

Técnicas de Clusterização para agrupamento de eventos de incêndios

RESUMO

Camila da Silveira Machado

camimachado@gmail.com

Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, Rio Grande do Sul, Brasil

Incêndios em ambientes domésticos respondem por um quarto das ocorrências anuais totais de incidentes com fogo nos EUA, sendo responsáveis por 81% das mortes e 77% das lesões em civis ao ano. O estudo dessas estatísticas tem possibilitado o desenvolvimento de diretrizes para prevenção, proteção e combate ao fogo. Para tanto, percebe-se crescente emprego de ferramentas multivariadas para o reconhecimento de tendências das ocorrências de incêndios domiciliares, em particular a análise de cluster. Este artigo propõe o uso de técnicas multivariadas para a formação de grupos homogêneos de ocorrências de incêndios residenciais. A metodologia proposta combina procedimentos hierárquico e não hierárquico de clusterização, bem como a Análise de Componentes Principais (ACP) com vistas a aprimorar a qualidade da clusterização avaliada através do Silhouette Index (SI). A sistemática proposta foi aplicada ao conjunto de observações de incêndios residenciais reportados nos estados americanos da Flórida e do Texas, no período de 2010 a 2014. O SI médio obtido com as observações de incêndio na Flórida foi de 0,8481 e de 0,9141 no Texas. Os agrupamentos gerados pelos dados dos dois estados apresentaram semelhanças, sendo majoritariamente justificados pela presença e efetividade de operação de detectores automáticos de fogo.

PALAVRAS-CHAVE: Classificação. Análise de clusterização. Análise de Componentes Principais. K-Means. Índice Silhouette. Incêndio.

INTRODUÇÃO

Embora o número de incêndios tenha diminuído consideravelmente em anos recentes, tal evento ainda responde por muitas fatalidades e perdas econômicas, justificando o desenvolvimento de estratégias para proteger a vida e a propriedade, bem como iniciativas para extinguir ou retardar o fogo por meio de estratégias de combate, prevenção e proteção. Regulamentos de proteção e prevenção contra o fogo remontam à Roma antiga, dada a magnitude dos prejuízos e perdas decorrentes de grandes incêndios (TROITZSCH, 2016). A literatura traz vasto relato acerca de catástrofes causadas por incêndios que destruíram grandes centros urbanos nos Estados Unidos. Em Pittsburgh (Pensilvânia), em 1845, cerca de 1.200 edifícios foram destruídos, 12.000 pessoas ficaram desabrigadas, e as perdas foram estimadas em 12 milhões de dólares na época (COOK JR., 1968). Em Chicago, em 1871, um incêndio destruiu cerca de 200 acres, desabrigando 100 mil habitantes, ocasionando uma perda de aproximadamente 200 milhões de dólares em propriedades e 300 mortos (MILLER, 1996). Em 1872, em Boston, um incêndio extinguiu cerca de 26 hectares do centro da cidade, destruindo grande parte do distrito financeiro, causando danos de US\$ 75 milhões (HORNBECK; KENISTON, 2017). Apesar de terem sido causados por um grande terremoto, em 1906, em São Francisco (Califórnia), incêndios devastadores irromperam na cidade e duraram vários dias, resultando na morte de mais de 3.000 pessoas e na destruição de mais de 80% da cidade; perdas econômicas foram estimadas em 500 milhões de dólares (SCAWTHORN et al., 2005).

Com vistas a evitar ocorrências de tamanho impacto, tecnologias avançadas em proteção contra incêndio e novas técnicas e materiais de construção têm auxiliado na prevenção contra o fogo. Além disso, países desenvolvidos e emergentes têm progressivamente adotado legislações de segurança contra incêndios. Embora as estatísticas mostrem que a maioria das perdas de incêndio ocorre em edifícios residenciais - correspondendo a 26% das ocorrências anuais totais de incidentes com fogo, incêndios domésticos contabilizam, em média, 81% das mortes e 77% das lesões em civis no ano (UFSA, 2017) - as regulamentações, na prática, são mais severas com sistemas de prevenção e combate ao fogo em edifícios altos, construções multi-familiares e comerciais do que com edificações residenciais (JENNINGS, 2013; TROITZSCH, 2016).

Amplamente utilizadas, as análises estatísticas de dados de incêndios se tornaram um importante meio para compreensão de tais eventos (LIZHI; AIZHU, 2008). A gestão de risco de incêndio objetiva compreender as tendências das ocorrências e fundamentar a seleção de medidas de prevenção e combate ao fogo, bem como visam ao desenvolvimento de estratégias para alocação eficiente dos recursos (CEYHAN et al., 2013). Essa gestão, ao envolver o estudo das estatísticas de incêndio à adaptação de regulamentos de segurança contra incêndios, tem como propósito a redução das perdas por incêndio (TROITZSCH, 2016).

No que diz respeito a abordagens quantitativas voltadas à análise de dados de incêndios, percebe-se crescente utilização de ferramentas multivariadas

objetivando diminuir a complexidade do problema, como Lizhi e Aizhu (2008), Palamara et al. (2011), Higgins et al. (2013; 2014) e Chen et al. (2017) aplicaram *K-means* para identificar a similaridade entre as ocorrências de incêndio coletadas e Orozco et al. (2012), Wuschke et al (2013) e Xin e Huang (2013) aplicaram ferramentas de mapeamento de *clusters* para identificar a significância estatística dos agrupamentos gerados. Dentre as principais abordagens multivariadas está a análise de agrupamentos (clusterização), a qual busca padrões similares em um conjunto de dados, agrupando observações (eventos de incêndios, no caso deste estudo) de forma que as mesmas sejam semelhantes dentro do mesmo grupo, mas mutuamente excludentes entre os grupos (HAIR JR. et al., 2009). Dada a simplicidade dos seus preceitos e flexibilidade de aplicação, a análise de clusterização tem sido vastamente aplicada na análise de dados descritivos de eventos de incêndios. Lizhi e Aizhu (2008) desenvolveram um método de classificação automática em relação ao risco de incêndio por meio de um modelo de agrupamento de *K-Means*, enquanto que Clare et al. (2012) detalharam a clusterização para avaliar o impacto do serviço de bombeiros e da campanha de educação pública de prevenção de incêndio na redução da frequência e gravidade de incêndios domésticos. Orozco et al. (2012), através do georreferenciamento, aplicaram a clusterização para detectar a distribuição espacial e temporal dos incêndios florestais e analisar suas causas de ignição. Wuschke et al. (2013) compararam as distribuições espaço-temporais de incêndios residenciais e assaltos a residências por agrupamentos, ao passo que Xin e Huang (2013) apresentaram um modelo de análise de risco baseado em *clusters* de cenários de incêndio, de supressão automática do fogo e de aspectos comportamentais.

Embora diversos estudos tenham se apoiado em técnicas de clusterização para encontrar padrões similares em dados descritivos de eventos de incêndios, percebe-se espaço para evoluir na adaptação de ferramentas com vistas a agrupar os incidentes de incêndio e permitir a interpretação de grandes quantidades de observações que podem não apresentar significado claro quando dispersas. Este artigo propõe uma sistemática apoiada em técnicas multivariadas para a formação de grupos homogêneos de ocorrências de incêndios residenciais. O método proposto é operacionalizado em cinco etapas: (i) geração do dendrograma para definição do número de *clusters* a serem gerados, (ii) clusterização via *K-Means* utilizando as variáveis originais que descrevem os eventos de incêndio, (iii) avaliação da qualidade dos agrupamentos gerados através do *Silhouette Index*, (iv) interpretação dos agrupamentos formados, e (v) repetição das etapas (i) a (iv) para dados transformados pela Análise de Componentes Principais (ACP). A transformação das variáveis originais via ACP visa a aprimorar a qualidade dos agrupamentos gerados através da utilização de novas variáveis não correlacionadas (scores da ACP) no procedimento de clusterização.

