

# Visual Analytics e Outlying Aspect Mining: contextualização de anomalias considerando questões temporais e multidimensionais

Felipe Marx Benghi<sup>1</sup>, Luiz Gomes-Jr<sup>1</sup>

<sup>1</sup>Programa de Pós-Graduação em Computação Aplicada  
Universidade Tecnológica Federal do Paraná (UTFPR) – Curitiba, PR – Brazil

`felipebenghi@alunos.utfpr.edu.br`, `gomesjr@dainf.ct.utfpr.edu.br`

**Abstract.** *Outlying Aspect Mining (OAM) is a new way of handling outliers that, instead of focusing solely on the detection, also provides an explanation. This is done by presenting a subspace of attributes that had the most abnormal behavior. Acknowledging this group of attributes is important but only listing them is not sufficient for a human specialist to comprehend the situation and take the necessary actions. A higher-level, visual approach can improve the process, providing better cognitive clues to experts. Here we describe a Visual Analytics platform developed to present data and OAM outputs in a human-friendly interface. A novelty available on this platform is a parallel coordinates plot that also display temporal multidimensional data. Such representation overcome human visual system limitations and helps in the outlier investigation. To explore the applicability of the developed tool, a locomotive operation user case is employed with focus on fault analysis in an OAM point of view.*

## 1. Introdução

A crescente integração entre sistemas e a maior disponibilidade e variabilidade de sensores faz com que conjuntos de dados de alta dimensionalidade e variantes com o tempo sejam cada vez mais comuns. Por exemplo, uma única locomotiva moderna em operação pode gerar dados para centenas de variáveis, com atributos discretos para a supervisão do acionamento de diferentes equipamentos e variáveis contínuas para a inspeção de grandezas físicas como a temperatura dos motores, tensão de geradores etc.

Uma das áreas importantes neste contexto é a detecção de anomalias ou *outliers*. O estudo dos elementos que destoam (*outliers*) das demais amostras em um conjunto (*inliers*) são aplicados para detecções de fraudes financeiras, identificação de invasores e até mesmo diagnósticos de câncer [Chandola et al. 2009]. Uma abordagem mais recente, *Outlying Aspect Mining* (OAM) [Vinh et al. 2016], não se limita a somente a encontrar anomalias, mas também busca apresentar uma explicação ou interpretação para o componente anômalo explorando a alta dimensionalidade disponível. Para tanto, tenta selecionar atributos que contextualizem de forma mais representativa as diferenças do ponto sob análise para os outros elementos.

Apesar dos avanços trazidos pela multidimensionalidade dos bancos de dados atuais, a análise de especialistas continua imprescindível porém mais complexa. Uma das dificuldades que surgem neste contexto decorre da limitação do sistema visual humano a três dimensões. Como a representação gráfica dos dados com dimensionalidade mais

alta em superfícies e monitores implica em perdas ou simplificações, soluções alternativas precisam ser consideradas para tornar dados palpáveis a pessoas.

Neste sentido, as técnicas de Visual Analytics ou Análise Visual (AV) frequentemente vêm ao auxílio dos especialistas pois não se limitam à geração de imagens, mas também buscam determinar o que e como os elementos devem ser exibidos a partir da vasta quantidade e variedade de dados disponíveis. Para isto, AV engloba pesquisas sobre a interação homem-máquina, gerenciamento/análise de dados e estatísticas [Keim et al. 2010] e tenta ampliar a capacidade humana de compreender e raciocinar sobre o mundo através de representações visuais [Thomas and Cook 2005].

A preocupação com a análise e identificação de anomalias está presente em diversos estudos de Análise Visual [Novotny and Hauser 2006, Webga and Lu 2015]. No entanto, existe um deficit em pesquisas que busquem endereçar a interpretação de anomalias do ponto de vista de OAM. Tal solução é relevante porque a simples identificação de falhas não é suficiente, também é necessário que a explicação dos eventos fornecida pelas técnicas de OAM sejam apresentadas de forma que especialistas possam tomar as atitudes necessárias para, por exemplo, prevenir situações similares, função que normalmente cabe às técnicas de Análise Visual.

