

Uso de árvores de decisão na determinação de equipamentos com falhas recorrentes

Adilson Pereira dos Santos
Companhia do Metropolitano de São Paulo - Metrô
adilson.pereira@uol.com.br

Prof. Dr. Maurício Almeida (orientador)
Centro Estadual de Ensino Tecnológico Paula Souza – CEETPS
malmeida@inteligenciaartificial.eti.br

Resumo

A análise de falhas recorrentes e as decisões tomadas para resolvê-las é de vital importância para a garantia da qualidade da manutenção realizada em equipamentos e sistemas que operam sob condições ambientais críticas. Uma das formas de determinar quais os equipamentos tem falhas recorrentes é determinar o tempo que este equipamento ficou em funcionamento entre duas falhas semelhantes. Este processo de descoberta pode ser feito manualmente, mas o uso de árvores de decisão facilitam grandemente este trabalho, oferecendo ainda maior precisão e flexibilidade. Este trabalho procurará discutir a viabilidade e os ganhos a serem obtidos com o uso de árvores de decisão para determinar equipamentos recorrentes.

Palavras chave: árvores de decisão, aprendizagem de máquina, inteligência artificial, análise de falhas.

Introdução

Manter um alto grau de disponibilidade e de confiabilidade dos equipamentos é muito importante para a garantia da qualidade dos serviços de transporte metro-ferroviários nos grandes centros urbanos.

Uma das formas de atingir estes objetivos, é garantir a qualidade da manutenção dos equipamentos através do uso de componentes sobressalentes de boa qualidade, de procedimentos adequados e de processos de monitoramento dos índices de falha e de reincidência dos equipamentos. Um equipamento é dito recorrente quando ele retorna para uma nova manutenção, antes de um determinado período de tempo, com o mesmo defeito ou sintoma. Esta tarefa de determinar se um equipamento é recorrente ou não é de grande importância para a elevação da qualidade da manutenção efetuada em equipamentos eletro-eletrônicos e mecânicos em oficinas de manutenção.

Durante o processo de análise para determinação dos equipamentos recorrentes, são analisados os períodos de tempo em que o equipamento ficou em operação e a descrições dos sintomas apresentado no sistema onde o equipamento estava instalado. Se considerarmos, o período de tempo em operação e a descrição dos sintomas, como atributos sobre os quais faremos algum tipo de teste, poderemos considera-los como os nós de uma árvore de decisão. Adicionalmente,

ser ou não recorrente correspondem a classes. Percebe-se então que o processo de análise, para decidir se o equipamento analisado é ou não recorrente, é muito semelhante ao processo de indução de árvores de decisão e aprendizagem de máquina (figura1).

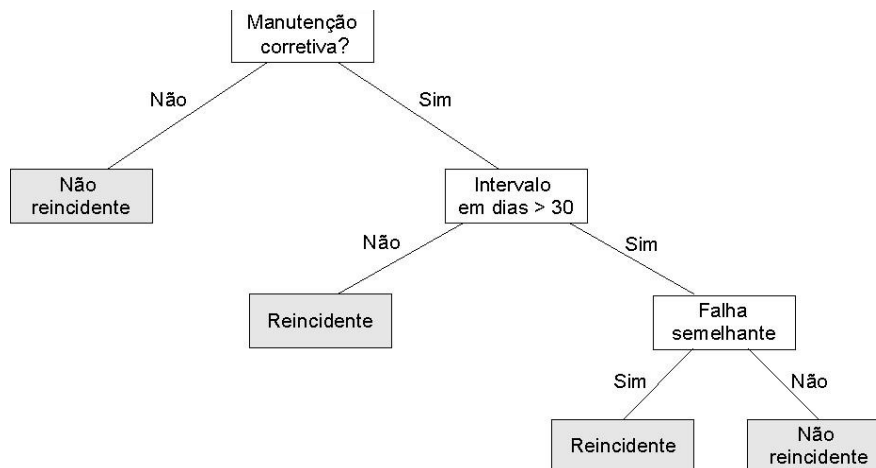


Figura 1 – Árvore de decisão para a determinação de equipamentos recorrentes

Baseado nesta semelhança, este trabalho terá por objetivo estudar a aplicabilidade de algoritmos geradores de árvore de decisão que permitam determinar, dado um conjunto de atributos, quais os equipamentos recorrentes em um dado período de tempo. Os dados históricos necessários serão obtidos diretamente do banco de dados, após um processo de limpeza, formatação e padronização para eliminar informações imprecisas, incoerentes ou incompletas, diretamente do sistema de gerenciamento informatizado que armazena estas informações das oficinas.

