SBrT 2019 1570550637

Proposta de um método de análise do desempenho de redes móveis com base na representação dimensional de KPIs usando PCA e clusterização

Felipe Augusto Machado Corrêa, Alan Petrônio Pinheiro e Willian Douglas Caixeta Nunes

***Resumo*— A complexidade, abrangência e quantidade de dados produzida pelas redes móveis têm exigido novos métodos para uma análise mais automatizada e capaz de buscar por padrões de anomalias não-óbvios. Neste sentido, esta pesquisa sugere um método de detecção de anomalias baseado em redução de dimensionalidade, clusterização de indicadores de desempenho e segmentação em diferentes períodos do dia. O método foi avaliado usando seis indicadores de uma rede em operação e os resultados encontrados encorajam o uso do método como ferramenta para nortear a correção de falhas, otimizar recursos e indicação de expansão de capacidade da rede.**

Palavras-Chave—Detecção de anomalias, redução de dimensionalidade, clusterização, indicadores de desempenho, rede celular.

Abstract— **The complexity, coverage and quantity of data produced by mobile networks have required new methods for a more automated analysis and able to search for patterns of non-obvious anomalies. In this sense, this research suggests a method of detection of anomalies based on reduction of dimensionality, clustering of performance indicators and segmentation in different periods of the day. The method was evaluated using six indicators of a network in operation and the results found encourage the use of the method as a tool to guide the correction of failures, optimize resources and indication of expansion of network capacity.**

Keywords—Anomaly detection, dimensionality reduction, clustering, key performance indicators, cellular network.

I. INTRODUÇÃO

As redes móveis têm como uma de suas características a geração de uma grande quantidade de dados de gerenciamento de desempenho da rede. Isso se deve, em grande parte, a característica de mobilidade e a transmissão de dados via interface aérea para um número geralmente grande e crescente de usuários. Esses dados são gerados continuamente a partir dos elementos da rede, como estações rádio base (ERB), e quantizam, por exemplo, a quantidade de tentativas de requisição [1]. As medições são associadas a contadores que são usados no cálculo dos principais KPIs (*Key Performance Indicators*) e possuem um papel fundamental no planejamento, otimização, monitoramento, detecção e recuperação de falhas da rede.

Tradicionalmente, o monitoramento de desempenho da rede é baseado em limiares estabelecidos com base na experiência e conhecimento do operador para identificar alguma degradação ou comportamento anormal. Como a rede possui diferentes configurações e apresenta um comportamento dinâmico que varia ao longo do dia, o monitoramento se torna limitado. Além disso, existe um grande número de KPIs armazenados nos OSS's (*Operations Support Systems*). Analisar toda esta quantidade de indicadores em um ambiente de operação é impraticável. Neste cenário, uma alternativa eficiente é automatizar o processo de monitoramento dos KPIs, desenvolvendo métodos ou modelos de análise que identifiquem padrões nos dados que não estão em conformidade com o comportamento esperado. Tal atividade é conhecida na área técnica como detecção de anomalias [2].

A detecção de anomalias pode ser realizada de diversas formas, seja por meio de análises estatísticas, mineração de dados, aprendizado de máquina, dentre outras. No contexto de redes móveis, diferentes abordagens têm sido propostas. Em [3] é desenvolvida uma plataforma de detecção de anomalias e análise de causa raiz (RCA), baseada na mineração de regras de associação combinadas com abordagens estatísticas para identificar as anomalias. Em [4], por exemplo, são comparados cinco algoritmos de aprendizado não supervisionado (*K-means, Fuzzy C-means, Local Outlier Factor, Local Outlier Probabilities e Kohonen's Self Organizing Maps* (SOM)) com o intuito de detectar falhas na rede. Os resultados mostraram que o SOM apresentou desempenho superior ao *K-means* e *Fuzzy C-means.*

No mesmo contexto e baseado em KPIs, alguns autores [5] combinam várias funções estatísticas para criar medidas como *score* e probabilidade para determinar a importância e certeza das anomalias detectadas. Em [6] é proposto um método baseado no perfil de comportamento do KPI usando informações estatísticas, seguido por uma etapa de clusterização baseada em densidade para a detecção das anomalias. Em [7-8], por exemplo, os autores propõem um *framework* com um método *ensemble* adaptativo para detecção de anomalias que indica a severidade da degradação das células por meio de diferentes métodos preditivos em conjunto com informações de gerenciamento de configuração. Por outro lado, vários autores [9-11] têm usado outras abordagens e bases de dados como CDR (*Call Detail Record*), transações coletadas das interfaces do *Core Network* e indicadores de qualidade de experiência (QoE) para detecção de anomalias.

