

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/267336463>

O teste de hipótese da restrição de escopo de uma Inferência de Redes Bayesianas pela utilização do GQM

Conference Paper · November 2008

DOI: 10.13140/2.1.2173.9528

CITATIONS

0

READS

63

4 authors:



Denis Montini

Instituto Tecnológico de Aeronautica

37 PUBLICATIONS **42** CITATIONS

[SEE PROFILE](#)



Paulo Tasinaffo

Instituto Tecnológico de Aeronautica

20 PUBLICATIONS **21** CITATIONS

[SEE PROFILE](#)



Luiz Alberto Vieira Dias

Instituto Tecnológico de Aeronautica

176 PUBLICATIONS **183** CITATIONS

[SEE PROFILE](#)



Adilson Marques da Cunha

Instituto Tecnológico de Aeronautica

190 PUBLICATIONS **243** CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



15th International Conference on Information Technology : New Generations (ITNG 2018) [View project](#)



SPOT: Sistema de Proteção Operacional e Tecnológica para Prevenção, Detecção e Controle de Fraudes e de Acessos Indevidos [View project](#)



O teste de hipótese da restrição de escopo de uma Inferência de Redes Bayesianas pela utilização do GQM

MSc. Denis Ávila Montini (ITA) denisavilamontini@yahoo.com.br

PhD. Paulo Marcelo Tasinaffo (ITA) tasinafo@ita.br

PhD. Luiz Alberto Vieira Dias (ITA) vdias@ita.br

Dsc. Adilson Marques da Cunha (ITA) cunha@ita.br

Resumo: Para a definição do escopo de uma rede bayesiana propôs-se a utilização de uma técnica complementar para restringir o alcance da modelagem pela utilização dos conceitos de GQM. Esta hipótese testa a possibilidade de se relacionar a causa e efeito com as probabilidades condicionais. O conceito de uma rede bayesiana pôde ser relacionado ao conceito de um caso particular de grafo o acíclico direcionado - GAD, no qual cada um dos nós representam variáveis discretas aleatórias e os arcos direcionados ilustrados pelas relações de causa e efeito entre tais variáveis. Assim, pôde-se considerar uma rede Bayesiana como um instrumento gráfico que restringe o problema que incorpora uma estrutura de dados relacionados através do aplicativo de computador Netica é um software básico para implementação de redes bayesianas.

Palavras-chave: Fábrica de Software, Células de Manufatura, GQM, Inspeção Final, Teste, Redes Bayesianas.

1. Introdução

Os elementos conceituais de **Redes Bayesianas (RB)** [NEAPOLITAN, 1990] são apresentados nesta pesquisa, de forma a aplicar a teoria em um contexto de fábrica de software, em uma das fases de teste de produção, assim como exemplificação do processo de confecção de um aplicativo de software utilizando da ferramenta Netica [NETICA, 1990]. A identificação da rede assim como os fundamentos envolvidos foram resolvidos através do emprego da técnica de Objetivos Questões e Métricas (**Goal Question and Metrics – GQM**) [BASILI, 1988].

A capacidade de representação gráfica da teoria da redes bayesianas tem sido cada vez mais explorada em estudos práticos, sendo esta área de conhecimento incerto sua utilização em diversas áreas como Engenharia de software e produção devido a elegância da solução quando codificado pela visão bayesianas. Nas atividades de manufatura humana existe a noção de previsão do comportamento da produção assim como o aplicação da probabilidade, e esta possibilidade torna possível as manipulações estatísticas. Este estudo desenvolve o assunto da probabilidade pelo teorema de bayes e um meio de definição do escopo de utilização através do uso do GQM para o entendimento das redes bayesianas.

Para se representar simbolicamente um sistema de informação sobre o conhecimento de um fenômeno [LAUNDON, 2007] [O'BRIEN, 2006], esquemas de árvores são utilizados pelas RB. A modelagem do mundo real é resolvida por processos de formalização matemática, no qual incorpora os elementos do mundo real, suas relações a fim de valer dos mecanismos matemáticos reconhecidamente verdadeiros para se estudar uma distribuição conjunta de probabilidades envolvendo variáveis aleatórias.

A natureza das variáveis selecionadas têm dependências estocásticas, organizadas por fatos e regras específicas, obtidas por meio de um estudo de sistemas especialista [STEIN, 2000], resultando em um base de dados especialista para RB. A organização deste conceitos baseiam a formação de sistemas especialistas probabilísticos no qual preconiza a



descoberta de conhecimento, contido em bases de dados classificados bayesianamente.

