

REDES AUTO-CODIFICADORAS COMO PRÉ-PROCESSAMENTO EM CALORIMETRIA DE ALTAS ENERGIAS COM FINA SEGMENTAÇÃO

Carlos Eduardo Covas Costa*,
Micael Veríssimo de Araújo,
João Victor da Fonseca Pinto,
José Manoel de Seixas
Laboratório de Processamento de Sinais
COPPE/POLI-UFRJ
Rio de Janeiro, Brasil.
Email: *caducovas@lps.ufrj.br

Werner Spolidoro Freund
Laboratório de Processamento de Sinais
COPPE/POLI-UFRJ
Rio de Janeiro, Brasil.
LPNHE, CNRS/IN2P3
Sorbonne Université
Paris, França.

Resumo—Ambientes que apresentam elevada taxa de amostragem, uma grande quantidade de canais de leitura e eventos raros imersos em uma grande quantidade de ruído de fundo são comuns em diversas aplicações de engenharia. Nestes ambientes, a filtragem do sinal de interesse se torna desafiadora devido à alta dimensionalidade no espaço de características original e ao grande volume de dados gerados. Na física de partículas, este contexto se faz presente, como no caso do sistema de filtragem do experimento ATLAS, situado no LHC (Large Hadron Collider), que tem como objetivo identificar partículas envolvidas em processos físicos de interesse de estudo. Desde 2017, o algoritmo *NeuralRinger*, que consiste em um conjunto de redes neurais do tipo perceptron de múltiplas camadas, é utilizado no sistema de filtragem *online* do ATLAS com o objetivo de discriminar elétrons (sinal de interesse) de jatos (ruído). O *NeuralRinger* oferece redução de dimensionalidade através da representação da energia depositada nas células do sistema de calorimetria (medidores responsáveis pela estimação da energia associada aos eventos) com o uso de anéis concêntricos (total de 100) que alimentam as redes neurais. Este trabalho avalia a aplicação de técnicas de codificação para encontrar uma representação da informação anelada de calorimetria com um menor número de componentes e capaz de manter a capacidade discriminante. Redes auto-codificadoras discriminantes se mostraram capazes de representar a informação anelada de calorimetria com apenas um componente, mantendo a capacidade de discriminação elétron-jato. Os resultados foram obtidos utilizando uma base de dados de simulação de Monte Carlo, que representa as condições de colisões de prótons a uma energia de 13 TeV no centro de massa.

Keywords—Auto-codificadores; Inteligência computacional; Estatística de ordem superior; Física de altas energias

I. INTRODUÇÃO

Em diversas áreas da engenharia, é comum encontrar ambientes com elevada taxa de amostragem e alta dimensionalidade, onde busca-se um sinal raro imerso em uma grande quantidade de ruído indesejável. Estas características podem ser identificadas, por exemplo, em diversas aplicações de

telecomunicações, controle, óleo e gás, eletricidade, segurança, processamento de áudio, etc. Sendo assim, muitos dos problemas atuais de engenharia se beneficiam do uso de sistemas inteligentes baseados em processamento de sinais e inteligência computacional.

Na física de altas energias tais características também estão presentes. Esta área da ciência busca uma melhor compreensão acerca dos constituintes básicos da matéria que compõem o universo e das relações que os governam [1]. Com o objetivo de gerar dados que permitam este estudo, as colaborações internacionais têm feito uso de um aparato experimental complexo que permite recriar as condições existentes no universo instantes após seu nascimento. Dentre os principais instrumentos que compõem tal infraestrutura, encontram-se os aceleradores de partículas e os detectores [2].

O Large Hadron Collider (LHC), está situado em um túnel de 27Km de circunferência, 175 metros abaixo do solo da fronteira entre a França e a Suíça e, até o momento, é considerado o maior colisor de partículas e permite-nos observar fenômenos de interesse nas menores escalas já atingidas [3]. O experimento ATLAS [4] é um dos dois experimentos de propósito geral do LHC, no CERN (Conseil européen pour la recherche nucléaire), ou seja, estuda todos os tópicos de física acessíveis às energias de centro de massa do LHC. O ATLAS é composto por detectores como pode ser visto na Figura 1. Estes detectores registram a passagem de partículas carregadas e neutras, usando as características específicas destas quando interagem com a matéria, para obter medidas muito precisas da energia e da quantidade de movimento de cada uma delas.

Os dados gerados no experimento ATLAS não estão prontos para análise física, entretanto, antes de passarem por um sistema de filtragem *online* dos eventos de interesse (*trigger*). Esta filtragem torna-se fundamental devido à quantidade de dados originados nas colisões, que chegam a uma taxa da or-

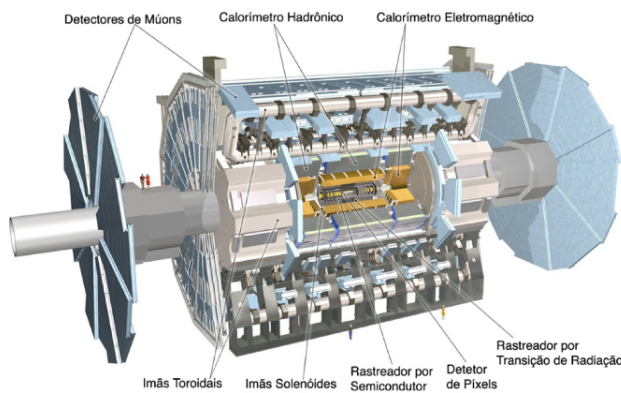


Figura 1. Diagrama ilustrativo do detector ATLAS. Fonte: [5]

dem de 70TB/s [6]. Como o armazenamento desta quantidade de dados não é trivial e apenas uma pequena parcela destes eventos constituem reações físicas de interesse, um sistema de filtragem *online* atua em uma primeira etapa com requisitos de baixo tempo de latência.