Um aspecto importante na detecção de anomalias em KPIs é que, ao longo do dia, estes valores podem apresentar variação especialmente em virtude da dinâmica que estas redes têm com seus usuários. Além disso, a anomalia pode estar relacionada ao perfil de comportamento entre os KPIs e a análise multivariada pode identificar anomalias não encontradas na análise univariada. Neste cenário, este trabalho propõe um método para detecção de anomalias com base em um conjunto de KPIs

Felipe Augusto Machado Corrêa, Uberlândia – MG, e-mail: felipeaugustomc@ufu.br; Alan Petrônio Pinheiro, Uberlândia – MG, e-mail: alanpetronio@ufu.br; Willian Douglas Caixeta Nunes, Uberlândia – MG, e-mail: willian_douglas_nunes@outlook.com. Universidade Federal de Uberlândia. Trabalho parcialmente financiado pelo projeto ANEEL/CEB PD-05160-1805/2018.

levando em consideração a variação temporal destes indicadores. Para isto, foi considerada nesta pesquisa uma segmentação feita por cada hora do dia. Os resultados do método de detecção podem ser usados como instrumento de correção de falhas, otimização de recursos, expansão de capacidade da rede, dentre outros. Vale destacar que o método indica apenas a existência de uma anomalia e não qual é o seu tipo ou causa. Esta identificação de causa deve ser feita pela análise de um profissional experiente da área.

O artigo está organizado da seguinte forma: na Seção II é descrito o método proposto para a detecção de anomalias. Na Seção III é apresentado um estudo de caso utilizando KPIs da rede 4G de uma operadora de telecomunicações. Algumas discussões também são tecidas. Por fim, na Seção IV, são apresentadas as conclusões deste trabalho.

II. METODOLOGIA PROPOSTA PARA DETECÇÃO DE ANOMALIAS

A detecção de anomalias tem como objetivo identificar padrões no conjunto de dados que não seguem o comportamento esperado. A identificação destes padrões envolve algumas etapas desde a seleção dos atributos, que no caso são os KPIs, até a detecção em si. O método proposto neste trabalho é baseado na (i) redução da dimensionalidade da base de dados e (ii) clusterização baseada na densidade dos dados. Tais abordagens não exigem dados de treinamento rotulados (*i.e.*, não supervisionado) e podem avaliar os objetos com base na vizinhança dos dados. Na sequência, os elementos da rede (células, no caso) com maior quantidade de amostras identificadas como anomalias são analisadas. A Figura 1 apresenta a estrutura do método de detecção de anomalias desenvolvido aqui proposto.

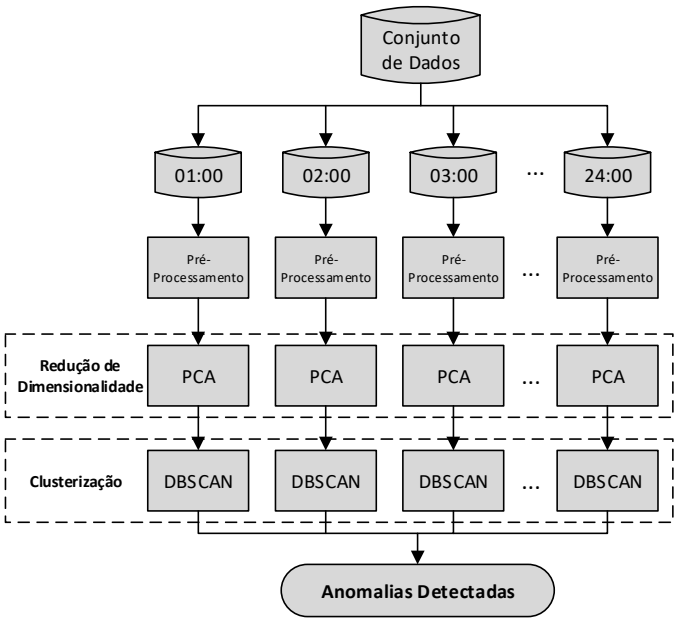

Fig. 1. Estrutura do método de detecção de anomalias.

