



MINISTÉRIO DA CIÊNCIA E TECNOLOGIA

INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS

CLASSIFICAÇÃO INTELIGENTE DE EVENTOS EXTREMOS UTILIZANDO SGD E HOG EM IMAGENS DE SATÉLITE

Luis Ricardo Arantes Filho

Emerson Jean da Silva

Trabalho Final da Disciplina CAP 378
Tópicos em Observação da Terra.

INPE
São José dos Campos
2018

RESUMO

Este trabalho apresenta uma proposta de análise de eventos extremos em imagens de satélite, especificamente a caracterização de eventos climáticos como chuvas torrenciais, tempestades, furacões e tornados. A classificação de eventos climáticos é de certa forma, algo de grande importância para todas as categorias sociais e governamentais, isto se dá devido aos efeitos de eventos extremos em diversos setores da economia, da agricultura, da saúde pública e setores sociais. Neste sentido este trabalho tem por objetivo, utilizando técnicas de *machine learning* baseadas no SGD (*Stochastic Gradient Descent*) e extração de características em imagens por HOG (*Histogram of Oriented Gradients*) apresentar uma abordagem inicial de classificação de eventos extremos em imagens de satélite sobre um grande volume de dados de forma a identificar com tempo reduzido estes eventos para que as tomadas de decisão sejam feitas de maneira otimizada.

Palavras-chave: Eventos extremos, imagens de satélite, *machine learning*, SGD, HOG, classificação.

LISTA DE FIGURAS

	<u>Pág.</u>
Figura 2.1 – Funções de custo utilizadas	6
Figura 2.2 – Cálculo de HOG para regiões em imagens.....	7
Figura 2.3 – Cálculo de Imagem HOG	8
Figura 3.1 – Imagem de satélite com eventos extremos	10
Figura 3.2 – Segmentação dos padrões de eventos	10
Figura 3.3 – Imagens negativas para não-evento.....	11
Figura 3.4 – Imagens artificiais para treinamento.....	11
Figura 5.1 – Curvas de aprendizado para o algoritmo SGD.....	14

SUMÁRIO

	<u>Pág.</u>
1 INTRODUÇÃO	1
2 FUNDAMENTAÇÃO TEÓRICA.....	3
2.1. ALGORITMO SGD.....	3
2.2. EXTRAÇÃO DE CARACTERÍSTICAS UTILIZANDO O HOG.....	7
3 DESCRIÇÃO DAS IMAGENS DE SATÉLITE	9
4 CLASSIFICADOR DE EVENTOS EXTREMOS.....	12
5 RESULTADOS E CONCLUSÕES.....	13
REFERÊNCIAS	15

1 INTRODUÇÃO

Eventos climáticos, principalmente os considerados extremos, como precipitações de grandes proporções, ciclones e tornados, fortes períodos de seca, inundações, são determinantes no comportamento econômico e social, pois eventos desta proporção geram prejuízos em grande escala. As mudanças climáticas são um fator determinante para o aumento da frequência de eventos extremos, identificar e emitir alertas em um curto período de tempo para que ações que reduzam os efeitos destes eventos sejam tomadas de melhor maneira é um desafio. O Instituto Nacional de Pesquisas Espaciais tem como uma de suas atividades, promover a análise e o monitoramento do clima, identificando também eventos desta magnitude para tomada de decisões.

A análise de dados em observação da terra pode ser feita de maneira visual indicando nas imagens de satélites regiões propensas a ocorrência de eventos extremos. Entretanto, avaliar um grande volume de dados gerados diariamente é uma tarefa não-trivial, ainda mais, quando é necessária a avaliação de especialistas para indicar qual tipo de evento ocorre em determinada região. Desta forma, o emprego de técnicas em *machine learning* torna-se justificável, pois modelos baseados no aprendizado de máquina podem reduzir a complexidade de avaliação e análise em problemas não lineares. Estes modelos podem ainda ser aplicados a um grande conjunto de dados e gerar avaliações em tempo reduzido.

O emprego de técnicas em *Machine Learning* para resolução ou aproximação de problemas transcorre pela escolha de paradigmas adequados e pela forma em que o conhecimento sobre determinado problema pode ser modelado computacionalmente. Os conceitos que envolvem esta área da computação indicam que não apenas escolher determinados paradigmas ou algoritmos pode ser suficiente para se desenvolver soluções adequadas e que gerem resultados aceitáveis, existe certa tendência ao experimento de diversas técnicas e a combinação das mesmas. O uso de técnicas de inteligência computacional unidas a outras técnicas, tais como, métodos numéricos, não é restrito, tais abordagens são explicitamente recomendadas para problemas não lineares e de difícil solução.

