

Proceeding Series of the Brazilian Society of Computational and Applied Mathematics

Modelagem do risco e perigo de incêndios florestais utilizando classificador e sistema fuzzy

Isaac D. B. Silva¹

Universidade Estadual de Campinas, UNICAMP, Campinas, SP

Laécio C. Barros²

Universidade Estadual de Campinas, UNICAMP, Campinas, SP

João F. C. A. Meyer³

Universidade Estadual de Campinas, UNICAMP, Campinas, SP

Resumo. Este trabalho visa modelar o perigo e o risco de incêndios florestais utilizando duas ferramentas matemáticas distintas: um sistema dinâmico fuzzy e um classificador do tipo k-NN (k-NN é o acrônimo de k-Nearest Neighbors). As variáveis de entrada do primeiro são umidade relativa do ar e precipitação pluvial e a saída gera uma série temporal que modela o perigo de incêndios. O segundo tem como entradas as variáveis geográficas: altitude, tipologia florestal e as distâncias para curso d'água e para estradas mais próximos enquanto a saída é a probabilidade condicional da classe ω_1 (que representa a presença de focos de calor), dado que o padrão de teste x_t foi classificado. O estudo é realizado para o Estado do Acre com informações geo-referenciadas dos anos de 2003 a 2014. A acurácia do classificador k-NN foi de 84,3% junto a um conjunto de teste e apenas para um percentual de 4,52% de focos de calor ocorridos em 2014 foi atribuído risco zero. A estimativa do risco temporal para o ano de 2014 foi avaliada por meio de uma classificação binária, na qual obteve-se acurácia de 73,97% com área abaixo da curva ROC de 87,94%.

Palavras-chave. Incêndios, Classificadores, Lógica Fuzzy.

1 Introdução

A floresta amazônica possui grande importância não só para a população local, mas também para o restante do mundo. Os incêndios florestais que ocorrem na Amazônia como um todo, são o resultado de severas secas e das práticas de corte-queima para a limpeza de roçados que muitas vezes se tornam incontroláveis. O impacto ambiental gerado pelas queimadas atinge a qualidade do solo e do ar e, conseqüentemente, a biodiversidade é alterada [1].

A literatura diferencia os conceitos de *índice de risco* e *perigo de incêndios*. Enquanto o primeiro está associado à probabilidade de um incêndio se iniciar em função da presença e/ou atividades de agentes causadores (nesse trabalho: altitude, tipologia florestal,

¹isaacdb@yahoo.com.br

²laeciob@ime.unicamp.br

³joni@ime.unicamp.br

distâncias para estrada e curso d'água); o segundo relaciona-se às características do material combustível (caráter permanente) e condições climáticas (caráter variável) - nesse trabalho foram utilizadas apenas umidade relativa do ar e precipitação pluvial - que pre-dispõem à ignição do incêndio ou representam dificuldade para a extinção [2].

O objetivo central deste trabalho é modelar o risco e o perigo de incêndios: o perigo de incêndio vai ser caracterizado apenas de fatores climáticos e que denominaremos de risco temporal R_t e o outro depende apenas de variáveis espaciais e será denominado risco espacial R_e . Foi elaborado um estudo de caso para o Estado do Acre com dados geo-referenciados dos anos de 2003 a 2014.

2 Risco de incêndio temporal e sistema fuzzy

Seja U um conjunto clássico; um subconjunto fuzzy F de U é caracterizado por uma função $\varphi_F : U \rightarrow [0, 1]$, denominada função de pertinência de F , associando a cada $x \in U$ o grau de pertinência $\varphi_F(x)$ de x em F [3].

Na teoria de conjuntos fuzzy, as variáveis linguísticas têm papel importante. Variável linguística é um substantivo e seus valores são adjetivos. Os conjuntos fuzzy representam os estados (adjetivos) da variável linguística que, em geral, são expressos por termos subjetivos como *grande*, *pouco*, *baixo*, entre outros e modelados por conjuntos fuzzy [4].