Os KPIs usados nesta pesquisa foram extraídos diretamente da OSS, que armazena as informações de gerenciamento de desempenho, configuração e alarmes dos elementos da rede móvel. Neste caso, os KPIs são formados por um conjunto de medições realizadas periodicamente pelas células das eNodeBs (estações rádio base das redes 4G). Os dados foram disponibilizados pela operadora Algar Telecom (Minas Gerais, Brasil).

A. Pré-processamento e redução de dimensionalidade

Antes de se iniciar qualquer análise é importante que seja realizado um pré-processamento dos dados com o intuito de se obter resultados significativos e úteis. Nesta aplicação isto é importante porque as bases de dados podem apresentar valores inconsistentes, nulos e *outliers*. No caso dos valores inconsistentes e nulos, as instâncias dos dados que apresentarem pelo menos um valor destes casos serão removidas. No caso dos *outliers*, os valores serão mantidos por fazerem parte do escopo de estudo deste trabalho.

Um ponto relevante a ser verificado é a variação dos dados. Os KPIs que possuem alta variância dominam as técnicas baseadas em distância e variância, como as que serão utilizadas no método proposto. Assim, é necessário fazer um dimensionamento adequado para tornar os KPIs iguais em sua importância. Uma forma de se fazer isso é normalizando os dados com base na média aritmética como referência (*i.e.*, valor zero do eixo) e o desvio padrão como medida de variação do eixo/dimensão em questão. Este tipo de normalização mantém a forma da curva dos KPIs sem perder informações importantes. A Figura 2a mostra a variação temporal de dois KPIs. Nota-se que o KPI 1 apresenta uma variação de amplitude maior que o KPI 2. Após a normalização (vista na Figura 2b), a variação dos KPIs se aproxima e é possível observar com mais detalhes o comportamento do KPI 2 sem distorções de magnitude inerentes a própria grandeza do KPI em questão. No caso desta figura, o valor 0 (zero) do eixo vertical indica os valores que estão na média aritmética. Já a cada B unidades que se move neste eixo, indica uma variação de B vezes o desvio padrão deste KPI em questão.

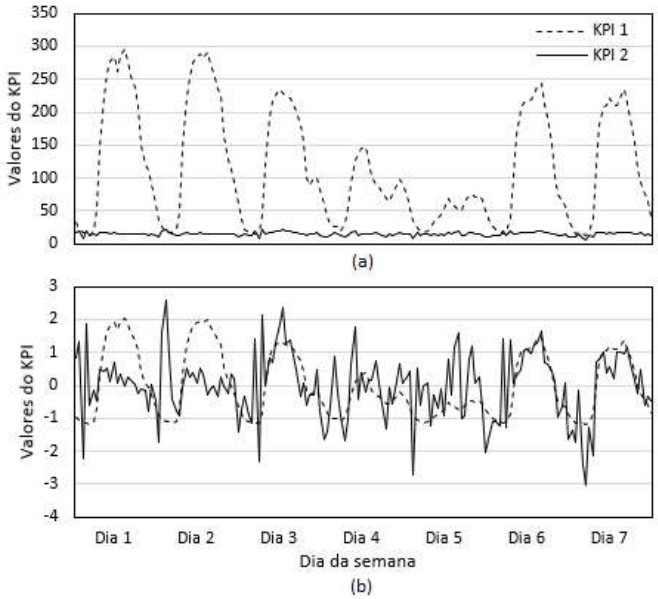

Fig. 2. Representação gráfica de dois KPIs de variações e magnitudes diferentes. (a) valores originais e (b) normalizados.

Outra questão é a quantidade de KPIs tratados que pode gerar uma grande dimensão de análise. Para evitar isto é importante ter uma redução de dimensionalidade a fim de potencialmente evidenciar de forma mais clara informações "escondidas" nos dados. Para tanto, o método aqui proposto usa a técnica de PCA (*Principal Component Analysis*) [10] para transformar um conjunto de dados multivariados de possíveis

variáveis correlacionadas em um conjunto ortogonal de componentes.

