

A descoberta de fatores associados à condição de risco dos transformadores elétricos usando Redes de Regras de Associação Filtrada

Alexandre Rosa Carvalho¹, Dario Brito Calçada¹

¹Universidade Estadual do Piauí – Parnaíba – PI – Brazil

alexandre.rosa.carvalho@gmail.com, dariobcalcada@gmail.com

Abstract. *Electrical transformers are essential equipment of an electrical system as they facilitate the transmission of electrical energy. Interruptions in the operation of this equipment may cause serious damage to the most diverse spheres. Computational Intelligence and Data Mining techniques have been used in the electrical sector to improve process prediction processes for at-risk transformers. Building Association Rules is one of the techniques used to mine data from a set, which assists in pattern detection and makes it possible to build a network for viewing them. The Filtered Association Rules Network employs filtering criteria to find the best extracted rules for easy knowledge extraction. This research aims to study the factors associated with the risk condition in electrical transformers with the use of Filtered Association Rules Networks. A comparison was made to results generated by an Association Rules Network to verify filtering efficiency. The filtering approach through objective measures proved to be able to eliminate the Association Rules that do not directly interfere with the item selected for study, facilitating the identification of items effectively linked to the situation of interest of this work.*

Resumo. *Os transformadores elétricos são equipamentos essenciais de um sistema elétrico, pois viabilizam a transmissão de energia elétrica. As interrupções do funcionamento deste equipamento podem implicar em grandes prejuízos nas mais diversas esferas. A Inteligência Computacional e técnicas de Mineração de Dados têm sido usadas no setor elétrico para melhorias em processos de previsão de situação de transformadores em risco. A construção de Regras de Associação é uma das técnicas utilizada para minerar dados de um conjunto, que auxilia na detecção de padrões e torna possível a construção de uma rede para visualização dos mesmos. A Rede de Regras de Associação Filtrada emprega critérios de filtragem para localizar as melhores regras extraídas a fim de que a extração de conhecimento seja facilitada. Esta pesquisa tem por objetivo o estudo de fatores associados à condição de risco em transformadores elétricos com o uso de Redes de Regras de Associação Filtrada. Foi feita uma comparação com os resultados gerados por uma Rede de Regras de Associação para verificar a eficiência da filtragem. A abordagem com filtragem através de medidas objetivas demonstrou-se capaz de eliminar as Regras de Associação que não interferem diretamente no item selecionado para estudo, facilitando a identificação de itens efetivamente ligados à situação de interesse deste trabalho.*

1. Introdução

É bastante comum em redes elétricas o uso de tensões e correntes alternadas devido a facilidade de geração destas condições. Os valores de tensão e corrente elétrica podem ter seus valores alterados para viabilizar a transmissão de energia elétrica, minimizando perdas. Os transformadores elétricos são equipamentos que realizam alterações nos módulos de tensão e corrente, sendo assim, um dos principais componentes em qualquer sistema elétrico [Xu 2017].

Por serem dispositivos que trabalham com tensões e correntes elevadas, os transformadores elétricos acabam dissipando bastante energia térmica, proporcionando problemas em seu funcionamento. Para evitar distorções térmicas, são utilizadas soluções formadas por óleo mineral para prover a refrigeração, além do isolamento elétrico e magnético. Durante o funcionamento, o óleo sofre degradação, sendo marcada por três fases: presença de gases dissolvidos, formação de peróxidos instáveis e de compostos oxo-celulósicos [Campos 2016].

Neste contexto, métodos que examinam as concentrações dos gases dissolvidos foram desenvolvidos e ganharam popularidade entre os profissionais da área de manutenção de transformadores elétricos [Brito et al. 2016]. A existência de casos em que a Análise de Gases Dissolvidos (DGA) apresentam diagnóstico imprecisos e os avanços na área da Mineração de Dados (DM) possibilitaram a aplicação de novas técnicas para aumentar a confiabilidade da avaliação dos parâmetros relacionados ao óleo utilizado no transformador e propiciar uma melhor tomada de decisão [Silva 2017].

Uma das técnicas de DM usada para encontrar padrões em conjuntos de dados (*datasets*) que permite a extração de informações e conhecimento pertinentes a um problema são as Regras de Associação (AR) [Kaur and Kang 2016, Calçada et al. 2019]. O ponto de partida do processo de DM vem das observações (eventos) que acionam o pesquisador para acelerar os estudos conceituais e chegar a uma estrutura na qual o processo subjacente (que está gerando os eventos) pode ser elucidado.

A Mineração de Regras de Associação (ARM) [Agrawal et al. 1994, Agrawal and Srikant. 1994, Agrawal and Shafer 1996] é usada a fim de encontrar padrões na forma de regras $A \Rightarrow B$, na qual A e B podem ser atributos, itens ou mais geralmente "objetos de dados". Considerando-se que fosse conhecido de antemão que A e B estão correlacionados, então a descoberta da regra $A \Rightarrow B$ apenas confirma o conhecimento prévio e não apresenta informação nova. Por outro lado, caso nunca tenha sido identificada a correlação entre A e B , a descoberta da regra $A \Rightarrow B$ sugere que A e B sejam pares candidatos a serem validados com o auxílio de medidas objetivas extraídas da regra.