Uma plataforma de Análise Visual adequada ao contexto de OAM possui algumas especificidades em relação a uma ferramenta tradicional:

- Priorização de atributos: em OAM, a explicação para uma anomalia se dá através da seleção dos atributos mais relevantes para tal.
- Necessidade de exibição do componente temporal: a plataforma deve guiar o especialista em questões como “Quando o problema começou?” ou “Existiu algum evento que desencadeou o processo?”.
- Preferência por técnicas que mantenham os atributos originais: a aplicação de técnicas como Principal Component Analysis (PCA) que constroem um novo espaço vetorial dificultariam a explicação das falhas por não representarem grandezas reais.

Este artigo descreve a implementação de uma plataforma de Análise Visual aplicada a OAM. O caso de uso para a demonstração da plataforma é o de operação de locomotivas de carga (Seção 3.1). A plataforma permite a exploração de eventos anômalos sob o ponto de vista temporal (Seção 3.3) e multidimensional (Seção 3.4).

## 2. Trabalhos relacionados

*Outliers* ou anomalias são definidos como amostras que se destoam do esperado, i.e. estão tão distantes das demais observações em um conjunto de dados (*inliers*), a ponto de ser razoável supor que foram originadas por um processo distinto ou que seguem uma lógica própria [Hawkins 1980, Cielen et al. 2016]. A detecção ou identificação das anomalias é um ramo de aprendizado de máquina bastante desenvolvido, com aplicações na previsão de falhas em plantas industriais e detecção de intrusos em redes.

Diferentemente, *Outlying Aspect Mining* (OAM) [Vinh et al. 2016] não se contenta em encontrar as amostras anômalas, mas procura também explicar o que objeto possui de diferente a ponto de ser considerado um *outlier*. Para tanto, buscam-se pelos atributos que apresentem um comportamento atípico e que, por isso, possuem maior

potencial para representarem uma interpretação para a condição anômala das amostras. Tipicamente, o resultado é fornecido através de uma classificação de subespaços, i.e. diferentes subconjuntos de atributos, que obtiveram maior pontuação em uma métrica de anormalidade.

*Isolation path* [Vinh et al. 2016] é um algoritmo de OAM derivado do *isolation forest* [Liu et al. 2008], originalmente focado na detecção de anomalias. Ambos algoritmos constroem árvores binárias através da sucessiva divisão do conjunto de dados e usam número de partições necessárias para isolar uma amostra como métrica de anormalidade. Esta estratégia considera que *outliers* são mais raros e possuem valores mais extremos que objetos normais, portanto, menos divisões seriam necessárias para o separá-los. Diferentes estratégias podem ser utilizadas para o particionamento, mas em sua forma original, ela é feita através da escolha aleatória de um valor dentro do intervalo possível para um atributo. Com base nesta referência, amostras com valores menores são direcionadas para um ramo da árvore binária e elementos com valores maiores, para outro ramo.

Dentre as vantagens de *isolation path*, destacam-se a ausência de cálculos custosos, como medidas de densidade e distância, e a manutenção de propriedades mesmo em subespaços de diferentes dimensionalidades. Tais características são importantes no contexto de OAM, pois a métrica de anormalidade precisa ser calculada para as diversas combinações de subespaços e cálculos custosos em termos de processamento seriam impeditivos dado a possível explosão combinatorial de possibilidades.

Técnicas de OAM são importantes para identificar subespaços de interesse, mas não são suficientes para a compreensão do contexto que determina a excepcionalidade do *outlier*. Esta proposta visa complementar tais técnicas com mais recursos visuais de suporte aos especialistas. *Isolation path* é a técnica escolhida para os testes desta proposta, mas nossa interface é adequada para qualquer técnica de OAM que liste subespaços de interesse.

