

SISTEMA SUPERVISÓRIO DE MONITORAMENTO DE INSPEÇÕES E DOS INSUMOS DE CLIENTES DO GRUPO A DAS EMPRESAS CPFL PAULISTA E PIRATINGA

Alarme do grupo A

Juliano Andrade Silva, Emerson Alexandre Pizzolito – Faculdade Anhanguera de Campinas 3

RESUMO: É notado que aproximadamente 60% da energia vendida pelas concessionárias CPFL Paulista e Piratininga são provenientes do Grupo A. O programa computacional alvo deste estudo busca um sistema supervisório no que tange a gestão de inspeções nas instalações com propósito de minimizar as perdas não-técnicas, prevenindo-as e localizando-as com maior brevidade. Os alarmes reagem a ligações novas, unidades com mais de dois anos sem inspeção, alteração de equipamentos de medição, alteração de constantes, variações de consumo e fator de carga, bem como sazonalidade dos consumos dos clientes. Há uma gama de algoritmos que exploram as bases cadastrais, equipamentos e históricos de insumos das empresas envolvidas. O sistema gera uma lista para análise de unidades que possuem maior propensão de terem perdas não técnicas. O programa supervisório está em funcionamento há um ano, com três meses de desenvolvimento, se mostrando robusto, e até então, sem falhas. A modicidade de programação no software Clementine, da SPSS, e a velocidade de processamento das bases foram pontos fortes para a adequação a novas situações de contorno, que vieram a surgir no decorrer deste período de um ano de produção.

ABSTRACT: Around 60% of solded energy that CPFL Paulista and Piratininga come from A Group. The computational program, focus of this paper, developed a supervisory system to inspection management of installations with the target to reduce and to prevent non-technical losses and identify theses losses in a short term. The alarm system react to new installations, units without two years inspection, change of measurement equipments, constants changeover, consumption variations and load factor, as well client consumption variations. There is a set of algorithms that explore the registers bases, equipments and consumption database of involved companies. The system generates a list of potential non-technical losses. The supervisory program was developed in 3 months and it is working out since one year without failures. The Clementine (SPSS) software programming affordability and the bases processing capacity were strongest topics to adequate to the new boundary conditions that were identified during the workout time.

PALAVRAS-CHAVE:
Alarme, Energia, Grupo A,
Inspeção, Supervisão

KEYWORDS:
Alarm, Energy, A group,
Inspection, Supervisory

Artigo Original
Recebido em: 28/10/2010
Avaliado em: 10/10/2011
Publicado em: 28/04/2014

Publicação
Anhanguera Educacional Ltda.

Coordenação
Instituto de Pesquisas Aplicadas e
Desenvolvimento Educacional - IPADE

Correspondência
Sistema Anhanguera de
Revistas Eletrônicas - SARE
rc.ipade@anhanguera.com

1. INTRODUÇÃO

A supervisão do Grupo A é muito importante no que tange ao faturamento das empresas CPFL Paulista e Piratinga, além de ser alvo essencial para a recuperação de energia. A missão básica do processo de recuperação de energia é assegurar que o montante de aproximadamente 60% da energia vendida e concentrada no Grupo A esteja resguardada de erros, que influenciem direta ou indiretamente, na receita das empresas.

Para supervisão computacional adotou-se o aplicativo da SPSS denominado Clementine. Este trabalha com KDD (Knowledge Discovery in Data Bases) além de possuir um avançado sistema da DM (Data Mining).

Segundo Fayyad et al, (1996) o primeiro passo para uma análise em KDD é aprender aplicação sob óptica da relevância do conhecimento e dos objetivos. Por isso há uma necessidade de se estudar as características do Grupo A, bem como diferenciá-lo dos demais grupos.

Pela portaria 456 da ANEEL (Nov, 2000) o grupo A caracteriza-se, também, por possuir ligação aérea de tensão superior a 2,3KV chegando até os limites de tensões de transmissão de energia. Isto mostra a relevância destes clientes para o concessionário bem como o potencial de venda de energia.