Como os eventos rejeitados pelo sistema de *trigger* não podem ser recuperados, a eficiência de detecção é aumentada com o custo de um aumento na quantidade de falso alarme. Como consequência, ao final desta primeira etapa de filtragem, ainda há uma quantidade significativa de ruído de fundo em meio aos canais físicos de interesse. Além disso, tais dados armazenados serão destinados a diferentes objetivos de estudo de física, como bóson de Higgs, supersimetria, dimensões extras, etc [7]. A fim de atender a estas necessidades diversas, sistemas de filtragem *offline* são empregados.

O sistema de filtragem *offline* conta com as informações provenientes do sistema de calorimetria e do detector de traços, no caso da análise de elétrons. Enquanto os calorímetros medem a energia incidente das partículas, bem como o perfil da deposição de energia, detectores de traço medem a trajetória de partículas eletricamente carregadas localizando espacialmente a ionização resultante [8]. Devido à alta dimensionalidade característica dos dados advindos de ambos os sistemas de detecção, as colaborações criaram variáveis discriminantes baseadas em conhecimento especialista da física do processo e, normalmente, a filtragem dos eventos de interesse, nesta etapa, vinha sendo feita através de cortes rígidos nestas variáveis. Mais recentemente, entretanto, os experimentos têm utilizado técnicas multivariadas como o método de verossimilhança [9], árvores de decisão (Boosted Decision Trees, BDT) e redes neurais, com o objetivo de aumentar seu potencial discriminante [10].

Um fator importante que afeta o desempenho de métodos de aprendizado de máquina é a escolha da representação de dados. Portanto, muito do esforço da aplicação de algoritmos de aprendizado de máquina se concentra na transformação dos dados e no teste de diferentes técnicas de pré-processamento [11]. Nos últimos anos, a área de aprendizado de representação, principalmente com a utilização de modelos

considerados profundos, tem experienciado um aumento de atividade científica de pesquisa. Alguns exemplos de sucesso apareceram nas áreas de reconhecimento de fala, reconhecimento de imagens, reconhecimento de objetos, processamento de linguagem natural, contribuindo positivamente, inclusive, para a área de físicas de altas energias [12]. Nesta última, tarefas como reconstrução de traço, seleção e identificação de objetos, análise física, estimação de energia e resiliência ao empilhamento dos sinais de diferentes eventos no detectores têm sido positivamente afetadas por tais modelos, uma vez que estes permitem aprendizado automático de representação de alto nível de dados, permitindo melhorar a exploração de todo o potencial do conjunto de dados coletados[10].

O *NeuralRinger* [6] é um algoritmo para identificação de elétrons, que utiliza anéis concêntricos (100) na representação da energia depositada nas células do sistema de calorimetria do detector ATLAS. Esta abordagem foi inicialmente proposta em 1996 [13], com base em [14], para separação de elétrons e jatos e entrou em operação no sistema de filtragem *online* (*trigger*) de elétrons do ATLAS, em sua camada de alto nível (HLT) [6], em 2017. Na versão atual, o teste de hipóteses é realizado através de *ensemble* de redes neurais especializadas em regiões do espaço de fase (posição da partícula incidente e sua energia (transversa, E_T)) medida pelo sistema de calorimetria.

O algoritmo *NeuralRinger* foi desenvolvido, originalmente, para o ambiente *online*, o que é incomum, pois, tradicionalmente, os modelos que atuam nas decisões da filtragem *online* consistem em degradações dos modelos utilizados na análise *offline* para adaptação aos requisitos de latência do *trigger*. A adoção do *NeuralRinger* para atuação *online* sem uma contrapartida na análise *offline*, portanto, inverte a lógica empregada nos experimentos. O presente trabalho estuda a possibilidade de se obter um espaço latente mais discriminante dos anéis, com informação de calorimetria, para atuação no ambiente *offline*, e de se obter uma representação mais compacta, para atuação no ambiente *online*, através da utilização de redes auto-codificadoras, com o objetivo de discriminar elétrons de jatos.

O resto deste artigo está organizado da seguinte forma: na Seção II o algoritmo *NeuralRinger* é descrito, na Seção III, a etapa de filtragem *offline* de Elétrons no ATLAS é descrita. As técnicas utilizadas para representação da informação são explicadas na Seção IV. O método de treinamento é descrito na Seção V, os resultados experimentais são analisados na Seção VI e as conclusões apresentadas na Seção V.

II. NEURALRINGER

Os chuveiros de partículas depositados nos calorímetros podem ser descritos por seus perfis lateral e longitudinal. Esta informação é crucial para a caracterização e discriminação de eventos de interesse. Os chuveiros gerados por elétrons possuem um formato cônico, estreito e com pouca penetração, enquanto as partículas provenientes de jatos (ruído) apresentam chuveiros mais largos e profundos [15]. Com o objetivo de preservar a informação referente à geometria do chuveiro, e, ao mesmo tempo, reduzir a dimensionalidade da informação de

energia das células do sistema de calorimetria, a representação da energia através de anéis concêntricos foi proposta [13].

O sistema de calorimetria do ATLAS conta com 7 camadas longitudinais de amostragem, sendo uma de pré-amostragem (PS, Pre-Sampler), três camadas eletromagnéticas (EM1, EM2 e EM3) e três camadas hadrônicas (HAD1, HAD2 e HAD3). Os anéis de energia utilizados pelo *NeuralRinger* são formados ao longo de todas estas sete camadas, totalizando 100 anéis, como pode ser visto na Tabela 1. A formação destes anéis se inicia na célula mais energética de cada camada, cuja energia é considerada o primeiro anel da camada, e continua através da soma da energia (E_T) das células externas em relação ao anel anterior, até que a Região de interesse (RoI) seja coberta ou um número máximo de anéis (100) seja atingido, como pode ser visto na Figura 2.