## **2.1. Visual Analytics e Parallel Coordinate Plots (PCP)**

Mesmo com o aumento da disponibilidade dos dados e de técnicas para extração de conhecimento, a atuação humana continua fundamental na interpretação do comportamento de sistemas complexos. No entanto, a quantidade de amostras e variáveis que uma pessoa consegue processar é limitada e, por isso, estes elementos precisam ser apresentados de forma adequada à capacidade cognitiva humana.

Neste sentido, para [Chen et al. 2011], a Análise Visual (AV) procura, através do uso de interface visual, criar um ambiente propício para a interação humana com os dados e, assim, gerar-se análises mais aprofundadas. A Análise Visual é mais ampla que a mera visualização das informações e engloba também transformação e interação com os dados.

Dentre as técnicas de visualização de dados, Parallel Coordinate Plots (PCPs) ou Gráficos de Coordenadas Paralelas [Inselberg 1985] são muito exploradas para a representação de dados multidimensionais. Esta técnica, na sua forma clássica, utiliza eixos paralelos (horizontais ou verticais) para representar as dimensões e polilinhas para indicar as amostras. O ponto em que uma polilinha intersecta um eixo reflete o valor daquela amostra para o respectivo atributo. Dentre as vantagens desta forma de visualização, salienta-se a possibilidade de representação teoricamente ilimitada de dimensões e a exibição da geometria das amostras ao invés de somente as pontos [Blaas et al. 2008].

Nestes mais de trinta anos desde a sua concepção, diversas variações de PCPs foram propostas com o objetivo de agregar elementos à forma clássica e facilitar a interpretação dos dados. Por exemplo, as polilinhas já foram substituídas por curvas (interpolações), representações de densidades e polígonos [Heinrich and Weiskopf 2013, Johansson et al. 2007]. Mais recentemente, imagens em 3D começaram a ganhar destaque, com [Tadeja et al. 2019] ou sem [Zhonghua and Lingda 2016] realidade virtual. Um dos problemas mais comuns nos gráficos de coordenadas paralelas é o excesso de polilinhas, comumente chamado de *overplotting*. Isto ocorre porque cada amostra em um conjunto de dados teoricamente pode ser representado por uma polilinha. No entanto, uma grande quantidade destes elementos pode tornar qualquer interpretação inviável, dada a confusão de formas.

A seleção de subconjuntos de dados, ou *brushing*, é uma das formas mais usuais e simples de enfrentar esta limitação. Por meio de plataformas interativas, usuários podem salientar ou mesmo criar novas imagens com amostras de interesse [Blaas et al. 2008, Sansen et al. 2017a, Tong et al. 2019]. No entanto, este tipo de análise é pouco escalável e pode requerer conhecimento prévio dos dados. Outra alternativa é a introdução de técnicas de clusterização. Neste âmbito, [Novotny and Hauser 2006] aplica *binning* aos dados *inliers* ao mesmo tempo que preserva *outliers* em sua forma original. [Sansen et al. 2017b] vai mais além ao exibir no próprio PCP elementos visuais que indiquem a distribuição de amostras e intervalos em um mesmo cluster.

Quanto à preocupação com a representação de *outliers*, pode-se dizer que ela é bastante recorrente em PCPs. Nestas propostas, anomalias são exibidas de forma diferenciada, o que graficamente se materializa através de cores chamativas ou texturas próprias [Blaas et al. 2008, Novotny and Hauser 2006, Glendenning et al. 2016].

Por sua vez, adaptações aos gráficos de coordenadas paralelas que possibilitem a ilustração da evolução temporal dos elementos é mais rara e diversa, sendo portanto um campo em aberto. [Johansson et al. 2007], por exemplo, propõem duas técnicas para a visualização de séries temporais com PCPs. Na primeira, *Temporal Density Parallel Coordinates*, as polilinhas são substituídas por polígonos, que de acordo com sua tonalidade comunicam a densidade dos valores através do período considerado, i.e. regiões com alta reincidência de valores através do tempo são representadas de forma mais vívida no gráfico. Na segunda técnica, *Depth Cue Parallel Coordinates*, um gradiente de cores é usado para indicar a diferença temporal das amostras, assim, amostras mais recentes são exibidas com cores mais fortes e aparentam estar no primeiro plano.