Neste contexto, o seguinte trabalho propõe uma avaliação de eventos extremos em imagens de satélite baseada em modelos de aprendizado de máquina para avaliar em um

curto período de tempo um grande volume de informações e inferir a ocorrência ou não de eventos extremos, tais como tempestades, ciclones, tornados e furacões. Os modelos gerados neste trabalho se baseiam no algoritmo de descida gradiente estocástica (*Stochastic Gradient Descent - SGD*) para o treinamento e aprendizado em padrões de imagens de satélite. As imagens utilizadas neste modelo foram obtidas do satélite landsat-8 e do cbers, com resoluções variadas. Para extração de características nas imagens foi utilizado o histograma de gradientes orientados (*Histogram of Oriented Gradients - HOG*), este algoritmo permite definir um vetor de características pela concatenação de histogramas normalizados dos gradientes presentes em uma janela de detecção.

O trabalho está dividido em cinco seções. Sendo nesta primeira, explicados os objetivos e a justificativa da pesquisa. Na seção 2 são definidos conceitos básicos sobre o SGD e sobre a extração de características em imagens utilizando o HOG. Na seção 3 são descritos os dados e as imagens utilizadas para o treinamento e teste do modelo. Na seção 4 é descrita a abordagem desenvolvida. Por fim, na seção 5 são indicados os resultados e conclusões tomadas da abordagem proposta.

2 FUNDAMENTAÇÃO TEÓRICA

2.1. ALGORITMO SGD

As principais técnicas de aprendizado de máquina são baseadas em redes neurais e modelos estatísticos tais como: Redes Perceptron de Múltiplas Camadas (MLP); Redes Kohonen; Modelos de Máxima Entropia (MaxEnt); Support Vector Machine (SVM); Regressão logística; Regressão de Kernel, etc. Estes modelos tem sido empregados na solução de problemas complexos e com restrições bem definidas. Estes algoritmos de aprendizado, são considerados como métodos relativamente simples de otimização, que têm como característica a lentidão na convergência para os pontos de mínimos ou máximos globais podendo em determinados casos convergir para um mínimo ou máximo local, gerando um resultado de falso positivo. Esta característica de convergência é explícita em curvas de nível do tipo alongadas. Neste sentido é clara a necessidade de adoção de técnicas que possibilitem a otimização destes algoritmos, uma delas é o algoritmo SGD.

Sendo algumas das principais ferramentas na solução de problemas não lineares, os métodos baseados em gradientes são clássicos exemplos de métodos iterativos. Um método iterativo propicia uma sequência de soluções aproximadas de uma solução ótima, em que cada uma das soluções são obtidas pelo mesmo processo que é repetido (iterativamente), um método iterativo é estacionário se cada solução da sequência de soluções é obtida do anterior sempre pelo mesmo processo, métodos iterativos também podem ser cíclicos (ATKINSON; HAN, 1985).

Um exemplo de problema em que métodos iterativos são empregados é a busca por mínimos e máximos globais de funções, por meio da determinação de raízes que podem ser reais ou complexas. Os métodos baseados em gradiente utilizam informações do gradiente de funções objetivo, com o intuito de direcionar a busca por estes mínimos ou máximos e definir a convergência ou não para solução ótima (FLETCHER, 1987).

Alguns critérios matemáticos são empregados em métodos iterativos e compreendem os seguintes itens:

- Dada uma aproximação inicial aleatória ou criteriosa, o método iterativo deve possuir um número finito de iterações para convergir ou não para a solução do problema, ou ao menos gerar uma solução aproximada.
- Um critério geral para a convergência é definido para cada tipo de método.

- Deve haver um critério de parada do método. Geralmente este critério é dado quando a solução é menor ou igual a um erro relativo pré-estabelecido.
- Um critério de parada genérico é dado pela seguinte equação 2.1:

$$\frac{\|x^{k+1} - x^k\|_{\infty}}{\|x^{k+1}\|_{\infty}} < \textit{erro relativo} \quad (2.1)$$

Onde x^{k+1} equivale a aproximação encontrada a partir da condição inicial x^k .

Dentro deste contexto, o SGD é definido como um método eficiente para a aprendizagem de maneira discriminativa, empregado em classificadores lineares com funções de perda convexas, como o SVM (*Support Vector Machine*) linear e a Regressão logística. Este algoritmo é amplamente aplicado no contexto de identificação de caracteres e processamento de linguagem natural (LIN et al., 2011; DAHL et al., 2012).

