

CADERNOS DO IME – Série Estatística

Universidade do Estado do Rio de Janeiro - UERJ
Rio de Janeiro - RJ - Brasil
ISSN impresso 1413-9022 / ISSN on-line 2317-4536 - v. 33, p. 35 - 50, 2012

ANÁLISES DE MÉTRICAS DE SIMILARIDADES EM SÉRIES TEMPORAIS PARA RECONHECIMENTO DE PADRÕES

João Francisco Monteiro Barragan
Universidade Federal da Bahia
J074@hotmail.com

Cristiano Hora de Oliveira Fontes
Universidade Federal da Bahia
cfontes@ufba.br

Otacílio José Pereira
Universidade Federal da Bahia
otaciliojpereira@gmail.com

Sérgio Torres Sá Barreto
Petrobras
sa.barretto@petrobras.com.br

Luciana de Almeida Pacheco
Petrobras
lucianapacheco.spassu@petrobras.com.br

Resumo

Este artigo apresenta uma metodologia para reconhecimento de padrões de operação de uma turbina a gás da unidade termelétrica (UTE) Rômulo Almeida, integrante do parque da Petrobrás. A distinção de diferentes padrões de operação, mais especificamente, partidas sem falha e com falha devido a desarme (trip) por dispersão de temperatura foi possível devido a um vasto banco de dados obtidos a partir de um Sistema de Gerenciamento de Informações de Plantas Industriais, ou Process Information Management System (PIMS), disponível na UTE. Os padrões identificados foram analisados utilizando métricas de similaridade univariáveis (distância euclidiana entre os coeficientes de autocorrelação, visando prever distinções entre as partidas sem e com trip por dispersão de temperatura. A possibilidade de gerar uma ferramenta que possa apontar para uma possível falha no equipamento com antecedência é um resultado potencial para a implementação de um sistema de controle inteligente.

Palavras-chave: Reconhecimento de Padrões; Análise de Similaridade; Turbina a Gás.

1. Introdução

O avanço da tecnologia da informação (TI) viabilizou que empresas de diversos setores passassem a coletar e armazenar, de forma sistemática, os dados provenientes de suas atividades diárias. Contudo, a maior parte delas ainda não usa essa grande quantidade de dados para transformá-los em informação útil, de forma a garantir uma melhor gestão e melhorias operacionais. Métodos associados à geração de conhecimento (*Knowledge Discovery on data* - KDD) e mineração (*Data Mining* – DM) são métodos que utilizam dados para a geração de conhecimento útil, de forma a apontar possíveis melhorias nos processos. (FAYYAD *et al.*, 1996).

A partir da mineração de dados se pode analisar dados comportamentais, obtendo conhecimento que estava “escondido” na base de dados, gerando aprendizado e dados complementares que podem influenciar no desenvolvimento de estratégias na organização. Esta técnica está sendo usada para descrever características do passado assim como prever tendências para o futuro. Sua utilização permite avanços tecnológicos e descobertas científicas, além de garantir uma vantagem competitiva invejável (SFERRA; CORRÊA, 2003). Nas empresas de infraestrutura, como exemplo as ligadas ao fornecimento de energia elétrica, a gestão deve potencializar o rendimento da planta, sua lucratividade proveniente da venda de energia e evitar pesadas penalidades contratuais em eventuais ocorrências de cessão do fornecimento (ANEEL, 2009).

Este artigo apresenta uma metodologia para reconhecimento de padrões de operação de uma turbina a gás de escala comercial, com capacidade de geração de energia de 27MW, representando o principal equipamento da unidade termelétrica (UTE) Rômulo Almeida, integrante do parque da Petrobrás. A aquisição de conhecimento e a mineração de dados foram realizadas a partir de um Sistema de Gerenciamento de Informações de Plantas Industriais, ou *Process Information Management System* (PIMS), disponível na UTE. Os padrões identificados foram analisados utilizando métricas de similaridade univariáveis (distância euclidiana entre os coeficientes de autocorrelação e análise no domínio da frequência), visando prever distinções entre as partidas sem e com trip por dispersão de temperatura.

A identificação de diferentes padrões de operação da turbina a gás consiste, em uma primeira etapa, na criação de um sistema inteligente baseado em lógica fuzzy

capaz de prever situações indesejáveis na planta. O sistema inteligente servirá de suporte aos operadores e gestores da unidade termelétrica, principalmente no que se refere a evitar a interrupção no fornecimento de energia, e consequentemente, as penalidades contratuais decorrentes dessa cessão.

Na seção 2 são apresentados o cenário de aplicação, o sistema de estudo e os aspectos envolvidos na predição de falhas em plantas industriais. A seção 3 demonstra a metodologia utilizada para o reconhecimento de padrões e predição de falhas. Na seção 4 são abordadas as medidas de similaridade convencionais, destacando os coeficientes de autocorrelação. A seção 5 apresenta os resultados obtidos, desde a coleta dos dados, até o processamento de dados e a análise estatística. Na última seção são tratadas as conclusões finais, indicando possibilidades de trabalhos futuros gerados por este projeto.

