

Monitoramento Ambiental de Cidades Urbanas: Detectando Outliers via Análise Fatorial Exploratória

Thiago I. A. Souza¹, Andre L. L. Aquino², Danielo G. Gomes¹

¹Universidade Federal do Ceará (UFC)
Grupo de Redes de Computadores, Engenharia de Software e Sistemas (GREAt)
Av. Mister Hull, s/n – Campus do Pici – Bloco 942–A
60455-760 – Fortaleza – CE – Brasil

²Instituto de Computação – Universidade Federal de Alagoas (UFAL)
Caixa Postal 57.072–970 – Maceió – AL – Brasil

thiagosouza@great.ufc.br, alla@laccan.ufal.br, dgomes@great.ufc.br

Abstract. *In recent years, smart cities have emerged as a vast repository of data. Thus, the need arises to detect important events that are outside the standard of normality, called outliers. In this paper, we present a new outliers detection approach to intelligent urban environment monitoring data based on the Exploratory Factor Analysis (EFA), using the following procedures: first, we apply EFA generating a factorial-base structure; in the sequence, the distance of Mahalanobis is calculated on the factors extracted for the outliers detection. Real data from the Spanish cities of Elda and Rois validated our proposal and the EFA revealed the most influential factors in the detected outliers patterns.*

Resumo. *Nos últimos anos, as cidades inteligentes têm emergido como um vasto repositório de dados. Assim, surge a necessidade de detectar importantes eventos que estão fora do padrão de normalidade, os chamados outliers. Neste artigo, propomos uma nova abordagem de detecção de outliers para dados de monitoramento de ambientes urbanos inteligentes baseada na Análise Fatorial Exploratória (AFE), através dos seguintes procedimentos: primeiro, aplicamos AFE gerando uma estrutura fatorial-base; na sequência, a distância de Mahalanobis é calculada sobre os fatores extraídos para a detecção de outliers. Dados reais das cidades espanholas de Elda e Rois validaram nossa proposta e a AFE revelou os fatores mais influentes nos padrões de outliers detectados.*

1. Introdução

De acordo com a Perspectiva de Urbanização Mundial¹ publicada pelas Nações Unidas, mais da metade da população vive atualmente em áreas urbanas e em até 2050 cerca de 70% da população global residirá nas cidades. Ao mesmo tempo que a urbanização cresce em um ritmo acelerado modificando o tecido urbano das grandes cidades juntamente com sua geometria, o fenômeno da Tecnologia da Informação e Comunicação (TIC) também cresce de forma extraordinária, sendo estimado pela CISCO² que até o final de 2018 será mais de 10 bilhões de dispositivos mobile conectados, o que

¹<http://esa.un.org/unpd/wup/>

²<http://www.cisco.com/c/en/us/solutions/service-provider/visual-networking-index-vni/index.html>

viabiliza a integração de múltiplas disciplinas, incluindo redes celulares de quinta geração (5G), redes ad hoc heterogêneas, redes móveis híbridas, redes de sensores sem fio, dentre outras, que englobam a chamada Internet das Coisas (*Internet of Things*, IoT). Como resultado da evolução dessa infraestrutura tecnológica, surge o paradigma da cidade inteligente, que integra todos os serviços urbanos habilitados pela TIC em um sistema combinado, de modo que a cidade possa ser inteligente e seja capaz de facilitar o acesso a diferentes tipos de serviços, compartilhamento de informações, monitoramento urbano em tempo real, e assim por diante [Bi et al. 2017].

IoT é uma facilitadora-chave do desenvolvimento de uma vasta gama de aplicações em cidades inteligentes, viabilizando a implementação de uma grande quantidade de nós sensores dentro de uma determinada área, uma vez que esses nós se comunicam entre si para coletar dados e fornecer referências para cidades inteligentes em seus mais diversos serviços ofertados aos cidadãos como indústria, agricultura, segurança, transporte, saúde, ambiente, dentre outros [Qiu et al. 2017]. Dessa forma, IoT está produzindo cada vez mais uma grande quantidade de diferentes tipos de dados, que inclui a localização geográfica do nó sensor, monitoramento da qualidade de água, níveis de poluição sonora e ambiental, controle de resíduos, iluminação, tráfego de veículos, dados de detecção de eventos, e assim por diante. Entretanto, com o aumento do número de dados, padrões de comportamento que se desviam da normalidade do conjunto de observações também tem crescido, denominados na literatura de *outliers* [Camacho et al. 2016]. Em particular, melhorar a robustez da detecção de eventos com base em dados estratégicos, coletados a partir do monitoramento dos diversos setores das cidades torna-se um problema crítico dada a enorme quantidade e heterogeneidade dos dados.