Tabela I

NÚMERO DE ANÉIS POR CAMADA DOS CALORÍMETROS. SENDO PS O CALORÍMETRO PRÉ-AMOSTRADOR; EM1,2,3 REFERENTES ÀS CAMADAS DO CALORÍMETRO ELETROMAGNÉTICO; E HAD1,2,3 REFERENTE ÀS CAMADAS DO CALORÍMETRO HADRÔNICO.

PS	EM1	EM2	EM3	HAD1	HAD2	HAD3	Total
8	64	8	8	4	4	4	100

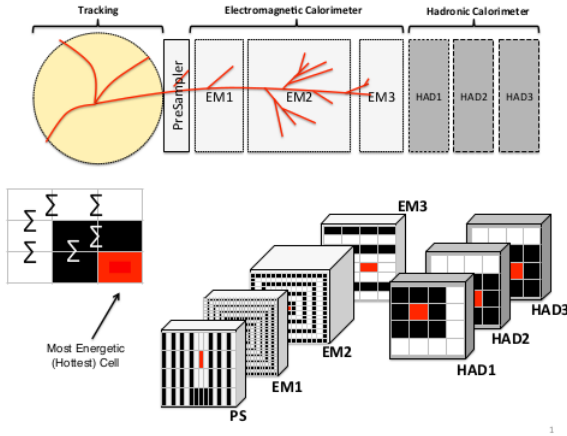


Figura 2. Construção dos anéis pelo algoritmo *NeuralRinger* através das seções do calorímetro. Fonte: [16]

Após o processo de criação dos anéis, o teste de hipótese é realizado através de um *ensemble* de classificadores compostos por redes neurais do tipo Perceptrons de Múltiplas Camadas (Multilayer Perceptron, MLP) [17], que são projetadas para regiões do plano formado pela energia transversa e posição da partícula, para discriminação elétron-hádron. A Tabela II apresenta as divisões deste plano, onde E_T representa a energia transversa e η representa a pseudorapidez [4], que identifica a posição da partícula definida em função do ângulo polar utilizado nas coordenadas cilíndricas do experimento.

Tabela II
FRONTEIRAS PARA AS REGIÕES DE η E E_T EMPREGADO NO AJUSTE DOS MODELOS PARA O *NeuralRinger*. UM TOTAL DE 25 REGIÕES SÃO UTILIZADAS PARA COMPOR O *ensemble* DE REDES NEURAIS.

Ajuste do Modelo					
Fronteira das Regiões em E_T [GeV]					
15	20	30	40	50	∞
Fronteiras das Regiões em $ \eta $					
$0 \leq \eta < 0.8$					
$0.8 \leq \eta < 1.37$					
$1.37 \leq \eta < 1.54$					
$1.54 \leq \eta < 2.5$					
$ \eta \geq 2.5$					

III. FILTRAGEM *Offline* DE ELÉTRONS NO ATLAS

O experimento ATLAS possui interesses de estudo diversos. O sistema de filtragem *offline* valida a decisão do *trigger* e permite uma filtragem de acordo com os diferentes objetivos de estudo da física. Normalmente, a colaboração fornece um conjunto de menus que seguem diferentes critérios de aquisição dos dados. No caso da identificação de elétrons, os menus são descritos como *Tight*, *Medium*, *Loose* e *Very Loose*, os quais se diferenciam pela relação entre a pureza da amostra e a quantidade de amostras disponíveis do processo de interesse. Ao passo que o critério *Tight* prioriza a pureza da amostra, o *Very Loose* obtém um grande quantidade de amostras, com maior contaminação, sendo utilizado, por exemplo, na busca de processos que contemplem a física nova. O critério *Medium* apresenta equilíbrio entre a pureza e a estatística do processo.

Na Física de Altas Energias, é comum, por parte das colaborações, criar grandezas altamente discriminativas para a aplicação de cortes lineares. De 2010 a 2012, os elétrons, no ATLAS foram identificados pelo uso de cortes simples em um conjunto de variáveis (referido como *Cut-based menu*) [18].

A fim de obter estatística suficiente para observação de processos raros, o LHC vem elevando a taxa de colisões próton-próton. Apesar de positiva, esta estratégia aumenta significativamente a ocupação dos detectores, provocando empilhamento de sinais, o que aumenta a demanda por métodos mais eficientes e que permitam melhor explorar todo o potencial de análise disponível. A partir de 2012, portanto, o ATLAS alterou sua referência, nas análises envolvendo elétrons, para uma abordagem utilizando técnicas multivariáveis [9]. A estratégia multivariável para filtragem *offline* de elétrons será apresentada abaixo.

A. Likelihood

O método de verossimilhança faz uso de funções de densidade de probabilidade (PDF), de sinal e de ruído, de variáveis discriminantes de calorimetria e traço. Com base nestas PDFs unidimensionais, a probabilidade condicional total, assumindo-se que as variáveis são descorrelacionadas, é o produto das PDFs [9].

$$L_{S(B)}(\vec{x}) = \prod_{i=1}^n P_{S(B),i}(x_i) \quad (1)$$

onde \vec{x} é representado pelas variáveis discriminantes na identificação de elétrons e $P_{S,i}(x_i)$ é o valor da PDF referente ao sinal (elétron) da i -ésima variável dada por x_i . Da mesma forma, $P_{B,i}(x_i)$ refere-se ao valor da PDF para o ruído [9].

Tais probabilidades são, então, combinadas em um discriminante no qual um corte é aplicado, como pode ser visto na equação abaixo:

$$d_L = \frac{L_S}{L_S + L_B} \quad (2)$$

A escolha do valor de corte (ponto de operação) determina a eficiência do sinal ou rejeição do ruído de fundo atingida pelo método.

Este método apresenta vantagens, quando comparado aos métodos tradicionais de cortes rígidos em variáveis físicas, com relação ao seu potencial de discriminação, uma vez que a construção da probabilidade permite que as caudas de variáveis discriminantes sejam parcialmente recuperadas, resultando em uma maior eficiência com a mesma rejeição ao ruído. Além disso, este método permite a utilização de variáveis que, apesar de apresentarem poder discriminativo significativo, não eram tradicionalmente utilizadas pelos métodos de cortes rígidos por apresentarem uma sobreposição considerável entre as distribuições de sinal e ruído [18].