[Blaas et al. 2008] não exibe dados coletados em instantes diferentes em um mesmo PCP. Entretanto, reconhece que a comparação entre diferentes instantes de tempo é relevante e possibilita que o usuário de sua interface interativa avance/regrida no tempo deslizando um marcador. Similarmente, em [Barlow and Stuart 2004] permite-se que sejam escolhidos os instantes para os quais PCP é exibido ou mesmo que seja observada a evolução de valores por meio de uma animação.

[Zhonghua and Lingda 2016] exibe simultaneamente múltiplos PCPs referentes a instantes diferentes por meio de visualizações 3D. Para tanto, os gráficos são alinhados paralelamente ao ponto de vista do observador, com amostras mais antigas sendo exibidas mais distantes e com cores mais transparentes. Os eixos para cada atributo foram

**Tabela 1. Variáveis da locomotiva analisadas ou controladas.**

Equipamento	Variável contínua
Motor diesel	Temperatura óleo antes [TOA] e depois [TOD] do arrefecimento (°C) Temperatura da água (°C) [TH2O] Pressão lubrificante [PLUB] e combustível [PCOMB] (psi) Rotação (rpm) [ROT] / Potência [POT]
Gerador Principal	Tensão (V) [VGP] / corrente (A)[IGP]
Gerador Companheiro	Tensão (V) [VGA] / corrente (A) [IGA]
Motores de tração	Corrente (A) [MT1]
Geral	Velocidade de deslocamento [VEL], Ponto de operação [PTO]

substituídos por planos que permitem a transformação dos valores para cada ponto com o tempo. Na opinião dos próprios autores, o excesso de elementos gráficos e sua consequente confusão de formas, dificultam a interpretação e obtenção de análises relevantes para o que foi proposto.

Mesmo com esta grande diversidade de topologias e aplicações de PCPs apresentados, os autores deste artigo desconhecem outros trabalhos que proponham a representação de séries temporais juntamente com anomalias em um único gráfico de coordenadas paralelas. Além disso, esta contribuição mostra-se relevante no contexto de análise de anomalias uma vez que, a partir de uma perspectiva multidimensional, permitiria a análise da degradação temporal de um sistema culminante em uma falha.

### **3. Interface de análise visual para interpretação de outliers**

#### **3.1. Caso de estudo**

Para demonstrar a aplicabilidade da plataforma desenvolvida, escolheu-se uma situação de falha em locomotiva relacionada à temperatura do motor diesel. Esta anomalia normalmente começa a aparecer a medida que o ciclo de vida dos trocadores de calor se aproxima do fim.

Dentre as variáveis disponíveis para análise (apresentadas na Tabela 1), a aplicação de OAM revelou corretamente que a temperatura do óleo antes/depois do arrefecimento (TOA e TOD) e temperatura da água (TH2O) seriam mais relevantes à interpretação desta falha.

#### **3.2. Visão Geral**

A interface proposta enfatiza os seguintes aspectos para auxiliar no processo de interpretação de *outliers*: (i) exibição temporal de múltiplas variáveis, (ii) marcação clara do subconjunto de dados no momento da anomalia, (iii) sugestão de subconjuntos de dados relevantes e (iv) possibilidade de cambiamento para visualização de séries temporais. O desenvolvimento ocorreu em Python, com Matplotlib como biblioteca gráfica. A configuração das imagens é feita por arquivos de configuração. No futuro planeja-se implementar uma abordagem mais interativa e disponibilizá-la em repositórios abertos.