2. Cenário de aplicação

A pesquisa em questão possui aplicação na Unidade Termelétrica Rômulo Almeida (UTE-RA), situada no município de Camaçari (Bahia). A UTE-RA (Figura 1) é uma unidade integrante do parque industrial da Petrobrás, o qual possui três turbinas a gás (TG) com capacidade de geração de 27MW e uma caldeira de recuperação (*Heat Recovery Steam Generator* – HRSG). A unidade ainda possui uma Caldeira Auxiliar (CA) e uma turbina a vapor, com capacidade de geração de até 56MW. A unidade opera em ciclo combinado, tem como principal insumo o gás natural, produz 260,3 t/h de vapores de alta e baixa pressão e até 137MW de energia elétrica.

Figura 1 – Unidade Termelétrica Rômulo Almeida



Fonte: BARRETTO, 2010.

A energia gerada pela UTE é fornecida ao Sistema Interligado Nacional (SIN) no qual o Operador Nacional do Sistema Elétrico (ONS) coordena e controla a operação e transmissão de energia elétrica. Conforme acordado com o ONS, havendo qualquer ocorrência que impeça o fornecimento de energia por parte da UTE-RA, a unidade estará sujeita a penalidades significativas, que terão impacto direto no desempenho financeiro da Unidade.

Os vapores de alta e baixa pressão produzidos pela UTE são fornecidos a uma série de indústrias do pólo petroquímico de Camaçari. Também neste caso, o fornecimento é regido por regras previstas em contratos. Assim, o não cumprimento de certas cláusulas implicará na aplicação de severas multas contratuais.

Este cenário permite perceber claramente que é essencial para o gestor ter pleno domínio sobre o comportamento da planta para, no início do processo, anunciar sua capacidade produtiva com mínimo risco de interrupção; e depois, para ter meios de evitar ou se antecipar às falhas e assim garantir o fornecimento acordado. Assim, os riscos de sofrer severas penalidades e grandes prejuízos financeiros são reduzidos de forma significativa.

A análise até este ponto teve como enfoque os riscos financeiros provenientes de uma interrupção no fornecimento. Vale ainda destacar que existem outros riscos que podem ser gerenciados com uma solução de predição de falhas: riscos de danos aos equipamentos, que geram custos de reposição e prejuízos pelo tempo ocioso do equipamento, além dos riscos de acidentes que impactam a saúde e integridade dos funcionários.

A decisão em concentrar os esforços do estudo sobre as turbinas a gás se deu devido à maior quantidade de energia elétrica da unidade ser gerada pelas turbinas a gás (81MW são gerados pela turbina a gás e 56MW pela turbina a vapor). Na planta, o modelo da turbina a gás instalada é o RB211-G62 DF, do fabricante Rolls-Royce.

Uma turbina a gás, cuja configuração típica é mostrada na Figura 2, é uma turbomáquina que tem como insumos o ar e um combustível. No caso da UTE-RA, o combustível utilizado é o gás natural. A turbina a gás produz vapor e energia mecânica e, quando acoplada a um gerador elétrico, possibilita a geração de energia elétrica (SARAVANAMUTTOO, 1996).

Figura 2 – Turbina a Gás RB211 24G DLE



Fonte: ROLLS-ROYCE

A composição básica de uma turbina a gás é composta pelo compressor, câmara de combustão e expensor. O ar necessário à queima é admitido e comprimido pelo compressor, seguindo para a câmara de combustão, onde se mistura com o gás natural. Uma chama piloto é utilizada para dar início à combustão. Os gases formados na reação passam com altas velocidades e temperaturas pelo expensor da turbina, o qual é acoplado a um gerador de energia, transformando energia mecânica em energia elétrica. Este conjunto opera em um ciclo aberto, ou seja, o fluido de trabalho (ar) é admitido na pressão atmosférica e os gases de escape, após passarem pela turbina, são descarregados de volta na atmosfera sem que retornem à admissão.

As TG de modelo RB211 Rolls-Royce dispõem de nove combustores. O sistema de controle da RB211 foi projetado para desarmar o equipamento caso a temperatura de operação de alguma das câmaras se distancie em $\pm 150^{\circ}\text{C}$ da média de temperatura de todas as câmaras, de forma a proteger o equipamento de danos por dilatação diferencial. Quando este desarme ocorre, caracteriza-se um *trip* por dispersão de temperatura.

As turbinas podem falhar e parar (*trip*). Vários motivos podem causar uma falha (GIAMPAOLO, 2006), por exemplo: dispersão de temperatura, surgência, vibração, além de outras. Com base em entrevistas e análise dos relatórios de operação da turbina, selecionou-se investigar o desarme (ou *trip*) por dispersão de temperatura. A escolha desta forma específica de causa geradora se deu por ser um desarme com maior número de ocorrências desde 2009, por seu impacto na interrupção ser significativo e pelo potencial de predição com base nas variáveis coletadas.

3. Metodologia

O projeto como um todo compreende três etapas principais. A geração de amostras e classificação de padrões de processo, a análise de similaridade inter e intragrupos e o agrupamento e reconhecimento de padrões. Este artigo se enfoca com maior ênfase na segunda e terceira etapas do projeto, buscando medir a partir de técnicas estatísticas, similaridades dentre os grupos observados para o reconhecimento nos padrões de operação.

