

COMBINAÇÃO DE CLASSIFICADORES NEURAIIS SEGMENTADOS COM PRÉ-PROCESSAMENTO POR ANÁLISE DE COMPONENTES INDEPENDENTES PARA UM SISTEMA ONLINE DE FILTRAGEM

EDUARDO F. SIMAS FILHO^{†*}, RODRIGO C. TORRES*, JOSÉ M. DE SEIXAS*, LUIZ P. CALÔBA*

*Laboratório de Processamento de Sinais, COPPE/Poli/UFRJ
Rio de Janeiro, Brasil.

[†]Instituto Federal de Educação, Ciência e Tecnologia
Campus Simões Filho, Bahia, Brasil.

Emails: esimas@lps.ufrj.br, torres@lps.ufrj.br, seixas@lps.ufrj.br, caloba@lps.ufrj.br

Abstract— This work describes the application of multiple neural classifiers to an online triggering system. In ATLAS, a detector of the LHC particle collider, the calorimeter (energy measurement) system is very important for characterizing the events generated at the collisions. ATLAS calorimeters are segmented into seven layers, each one with different characteristics. In this work it is proposed the application of expert neural classifiers dedicated to each calorimeter layer. Segmented independent component analysis is applied, as a pre-processing step, to calorimeter signals. Some classifier combination methods were used to achieve global decision by combining the segmented classifiers outputs.

Keywords— Neural Classifiers, Independent Component Analysis, Combining Classifiers.

Resumo— Neste trabalho está sendo descrita a aplicação de múltiplos classificadores neurais a um sistema online de filtragem (*trigger*). No ATLAS, um dos detectores do colisionador de partículas LHC, o sistema de calorímetros (medidor de energia) é muito importante para a caracterização dos eventos gerados nas colisões. Os calorímetros do ATLAS são segmentados em sete camadas, cada uma com diferentes características. Neste trabalho está sendo proposto o uso de classificadores neurais especialistas em cada camada do calorímetro. Análise de componentes independentes foi aplicada de modo segmentado (por camada) aos sinais dos calorímetros como pré-processamento. Alguns métodos de combinação de classificadores foram utilizados para produzir a decisão global a partir das saídas dos diversos classificadores.

Keywords— Classificadores Neurais, Análise de Componentes Independentes, Combinação de Classificadores.

1 Introdução

O ATLAS (ATLAS Collaboration, 2008) é um dos detectores do colisionador de partículas LHC (*Large Hadron Collider*), que entrou em operação em 2008 no CERN (Centro Europeu para Estudos Nucleares). O LHC (Evans and Bryant, 2008) é o maior colisionador de partículas já construído (utilizando um túnel de 27 km de circunferência a 100 m abaixo do solo) e irá explorar novas fronteiras de energia em busca de informações sobre os constituintes fundamentais da matéria.

Devido à alta frequência de colisão do LHC (40 MHz) e à alta segmentação do detector (onde cada colisão é descrita por aproximadamente 1,5 MB), é esperada para o ATLAS uma taxa de informação da ordem de 60 TB/S, sendo que grande parte desta massa de dados não é relevante para o experimento. Um sistema de filtragem online e com alta eficiência precisa ser utilizado para garantir que a maior parte do ruído de fundo da experiência seja descartado e que os eventos de interesse (que são extremamente raros, ocorrendo com frequência de aproximadamente 1Hz) não sejam perdidos.

Conforme ilustrado na Figura 1, o sistema de filtragem do ATLAS (ATLAS Collaboration, 2008)

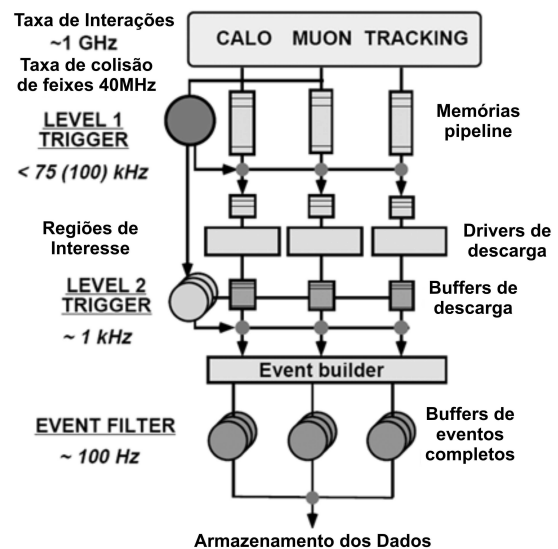


Figura 1: Diagrama em blocos do sistema de filtragem e aquisição de dados do ATLAS.