Os métodos baseados em descida gradiente como a máxima descida e os gradientes conjugados têm como característica seguir uma condição inicial e a partir disso executar as iterações até que a função objetivo seja atingida, ou seja, os pontos próximos ao de máximo ou de mínimo global sejam encontrados. No método da máxima descida essa direção de busca é dada pela direção contrária a direção do vetor gradiente da função objetivo. Já no método dos gradientes conjugados, a direção de busca utiliza o conceito de direções conjugadas, obtidas a partir do vetor gradiente da função objetivo na iteração atual e iterações anteriores. Ambos os algoritmos são recursivos, cujo objetivo é gerar trajetórias contínuas que sejam capazes de escapar de mínimos locais para se estacionarem em um mínimo global da função ou sistema de funções não lineares.

No sentido de minimizar o número de iterações do método e gerar a otimização para a convergência o método SGD propõe que, dada uma aproximação inicial aleatória da função em até n iterações e seguindo uma direção de busca é possível encontrar os pontos de objetivo (mínimos e máximos) (FLETCHER; REEVES, 1964). O objetivo deste algoritmo de aprendizagem supervisionada é estimar uma função objetivo f que mapeia os dados de treinamento X para as variáveis de saída Y .

O algoritmo SGD é executado da seguinte maneira (SCIKIT-LEARN, 2018):

- Dado um conjunto de exemplos de treinamento $(x_1, y_1), \dots, (x_n, y_n)$ onde $x_i \in \mathbf{R}^m$ e suas respectivas variáveis de saída (classes) $y_i \in \{-1, 1\}$ (caracteriza a classificação binária), o objetivo é encontrar os pontos objetivo para a função $f(x) = w^T x + b$.
- As predições (classificações) são identificadas pelo sinal de $f(x)$, para as duas classes definidas (-1, 1), falso e verdadeiro. Uma escolha comum para encontrar os parâmetros do modelo é minimizar o erro de treinamento. Dado pela equação 2.2.

$$E(w, b) = \frac{1}{n} \sum_{i=1}^n L(y_i, f(x_i)) + \alpha R(w) \quad (2.2)$$

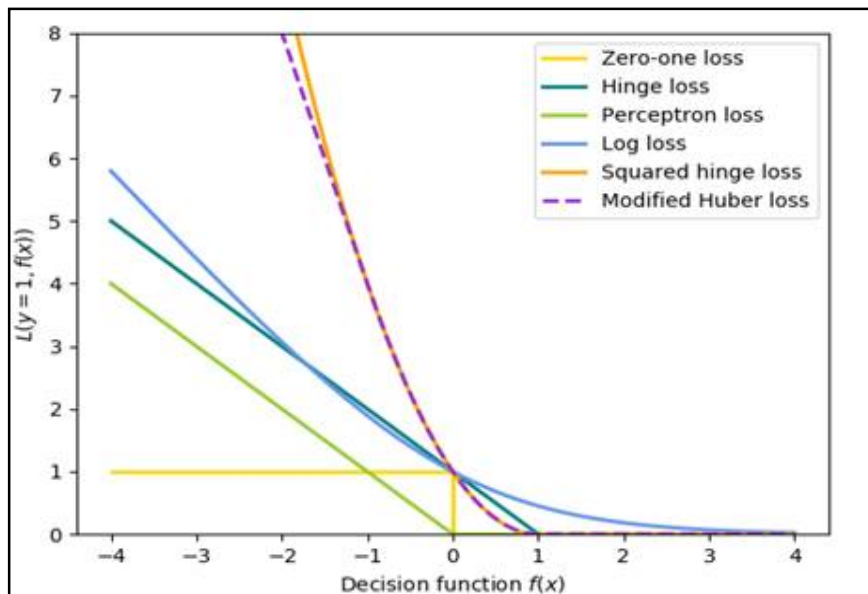
- Onde L é a função de custo que mede o ajuste do modelo, R é o viés de penalização.
- As funções de custo podem ser caracterizadas como:
 - Hinge: Máquinas de vetores de suporte
 - Log: Regressão Logística
 - Perceptron
 - Modified Huber

O algoritmo SGD é definido da seguinte maneira:

- Escolha um ponto inicial, ou seja, inicialize os pesos (w)
 - Repita até obter o mínimo ou máximo aproximado:
 - Embaralhe os exemplos do conjunto de treinamento a cada época de treinamento
 - Para $i = 1 \dots n$, faça:
 - $W = w - \alpha R(w)$

Neste trabalho foram testadas as funções custo *Hinge*, *Log*, *Least-Square* e *Epsilon-Intesitive* para o treinamento do modelo. A figura 2.1 ilustra estas funções.