Devido ao grande fluxo de informações dentro do processo, utilizou-se como campo de estudo a área de KDD. Esta é uma área que visa a extrair conhecimento a partir do processamento de um grande volume de dados. A ideia chave é realizar um trabalho de análise que seria impossível ou impraticável por um usuário humano. Um processo típico bem disseminado dentro da área de KDD realiza as seguintes etapas: seleção de variáveis, pré-processamento, transformação, mineração de dados e interpretação, até a obtenção do conhecimento (FAYYAD, 1996; DE AMO, 2004). Estas etapas nortearam a seqüência de ações e atividades da pesquisa.

Na primeira etapa do projeto, através dos relatórios de operação e da base de dados proveniente do PIMS, todas ocorrências das três turbinas a gás foram levantadas e classificadas em partidas normais e com trip por dispersão de temperatura. O algoritmo de varredura linear (TRAUB, 1998) possibilitou a captura de cada evento de partida da turbina, contemplando várias variáveis importantes do processo, como temperatura de entrada do ar, vazão de gás natural, pressão na saída do compressor, temperatura dos gases de exaustão, potência bruta total gerada, dentre outros. Mais especificamente, os dados foram extraídos durante os anos de 2009 e 2010 (com período de coleta de 1 min.), sendo observado que a duração média de desarme da turbina devido a problemas de dispersão de temperatura após a partida é de 20 minutos.

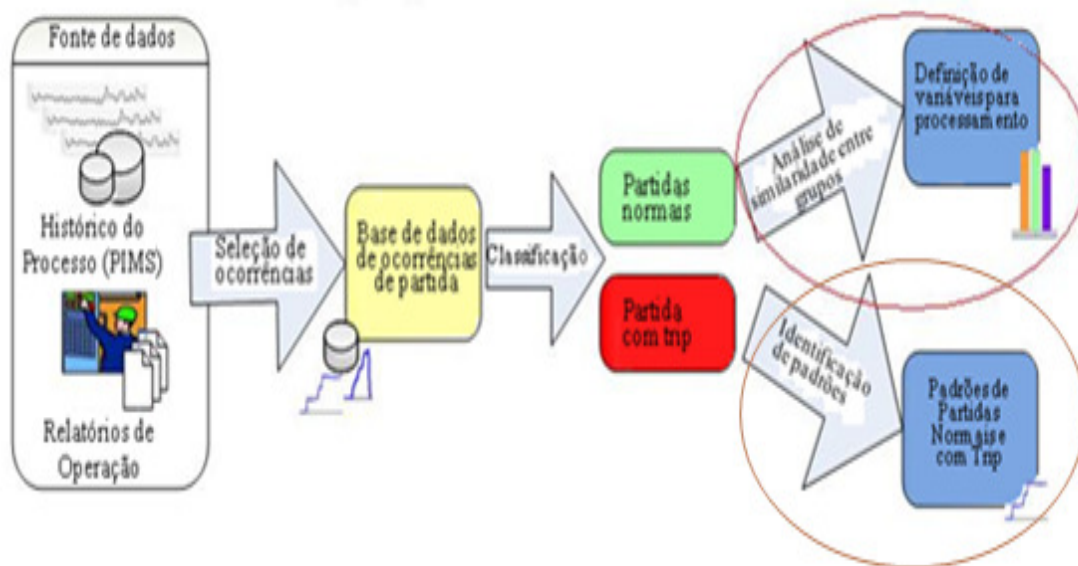
Em uma segunda etapa, os dados foram pré-processados, passando por um “enjanelamento” e normalização, sendo posteriormente aplicadas técnicas de medidas métricas de similaridade entre os grupos de partidas normais e partidas com trip, buscando evidenciar diferenças nas variáveis de processo que ocasionaram a ocorrência de parada da turbina. As métricas de similaridade foram aplicadas mais especificamente a quatro variáveis medidas e obtidas do banco de dados - temperatura de entrada dos

gases, vazão de gás natural, pressão na saída do compressor e temperatura dos gases de exaustão.

A terceira etapa do projeto consiste no agrupamento e reconhecimento de padrões. De posse de ferramentas matemáticas que possibilitem prever a probabilidade de ocorrência de falha, somado aos dados de operação da partida de uma turbina, é possível a construção de um sistema inteligente de controle baseado em lógica fuzzy, que poderá a partir da evolução temporal das variáveis da turbina, apontar a probabilidade de falha na partida da turbina.

A figura 3 a seguir apresenta um resumo do projeto, destacados pelos círculos os locais de aplicação deste artigo dentro do trabalho.

Figura 3. Visão Geral da Metodologia



Fonte: Pereira, 2012

4. Medidas de similaridade

A predição de valores está diretamente associada a uma busca no passado por padrões de informações similares aos encontrados no momento atual (H. WU *et al.*, 2005). Uma forma de se obter a predição de valores é a partir de técnicas de regressão numérica. Elas são capazes de gerar parâmetros que caracterizam curvas definidas pelos dados que pertencem à série de dados em análise (RAFIEI; MENDEL, 1997; B YI *et al.*, 2000). Uma destas técnicas aplicadas a séries univariadas é a utilização da autocorrelação (DUNN, 2005). O vetor autocorrelação contempla os valores de

autocorrelação entre pontos de uma série temporal para vários atrasos (deslocamentos de tempo) e representa uma alternativa de extração de características da série original.

$$R(k) = \frac{E[(X_t - \mu)(X_{t+k} - \mu)]}{\sigma^2} \quad (1)$$

A expectância (E) é estimada através da média aritmética. k é o deslocamento no tempo (lag) e σ^2 é a variância da série X_t .