Diante dessa problemática, propomos neste artigo uma nova abordagem de detecção de *outliers* para dados de monitoramento de ambientes urbanos inteligentes baseada na Análise Fatorial Exploratória (AFE), através dos seguintes procedimentos: 1) AFE é realizada sobre os dados para obter uma estrutura fatorial base na qual analisamos e nomeamos os fatores a partir dos padrões identificados; 2) a distância de Mahalanobis é calculada sobre os fatores extraídos para a detecção de eventos fora dos padrões de normalidade do conjunto observado. Dados reais das cidades da Espanha, Elda e Rois, coletados da plataforma de monitoramento ambiental Smart Citizen, foram utilizados para validar nossa abordagem.

O restante deste artigo está organizado como se segue. A Seção 2 apresenta os trabalhos relacionados. Na Seção 3 a metodologia proposta para a nova abordagem de detecção de *outliers* via AFE é apresentada. A metodologia proposta é validada para um conjunto de dados reais na Seção 4. Por fim, na Seção 5, são apresentadas as considerações finais e trabalhos futuros.

2. Trabalhos Relacionados

Vários esforços na literatura foram realizados na perspectiva de propor metodologias e esquemas de detecção de *outliers* a partir do monitoramento de determinadas variáveis ambientais no intuito de contribuir para o desenvolvimento das cidades inteligentes. Por exemplo, [Zanella et al. 2014] apresentou um levantamento abrangente de arquiteturas, protocolos e tecnologias para a proposta de uma metodologia baseada

em serviços web para o projeto da cidade inteligente de Padova (Itália), em que sua implementação engloba soluções que vislumbram o monitoramento e detecção de eventos anormais a partir de dados de iluminação pública e qualidade do ar. Um estudo sobre os fundamentos de IoT no desenvolvimento das cidades inteligentes foi realizado em [Jin et al. 2014], que analisou a ocorrência de *outliers* extraídos a partir do monitoramento de ruído. Já [Filipponi et al. 2010], em seu trabalho apresentou um esquema para implementação de serviços de informação para o monitoramento de áreas públicas e infraestruturas, testando sua abordagem em um cenário de transporte público (metrô) apresentando um sistema que auxilia na detecção de *outliers* e simplifica as comunicações em casos de emergência no transporte. [Souza et al. 2017] propôs uma nova abordagem de detecção de *outliers* a partir do monitoramento ambiental de dados coletados de diferentes e heterogêneos sensores através da aplicação de técnicas da estatística multivariada, como a Análise de Componentes Principais (PCA), trazendo como contribuição o diagnóstico das causas que geraram os eventos discrepantes, indicando quais variáveis ambientais contribuíram para o comportamento anômalo dos dados.

O presente trabalho diferencia-se dos anteriores por combinar a técnica multivariada Análise Fatorial Exploratória com a distância de Mahalanobis no intuito de extrair os fatores latentes mais representativos para a detecção de *outliers* a partir do sensoriamento de medidas ambientais heterogêneas de ambientes urbanos inteligentes.

3. Material e Métodos

Esta seção apresenta o delineamento metodológico no qual informamos como foram realizadas a coleta e a organização dos dados, a decorrente modelagem da análise fatorial exploratória e a detecção de *outliers*.

3.1. Coleta e Organização de Dados

A plataforma Smart Citizen³, utilizada em pesquisas anteriores [McKercher et al. 2017, Souza et al. 2017], é uma ferramenta de *crowdsourcing* que coleta dados ambientais urbanos e tem a capacidade de enviar dados para a Internet. O dispositivo mede variáveis ambientais como temperatura, umidade, luminosidade, monóxido de carbono, dióxido de nitrogênio e níveis de ruído, podendo enviar dados via Wi-Fi que podem ser visualizados online em tempo real. Diante do exposto, as medidas de tais variáveis (registradas por um período de 31 dias do mês de Julho de 2017) de duas cidades da Espanha, Elda e Rois, foram coletadas e organizadas em matrizes, cada uma representando uma respectiva cidade. Assim, a análise fatorial exploratória é realizada sobre as matrizes geradas, cada uma com dimensões, 31 (tempo em dias) \times 6 (variáveis ambientais).

