



Grupo de Estudo de Sistemas de Distribuição-GDI

Aplicação de Inteligência Artificial no processo de triagem de chamados emergenciais da distribuidora visando redução de deslocamentos improcedentes no serviço de campo

WENDELL WILLIAM TEIXEIRA(1);
JOÃO PEDRO FERREIRA DOS SANTOS FREITAS(1);
RAUL SENA FERREIRA(2);
ANDRÉ LUIZ DE MESQUITA MELO(2);
TÁSSIO SIMIONI(2);
CPFL ENERGIA(1);
RADIX ENGENHARIA(2);

RESUMO

No dia a dia da operação de uma distribuidora de energia recebe-se muitos contatos de clientes para solicitar manutenção de rede ou relatar problemas no suprimento de energia.

Geralmente, quando uma ligação é recebida pelos canais apropriados, o problema relatado passa por uma fase de triagem e, no final, uma equipe de manutenção é enviada para o local com o objetivo de resolver o problema do cliente.

Entretanto, nem todos os problemas são de responsabilidade da empresa, este tipo de situação é uma das causadoras do denominado “Deslocamento Improcedente”, que se trata de um deslocamento desnecessário para a equipe de campo, que gera custos de combustível e impactos logísticos.

Um volume elevado de deslocamentos improcedentes pode resultar em penalidades pesadas para a empresa, uma vez que a equipe não estará disponível para atender clientes que realmente precisariam de assistência.

Para resolver esse problema, propomos uma solução de aprendizado de máquina supervisionada que usa informações relatadas pelo cliente e dados da distribuidora de energia para prever a probabilidade de uma determinada solicitação do cliente gerar um deslocamento improcedente. Nossos primeiros resultados indicam que o modelo alcançou 84% de assertividade dentro conjunto de dados avaliados (100 mil eventos).

Neste trabalho, mostramos como construímos esse modelo, pré-processamos as informações e como essa solução pode ser aplicada para diminuir os custos de manutenção dentro de uma empresa de distribuição de energia, parte de um projeto de Pesquisa e Desenvolvimento ANEEL DE3028 “Ferramenta em tempo real para redução de deslocamentos improcedentes no serviço de campo das distribuidoras do grupo CPFL Energia”, proposto pela CPFL Piratininga em cooperação com CPFL Paulista, CPFL Santa Cruz, Rio Grande Energia Sul.

PALAVRAS-CHAVE

Deslocamento Improcedente, Eficiência Operacional, Machine Learning, Inteligência Artificial, Chamados Emergenciais, Call back

1.0 - INTRODUÇÃO

As empresas de distribuição elétrica possuem ativos dispersos geograficamente em suas áreas de concessão. Tal característica exige que serviços de campo sejam executados para garantir o fornecimento de energia com segurança, qualidade e eficiência.

Durante o dia a dia da operação da rede de distribuição, há atividades programadas e emergenciais. Atividades programadas partem de planejamento prévio e não são alvo de investigação neste trabalho. Já as atividades emergenciais ocorrem em tempo real a partir de algum problema detectado e informado pelo cliente, e é o objeto de estudo do presente trabalho.

Os deslocamentos emergenciais são diversos e as demandas são iniciadas através de solicitações feitas por clientes via *Call Center* ou outros meios. Esses deslocamentos são classificados em dois tipos: Procedente ou Improcedente. Um deslocamento procedente é caracterizado por ser um serviço necessário ou obrigatório. Portanto, resulta em serviço e/ou agrega valor à empresa. Já um improcedente não agrega valor à empresa e/ou resulta serviço. Consequentemente, gera gastos e exposição dos colaboradores a riscos desnecessários. Um exemplo de deslocamento deste tipo é falta de luz causada por problema interno do cliente, como fusível queimado, que não é responsabilidade da companhia de distribuição.

Além do gasto resultante de um deslocamento improcedente com o despacho de uma equipe de manutenção, atender a esse tipo de demanda pode causar atrasos no atendimento a clientes que de fato precisam de atendimento, o que por sua vez resulta em multas à companhia. Isso acontece porque o número de equipes de manutenção disponíveis é limitado.