opera com três níveis de seleção de eventos, acessando informações de três sub-detectores: os calorímetros (medidores de energia), a câmara de múons e os detectores de traços (*tracking*). O primeiro nível de filtragem é implementado em

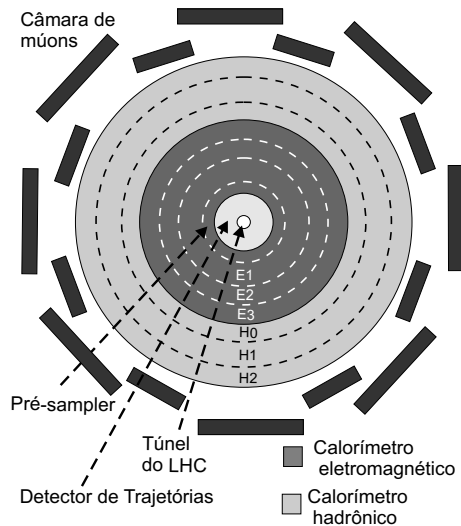


Figura 2: Sistema de calorímetros do ATLAS, composto por 7 camadas (PS, E1 a E3, H0 a H2).

hardware e opera com severas restrições de tempo de processamento (latência de $2,5 \mu s$). Para atender a esta demanda, apenas parte da informação é acessada, com forte participação da calorimetria. O segundo nível recebe do primeiro a localização das regiões onde possivelmente ocorreram eventos de interesse (conhecidas como RoI - *Regions of Interest*). Este nível é implementado em software e utiliza um conjunto de 500 computadores operando em paralelo para produzir a decisão em menos de 10 ms. O terceiro nível (*Event Filter*) também é implementado em software e dispõe de poucos segundos para produzir a decisão final de aceitação/rejeição dos eventos.

O objetivo deste trabalho é a identificação de elétrons na filtragem de segundo nível, a partir de informações do sistema de calorímetros. Conforme Figura 2, os calorímetros do ATLAS são segmentados em sete camadas, sendo quatro eletromagnéticas (PS, E1, E2 e E3) e três hadrônicas (H0, H1 e H2), cada uma com diferentes características. Os elétrons são assinaturas interessantes para diversos eventos importantes ao experimento (incluindo os decaimentos do bóson de Higgs) (ATLAS Collaboration, 2008). As assinaturas de energia de elétrons podem ser semelhantes a perfis gerados por jatos, que representam um intenso ruído de fundo para o experimento. Sendo assim, o canal elétron/jato é importante para o desempenho do detector.

Neste trabalho, é proposto um sistema de classificação que explora a segmentação da informação das diferentes camadas do calorímetro. Classificadores neurais supervisionados (Haykin, 2008) são treinados para os sinais de cada camada. A Análise de Componentes Independentes (*Independent Component Analysis* - ICA) (Hyvarinen et al., 2001) é utilizada como pré-processamento para os classificadores neurais. A partir da in-

formação dos classificadores segmentados, diversas estratégias para combinação são utilizadas com o objetivo de produzir alta eficiência global de discriminação (Kuncheva, 2004).

2 Classificação Segmentada

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.

Em um trabalho anterior (Simas Filho et al., 2009), foi desenvolvido um sistema de classificação com pré-processamento por ICA para o problema em questão. Alta eficiência de discriminação foi obtida através da combinação de extração de características segmentadas (através de ICA) e classificação neural global (utilizando informação de todas as camadas). O objetivo do presente trabalho é verificar se o treinamento de classificadores segmentados (especialistas na informação de cada camada do calorímetro) produz maior eficiência.

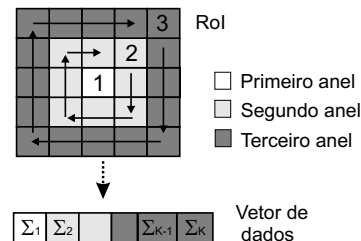


Figura 3: Processo de construção dos anéis.

Conforme proposto em (dos Anjos et al., 2006), os sinais de cada camada são inicialmente formatados em anéis concêntricos de deposição de energia. Em seguida a energia das células que pertencem a um dado anel é somada e normalizada (ver Figura 3). A dimensão dos sinais é reduzida de aproximadamente 700 células detectoras de uma RoI, para 100 anéis, que ficam assim distribuídos pelas camadas do calorímetro: PS-8, E1-64, E2-8, E3-8, H0-4, H1-4 e H2-4.

A Figura 4 ilustra o método proposto. De modo segmentado, após a construção dos anéis, é realizada a extração de características através de ICA e em seqüência a classificação neural. As saídas dos classificadores segmentados são combinadas visando produzir a decisão final para o canal de filtragem.