As componentes da PCA quantificam a importância de cada variável para descrever a variabilidade de um conjunto de dados. A partir da matriz de covariância das variáveis (ou KPIs) são obtidos os autovalores e autovetores das componentes, que representam a magnitude e direção da variância dos dados, respectivamente. Dessa forma, cada componente carrega consigo uma parte da variância de cada KPI e as componentes com maior representação da variância são usadas. Geralmente, as duas primeiras componentes representam pelo menos metade da variância total das amostras. A Figura 3 mostra um gráfico de dispersão das duas primeiras componentes PCA geradas de um conjunto de dados com seis KPIs.

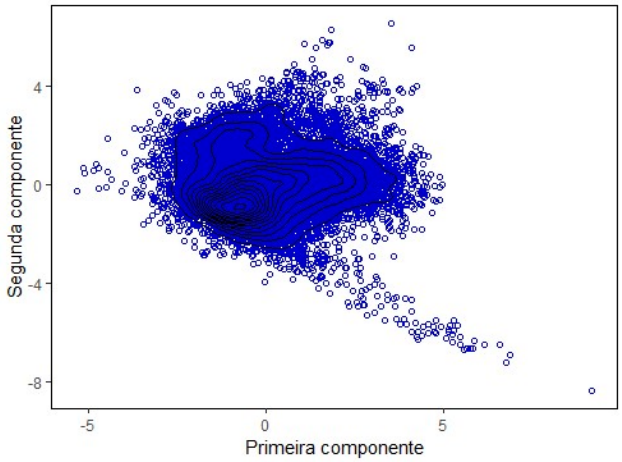

Fig. 3. Representação gráfica da dispersão das duas primeiras componentes geradas pela técnica PCA. As curvas em preto representam a distribuição da densidade dos pontos apresentadas na próxima seção.

B. Clusterização baseada em densidade

A clusterização é uma técnica geralmente não supervisionada empregada para agrupar dados com base em uma medida de dissimilaridade. A distância entre os dados, por exemplo, pode ser uma das métricas adotadas para se medir esta dissimilaridade e formar grupos comumente chamados de *clusters*. Os dados que fazem parte do mesmo *cluster* tendem a possuir alguma característica ou propriedade em comum e no cenário de detecção de anomalias, podem representar o padrão normal ou anormal dos dados.

Na clusterização baseada em densidade, um *cluster* é formado em qualquer direção conduzido pela densidade dos pontos (leia-se dados). Os algoritmos baseados em densidade são capazes de identificar *clusters* de formas arbitrárias e procuram por regiões de alta densidade separados por regiões de baixa densidade. Um dos primeiros algoritmos implementados neste contexto foi o DBSCAN (*Density-Based Spatial Clustering of Applications with Noise*) [13].

A ideia central do DBSCAN está associada à noção de *density-reachability* (acessibilidade em densidade) e depende dos parâmetros *Eps* (raio da vizinhança de um ponto *p*) e *MinPts* (número mínimo de pontos da vizinhança). Tais parâmetros são matematicamente associados como:

$$V_{Eps}(p) = \{q \in D \mid dist(p,q) \leq Eps\} \quad (1)$$

$$p \in V_{Eps}(q) \quad (2)$$

$$|V_{Eps}(q)| \geq MinPts \quad (3)$$

A vizinhança de um ponto *p* é definida conforme a Equação (1), em que *D* é o conjunto de pontos com distância (Euclidiana, no caso) menor ou igual ao parâmetro *Eps*. Um ponto *p* é considerado *directly density-rechable* (DDR - diretamente alcançável em densidade) de um ponto *q* com base nas Equações (2) e (3). Se o número de pontos DDR de um ponto *p* é maior que *MinPts*, este ponto é um *core point* (ponto de centro). Caso o número de pontos seja menor que *MinPts*, este é um *border point* (ponto de borda). Um ponto *p* é *density-rechable* (DR - alcançável em densidade) de um ponto *q* se existe uma sequência de pontos $p_1, p_2, ..., p_n, p_1 = q, p_n = p$ em que p_i+1 é DR de p_i. Dois pontos *p* e *q* são *density-connected* (DC - conectados em densidade) se são DR de um ponto *o*.