A distância ou medida de similaridade entre dois vetores de autocorrelação (genericamente p e q) referentes a duas séries distintas pode ser obtida através da distância Euclidiana (ROSENFELD, 1968):

$$D = \sqrt{\sum_{i=1}^n (p_i - q_i)^2} \quad (2)$$

A distância Euclidiana, enquanto medida de similaridade, pode ser usada diretamente para o reconhecimento de diferentes padrões de partida para o caso em estudo. Neste caso, utilizou-se um algoritmo de agrupamento (“clustering”). No agrupamento, um conjunto de elementos (objetos) é analisado através de métricas de similaridade e grupos de objetos similares são obtidos. O algoritmo de agrupamento implementado foi o Fuzzy C-Means, que é baseado no seguinte problema de otimização:

Considerando-se n objetos $\{x_k | k = 1, \dots, n\}$, o problema consiste em obter c centros (ou padrões) associados a cada grupo (“cluster”) $\{v_i | i = 1, \dots, c\}$ e uma matriz de partição U que contempla o nível de pertinência de cada objeto a cada um dos grupos obtidos. O problema de otimização consiste em:

$$\min J_m(U, V) = \sum_{i=1}^c \sum_{k=1}^n u_{ik}^m \|x_k - v_i\|^2$$

Sujeito a

$$\sum_{i=1}^c u_{ik} = 1 \quad \forall k$$

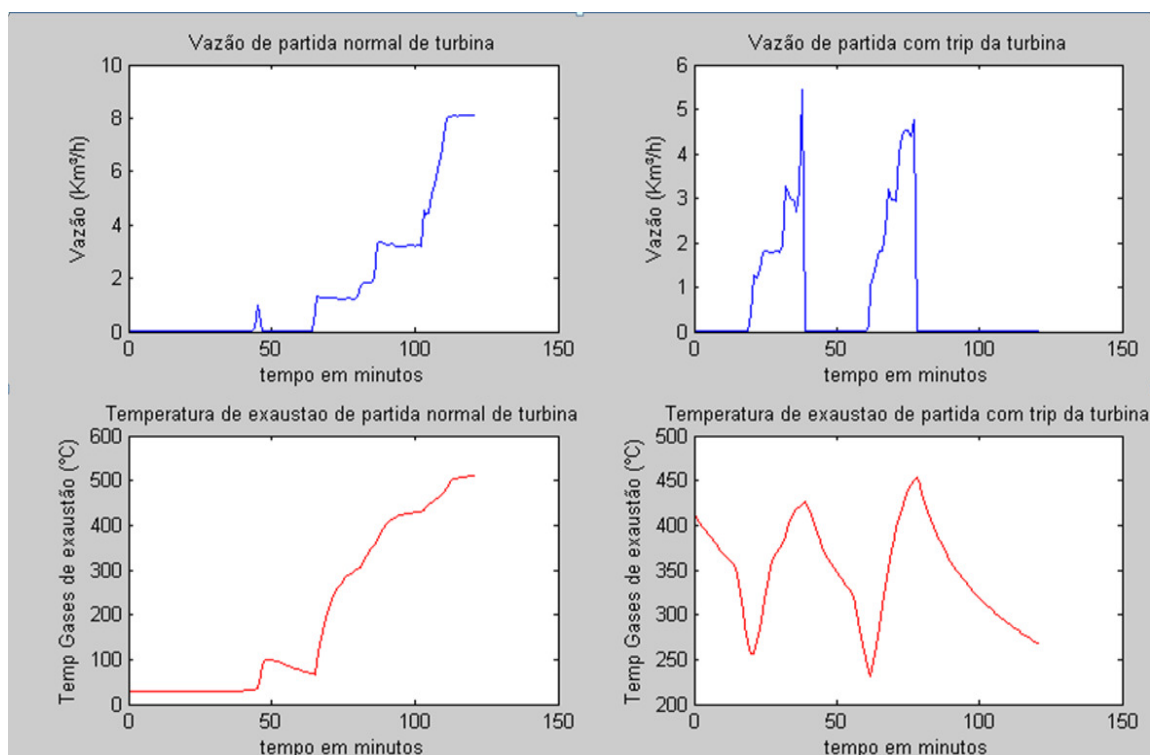
$u_{ik} \in [0,1] \quad \forall i, k$ é a *pertinência do objeto k ao cluster i* e m ($1 \leq m \leq \infty$) é o parâmetro fuzzificador.

5. Resultados e discussão

Primeiramente, a partir dos relatórios técnicos, e do banco de dados do PIMS foi possível identificar 80 eventos de partida da turbina no período de 2009 a 2010. A

figura 4 ilustra dois exemplos de medições de vazão (azul) e temperatura de exaustão (vermelho) em partidas da turbina, sendo uma delas classificada como partida normal e a outra como partida com trip por dispersão de temperatura. O início da partida da turbina está representado aproximadamente no instante de 60 minutos e, como já discutido, a ocorrência de cessão na vazão se dá aproximadamente após 20 minutos da partida da turbina.

Figura 4 - Dados de vazão e temperatura de partida normal e com trip por dispersão de temperatura.