Para exibição multidimensional usamos PCP como pode ser visto na Figura 1. Para indicar o momento de identificação do *outlier*, usamos um esquema de cores, com verde representando o período antes da anomalia e vermelho o período posterior.

Para casos complexos, pode ser interessante analisar os dados como séries temporais. Porém, é preciso reduzir a dimensionalidade para que a visualização não fique sobrecarregada. Neste caso oferecemos ao especialista os subconjuntos de variáveis identificadas pelo OAM. Ao escolher um subconjunto, o especialista passa a visualizar os dados numa série temporal como na Figura 2.

Dentre os parâmetros que o usuário pode determinar, destacam-se as janelas de tempo prévio ou posterior ao momento de detecção da falha. Estas configurações são importantes porque, de acordo com o tipo de anomalia analisada, diferentes intervalos são necessários para que padrões fiquem evidentes. Por exemplo, um problema relacionado a um evento extremo pode ser visível imediatamente. Por sua vez, a degradação de certos equipamentos pode durar anos.

A janela de tempo escolhida possui consequências diretas na quantidade de amostras a serem exibidas. Dentre as ferramentas disponíveis para reduzir o *overplotting*, destaca-se a possibilidade de sub-amostrar as variáveis com granularidade de segundos até anos. Além disso, pode-se optar pela utilização de funções como máximo, mínimo e média dos valores conforme a necessidade. Nos testes realizados, as funções de máximo e mínimo mostraram-se mais adequadas devido à natureza extrema das anomalias.

Para as imagens exibidas nesta seção, foi considerado o período de sete dias antes da identificação da falha (o que ocorreu em 08/08) e um dia após esta detecção. Quanto a amostragem dos dados, optou-se por 90 minutos, representando-se o valor máximo para cada variável no intervalo amostrado.

### 3.3. Visualização temporal

A representação de séries temporais são extremamente comuns, de fácil de interpretação e manipulação, sendo por isso incluídas na plataforma. Para adequá-la ao contexto de OAM, o usuário pode escolher o subespaço de variáveis a ser representado. Além disso, como complemento, também exibe-se a evolução de uma métrica de anormalidade calculada diariamente para o subespaço analisado, no caso, optou-se pelo *Local Outlier Factor* (LOF) [Breunig et al. 2000], muito aplicada a OAM [Vinh et al. 2016]. O valor obtido para LOF foi normalizada para o intervalo de 0 a -1, em que -1 representa o máximo de anormalidade para o subespaço. O gradiente de cores utilizado para a exibição da métrica de anormalidade segue o mesmo padrão adotado para a representações de amostra no

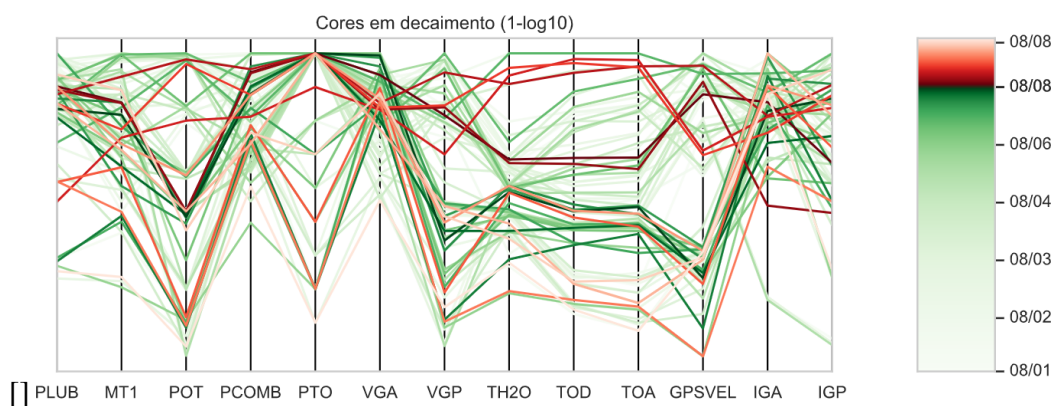


Figura 1. Visualização multidimensional PCP

gráfico de coordenadas paralelas, com valores amostrados antes da falha em verde e, após a falha, em vermelho.