Figura 2.1 – Funções de custo utilizadas



Fonte: SCIKIT-LEARN (2018)

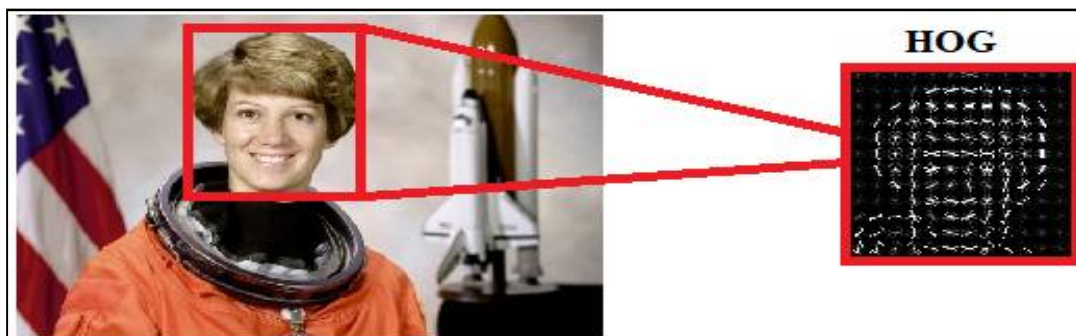
2.2. EXTRAÇÃO DE CARACTERÍSTICAS UTILIZANDO O HOG

Este trabalho visa a partir de imagens de satélite definir a presença ou não de eventos extremos como furacões, tempestades, etc. Uma das maneiras de avaliar as principais características de imagens é o Histograma de Gradientes Orientados (HOG), esta técnica permite que se obtenham apenas informações importantes das imagens diminuindo a complexidade de análise (DAVID, 2004).

Esta técnica foi desenvolvida por Dalal & Triggs (2005), para detecção de objetos em imagens principalmente voltadas ao reconhecimento facial. A idéia fundamental do HOG é identificar as características de borda e formatos nas imagens através da distribuição dos gradientes de intensidade dos pixels e pelas direções das bordas.

Para o calculo do HOG uma imagen é dividida em blocos, ou regiões, por exemplo, em matrizes de 16x16 pixels. Para cada um destes blocos a magnitude do gradiente e um número de direções é calculado. A figura 2.2 indica como é calculado o HOG para um bloco de uma imagem.

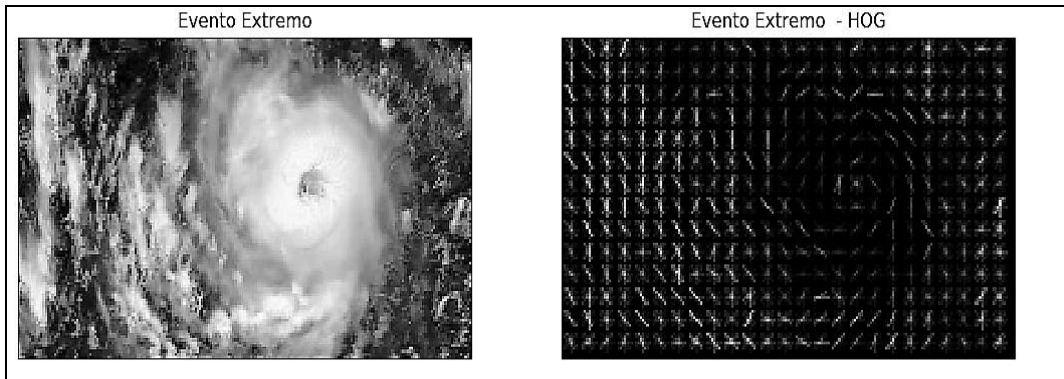
Figura 2.2 – Cálculo de HOG para regiões em imagens



Fonte:Produção dos autores

Este tipo de extração de características em imagens é útil para a observação apenas de objetos e formas relevantes nas imagens. O HOG transforma uma imagem RGB em um vetor de características (a distribuição e as direções dos gradientes na imagem). Os gradientes (derivadas de x e y) são uteis pois a magnitude dos gradientes é elevada em torno das bordas dos objetos, desta forma estas bordas possuem mais informações sobre objetos presentes na imagem. A figura 2.3 ilustra o cálculo do HOG para uma imagem de satélite de um evento extremo, evidenciando as características de forma do objeto (evento) encontrado.

Figura 2.3 – Cálculo de Imagem HOG



Fonte: produção dos autores.