Qualquer alteração na medição deste cliente pode provocar imensos prejuízos que afetam ambas as partes (concessionário e cliente), mútuas e excludentemente, ou seja, sempre um deles paga pelo erro, seja um acréscimo para o cliente ou um decréscimo para o concessionário.

Como amparo para esta aplicação há um banco de dados dedicado e integrado, utilizado para grandes volumes de dados e processamento, conforme Fayyad et al (1996) é chamado Data Warehouse. Há um servidor próprio para aplicação que aloja todas as bases envolvidas no processo de KDD e o software fica alojado em outro servidor.

É necessário, contudo, criar um ambiente com dados limpos, antes de se iniciar o uso dos algoritmos. A depuração segundo Sferra et al, (2003), chega a 80% do tempo de processamento e deve ser feita com imenso cuidado e conhecimento da base de dados. Alguns dados são resultados de operações entre bancos, com informações com formatos diferentes. Assim, após algumas relações de união (append) ou intersecção (merge) deve-se criticar os dados obtidos nas operações, segundo o manual User's Guide (2006), pois com fontes diferentes corre-se o risco de criar um campo com formatações múltiplas. Campos com valores missing (não encontrados, devido a operações entre bancos) ou null (nulos), devem ser tratados antes das operações realizadas pelos algoritmos.

Desta forma, criadas as condições necessárias, após as depurações das informações, o Alarme do Grupo A “vigia” as condições dos clientes: seu estado, sua classificação, as inspeções realizadas, as migrações tarifárias, as trocas de equipamentos, as alterações de constantes de medição, as variações dos consumos, as variações dos fatores de carga, variações de comportamento entre inspeções, variações de consumo fora da sazonalidade e consumos anormais comparados com históricos entre inspeções. Cada item descrito a pouco se caracteriza por um algoritmo implementado dentro do ambiente Clementine SPSS.

Existe como passos finais, uma listagem que reflete todas as inconsistências, cabendo ao analista liberá-la ou não para inspeção. Nesta etapa final tem de se imperar muito mais noções comportamentais e situações comerciais do que a mineração de dados feita até então. O chamado feeling do analista que programou o modelo é denotado nesta análise final, como complemento da inteligência já despreendida com o recurso computacional. Ao final do processo, o KDD mostra dados plenamente confiáveis, mas a decisão de negócio compete de fato ao analista de recuperação de energia.

2. DESENVOLVIMENTO

2.1. Necessidade da supervisão – entendimento da aplicação e objetivos

Através da classificação da ANEEL, o Grupo A compreende uma gama de clientes, no caso específico das CPFL Paulista e Piratininga, de 17.456 (clientes ativos). Estes por sua vez são responsáveis por 1.871 GWh/ano, em média. A grandiosidade deste número dá ideia da representatividade do Grupo A independente do total do varejo das duas empresas.

Esta energia associada deve ser resguardada e a gestão pode ser tornar tão mais complexa quanto as necessidades envolvidas. O método de gestão utilizado usa KDD como ferramenta essencial no processo.

Foi eleito como software o Clementine da SPSS, que possui uma interação com as bases SQL, além de apresentar uma linguagem de programação simples e orientada ao objeto. A figura 1 mostra a interface de programação do Clementine 12.0, que possui ferramentas de manipulação de dados e recursos de modelagem.