O algoritmo mais difundido para a ARM é o *Apriori*. Este utiliza uma busca em profundidade para obter os padrões candidatos e eliminar itens não frequentes. A descoberta dos padrões é dividida em duas partes: obtenção dos conjuntos de itens frequentes e geração das AR com base nos conjuntos frequentes. O número de regras encontradas varia de acordo com a quantidade de registros e o número de itens contidos no *dataset*. Como os *datasets* geralmente estão aumentando, tanto em quantidade quanto em dimensionalidade, o uso do *Apriori* fica comprometido, já que este precisa percorrer o *dataset* várias vezes para enumerar todas as combinações possíveis de A e B e depois verificar sua

correlação, o que não é computacionalmente viável [Calçada 2019]. O algoritmo *Apriori-TID* elimina a necessidade de percorrer o *dataset* várias vezes, pois utiliza uma estrutura para auxiliar no processamento das AR, aumentando o desempenho em relação ao *Apriori* [Calçada 2019].

Para facilitar a extração do conhecimento, muitos processos de mineração utilizam técnicas de Redes para visualização dos dados [Kim et al. 2011]. O uso de Redes facilita o processo de identificação do conhecimento e o uso de sistemas de suporte à decisão, motivo pelo qual estas técnicas podem ser aplicadas em pesquisas sobre transformadores elétricos. Assim, no presente trabalho é empregado um método de ARM que utiliza medidas objetivas aliadas a uma estrutura de Rede para otimização da formação de hipóteses.

A extração de padrões da pesquisa tem por base a Rede de Regras de Associação Filtradas (*Filtered-ARN*) proposta por [Calçada et al. 2018] para determinar os parâmetros diretamente relacionados à situação de risco de transformadores elétricos. Foi feita a comparação dos resultados obtidos com aqueles produzidos pela técnica de Rede de Regras de Associação (ARN) proposta por [Pandey et al. 2009] para validação do método. Essa comparação foi realizada para verificar os resultados e determinar qual dos métodos poderia descrever, de maneira mais apropriada, o *dataset* com base no item objetivo do estudo, facilitando a identificação de itens efetivamente ligados a este e eliminando os que não possuem ligação com a situação escolhida.

Este artigo apresenta a seguinte estrutura: pesquisas relacionadas na Seção 2. Na Seção 3, é descrito o método das *Filtered-ARNs* utilizado para análise dos parâmetros relacionados a transformadores elétricos. Na Seção 4, são apresentados os métodos usados para gerar o *dataset*, bem como o processo de ARM e a técnica empregada para construção da ARN e *Filtered-ARN*. Os resultados são apresentados e discutidos na Seção 5 e as conclusões estão na Seção 6.

2. Trabalhos Relacionados

A busca de soluções que otimizem o processo de tomada de decisão vem crescendo exponencialmente com o uso de técnicas de inteligência artificial, em especial na área de manutenção de elementos de sistemas elétricos [Xu 2017]. Sendo assim, a descoberta dos fatores que contribuem para uma melhor análise dos dados é muito importante.

O custo de um transformador é diretamente proporcional a potência elétrica que este é capaz de atender. Os prejuízos decorrentes de uma falha no funcionamento podem chegar facilmente na casa dos milhões de dólares [Yaacob et al. 2015, Xu 2017]. O método de DGA foi pensado inicialmente para estudo de um transformador e logo tornou-se popular. Em condições normais, o óleo mineral se degrada lentamente. Se o dispositivo encontra situações adversas do ponto de vista térmico ou elétrico, então ocorrem instabilidades na estrutura do óleo e dos isolantes sólidos utilizados nos transformadores. Nestes casos, o óleo dissolve uma série de gases que podem provocar problemas, sendo classificados como combustíveis e não combustíveis (Tabela 1) [Xu 2017].

Métodos tradicionais de avaliação de transformadores, como o gás-chave, o triângulo de Duval e o do *Institute of Electrical and Electronics Engineers* (IEEE), tem por base a concentração de alguns destes gases para determinação da condição do

Tabela 1. Gases que podem ser encontrados dissolvidos no óleo isolante e as representações destes adotadas neste trabalho.