GQM								
Objetivo de Negócio	Assegurar que todos os defeitos são corrigidos antes do software ser liberado para uso.							
Objetivo da Medição	Remover todos os defeitos do software pelo processo de teste.							
Perguntas	7-8 Qual a numero Planejado e Realizado de requisitos testados?		5-6 Qual a cobertura Planejado e Realizado dos testes?		3-4 Quantos defeitos Unitário, Integrado, Sistema temos atualmente?		1-2 - Qual número de defeitos Sistema?	
Medidas	8 - Número de requisitos Testados Quantitativo	7 - Número de requisitos Testados Qualitativo	5 - Número de requisitos Testados Quantitativo	6 - Número de requisitos Testados Qualitativo	4 - Número de defeitos Quantitativo	3 - Número de defeitos Qualitativo	1 - Número de defeitos por status Quantitativo	2 - Número de defeitos Qualitativo

Quadro 01 – Um caso prático de aplicação da GQM.

O quadro um representa o estudo prático de um GQM, e a explicação de sua estrutura se dá da seguinte forma:

Objetivo de Negócio: Assegurar que todos os defeitos são corrigidos antes do software ser liberado para uso.

Objetivo da Medição: Remover todos os defeitos do software pelo processo de teste.

Perguntas: “7-8 Qual a numero Planejado e Realizado de requisitos testados?”, “5-6 Qual a cobertura Planejado e Realizado dos testes?”, “3-4 Quantos defeitos Unitário, Integrado, Sistema temos atualmente?” e “1-2 - Qual número de defeitos Sistema?”.

Medidas: “8 - Número de requisitos Testados Quantitativo”, “7 - Número de requisitos Testados Qualitativo”, “5 - Número de requisitos Testados Quantitativo”, “6 - Número de requisitos Testados Qualitativo”, “4 - Número de defeitos Quantitativo”, “3 - Número de defeitos Qualitativo”, “1 - Número de defeitos por status Quantitativo” e por último “2 - Número de defeitos por status Qualitativo”.

As RBs e o GQM são podem ser compostos estruturalmente por duas parte, sendo uma qualitativa e outra quantitativa [SIMÕES, 2001] [BASILI, 1988]. A organização do sistema de informações pode dividido em duas partes quantitativa e qualitativas, e com a aplicação do GQM podem ser visualizada esquematicamente o resultado da implementação deste teoria no Quadro 01. Um grafo é um modelo gráfico [BIGGS, 1976], que constitui um caso particular de grafo acíclico direcionado, GAD, [PEARL, 1988], no qual as variáveis são representadas por nós, e os arcos que os ligam significam dependências diretas entre essas variáveis. Ele é acíclico no sentido de que não há, no grafo, arcos ligando a saída de um nó à sua própria entrada. Na figura 01 Apresenta-se a resolução de um Grafo implementação com a solução de GQM do Quadro 01.

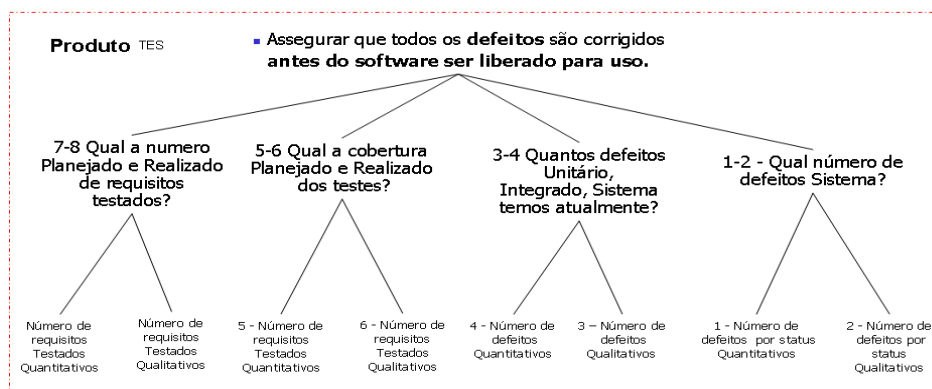


Figura 01 – Um caso prático de aplicação da GQM através da teoria dos grafos.

O trabalho busca como apresentar como as bayesianas foram propostas, definidas e construídas assim como o processo de decisão e as regras de inferências para a sua utilização. A redes bayesianas proposta é mostrada como um exemplo no diagnóstico da fase de teste de



um produto, sendo que o processo de desenvolvimento deste conceito foi resolvido com a ferramenta Netica [NETICA, 1990].

2. Metodologia

O sistema de apoio à decisão ao gerenciamento de projeto tem a intenção de ser uma ferramenta utilizada no processo de triagem à contingência a problemas no projeto, com o intuito disciplinar as operações humanas a fim de diminuir os defeitos residuais do processo resultantes da atividade de construção de software. Para a formulação de um processo de planejamento e controle eficaz, é necessário identificar as formas de raciocínio, mapas mentais, que os profissionais de desenvolvimento de software precisam ter durante a construção. Esta organização de ações e procedimento tem que visar a viabilização de conhecimentos a respeito do projeto para a tomada de decisão.