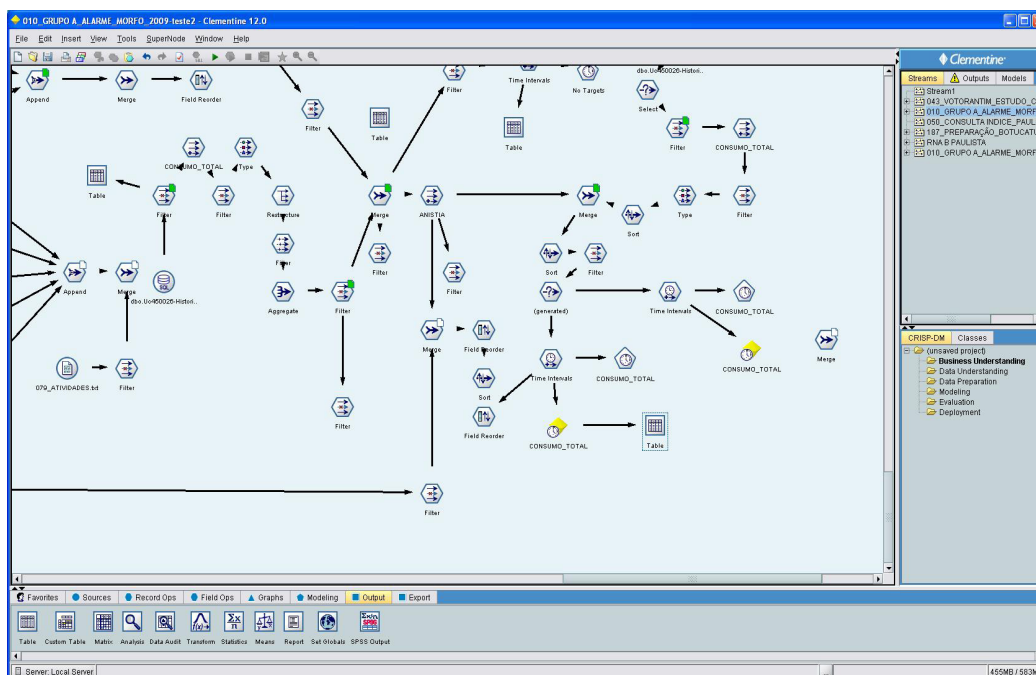


Figura 1 – Interface de programação de Clementine 12.0.

Pode-se com este recurso computacional tratar as bases disponíveis para o melhor entendimento das informações. Conforme Fayyad et al (1996), o uso do KDD é vital quando a necessidade de exploração extrapola a capacidade humana, isto não apenas pela quantidade de registros envolvidos mas pela associação de padrões que os modelos podem encontrar. A descoberta de padrões e a compreensão do significado destes é o principal mote do estudo, bem como a distinção daquelas unidades consumidoras que “fogem” da normalidade. Esta primeira etapa do KDD seria o entendimento da massa de dados, a definição do objetivo a ser atingido que é monitoramento.

Mesmo com apoio computacional seria praticamente impossível fazer o monitoramento sem o uso do método. A quantidade de informações e o cruzamento destas revelam nichos interessantes para estudo e muitas vezes destinados para inspeção.

2.2. Depuração dos dados, agrupamento organizado e pré-processamento.

Nesta etapa do alarme é feita a segregação dos dados necessários para seleção do grupo A. A extração da base cadastral UC-068 é feita diretamente selecionando um campo tipo flag, sinalizador do grupo A. Um cuidado inicial é preservar informações importantes como Classe, Situação, Região Geográfica e Data da Ligação.

Neste pré-processamento seguem-se as etapas:

1. Seleção dos clientes ativos do Grupo A.
2. Definição das datas de ligação.
3. Definição das datas de troca de equipamentos.
4. Exclusão dos Clientes Livres.
5. Exclusão de cliente que já possuem inspeções do Grupo A abertas e não concluídas.

6. Agrupamento por Município, EA, UNEG e Empresa.
7. Alocação dos históricos das quatro últimas inspeções válidas.
8. Quantificação do tempo sem inspeção.
9. Separação e criação de base de constantes.
10. Separação de clientes que historicamente apresentam sazonalidade bem definidas (escolas).
11. Separação de clientes que são tidos como imprevisíveis (Rural, agro-negócio, etc.).
12. Inserção do histórico de insumos para futuras análises.

A Data de Ligação possui uma particularidade importante que somente a inspeção de dados pode mostrar. Há dados que possuem o formato ddmmaa e outros aaaamm. Para esta correção e padronização necessitou-se criar uma regra de identificação e mudança de todos os dados para aaaamm. A data de troca dos equipamentos será fundamental para a avaliação do algoritmo de queda após troca e foi inserida no mesmo campo, seguindo como regra a prevalência da mais recente. A figura 2 mostra explicitamente uma das etapas da depuração de dados envolvendo regras no campo data de ligação.