Combustíveis	Não Combustíveis
Hidrogênio [H ₂]	Oxigênio [O ₂]
Metano [CH ₄]	Nitrogênio [N ₂]
Monóxido de Carbono [CO]	Dióxido de Carbono [CO ₂]
Etileno [C ₂ H ₄]	Vapor [W]
Etano [C ₂ H ₆]	
Acetileno [C ₂ H ₂]	

transformador, enquanto as técnicas de Doernenburg, Rogers e do *International Electrotechnical Commission* (IEC) utilizam razões entre os mesmos para o diagnóstico [Brito et al. 2016, Mirowski and LeCun 2012]. Uma outra análise considerada para avaliar o surgimento de problemas em transformadores é a das propriedades físico-químicas (PHCH) do óleo isolante [Veloso 2017]. De acordo com [Carrapato et al. 2015], combinar esta avaliação com a dos gases dissolvidos permite planejar melhor a manutenção do equipamento. Para [Nogueira and Santos 2016], as características avaliadas por esta análise são:

1. Cor [C]: parâmetro usado para verificar a degradação do óleo com base na presença de partículas sólidas e de água. A classificação é feita com base em uma escala que vai de 0 a 8.
2. Densidade [G]: item que pode indicar contaminação por outros líquidos. Também influencia na condução de calor. Em boas condições, apresenta valores entre 0,8 e 0,9.
3. Rigidez dielétrica [DE]: verifica a capacidade de prover isolamento elétrico para a parte ativa do equipamento.
4. Tensão interfacial [IT]: denota o surgimento de partículas que alteram sua viscosidade, o que pode impactar na dissipação de calor.
5. Teor de água [W]: este pode afetar tanto o óleo isolante quanto as estruturas celulósicas da parte ativa.
6. Número de neutralização [NN]: indica o nível de acidez do óleo. Este fator pode afetar principalmente a celulose da parte ativa.
7. Fator de potência [PF]: também considerado para medir a contaminação do óleo por água e contaminantes sólidos.

Todos os métodos citados são de fácil implementação e não necessitam de um histórico de dados para um diagnóstico da condição do equipamento. Os aspectos PHCH analisados são baseados na experiência do monitoramento, sendo amplamente reconhecidos e aplicados [Silva 2017]. Embora a DGA e a análise das propriedades PHCH sejam bem disseminados, alguns resultados são inconclusivos, gerando diagnósticos equivocados e sem credibilidade. Desta forma, pode-se buscar o aumento da confiabilidade das análises com o uso de técnicas de inteligência artificial.

[Carrapato et al. 2015] e [Silva 2017] aplicaram o método da lógica *Fuzzy* para a DGA. Este é amplamente utilizado e expressa os resultados com base em uma entrada usando uma linguagem de fácil compreensão para humanos. Embora apresente vantagens em relação aos tradicionais, ainda existem alguns problemas como a falta de dinamismo e

a mutabilidade da função de associação base para a modelagem da rede. Desta forma, a lógica *Fuzzy* não permite a descrição de todos os casos de associação possíveis entre os gases dissolvidos e as propriedades PHCH do óleo isolante [Xu 2017].

Nas pesquisas de [Yaacob et al. 2015] e [Hussein et al. 2015], observa-se o uso de um sistema especialista para avaliação dos gases dissolvidos. O sistema realiza a simulação da tomada de decisão por um ser humano sustentada pelo uso de uma base de conhecimento. A maior desvantagem deste método é a incapacidade de aprendizado, portanto não sendo capaz de prever novas situações que podem ocorrer nos dados coletados. Sua base de conhecimento também pode gerar inconsistências, uma vez que dados incompletos ou errados levam a decisões equivocadas.

Outra técnica computacional bastante utilizada no processo de análise da situação de transformadores elétricos, são as Redes Neurais Artificiais (ANN). Algumas pesquisas empregam ANNs devido ao fato desta possuir a capacidade de processar dados não lineares, sendo um fator positivo e que pode favorecer o uso em aplicações direcionadas a tomada de decisão [Moraes and Castro 2018, Zhang 2014]. Mesmo com todo este potencial, a ANN ainda apresenta dificuldades em relação a sua estrutura, topologia, bem como o número de neurônios e camadas necessárias para um resultado otimizado. Dependendo dos valores, o tempo gasto para o treinamento pode ter um valor alto em relação a outras técnicas, além de impossibilitar a análise dos parâmetros de modo individualizado [Xu 2017].

A máquina de vetor de suporte (SVM) também pode ser utilizada em processos de classificação de transformadores elétricos conforme sua situação de risco [Xu 2017, Sahri and Yusof 2014]. SVM é baseada na técnica do aprendizado estatístico, a qual propicia a criação de modelos mais eficientes para classificação contornando alguns problemas encontrados na ANN. Fatores como a estrutura, parâmetros de construção e de entrada junto com as funções de *kernel* podem afetar o desempenho do modelo.

[Shrivastava and Choubey 2013] usaram a abordagem da ARM para diagnosticar a condição do transformador. Os atributos selecionados para a construção das regras consistem nas concentrações dos gases dissolvidos e os valores obtidos por meio da técnica do IEC. Os valores foram relacionados como base para geração de códigos que emitem o estado em que o equipamento se encontra. Porém, nenhum critério objetivo de seleção das regras para obtenção dos padrões mais relevantes para diagnóstico dos transformadores elétricos foi utilizado.