Por outro lado, deixar de atender um cliente por achar que o deslocamento será improcedente poderá causar uma multa caso ele realmente precise do atendimento. Além disso, tipo de cliente pode fazer com que a multa varie. Um hospital, por exemplo, pode ter a multa associada muito maior do que um cliente residencial.

Dessa forma, é necessário priorizar o envio de equipes de acordo com os seguintes critérios:

- 1) Qual a chance de ser improcedente?
- 2) Qual o custo de não enviar a equipe de manutenção?

Levando em conta essas duas questões, nós propomos o sistema de Redução de Deslocamentos Improcedentes (RDI), um sistema de informação em tempo real que utiliza aprendizado de máquina supervisionado em conjunto com uma técnica de modelo preditivo de controle (MPC), para auxiliar essa tomada de decisão. O algoritmo de aprendizado de máquina calcula a probabilidade de um chamado resultar em deslocamento improcedente, enquanto o MPC calcula o risco associado com o possível custo de não responder ao chamado. O objetivo da ferramenta proposta é reduzir o custo operacional causado pelos deslocamentos improcedentes.

O restante deste trabalho está dividido da seguinte forma: a seção 2 apresenta os trabalhos relacionados. A seção 3 traz a implementação da ferramenta, explicando como é feito o modelo matemático e como pode ser usada. A seção 4 apresenta os experimentos e os resultados que validam a solução proposta. Já a seção 5 discute as principais características da solução e possibilidades de trabalhos futuros.

2.0 - TRABALHOS RELACIONADOS

Podemos tratar o envio de equipes de manutenção como um problema de alocação de recursos. Esse tipo de problema pode ser resolvido utilizando técnicas como MPC (Castanon & Wholetz, 2009), (Zhang & Xia, 2011), e aprendizado de máquina (Maghrebi, et al., 2013), (Roy, 2017), (Ferrari-Trecate, et al., 2004).

Alguns destes trabalhos envolvem envio de equipes: (Maghrebi, et al., 2013) trata do despacho de caminhões com concreto para as áreas que precisam do material. Já (Roy, 2017) compara estratégias baseadas em máxima recompensa e processo de decisão de Markov (Papadimitriou & Tsitsiklis, 1987) no envio de ambulâncias em um serviço de emergência. Outros trabalhos também envolvem envio de equipes de manutenção, porém sem uso de chamados: (Barriquello, et al., 2017) propõe um sistema de apoio à decisão para manutenção preditiva em redes de distribuição de energia e (Yamato, et al., 2017) utiliza o som coletado por microfones próximos a máquinas e/ou outros equipamentos para identificar situações anômalas e então enviar uma equipe de manutenção.

Existem outros trabalhos que focam na classificação automática de chamadas em um *Call Center*. Um dos primeiros trabalhos (Busemann, et al., 2000) utilizou processamento de texto em conjunto com máquinas de vetores de suporte (Hearst, et al., 1998) para identificar qual problema está sendo reportado via e-mail e sugerir uma solução. Já em (Tang, et al., 2003), os autores propuseram métodos para classificar chamados de um *Call Center* através da gravação da conversa.

3.0 - SISTEMA DE REDUÇÃO DE DESLOCAMENTOS IMPROCEDENTES

O sistema de RDI foi pensado como um sistema de apoio à decisão para que o operador identifique chamados com alta chance de serem improcedentes e possa tomar alguma atitude em relação a isso. Ele pode, por exemplo, tentar em contato com o cliente para verificar causas comuns de improcedência.

No computador do operador fica o aplicativo do RDI que se comunica a um servidor onde fica um serviço de API. Este serviço por sua vez busca os chamados de um banco de dados centralizado e utiliza os modelos matemáticos já salvos para obter a chance de cada chamado ser improcedente. O banco de dados centralizado reúne informações de diferentes bases através de um processo de ETL que atualiza o banco em quase tempo real.