Os clientes livres foram excluídos deste monitoramento, pois já se beneficiam de outros métodos com outras tratativas.

Agrupam-se as unidades pensando na logística de execução, já que a simples segregação nos dá informações válidas.

Clientes os quais não possuem inspeção na classificação de Grupo A, devem imediatamente serem alocados para inspeção. Há alguns critérios que exigem separação e envio imediato para inspeção:

1. Ligações novas.
2. Trocas de equipamentos por obsolescência ou defeito.
3. Aumento de carga.
4. Migração tarifária.
5. Última inspeção com prazo maior ou igual a dois anos.

Estas etapas descritas neste item 2.2 garantem que as unidades restantes não possuam óbice algum ao processamento de algoritmos para detecção de irregularidades, e podem, sem restrições, sofrerem o processo de inspeção em campo.

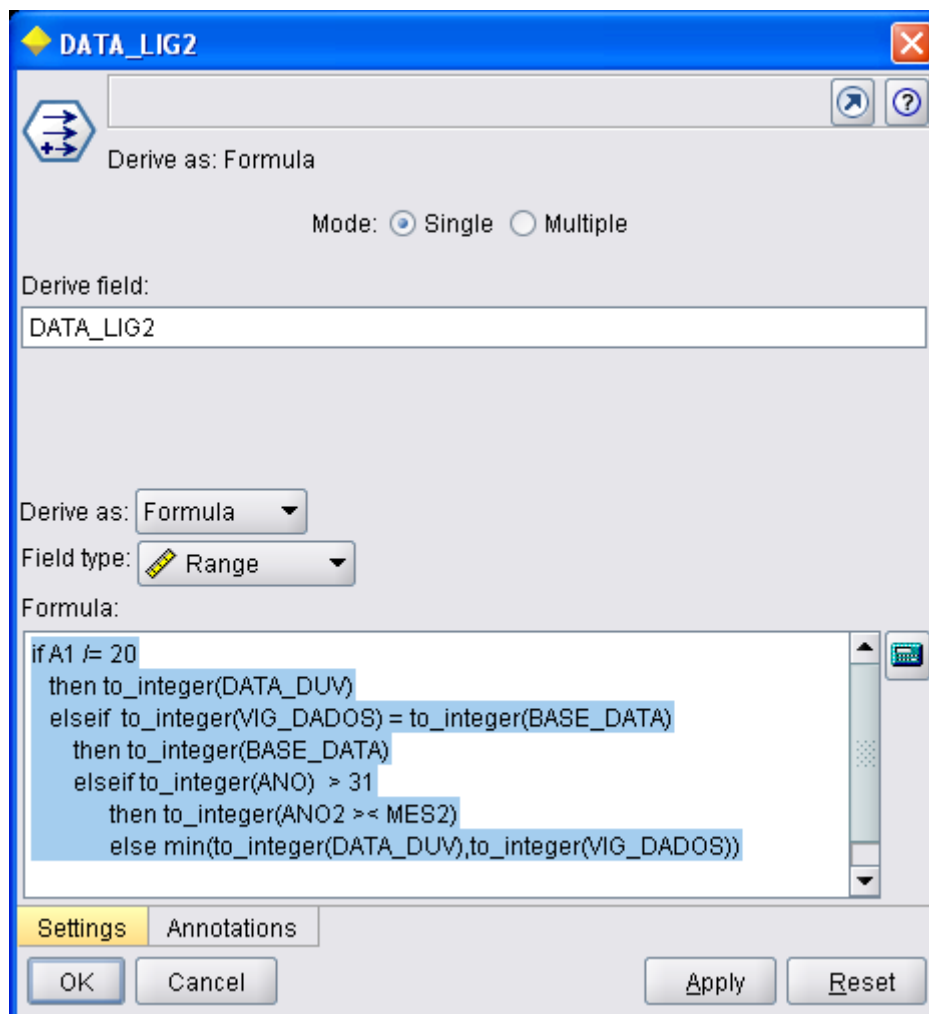


Figura 2 – Programação para depuração do campo DATA.