Desta forma, com os vetores HOG calculados para as imagens é possível gerar um conjunto de treinamento (X) e definir as saídas (Y) de não-evento e evento $(-1,1)$. Desta maneira é possível executar um treinamento com o algoritmo SGD e realizar as predições da presença ou ausência destes eventos nas imagens de satélite.

3 DESCRIÇÃO DAS IMAGENS DE SATÉLITE

As imagens utilizadas neste trabalho são categorizadas como imagens de satélite que contém eventos extremos e imagens sem estes eventos. Foram pesquisadas diversas bases de dados e selecionadas imagens do acervo do INPE dos satélites LANDSAT-8 e CEBERS, e do acervo do NOAA na identificação dos eventos climáticos.

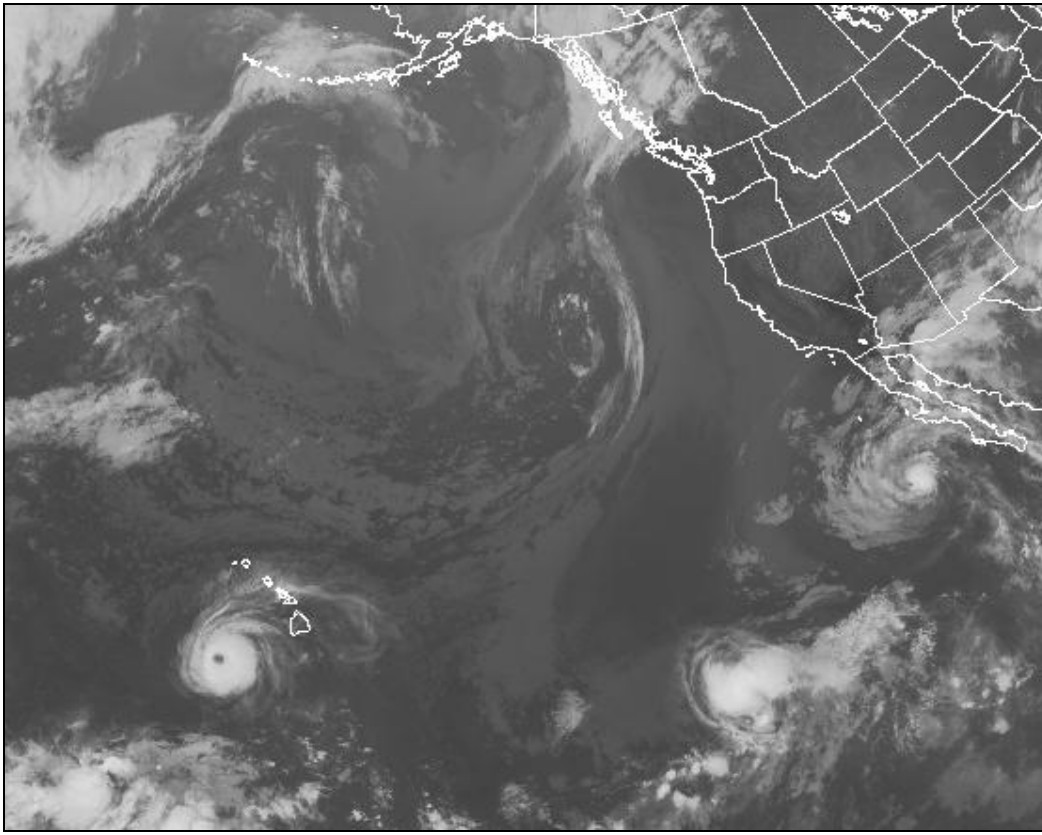
Assim como o INPE, o NOAA (*National Oceanic And Atmospheric Administration*) gerencia bases de dados com diversas imagens que caracterizam eventos meteorológicos de forma a identificar, analisar e monitorar eventos climáticos para possíveis alertas e tomadas de decisão (INPE, 2018; NOAA, 2018).

O NOAA usa um sistema de biblioteca on-line chamado de Sistema de Administração de Dados de Matriz Ampla Compreensiva (Comprehensive Large Array-data Stewardship System - CLASS) para armazenar uma infinidade de dados ambientais. Os dados provêm do Satélite Operacional Ambiental (POES) do Departamento de Defesa dos EUA (DoD), do Satélite Ambiental Operacional Geoestacionário da NOAA (GOES) e de outros dados derivados.

Atualmente, os Data Centers Nacionais da NOAA suportam dados POES, DMSP, GOES, MetOp, Jason-2 e selecionam dados de reanálise de modelos. Estes sistemas arquivam coleções de dados do NPP, JPSS (anteriormente NPOESS), GOES-R, Jason-3, e os sistemas de observação baseados na Terra planejados incluem produtos NEXRAD.