Na Figura 2, as três variáveis exibidas foram selecionadas tomando-se por base resultados anteriores fornecidos pelo OAM: temperatura do óleo depois do arrefecimento (TOD), temperatura do óleo antes do arrefecimento (TOA) e temperatura da água do trocador de calor (TH2O). As lacuna existentes na representação do LOF ocorrem para dias em que a locomotiva não gerou dados. Ademais, o momento em que a falha foi detectada é indicado através de uma linha pontilhada no gráfico.

### 3.4. Visualização multidimensional

Além da representação de séries temporais, a plataforma gera também gráficos de coordenadas paralelas. Nelas, apesar do usuário poder escolher a sequência de variáveis a serem representadas nos eixos, o padrão inicial é a exibição das variáveis consideradas mais representativas por técnicas de OAM.

Uma das inovações deste trabalho é comunicação da evolução temporal em um gráfico de coordenadas paralelas. Neste sentido, amostras obtidas antes da identificação do *outlier* são representadas em verde e, após a detecção, em vermelho. Isto no entanto não é o bastante. Por isso, como complemento, adaptou-se a estratégia adotada por [Johansson et al. 2007, Zhonghua and Lingda 2016], em que amostras mais antigas são exibidas de forma mais apagada. Contudo, como para esta aplicação a representação se estende também para após o instante de detecção da anomalia, decidiu-se pelo aumento gradual da transparência das linhas após a falha também. Desta forma, salientam-se as amostras situadas temporalmente mais próximas ao momento crítico em que a anomalia foi encontrada.

Diferentes funções para o gradiente de saturação das cores foram testadas para o período anterior à falha:  $f(x) = x$ ,  $f(x) = \log(x)$  e  $f(x) = 1 - \log(x)$ . No entanto, por enfatizar o período mais próximo à anomalia e reduzir a confusão entre as linhas, a função  $f(x) = 1 - \log(x)$  foi a preferida. Para o período após a falha, foi adotado um decaimento linear, pois o tempo representado costuma ser mais reduzido. Esta combinação de escolhas é apresentada na Figura 1. Na plataforma desenvolvida, todas as alternativas

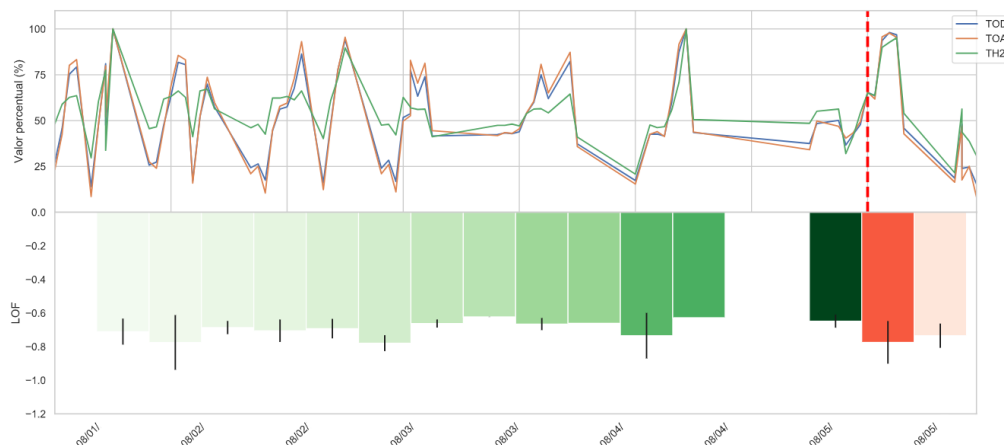
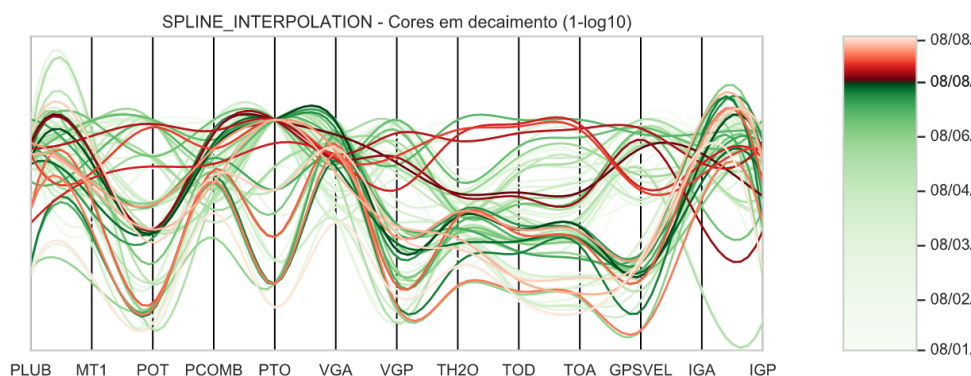