O artigo se divide em cinco seções. A presente introdução aponta o contexto que se insere o estudo e o objetivo proposto. A Seção 2 apresenta a fundamentação teórica das técnicas multivariadas utilizadas no estudo. A Seção 3 apresenta a metodologia empregada, enquanto a Seção 4 descreve o conjunto de observações utilizadas na pesquisa e discorre sobre os resultados obtidos. A Seção 5 conclui o estudo e sugere possíveis desdobramentos desse trabalho.

REFERENCIAL TEÓRICO

ANÁLISE DE AGRUPAMENTOS

A análise de clusterização é método estatístico para alocar observações em grupos (*clusters*), de forma que as observações inseridas em um mesmo grupo sejam similares entre si e diferentes das inseridas em outro. Os agrupamentos gerados devem exibir elevada homogeneidade interna e externa; se representados graficamente, os objetos dentro dos *clusters* deverão estar próximos e agrupamentos distintos estarão distantes. Em termos práticos, a clusterização subsidia a interpretação de grandes quantidades de observações, que podem não apresentar significado evidente quando dispersas, com a descrição concisa das observações através da construção de agrupamentos (HAIR et al., 2009).

Existem dois procedimentos para clusterização de observações: hierárquicos e não hierárquicos. Procedimentos hierárquicos geram agrupamentos através do dendrograma, não se fazendo necessário pré-estabelecer o número de *clusters*, e avaliam progressivamente a similaridade entre os grupos e observações. Por sua vez, procedimentos não hierárquicos inserem objetos em um único movimento com base nas distâncias entre as observações, pressupondo o conhecimento prévio do número de *clusters* para alocação das observações em grupos (HAIR et al., 2009). Jain e Dubes (1988) realçam que o impacto visual que o dendrograma gera é uma característica importante dos métodos de agrupamento hierárquico, possibilitando ao pesquisador ver como os objetos estão sendo mesclados em *clusters* ou divididos em níveis sucessivos de proximidade. O dendrograma é um tipo específico de diagrama em estrutura de árvore que consiste em camadas de nós, cada um representando um cluster, onde as linhas conectam os nós que representam os *clusters* que estão aninhados uns aos outros e o corte horizontal do dendrograma define o agrupamento.

Embora os resultados estejam diretamente relacionados à qualidade dos dados e ao número de agrupamentos gerados, percebe-se que, no geral, métodos não hierárquicos propõem melhor desempenho para grandes conjuntos de dados, visto que dispensam o cálculo de matrizes de semelhança entre todas as observações, requisitando apenas o cálculo da semelhança entre cada observação e o centroide do cluster. Além disso, são considerados menos suscetíveis a dados atípicos, à medida de distância e à inclusão de variáveis irrelevantes (HAIR et al., 2009). No entanto, conforme Hair Jr. et al (2009), a combinação de ambos procedimentos se faz conveniente para aprimorar o desempenho da clusterização: usa-se o método hierárquico para definição do número adequado de *clusters* a serem gerados, e o método não hierárquico para efetiva realização dos agrupamentos. Tal sistemática foi adotada nessa pesquisa, utilizando o método *K-Means* para clusterização das observações (RENCHE, 2002).

Operacionalmente, o método *K-Means* agrupa as observações em *K clusters*, sendo esse *K* um valor previamente conhecido (NALDI et al., 2011; KAUFMAN; ROUSSEEUW, 2005). O agrupamento é feito calculando o centroide de cada

grupo e atribuindo cada observação ao grupo com o centroide mais próximo. O método fornece um algoritmo que minimiza a soma das distâncias euclidianas entre as observações e o centroide mais próximo (TABOADA; COIT, 2007). Os *K-Means* apoia-se em: (i) selecionar uma partição inicial com *clusters* K, (ii) calcular os centroides para cada cluster e a distância euclidiana dos centroides para cada observação do conjunto de dados, (iii) agrupar as observações aos *clusters* cujos centroides estejam mais próximos e (iv) retornar ao passo (ii) até que a associação ao cluster se estabilize, ou seja, até quando não ocorrer variação significativa na distância mínima entre cada observação a cada um dos centroides dos K *clusters* (JAIN; DUBES, 1988; JAIN, 2010; NALDI et al., 2011).

A qualidade dos agrupamentos gerados através de ferramentas de clusterização pode ser avaliada através do *Silhouette Index* (SI), que afere a qualidade dos agrupamentos formados por qualquer técnica de clusterização. O SI mede o grau de similaridade de uma observação em relação às demais inseridas em seu próprio cluster, comparando-a com as observações alocadas em outros *clusters* (KAUFMAN; ROUSSEEUW, 2005). O SI é calculado através da equação (1), onde $a(n)$ é a distância média da n -ésima observação em relação às demais observações do grupo em que foi alocada e $b(n)$ é a distância média entre a n -ésima observação em relação às observações do grupo vizinho mais próximo.

$$SI_n = \frac{b(n) - a(n)}{\max\{b(n), a(n)\}} \quad (1)$$

Cada observação apresenta um SI_n no intervalo $[-1; 1]$, onde n representa a observação avaliada, sendo $n = 1, \dots, N$. Quanto mais elevado o valor do índice SI, melhor a alocação do objeto em seu *cluster* de destino. Dessa forma, valores próximos a -1 apontam que a observação foi equivocadamente inserida no grupo; valores intermediários, próximos a zero, denotam observações que a observação poderia estar tanto em seu grupo atual quanto em algum outro grupo; e valores de SI próximos a 1 revelam que a distância, ou dissimilaridade, entre a observação e demais observações inseridas em outros grupos é pequena, considerando-se, portanto, que a observação foi adequadamente alocada em seu *cluster* de destino (ROUSSEEUW, 1987).

ANÁLISE DE COMPONENTES PRINCIPAIS

O objetivo central ACP é reduzir a dimensionalidade de um conjunto de observações que consiste em um grande número de variáveis inter-relacionadas, mantendo o máximo possível a variabilidade do sistema. Isso é alcançado por meio da conversão em um novo conjunto de variáveis, os componentes principais (CPs), que não são correlacionados e que são ordenados de forma que os primeiros retenham a maior parte da variação presente em todas as variáveis originais (JOLLIFFE, 2002).

Conforme Rencher (2002), o primeiro componente principal é a combinação linear com variância máxima – onde se busca uma dimensão ao longo da qual as observações sejam separadas ou dispersas ao máximo; já o segundo componente

principal é a combinação linear com variância máxima em uma direção ortogonal ao primeiro componente principal e assim por diante.

A operacionalização da ACP ocorre por meio do cálculo dos autovalores e dos autovetores da matriz de covariância dos dados originais (JOLLIFFE, 2002). Considere x um vetor de P variáveis. O primeiro componente principal é definido como $\alpha_1^T x$ tal que os elementos em x apresentem máxima variância, onde $\alpha_1^T = [\alpha_{11} \alpha_{12} \dots \alpha_{1p}]$ são atribuídos como pesos. O segundo componente é estabelecido como $\alpha_2^T x$, não correlacionado com $\alpha_1^T x$, e com os elementos de x tendo a máxima variância possível. Os vetores α_j são autovetores da matriz Σ , considerada como a matriz de variâncias e covariâncias de x . Finalmente, impõe-se à formulação de maximização de variância entre os componentes a restrição $\alpha_j^T \alpha_j = 1$, impondo o comprimento unitário nos autovetores. Nessa notação, cada autovetor α_j está relacionado com λ_j , o j -ésimo maior autovalor da matriz Σ . O problema se resume a maximizar a variância de $\alpha_1^T x = \alpha_1^T \Sigma \alpha_1$, sujeito à restrição $\alpha_1^T \alpha_1 = 1$ (CERVO; ANZANELLO, 2015).