Um sistema baseado em regras fuzzy, essencialmente, possui quatro componentes principais: um processador de entrada (ou fuzzificador), uma base de regras fuzzy, um método de inferência fuzzy e um processador de saída (ou defuzzificador), gerando um número real como sua saída [5].

O método de inferência utilizado neste trabalho é o de Mamdani e a defuzzificação foi realizada através do método do Centro de Gravidade [3].

O sistema proposto neste trabalho foi baseado no sistema dinâmico p-fuzzy estudado em [6], e dado por

$$R_{t+1} = \max(0, R_t + \Delta(R_t)) \tag{1}$$

onde R_t é o risco temporal no instante t , $\Delta(R_t) = f(U_t, P_t)$ é a variação do risco e f é obtida por um sistema baseado em regras fuzzy, no qual as variáveis linguísticas de entrada são a umidade relativa do ar U_t e precipitação pluvial total diária P_t no instante t , como pode ser visualizado no diagrama da Figura 1.

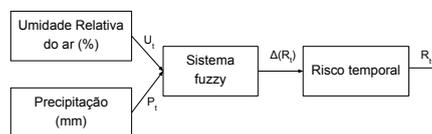


Figura 1: Diagrama em blocos do sistema fuzzy para o risco temporal

O risco temporal definido dessa forma é um valor real e positivo para cada t . A seguir, encontra-se estudo para o Estado do Acre a partir desse modelo.

2.1 Estudo de Caso 1: Estado do Acre

As variáveis linguísticas de entrada *umidade relativa do ar* e *precipitação pluvial* assumem as classificações: MUITO BAIXA, BAIXA, MÉDIA, ALTA, MUITO ALTA.

Tanto os suportes dos conjuntos fuzzy para as variáveis linguísticas quanto a base de regras do sistema fuzzy foram ajustados a partir de informações fornecidas por especialistas e das informações de umidade, precipitação e focos de calor da cidade de Rio Branco nos anos de 2003 a 2013 [7, 8]. Por simplificação e ausência de um número considerável de estações meteorológicas (o Estado possui apenas três estações meteorológicas registradas em [7]), os dados climáticos da capital Rio Branco foram considerados para o Estado do Acre inteiro.

A simulação realizada para validar o ajuste realizado considerou os dados climáticos e de focos de calor de Rio Branco-AC para os 365 dias do ano de 2014. Após a simulação, a série temporal gerada foi normalizada (norma do máximo) para que pudesse ser realizada a comparação com o percentual de focos de calor ocorridos no mesmo período. A Figura 2 ilustra esse fato.

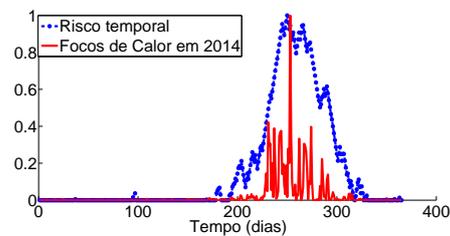


Figura 2: Focos de calor e risco temporal para o ano de 2014.

Para avaliar a qualidade da estimativa do risco temporal, considerou-se os resultados obtidos como uma classificação binária, na qual, para cada instante de tempo t , R_t pertence à classe 1 (ω_1) se há ocorrência de foco de calor no dia t e se não há ocorrência de foco de calor no dia t , R_t pertence à classe 0 (ω_0). O verdadeiro positivo (vp) é caracterizado quando R_t pertence à classe ω_1 e é positivo. Da mesma forma, o verdadeiro negativo (vn) ocorre quando R_t pertence à classe ω_0 e é nulo. Para o falso negativo (fn) temos $R_t = 0$ e $R_t \in \omega_1$ e o falso positivo (fp) possui a característica de $R_t > 0$ e $R_t \in \omega_0$. A Tabela 1 resume essas ideias.