3.2. Modelagem da Análise Fatorial Exploratória

A Análise Fatorial Exploratória (AFE) é um dos métodos da estatística multivariada que tem como objetivo principal a redução de dimensionalidade e identificar as relações subjacentes entre as variáveis medidas, determinando o número e a natureza apropriada dos fatores comuns (fatores latentes) necessários para explicar a matriz de correlação observada [Bartholomew and Knott 1999]. Assim, podemos distinguir

³<https://smartcitizen.me/>

ou classificar as variáveis de acordo com as contribuições dos fatores latentes para cada variável individualmente. O modelo expressa um vetor \mathbf{x} m -dimensional (variáveis ambientais) como

$$\mathbf{x} = \mathbf{\Lambda}\mathbf{f} + \mathbf{e}, \quad (1)$$

em que $\mathbf{\Lambda}$ ($m \times k$) é a matriz de coeficientes ou fatores de carregamento e \mathbf{f} é o vetor k -dimensional dos fatores comuns (fatores latentes), com $k \leq m$, e \mathbf{e} é um vetor m -dimensional dos termos residuais do modelo (fatores específicos).

Para simplificar ainda mais nosso modelo original, adotamos um modelo de fatores ortogonais com base em três pressupostos [Basilevsky 2009]: (i) a média do vetor dos fatores comuns \mathbf{f} é zero e a matriz de covariância é a identidade; (ii) a média dos termos do vetor de erro \mathbf{e} é zero e a matriz de covariância é diagonal; e (iii) os termos do vetor de erro não tem correlação com os fatores comuns.

Após a extração dos fatores latentes baseados no modelo da análise fatorial, um índice importante a ser obtido é a comunalidade. A comunalidade para qualquer variável pode ser interpretada como a proporção da variabilidade dessa variável explicada por todos os fatores comuns [Basilevsky 2009]. Assim, a comunalidade h_i para a i -ésima variável é definida como

$$h_i = \sum_{j=1}^k \lambda_{ij}^2, \quad (2)$$

na qual λ_{ij} é o coeficiente do fator de carregamento da i -ésima variável no j -ésimo fator comum.

Além das comunalidades, calculamos os escores dos fatores que são definidos como combinações lineares das variáveis observadas. A estimativa dos escores dos fatores com base na regressão é dada diretamente aqui sem derivação [Bartholomew and Knott 1999]

$$\mathbf{S} = \mathbf{\Lambda}\mathbf{R}^{-1}\mathbf{X}, \quad (3)$$

em que \mathbf{S} denota os escores dos fatores e \mathbf{R} é a matriz de correlação das variáveis observadas, representadas pela matriz \mathbf{X} .

3.3. Detecção de *Outliers*

A distância de Mahalanobis é uma estatística responsável por indicar variações desiguais, identificando valores ou eventos discrepantes de um determinado conjunto de observações [Hotelling 1947]. Portanto, uma vez que a detecção de *outliers* visa estabelecer se um novo padrão é significativamente diferente de um padrão anterior, utilizamos tal medida combinada com a AFE na detecção de eventos discrepantes no cenário de ambientes urbanos inteligentes. A distância de Mahalanobis nesta pesquisa é dada por

$$D_M^2 = (\mathbf{s} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{s} - \boldsymbol{\mu}), \quad (4)$$

na qual s é o vetor de escores dos fatores calculados na Equação 3, μ é a média dos escores dos fatores, Σ é a matriz de covariância dos escores e D_M^2 é a estatística de desvio.

Portanto, neste trabalho a distância de Mahalanobis D_M é calculada sobre os vetores s de escores dos fatores extraídos da AFE de cada cidade e comparada com um valor limite calculado. Os limites aproximados de controle da distância de Mahalanobis, com um nível de confiança α podem ser determinados de diferentes maneiras, aplicando-se os pressupostos de distribuição de probabilidade [Tracy et al. 1972]

$$T_\alpha = \frac{d(n^2 - 1)}{n(n - d)} F_\alpha(d, n - d), \quad (5)$$

em que $F_\alpha(d, n - d)$ é o limite superior do percentil da distribuição F com d e $n - d$ graus de liberdade. Assim, se $D_M^2 > T_\alpha$, então as observações são consideradas *outliers*, caso contrário, normais.