Por conta de sua reconhecida habilidade na análise de dados, a ACP encontra aplicabilidade em dados de incêndios e outras naturezas de acidentes. Miguel et al. (2010) utilizaram a ACP para avaliar a influência de um período de treinamento no comportamento do trabalhador e, conseqüentemente, no tempo de evacuação, em caso de emergência. Hastie e Searle (2016), por ACP e regressão de mínimos quadrados ordinários, desenvolveram um modelo para explicar a variação nas taxas de incêndio acidental usando variáveis socioeconômicas. Conedera et al. (2018) utilizaram a análise de redundância (RDA), descrita como uma ACP em conjunto da análise de agrupamento hierárquico baseado no índice de dissimilaridade de *Bray-Curtis*, visando a identificar as características climáticas, ambientais e socioeconômicas dos *clusters* gerados para os regimes de incêndio nas Região dos Alpes. Caleffi et al. (2017) apresentaram uma abordagem multivariada para seleção de variáveis aplicando em conjunto a Distância de Bhattacharyya (BD) e a ACP a um conjunto de observações composto de variáveis que potencialmente ajudam a explicar a ocorrência de conflitos de tráfego, usando também um modelo de Análise Linear Discriminante (LDA) com as variáveis selecionadas para estimar a ocorrência de conflitos. Rundmo e Hale (2003) valeram-se da ACP para identificar as dimensões das atitudes de segurança e de prevenção de acidentes entre gestores de uma empresa. Dimitriou et al. (2018), para demonstrar uma abordagem de investigação da magnitude das inconsistências de dados estatísticos de segurança rodoviária, utilizaram em conjunto a ACP, a análise de clusterização hierárquico e a Modelagem de Equações Estruturais (SEM).

TÉCNICAS MULTIVARIADAS APLICADAS A DADOS DESCRITIVOS DE INCÊNDIO

A literatura traz vasto número de aplicações de ferramentas de clusterização em dados associados a eventos de incêndios. Higgins et al. (2013) utilizaram *K-Means* para identificar os perfis de cidades inglesas por meio da coleta de informações demográficas e de estilo de vida dos indivíduos; a pesquisa integrou técnicas estatísticas e um modelo espacial para compreender o risco de incêndio e identificar o índice de vulnerabilidade dos indivíduos a incêndios. Higgins et al. (2014) também se valeram de *K-Means* para identificar as características

socioeconômicas e níveis de risco de incêndio dos segmentos populacionais dentro da região estudada visando a examinar a evolução de um sistema de informação geográfica para apoio à prevenção de incêndios.

Guldaker e Hallin (2014), para identificar as causas da distribuição espacial e espaço-temporal de incêndios intencionais e analisar como as diferentes condições de vida podem determinar a frequência dessas ocorrências, aplicaram o método de densidade de *Kernel (Kernel Density Estimation)* agregado à análise de clusterização para revelar *clusters* entre as subáreas geradas. Tung e Kim (2011) propuseram um algoritmo para detecção de fumaça usando imagens de vídeo, aplicando a técnica de clusterização *Fuzzy C-Means* para extrair parâmetros de características da fumaça que foram utilizados como vetores para máquina de suporte vetorial (*Support Vector Machine - SVM*). Na mesma linha, Alamgir et al. (2018) desenvolvem um algoritmo de detecção de fumaça baseado em vídeo combinando Padrões de Co-ocorrência Binária Local com a análise de *clusters Fuzzy C-Means* e SVM para explorar as características da textura da fumaça. Palamara et al. (2011) desenvolveram uma abordagem utilizando em conjunto os algoritmos Mapa Auto-Organizacional de *Kohonen (Kohonen's Self-Organizing Map - SOM)* e clusterização *K-Means* a fim de agrupar os acidentes ocupacionais em diferentes classes, comparando os resultados da sistemática proposta com a clusterização hierárquica. Por fim, Chen et al. (2017) aplicaram *K-Means* para investigar as características do fogo florestal e seus potenciais impulsionadores.

Lizhi e Aizhu (2008) desenvolveram um método para classificação automática do risco de incêndio urbano o qual apoia-se em número de ocorrências de incêndio, perdas econômicas diretas e baixas de incêndio - ponderados conforme sua importância para avaliação do risco. O estudo baseou-se nas estatísticas de incêndio em Shenyang, na China, no período de 2000 a 2004, realizando a análise de agrupamento em relação à estatística de incêndio com o método de agrupamento *K-Means*.

Orozco et al. (2012) agregaram os incêndios florestais registrados em Cantão Ticino (Suíça) de 1969 a 2008, em três subperíodos (*clusters*), conforme as disposições legais preventivas à época, e em duas causas de ignição. A metodologia usou permutação estatística de varredura, no espaço e do tempo, e sistema georreferenciado para visualização dos dados e resultados; testes de hipóteses de Monte Carlo foram aplicados para avaliar a significância estatística de cada cluster.

A relação de incêndios e de roubos em residência, de 2004 a 2006, em Surrey (Canadá), serviu para a aplicação de ferramentas de mapeamento de *clusters*, identificando as localidades estatisticamente significativas. Foram analisados os padrões espaciais e temporais de incêndios e de eventos de roubo a residências. Percebeu-se que incêndios, como crimes, não são uniformemente distribuídos no tempo e no espaço, ocorrendo de forma previsível e não aleatória (WUSCHKE, K. et al., 2013).

Análises de risco foram realizadas nos incêndios ocorridos em edifícios residenciais na China de 2007 a 2010. As características estruturais e de ocupação, os sistemas de prevenção e proteção contra o fogo e os perigos de incêndio presentes nas edificações constituíram os cenários; o número de mortes e as perdas financeiras compuseram os índices de risco. Os agrupamentos de

cenários de incêndio especificaram elementos relativos ao fogo, como desenvolvimento e propagação. O *cluster* de cenário de supressão automática de fogo descreveu o processo de extinção ou de controle desse. O *cluster* de cenários comportamentais descreveu a reação dos ocupantes em resposta ao início do fogo e a intervenção dos serviços de combate e resgate (XIN; HUANG, 2013).

MÉTODO

O método proposto é dividido em cinco etapas: (i) geração do dendrograma para definição do número de *clusters* a serem formados, (ii) clusterização via *K-Means* utilizando as variáveis originais que descrevem os eventos de incêndio, (iii) avaliação da qualidade dos agrupamentos gerados através do *Silhouette Index*, (iv) interpretação dos agrupamentos formados, e (v) repetição das etapas (i) a (iv) para dados transformados pela ACP.

Etapla 1: Geração do dendrograma para definição do número de *clusters* a serem formados

Consideram-se as x variáveis originais do conjunto de observações. Aplica-se a elas o método hierárquico de dendrograma para definir o número K de *clusters* recomendados. O dendrograma apresenta visualmente as características das observações avaliadas, representando, por meio do gráfico tipo árvore, a similaridade entre os grupos e observações.

Etapla 2: Clusterização via *K-Means* utilizando as variáveis originais que descrevem os eventos de incêndio

Conhecendo o número adequado de *clusters* de acordo com o dendrograma resultante na etapa anterior, aplica-se a ferramenta não hierárquica *K-Means* às variáveis originais para agrupar as observações em K *clusters*.

Etapla 3: Avaliação da qualidade dos agrupamentos gerados através do *Silhouette Index*

A qualidade dos grupos gerados é avaliada através da média do SI, fazendo uso da distância Euclidiana para medir o grau de similaridade de uma observação em relação às demais alocadas no mesmo cluster e comparando-a com as observações inseridas em outros *clusters*.

Etapla 4: Interpretação dos agrupamentos formados

A quarta etapa consiste em interpretar qualitativamente os resultados da clusterização, identificando particularidades que definem cada grupo, apontando os fatores mais relevantes, similaridades e dissimilaridades entre as observações de cada grupo. Para tanto, essa etapa apoia-se na literatura específica da área, buscando compreender os padrões de comportamento das ocorrências de incêndio residencial dos bancos de dados estudados.