Figura 2. Modo de visualização temporal

apresentadas estão disponíveis para seleção do usuário.



**Figura 3. PCPs com curvas.**

Conforme demonstrado em [Heinrich and Weiskopf 2013], se os valores de duas amostras forem iguais para um mesmo atributo, o senso de continuidade da polilinha pode ser perdido. Dentre as alternativas existentes para combater esta ambiguidade, ressaltam-se a utilização de diferentes cores e a substituição de linhas retas por curvas. No modelo proposto, apesar de adotar-se um gradiente de cores, dependendo da proximidade temporal, é possível que os elementos gráficos se confundam. Por isso, o usuário pode alternar entre as duas formas de exibição: linhas curvas e retas. A representação com linhas retas foi mantida, pois a interpolação pode gerar distorções dos valores exibidos. Dentre as funções de interpolação testadas, preferiu-se a spline.

#### 4. Discussão e conclusão

As imagens apresentadas na seção anterior demonstram como a plataforma desenvolvida tem potencial para auxiliar na explicação das falhas. A Figura 2, por exemplo, demonstra como o subespaço levado em consideração apresenta valores altos para a métrica de anormalidade próximos à identificação da anormalidade. Além disso, a representação em coordenadas paralelas também indica com clareza que as variáveis relacionadas à temperatura do motor diesel tiveram valores mais altos próximos à detecção da falha, indo ao encontro dos resultados fornecidos pela técnica de OAM e à condição de falha.

Dentre as dificuldades que citamos no uso do gráfico de coordenadas paralelas, salientamos a necessidade de se escolher taxas de amostragem adequadas para se evitar o *overplotting*. Uma alternativa seria a substituição da frequência de amostragem contínua por uma análise que levasse em consideração a relevância das amostras, possivelmente baseada na variabilidade dos dados ou em uma métrica de anormalidade. Com isso, os períodos representados poderia ser aumentados sem perda de informações importantes e sem prejuízo à identificação de padrões visuais.

Quanto à utilização de gradiente de cores para a representação de tempo em coordenadas paralelas, consideramos que esta estratégia comunicou com sucesso a evolução do estado da máquina através do tempo. No entanto, testes com usuários reais ainda precisariam ser realizados para se determinar a curva de aprendizado ou mesmo a efetividade de tal modelo.