4. Resultados

A análise fatorial exploratória foi realizada sobre o conjunto de dados que engloba as 6 variáveis observadas durante um período de 31 dias do mês de Julho de 2017 de duas cidades da Espanha, Elda e Rois, em que uma solução de dois fatores foi derivada para cada cidade. Nas subseções seguintes serão apresentados e discutidos os critérios de validação da AFE bem como a rotulagem de cada fator extraído e suas interpretações, e a etapa de detecção de *outliers* baseada na análise fatorial. Todas as simulações realizadas nesta pesquisa foram implementadas no *software* MATLAB, utilizando um processador *Intel Core I5* com velocidade de 2.66GHz e 4GB de memória.

4.1. Validação dos Resultados

Análises preliminares foram realizadas para examinar a adequação dos dados à AFE (Tabela 1). Para tanto, no intuito de verificar se as variáveis analisadas são correlacionadas entre si, gerando a hipótese de a matriz de correlação das variáveis ser identidade, os testes de Kaiser-Meyer-Olkin (KMO) e Bartlett foram aplicados sobre os dados de cada cidade. Conforme observado na Tabela 1, o valor do teste KMO se mostrou significativo tanto para os dados da cidade de Elda (0,71), quanto da cidade de Rois (0,75), garantindo uma boa adequação da amostra a aplicação da AFE, uma vez que possui valor superior a 0,6 [Bartholomew and Knott 1999]. Já o teste de Bartlett rejeitou a hipótese de que a matriz de correlação seria a matriz identidade [Basilevsky 2009]. Desta forma, ambos os métodos mostraram que os dados são adequados para aplicação da AFE.

Tabela 1. Testes de Validação da AFE.

Cidade	Adequação da Amostra - KMO	Esfericidade de Bartlett
Elda	0,71	489,08
Rois	0,75	435,89

4.2. Análise das Comunalidades

A variância extraída de cada nova variável foi comparada com as variâncias das variáveis originais, verificando o quanto de variância comum (comunalidade), existe entre

as variáveis observadas e as que foram obtidas através da AFE. Desta forma, conforme observado na Tabela 2, os valores de comunalidade das variáveis apresentam valores superiores a 0,6, indicando que todas as variáveis, para ambas as cidades, apresentam elevada representatividade dentro dos fatores extraídos pela AFE.

Tabela 2. Valores de Comunalidade das Variáveis Ambientais das Cidades.

Cidade	Variáveis	Extração	Cidade	Variáveis	Extração
Elda	Temperatura	0,84	Rois	Temperatura	0,74
	Umidade	0,89		Umidade	0,69
	Luminosidade	0,81		Luminosidade	0,67
	NO ₂	0,93		NO ₂	0,85
	CO	0,95		CO	0,92
	Ruído	0,99		Ruído	0,74

4.3. Seleção e Interpretação dos Fatores

Para a seleção do número de fatores, foi utilizado o critério da variância explicada [Sundberg and Feldmann 2016, Souza et al. 2017] cujos primeiros dois fatores explicam cerca de 73% da variância total para a cidade de Elda (Tabela 3), e cerca de 83% da variância para a cidade de Rois (Tabela 4). Além da variância explicada, foi utilizado também o critério de Kaiser [Kaiser 1966, Souza et al. 2017], que diz que os fatores a serem considerados devem apresentar autovalores acima da unidade ($\lambda > 1$, segunda coluna da Tabela 3 e Tabela 4).

Tabela 3. Distribuição da Variância Explicada da AFE - Cidade de Elda.

Fatores	Autovalores (λ)	% Variância	% Variância Acumulativa
1	2,89	48,17	48,17
2	1,51	25,18	73,35
3	0,99	16,69	90,05
4	0,32	5,44	95,49
5	0,18	3,13	98,62
6	0,08	1,39	100

Tabela 4. Distribuição da Variância Explicada da AFE - Cidade de Rois.