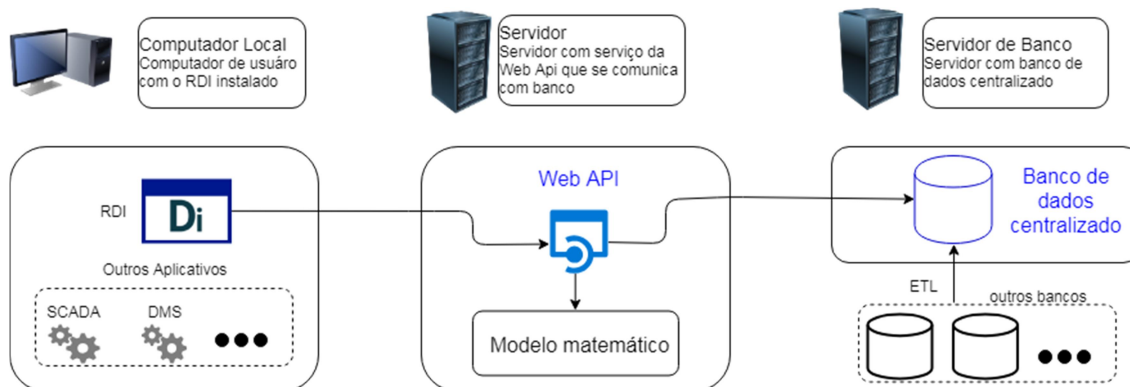


Figura 1- Comunicação do sistema

O modelo matemático usado é dividido em duas partes: uma calcula a chance de o chamado ser improcedente usando aprendizado de máquina e a outra usa o MPC para calcular o risco de não atender àquele chamado.

O algoritmo usado para calcular a chance de o chamado ser improcedente foi o algoritmo de aprendizado supervisionado *extreme gradient boosting* (Chen & Guestrin, 2016). Ele foi escolhido por ter uma alta escalabilidade mesmo com bilhões de exemplos em ambientes distribuídos e limitados por memória. Ele também é executado até dez vezes mais rápido que suas alternativas mais populares.

O módulo de MPC recebe como insumo a probabilidade de um chamado ser improcedente e a utiliza para calcular o risco de não atender o cliente. Para construir um MPC são necessários três itens: um modelo preditivo, uma função objetivo e um procedimento para obter a lei de controle. Neste trabalho, o modelo preditivo foi feito usando um processo de cadeia de Markov. Este é um processo onde o próximo estado depende apenas do estado atual, ou seja, é um processo sem memória. Desta forma a equação que descreve tal processo é:

$$P(X_t=i_t|X_0=i_0, X_1=i_1, \dots, X_{t-1}=i_{t-1})=P(X_t=i_t|X_{t-1}=i_{t-1})$$

Esse modelo de otimização pode ser calculado como uma função de custo, representando os custos de alocação e não-alocação de um recurso para uma tarefa. Custos de alocação podem ser representados, por exemplo, como gasolina e homem-hora em uma ambulância. Outro exemplo são times de manutenção técnica. Por outro lado, custos de não alocar um recurso geralmente são associados a penalidades por não se realizar determinadas tarefas. Por exemplo, em uma distribuidora de energia, agências regulatórias podem aplicar multas em caso de longos períodos de falta de energia. Outros exemplos incluem não atender determinadas casas ou hospitais onde existem pacientes que necessitam diretamente de energia elétrica para que seus aparelhos continuem a funcionar. Para definir quando um chamado deve ser classificado em “atendível” ou “possivelmente não atendível”, usamos o módulo de *machine learning* para classificar tais chamados.