IV. TÉCNICAS DE CODIFICAÇÃO DA INFORMAÇÃO

Com o objetivo de avaliar redes auto-codificadoras para representação da informação anelada de calorimetria, a fim de encontrar um espaço latente mais discriminante para atuação no ambiente *offline* ou uma codificação com um menor número de componentes, para atuação no ambiente *online*, diferentes arquiteturas de redes auto-codificadoras foram avaliadas, sendo comparadas com formulações conhecidas como análise de componentes principais (PCA) e análise de componentes principais não-lineares (NLPCA). Estes modelos serão descritos a seguir.

A. PCA e PCA Não-Linear (NLPCA)

A análise de componentes principais (PCA) projeta dados N-dimensionais em M direções ortogonais maximizando a variância. Essas M direções principais formam um sub-espaço de dimensão menor, sendo possível representar amostras N-dimensionais por suas projeções nas M direções principais. Uma generalização, de forma não-linear, da PCA foi proposta por Kramer em 1991 [19]. Kramer sustenta que, para alcançar extração não-linear de características, três camadas escondidas são necessárias, sendo uma camada de mapeamento, uma de gargalo e uma de mapeamento. Os resultados de Cybenko [20], de que uma camada escondida é suficiente para modelar qualquer função contínua não linear em subconjuntos compactos e com um número suficiente de unidades escondidas, aplica-se para as funções G (mapeamento) e H (de-mapeamento) individualmente. Ao explicar a motivação do uso de três camadas escondidas na sua proposta, Kramer afirma que, no caso da utilização de apenas uma

camada escondida (retirando as camadas de mapeamento e de-mapeamento propostas), se as ativações da camada de gargalo fossem lineares, os resultados seriam similares à PCA, e, no caso de ativações sigmoidais, as formas funcionais da codificação e decodificação ainda seriam severamente restringidas, sendo possível representar apenas uma combinação linear das entradas comprimida pela sigmoide no intervalo (-1,1).

B. Redes neurais auto-codificadoras

Uma rede auto-codificadora (Autoencoder, AE) é uma Rede Neural Feed-Forward treinada para reproduzir a informação de entrada na camada de saída, combinando uma função de codificação, que converte os dados de entrada em uma representação latente, e uma função de decodificação, que converte a nova representação de volta ao formato original, como pode ser visto na Figura 3 [21]. As redes auto-codificadoras são treinadas utilizando o algoritmo de Backpropagation[17], para preservar o máximo de informações possível quando uma entrada é executada através do codificador e, em seguida, do decodificador, sendo um algoritmo com perdas treinado de forma não supervisionada, a fim de obter codificação eficiente (menor erro de reconstrução) [21]. Entre seus principais objetivos de uso estão a redução de dimensionalidade de forma não-linear, e a extração de característica significativas dos dados, à medida que diferentes restrições são postas na função custo que norteia a reconstrução.

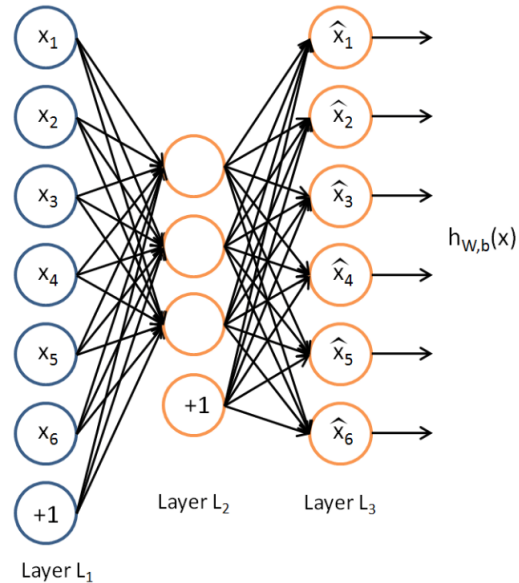


Figura 3. Diagrama ilustrativo da arquitetura de uma rede Auto-codificadora. Fonte: [22]

C. Redes neurais auto-codificadoras com regularização

Diferentes modelos de redes neurais auto-codificadoras com regularização foram propostas nos últimos anos. Uma proposta de penalidade utilizando a norma Frobenius quadrática da matriz jacobiana de derivadas parciais associadas a função de codificação é apresentada em [23]. Essa abordagem,

chamada de Contractive Autoencoder (CAE), tem conexões teóricas com as redes auto-codificadoras de aprendizado múltiplo e modelagem probabilística, e permite encontrar uma representação que captura melhor as direções locais de variação ditadas pelos dados, correspondendo a uma variedade não-linear de menor dimensão, sendo mais invariável para a grande maioria das direções ortogonais à variedade. Tais direções encontradas supostamente se aproximam dos planos tangentes à variedade.

D. Redes neurais auto-codificadoras profundas

Autoencoders profundos foram pensados na década de 1980, mas não era possível treiná-los de maneira a conseguir resultados significativamente melhores que a PCA. Nesta época, os treinamentos não convergiam devido ao problema do desaparecimento do gradiente e à demora ou inviabilidade do treinamento, o que foi superado em 2006 com a idéia do pré-treino camada a camada e empilhamento de redes auto-codificadoras (Stacked Autoencoders, SAE), com uma etapa não-supervisionada e uma etapa de ajuste fino dos pesos, de acordo com a tarefa final, apresentada por Hinton e Salakhutdinov [24].