A seguir, serão abordados brevemente a análise de

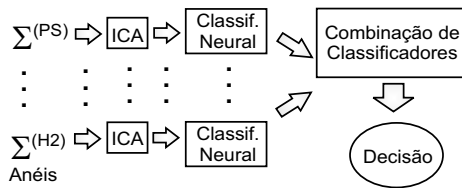


Figura 4: Diagrama do método proposto.

componentes independentes e os métodos de combinação de classificadores.

2.1 Análise de Componentes Independentes

A análise de componentes independentes (*Independent Component Analysis* - ICA) considera que um conjunto de sinais observados $\mathbf{x} = [x_1, x_2, \dots, x_N]$ é gerado a partir da combinação linear de fontes independentes $\mathbf{s} = [s_1, s_2, \dots, s_N]$ (Hyvarinen et al., 2001):

$$\mathbf{x} = \mathbf{A}\mathbf{s} \quad (1)$$

onde a matriz \mathbf{A} ($N \times N$) é conhecida como a matriz de mistura. Uma estimativa das fontes pode ser obtida pela transformação linear inversa:

$$\mathbf{y} = \mathbf{W}\mathbf{x} \quad (2)$$

onde os componentes de \mathbf{y} são estatisticamente independentes (ou o mais independente possível).

A ICA vêm sendo aplicada em diversos problemas de processamento de sinais (Hyvarinen et al., 2001) e extração de características para classificação (Poblador et al., 2004; Simas Filho et al., 2009).

Neste trabalho, foi utilizado o algoritmo JADE (*Joint Approximate Diagonalization of Eigenmatrices*) (Cardoso and Souloumiac, 1993) para estimação dos componentes independentes. No JADE, as informações estatísticas de segunda e quarta ordem são utilizadas a partir de uma abordagem tensorial. Tensores (Michal, 2008) são generalizações de alta-dimensão das matrizes. O tensor cumulante de quarta ordem (\mathbf{T}_4) é uma matriz de 4 dimensões que contém toda a informação de quarta-ordem dos dados (Cardoso and Souloumiac, 1993). Sabe-se que a diagonalização da matriz de correlação (\mathbf{C}_y) produz a decorrelação entre os componentes de \mathbf{y} (Hyvarinen et al., 2001). O método JADE propõe a diagonalização de \mathbf{T}_4 para alcançar a independência estatística (Cardoso and Souloumiac, 1993).

2.2 Combinação de 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 (Kuncheva, 2004).

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

$$\mu(\mathbf{x}) = \sum_{k=1}^K u_k(\mathbf{x}) \quad (3)$$

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 :

$$\mu(\mathbf{x}) = \sum_{k=1}^K \alpha_k u_k(\mathbf{x}) \quad (4)$$

Outra forma para a combinação de classificadores de saídas contínuas u_k é o cálculo da média geométrica:

$$\mu(\mathbf{x}) = \sqrt[\kappa]{\prod_{k=1}^K u_k(\mathbf{x})} \quad (5)$$

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 (Kuncheva, 2004). 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.

3 Resultados

Neste trabalho, foram utilizados dados obtidos a partir de simuladores de Monte Carlo para colisões próton-próton no LHC (ATLAS Collaboration, 2008). As características do detector e os efeitos do primeiro nível de filtragem são levadas em conta nas simulações. Estão disponíveis assinaturas (RoI de calorimetria) de aproximadamente 470.000 elétrons e 310.000 jatos, que foram divididas igualmente em conjuntos de treino, teste e validação.

Para avaliar o desempenho de discriminação são utilizados a curva ROC (Receiver Operating Characteristics) (Van-Trees, 2001), que mostra como as probabilidades de detecção (para elétrons) e falso

Tabela 1: Características dos classificadores segmentados (Máximo SP×100, Número de Neurônios Ocultos e PF p/ PD=97%).