Um *cluster* é formado se o número de pontos na vizinhança *Eps* de *p* for maior que *MinPts.* Neste cenário, os pontos DC de *p* são adicionados ao *cluster*. Caso contrário, um ponto não processado na base de dados é selecionado. Um ponto será considerado como ruído (*noise point*) se não for DC de outro ponto e, portanto, não pertencer a algum *cluster*. Ao fim do processo, os pontos categorizados como ruídos tendem a apresentar diferentes propriedades dos pontos pertencentes a um *cluster* e podem ser considerados como anomalias.

O algoritmo DBSCAN depende diretamente da escolha dos parâmetros *Eps* e *MinPts*. Geralmente, um estimador razoável de densidade deve contar com as distâncias dos *k*-vizinhos mais próximos, em que *k* é igual *MinPts*. O valor de *Eps* é o primeiro ponto do primeiro "vale" do gráfico das distâncias ordenadas e separa os pontos que estarão em algum *cluster* dos pontos categorizados como ruídos. Na Figura 3 as curvas em preto representam a distribuição de densidade dos pontos na região. Nota-se que a densidade é maior ao centro e diminui próximo às bordas, onde tendem a se localizar os pontos categorizados como "ruídos".

No método proposto, os parâmetros *Eps* e *MinPts* foram definidos com o objetivo de se obter um *cluster* único, que represente o comportamento normal da rede e ao redor os pontos ruídos que representam potenciais anomalias. Para tanto, o valor de *MinPts* deve ser alto e indexado ao desvio padrão (σ) dos dados, já que o comportamento dos KPIs varia durante o dia. Dessa forma, considera-se *MinPts* = 10.σ e o valor de *Eps* é o ponto de "vale" das distâncias dos *k*-vizinhos mais próximos.

Ao fim das etapas do método de detecção, tem-se os pontos categorizados como anomalia das células da rede. Com base nessas informações, podem ser realizadas várias análises do que está motivando o comportamento anormal das células e os resultados podem iniciar ações de otimização e necessidade de expansão de rede, por exemplo.

III. ESTUDO DE CASO

O estudo de caso envolve um conjunto de dados com KPIs relacionados à qualidade do sinal e utilização da rede, coletados da rede celular 4G de uma operadora com 1.054 células no período de 01/03/2019 à 14/03/2019. O conjunto de dados é composto por instâncias formadas pelo nome da célula, dia/hora e os valores dos KPIs no período, num total de 336 instâncias por célula (14 dias x 24 horas). Estes dados são usados na avaliação do método de detecção de anomalias proposto, que identificará as células que possuem comportamento anormal. O método como um todo foi implementado na linguagem R

usando o ambiente RStudio e usando o algoritmo DBSCAN da biblioteca *dbscan* [14]. Foram selecionados os seguintes KPIs para indicar o comportamento dinâmico da rede:

- Relação sinal-ruído-interferência (SINR) média no canal PUCCH (*Physical Uplink Control Channel*);
- Relação sinal-ruído-interferência (SINR) média no canal PUSCH (*Physical Uplink Shared Channel*);
- Indicador de intensidade do sinal recebido (RSSI) média no canal PUCCH;
- Indicador de intensidade do sinal recebido (RSSI) média no canal PUSCH;
- Número médio de usuários ativos;
- *Throughput* médio no *downlink*.

As instâncias de dados que apresentavam pelo menos um KPI com valor nulo (aproximadamente 2,5% do total) foram removidas durante o pré-processamento dos dados. Já na etapa de redução de dimensionalidade, foram consideradas as duas primeiras componentes pela simplificação da representação dimensional e custo computacional. Estas duas componentes são capazes de representar, em média, 65% da variância dos KPIs.

A Figura 4 apresenta os resultados da detecção das anomalias em dois períodos diferentes do dia.