E. Redes neurais auto-codificadoras discriminantes

As redes neurais auto-codificadoras discriminantes propostas em [25] tentam aumentar a compactação intra-classe e a separação inter-classe, ao mesmo tempo que apresentam uma codificação latente em dimensão reduzida. Para atingir este objetivo, a rede auto-codificadora reconstrói, no lugar da própria entrada, uma transformação da entrada mais discriminante, de forma que as amostras estejam mais próximas de outras amostras da mesma classe e mais distantes das amostras das outras classes, através de um processo iterativo de mudança de alvo. Desta forma, este modelo minimiza a função:

$$l = \|\hat{x} - \tilde{x}^{(t)}\|_2^2 \quad (3)$$

onde \hat{x} representa o vetor de entradas e $\tilde{x}^{(t)}$ o vetor de entradas após t iterações. No lugar das entradas, o novo alvo é transformado de maneira que as amostras se movam na direção dos centros de classe, por um valor pequeno de α :

$$\tilde{x}_i^{(t+1)} = (1 - \alpha)\tilde{x}_i^{(t)} + \alpha \left(\frac{1}{|C_i|} \sum_{\tilde{x}_j^{(t)} \in C_i} \tilde{x}_j^{(t)} \right) \quad (4)$$

onde α é um parâmetro que permite ajustar a magnitude do deslocamento e C_i o conjunto de amostras pertencentes a class i . De forma similar, as distâncias entre as amostras e centros de outras classes podem ser intensificadas afastando cada amostra da média das amostras pertencentes a outras classes:

$$\tilde{x}_i^{(t+1)} = (1 + \alpha)\tilde{x}_i^{(t)} - \alpha \left(\frac{1}{N - |C_i|} \sum_{\tilde{x}_j^{(t)} \notin C_i} \tilde{x}_j^{(t)} \right) \quad (5)$$

onde N representa o total de amostras.

V. MÉTODO

O conjunto de dados utilizado foi obtido através de simulação Monte Carlo, que representa as condições de colisão a uma energia de 13TeV no centro de massa. Este conjunto de dados é dividido em 36 bins, de acordo com a posição do detector e com a energia da partícula incidente. Para este trabalho, apenas o bin correspondente à posição central do detector, contendo eventos com energia de 20 a 30 GeV, foi utilizado, totalizando 228339 amostras.

Os treinamentos foram feitos em duas etapas, sendo uma primeira etapa de pré-processamento e uma segunda de classificação, onde a representação obtida pelos modelos utilizados na primeira etapa serviram como entrada para o treinamento do classificador.

O Erro Médio Quadrático (MSE) foi usado como função custo no treinamento das redes auto-codificadores, do modelo NLPCA e do classificador. Nestes treinamentos, o algoritmo Adam [26] foi utilizado para otimização com taxa de aprendizagem adaptativa, com 2048 amostras como tamanho da batelada. As redes auto-codificadoras foram treinadas com função de ativação tangente hiperbólica na camada intermediária e linear na camada de saída, enquanto o classificador foi treinado com tangente hiperbólica em todas as camadas. Um número máximo de 5000 épocas foi utilizado em todos os treinamentos, com parada antecipada baseada no erro no conjunto de validação (MSE).

Para que fosse possível estudar o comportamento dos codificadores para compactações cada vez mais intensas, os modelos de pré-processamento foram avaliados para diferentes números de dimensões codificadas. Considerando que o espaço original possui 100 dimensões, variou-se a dimensão codificada em intervalo de 10 unidades e em intervalos de 1 unidade abaixo de 30 dimensões. Sendo assim, diferentes estratégias foram adotadas para escolha das arquiteturas do modelos.

Os modelos baseados em redes neurais possuem um camada de gargalo, cuja saída representa a codificação aprendida e, portanto, a quantidade de neurônios é igual a dimensão da codificação. No caso das redes auto-codificadoras simples (AE) e as redes contrativas (CAE), que possuem apenas uma camada escondida, o número de neurônios desta camada é igual ao número de dimensões codificadas, uma vez que esta é a camada de gargalo. As redes auto-codificadoras profundas e o modelo NLPCA, por outro lado, possuem outras camadas além da camada de gargalo e foram treinadas com um número de neurônios incremental ao longo destas camadas. O modelo NLPCA utilizou 30 neurônios a mais nas camadas de mapeamento e de-mapeamento do que na camada de gargalo para todas as codificações, estrutura similar à utilizada nas redes auto-codificadoras profundas com duas camadas. A diferença no treinamento de ambas, portanto, consistiu no fato de que as redes auto-codificadoras foram treinadas camada a camada, com a possibilidade de ajuste fino, enquanto o modelo NLPCA foi treinado de uma só vez. As redes auto-codificadoras profundas com 3 camadas apresentaram acréscimo de 20 e 30 neurônios com relação à camada de

gargalo e as redes profundas com 5 camadas apresentaram acréscimos de 20, 30, 40 e 60. Após o treinamento destas redes, a projeção na saída da camada de gargalo foi utilizada como representação e fornecida ao classificador.

As redes auto-codificadoras profundas foram treinadas em duas etapas. Primeiramente, com um processo de treinamento camada a camada e, posteriormente, com um ajuste fino dos pesos de toda a rede na tarefa de compressão dos dados (reconstrução da entrada na saída da rede após a camada de gargalo).

No caso das redes auto-codificadoras discriminantes, como seu treinamento consiste em reconstruir uma transformação da entrada, em que as amostras se encontram mais próximas das amostras da mesma classe e distantes das amostras da outra classe, estes modelos foram treinados variando o número de iterações em que os dados eram transformados de 1 até 20 iterações. O parâmetro α , que pondera a intensidade da transformação, foi escolhido igual a 0.2 em todas as transformações.

No caso das redes MLP, treinadas para discriminação, sua arquitetura consistiu em uma rede neural com 5 neurônios na camada escondida e 1 neurônio na camada de saída. O índice SP foi utilizado como figura de mérito para comparação dos modelos, uma vez que no máximo (SP_{max}) o mesmo apresenta um equilíbrio entre a probabilidade de detecção (P_D) e o falso alarme (P_F), conforme equação:

$$SP = \sqrt{\sqrt{P_D \cdot (1 - P_F)} \cdot \frac{P_D + (1 - P_F)}{2}} \quad (6)$$

onde P_D é a probabilidade de detecção de elétrons e P_F é a probabilidade de classificar como elétron um jato hadrônico (ruído).