## Referências

- Barlow, N. and Stuart, L. J. (2004). Animator: A tool for the animation of parallel coordinates. In *Proceedings. Eighth International Conference on Information Visualisation, 2004. IV 2004.*, pages 725–730. IEEE.
- Blaas, J., Botha, C., and Post, F. (2008). Extensions of parallel coordinates for interactive exploration of large multi-timepoint data sets. *IEEE Transactions on Visualization and Computer Graphics*, 14(6):1436–1451.
- Breunig, M. M., Kriegel, H.-P., Ng, R. T., and Sander, J. (2000). Lof: identifying density-based local outliers. In *Proceedings of the 2000 ACM SIGMOD international conference on Management of data*, pages 93–104.
- Chandola, V., Banerjee, A., and Kumar, V. (2009). Anomaly detection: A survey. *ACM computing surveys (CSUR)*, 41(3):1–58.
- Chen, M., Trefethen, A., Banares-Alcantara, R., Jirotko, M., Coecke, B., Ertl, T., and Schmidt, A. (2011). From data analysis and visualization to causality discovery. *Computer*, (10):84–87.
- Cielen, D., Meysman, A., and Ali, M. (2016). *Introducing data science: big data, machine learning, and more, using Python tools*. Manning Publications Co.
- Glendenning, K., Wischgoll, T., Harris, J., Vickery, R., and Blaha, L. (2016). Parameter space visualization for large-scale datasets using parallel coordinate plots. *Electronic Imaging*, 2016(1):1–8.
- Hawkins, D. (1980). *Identification of Outliers*. Springer Netherlands, 1st edition.
- Heinrich, J. and Weiskopf, D. (2013). State of the art of parallel coordinates. In *Eurographics (STARs)*, pages 95–116.
- Inselberg, A. (1985). The plane with parallel coordinates. *The visual computer*, 1(2):69–91.
- Johansson, J., Ljung, P., and Cooper, M. D. (2007). Depth cues and density in temporal parallel coordinates. In *EuroVis*, volume 7, pages 35–42.
- Keim, D. A., Mansmann, F., and Thomas, J. (2010). Visual analytics: How much visualization and how much analytics? *SIGKDD Explor. Newsl.*, 11(2):5–8.
- Liu, F. T., Ting, K. M., and Zhou, Z.-H. (2008). Isolation forest. In *2008 Eighth IEEE International Conference on Data Mining*, pages 413–422. IEEE.
- Novotny, M. and Hauser, H. (2006). Outlier-preserving focus+context visualization in parallel coordinates. *IEEE Transactions on Visualization and Computer Graphics*, 12(5):893–900.
- Sansen, J., Richer, G., Jourde, T., Lalanne, F., Auber, D., and Bourqui, R. (2017a). Visual exploration of large multidimensional data using parallel coordinates on big data infrastructure. In *Informatics*, volume 4, page 21. Multidisciplinary Digital Publishing Institute.
- Sansen, J., Richer, G., Jourde, T., Lalanne, F., Auber, D., and Bourqui, R. (2017b). Visual exploration of large multidimensional data using parallel coordinates on big data

- infrastructure. In *Informatics*, volume 4, page 21. Multidisciplinary Digital Publishing Institute.
- Tadeja, S. K., Kipouros, T., and Kristensson, P. O. (2019). Exploring parallel coordinates plots in virtual reality. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*, pages 1–6.
- Thomas, J. and Cook, K. A. (2005). Illuminating the path: The r&d agenda for visual analytics national visualization and analytics center. *National Visualization and Analytics Center*.
- Tong, C., Zhang, J., Chowdhury, A., and Trost, S. G. (2019). An interactive visualization tool for sensor-based physical activity data analysis. In *Proceedings of the Australasian Computer Science Week Multiconference*, pages 1–4.
- Vinh, N. X., Chan, J., Romano, S., Bailey, J., Leckie, C., Ramamohanarao, K., and Pei, J. (2016). Discovering outlying aspects in large datasets. *Data Mining and Knowledge Discovery*, 30(6):1520–1555.
- Webga, K. and Lu, A. (2015). Discovery of rating fraud with real-time streaming visual analytics. In *2015 IEEE Symposium on Visualization for Cyber Security (VizSec)*, pages 1–8. IEEE.
- Zhonghua, Y. and Lingda, W. (2016). 3d-parallel coordinates: Visualization for time varying multidimensional data. In *2016 7th IEEE International Conference on Software Engineering and Service Science (ICSESS)*, pages 655–658. IEEE.