Etapla 5: Repetição das etapas 1 a 4 para dados transformados pela ACP

A fim de avaliar a influência que os níveis de correlação das variáveis exercem sobre a formação dos agrupamentos, converte-se o conjunto original de variáveis em um novo conjunto de variáveis latentes t (PCA scores). Passa-se então a utilizar tais variáveis nas 4 etapas descritas acima.

ESTUDO DE CASO

ÁREA DE ESTUDO

O *National Fire Incident Reporting System* (NFIRS) é um relatório utilizado por bombeiros americanos para reportar atividades e eventos de incêndios. Anualmente, mais de 30.000 departamentos de bombeiros reportam-se ao NFIRS, relatando quase 1 milhão de incidentes de incêndio. Ressalta-se, no entanto, que por se tratar de um sistema de preenchimento voluntário, inclui apenas os incidentes de incêndio notificados pelos departamentos de bombeiros que se reportam ao NFIRS, não se constituindo em um relato de todas as ocorrências de incêndio. Assim, a NFIRS não é representativa da totalidade de incidentes de incêndio nos Estados Unidos, não servindo como um recenseamento de incidentes de incêndio ou acidentes (USFA, 2015; 2017).

Os dados utilizados nesse estudo são provenientes de dois dos onze relatórios disponibilizados pela NFIRS: o Módulo de Incêndio (NFIRS-2), que descreve cada evento de incêndio, e o Módulo de Incêndio de Estrutura (NFIRS-3), que especifica a estrutura que sofreu essa ocorrência. Os dados coletados referem-se aos incêndios residenciais ocorridos nos estados da Flórida e do Texas no período de 2010 a 2014 associados a cinco causas: incêndio intencional, descuido ao cozinhar, mau funcionamento na rede elétrica, falhas em eletrodomésticos e incêndio a partir de chamas abertas. Optou-se por realizar o estudo sobre os dados desses estados devido à representatividade de seus registros. As cinco causas de incêndio foram selecionadas para a investigação por representarem grande parte da ocorrência de incêndio doméstico no contexto analisado.

As 14 variáveis que compõem os bancos de dados e explicam os eventos são: (i) área de origem do fogo, (ii) fonte de calor, (iii) primeiro item a sofrer ignição, (iv) causa da ignição, (v) fator humano, (vi) propagação do fogo, (vii) condições de uso e ocupação da edificação, (viii) grau do dano causado pelo fogo ao pavimento, (ix) presença de sistema de detecção automática do fogo, (x) tipo de detector automático do fogo, (xi) sistema de alimentação do detector automático do fogo, (xii) operação do detector automático do fogo, (xiii) efetividade do detector automático do fogo e (xiv) presença de sistemas automáticos de extinção do fogo.

Em termos da composição e pré-processamento dos bancos de dados analisados, foram selecionados exclusivamente os eventos de incêndio em edificações residenciais de um pavimento que ocorreram de forma isolada, envolvendo apenas essa edificação no incidente. As observações contam somente com as edificações identificadas como construção finalizada e fechada. Considera-se edificação fechada aquela que não cumpre os requisitos para ser tratada como aberta ou parcialmente aberta; edificações abertas são aquelas que cada parede contém, pelo menos, 80% da área aberta. Classificam-se parcialmente abertas as edificações em que a área total de aberturas em uma parede, que recebe pressão externa positiva, excede em mais de 10% a soma das áreas de aberturas no balanço do envoltório do edifício (paredes e telhado) e que a área total de aberturas em uma parede, que recebe uma pressão externa positiva, excede 0,37 m² ou 1% da área dessa parede, o que for menor, e a

porcentagem de aberturas no balanço do envoltório do edifício não excede 20% (ASCE, 2010).

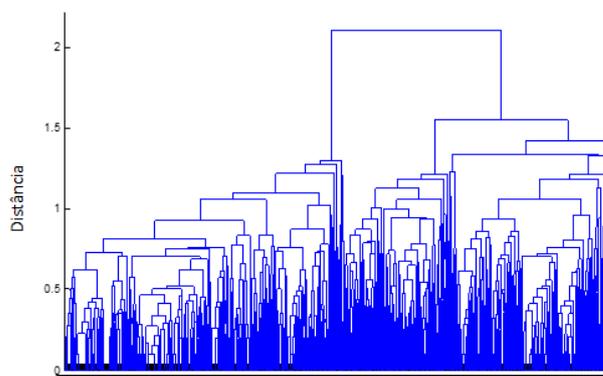
No tratamento das observações, também foram eliminadas aquelas identificadas como espúrias, ora porque continham registros indeterminados nas variáveis, ora porque os registros eram conflitantes entre as variáveis, restando então 782 observações para compor o banco de dados da Flórida e 1596 para o banco do Texas. Além disso, as ocorrências registradas de forma incompleta ou inconsistente foram excluídas do conjunto de dados. Tashakkori e Teddlie (2002) e Sandelowski (2000; 2001) sugerem que temas qualitativos sejam representados numericamente em escores ou escalas para facilitar a interpretação do fenômeno. Dessa forma, as observações qualitativas da amostra foram quantificadas através no intervalo 0 a 1, pontuados através da lógica “maior-é-melhor”, conforme Anexo.

RESULTADOS

BANCO DE DADOS DA FLÓRIDA

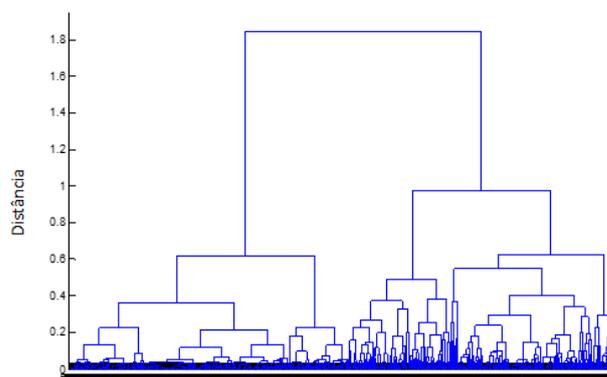
As Figuras 1, 2 e 3 apresentam, respectivamente, os dendrogramas gerados pelo método hierárquico nas 14 variáveis originais e com 2 e 3 CPs obtidos pela ACP, respectivamente. Percebe-se, nos 3 casos, consistente sugestão para formação de 2 ou 3 *clusters*. Por se tratar de uma análise exploratória, optou-se pela formação de 3 agrupamentos (K=3).

Figura 1 - Dendrograma gerado a partir das 14 variáveis originais do banco de dados da Flórida



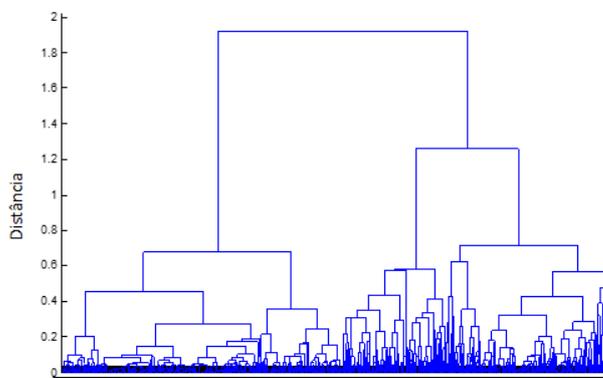
Fonte: Autoria própria (2019)

Figura 2 - Dendrograma gerado a partir dos 2 CPs obtidos pela ACP do banco de dados da Flórida



Fonte: Autoria própria (2019)

Figura 3 - Dendrograma gerado a partir dos 3 CPs obtidos pela ACP do banco de dados da Flórida



Fonte: Autoria própria (2019)

Na sequência, os eventos descritos pelas 14 variáveis originais foram clusterizadas através do método *K-Means*, assim como os conjuntos de 2 e 3 CPs gerados pela ACP. A qualidade da clusterização foi avaliada através do SI médio, conforme apresentado na Quadro 1; o melhor valor foi 0,8481, obtido tanto para o agrupamento com as 14 variáveis originais como pela clusterização realizada com os 3 CPs. De tal forma, percebe-se que as correlações tipicamente verificadas nas variáveis originais não comprometem a qualidade da clusterização, tendo em vista que a qualidade do procedimento é equivalente ao agrupamento gerado com base em 3 CPs, o que é amparado pelo fato da análise de *clusters* estar baseada em distâncias e não em correlações.