Antes da aplicação de técnicas de extração de características, fez-se necessário a realização de um pré-tratamento das séries temporais demonstradas acima. Cada evento foi “enjanelado”, ou seja, fez-se o estudo da evolução das variáveis da turbina, desde o seu momento de partida até os 5, 8, 11, 20 e 33 minutos posteriores, gerando “retratos” de partida para diferentes intervalos. Cada um destes “retratos” (de diferentes intervalos de tempo) passou a representar eventos de partida sem ou com trip. Tais dados foram então normalizados, gerando o resultado ilustrado nas figuras 5 e 6 a seguir, que representam os mesmos eventos de partida da turbina da figura 4 obtidos após o devido tratamento dos dados para intervalos de 11 e 20 minutos.

Figura 5- Dados após pré-processamento para janela de 11 minutos.

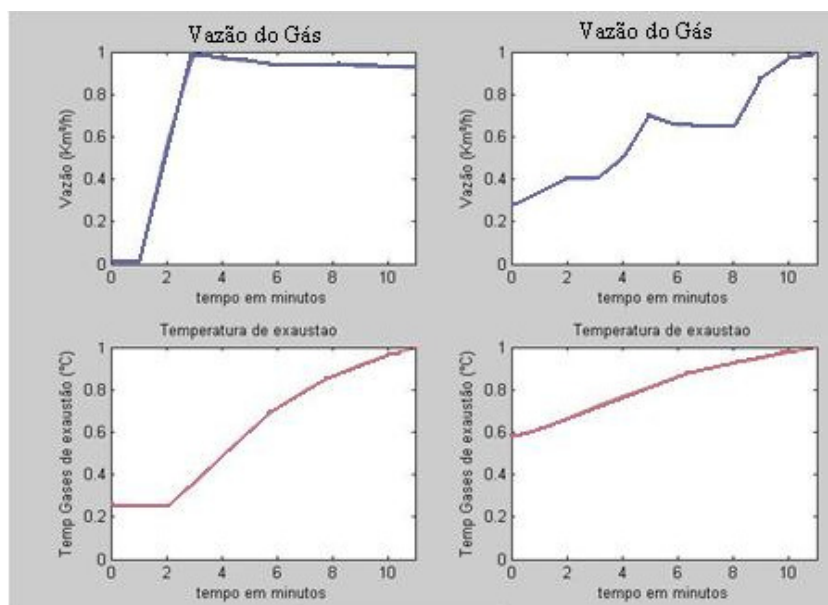
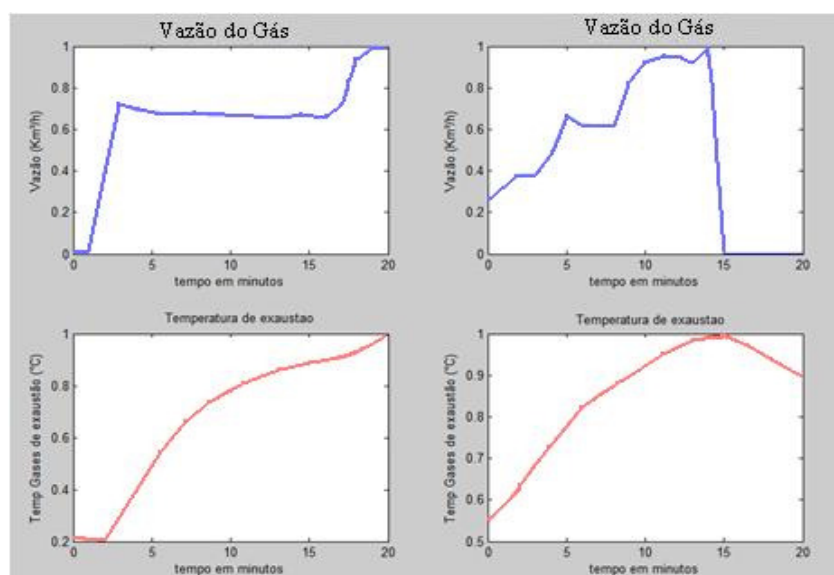


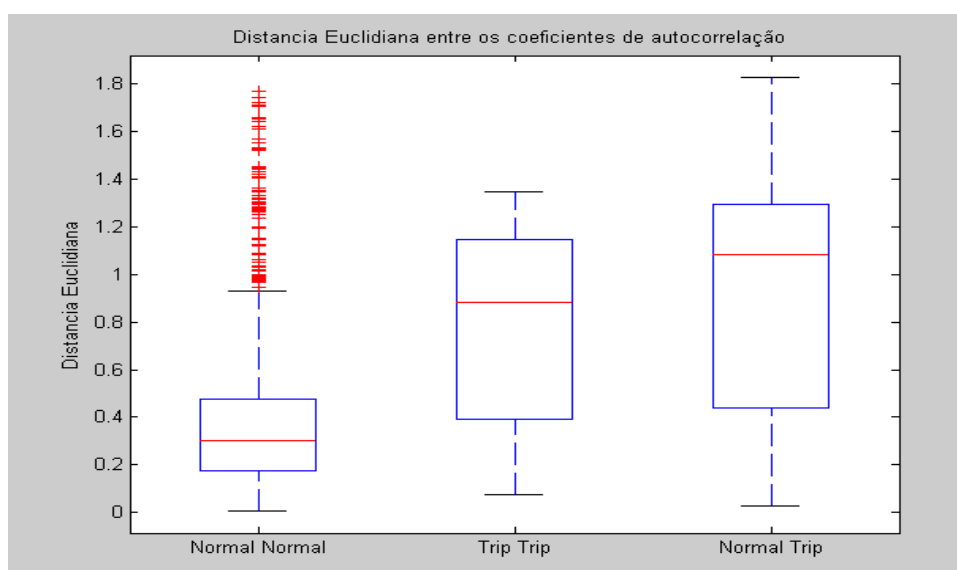
Figura 6- Dados após pré-processamento para janela de 20 minutos.