[Mirowski and LeCun 2012] informam os problemas enfrentados para o uso de métodos estatísticos na análise de gases dissolvidos, que são: a falta e a insuficiência de dados públicos para teste, o nível de complexidade das técnicas híbridas e a não existência de um estado da arte para comparação. Estes autores mencionam que a falta de balanceamento dos dados desfavorece o uso de algoritmos baseados em aprendizado estatístico para análise nesta área.

[Newman 2010] afirma que vários sistemas podem ser representados como redes em que os dados podem ser reunidos seguindo algum critério. A função do sistema que a rede representa pode indicar qual a forma ideal desta. Algumas abordagens promovem a análise de uma rede de acordo com um item (nó) objetivo. A fim de otimizar a identificação do conhecimento obtido em processos de ARM, [Calçada et al. 2018] de-

envolveram a *Filtered-ARN*, que permite a exploração de um nó objetivo com análise de dependência entre os elementos das AR. Por trabalhar com medidas objetivas assimétricas, a *Filtered-ARN* lida com o problema dos *datasets* desbalanceados vistos em [Mirowski and LeCun 2012].

3. *Filtered-ARN*

Em geral, uma rede R é representada como $R = (V, E)$, na qual V é um conjunto de vértices (ou nós) e E é um conjunto de arestas (ou links), que ligam alguns pares de vértices em V . Estatisticamente, um grafo pode ser caracterizado por valores derivados, como o grau médio dos nós e o comprimento médio (caminho) entre os nós. Características adicionais como: diâmetro da rede, número de triângulos, número de isomorfismos e o coeficiente de agrupamento, também podem ser analisados [Nettleton 2013].

Dada uma rede $R = (V, E)$, vários links e auto-conexões não são permitidas dependendo do tipo de rede que está sendo implementada. Se R é uma Rede Dirigida (RD), considere o conjunto universal, denotado por U , contendo todas as $|V| * (|V| - 1)$ potenciais ligações dirigidas entre par de nós em V , no qual $|V|$ denota o número de elementos em V . Se R não é uma RD, U contém $|V| * (|V| - 1)/2$ links. Deste modo, a representação da rede está relacionada diretamente ao tipo de dado que ela representa [Valverde-Rebaza and Lopes 2014].

Existem casos em que a ARM é feita com o objetivo de explicar itens predeterminados. A *Filtered-ARN* apresenta uma exploração das AR guiada por um único item objetivo. Esta exploração remove todas as AR que não são interessantes no contexto do item objetivo, de acordo com as métricas mínimas de *Suporte* e *Confiança*, além de filtros das medidas objetivas chamadas de *Added Value* e *Gain*, mostrando apenas as regras relevantes com a certeza da dependência entre os elementos das regras.

Definição 1 *Added Value (AV)* [-1;1]: descrita na Equação 1, a medida indica o quanto a frequência do consequente (também chamado de *RHS*) aumenta na presença do antecedente (conhecido também como *LHS*), ou seja, mede o ganho de *RHS* na presença de *LHS* [Sahar 2003]. Se $AV > 0$, a frequência de *RHS* aumenta na presença de *LHS*. Sendo $AV < 0$, a frequência de *RHS* diminui na presença de *LHS*. Se $AV = 0$, tem-se uma coincidência aleatória, ou seja, a frequência de *LHS* não altera a frequência de *RHS*.

$$AV = Confiança(LHS \Rightarrow RHS) - Suporte(RHS) \quad (1)$$

Definição 2 *Gain* [0;1]: é uma medida proposta por [Fukuda et al. 1996] (Equação 2) que forma um *trade-off* entre suporte e confiança, auxiliando na seleção das regras de acordo com a frequência da mesma e o valor da *Confiança Mínima*.

$$Gain = [Confiança(LHS \Rightarrow RHS) - ConfiançaMínima]Suporte(LHS) \quad (2)$$

Definição 3 Dado um conjunto de AR R , contendo regras de subconjuntos unitários, e um item objetivo Z , a *Filtered-ARN* é uma RD que modela todas as regras relacionadas ao item em Z [Calçada et al. 2018], como:

1. Cada aresta modela uma regra $r \in R$.
2. A partir de qualquer ponto da Rede, é sempre possível alcançar pelo menos 1 vértice representando um item Z .
3. Dado um vértice $v \in Filtered-ARN$, como $v \notin Z$. Não há caminho de qualquer item Z para v .
4. Se existe uma regra r como $RHS(r) \subset Z$, então a regra $r \in Filtered-ARN$.