Quadro 1 – Índices *Silhouette* médios obtidos para banco de dados da Flórida com a aplicação do método *k-means* com 14 variáveis originais e para 2 e 3 CPs obtidos pela ACP

	Clusterização utilizando método <i>K-means</i> – banco de dados Flórida		
	Com o uso das 14 variáveis originais	Com o uso da ACP	
		2 CPs	3 CPs
k	3	3	3
SI médio	0,8481	0,7660	0,8481

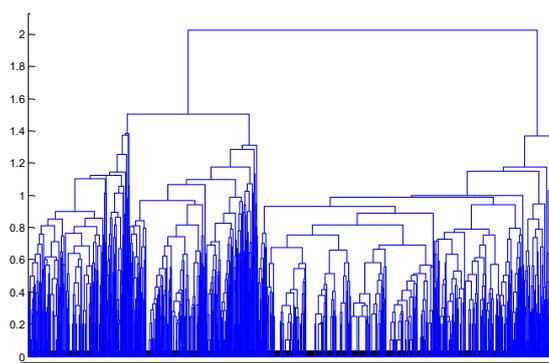
Fonte: Autoria própria (2019)

A análise dos agrupamentos formados pelas 14 variáveis e pelas 3 CPs mostrou formação de agrupamentos idênticos, sendo a proporção de eventos em cada cluster igual a 20%, 29% e 51%. O maior dos *clusters* gerados, com 51% das observações agrupadas, contém todos os incidentes de incêndio que não contava com sistema de detecção automática do fogo, ou seja, a variável **presença de sistema de detecção automática do fogo** é nula e, por conseguinte, as variáveis **tipo de detector automático do fogo, sistema de alimentação do detector automático do fogo, operação do detector automático do fogo e efetividade do detector automático do fogo** também não possuem valor nesse grupo. O *cluster* que abrange 29% dos eventos concentra eventos nos quais o sistema de detecção automática do fogo operou plenamente. O *cluster* com 20% das observações é composto pelas demais observações que possuem sistema de detecção automática do fogo. No entanto, percebe-se que, na maioria dos eventos, esse sistema não operou plenamente (97%) e, portanto, a variável **efetividade do detector automático do fogo** tenha sido registrada apenas para 3% dos incidentes.

BANCO DE DADOS DO TEXAS

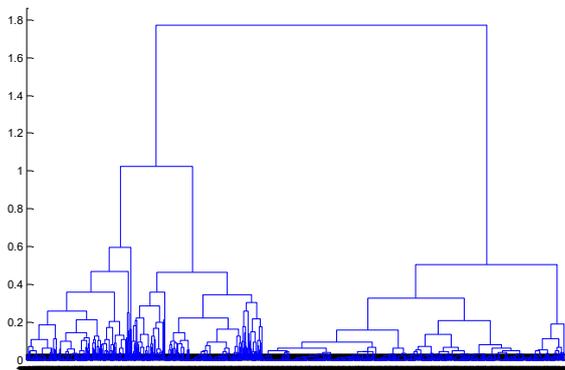
A Figura 4 apresenta o dendrograma gerado com aplicação do método sobre as 14 variáveis originais, enquanto que as Figuras 5 e 6 mostram os dendrogramas aplicados sobre sobre 2 CPs e 3 CPs. Novamente, entende-se que a formação de 3 *clusters* é satisfatória para evidenciar os diferentes padrões dos eventos de incêndio.

Figura 4 - Dendrograma gerado a partir das 14 variáveis originais do banco de dados da Texas



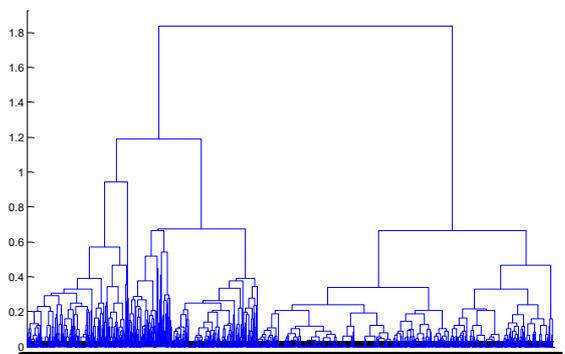
Fonte: Autoria própria (2019)

Figura 5 - Dendrograma gerado a partir dos 2 CPs obtidos pela ACP do banco de dados da Texas



Fonte: Autoria própria (2019)

Figura 6 - Dendrograma gerado a partir dos 3 CPs obtidos pela ACP do banco de dados da Texas



Fonte: Autoria própria (2019)

A Quadro 2 mostra os valores médios de SI obtido pela clusterização via *K-Means* aplicada ao conjunto de variáveis originais do banco de dados do Texas e aos conjuntos de 2 e 3 CPs gerados pela ACP nesse mesmo banco. O melhor valor de SI médio encontrado foi de 0,9141 para a clusterização aplicada aos conjuntos de 2 e 3 CPs obtidos pela ACP, sugerindo satisfatória estratificação das observações.

Quadro 2 – Índices *Silhouette* médios obtidos para banco de dados do Texas com a aplicação do método *k-means* com 14 variáveis originais e para 2 e 3 CPs obtidos pela ACP

	Clusterização utilizando método <i>K-means</i> – banco de dados Texas		
	Com o uso das 14 variáveis originais	Com o uso da ACP	
		2 CPs	3 CPs
k	3	3	3
SI médio	0,7875	0,9141	0,9141

Fonte: Autoria própria (2019)

Os *clusters* gerados a partir de 2 e 3 CPs conduziram a alocações idênticas dos eventos nos *clusters* de destino. A proporção do agrupamento das observações para o banco de dados do Texas foi de 19%, 24% e 56%. O *cluster* com 19% dos incidentes é composto por observações que apresentavam sistema

de detecção automática do fogo embora, em 98% dos casos, esse sistema não tenha operado plenamente e, por isso, a variável efetividade do detector automático do fogo foi registrada em apenas 2% das observações (7 eventos). O *cluster* constituído por 24% dos eventos detém quase todas observações onde o detector automático do fogo operou plenamente; a exceção é encontrada em apenas uma única observação em que o detector não operou porque o fogo foi considerado pequeno para ativação. O maior dos *clusters*, com 56% das observações, traz as observações que não apresentam sistema de detecção automática do fogo, ou seja, a variável **presença de sistema de detecção automática do fogo** é nula e, em consequência, as variáveis **tipo de detector automático do fogo**, **sistema de alimentação do detector automático do fogo**, **operação do detector automático do fogo** e **efetividade do detector automático do fogo** também são nulas nesses registros.

DISCUSSÃO DOS RESULTADOS

A aplicação da abordagem proposta conduziu a resultados alinhados com as pesquisas e estudos relacionados ao gerenciamento de risco de incêndio trazidos pela literatura. Amplamente tratados em relatórios oficiais de incêndios, os sistemas de detecção do fogo fazem parte das exigências entre os sistemas de segurança contra o fogo nos EUA, sendo foco de diversas campanhas de prevenção contra incêndio e testados exaustivamente em estudos, tanto científicos (acerca de sua eficiência), como também comportamentais (acerca da conduta dos ocupantes da edificação em caso de incêndio). Os agrupamentos das observações foram gerados com relação às informações acerca do sistema de detecção automática do fogo: sua presença e sua efetividade de operação. Os agrupamentos resultantes em ambos bancos de dados encontraram padrões em torno das variáveis **presença de sistema de detecção automática do fogo** e **efetividade do detector automático do fogo** - se a detecção automática operou plenamente ou não.