Metodologia

Para a determinação dos equipamentos recorrentes usando árvores de decisão, a busca será feita diretamente nas tabelas do banco de dados transacional, do sistema de gestão de informações adotado pelo departamento de oficinas de manutenção. Estas tabelas contêm milhares de registros com as informações relativas aos diversos equipamentos que passam por processos de manutenção diariamente. Nota-se que o trabalho de agrupamento destas informações poderá ter um custo de processamento computacional alto, principalmente para grandes volumes de informação, pois os valores dos atributos a serem analisados, geralmente estão dispersos no conjunto de dados aonde vão se aplicar as regras da árvore de decisão.

Felizmente os bancos de dados relacionais possuem mecanismos de indexação que permitem criar estes agrupamentos de forma rápida e confiável, deixando ao algoritmo de indução da árvore de decisão a aplicação das regras de classificação.

A fase de pré-processamento do banco de dados, assim como nos processos de descoberta de conhecimento diretamente em banco de dados [1], é importante para preparar os dados antes de submetê-los ao processo de classificação da

árvore de decisão. Este pré-processamento pode envolver fases como o entendimento, seleção, limpeza e transformação dos dados, conforme a figura 2.

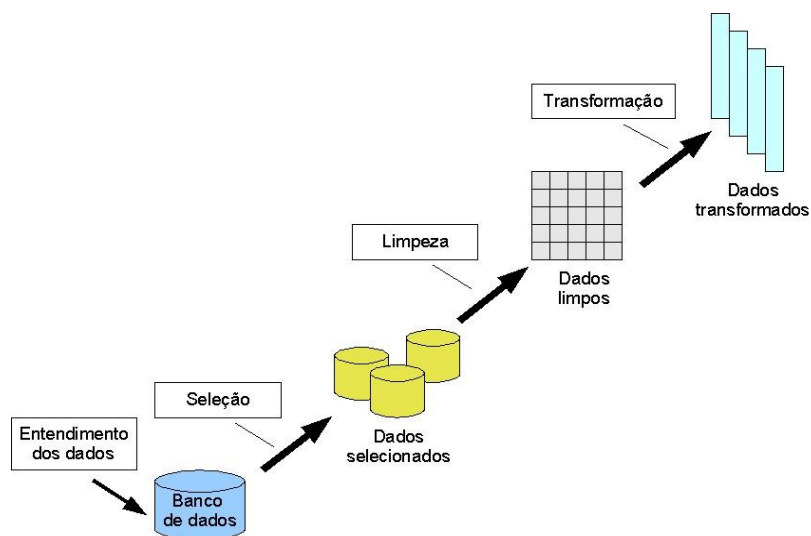


Figura 2 – Subfases do pré-processamento
Fonte:[1]NEVES, 2003. p.31

A seleção dos dados envolverá a escolha das tabelas e atributos a serem testados, podendo, dependendo da necessidade, ser preciso uma manipulação das informações contidas nas tabelas, de modo a obter-se um conjunto único de instâncias [1] sobre o qual serão aplicados os testes da árvore de decisão.

A limpeza dos dados procurará padronizar os dados, eliminar dos dados duplicados e incoerentes e os registros com instâncias de atributos ausentes ou incompletos [1].

A transformação de dados abrangerá operações que tornem os dados mais fáceis de serem testados pelo algoritmo de indução da árvore de decisão. Isto poderá envolver a discretização de valores contínuos e composição ou decomposição de atributos, dentre outras.