No algoritmo *Filtered-ARN* faz-se uso de filtros com medidas objetivas assimétricas (*AV* e *Gain*), construindo o gráfico de acordo com as regras selecionadas. O algoritmo pode ser descrito em 3 passos:

- (1) Passo A: similar ao primeiro passo de todos os processos de ARM. Extração das regras com corte por suporte e confiança mínimos. A única restrição adicionada a essa etapa, se comparada a uma ARM convencional, é que as regras devem possuir conjuntos unitários no *LHS* e *RHS*. Essa restrição foi adicionada para facilitar a modelagem da *Filtered-ARN*.
- (2) Passo B: Efetuar o cálculo das medidas objetivas assimétricas *Added Value* e *Gain*, realizando a exclusão de todas as regras com $AV = 0$ e por um ganho mínimo (*MinGain*). Esta etapa guiará toda a exploração, pois definirá as regras de interesse que serão utilizadas com o item objetivo do qual a Rede será construída.
- (3) Passo C: escolher um item frequente Z , que será representado no conjunto de regras como o nó objetivo, e construir um *B-graph* (do inglês *Backward-directed hypergraph*) que flui de forma recursiva para Z . Logo após a escolha do item objetivo, é efetuada a construção da *Filtered-ARN*. Esta etapa é responsável por obter todas as regras que estão direta ou indiretamente relacionadas ao item objetivo e modelá-las. Primeiro, o item selecionado como objetivo é modelado no gráfico (nível = 0). Então, todas as regras que o item *LHS* não estão no gráfico e possuem o item do *RHS* no nível 0 são modelados na rede. O mesmo processo é feito para todos os itens no nível 1, nível 2 e assim por diante. Até que não haja mais regras a serem modeladas.

4. Materiais e Métodos

Todo o processo de mineração de dados foi executado, compreendendo desde a preparação e pré-processamento do conjunto de dados, antes de analisar as AR extraídas e construir a *Filtered-ARN* para identificação sobre os fatores associados a transformadores elétricos em situação de risco. As etapas podem ser vistas na Figura 1.

Figura 1. Fluxograma com as etapas do processo de construção das Redes.



O *dataset*¹ elaborado por [Velásquez and Lara 2018] serviu de base para o conjunto usado neste trabalho. Neste *dataset*, os transformadores são classificados de acordo com problemas causados pela presença de enxofre corrosivo. O conjunto de dados possui 61 instâncias que representam amostras isoladas de óleo isolante extraídas de transformadores de distribuição e de potência. Também são encontrados 53 atributos, os quais incluem concentrações de gases dissolvidos, aspectos físico-químicos, condição do equipamento, análises dos efeitos do enxofre corrosivo, recomendações para a recuperação do equipamento e resultados para os métodos tradicionais. Os motivos que justificam a manutenção dos atributos relacionados aos gases dissolvidos, propriedades PHCH e a condição do equipamento no novo *dataset* se dão pelo fato de que estes possuem valores para todas as instâncias e por indicação de profissionais² do setor elétrico, que se fundamentaram nas normas adotadas pela empresa para a manutenção destes equipamentos.

Foi adicionado um atributo chamado de [TGCD] cujo o valor é determinado pela soma das concentrações dos gases combustíveis mencionados na Tabela 1 e outro, denominado [OQIN], obtido por meio da razão entre os valores de [IT] e [NN] mencionados na Seção 2. Também foram inseridas as razões [C2H2/C2H4], [CH4/H2], [C2H4/C2H6], [CO2/CO], [O2/N2] e [C2H2/H2] vistas no método do IEC para diagnóstico do equipamento. O *dataset*³ usado para esta pesquisa possui 23 atributos e 61 instâncias.

Em relação aos gases dissolvidos, a classificação foi feita em quatro categorias como no critério adotado pelo IEEE. O valor 1 informa que o nível daquele gás é considerado normal, o valor 2 indica a possibilidade de um defeito, o valor 3 já mostra uma alta decomposição do gás no óleo e o valor 4 estabelece que a continuação da operação resultará em falha do transformador.

Os dados referentes às propriedades físico-químicas citadas na Seção 2 foram organizados em oito categorias de acordo com os valores obtidos. Os intervalos para os parâmetros foram estabelecidos usando o método de categorização estatística descrito por [Barbetta et al. 2010], no qual calcula-se o intervalo (I) dos valores para cada variável (Equação 3), e determina-se Max e Min como os valores máximo e mínimo da classe, respectivamente.

$$I = \frac{Max - Min}{8} \quad (3)$$

As razões do IEC foram classificadas também em quatro categorias. O código 0 informa divisão por zero para a mesma, o código 1 indica valor inferior a 0.1, o código 2 sinaliza um valor entre 0.1 e 1 e o código 3 é usado para informar um valor maior que 3.

A ARM foi feita usando o algoritmo *Apriori-TID*. O suporte mínimo e a confiança mínima considerados foram, respectivamente 0.0, 0.01 para que todas as regras de antecedente e consequente unitários fossem extraídas. Após a etapa de extração, foram aplicados os filtros $AV = 0$ e $mingain = 0.001$ sendo excluídas as regras que não apresentam comprovação de influência. Para comparação, a ARN proposta por [Pandey et al. 2009] também foi construída. Em ambas, o parâmetro “[Situacao]=Risco” foi selecionado como

¹<https://data.mendeley.com/datasets/6ng342f32t/1>

²Engenheiros Fábio Mendes da Silva e Jorge Farid Amate da empresa Eletronorte

³<https://drive.google.com/open?id=1A01T7dj81CHHP1IuBSLdei-1ywE17Zu>

Figura 3. ARN com “[Situacao]=Risco” como item objetivo. Os itens de nível 1 em azul não possuem LHS.