Nos EUA, cada Estado tem seu próprio código de edificações, que na sua maioria adota os códigos da NFPA como base para a legislação local de segurança contra incêndio. A *National Fire Protection Association* (NFPA), uma organização global sem fins lucrativos, dedica-se à redução de mortes, ferimentos, propriedades e perdas econômicas decorrentes de eventos de incêndio. Entre suas diversas publicações, o NFPA 72, *National Fire Alarm and Signaling Code*, e o NFPA 101, *Life Safety Code*, exigem que residências novas e existentes de uma e duas famílias tenham detectores de fumaça dentro de todos os dormitórios e fora deles, estando presentes em todos os pavimentos. Além disso, exige-se que os detectores de incêndio estejam ligados diretamente à rede elétrica, bem como estarem interconectados e possuírem bateria auxiliar, de forma que todos soem simultaneamente alertando todos os ocupantes da edificação (AHRENS, 2015).

Os agrupamentos gerados em ambos conjuntos de dados apresentaram semelhança entre si, tanto pelas características das observações em relação às variáveis como pela proporção de distribuição das observações. Pode-se intitular os *clusters* gerados como: *cluster* de ocorrências de incêndio que não possuíam sistema de detecção automática do fogo, *cluster* de ocorrências de incêndio em que o sistema de detecção automática do fogo operou plenamente e *cluster* de

ocorrências de incêndio em que a operação do sistema de detecção automática do fogo falhou.

Embora as observações coletadas para esse estudo sejam referentes aos anos de 2010 a 2014, a proporção dos *clusters* gerados diverge bastante das estatísticas registradas no período de 2009 a 2013, quando os detectores de fumaça - incluindo os eventos em que o fogo foi considerado muito pequeno para ativá-los, as situações em que operaram plenamente e os casos em que falharam-, estavam presentes em 73% dos incêndios domésticos registrados à NFIRS, tendo sido acionados em 53% dos incêndios domiciliares reportados. O relatório oficial da NFIRS também traz que, quando presentes em incêndios domésticos considerados grandes para ativá-los, os detectores de fumaça operaram em 87% das ocorrências (AHRENS, 2015).

Para a Flórida, as observações nos *clusters* gerados agruparam-se de forma que 51% das ocorrências de incêndio não possuíam sistema de detecção automática do fogo, 29% das ocorrências de incêndio contaram com o sistema de detecção automática do fogo operando plenamente e 20% das ocorrências de incêndio apresentaram falha na operação do sistema de detecção automática do fogo. Para o Texas, essa distribuição de agrupamento das observações foi, respectivamente, 56%, 24% e 19%. Uma das explicações para essa discrepância no tamanho dos agrupamentos com as estatísticas registradas é o fato que detectores de fumaça ligados diretamente à rede elétrica oferecem maior efetividade de operação quando comparados a alarmes alimentados por bateria. Isso se evidencia nas características do *cluster* de ocorrências de incêndio que apresentaram falha na operação do sistema de detecção automática do fogo: na Flórida, percebe-se que 27% das observações contavam com o sistema ligado à rede elétrica com ou sem bateria, enquanto 36% possuíam apenas alimentação à bateria; no Texas, 61% das ocorrências possuíam sistema de detecção alimentado somente por bateria e apenas 18% contavam com a interligação à rede elétrica, em conformidade às normas de segurança.

As pesquisas publicadas corroboram esse quadro. Em 2011, a *American Housing Survey* (AHS) realizou uma pesquisa onde três a cada cinco entrevistados relataram possuir alarmes de fumaça alimentados apenas por baterias, enquanto um terço disse que seus alarmes eram alimentados por eletricidade e baterias e 7% tinham alarmes alimentados apenas por eletricidade. A pesquisa revelou também que, apesar das recomendações dos códigos da NFPA, 30% das residências com menos de cinco anos de idade que possuíam alarmes de fumaça contavam apenas com o sistema de alimentação por bateria. Outra pesquisa, realizada em 2010 pela *Harris Interactive* e NFPA, revelou que aproximadamente dois em cada cinco domicílios possuíam alarmes de fumaça em todos os dormitórios (como recomendado pelo NFPA 101), mas apenas um quarto das residências contavam com o sistema de detecção interconectado à rede elétrica.

Com relação ao *cluster* de ocorrências de incêndio que não possuíam de sistema de detecção automática do fogo, esperava-se que a variável **presença de sistema de detecção automática do fogo** fosse nula e, por conseguinte, as variáveis **tipo de detector automático do fogo**, **sistema de alimentação do detector automático do fogo**, **operação do detector automático do fogo** e **efetividade do detector automático do fogo** também não possuíssem valor nesse grupo, o que ocorreu em ambos conjuntos de dados analisados.

Para os *clusters* de ocorrências de incêndio em que o sistema de detecção automática do fogo operou plenamente e de ocorrências de incêndio em que a operação do sistema de detecção automática do fogo falhou, percebe-se que a maioria das observações contavam com sistemas de alarmes de fumaça, em comparação a sistemas de detecção de calor ou sistemas combinados de detecção de fumaça e calor. Esse comportamento vai de encontro aos números consolidados por Ahrens (2015) de 2009 a 2013, quando 88% dos dispositivos de detecção automático do fogo nos incêndios domiciliares relatados foram projetados para detectar apenas a fumaça.

As demais variáveis que compõem os bancos analisados, como **área de origem do fogo, fonte de calor, primeiro item a sofrer ignição, causa da ignição** e outras, embora expressem características das observações, têm seus valores extremamente pulverizados nos *clusters*, não apresentando relevância que auxilie na caracterização dos agrupamentos gerados. Esse comportamento era esperado para algumas variáveis, a exemplo da variável **presença de sistemas automáticos de extinção do fogo**: no período de 2010 a 2014, apenas 8% dos incêndios residenciais relatados à NFIRS contavam com a presença de *sprinklers*. Além disso, os Estados da Flórida e do Texas não possuem regulamentação estadual que exija o uso de *sprinklers* em ambientes residências, embora a Flórida permita que jurisdições locais prescrevam *sprinklers*, enquanto o Texas não autoriza legislações locais a disporem sobre esse assunto (AHRENS, 2017; NFPA, 2017).

CONCLUSÃO

A análise de *clusters* tem sido reiteradamente aplicada em diversas pesquisas que buscam a melhor compreensão das tendências dos incidentes de incêndio, fundamentando a escolha desse artigo em agrupar as observações para identificar as particularidades predominantes que definem cada grupo.

A abordagem proposta associou técnicas hierárquica e não-hierárquica de clusterização, aplicando o método *K-Means* para aprimoramento das técnicas hierárquicas e a ACP para melhorar a qualidade da análise de clusterização, avaliada do através do SI. As etapas dessa pesquisa compreenderam em (i) geração do dendrograma para definição do número de *clusters* a serem formados, (ii) clusterização via *K-Means* utilizando as variáveis originais que descrevem os eventos de incêndio, (iii) avaliação da qualidade dos agrupamentos gerados através do *Silhouette Index*, (iv) interpretação dos agrupamentos formados; e (v) repetição das etapas (i) a (iv) para variáveis transformadas pela ACP. A utilização dos CPs como parâmetros de entrada da clusterização (substituindo as variáveis originais) objetiva avaliar um eventual aprimoramento da qualidade dos grupos gerados com base na utilização de variáveis não correlacionadas (caso dos CPs).