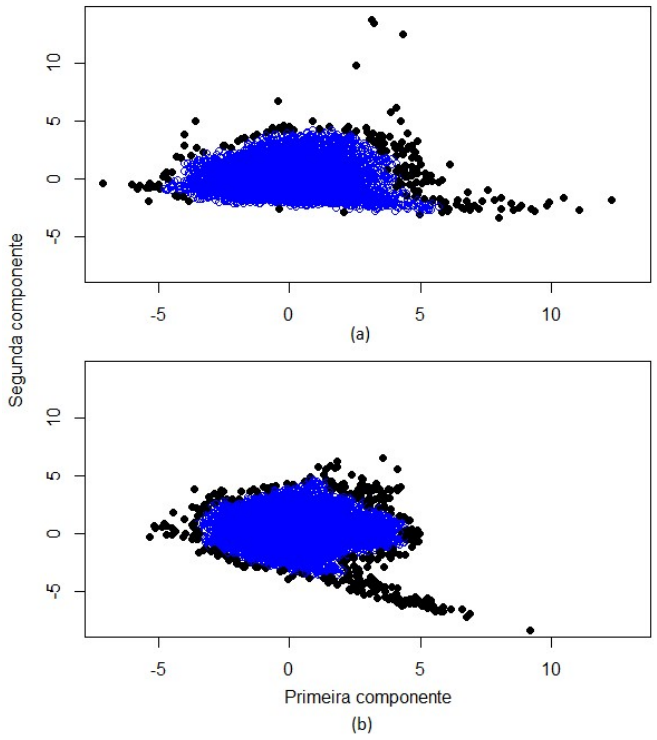

Fig. 4. Representação gráfica das anomalias detectadas (em preto) pelo método em duas horas do dia, 04:00 (a) e 20:00 (b).

A Figura 4a se refere a um período de baixo tráfego (04:00) e a Figura 4b a um período de alto tráfego (20:00). Os pontos em azul (círculo não preenchido) formam um *cluster* que representa o comportamento normal dos KPIs e os pontos em preto (círculo preenchido) representam as anomalias detectadas. Como o método proposto segmenta o conjunto de dados em horas, devido ao comportamento dinâmico dos KPIs, é esperado que a dispersão dos dados seja diferente nos dois períodos.

O método de detecção apontou pelo menos uma potencial anomalia em 389 células, que representa cerca de 37% do número total de células. A Figura 5 mostra o número de anomalias detectadas por célula, ordenado de forma decrescente. Observa-se que grande parte das anomalias foram detectadas em uma pequena parcela de células, em que 80% do total de anomalias são referentes à 87 células ou cerca de 8,2% do grupo de estudo. Neste caso, evidencia-se que são casos críticos e o comportamento tido como anormal é recorrente.

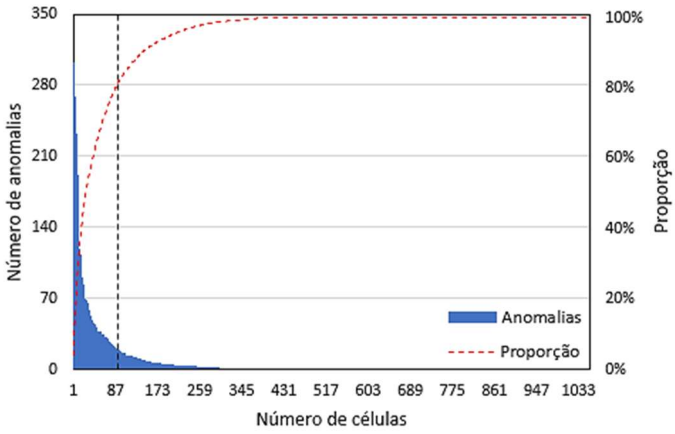

Fig. 5. Representação gráfica da quantidade de anomalias detectadas por célula (barras em azul) e a proporção acumulada (linha tracejada).

As anomalias detectadas pelo método proposto possuem relação direta com os KPIs selecionados. Se o comportamento de um ou vários KPIs em um horário não seguir o comportamento da rede, é muito provável que seja categorizado como anomalia. Além disso, o motivo real da célula ter um comportamento anormal pode também estar relacionado com os KPIs em si. Com base nos resultados apresentados na Figura 5, foram analisadas individualmente as vinte células que apresentaram a maior quantidade de anomalias, que juntas representam cerca de 44% do total de anomalias detectadas. Esta análise das células selecionadas (tidas como anômalas) foi feita pela análise direta de um profissional habilitado para identificar qual seria a possível anormalidade. Os resultados são apresentados na Tabela I.