As imagens obtidas neste trabalho para identificação de eventos extremos possuem uma resolução espacial variando entre 10 km a 20 km no canal infravermelho, Banda 4 infravermelho termal. A figura 3.1 ilustra um exemplo de imagem positiva, em que eventos climáticos extremos são encontrados.

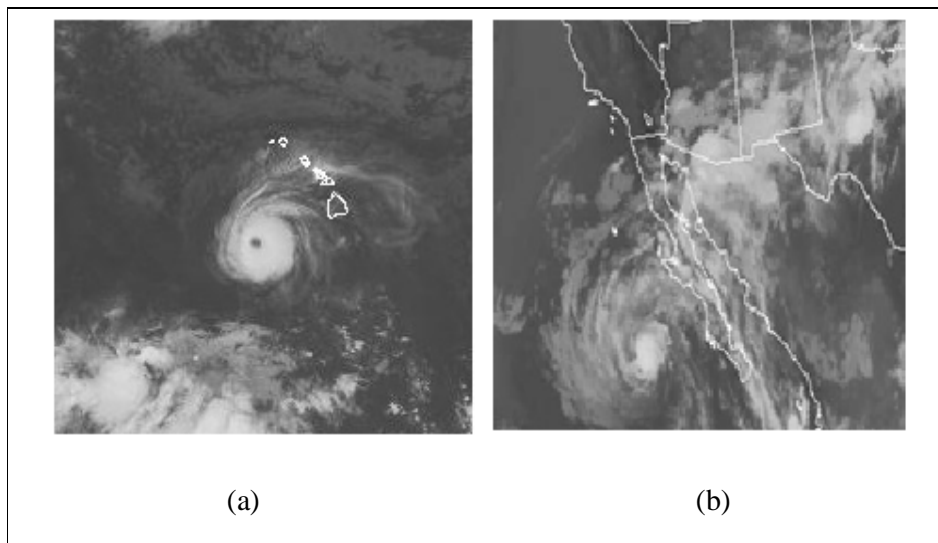
Figura 3.1 – Imagem de satélite com eventos extremos



Fonte: produção dos autores.

Para gerar um vetor de características que fosse capaz de identificar em uma imagem os eventos em regiões isoladas as imagens foram segmentadas em blocos com resoluções variadas para identificar estes padrões, conforme ilustra a figura 3.2.

Figura 3.2 – Segmentação dos padrões de eventos



Fonte: produção dos autores.

Para a identificação das imagens negativas, ou seja, imagens sem a ocorrência de eventos extremos, foram coletadas imagens do acervo do INPE com resoluções espaciais variando entre 10 km a 16 km com a cobertura de nuvens mínima, evitando a análise de um falso positivo. A figura 3.3 ilustra algumas das imagens sem a ocorrência de eventos.

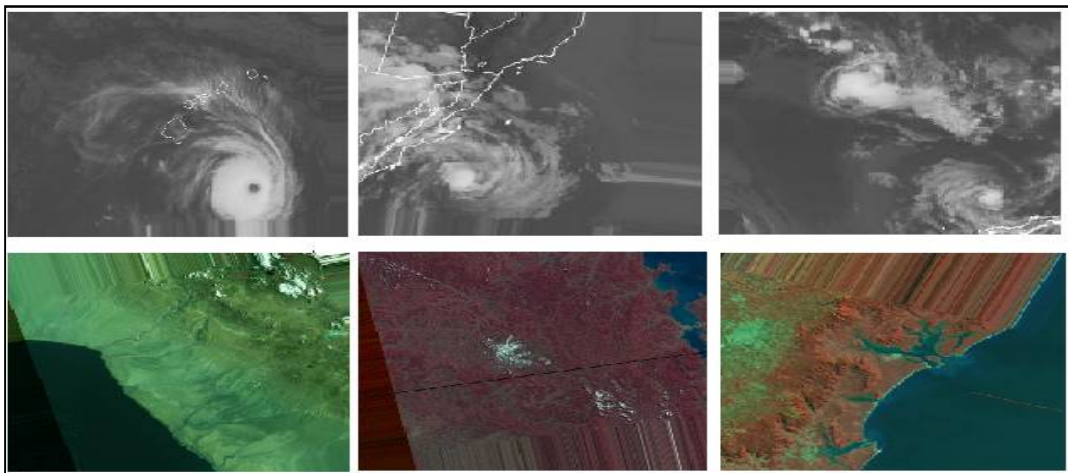
Figura 3.3 – Imagens negativas para não-evento



Fonte: produção dos autores.