Tabela 1: Classificação utilizada para medir a eficiência do risco temporal

Classificação	Risco temporal (R_t)	Classe de R_t
Verdadeiro positivo (vp)	positivo	ω_1
Verdadeiro negativo (vn)	nulo	ω_0
Falso positivo (fp)	positivo	ω_0
Falso negativo (fn)	nulo	ω_1

Com essa metodologia, após serem preditos os riscos temporais para o ano de 2014, obteve-se acurácia=73,97% e a área abaixo da curva ROC é 0,8794. As curvas ROC

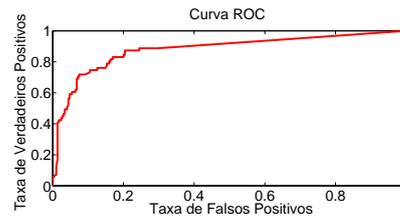


Figura 3: Curva ROC para o risco temporal

(do inglês *Receiver Operator Characteristic*) são comumente usadas para apresentar os resultados de problemas de decisão binária em aprendizado de máquina [9]. Quanto mais próximas de 1 (um) essas duas medidas, melhor o desempenho do classificador. A Figura 3 apresenta a curva ROC obtida.

3 O risco espacial e o problema de classificação

Outro estudo de interesse foi obter um mapeamento R_e que associou variáveis espaciais de um domínio $\Omega \subset \mathbb{R}^n$ ao conjunto imagem \mathbb{R} tal que para cada $x \in \mathbb{R}^n$, $y = R_e(x) \in [0, 1]$ é o risco de incêndio espacial associado a x . Para atingir esse objetivo, o problema de obter esse mapeamento foi reformulado para um problema de classificação binária e, a partir de informações geo-referenciadas sobre as variáveis independentes e dependentes, utilizou-se de técnicas de aprendizado computacional para obter um processo de classificação eficiente e o risco espacial foi definido como a probabilidade condicional da classe ω_1 (presença de focos de calor) dado que o padrão x_t foi classificado.

As variáveis de entrada escolhidas para o classificador são: altitude, tipologia, distância para curso d'água e distância para a estrada mais próximos, todas geo-referenciadas. A variável de saída é binária e representa a presença (classe ω_1) ou não (classe ω_0) de foco na célula considerada.

O diagrama da Figura 4 ilustra o modelo do classificador proposto.

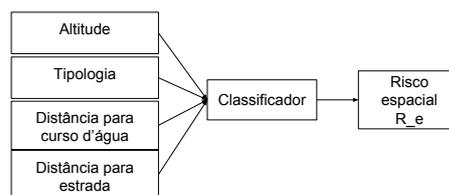


Figura 4: Diagrama em blocos do classificador

Após a separação dos dados em conjunto de treinamento e teste, realizou-se o treinamento com o classificador escolhido e averiguou-se o desempenho do mesmo a partir do cálculo do erro junto ao conjunto de teste. Quanto menor o erro de classificação junto ao conjunto de teste, melhor será a estimativa do risco espacial. Efetuou-se a predição da classe para o conjunto completo dos dados que recobrem o domínio Ω , e, assim, obteve-se um mapeamento geográfico do risco espacial, no qual, para cada ponto da discretização do

domínio associou-se a probabilidade condicional em relação à classe ω_1 dado que o ponto foi classificado, o que sugere a seguinte definição.

Define-se aqui o risco incêndio espacial pela Fórmula:

$$R_e = P(\omega_1|x_t) \quad (2)$$

onde $P(\omega_1|x_t)$ é a probabilidade condicional da classe ser ω_1 dado que o padrão x_t foi classificado. A probabilidade condicional da Fórmula 2 pode ser calculada pelo teorema de Bayes [10].

3.1 Estudo de Caso 2: Estado do Acre

Com a metodologia da seção anterior, foi obtido o risco espacial para o Estado do Acre. Os mapas geo-referenciados foram obtidos a partir do Zoneamento Ecológico Econômico do Estado do Acre fornecidos pela Fundação de Tecnologia do Estado do Acre - FUNTAC. O banco de dados que originou os conjuntos de treinamento e validação foi obtido a partir do cruzamento de informações geo-referenciadas de uma malha de retângulos (células) que recobre o Estado do Acre com focos de calor ocorridos nesse Estado no período de 2003 a 2013 [1].