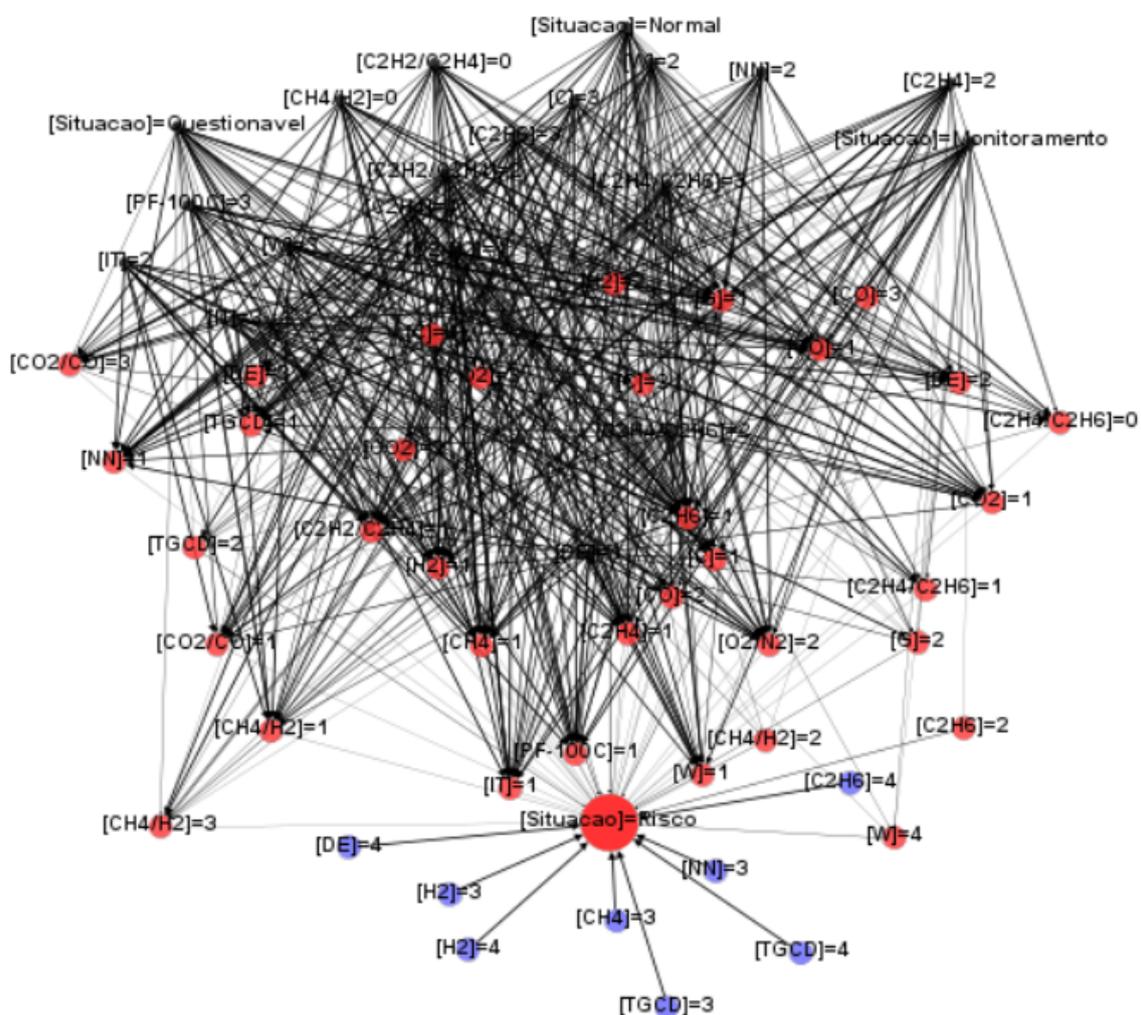


Tabela 2. Itens de nível 1 de ambas as redes que mais influenciam em “[Situacao]=Risco”.

ARN	<i>Filtered-ARN</i>
[TGCD]=4	[TGCD]=4
[TGCD]=3	[TGCD]=3
[H2]=4	[H2]=4
[H2]=3	[H2]=3
[CH4]=3	[CH4]=3
[C2H6]=4	[C2H6]=4
[DE]=4	[DE]=4
[NN]=3	[NN]=3
[PF-100C]=1	
[W]=1	
[IT]=1	

com [Carrapato et al. 2015]. Se analisados em conjunto, os mesmos gases são associados ao problema das descargas parciais [Américo et al. 2017, Bechara 2010], que também corroboram para a validação de “[DE]=4”, indicando que o óleo não está provendo o isolamento elétrico e magnético adequado pela presença de algum contaminante e “[NN]=3” que informa acidez elevada, comprometendo o isolante sólido, que não fornece o enclausuramento apropriado para a parte ativa. Observações sobre estes valores são informados no atributo recomendações e pelos resultados de métodos não computacionais presente no *dataset* elaborado por [Velásquez and Lara 2018], sendo um conhecimento já esperado.

