neural para a classificação. O procedimento para o cálculo da relevância consiste em substituir cada variável de entrada do classificador por sua média no conjunto de treino, e calcular o desvio médio na saída da rede. A relevância da variável i é igual à média das variações medidas na saída rede, obtidas quando substitui-se a variável pela sua média no processo, como mostra a equação (5) (Seixas et al., 1996).

$$R_i = \frac{1}{N}\sum_{j=1}^{N}\left(S\left(\mathbf{x}_j\right) - S\left(\mathbf{x}_j|x_{ji} = \bar{x}_i\right)\right)^2 \qquad (5)$$

onde $S(\cdot)$ é a saída do classificador, e N o número de amostras. Quanto maior o valor obtido para R_i, mais relevante é a variável.

Partindo do classificador neural implementado na abordagem NLPCA segmentada, que extraiu dos dados originais 33 componentes, a equação (5) foi aplicada. A relevância R_i da maioria das variáveis esteve dentro de um intervalo muito próximo. A tabela 4 mostra esse comparativo em função do percentual de relevância, $100 \cdot R_i/R$ onde $R = \sum R_i$.

Tabela 4: MSE de reconstrução em função do número de componentes principais não-lineares.

Camada	% de R
1 Comp. *EM front layer*	15,04
2 Comp. *EM front layer*	11,08
1 Comp. *EM back layer*	15,01
1 Comp. *HAD layer* 0	17,3
1 Comp. *HAD layer* 1	13,05
1 Comp. *HAD layer* 2	16,22

Esses valores indicam que a informação mais relevante de cada um dos processos estudados (camadas do calorímetro) nessa abordagem se encontra nas primeiras componentes principais não-lineares. Além disso, pode-se criar relações entre a informação contida em cada camada do calorimetro, assim como suas propriedades de interação com a física de interesse e que baseia trabalhos teóricos nessa área, com quanto o classificador faz uso disso na sua tomada de decisão.

7 Conclusão e Trabalhos Futuros

A aplicação da NLPCA obteve um poder de compactação bastante expressivo, independente de qualquer informação a priori sobre o problema tratado. O seu uso em cada uma das camadas, mostrou uma redução de dimensionalidade de aproximadamente 70%.

A análise de relevância mostrou uma boa interpretação das componentes principais não-lineares de cada camada na abordagem segmentada e sua relação com o problema de classificação analisado.

Para trabalhos futuros, uma análise de relevância mais detalhada sobre as componentes extraídas com a técnica não-linear, após o processamento dos dados para cada camada dos calorímetros, pode comparar diretamente as componentes NLPCA com as componentes de discriminação e buscar interpretações físicas para as componentes extraídas.

Pretende-se ainda avaliar uma implementação desta técnica no ambiente de processamento distribuído (2000 processadores duais) do sistema de segundo nível de *trigger* e a extração NLPCA a partir de algoritmos que implementam a análise de componentes independentes (ICA).

8 Agradecimentos

Agradecemos ao apoio financeiro do CNPQ e FAPERJ (Brasil), e CERN (Suiça). Agradeçemos, também, aos colegas da colaboração pela base de dados.

Referências

Collaboration, A. (1999a). *Atlas Detector and Physics Performance, Technical Design Report I*, CERN/LHCC 99-14.

Collaboration, A. (1999b). *Atlas Detector and Physics Performance, Technical Design Report II*, CERNLHCC 99-15.

dos Anjos, A., Torres, R., Seixas, J., Ferreira, B. and Xavier, T. (2006). Neural triggering system operating on high resolution calorimetry information, *Nuclear Instruments and Methods in Physics Research* **559**: 134–138.

Group, A. (2003). *ATLAS High-Level Trigger, Data Acquisition and Controls, Technical Design Report*, ATLAS TDR-016.

Hyvärinen, A., Karhunen, J. and Oja, E. (2001). *Independente Component Analysis*, John Wiley & Sons, Inc.

Karhunen, J. and Joutsensalo, J. (1995). Generalizations of principal component analysis, optimization problems, and neural networks, *Neural Networks* **8**: 549–562.

Kramer, A. (1991). Nonlinear principal component analysis using autoassociative neural networks, *AIChE Journal* **37**: 233–243.

Seixas, J., Calôba, L. and Delpino, I. (1996). Relevance criteria for variables selection in classifier design, *International Conference on Engineering Applications of Neural Networks*, Londres, Inglaterra, pp. 451–454.

Seixas, J. and Júnior, H. L. (2006). A segmented principal component analysis applied to calorimetry information at atlas, *Nuclear Instruments and Methods in Physics Research* **559**: 129–133.