Esta iteração será realizada até que não existam mais novos atributos para serem testados ou que os agrupamentos não tenham mais elementos de classes diferentes (tabela 1). Vale ressaltar que, durante o processo construção da árvore de decisão a partir de um conjunto de exemplos, poderá realizar-se uma poda para melhorar o desempenho para exemplos não vistos, sem perda de informação relevante, principalmente quando o aprendizado ocorrer em exemplos contendo ruídos [3].

A escolha dos atributos a serem utilizados no agrupamento deverá levar em conta quais os principais atributos para determinar se um equipamento é reincidente ou não. De acordo com NORVIG [2], deverá ser testado primeiro o atributo mais importante - aquele que faz maior diferença para a classificação - para podermos chegar à classificação correta com o menor número de testes. Baseado nesta recomendação e tendo como exemplo a tabela 1, o atributo mais importante poderá ser o tipo de manutenção solicitada - programada ou corretiva - já que esta permite uma classificação binária, já que para estudo dos equipamentos reincidentes só interessam os que retornaram para manutenção corretiva. O segundo teste mais importante a ser realizado diz respeito ao intervalo entre falhas, cujo significado é a diferença, em número de dias, entre a data de liberação e a data de retorno à

oficina, para reparo, do equipamento sob análise. No caso em estudo, estes atributos estarão contidos em uma mesma tabela, não sendo necessário qualquer tipo de junção com as demais tabelas do sistema.

Tabela 1 – Exemplo de agrupamento dos atributos

<i>Atributos</i>			<i>Classe</i>
Tipo de manutenção	Intervalo em dias > 30	Falha semelhante	Reincidente
programada	não	sim	não
programada	não	não	não
corretiva	não	sim	sim
corretiva	não	sim	sim
corretiva	não	não	sim
corretiva	não	não	sim
corretiva	sim	sim	sim
corretiva	sim	sim	sim
corretiva	sim	não	não
corretiva	sim	não	não

Estes dois atributos permitirão, no entanto, uma classificação apenas satisfatória já que o motivo, ou a falha que provocou o retorno do equipamento não foi testado. Para que este atributo possa ser utilizado para teste faz-se necessário o uso adicional de ferramentas ou algoritmos de mineração de textos. A função destes algoritmos de mineração de textos será identificar, na descrição da falha informada pelas equipes de campo e registrada no banco de dados, termos ou palavras iguais ou semelhantes dentro das descrições atual e anteriores, e que podem indicar que um determinado equipamento retornou para reparo com a mesma falha anteriormente descrita.

Resultados

Espera-se com a utilização de árvores de decisão, a classificação de forma clara e precisa de um determinado equipamento como reincidente, dada a facilidade de interpretação de uma árvore de decisão. Além da precisão, o uso deste mecanismo permitirá um levantamento mais rápido. Hoje o levantamento dos equipamentos reincidentes é feito consultando longas listagens com as datas de entrada e as descrições das falhas ou motivos de remoção dos equipamentos, anotadas pelas equipes de campo. Isto consome tempo para ser realizado além de ser trabalhoso, o que desestimula a realização rotineira deste tipo de levantamento. Com a automatização deste procedimento, as consultas estarão disponíveis a qualquer momento para qualquer um que necessite ou tenha curiosidade sobre o desempenho de um determinado equipamento.

O uso de tabelas de um banco de dados para acessar as informações a serem analisados facilitará o trabalho de preparação dos dados, pois será possível utilizar muitos dos mecanismos disponíveis nos sistemas gerenciadores de banco de dados, SGBDs¹, para eliminação dos registros que apresentam algumas das

1 Sistema Gerenciador de Banco de Dados

características indesejáveis, como mencionadas na metodologia, na a fase de pré-processamento. O uso de tabelas de SGBDs implicará também na não necessidade de conversão dos atributos para outros formatos, como por exemplo, para texto puro, como exigido em outros algoritmos.