O método de geração da *Filtered-ARN* consegue excluir as regras que não propiciam associações com validação matemática de influência, portanto possibilita uma identificação otimizada do conhecimento proporcionando a elaboração de hipóteses de estudo com maior probabilidade de serem verdadeiras. Os resultados obtidos pelas duas Redes foram apresentados a profissionais² do setor elétrico, que analisaram as informações e validaram os *LHS* das regras apresentadas na *Filtered-ARN* com base nas normas adotadas pela empresa para manutenção destes equipamentos. Ao comparar com a *ARN*, estes profissionais² apontaram que as regras exclusivas do nível 1 desta Rede podem estar presentes em situações de risco, mas não são determinantes para tal situação.

6. Conclusões

A pesquisa foi desenvolvida para verificar os fatores associados à condição de risco em transformadores elétricos com o uso da *Filtered-ARN*, que apontou 40 fatores associados ao item alvo, dos quais 8 estão diretamente conectados ao nó alvo.

Uma primeira confirmação dos resultados foi feita com base no campo recomendações do *dataset* de [Velásquez and Lara 2018]. A *Filtered-ARN* permitiu que a análise ocorresse de maneira bem mais simples. Os resultados foram comparados com os de uma *ARN* para comprovação de melhorias na formulação de hipóteses que levam o transformador à situação alvo, provando que a *ARN* não foi capaz de eliminar algumas das *AR* produzidas. Em um segundo momento, os *LHS* das *AR* conectadas diretamente ao item alvo foram apresentados a profissionais² do setor elétrico e confirmados pelos mesmos com base nas normas que a empresa adota na manutenção de seus equipamentos. Portanto, a eliminação das *AR* pelo critério de filtragem impactou positivamente no processo de identificação do conhecimento.

É importante lembrar que o caso visto na pesquisa leva em consideração os dados de amostras isoladas de óleo isolante de transformadores de distribuição e potência. Para pesquisas futuras, considera-se o acompanhamento deste tipo de análise utilizando um histórico de amostras de óleo isolante dos equipamentos e a inclusão de mais atributos para refinar ainda mais os diagnósticos.

Agradecimentos

Os autores agradecem a Ricardo Manuel Arias Velásquez e Jennifer Vanessa Mejía Lara, pela criação do conjunto de dados usado como base e pela disponibilização dos mesmos. Aos funcionários da subestação São Luís II da empresa Eletronorte, em especial aos engenheiros Fábio Mendes da Silva e Jorge Farid Amate, por toda assistência prestada

no entendimento dos dados avaliados.

Referências

- Agrawal, R., Imielinski, T., and Swami, A. (1994). Mining Association Rules between Sets of Items in Large Databases. *Special Interest Group on Management of Data*, 22(2):207–216.
- Agrawal, R. and Shafer, J. C. (1996). Parallel mining of association rules. *IEEE Transactions on Knowledge and Data Engineering*, 8(6):962–969.
- Agrawal, R. and Srikant., R. (1994). Fast algorithms for mining association rules. In *Proceedings of Twentieth International Conference on Very Large Data Bases (VLDB)*, pages 487–499, San Francisco, CA, USA. Morgan Kaufmann Publishers Inc.
- Américo, J., Cabral, S., Stefenon, S. F., Salvador, M., Oliboni, C. P., and Madruga, G. G. (2017). Um estudo qualitativo das técnicas utilizadas para medição e localização de descargas parciais em transformadores de potência. *Espacios*, 38:14–25.
- Barbetta, P. A., Reis, M. M., and Bornia, A. C. (2010). *Estatística para cursos de engenharia e informática*. Atlas, São Paulo, 3rd edition.
- Bechara, R. (2010). Análise de falhas de transformadore de potência. Master’s thesis, Escola Politécnica, Universidade de São Paulo.
- Brito, L. C., Marques, A., Ribeiro, C., Rocha, A., Dias, Y., Moura, N., Palhares, P., Azevedo, C. H., and Santos, J. (2016). Abordagem combinada dos guias ieee e iec para interpretação de análises de gases dissolvidos em óleo mineral isolante. In *VI Simpósio Brasileiro de Sistemas Elétricos–SBSE*, pages 1–6. 0.
- Calçada, D. B., Rezende, S. O., and de Padua, R. (2018). Asymmetric Objective Measures applied to Filter Association Rules Networks. In *Latin American Computing Conference - CLEI-LACLO*, pages 1–10, São Paulo. 0.
- Calçada, D., Rezende, S., and Teodoro, M. (2019). Analysis of green manure decomposition parameters in northeast brazil using association rule networks. *Computers and Electronics in Agriculture*, 159:34–41.
- Calçada, D. B. (2019). *Redes de regras de associação filtradas e multialvo*. PhD thesis, Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo.
- Campos, F. d. S. (2016). Avaliação da eficiência dos Óleos isolantes em transformadores em rede de transmissão elétrica. Monografia (Bacharel em Engenharia de Energia), Universidade de Brasília - UnB.
- Carrapato, M. A., Flauzino, R. A., Costa, P. I. d., Silva, I. N. d., and Gonzalez, C. G. (2015). Identificação da condição operativa de transformadores de potência por meio de mapeamento nebuloso dos gases dissolvidos no óleo mineral isolante. In *Simpósio Brasileiro de Automação Inteligente - SBAI*. SBA.
- Fukuda, T., Morimoto, Y., Morishita, S., and Tokuyama, T. (1996). Data Mining Using Two-Dimensional Optimized Association Rules: Scheme, Algorithms, and Visualization. In *Proceedings of the 1996 ACM SIGMOD international conference on Management of data*, pages 13–23, New York, NY, USA. ACM.