Assim como a maioria dos algoritmos de *machine learning*, para executar uma boa predição ou classificação é necessária uma análise e aprendizado sobre um grande volume de dados, neste caso, uma grande quantidade de imagens. Desta forma, foram geradas imagens artificiais para o treinamento do modelo. Estas imagens artificiais variam em rotação, inclinação e posição de forma a gerar sobre uma imagem original diversas variações da mesma. A figura 3.4 ilustra algumas imagens artificiais geradas neste trabalho, tanto para imagens positivas como para negativas.

Figura 3.4 – Imagens artificiais para treinamento



Fonte: produção dos autores.

Foram geradas um total de 1810 (50% para evento e 50% para não-evento) imagens para o treinamento e teste do modelo de aprendizado computacional.

4 CLASSIFICADOR DE EVENTOS EXTREMOS

O classificador de eventos extremos desenvolvido neste trabalho, busca identificar em imagens de resoluções variadas a ocorrência ou não de ciclones, tempestades, precipitações, etc.

As 1810 imagens de eventos foram submetidas a um pré-processamento de dados que visou normalizar todas as amostras. Esta normalização compreendeu redimensionar todas as imagens em um único padrão de 300x300. Após esta normalização todos os dados foram submetidos a transformação do canal RGB para escala de cinza. Desta forma todas as diferenças marcantes nas imagens estariam voltadas a presença ou não de eventos extremos.

Após esta normalização foram calculados os vetores de características utilizando o algoritmo HOG. O HOG utilizado nas imagens foi dividido em blocos de 12x12 pixels gerando assim os gradientes para cada bloco nas imagens. Após a normalização das imagens os dados foram gravados em um Dataframe Python, utilizando pandas para a melhor manipulação das informações. As informações gravadas continham:

- Tipo de Evento {Evento, Não-Evento}
- Vetor RGB da Imagem Original de 3 dimensões {altura, largura, profundidade}
- Vetor de características HOG de 2 Dimensões {As coordenadas e intensidades dos gradientes}

Os conjuntos de treinamento e testes foram preparados conforme é indicado na literatura para 80% dos dados para treinamento e 20% para testes. É importante salientar que estas amostras são balanceadas no que se refere a quantidade de exemplos nos conjuntos de forma a evitar o *overfitting* e o *underfitting*.

O conjunto de treinamento foi inserido no modelo SGD (explicado na seção 2) para realizar a predição e classificação binária dos eventos. O modelo SGD foi testado para as funções de custo Hinge, Log, modified Huber e perceptrons. O viés de penalização é descrito na equação 4.1.

$$R(w) := \frac{1}{2} \sum_{i=1}^n w_i^2 \quad (4.1)$$

O modelo foi desenvolvido em linguagem de programação Python. Os termos relacionados a custo computacional não foram computados neste trabalho pois o

treinamento foi eficiente e pouco custoso computacionalmente, ou seja, foram necessários poucos segundos para executar 1000 épocas de treinamento testando cada uma das funções de custo.

5 RESULTADOS E CONCLUSÕES

Os resultados obtidos com a abordagem são descritos na tabela 5.1. Os testes foram feitos para as funções de custo citadas, desta forma foram generalizadas 362 imagens sendo 50% para eventos e 50% para não-eventos.

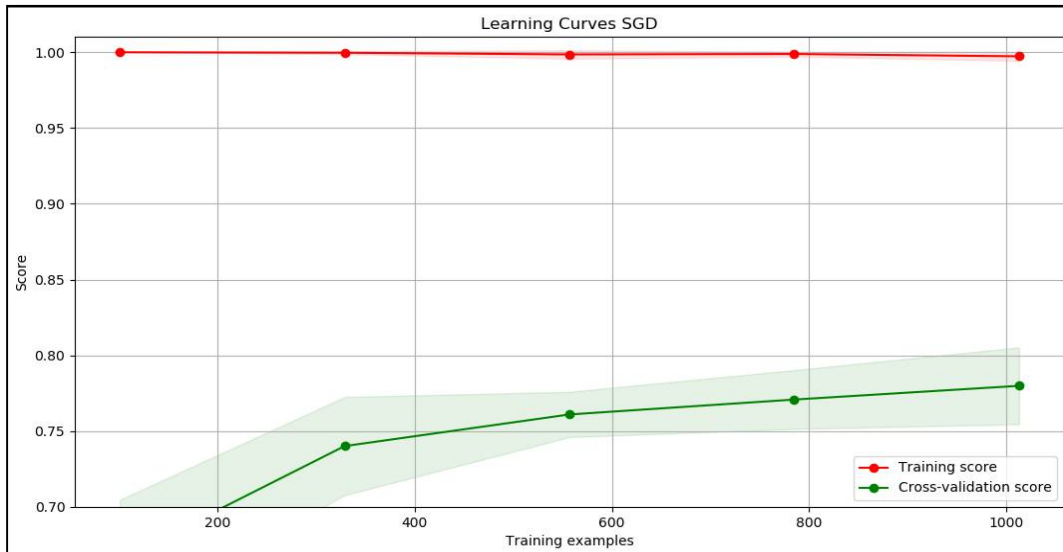
Tabela 5.1 – Resultados de Classificação
Classificação de eventos Extremos em imagens de Satélite