O módulo de *machine learning* é composto por um algoritmo supervisionado denominado *extreme gradient boosting* (Chen and Guestrin, 2016). O algoritmo funciona baseado no conceito de árvores de decisão, onde baseado nos dados treinados, o algoritmo descobre quais são os melhores atributos para se usar na tarefa de classificação como também descobre quais padrões nos dados levam as maiores probabilidades de ser uma classe ou outra. Este algoritmo foi escolhido para este trabalho não somente por ser amplamente usado no mercado, mas também por ser altamente escalável, podendo trabalhar de forma paralelizável com bilhões de exemplos ao mesmo tempo mesmo com limitações de memória (Chen and Guestrin, 2016).

O sistema foi desenvolvido usando-se a arquitetura lambda, descrito na Figura 2. A camada *speed layer* serve para poder lidar com os dados em tempo real, ou seja, os chamados dos clientes que estão chegando naquele

momento e precisam ser classificados. Já a camada *batch layer* é responsável por atualizar o algoritmo de *machine learning* com os dados históricos dos chamados juntamente com os resultados desses chamados, respondidos pela equipe de campo. A camada *serving layer* faz a interface entre o modelo já treinado com os dados históricos e a interface, como também grava os resultados da classificação dada pelo *machine learning* e os resultados dos chamados dados pelos operadores em campo.

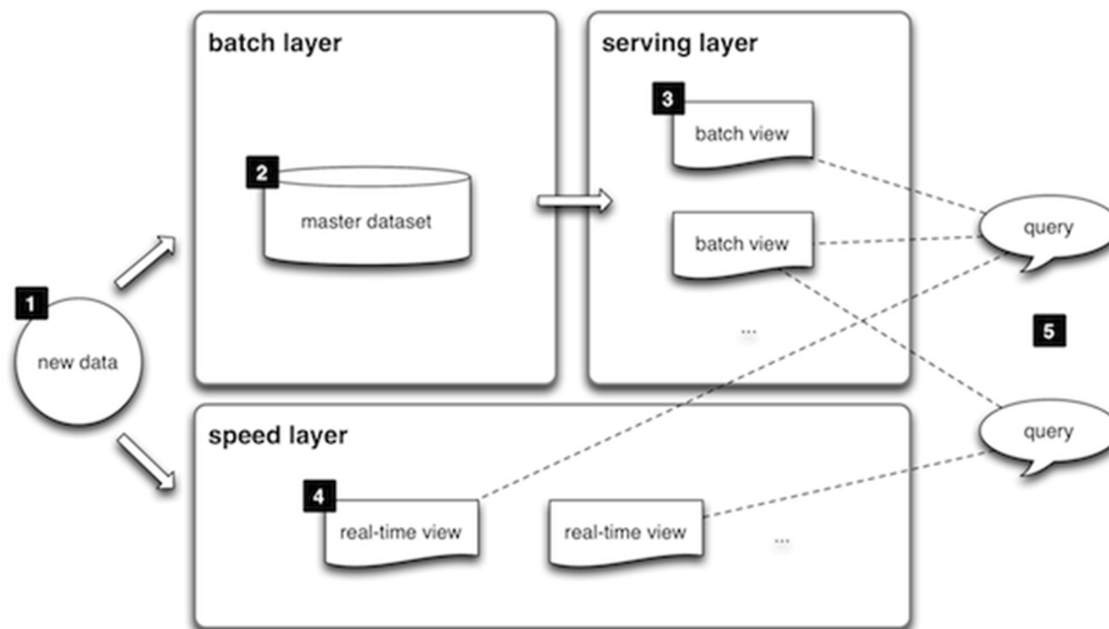


Figura 2- Arquitetura do sistema

4.0 - EXPERIMENTOS E RESULTADOS

Para validar a aplicação do RDI em um cenário real, foram usados dados de um centro de distribuição de energia da CPFL Energia. O conjunto de dados foi extraído do banco de dados do *Call Center* da CPFL Energia e foram obtidos 982.868 registros de chamados de clientes no ano de 2018. Os dados são rotulados entre procedentes e improcedentes, com desequilíbrio de classes de 89,3% procedentes e 10,7% improcedentes.