2.3. Processamento, algoritmos e Data-Mining

Pode-se agora, após as depurações, trabalhar com algoritmos que rastreiam anormalidades nos consumos e fatores de carga do grupo A. Cada algoritmo pontua as unidades consumidoras, que ao serem analisadas, acumulam pontos que explicitam a maior necessidade de verificação.

Algoritmo de variação de constante

É criada tabela que possui a chave primária (UC), o mês referência, a constante de medição e o consumo total daquele mês (Ponta, Fora de Ponta, Reservado e Convencional). É importante definir que o histórico está na vertical, pois desta forma, consegue-se melhor performance nas programações efetuadas.

Enquanto a chave primária permanecer a mesma, o processo identifica alterações de constante e se houve queda no consumo total referenciado a média de dois meses anteriores. Busca-se neste caso erros de ligação dos equipamentos e erro na constante da medição.

Algoritmo de comparação entre períodos de inspeção

Neste algoritmo é comparado o comportamento do cliente entre duas inspeções realizadas. Podem-se definir quatro situações:

1. Duas últimas inspeções regulares: estuda-se o período compreendido entre estas inspeções e supõe-se que aqueles insumos totais são válidos e tidos como corretos. Os insumos que não estão compreendidos neste período de tempo, após a última inspeção, não podem ser menores que o menor (queda de 20%) e nem atingirem quantidades de picos médios menores do que o período correto (20% menor da frequência).
2. Penúltima inspeção irregular e última regular: usa-se o método de queda do fator de carga e degrau de demanda. Caso haja significativa queda nestes campos a unidade ganha pontuação para o alarme.
3. Última inspeção irregular e penúltima regular: é percebido se houve aumento no consumo após a correção da irregularidade (com influência no faturamento). Além disso, há a análise dos fatores de carga, degrau de demanda, consumo mínimo inferior. O histórico é cortado após a data da regularização, isto para que não seja influenciado pelo degrau que gerou a última irregularidade.
4. Duas últimas inspeções irregulares: simplesmente a unidade deve ser inspecionada prontamente, para que se tenha uma última inspeção regular.

Algoritmo de troca de equipamento

Após a troca de equipamento é calculado o incremento. Usa-se a média do segundo mês após a troca, isso para não se usar consumos fracionados, com o terceiro mês. Estas médias são comparadas com os dois meses imediatamente anteriores a troca. (usa-se o menor deles).

Outro método usado é que o fator de carga não pode ser menor que 20% do menor dos meses anteriores. Isto demonstraria que o equipamento foi responsável pela mudança do perfil de consumo do cliente. É possível que esta mudança seja lícita, mas improvável, portanto a unidade consumidora acumula mais pontos para a seleção final.

Algoritmo de queda brusca e consumos nulos

Todas as quedas bruscas são analisadas bem como consumos que beiram zero. Há uma confrontação se este consumo é possível analisando uma região de proximidade da última inspeção. Mesmo sendo possível, a unidade recebe pontos para seleção final.

Algoritmo de anistia

Em casos de quedas bruscas, mas que por análise de proximidade sejam factíveis, há um set denominado ANISTIA, que quanto positivo indica que os consumos envolvidos na queda são cabíveis e devem ser repensados em serem inspecionados. São casos especiais que envolvem tratativas mais complexas e muitas vezes devem ser analisadas individualmente.

Algoritmo de predição de consumo

Para caso de dúvidas quanto a sazonalidade ou a veracidade do último consumo, usa-se algoritmo de Data-Mining, modelo de séries temporais, para prever o próximo consumo. Podem-se usar vários métodos de previsão: SARIMA, Exponencial Smoothing, Expert Modeler.

A figura 3 mostra a implementação da série temporal.