A abordagem foi aplicada a dois conjuntos de eventos de incêndios descritos pelas mesmas 14 variáveis. Para a Flórida, o melhor valor de SI médio foi de 0,8481, obtido tanto a partir do agrupamento gerado pelo *K-Means* com as 14 variáveis originais como pela clusterização realizada com os 3 CPs. Para o Texas, o SI médio encontrado foi de 0,914, resultante da clusterização aplicada aos conjuntos de 2 e 3 CPs obtidos pela ACP. Os dendrogramas de ambos os bancos

de dados recomendaram o uso de 3 *clusters* para a aplicação do método *K-Means*. A distribuição das observações nesses *clusters* foi similar para Flórida e Texas, sendo os 3 grupos rotulados como (i) ocorrências de incêndio que não possuíam de sistema de detecção automática do fogo, (ii) ocorrências de incêndio em que o sistema de detecção automática do fogo operou plenamente, e (iii) ocorrências de incêndio em que a operação do sistema de detecção automática do fogo falhou. Os agrupamentos gerados pela aplicação do método atendem a expectativa, haja vista a vasta recorrência do assunto na literatura especializada e nos relatórios oficiais acerca de incêndio.

Para trabalhos futuros, sugere-se o emprego de diferentes ferramentas multivariadas para a formação dos agrupamentos, comparando os resultados obtidos com a metodologia proposta por esse artigo. Recomenda-se ainda o uso de outras métricas para aferir a qualidade dos *clusters* gerados.

Cluster analysis to grouping fire events

ABSTRACT

Residential fires answer for one quarter of the total annual occurrences of fire incidents in the US, accounting in recent years for approximately eighty percent of deaths and injuries in civilians in the United States. The study of these statistics has enabled the development of directives for prevention, protection and firefighting. In this regard, it is possible to identify the successive use of multivariate analysis techniques for the recognition of residential fire patterns, in particular, the cluster analysis. This article proposes the use of multivariate techniques to establish of homogeneous groups of residential fires occurrences. The proposed method combines hierarchical and non-hierarchical procedures of clustering, Principal Component Analysis (PCA) and Silhouette Index (SI) to evaluate the quality of the generated clusters. The presented system was applied to the group of residential fires incidents reported in the states of Florida and Texas between 2010 and 2014, associating these occurrences to five main causes. The average SI obtained with the fire observations in Florida was 0.8481 and 0.9141 in Texas. The clusters generated by the data of the two states showed similarities, being mainly justified by the presence and effectiveness of the automatic fire detectors operation.

KEYWORDS: Classification. Cluster analysis. Principal component analysis. K-Means. Silhouette Index. Fire.

REFERÊNCIAS

AHRENS, M. Smoke Alarms in U.S. Home Fires. **NFPA Research, Data and Analytics Division**, set. 2015.

AHRENS, M. U.S. Experience with Sprinklers. **NFPA Research, Data and Analytics Division**, jul. 2017.

ALAMGIR, N. Combining multi-channel color space with local binary Co-occurrence feature descriptors for accurate smoke detection from surveillance videos. **Fire Safety Journal**, disponível on line em 24. set. 2018. **crossref**

AMERICAN SOCIETY OF CIVIL ENGINEERS. **Minimum Design Loads for Buildings and Other Structures**, ASCE/SEI Standard 7-10, 2010.

CALEFFI, F.; ANZANELLO, M. J.; CYBIS, H. B. B. A multivariate-based conflict prediction model for a Brazilian freeway. **Accident Analysis and Prevention**, v. 98, p. 295-302, jan. 2017. **crossref**

CERVO, V. L.; ANZANELLO, M. J. Seleção de variáveis para clusterização de bateladas produtivas através de ACP e remapeamento kernel. **Production**, v. 25, n. 4, p. 826-833, out./dez. 2015. **crossref**

CEYHAN, E.; ERTUGAY, K.; DÜZGÜN, S. Exploratory and inferential methods for spatio-temporal analysis of residential fire clustering in urban areas. **Fire Safety Journal**, v. 58, p. 226-239, maio 2013. **crossref**

CLARE, J.; GARIS, L.; PLECAS, D.; JENNINGS, C. Reduced frequency and severity of residential fires following delivery of fire prevention education by on-duty fire fighters: Cluster randomized controlled study. **Journal of Safety Research**, v. 43, p. 123-128, abr. 2012. **crossref**

CHEN, D.; PEREIRA, J. M. C.; MASIEIRO, A.; PIROTTI, F. et al. Mapping fire regimes in China using MODIS active fire and burned area data. **Applied Geography**, v. 85, p. 14-26, ago. 2017. **crossref**

CONEDERA, M.; KREBS, P.; VALESE, E.; COCCA, G.; SCHUNKD, C.; MENZEL, A.; VACIK, H.; CANEGH, D.; JAPELJ, A.; MURI, B.; RICOTTA, C.; OLIVERI, S.; PEZZATTI, G. B. Characterizing Alpine pyrogeography from fire statistics. **Applied Geography**, v. 98, p. 87-99, set. 2018. **crossref**

COOK JR, D. E. The Great Fire of Pittsburgh in 1845 or How a Great American City Turned Disaster into Victory, **The Western Pennsylvania Historical**, v. 51, n. 2, p. 127-153, abril 1968.

DIMITRIOU, L.; NIKOLAOU, P.; ANTONIOU, C. Exploring the temporal stability of global road safety statistics. **Accident Analysis and Prevention**, disponível on line em 21 fev. 2018. [crossref](#)

GULDAKER, N.; HALLIN, P. O. Spatio-temporal patterns of intentional fires, social stress and socio-economic determinants: A case study of Malmö, Sweden. **Fire Safety Journal**, v. 70, p. 71-80, nov. 2014. [crossref](#)

HAIR JR., J. F.; BLACK, W. C.; BABIN, B. J.; ANDERSON, R. E.; TATHAM, R. L. **Análise multivariada de dados**. Recurso eletrônico. Tradução Adonai Schlup Sant'Anna. 6. ed. Porto Alegre: Bookman, 2009.

HASTIE, C.; SEARLE, R. Socio-economic and demographic predictors of accidental dwelling fire rates. **Fire Safety Journal**, v. 84, p. 50-56, ago. 2016. [crossref](#)

HIGGINS, E.; TAYLOR, M.; FRANCIS, H.; JONES, M.; APPLETON, D. The evolution of geographical information systems for fire prevention support. **Fire Safety Journal**, v. 69, p. 117-125, out. 2014. [crossref](#)

HIGGINS, E.; TAYLOR, M.; JONES, M.; LISBOA, P. J. G. Understanding community fire risk - A spatial model for targeting fire prevention activities. **Fire Safety Journal**, v. 62, parte A, p. 20-29, nov. 2013. [crossref](#)

HORNBECK, R.; KENISTON, D. Creative Destruction: Barriers to Urban Growth and the Great Boston Fire of 1872, **American Economic Review**, v. 107, n. 6, p. 1365–1398, jun. 2017. [crossref](#)

JAIN, A. K. Data clustering: 50 years beyond K-means. **Pattern Recognition Letters**, v. 31, ed. 8, p. 651-666, jun 2010. [crossref](#)

JAIN, A. K; DUBES, R. C. **Algorithms for clustering data**. Englewood Cliffs: Prentice Hall; 1988.

JENNINGS, C. R. Social and economic characteristics as determinants of residential fire risk in urban neighborhoods: A review of the literature. **Fire Safety Journal**, v. 62, parte A, p.13–19, nov. 2013. [crossref](#)

JOLLIFFE, I. T. **Principal Component Analysis**. 2. ed. New York: Springer-Verlag, 2002.

KAUFMAN, L.; ROUSSEEUW, P. **Finding groups in data: an introduction to cluster analysis**. New Jersey: Wiley Interscience; 2005.

LIZHI, W.; AIZHU, R. Urban Fire Risk Clustering Method Based on Fire Statistics. **Tsinghua science and technology**, v. 13, suplemento 1, p. 418-422, out. 2008.

crossref

MIGUEL, A. S.; GÓIS, J.; SILVA, J. Study on workers' evacuation in an industrial company. **Safety Science**, v. 48, p. 1050-1053, out. 2010. **crossref**

MILLER, D. City of the Century: The Epic of Chicago and the Making of America. In: _____. **My Lost City, The Great Fire**. New York: Simon & Schuster, 1996. cap. 6. p. 143-157.