Para estimar a incerteza associada aos modelos, empregou-se o método jackknife de validação cruzada [27], onde a partir do conjunto de treinamento são geradas 10 partições, onde 9 são usadas no treinamento e 1 como validação/teste.

O pré-processamento utilizado para os anéis foi a norma-1 descrita na equação abaixo:

$$r'_i = \frac{r_i}{\sum_{j=1}^N r_j} \quad \forall i = 1, 2, 3 \dots N \quad (7)$$

onde N é o número de anéis (100).

VI. RESULTADOS

Inicialmente, diferentes números de iterações, para a rede auto-codificadora discriminante, foram comparados, como pode ser visto na Figura 4. Como o índice SP se estabiliza após 9 iterações, o valor de 10 iterações foi escolhido para comparação com os demais modelos.

A Figura 5 mostra a projeção dos 100 anéis de energia em duas componentes com uso de análise de componentes principais (PCA). Nota-se que a capacidade discriminativa se torna restrita, a medida que as componentes desta projeção apresentam uma área de sobreposição significativa. A Figura 6 mostra a projeção em duas componentes, obtidas através da

rede auto-codificadora discriminativa com três camadas (50-25-2), para diferentes iterações de transformação dos dados. As componentes apresentadas são obtidas através da projeção do espaço de entrada na camada intermediária, após um número de 1, 4, 6 e 10 iterações de transformação dos dados de entrada.

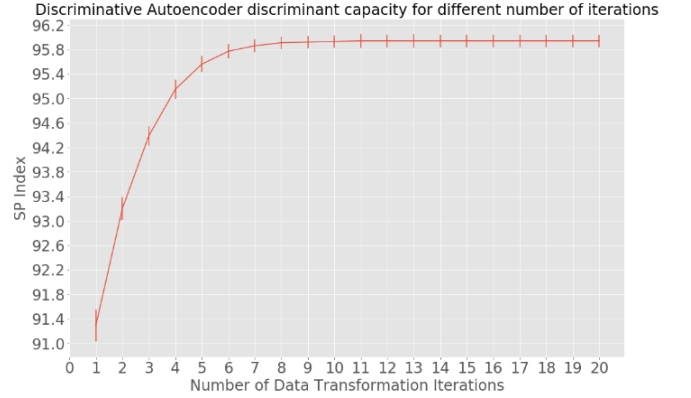


Figura 4. Valor do índice SP para diferentes números de iterações de transformações de dados para a rede auto-codificadora discriminante, considerando apenas uma dimensão na codificação. $20\text{GeV} < E_t < 30\text{GeV}$, $0.6 < \eta < 0.8$

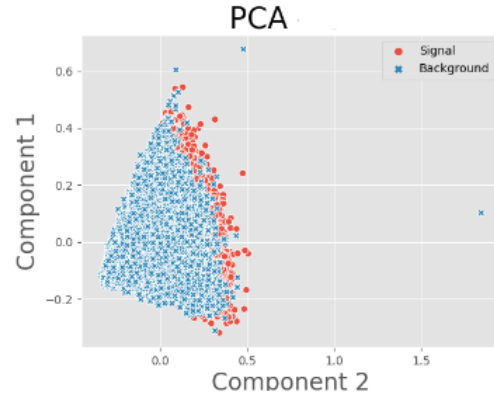


Figura 5. Projeção da informação anelada de calorimetria em duas componentes obtidas através da análise de componentes principais (PCA).

Nas Figuras 7 e 8, diferentes modelos de codificação são comparados com relação à capacidade de discriminação fornecida pela estrutura latente encontrada. Na Figura 7, nota-se a importância da etapa de ajuste fino para melhoria da capacidade de discriminação elétron-hádron em codificações com poucas dimensões. Entretanto, os modelos apresentaram instabilidade com uma barra de erro relativamente larga, considerando os diferentes resultados de validação cruzada.

Na figura 8 verifica-se que o modelo de rede auto-codificadora discriminante obtém os melhores resultados, o que era esperado, a medida que este é o único modelo que utiliza informação de classe durante o treinamento, não consistindo, exatamente, de um modelo de caráter não-



Figura 6. Projeção da informação anelada de calorimetria em duas componentes, obtidas através da rede auto-codificadora discriminativa com três camadas (50-25-2), para diferentes iterações.

supervisionado como os outros. Os demais modelos apresentam degradação no índice SP para codificações abaixo de 25 dimensões. Os modelos NLPCA e auto-codificador profundo com 2 camadas apresentaram, entretanto, uma melhor performance, para dimensões abaixo de 3, quando comparados aos demais modelos. Sendo assim, nota-se uma vantagem na utilização de codificações não-lineares que exploram a estatística de ordem superior para baixas dimensões. Verificou-se que o índice SP se mantém constante para codificações acima de 25 dimensões.