	Anéis			Anéis+ICA		
	SP	NN	PF	SP	NN	PF
PS	64,9	6	40,0	64,9	6	40,0
E1	95,1	13	2,4	95,1	15	2,5
E2	88,8	6	9,6	88,8	6	9,7
E3	58,7	6	55,5	59,0	6	54,4
H0	66,7	3	49,9	64,6	3	49,1
H1	49,6	3	69,4	48,9	3	68,9
H2	37,8	2	85,9	36,8	2	85,6

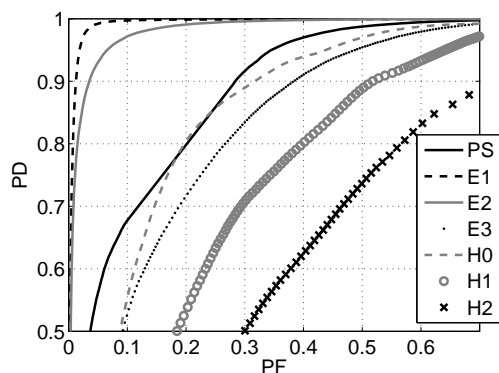


Figura 5: Curvas ROC para os classificadores segmentados (sem pré-processamento por ICA).

alarme (respectivamente PD e PF) variam conforme se ajusta o patamar de decisão, e o índice SP (dos Anjos et al., 2006), que é definido por:

$$SP = \left[(Ef_e + Ef_j) \times \sqrt{Ef_e \times Ef_j} \right] / 2 \quad (6)$$

onde Ef_e e Ef_j são as eficiências obtidas, respectivamente, para elétrons e jatos. O produto SP busca definir o projeto do classificador através do equilíbrio da eficiência em ambas as classes.

Para os classificadores segmentados foram utilizadas redes MLP (*Multi-Layer Perceptron*) (Haykin, 2008) com uma camada escondida, um neurônio de saída e função de ativação tangente hiperbólica. O algoritmo de treinamento foi o *Back-propagation* (Haykin, 2008). O número de neurônios da camada escondida foi obtido após testes exaustivos com várias configurações buscando máxima eficiência. Foram utilizados, como entrada, os sinais formatados em anéis com e sem pré-processamento por ICA. A Tabela 1 e a Figura 5 mostram as características dos classificadores segmentados.

Analisando a Tabela 1, percebe-se que para algumas camadas, a probabilidade de falso alarme (para PD=97%) foi levemente reduzida com o uso da ICA. O número de neurônios ocultos em cada classificador também não foi afetado de modo significativo pelo uso da ICA. Os discriminadores mais

Tabela 2: Desempenho dos métodos de combinação de classificadores (Máximo SP×100 e PF p/ PD=97%).

	Anéis		Anéis+ICA	
	SP	PF	SP	PF
Média	96,0	1,8	96,1	1,7
Méd. Pond.	96,5	1,5	96,5	1,4
Méd. Geom.	33,5	93,7	32,4	94,3
Votação	90,8	-	90,6	-
Vot. Pond.	94,2	-	94,0	-

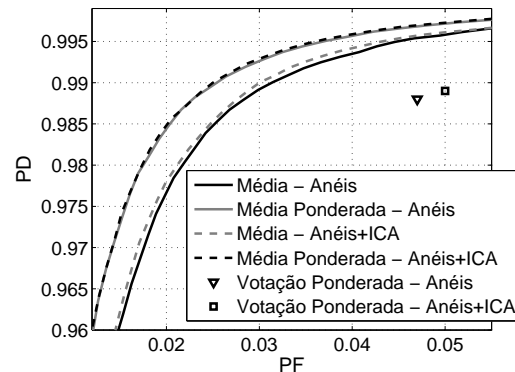


Figura 6: Desempenho dos métodos de combinação de classificadores.

eficientes foram os obtidos nas camadas E1 e E2 (camadas mais profundas e com maior número de células detectoras), e os de pior desempenho em H1 e H2 (camadas com baixa absorção de energia e por isso mais sensíveis ao ruído).

Para combinação da informação dos classificadores segmentados, foram utilizados a média, a média ponderada, a média geométrica, a votação e a votação ponderada (para a camada k , o fator de ponderação utilizado foi $\alpha_k = SP_k^{(Norm)}$, onde $SP_k^{(Norm)} = SP_k / (\sum_k SP_k)$, assim $\sum \alpha_k = 1$). Na Figura 6 e na Tabela 2, compara-se o desempenho dos métodos propostos. Percebe-se que, para a combinação por votação, existe apenas um par PD × PF. Considerando os diversos métodos, o mais eficiente foi a média ponderada, seguido da média e da votação ponderada. A média geométrica produziu resultados muito inferiores aos demais. O uso de ICA como pré-processamento não produziu aumento da eficiência. Para a combinação por votação há apenas um par PD × PF, não sendo possível preencher completamente a Tabela 2.

Considerando a Média Ponderada, foi testada a relevância da informação de cada camada na decisão global. Para isso, as saídas dos classificadores segmentados foram, uma a uma, removidas e o desempenho de discriminação novamente calculado. Observa-se, na Tabela 3, que a remoção da informação de algumas camadas (exceto E1) produz pouca ou nenhuma perda de eficiência.