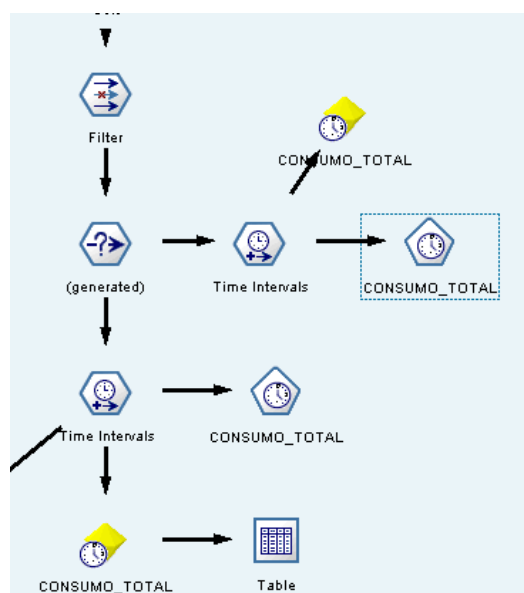


Figura 3 – Série temporal de predição de consumo.

2.4. Cômputo final

No final de todas as análises computacionais as unidades são escolhidas conforme a gravidade dos campos e pelo score final destas. Cabe agora ao analista verificar as anormalidades e encaminhá-las para seleção. É necessário ressaltar que casos especiais acontecem e feeling do analista é essencial para a validação. O conhecimento gerado destes métodos bem como a capacidade de identificar melhorias ajuda na realimentação e pode sugerir novas versões da programação. Gera-se uma listagem que é inserida nos sistema de inspeções. O processo é realimentado a partir das respostas das inspeções e dos novos históricos de consumo e demanda.

3. CONSIDERAÇÕES FINAIS

O uso do KDD é essencial para o sistema supervisorio do Grupo A. A necessidade advém da grande massa de dados envolvidas, bem como na interação destas gerando conhecimentos não triviais. É amplamente necessário que o objetivo esteja bem definido, assim como o campo de atuação e as bases envolvidas.

A depuração de dados é essencial para que se tenha veracidade nas respostas. Deve-se entender os dados, criticá-los e estressá-los para que se consiga auferir nos algoritmos o máximo possível. Na depuração têm-se várias seleções (ligações novas, trocas de equipamentos, etc.) que são essenciais para o funcionamento do alarme.

Os algoritmos por vezes abrangem variações de insumos, contudo, tratados de formas diferentes: usam estatísticas convencionais, programações lógicas, comparações e o sofisticado Data-mining. Não apenas um destes métodos, mas somente a interação completa gera a inteligência que será capaz de cercar os possíveis desvios. A robustez do método denota que muitas das possíveis situações do contorno foram sanadas, contudo a dinâmica do mercado e das empresas incentivam a uma constante realimentação das respostas em prol da melhoria dos métodos.

Cabe ao analista, a palavra final sobre o que será ou não inspecionado. É essencial que o mesmo tenha intimidade com o sistema e se lembre que os insumos refletem atividades humanas, e por vezes, nos levam a respostas intangíveis. A análise comercial final, orientada pelo KDD, mostra-se altamente eficiente, e em seis meses de produção não houve caso provado que contrariasse a eficiência do método.

O alarme, como ferramenta de análise, propiciou em seis meses um aumento no índice de acerto das inspeções dirigidas (chamadas de 'ALARME') em 23,73%, com relação ao índice das inspeções de rotina ('CRITÉRIO TEMPO'). Foram encontradas no último semestre duas ligações novas com irregularidades e vinte e cinco unidades com irregularidades por critério tempo (rotina).

REFERÊNCIAS

CLEMENTINE® 8.0 User's Guide, Copyright © 2003 by SPSS Inc. Printed in the United States of America, 2003.

FAYYAD, U.M. et al. **The KDD Process for Extracting Useful Knowledge from Volumes of Data.** In: Advances in Knowledge Discovery in Data Mining. Menlo Park: AAAI Press, 1996.

LOPES, C. H. P. **Classificação de registros em bancos de dados por evolução de regras de associação utilizando algoritmos genéticos.** Pontifícia Universidade Católica do Rio de Janeiro. Abril de 1999.

SFERRA, H. H. CORRÊA, A. M. C. J. **Conceitos e Aplicações de Data Mining.** Universidade Metodista de Piracicaba. Revista Ciência e Tecnologia, v.11.Nº 22, pp 19-34. Universidade Metodista de Piracicaba, 2003.

Resolução 456 ANELL, Novembro de 2000.