Para que qualquer usuário possa criar a árvore de decisão e utilizar o resultado apresentado, o desenvolvimento de uma interface simples e funcional facilitará a operação, a interpretação dos resultados, a reconfiguração da árvore e a troca dos atributos analisados e das classes. Com o uso de uma interface deste tipo poderá ser dispensável a necessidade de acompanhamento de um especialista da área de IA, o que deixará o usuário muito mais a vontade para realizar suas pesquisas e estender este trabalho para quantos equipamentos e sistemas ele achar necessário.

Esta independência de um especialista de IA para a interpretação dos resultados permitirá o uso de árvores de decisão no dia-a-dia das equipes diretamente envolvidas nas atividades de manutenção, como um auxiliar para a melhoria da qualidade dos serviço executado.

A facilidade de operação e interpretação esperada como resultado, vai permitir a identificação dos equipamentos ou sistemas que deverão ser acompanhados mais de perto. Para o caso da análise sobre equipamentos isto irá subsidiar a tomada de decisão de segregar ou não o mesmo, com o intuito de evitar ainda mais transtornos operacionais. Esta segregação poderá implicar em um tratamento individualizado durante os testes e reparo ou até mesmo no sucataamento do equipamento, se esta for a solução economicamente mais viável.

Adicionalmente, o uso da linguagem Java para a implementação irá permitir o uso da aplicação em qualquer plataforma que suporte esta linguagem.

Discussão e conclusões

Como mencionado a utilização de árvore de decisão para a determinação de equipamento com falhas recorrentes irá facilitar o trabalhos dos profissionais diretamente envolvidos na manutenção e na gestão das oficinas de manutenção da companhia. O trabalho realizado de forma manual, com base em relatórios impressos, por ser um processo cansativo e demorado, é realizado somente uma vez por mês para a atualização de um dos indicadores de qualidade dos serviços realizados nas oficinas de manutenção. Como estes indicadores, tem por finalidade parametrizar a qualidade da manutenção realizada, poderão refletir muito claramente os ganhos oriundos do uso desta técnica. Dentre os ganhos podem ser citados, o aumento da confiabilidade através do aumento do tempo médio entre falhas, também conhecido como MTBF², e a diminuição no número de viagens canceladas como consequência de falhas nos trens, durante a operação em horário comercial.

Com base nos resultados da árvore de decisão poderão ser propostas a revisão de procedimentos e processos de manutenção de forma a possibilitar a detecção do defeito causador da falha operacional. Nesta revisão poderão estar incluídas, por exemplo, modificações nos simuladores e jigas de teste, para simular ou testar condições de funcionamento não previstas até o momento, erros de especificação ou mesmo de projeto.

Para confirmar a indicação de que a solução passará por uma revisão nos

2 Do inglês: Mean Time Between Failures

procedimentos, simuladores e jigas de teste, com uma conseqüente modificação dos mesmos, serão necessários testes complementares para indicar onde e como atuar. Para subsidiar esta decisão, poderá ser adicionado como atributo para análise, se foi ou não constatado, em oficina, algum tipo de falha ou se os diagnóstico anteriores foram repetidos “nada constatado”.

Pelo exposto, espera-se que o uso de árvores de decisão para analisar os processos de manutenção, possibilite um ganho de qualidade e produtividade, não só na manutenção dos equipamento mas também na análise dos sobressalentes aplicados na manutenção dos diversos sistema existente na companhia.

Referências

- [1] NEVES, R. C. D. **Pré-processamento no processo de descoberta de conhecimento em banco de dados**. 2003. 137p. Dissertação de mestrado – Instituto de Informática, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2003.
- [2] NORVIG, P; RUSSELL, S. **Inteligência Artificial**. 2. ed. São Paulo. Campus, 2003.
- [3] RESENDE, S. O. (Org.) **Sistemas Inteligentes: fundamentos e aplicações**. São Paulo. Manole, 2005.

Contato

Adilson Pereira dos Santos
Mestrando em Tecnologia da Informação Aplicada – Centro Paula Souza
Cia. do Metropolitano de São Paulo
Rua Francisco de Paula Quintanilha Ribeiro, 134
GMT/MTO/OFJ
São Paulo - SP
telefone: (11)3444-1234
e-mail: adilson.pereira@uol.com.br