Tabela 3: Desempenho obtido (Máximo SP×100 e PF p/ PD=97%) através da média ponderada, removendo-se algumas camadas (C. Rem)

C. Rem.	SP	PF	C. Rem.	SP	PF
PS	96,3	1,6	E3, H2	96,5	1,4
E1	92,8	4,3	E3, H1, H2	96,4	1,5
E2	96,3	1,6			
E3	96,5	1,4			
H0	96,2	1,6	PS, E3, H1, H2	96,2	1,7
H1	96,4	1,5			
H2	96,5	1,4			

Em (dos Anjos et al., 2006) foi mostrado que a maior parte do tempo de processamento (cerca de 90%) no sistema de classificação neural operando sobre anéis é gasta na montagem dos sinais em anéis, a partir dos dados brutos. Então, percebe-se que, através do método proposto, grande quantidade de recursos computacionais pode ser poupada a partir da montagem apenas dos anéis provenientes das camadas mais discriminantes.

Na Figura 7, o método proposto (CS - Classificação Segmentada) é comparado com um classificador neural operando diretamente sobre os sinais em anéis (Neural-Ringer, discriminador já pronto para operar no sistema de filtragem do ATLAS) (dos Anjos et al., 2006) e também com um único classificador neural operando sobre todos os componentes independentes extraídas de modo segmentado (SICA) (Simas Filho et al., 2009). Percebe-se que o método proposto é mais eficiente, mesmo sem a utilização de informação das camadas E3 e H2. Retirando-se quatro camadas, ainda assim é obtido melhor desempenho que o Neural-Ringer.

4 Conclusões

Foi proposto um sistema de classificação neural segmentado para o sistema de filtragem online do detector ATLAS. Com a segmentação, foi possível identificar que parte da informação disponível não apresenta características discriminantes significativas e pode ser eventualmente desprezada, contribuindo para a redução do tempo necessário para a decisão final. Através do método proposto foi obtida uma eficiência na identificação de elétrons de 97% para falso alarme (ruído de fundo erroneamente aceito) de apenas 1,4%.

Agradecimentos

Os autores agradecem ao CNPQ, CAPES, e FAPERJ pelo apoio financeiro, e à Colaboração ATLAS/DAQ/Trigger pelas sugestões e pelos dados experimentais utilizados.

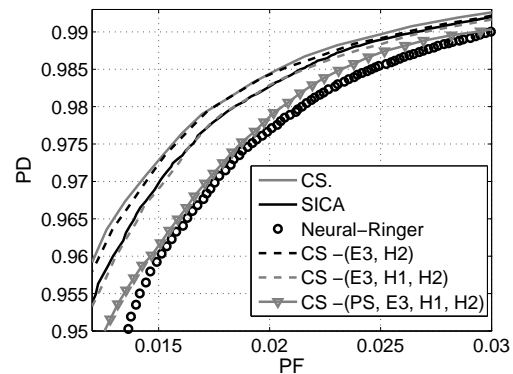


Figura 7: Comparação do método proposto com outros discriminadores. Entre parênteses as camadas removidas na combinação.

Referências

ATLAS Collaboration (2008). ATLAS experiment at CERN Large Hadron Collider, *Journal of Instrumentation* **3**(S08003).

Cardoso, J.-F. and Souloumiac, A. (1993). Blind beamforming for non-gaussian signals, *IEEE Proceedings-F* **140**(6): 362–370.

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

Evans, L. and Bryant, P. (2008). LHC machine, *Journal of Instrumentation* **3**(S08001).

Haykin, S. (2008). *Neural Networks and Learning Machines*, 3rd edn, Prentice Hall.

Hyvarinen, A., Karhunen, J. and Oja, E. (2001). *Independent Component Analysis*, Wiley.

Kuncheva, L. (2004). *Combining Pattern Classifiers, Methods and Algorithms*, Wiley.

Michal, A. D. (2008). *Matrix and Tensor Calculus*, 1st edn, Dover.

Poblador, V. S., Moreno, E. M. and Casals, J. S. (2004). ICA as a preprocessing technique for classification, *Ind. Comp. Analysis And Signal Separation, 5th Int. Conf.*, LNCS, Springer, Charleston, USA, pp. 1165–1172.

Simas Filho, E. F., Seixas, J. M. and Caloba, L. P. (2009). Optimized calorimeter signal compaction for an ICA based ATLAS electron/jet second-level trigger, *Proceedings of Science* **ACAT08**(102): 1–10.

Van-Trees, H. L. (2001). *Detection, Estimation, and Modulation Theory, Part I*, Wiley.