Hinge (%)	Log (%)	Modified Huber (%)	Perceptron (%)
82.13	81.40	81.40	82.13

Fonte: Produção dos autores.

As figuras 5.1 indica o comportamento do modelo no treinamento e na validação respectivamente. A curva revela que o modelo generaliza o conjunto de treinamento em quase 99% das amostras, entretanto pelo número de épocas de treinamento os dados de teste e validação encontram-se em média 82%, ou seja, o modelo generalizou para amostras que não foram treinadas a classificação correta de 82% das imagens de satélite.

Figura 5.1 – Curvas de aprendizado para o algoritmo SGD



Fonte: Produção dos autores.

É importante ressaltar que a proposta apresentada neste trabalho não é definitiva, a generalização em 82% para correta classificação dos eventos extremos é uma aproximação inicial. A estrutura gerada neste trabalho permite alternativas a otimização e variação dos modelos de *machine learning* para o treinamento e testes de classificação. A curva apresentada na figura 5.1 indica que houve um crescimento na validação do modelo conforme as épocas de treinamento avançam, ou seja, é possível o aumento da precisão para um número maior de épocas.

Outro ponto importante é a qualidade das amostras obtidas. Pela dificuldade em encontrar imagens padronizadas que representassem os eventos extremos foram geradas imagens artificiais para o treinamento do modelo. Estas transformações influem no desempenho do modelo já que as variações existentes são propagadas como erro, desta forma, é importante executar um treinamento para uma grande quantidade de dados originais.

Devido a abordagem proporcionada este modelo inicial cumpre os requisitos solicitados em relação a geração de uma estrutura computacional, com o custo de tempo reduzido para a identificação de eventos. Espera-se que com o aprimoramento deste modelo inicial seja possível identificar com acurácia e precisão elevada os eventos extremos presentes em imagens de satélite para que tomadas de decisão sejam feitas em tempo reduzido permitindo a otimização e melhor desempenho do processo.

REFERÊNCIAS

- ATKINSON, K. E.; HAN, W. **Elementary numerical analysis**. New York et al.: Wiley, 1985.
- DAHL, G. E. et al. Context-dependent pre-trained deep neural networks for large-vocabulary speech recognition. **IEEE Transactions on audio, speech, and language processing**, v. 20, n. 1, p. 30-42, 2012.
- DALAL, N.; TRIGGS, B. Histograms of Oriented Gradients for Human Detection. 2005. **IEEE Computer Society Conference on Computer Vision and Pattern Recognition**, 2005, San Diego, CA, USA.
- DAVID, G. L. Distinctive image features from scale-invariant keypoints. 2004. **International Journal of Computer Vision**, 60, 2 (2004), pp. 91-110.
- FLETCHER, R. **Practical Methods of Optimization**. 2. ed. University of Michigan: John Wiley and Sons, 1987. 1, 11.
- FLETCHER, R.; REEVES, C. M. **Function minimization by conjugate gradients**. *Computer Journal*, v. 7, p. 149-154, 1964. 12, 13.
- INPE – INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS. Catálogo de Imagens. 2018. Disponível em:<<http://www.dgi.inpe.br/CDSR/>> Acesso em: 15 de julho de 2018.
- LIN, Y. et al. Large-scale image classification: fast feature extraction and svm training. In: **Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on**. IEEE, 2011. p. 1689-1696.
- LUENBERGER, David G. et al. **Linear and nonlinear programming**. Reading, MA: Addison-wesley, 1984.
- NOAA - NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION. Geostationary Satellite Server. 2018. Disponível em: <<http://www.goes.noaa.gov/browse.html>> Acesso em: 15 de julho de 2018.
- SCIKIT-LEARN. **1.5. Stochastic Gradient Descent**. 2018. Disponível em: <<http://scikit-learn.org/stable/modules/sgd.html>> Acesso em: 3 de Julho de 2018.