Para cada variável de partida da turbina, o cálculo dos coeficientes de autocorrelação foi realizado, tendo como produto duas matrizes. Uma das matrizes representa os coeficientes de autocorrelação das partidas normais e a outra os coeficientes das curvas que apresentaram *trip*. Estes coeficientes exprimem uma característica das curvas que o geraram, podendo ser comparadas a partir de funções distância, no caso utilizado, a distância euclidiana. É esperado que as curvas normais apresentem entre si, baixas distâncias euclidianas de seus coeficientes de

autocorrelação, apresentando uma tendência comum entre elas. Da mesma forma, espera-se uma pequena distância entre partidas com *trip*. Já a comparação entre partidas normais com partidas com *trip* tenderia a gerar maiores distâncias, provando as dissimilaridades entre os padrões. Este procedimento foi repetido para cada uma das quatro variáveis (temperatura de entrada, vazão, pressão e temperatura de exaustão) em diferentes janelas de tempo (5, 8, 11 e 20 minutos). O gráfico a seguir sintetiza os resultados obtidos para a variável temperatura de exaustão dos gases em janelas de 20 minutos:

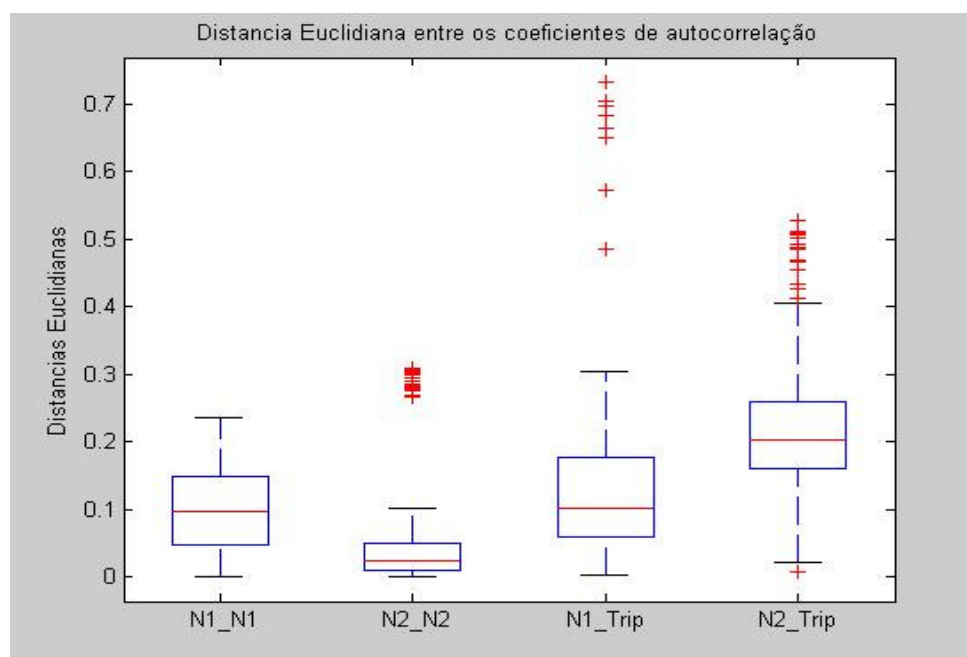
Figura 7. Gráfico *Boxplot* resultante das distâncias euclidianas inter e intragrupos



A partir deste resultado, é possível perceber uma segregação nas características das curvas com *trip* e normais (a última caixa, entre curvas normais e com *trip* apresentou maiores distâncias, portanto, maiores diferenças entre si). Entretanto, pôde se observar uma grande quantidade de pontos que se distanciaram muito do valor central dentro do grupo normal. Isto motivou um estudo mais aprofundado para a verificação da existência de dois grupamentos (clusters) dentro do grupo de partidas normais, ou seja, a ocorrência de duas formas distintas de partida da planta gerando um evento normal. A função Fuzzy C-means foi utilizada, buscando gerar dois grupos distintos entre as partidas normais da turbina. Foram obtidos então dois grupos de partidas normais, onde estes grupos se distinguiram basicamente pela temperatura inicial no momento da partida da planta. As tentativas de partidas de planta realizadas

logo após uma *trip* ainda possuíam altas temperaturas dentro da câmara de combustão, já que não houve tempo para o sistema esfriar. O outro grupo de partidas apresentou como característica própria a de possuir temperaturas iniciais na faixa da temperatura ambiente. Todo o procedimento de cálculo dos coeficientes de autocorrelação e suas respectivas distâncias euclidianas foi novamente realizado, utilizando agora dois grupos de partida normal distintos. Os resultados obtidos são apresentados na figura 8.

Figura 8- Boxplot resultante das distâncias euclidianas com dois grupos de partida normal.