Para experimentos do cálculo de probabilidade de um chamado ser improcedente, o conjunto foi dividido em 70% para treino e 30% para teste. As métricas escolhidas para avaliar os resultados foram o macro-f1 e o coeficiente de correlação de Matthews (MCC) (Matthews, 1975).

A acurácia obtida foi de 83,93%, com MCC de 0,4607. A matriz de confusão pode ser vista na tabela 1 abaixo:

Matriz de confusão			
Classe verdadeira	Procedente	0,84	0,16
	Improcedente	0,18	0,82
		Procedente	Improcedente
		Classe prevista	

Tabela 1. Matriz de confusão

Os resultados do uso do módulo do MPC para ordenar corretamente os resultados de despacho das equipes previamente classificado pelo módulo de *machine learning* foram promissores. MPC decidiu não enviar 5 de 16 times de manutenção baseado no custo de envio dessa equipe e da incerteza da causa (procedente / improcedente). No lugar o MPC deu a prioridade desses 5 times para outros 5 eventos mais importantes (ou com risco maior no caso de não atendimento). Tal solução, diminuiu em 41,61% os custos de despacho de equipe.

Abaixo, na Figura 3, estão as telas da aplicação sobre as telas usuais dos operadores em produção. Desta forma a equipe de tomada de decisão (enviar ou não um despacho) pode consultar a ferramenta sem que haja interferência visual nas telas de outros softwares.

Na Figura 3 a) é a tela do sistema, da esquerda para a direita: Risco associado ao não atendimento do evento fornecido pelo módulo de MPC, do vermelho ao verde, ou seja, do maior para o menor risco; A coluna do meio diz respeito ao ID do evento e é feito para que a célula de callback tenha acesso aos dados do evento e possa fazer maiores análises; por último, a terceira coluna é a probabilidade do evento ser improcedente, fornecidas pelo módulo de machine learning. Na Figura 3 b), a tela do sistema sobre as telas dos operadores de callback, para poderem tomar suas decisões baseadas em suas ferramentas atuais e na ferramenta desenvolvida neste trabalho.

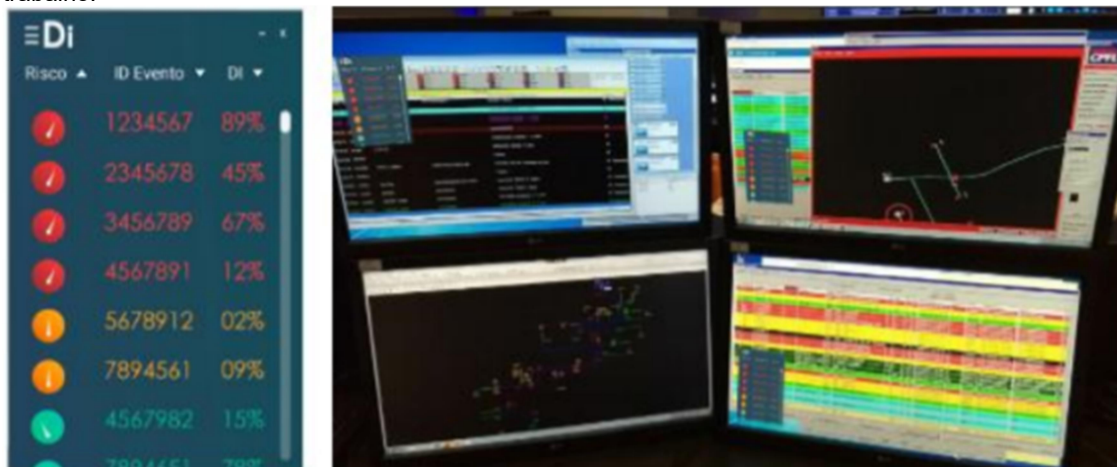


Figura 3 - a) Tela do sistema b) Tela do sistema sobre as telas da equipe de callback

5.0 - CONCLUSÃO

Neste trabalho foi proposto uma integração entre duas grandes áreas (*machine learning* e *model predictive control*) funcionando dentro de uma arquitetura lambda, no intuito de classificar os eventos (procedentes ou improcedentes) e organizar a prioridade de atendimento desses eventos. O sistema funciona recebendo os dados textuais de chamados dos clientes e analisando a probabilidade desse chamado ser ou não improcedente. A solução foi validada em ambientes de teste e também em produção dentro da empresa.