A avaliação das vinte células que apresentaram a maior quantidade de anomalias apontou as seguintes causas:

- *Overshooting*: a cobertura das células 2, 5, 7, 8, 11, 18 e 19 estavam além do ideal. Neste caso, existe uma degradação na qualidade do sinal e sobrecarga da rede;
- Alta utilização ou sobrecarga: aconteceram nas células 6, 9, 12, 14, 16, 17 e 20 gerando degradação nos KPIs de SINR e alto número médio de usuários ativos;
- Congestionamento no *backhaul*: o enlace de transmissão entre a ERB e o *Core Network* apresentava congestionamento, gerando degradação do KPI de *throughput* médio no *downlink* das células 1 e 3. Ambas células pertencem à mesma ERB;
- Subutilização: as células 4 e 15 apresentavam baixa utilização e desviou do restante da rede pelo desempenho superior que obtiveram;
- Falha no sistema irradiante: a célula 10 apresentava degradação nos KPIs de qualidade do sinal e *throughput* médio no *downlink;*
- Interferência externa: a célula 13 apresentava degradação acentuada nos KPIs de qualidade do sinal e *throughput* médio no *downlink* em períodos característicos.

TABELA I. LISTA DAS VINTE CÉLULAS QUE APRESENTARAM A MAIOR QUANTIDADE DE INSTÂNCIAS CATEGORIZADAS COMO ANOMALIA.

Célula	Anomalias	Avaliação
1	303	Congestionamento no *backhaul*
2	269	*Overshooting*
3	258	Congestionamento no *backhaul*
4	251	Subutilização
5	232	*Overshooting*
6	228	Alta utilização
7	218	*Overshooting*
8	191	*Overshooting*
9	154	Alta utilização
10	127	Falha no sistema irradiante
11	119	*Overshooting*
12	119	Alta utilização
13	118	Interferência externa
14	113	Alta utilização
15	112	Subutilização
16	106	Alta utilização
17	91	Alta utilização
18	90	*Overshooting*
19	90	*Overshooting*
20	83	Alta utilização

Com base nas informações disponibilizadas pelo método foi possível identificar um comportamento anormal das células usando, simultaneamente, vários KPIs. Naturalmente que o diagnóstico (confirmação) final deve ser feito por pessoal especializado. Contudo, o método proposto ajuda a melhorar a seleção, diagnóstico e assertividade de análise das células que entende ter comportamentos mais "específicos" (ou anômalos). As informações produzidas tem potencial de proporcionarem vários ganhos em um ambiente operacional, já que todo o processo é automatizado e pode ser executado periodicamente. Até mesmo um especialista altamente qualificado, pode não identificar estes comportamentos em meio a tantos dados e parâmetros. Por isto a necessidade de métodos como o aqui proposto.

Em específico, os resultados aqui encontrados no estudo de caso apresentado, foram usados em ações para correção das falhas, otimização de recursos e expansão de capacidade da rede da operadora que disponibilizou os dados para a pesquisa. Adicionalmente, não foi possível comparar o desempenho do método proposto com outros, uma vez que não se encontrou disponível outro método que pudesse tratar na mesma amplitude (*i.e.*, quantidade de KPIs) os dados usados nesta pesquisa.

IV. CONCLUSÕES

Este trabalho propôs um método de detecção de anomalias baseado em agrupamento de informações que segmenta o conjunto de dados em períodos temporais e tem como objetivo fornecer subsídios de análise para correção de eventuais falhas, otimização de recursos, planejamento e expansão de capacidade da rede móvel. O estudo de caso abordado usou um conjunto de KPIs de uma rede 4G em operação, relacionados à qualidade do sinal e utilização da rede e foi aplicado a 1054 células para identificar quais tinham comportamentos potencialmente anômalos. As vinte células que apresentaram o maior número de anomalias detectadas foram analisadas e os resultados encontrados foram usados com sucesso na correção de falhas, otimização de recursos e expansão de capacidade da rede pela operadora nestas mesmas vinte células com base no método/ferramenta aqui proposto.