Se observa a partir do gráfico anterior que o comportamento esperado para a turbina é o padrão de partida normal 2 (N2), pois este comportamento é mais distante dos eventos que levaram a posteriores paradas na planta. Isto se confirma visualizando no gráfico os maiores valores de distância euclidiana entre o grupo de partida normal 2 e partida com trip (representado pelo boxplot “N2_Trip”). A quantidade de *Outliers* foi significativamente diminuída com a inclusão de dois grupos, o que reforça a teoria da existência de dois “clusters”.

Foi gerado, a partir das curvas normais tipo 1 e 2, modelos de partida típicos de cada padrão. Comparando a distância entre as curvas de partida normais tipo 1 e 2 com o modelo de *trip* (figura 9) foi observado novamente que o padrão de partida numero 2 possui menor semelhança (maior distância) do que aquelas partidas normais do tipo 1. A figura 10 apresenta o modelo padrão para a partida do tipo 1 e do tipo 2 obtidos através

da média dos respectivos eventos observados. Como já foi discutido, um dos padrões de partida (padrão normal 1) se inicia a temperaturas muito mais altas do que o outro padrão encontrado. Este padrão de operação, em que a turbina é religada ainda quente após uma trip (padrão 1) se mostrou muito mais propenso a ocasionar em futuras reincidências de falhas. Sendo assim, é aconselhável manter uma evolução da variável temperatura semelhante ao padrão 2, mais confiável para o funcionamento do equipamento.

Figura 9- Distâncias euclidianas entre padrão de trip e padrões normais.

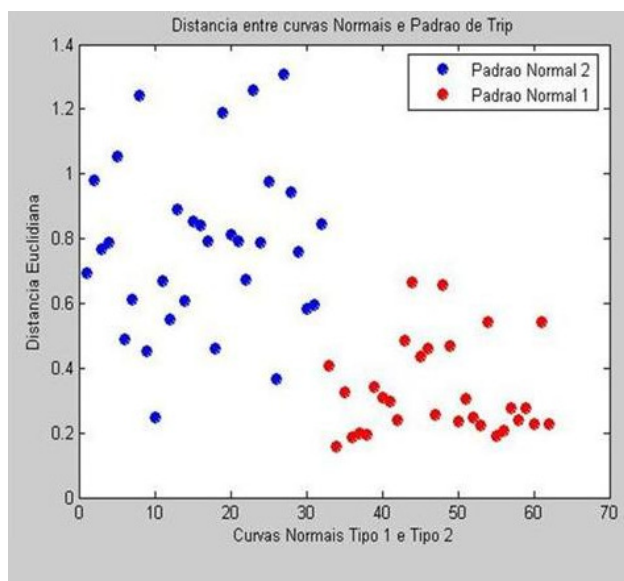
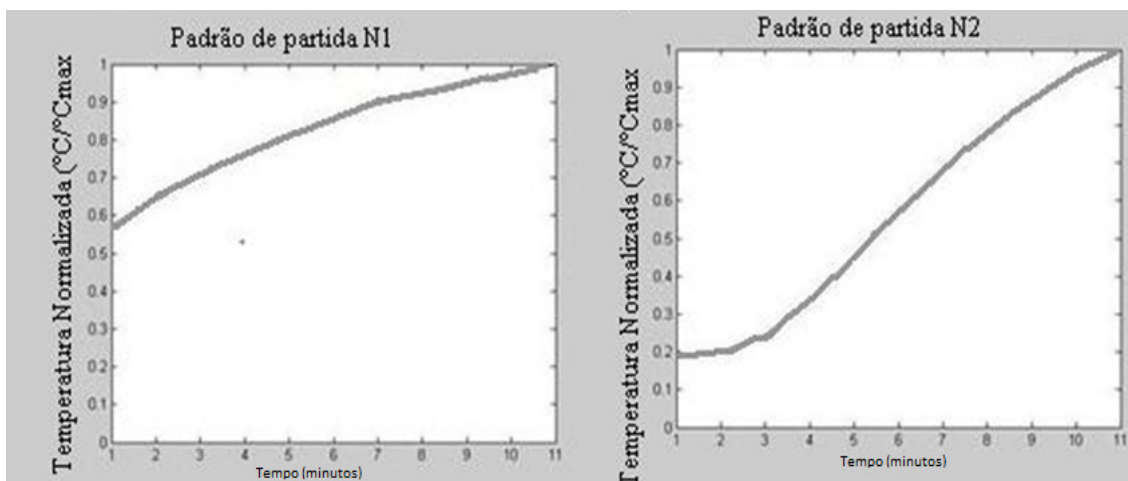


Figura 10- Definição de padrões de partida.



Desta forma, ao início de uma futura partida da turbina, pode-se extrair a característica dos coeficientes de autocorrelação gerados pela variável temperatura dos

gases de exaustão. Este vetor de característica próprio da partida da planta deve ser comparado à base de dados de partidas normais. Caso a partida se caminhe para um evento de partida normal do tipo 2, seus coeficientes de autocorrelação devem se assemelhar aos casos desejados, apresentando uma alta distância comparada aos dados de partidas com trip. Caso a distância tenda a valores maiores, há uma tendência de a partida estar se encaminhando para um caso de trip por dispersão de temperatura, devendo o operador tomar medidas preventivas.