Fatores	Autovalores (λ)	% Variância	% Variância Acumulativa
1	2,17	46,23	46,23
2	1,61	36,90	83,14
3	0,98	7,41	90,55
4	0,57	5,66	96,22
5	0,47	3,13	99,35
6	0,17	0,67	100

Os fatores de carregamento permitem que uma correlação possa ser estabelecida entre as variáveis observadas e os fatores extraídos. Desta forma, tanto para a cidade de Elda (Tabela 5) quanto para a cidade de Rois (Tabela 6), todas as cargas com valores superiores a 0,6 estão destacadas em negrito. Nesta pesquisa, a análise desses fatores se

estabeleceu a partir do cruzamento de cargas elevadas com as demais variáveis. Assim, a partir do padrão observado deste cruzamento, nomeamos os fatores extraídos de cada cidade que obtiveram os maiores valores, conforme discutido a seguir:

- **Fatores de Carregamento da Cidade de Elda** - Analisando os valores dos fatores das cargas desta cidade (Tabela 5), ambos têm em comum o fato de se referirem prioritariamente às variáveis climáticas, uma vez que para as variáveis, temperatura e umidade, os fatores apresentaram os maiores valores para o Fator I. Dessa forma, para fins de análise nomeamos o Fator I como **Condições Climáticas**.
- **Fatores de Carregamento da Cidade de Rois** - Os fatores de carregamento para esta cidade (Tabela 6) apresentam características relacionadas à poluição da cidade. Isto devido ao fato das variáveis monóxido de carbono (CO) e dióxido de nitrogênio (NO_2) apresentarem os maiores fatores de carregamento para o Fator II. Assim, o Fator II recebeu a denominação de **Qualidade do Ar**.

Tabela 5. Fatores de carregamento da AFE - Cidade de Elda.

Variáveis Ambientais	Fator I	Fator II
Temperatura	0,99	0,59
Umidade	0,95	0,53
Luminosidade	0,03	0,15
NO_2	0,11	0,167
CO	0,41	0,09
Ruído	0,20	0,11

Tabela 6. Fatores de carregamento da AFE - Cidade de Rois.

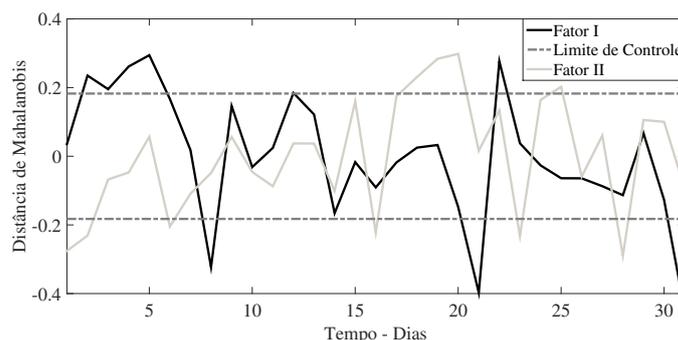
Variáveis Ambientais	Fator I	Fator II
Temperatura	0,39	0,38
Umidade	0,43	0,56
Luminosidade	0,19	0,15
NO_2	0,54	0,91
CO	0,49	0,95
Ruído	0,01	0,32

Os fatores derivados do modelo exploratório fatorial obtidos neste estudo podem ser utilizados como ferramenta para identificar padrões de eventos relacionados as variáveis ambientais que estejam fora do padrão de normalidade. Como é amplamente conhecido que uma cidade com cidadãos pouco saudáveis dificilmente se torna uma cidade inteligente, uma vez que tais variáveis impactam diretamente na vida dos cidadãos, esses fatores podem revelar padrões de anormalidade que permaneceriam invisíveis frente a uma análise de dados que apenas explorasse a natureza descritiva dos dados.