NALDI, M. C.; CAMPELLO, R.J.G.B.; HRUSCHKA, E., R.; CARVALHO, A.C.P.L.F. Efficiency issues of evolutionary k-means. **Applied Soft Computing**, v. 11, ed. 2, p. 1938-1952, mar. 2011. **crossref**

NATIONAL FIRE PROTECTION ASSOCIATION. Fire Sprinkler Initiative. **Home fire sprinkler requirements at a glance**. Disponível em: <<http://www.firesprinklerinitiative.org/legislation/sprinkler-requirements-by-state.aspx>>. Acesso em 30 set. 2017.

OROZCO, C. V.; TONINI, M.; CONEDERA, M.; KANVESKI, M. Cluster recognition in spatial-temporal sequences: the case of forest fires. **Geoinformatica**, v. 16, ed. 4, p. 653-673, out. 2012. **crossref**

PALAMARA, F.; PIGLIONE, F.; PICCININI, N. Self-Organizing Map and clustering algorithms for the analysis of occupational accident databases. **Safety Science**, v. 49, p. 1215-1230, out. 2011. **crossref**

RENCHER, A. C. **Methods of Multivariate Analysis**. 2. ed. New York: John Wiley & Sons, 2002. **crossref**

ROUSSEEUW, P. Silhouettes: a Graphical Aid to the Interpretation and Validation of Cluster Analysis. **Journal of Computational and Applied Mathematics**, v. 20, p. 53-65, nov. 1987. **crossref**

RUNDMO, T.; HALE, A. R. Managers' attitudes towards safety and accident prevention. **Safety Science**, v. 41, p. 557-574, ago. 2003. **crossref**

SANDELOWSKI, M. Combining Qualitative and Quantitative Sampling, Data Collection, and Analysis Techniques in Mixed-Method Studies. **Research in Nursing & Health**, v. 223, p. 246-255, jun. 2000. **crossref**

SANDELOWSKI, M. Real Qualitative Researchers Do Not Count: The Use of Numbers in Qualitative Research. **Research in Nursing & Health**, v. 24, p. 230–240, jul. 2001. [crossref](#)

SCAWTHORN, C.; EIDINGER, J.; SCHIFF, A. J. The 1906 San Francisco, California Earthquake and Fires, **Fire Following Earthquake**. v. 11. Reston, VA: American Society of Civil Engineers, 2005.

TABOADA, H. A.; COIT, D. W. Data clustering of solutions for multiple objective system reliability optimization problems. **Quality Technology & Quantitative Management**, v. 4, ed. 2, p. 191–210, jan. 2007. [crossref](#)

TASHAKKORI, A.; TEDDLIE, C. **Handbook of Mixed Methods in Social & Behavioral Research**. Thousand Oaks: Sage Publications, 2002.

TROITZSCH, J. H. Fires, statistics, ignition sources, and passive fire protection measures. **Journal of Fire Sciences**, v. 34 (3), p. 171-198, fev. 2016. [crossref](#)

TUNG, T. X.; KIM, J. M. An effective four-stage smoke-detection algorithm using video images for early fire-alarm systems. **Fire Safety Journal**, v. 46, p. 276-282, jul. 2011. [crossref](#)

UNITED STATES FIRE ADMINISTRATION. Federal Emergency Management Agency. **Fire in the United States, 2005-2014**. 18 ed., jan. 2017.

UNITED STATES FIRE ADMINISTRATION. National Fire Data Center. **National Fire Incident Reporting System 5.0** - Complete Reference Guide, 515 p., jan. 2015.

WUSCHKE, K.; CLARE, J.; GARIS, L. Temporal and geographic clustering of residential structure fires: A theoretical platform for targeted fire prevention. **Fire Safety Journal**, v. 62, parte A, p. 3-12, nov. 2013. [crossref](#)

XIN, J.; HUANG, C. Fire risk analysis of residential buildings based on scenario clusters and its application in fire risk management. **Fire Safety Journal**, v. 62, parte A, p. 72-78, nov. 2013. [crossref](#)

Recebido: 28 Mai. 2019

Aprovado: 19 Set. 2019

DOI: 10.3895/gi.v15n3.10165

Como citar:

MACHADO, C. S. Técnicas de Clusterização para agrupamento de eventos de incêndios. **R. Gest. Industr.**, Ponta Grossa, v. 15, n. 3, p. 76-99, Jul./Set. 2019. Disponível em:

<https://periodicos.ufpr.edu.br/revistaqi>. Acesso em:.

Correspondência:

Camila da Silveira Machado

Rua Afonso Arinos, 291, ap 514. Bairro Camaquã. Porto Alegre/RS. Brasil. CEP 91910-30

Direito autoral: Este artigo está licenciado sob os termos da Licença Creative Commons-Atribuição 4.0 Internacional.



APÊNDICE

Quantificação das observações para transformar dados qualitativos em dados quantitativos

Variável	Quantificação da observação	
	Qualitativo	Quantitativo
área de origem do fogo	áreas estruturais	0,10
	áreas de ingresso	0,30
	áreas funcionais	0,50
	áreas de estocagem	0,80
fonte de calor	equipamentos em funcionamento	0,30
	objetos aquecidos	0,50
	fumo e assemelhados	0,70
primeiro item a sofrer ignição	componentes estruturais da edificação	0,10
	suplementos estocados	0,30
	mobiliário	0,40
	vestuário	0,50
	adorno, enfeite e material decorativo	0,60
	materiais gerais pesados (como fios elétricos, transformadores ou pneus)	0,70
causa da ignição	materiais gerais leves (como livros, revistas, papéis ou tecidos)	0,80
	incêndios intencionais	0,10
	causas sob investigação	0,30
	falhas do equipamento ou da fonte de energia	0,50
fator humano	causas naturais	0,80
	causas não intencionais	1,00
	possível incapacidade mental	0,10
	incapacidade física	0,20
	possivelmente sob efeito de álcool ou drogas	0,30
propagação do fogo	idade foi um fator agravante	0,40
	adormecido	0,50
	desacompanhado ou sozinho	0,70
	múltiplas pessoas envolvidas no incidente	0,80
	não houve nenhum fator humano relacionado ao incêndio	1,00
	chamas foram além da edificação de origem	0,00
condições de uso e ocupação da edificação	fogo permaneceu confinado na edificação de origem	0,25
	fogo permaneceu confinado no pavimento de origem	0,50
	fogo permaneceu confinado na dependência de origem	0,75
	chama não se propagou e permaneceu confinada no objeto que ocorreu a ignição	1,00
grau do dano causado pelo fogo ao pavimento	em demolição	0,30
	desabitada e em condições inseguras (abandonada)	0,40
	em construção	0,50
	em reforma	0,60
	desabitada e segura	0,70
	desocupada (não utilizada frequentemente)	0,80
presença de sistema de detecção automática do fogo	em uso e ocupação normal	1,00
	danos extremos	0,10
	danos severos	0,30
	danos moderados	0,60
tipo de detector automático do fogo	danos leves	0,90
	ausente	0,00
	presente	1,00
sistema de alimentação do detector automático do fogo	detecção de calor	0,60
	combinação de detecção de calor e fumaça	0,90
	detecção de fumaça	1,00
	somente a bateria	0,20
operação do detector automático do fogo	ligado à tomada	0,40
	ligado direto à fiação elétrica	0,60
	ligado à tomada com bateria	0,80
	ligado direto à fiação elétrica com bateria	1,00
efetividade do detector automático do fogo	falhou	0,10
	não foi ativado devido à baixa magnitude do fogo	0,50
	operou	1,00
	houve falha no alerta aos ocupantes	0,10
presença de sistemas automáticos de extinção do fogo	não havia ocupantes na edificação durante o incêndio	0,50
	sistema alertou os ocupantes, mas ocupantes falharam na resposta	0,80
	sistema alertou os ocupantes e os ocupantes responderam	1,00
presença de sistemas automáticos de extinção do fogo	ausente	0,20
	parcialmente instalado	0,90
	presente	1,00