6. Conclusão

O projeto contemplou uma metodologia que possibilita identificar padrões de operação em uma turbina a gás, de forma a gerar um sistema inteligente capaz de prever falhas, potencializando a capacidade de geração elétrica e evitando multas contratuais. Esta metodologia pode ser aplicada a quaisquer outros tipos de equipamentos, desde que haja uma disponibilidade de dados de operação do sistema, além de uma sistematização na coleta dos dados. Esta metodologia é fruto de um projeto de pesquisa que tem como cenário a UTE-RA. A primeira parte do trabalho contemplou a extração automática de todas as ocorrências de partidas da turbina desde 2009. Na segunda parte, os padrões de operação das turbinas durante a partida foram mapeados com ênfase na determinação de trip por dispersão de temperatura e de partida normal.

Uma sugestão interessante observada é a possibilidade da instalação de termopares nas diversas câmaras de combustão e não apenas na avaliação de um único valor. Como a trip é gerada pela dispersão nas temperaturas entre as câmaras, é importante a informação dos valores específicos de temperatura em cada câmara; dados estes que poderiam ser passados para o PIMS e esclareceriam de uma melhor forma os fatos geradores das paradas de operação da planta.

A condução do projeto, as práticas exercidas e os resultados obtidos permitem constatar que o bom uso da tecnologia de informação e da engenharia de controle e automação podem gerar soluções que permitam a melhora nos resultados operacionais, que neste contexto gera uma melhora da disponibilidade. Consequentemente, obtém-se um aumento na lucratividade e, sobretudo, a redução de prejuízos indesejáveis, decorrentes de paradas da planta e penalidades contratuais.

Referências

- AGÊNCIA NACIONAL DE ENERGIA ELÉTRICA – ANEEL. **Resolução Normativa nº385**. 2009.
- BALASKO, B.; ABONYI, J.; FEIL, B. Fuzzy Clustering and Data Analysis Toolbox. **Department of Process Engineering** University of Veszprem, 2002.
- CARVALHO L. A. **Data Mining – A Mineração De Dados no Marketing, Medicina, Economia, Engenharia e Administração**. Ed. Ciência Moderna, 2005.
- DUNN, F. P. **Measurement and Data Analysis for Engineering and Science**, New York: McGraw-Hill, 2005.
- FAYYAD, U. M.; PIATETSKY, G.; SMYTH, P.; UTHURUSAMY, R. **Advances in Knowledge Discovery and Data Mining**. MIT Press, 1996.
- GIAMPAOLO, A. **Gas Turbine Handbook, Principles and Practices**. Georgia: The Farmount Press, 2006.
- HOEPPNER, F.; KLAWONN, F.; KRUSE, R.; RUNKLER, T. **Fuzzy Cluster Analysis: Methods for Classification, Data Analysis and Image Recognition**. Chichester: John Wiley & Sons, 1999.
- KORBICZ, J.; KOSCIELNY, J. M.; KOWALCZUK, Z.; CHOLEWA, W. **Fault Diagnosis: Models, Artificial Intelligence, Applications**. 1.ed. Berlim: Springer, 2004.
- LIAO, T. W. Clustering of Time Series Data - a survey, **Science Direct**. Pattern Recognition 38, 1857 – 1874, Elsevier, 2005.
- RAFIEL, D.; MENDELZON, A. **Similarity-Based Queries for Time Series Data**, Vol 1, Pág 13- 15, 1997.
- ROLLS-ROYCE. **Material de Treinamento do Conjunto RB 211-G62**, DF, 2010.
- ROSENFELD, A.; PFALTZ, J. **Distance Functions on Digital Pictures, Pattern Recognition**, vol. 1, pp. 33-61. Ed. Elsevier, 1968.
- SARAVANAMUTTOO, H. I. H.; ROGERS, G. F. C.; COHEN, H. **Gas Turbine Theory**. 5.ed. Dorchester: Prentice Hall, 1996.
- SFERRA, H. H.; CORRÊA, M. C. A. Conceitos e Aplicações de Data Mining. **Revista Ciência e Tecnologia**, Vol. 11 nº 22 pág 20, Universidade Metodista de Piracicaba. 2003.
- TRAUB, O.; HOLLOWAY, G.; SMITH, M. D. 1998. Quality and Speed in Linear-scan Register Allocation. **Proceedings of the ACM SIGPLAN**, 1998
- YI, B. K.; SIDIROPOULOS, D.; JOHNSON, T.; JAGADISH, H. V.; FALOUTSOS, C.; BILIRIS, A. **Online Data Mining for Co-Evolving Time Sequences**. National Science Foundation, Vol. 1, Pág 1-3, 2000.

ANALYSIS OF SIMILARITY METRICS IN TIME SERIES FOR PATTERN RECOGNITION

Abstract:

This paper presents a methodology for pattern recognition of gas turbines operation in a thermal power plant thermoelectric unit belonging to the industrial park of the Brazilian Oil Company. The distinction of different patterns of operation, more specifically, starts with and without failure caused by temperature dispersion was accomplished using a large database obtained from the System Information Management of Industrial Plants, available in the unit. The patterns recognized were analyzed using similarity metrics applied to point-prototype problems (Euclidean distance between the autocorrelation coefficients), aiming to provide distinctions between events with and without failure. The results show the potentiality of the methodology to predict failures and the possibility of use the patterns as inputs of classifiers based on intelligent control systems.

Key-Words: Pattern Recognition; Similarity Analysis; Gas Turbine