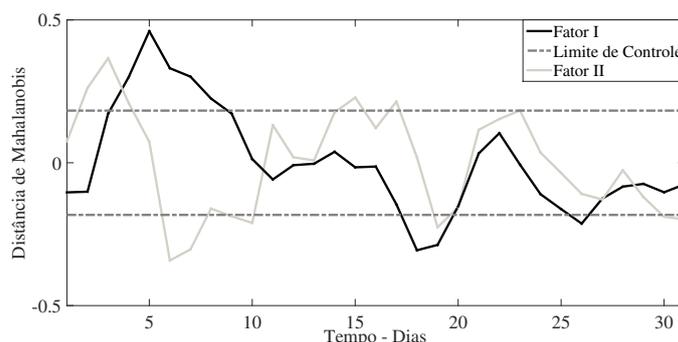
4.4. Detecção de *Outliers* via Análise Fatorial Exploratória

A partir dos fatores extraídos, de acordo com os critérios de variância explicada e Kaiser, calculamos a distância de Mahalanobis para ambos os fatores extraídos de cada cidade, no intuito de serem utilizados como uma distribuição de referência empírica para

estabelecer uma região gráfica de controle para o monitoramento do comportamento das variáveis ambientais. Assim, se os valores da estatística permanecem dentro das regiões de controle, não há evidências de que o processo em análise sofreu algum tipo de mudança. Entretanto, caso os valores em algum instante sejam traçados fora do limiar de controle, há evidências de que o processo sofreu alguma tipo de alteração. Portanto, quando os fatores extraídos têm uma interpretação clara, o gráfico estatístico que descreve o comportamento dos fatores fornece uma ilustração visual útil para analisar os perfis das variáveis ambientais das cidades.



(a)



(b)

Figura 1. Detecção de *outliers* para: (a) Cidade de Elda; (b) Cidade de Rois.

Os resultados da detecção de *outliers* através da distância de Mahalanobis aplicada sobre os dois fatores extraídos de cada cidade são apresentados na Figura 1 e Tabela 7. É importante destacar que o parâmetro T_α , calculado na Equação 5 da Seção 3.3, foi utilizado nesta pesquisa para delimitar a fronteira de monitoramento, no intuito de inferir qual região do gráfico está fora dos limites de controle. Para a cidade de Elda, verificamos conforme a Tabela 7 que 32,25% e 29,03% de eventos do Fator I e Fator II, respectivamente, estavam fora das fronteiras estabelecidas pelo limite de controle. Já a cidade de Rois, para o Fator I e Fator II, respectivamente, os percentuais foram de 25,80% e 41,94% (Tabela 7) de eventos que ultrapassaram os limites de controle. Esses resultados permitem-nos analisar o dia em que determinado evento anormal aconteceu, e assim verificar o instante do ocorrido, podendo a informação ser utilizada pelo órgão público responsável por monitorar padrões ambientais bem como servir de *insights* para as tomadas de decisões futuras por parte dos gestores públicos.

Na Figura 1a, T_α é o parâmetro delimitador dos valores da distância de Mahala-

Tabela 7. Distribuição de *outliers* por fator.

	Fator I (%)	Fator II (%)
Elda	32,25	29,03
Rois	25,80	41,94

nobis sobre os dois fatores extraídos. Observamos que na maioria dos dias não há alertas para eventos fora dos limites, contudo o Fator I supera em cerca de 3% no número de *outliers* em relação ao Fator II. Este resultado permite-nos inferir que o Fator Condições Climáticas (conforme nomeado na subseção 4.3) foi o responsável por influenciar na geração de *outliers* para a cidade de Elda, revelando um importante comportamento discrepante por parte das variáveis ambientais, temperatura e umidade, em determinados dias do mês de Julho. Este fator pode ser útil para revelar padrões de conforto ou desconforto climático da cidade.

Na Figura 1b, observa-se novamente determinados instantes em que o valor da distância de Mahalanobis é superior ao limite de controle T_α , sinalizando alertas de ocorrência de eventos discrepantes. Verificamos que apesar do Fator I registrar o maior pico no dia 5 de Julho, o Fator II supera em cerca de 16% na quantidade de *outliers* detectados, sendo o fator responsável por influenciar significativamente na geração de tais eventos. Portanto, o Fator Qualidade do Ar (conforme denominado na subseção 4.3), aponta para o comportamento discrepante das variáveis monitoradas relacionadas a poluição do ar, monóxido de carbono (CO) e dióxido de nitrogênio (NO_2), em determinados períodos do mês de Julho de 2017 da cidade de Rois. A análise deste fator pode ser útil para apontar tendências dos níveis de qualidade do ar, contribuindo para o monitoramento efetivo dos índices de poluição ambiental da cidade.