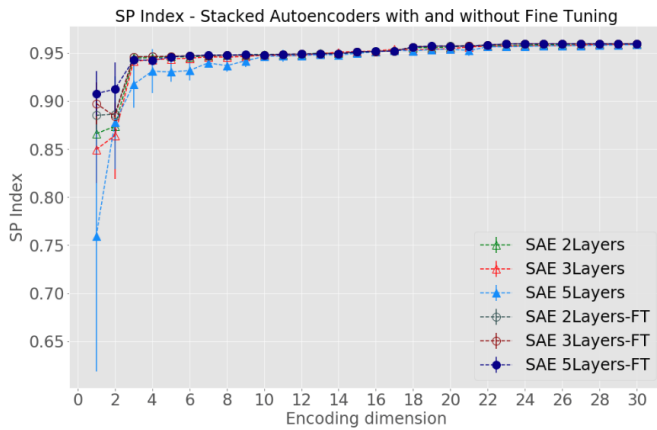


Figura 7. Valor do índice soma produto para diferentes redes auto-codificadoras com múltiplas camadas com e sem o uso de ajuste fino após o treinamento camada a camada. A dimensão da codificação variou de 30 a 1. $20\text{GeV} < E_t < 30\text{GeV}$, $0.6 < \eta < 0.8$

A rede auto-codificadora discriminante mantém a capacidade discriminativa mesmo com apenas 1 dimensão codificada, o que apresenta a possibilidade de codificar os anéis de forma simples. A Figura 9 apresenta uma análise dos parâmetros

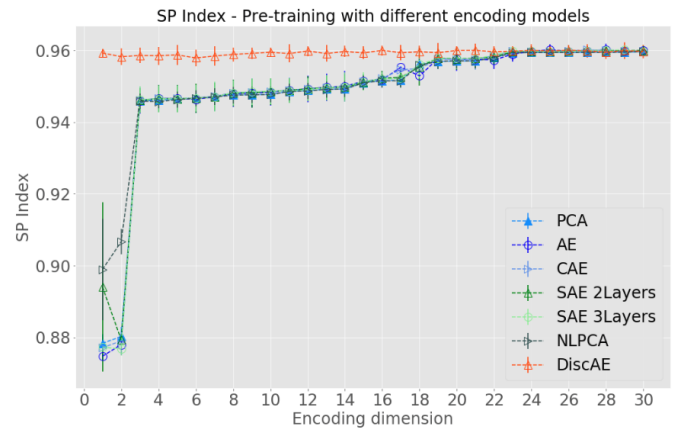


Figura 8. Valor do índice soma produto para diferentes modelos de compactação, a dimensão da codificação variando de 30 a 1. $20\text{GeV} < E_t < 30\text{GeV}$, $0.6 < \eta < 0.8$

aprendidos por tal rede. Com base na figura, é possível verificar que a rede fornece valores de peso positivos, em verde aos anéis que apresentam, em média, valor de energia maior para a classe ruído que para a classe sinal e, por outro lado, valores negativos para os anéis onde a classe sinal deposita mais energia. Nota-se também que a magnitude dos parâmetros tem relação com a diferença média de energia entre ambas as classes. Na parte inferior da Figura 9, a multiplicação dos pesos pela média de cada classe é feita e somada. Nota-se que após a aplicação da função tangente hiperbólica, na soma obtida, o valor para a classe sinal é convertido para -0.94, enquanto para a classe ruído o valor é de 0.99. Isso mostra que o modelo de rede auto-codificadora discriminativa alcança um separação significativa das classes, mas que também é interpretável. Sendo assim, é possível que esta parametrização sirva para codificação de uma variável discriminante representando a informação anelada de calorimetria e auxilie na análise *offline*, uma vez que se conhece o processo de tomada de decisão.

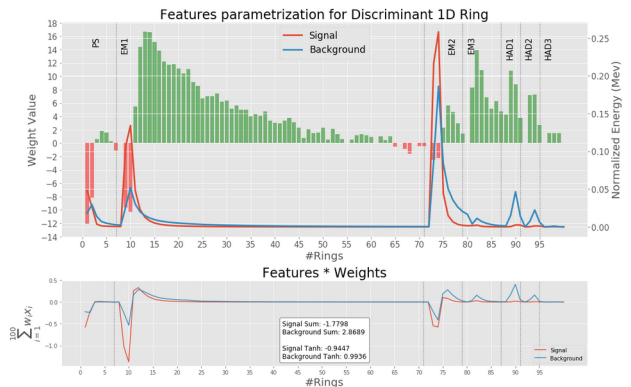


Figura 9. Valor dos parâmetros associados a cada anel pela codificação discriminante. Valores positivos (verde) são atribuídos para anéis que apresentam um valor energético apresentado pela classe ruído maior que a classe sinal, enquanto valores negativos (vermelho) são atribuídos para anéis em que a classe sinal deposita mais energia.

Tabela III

VALORES DE EFICIÊNCIA PARA CODIFICAÇÃO EM 1 DIMENSÕES,
OPERANDO NO PONTO DE MÁXIMO SP

Codificação	SP[%]	P_D [%]	P_f [%]
100anéis(Ref)	95.98±0.11	97.19±0.34	5.22±0.31
AE	87.70±0.25	91.92±0.97	16.41±1.14
AE,2Camadas	81.62±0.92	86.15±1.34	22.78±1.10
AE,2CamadasFT	90.15±1.55	95.48±2.05	15.02±1.27
AE,3Camadas	82.64±4.36	86.56±4.80	21.19±4.00
AE,3CamadasFT	88.53±1.68	92.58±2.18	15.42±1.51
CAE	87.84±0.25	91.99±1.17	16.20±1.20
Disc10IterAE	95.93±0.10	96.94±0.39	5.07±0.39
NLPCA	88.42±0.99	92.58±1.87	15.64±1.08
PCA	87.83±0.25	91.61±1.01	15.86±1.07

A Tabela III apresenta os resultados para o algoritmo *NeuralRinger* sem nenhuma codificação (100 anéis), comparando com as demais codificações estudadas. Nota-se um bom indicativo que a rede auto-codificadora discriminante possibilita a redução da dimensionalidade em 99% sem perder a capacidade discriminativa, uma vez que esta se encontra dentro da barra de erro quando comparada com a referência (*NeuralRinger* com o espaço original de 100 anéis).