Os primeiros resultados mostraram um acerto do sistema de 84% no geral contra 47% de acerto usando-se a tradicional análise humana. Este modelo é genérico o suficiente para poder ser replicado em outras filiais da empresa ou ser usado como base para outras empresas da área de distribuição. Como trabalho futuro pode-se investir no re-aprendizado automático, sem a necessidade de re-treino (modelo semi-supervisionado adaptativo).

6.0 - REFERÊNCIAS BIBLIOGRÁFICAS

- (1) Barriquello, C. H., Garcia, V. J., Schmitz, M., Bernadon, D. P., & Fonini, J. S. (2017). A decision support system for planning and operation of maintenance and customer. Em *System Reliability*. IntechOpen.
- (2) Busemann, S., Schmeier, S., & Arens, R. G. (2000). Message classification in the call center. *Proceedings of the sixth conference on Applied natural language processing* (pp. 158-165). Association for Computational Linguistics.
- (3) Castanon, D. A., & Wholetz, J. M. (Agosto de 2009). Model Predictive Control for Stochastic Resource Allocation. *IEEE Transactions on Automatic Control*, pp. 1739-1750.
- (4) Chen, T., & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining* (pp. 785-794). ACM.
- (5) Ferrari-Trecate, G., Gallestey, E., Letizia, P., Spedicato, M., Morari, M., & Antonie, M. (2004). Modeling and control of co-generation power plants: a hybrid system approach. *IEEE Transactions on Control Systems Tchnology*, pp. 694-705.

- (6) Hearst, M. A., Dumais, S. T., Osuna, E., Platt, J., & Scholkopf, B. (1998). Support vector machines. *IEEE Intelligent Systems and their Applications*, pp. 18-28.
- (7) Maghrebi, M., Sammut, C., & Walker, T. (2013). Reconstruction of an Expert's decision making expertise in concrete dispatching by machine learning. *Journal of Civil Engineering and Architecture*.
- (8) Matthews, B. W. (1975). Comparison of the predicted and observed secondary structure of t4 page lysozyme. *Biochimica et Biophysica Acta (BBA)-Protein Structure*, (pp. 442-451).
- (9) Papadimitriou, C. H., & Tsitsiklis, J. N. (1987). The Complexity of Markov Decision Processes. *Mathematics of Operations Research*, pp. 441-450.
- (10) Roy, R. B. (2017). Equity-constrained dispatching models for emergency medical services. *Team Performance Management: An International Journal*, pp. 28-45.
- (11) Tang, M., Pellom, B., & Hacioglu, K. (2003). Call-type classification and unsupervised training for the call center domain. *2003 IEEE Workshop on Automatic Speech Recognition and Understanding (IEEE Cat. No. 03EX721)* (pp. 204-208). IEEE.
- (12) Yamato, Y., Fukumoto, Y., & Kumazaki, H. (2017). Predictive maintenance platform with sound stream analysis in edges. *Journal of Information Processing*, pp. 317-320.
- (13) Zhang, J., & Xia, X. (2011). A model predictive control approach to the periodic implementation of the solutions of the optimal dynamic resource allocation problem. *Automatica*, pp. 358-362.

7.0 - DADOS BIOGRÁFICOS



Wendell William Teixeira
Brasileiro, Solteiro, 30 anos
E-mail: wendell@cpfl.com.br
Fone: (19) 3756-8517

Formado em Engenharia Elétrica pela Universidade Estadual Paulista, mestrando em Engenharia Elétrica pela UNICAMP. Atua no setor elétrico em projetos de P&D com foco em novas tecnologias, como aplicação de Inteligência Artificial para melhorar a eficiência operacional nos processos do Grupo CPFL Energia.