Ao realizar testes de classificação com várias técnicas, o classificador que apresentou melhor desempenho, considerando os dados disponíveis, foi o classificador k-vizinhos mais próximos (k-NN do acrônimo k-Nearest Neighbors). Esse classificador possui aprendizagem baseada em memória e identifica a classe de um determinado vetor de teste x_t a partir do voto majoritário dos k vizinhos mais próximos de x_t [11].

Em virtude da ocorrência de focos de calor ser diferente entre o leste e oeste do Estado do Acre, houve a necessidade de separar os dados (conjuntos Leste e Oeste) antes do treinamento ser realizado. Para cada um desses dois conjuntos, separou-se os dados em dois conjuntos: 85% dos dados para o conjunto de treinamento e 15% para o de teste.

Para estimar o valor ótimo para k utilizou-se, com os dados de treinamento, a técnica de validação cruzada com dez pastas (10-fold cross validation) da seguinte forma: para cada valor de k ($k = 1, \dots, 50$), obteve-se o Hamming Loss de validação médio entre os 10 treinamentos. O valor de k escolhido é aquele que corresponde ao menor valor de Hamming Loss médio. Esse procedimento foi repetido para os dois conjuntos (Leste e Oeste). Os valor ótimo de k_d (nº de vizinhos para o conjunto direita) foi igual a 21 com um Hamming Loss médio de 16,74% e de k_e (nº de vizinhos para o conjunto esquerda) foi igual a 11 com um Hamming Loss médio de 12,44%.

O risco incêndio espacial é a probabilidade condicional da classe ω_1 (que representa a presença de focos de calor), dado que o padrão de teste x_t foi classificado, que para o classificador k-NN, é dada pela fórmula:

$$R_e = P(\omega_1|x_t) = \frac{n_{viz}}{k}, \quad (3)$$

onde n_{viz} é o número de vizinhos do padrão x_t que estão na classe ω_1 e k é o número total de vizinhos mais próximos [10].

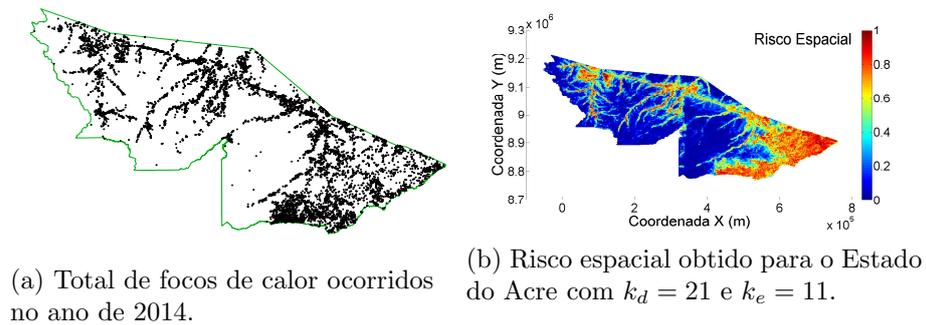


Figura 5: Mapa de ocorrência de focos de calor e a predição de risco espacial para o ano de 2014.

Obtidos os valores de k_e e k_d ótimos, eles foram utilizados para prever as classes dos dados do conjunto de teste. O Hamming Loss calculado junto a esse conjunto foi de 15,70%. Como a taxa de acerto (acurácia) do classificador k-NN foi em torno de 84,30%, pode-se perceber que dadas as condições espaciais de qualquer ponto do domínio, consegue-se determinar, com uma boa taxa de acerto, a ocorrência ou não de foco de calor para o período considerado. A Figura 5b ilustra a predição do risco espacial para todos os dados relativos à discretização do domínio considerado.