Apesar do estudo de caso considerar seis KPIs, o método proposto pode ser aplicado em qualquer conjunto de KPIs na forma de séries temporais, independentemente da quantidade, desde que apresentem uma variância não nula. Os resultados devem mostrar o que pode ser considerado como comportamento normal e anormal com base na proximidade matemática dos dados. Entre as vantagens do método, pode-se citar sua capacidade de avaliar simultaneamente várias KPIs e poder ser executado periodicamente para validação das tratativas das falhas encontradas e detecção de novos comportamentos destoantes da média populacional avaliada.

AGRADECIMENTOS

Os autores agradecem à operadora Algar Telecom por disponibilizar a base de dados para estudo e o suporte na avaliação dos resultados.

REFERÊNCIAS

[1] 3GPP TS 32.425 V15.1.0*, Performance measurements Evolved Universal Terrestrial Radio Access Network (E-UTRAN)*, Release 15, 2018.

[2] V. Chandola, A. Banerjee e V. Kumar, *Anomaly detection: A survey*. ACM Computing Surveys, Vol. 41, No. 3, Artigo 15, 2009.

[3] K. Yang, R. Liu, Y. Sun, J. Yang, X. Chen, *Deep Network Analyzer (DNA): A Big Data Analytics Platform for Cellular Networks*, IEEE Internet of Things Journal, Vol. 4, No. 6, pp. 2019-2027, 2017.

[4] U. S. Hashmi, A. Darbandi, A. Imran, *Enabling proactive selfhealing by data mining network failure logs*, International Conference on Computing, Networking and Communications (ICNC): Cloud Computing and Big Data, pp. 511–517, 2017.

[5] M. Wang e S. Handurukande, *Anomaly Detection for Mobile Network Management*. International Journal of Next-Generation Computing, Vol. 9, No. 2, pp. 80-97, 2018.

[6] L. Bodrog, M. Kajo, S. Kocsis, B. Schultz, *A robust algorithm for anomaly detection in mobile networks*, 2016 IEEE 27th Annual International Symposium on Personal Indoor and Mobile Radio Communications (PIMRC), pp. 1-6, 2016.

[7] G. Ciocarlie, U. Lindqvist, S. Novaczki, and H. Sanneck, *Detecting anomalies in cellular networks using an ensemble method*, 2013 9th international conference on Network and service management (CNSM), pp. 171–174, 2013.

[8] G. Ciocarlie, U. Lindqvist, K. Nitz, S. Novaczki, e H. Sanneck, *On the feasibility of deploying cell anomaly detection in operational cellular networks*, Network Operations and Management Symposium (NOMS), pp. 1–6, 2014.

[9] M. S. Hadi, A. Q. Lawey, T. E.H. El-Gorashi, J. M.H. Elmirghani, *Big data analytics for wireless and wired network design: A survey*, Computer Networks, Vol. 132, pp. 180-199, 2018.

[10] A. Asghar, H. Farooq, A. Imran, *Self-healing in emerging cellular networks: Review, challenges, and research directions*, IEEE Communications Surveys and Tutorials, Vol. 20, No. 3, pp. 1682-1709, 2018.

[11] P. V. Klaine, M. A. Imran, O. Onireti, R. D. Souza, *A Survey of Machine Learning Techniques Applied to Self-Organizing Cellular Networks*, IEEE Communications Surveys and Tutorials, Vol. 19, No. 4, pp. 2392-2431, 2017.

[12] J. Shlens, *A Tutorial on Principal Component Analysis*, Cornell University, pp. 1-12, 2014.

[13] M. Ester, H.P. Kriegel, J. Sander, X. Xu, *A density-based algorithm for discovering clusters in large spatial databases with noise*, Proceedings of the KDD, vol. 96, pp. 226–231, 1996.

[14] M. Hahsler, M. Piekenbrock, S. Arya e D. Mount, *dbscan: Density Based Clustering of Applications with Noise (DBSCAN) and Related Algorithms*, disponível em: <https://cran.r-project.org/web/packages/dbscan/index.html >, acesso em: 01/04/2019