VII. CONCLUSÃO

A rede auto-codificadora discriminante manteve a capacidade discriminativa com apenas uma dimensão, ao passo que a representação encontrada por este modelo permitiu um entendimento do processo de tomada de decisão. A análise de tal codificação é interessante, uma vez que os experimentos exigem uma compreensão e entendimento do comportamento dos modelos a fim de evitar a influência de erros sistemáticos. Sendo assim, os resultados apontam para uma possibilidade de se utilizar esta variável adicionando-a aos modelos empregados atualmente na filtragem *offline*, como a Likelihood, a fim de incorporar a informação anelada de calorimetria ao teste de hipótese empregado pela mesma, potencializando a capacidade de discriminação, com a inserção da informação que ainda não é utilizada no processo atual, sem aumentar a complexidade da solução. A representação encontrada permite, também, uma codificação intensa dos anéis para atuação no ambiente de filtragem *online*. Entretanto, cada mudança tanto no ambiente *online* como no *offline* precisa ser avaliada com profundidade e, portanto, novas análises ainda são necessárias para verificar, por exemplo, se a compactação encontrada cria tendências. Necessita-se, também, analisar se há ganhos reais, sem perda de informação relevante, ao se adicionar a variável encontrada na de decisão do *offline* pela Likelihood e, por último, é necessário analisar o comportamento do modelo ao longo das demais faixas de energia e posição da partícula no detector.

AGRADECIMENTOS

Os autores gostariam de agradecer à FAPERJ, CNPq, RE-NAFAE e CERN (Suíça) pelo apoio a este trabalho.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001

REFERÊNCIAS

- [1] Griffiths, David. Introduction to elementary particles. John Wiley & Sons, 2008.
- [2] Lincoln, Don. "The quantum frontier: The large Hadron collider." JHU Press, 2009.
- [3] Evans, Lyndon, and Philip Bryant. "LHC machine." Journal of instrumentation 3.08 (2008): S08001.
- [4] Collaboration, A. T. L. A. S. "The ATLAS experiment at the CERN large hadron collider." (2008).
- [5] Pequenaço, Joao. "Computer generated image of the whole ATLAS detector" . No. CERN-GE-0803012. 2008.
- [6] da Fonseca Pinto, Joao Victor. "Ring-shaped Calorimetry Information for a Neural Egamma Identification with ATLAS Detector." *J. Phys.: Conf. Ser.* Vol. 762. No. ATL-DAQ-PROC-2016-007. ATL-COM-DAQ-2016-015, 2016.
- [7] Atlas Collaboration. "ATLAS high-level trigger, data acquisition and controls." Technical Design Report, CERN/LHCC 22 (2003): 2003.
- [8] Albertsson, Kim, et al. "Machine learning in high energy physics community white paper." Journal of Physics: Conference Series. Vol. 1085. No. 2. IOP Publishing, 2018.
- [9] Alison, J., et al. "Description and Performance of the Electron Likelihood Tool at ATLAS using 2012 LHC Data." ATLAS Internal Communication ATL-COM-PHYS-2013-378 (2013).
- [10] Guest, Dan, Kyle Cranmer, and Daniel Whiteson. "Deep learning and its application to LHC physics." Annual Review of Nuclear and Particle Science 68 (2018): 161-181.
- [11] Bengio, Yoshua, Aaron Courville, and Pascal Vincent. "Representation learning: A review and new perspectives." IEEE transactions on pattern analysis and machine intelligence 35.8 (2013): 1798-1828.
- [12] Baldi, Pierre, Peter Sadowski, and Daniel Whiteson. "Searching for exotic particles in high-energy physics with deep learning." Nature communications 5 (2014): 4308.
- [13] Seixas, J. M., et al. "Neural second-level trigger system based on calorimetry." Computer Physics Communications 95.2-3 (1996): 143-157.
- [14] Acosta, D., et al. "The performance of a lead/scintillating-fiber calorimeter at lhc/ssc compatible gate widths." Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment 314.3 (1992): 431-449.
- [15] Wigmans, Richard, and Richard Wigmans. "Calorimetry: Energy measurement in particle physics." Vol. 107. Oxford University Press, 2000.
- [16] Spolidoro Freund, Werner. "An Ensemble of Neural Networks for Online Electron Filtering at the ATLAS Experiment". No. ATL-DAQ-PROC-2019-009. ATL-COM-DAQ-2019-067, 2019.
- [17] Haykin, Simon. "Neural networks: a comprehensive foundation." Prentice Hall PTR, 1994.
- [18] Brendlinger, Kurt. Physics with Electrons in the ATLAS Detector. Springer, 2018.
- [19] Kramer, Mark A. "Nonlinear principal component analysis using auto-associative neural networks." AICHE journal 37.2 (1991): 233-243.
- [20] Cybenko, George. "Approximation by superpositions of a sigmoidal function." Mathematics of control, signals and systems 2.4 (1989): 303-314.
- [21] Goodfellow, Ian, Yoshua Bengio, and Aaron Courville. Deep learning. MIT press, 2016
- [22] Maas, Andrew L., and Andrew Y. Ng. URL: <http://ufldl.stanford.edu/tutorial/unsupervised/Autoencoders/>
- [23] Rifai, Salah, Pascal Vincent, Xavier Muller, Xavier Glorot, and Yoshua Bengio. "Contractive auto-encoders: Explicit invariance during feature extraction." In Proceedings of the 28th International Conference on International Conference on Machine Learning, pp. 833-840. Omnipress, 2011.
- [24] Hinton, Geoffrey E., and Ruslan R. Salakhutdinov. "Reducing the dimensionality of data with neural networks." science 313.5786 (2006): 504-507.
- [25] Nousi, Paraskevi, and Anastasios Tefas. "Discriminatively trained auto-encoders for fast and accurate face recognition." International Conference on Engineering Applications of Neural Networks. Springer, Cham, 2017.
- [26] Kingma, Diederik P., and Jimmy Ba. "Adam: A method for stochastic optimization." arXiv preprint arXiv:1412.6980 (2014).
- [27] Efron, Bradley, and Gail Gong. "A leisurely look at the bootstrap, the jackknife, and cross-validation." The American Statistician 37.1 (1983): 36-48.