- Hussein, A., Yaacob, M., and Othman, F. (2015). An expert system for diagnosing faults and assessing the quality insulation oil of power transformer depending on the dga method. *Journal of Theoretical and Applied Information Technology*, 78:278–285.
- Kaur, M. and Kang, S. (2016). Market basket analysis: Identify the changing trends of market data using association rule mining. *Procedia Computer Science*, 85:78–85.
- Kim, C., Lee, H., Seol, H., and Lee, C. (2011). Identifying core technologies based on technological cross-impacts: An association rule mining (ARM) and analytic network process (ANP) approach. *Expert Systems with Applications*, 38(10):12559–12564.
- Mirowski, P. and LeCun, Y. (2012). Statistical machine learning and dissolved gas analysis: A review. *IEEE Transactions on Power Delivery*, 27(4):1791–1799.
- Moraes, H. R. S. and Castro, A. R. G. (2018). Deep neural networks with application to transformer failure diagnosis. In *Anais do XV Encontro Nacional de Inteligência Artificial e Computacional*, pages 287–298, Porto Alegre, RS, Brasil. SBC.
- Nettleton, D. F. (2013). Data mining of social networks represented as graphs. *Computer Science Review*, 7(1):1–34.
- Newman, M. (2010). *Networks: An introduction*, volume 55. Oxford University Press, New York, USA, 1st edition.
- Nogueira, I. and Santos, M. M. M. (2016). Descritores físico-químicos para ação preditiva de dielétricos líquidos em transformadores de energia: Estudo de caso da cpfl paulista. Trabalho Final de Graduação (Bacharel em Engenharia Elétrica), Universidade Federal de Itajubá.
- Pandey, G., Chawla, S., Poon, S., Arunasalam, B., and Davis, J. G. (2009). Association Rules Network: Definition and Applications. *Statistical Analysis and Data Mining*, 1(4):260–179.
- Sahar, S. (2003). *What Is Interesting: Studies on Interestingness in Knowledge Discovery*. PhD thesis, Tel-Aviv University.
- Sahri, Z. B. and Yusof, R. (2014). Support vector machine-based fault diagnosis of power transformer using k nearest-neighbor imputed dga dataset. *Journal of Computer and Communications*, 02:22–31.
- Shrivastava, K. and Choubey, A. (2013). Data mining approach with iec based dissolved gas analysis for fault diagnosis of power transformer [j]. *International Journal of Engineering Research and Technology*, 2(3):1–11.
- Silva, T. H. V. e. (2017). Desenvolvimento de uma metodologia de interpretação de análise de gases dissolvidos em Óleo para transformadores de potência usando lógica fuzzy. Master's thesis, Universidade Federal de Goiás.
- Valverde-Rebaza, J. C. and Lopes, A. d. A. (2014). Link prediction in online social networks using group information. In Murgante, B., Misra, S., Rocha, A. M. A. C., Torre, C., Rocha, J. G., Falcão, M. I., Tanir, D., Apduhan, B. O., and Gervasi, O., editors, *Computational Science and Its Applications – ICCSA 2014*, pages 31–45, Cham. Springer International Publishing.

- Veloso, D. P. (2017). Metodologia fuzzy aplicada ao triângulo de duval para diagnóstico de transformadores de potência por meio de análise de gases dissolvidos. Master's thesis, Universidade Federal de Goiás.
- Velásquez, R. M. A. and Lara, J. V. M. (2018). Corrosive sulphur effect in power and distribution transformers failures and treatments. *Engineering Failure Analysis*, 92:240–267.
- Xu, H. (2017). An intelligent fault diagnosis approach for power transformers based on support vector machines. Master's thesis, University of Alberta.
- Yaacob, M., Hussein, A., and Othman, F. (2015). Dga method-based anfis expert system for diagnosing faults and assessing quality of power transformer insulation oil. *Modern Applied Science*, 10:13.
- Zhang, L. (2014). Fault diagnosis of oil-immersed transformers using self-organization antibody network and immune operator. *Mathematical Problems in Engineering*, 2014:1–8.