5. Conclusão

Neste artigo propomos uma nova abordagem de detecção de *outliers* para dados de monitoramento de ambientes urbanos inteligentes baseada na Análise Fatorial Exploratória (AFE). Através dos resultados alcançados com a aplicação da AFE, uma estrutura fatorial-base foi obtida revelando os fatores latentes mais representativos, os quais foram nomeados, a saber: **Fator Condições Climáticas** e **Fator Qualidade do Ar**. A partir dos fatores latentes extraídos pelo modelo multivariado, a distância de Mahalanobis foi calculada sobre os escores dos fatores, tomando os valores da estatística como um recurso de identificação de eventos discrepantes, caso ultrapassem o limite de controle estabelecido. Padrões de *outliers* foram identificados para ambos os fatores: para o Fator Condições Climáticas, constatamos que as variáveis ambientais, temperatura e umidade, foram as responsáveis por gerar o comportamento discrepante; para o Fator Qualidade do Ar, as variáveis ambientais, monóxido de carbono e dióxido de nitrogênio, foram as que influenciaram no comportamento anômalo dos dados.

Como perspectivas de trabalhos futuros, sugere-se: (i) aumentar o número de amostras e de cidades; (ii) realizar uma análise de correlação entre variáveis ambientais; (iii) utilizar os fatores da AFE para a análise de *cluster*.

Agradecimentos

Thiago Iachiley agradece a CAPES, André L. L. Aquino agradece ao CNPq, FAPEAL e FAPESP e Danielo G. Gomes agradece ao CNPq (respectivamente, processos 88882.183548/2018-01, 311878/2016-4 e 432585/2016-8) pelo apoio financeiro.

Referências

- Bartholomew, D. J. and Knott, M. (1999). *Latent Variable Models and Factor Analysis*. Arnold Publishers.
- Basilevsky, A. T. (2009). *Statistical factor analysis and related methods*. John Wiley and Sons.
- Bi, Y., Lin, C., Zhou, H., Yang, P., Shen, X., and Zhao, H. (2017). Time-constrained big data transfer for sdn-enabled smart city. *IEEE Communications Magazine*, 55:44–50.
- Camacho, J., Villegas, A. P., Teodoro, P. G., and Fernandez, G. M. (2016). Pca-based multivariate statistical network monitoring for anomaly detection. *Computers and Security*, 59:118–137.
- Filipponi, L., Vitaletti, A., Landi, G., Memeo, V., Laura, G., and Pucci, P. (2010). Smart city: An event driven architecture for monitoring public spaces with heterogeneous sensors. In *Sensor Technologies and Applications, SENSORCOMM*, pages 281–286. Fourth International Conference.
- Hotelling, H. (1947). *Multivariate quality control*. In: Techniques of statistical analysis. New York: McGraw-Hill.
- Jin, J., Gubbi, J., Marusic, S., and Palaniswami, M. (2014). An information framework for creating a smart city through internet of things. *IEEE Internet of Things Journal*, 1:112–121.
- Kaiser, H. F. (1966). The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, 20:141–151.
- McKercher, G. R., Salmond, J. A., and Vanos, J. K. (2017). Characteristics and applications of small, portable gaseous air pollution monitors. *Environmental Pollution*, 223:102–110.
- Qiu, T., Liu, J., Si, W., Han, M., Ning, H., and Atiquzzaman, M. (2017). A data-driven robustness algorithm for the internet of things in smart cities. *IEEE Communications Magazine*, 55:18–23.
- Souza, T. I. A., Magalhães, D. M. V., and Gomes, D. G. (2017). Aplicando estatística multivariada para detecção e diagnóstico de anomalias em dados urbanos. *Anais do I Workshop de Computação Urbana (CoUrb)*, 1:72–85.
- Sundberg, R. and Feldmann, U. (2016). Exploratory factor analysis—parameter estimation and scores prediction with high-dimensional data. *Journal of Multivariate Analysis*, 148:49–59.
- Tracy, N. D., Young, J. C., and Mason, R. L. (1972). Multivariate control charts for individual observations. *Expert Systems With Applications*, 24:88–95.
- Zanella, A., Bui, N., Castellani, A., Vangelista, L., and Zorzi, M. (2014). Internet of things for smart cities. *IEEE Internet of Things Journal*, 1:22–32.