Para avaliar a estimativa do risco de incêndios, realizou-se a predição do risco para os 9327 focos de calor que ocorreram durante o ano de 2014 [8]. A Tabela 2 apresenta o percentual de focos de calor e do total de células do domínio que apresentam risco espacial zero e maior que 20%. Esses números indicam que, de fato, locais com maior risco espacial estão mais propensos à ocorrência de novos focos de calor.

Tabela 2: Percentuais de focos de calor ocorridos em 2014 e do total de células do domínio

Risco Espacial	Percentual	
	Focos de Calor	Total de Células
Zero	4,52%	48,12%
Maior que 20%	80,5%	29,24%

4 Conclusões

O trabalho desenvolvido modela o perigo e o risco de ocorrência de incêndios por técnicas diferentes. O primeiro está associado à variação dos fatores climáticos durante um ano completo para estimar o período no qual o risco temporal - aqui denominado perigo de incêndio - se torna positivo e começa a aumentar. O último - denominado aqui de risco espacial - é obtido a partir de um sistema de aprendizado computacional que identificou, baseado em um histórico de ocorrências de focos de calor, características físicas do meio estudado que estão associadas ao fenômeno dos incêndios. No risco temporal

as informações utilizadas no modelo foram obtidas de especialistas e de observação da ocorrência dos focos de calor associando-os aos fatores climáticos. Já no risco espacial, essas associações foram extraídas dos dados pelo classificador (aprendizado de máquina). A acurácia do classificador k-NN foi de 84,3% e apenas para um percentual de 4,52% de focos de calor ocorridos em 2014 foi atribuído risco zero. Em relação ao risco temporal, a estimativa para o ano de 2014 foi avaliada a partir de uma classificação binária, na qual obteve-se acurácia de 73,97% e o valor da área abaixo da curva ROC foi 0,8794.

Agradecimentos

O primeiro autor agradece a bolsa de doutorado da CAPES e o segundo agradece ao CNPQ (processo 305862/2013-8).

Referências

- [1] G. E. A. Acre. Zoneamento Ecológico-Econômico do Estado do Acre, Fase II (Escala 1:250.000): Documento Síntese. Ed. Rio Branco: SEMA, 2010.
- [2] A. A. Brown and K. P. Davis. *Forest Fire*. McGraw Hill, New York, 2 edition, 1973.
- [3] L. C. Barros and R. C. Bassanezi. *Tópicos de Lógica Fuzzy e Biomatemática*. Coleção IMECC. Textos Didáticos. IMECC - UNICAMP, Campinas, 2 edition, 2010.
- [4] M. S. Cecconello. *Modelagem alternativa para dinâmica populacional: Sistemas dinâmicos fuzzy*. Dissertação de Mestrado. IMECC- UNICAMP, Campinas, 2006.
- [5] W. Pedrycz and F. Gomide. *Fuzzy Systems Engineering*. John Wiley & Sons, Inc., New Jersey, 2007.
- [6] J. D. M. Silva. *Análise de Estabilidade de Sistemas Dinâmicos P-fuzzy com Aplicações em Biomatemática*. Tese de doutorado, Unicamp, 2005.
- [7] M. A. P. A. Brasil. Instituto Nacional de Meteorologia, 2011. Disponível em: <http://www.inmet.gov.br>. Acesso em: 20/06/2015.
- [8] D. P. I. Inpe. Divisão de processamento de imagens, 2011. Disponível em: <http://www.dpi.inpe.br/>. Acesso em: 20/03/2015.
- [9] J. Davis and M. Goadrich. The Relationship Between Precision-Recall and ROC Curves. In *Proceedings of the International Conference on Machine Learning*, pages 233–240, Pittsburgh PA, 2006.
- [10] A. R. Webb. *Statistical Pattern Recognition*. John Wiley & Sons, Ltd., Malvern, UK, 2 edition, 2002.
- [11] S. Haykin. *Redes Neurais: Princípios e Práticas*. Ed. Bookman, Porto Alegre-RS, 2001.