O conhecimento sobre o Planejamento e Control e Produção, (**PCP**), de projetos aplicados a linhas de produção, por exemplo celular de manufatura [MONTINI, 2005] [MONTINI, 2006a] [MONTINI, 2006b] [MONTINI, 2006c] [MONTINI, 2006d] [MONTINI, 2006e] [MONTINI, 2007] [MONTINI, 2008], estão ligados aos aspectos de conhecimento experiencial. O conhecimento necessário para o dimensionamento deve estar ligado as evidências disponíveis fundamentadas no conhecimento de engenharia de software e no conhecimento formal. Enfatizando que informações são os resultados do processamento, manipulações e organizações de dados de tal forma que represente um acréscimo ao conhecimento que o usuário final recebe [CUNHA, 2008]. O conhecimento experiencial pode ser sistematizado através de informações constantes em base de dados históricas.

O conhecimento baseado em evidências, por sua vez, é dado pelas evidências acumuladas ao longo do tempo e sumarizadas por processos estatísticos adequados. O conhecimento formal se baseia na literatura de engenharia, disponibilizada através de periódicos, livros e outras formas de armazenamento incluindo ainda processos de dedução [O'BRIEN, 2006].

Tomando como premissa a existência de uma base histórica para a calibração dos dados, contendo informações estruturadas e não estruturadas, pode-se obter, através de técnicas estatísticas, as características e a taxa de acerto dos analistas programadores [CMMI, 2006]. A restrição do escopo de acordo com os objetivos de negócio e a forma com que as atividade de programação têm seguindo as diretrizes da empresa e o contrato que firma as bases técnica do projeto, nas quais podem ser modeladas através de regras de produção GQM e submetidas à inferência Bayesianas para a formação de Sistemas de Informações Gerenciais, SIG's [O'BRIEN, 2006].

Nesta cenário no qual restringe a pesquisa operacional, os defeitos incutidos durante a fase de construção podem ser tratados através de técnicas de descobrimento de conhecimento em base de dados, Knowledge Discovery in Database (**KDD**) [SILVA, 2002], baseadas neste estudo de caso na modelagem Bayesiana [BARRETO, 2002]. A resolução desta hipótese propõem os alicerces para as técnicas de inteligência artificial para apoio à tomada de decisão [RUSSEL, 1995].

O GQM, também conhecido como sendo as regras de pesquisa qualitativa definidas por objetivos, questões e métricas, tem como finalidade modelar o raciocínio humano, para o desenvolvimento de sistemas computacionais para a tomada de decisão racional em ambientes em que haja incertezas [BASILI, 1988]. Nela, são modeladas os conjuntos de indicadores, nas quais se constituem em funções. O GQM pode comportar qualquer tipo de análise e por este motivo testou-se o uso de RB e foram mapeadas com um valor escalar entre "0" e "1", a fim de indicar o seu grau de pertinência a este conjunto. São também estabelecidas as regras de inferência, que normalmente são extraídas de especialistas humanos ou valores numéricos. A partir de sinais, sintomas e queixas apresentados pelo paciente, o processo de inferência determina como as regras serão ativadas e combinadas.



Para que se possa implementar o GQM é necessário discutir os métodos de RB empregados desenvolvendo os elementos da teoria de Bayes. Adotou-se nesta pesquisa a seguinte pré-condição que relaciona uma RB é a teoria dos Grafos restringindo os aos grafos acíclicos orientados. Este condição foi adotada pra que os nós representassem as variáveis aleatórias e o arco unindo dois nós representando a dependência probabilística entre as variáveis quantitativas e qualitativas associadas. Esta idéia é ratificada pela Figura 01.

A representação gráfica explicita as relações de dependências e constitui uma ferramenta poderosa no processo de verificação e na aquisição de conhecimentos [CMMi, 2006]. Quando uma RB está ligada à uma base de dados está pode ser construída de forma a se tornar adaptativa e se atualizar conforme as probabilidades estimadas a partir dos dados armazenados, sendo chamadas de **Redes Bayesianas Adaptativas (RBA)**. Esta capacidade oferece a possibilidade de implementar técnicas de KDD [SILVA, 2002].

Com esta composição de teorias, propõem-se que a teoria de Bayes no tratamento de múltiplas hipóteses e múltiplas evidências restritas por um delimitador organizacional. Esta abordagem é aderente ao conceito probabilístico de independência entre eventos distintos, nos quais os dados sobre um evento não traz informação adicional sobre outro evento, ou seja, dado que um evento que ocorreu, isto não traz informação para o evento 2, e vice-versa, no entanto a ocorrência dos eventual em uma estrutura conhecida corrige a probabilidade de êxito.

3.3. Noções de Probabilidade

Ao se utilizar a teoria das redes bayesianas se depara com os conceitos básicos de probabilidade. A elemento atômico da linguagem de probabilidade é o “evento”. Sendo que um “Evento” pode ser definido como um estado de alguma coisa em algum intervalo de tempo. O exemplo clássico de um dado “evento” é o lançamento de uma moeda: no qual se pode espera somente um resultado entre dois possíveis sendo um deles “Cara” e o outro “Coroa”.

A representação matemática seria “1” evento entre “2” possíveis, e é apresentado por uma fração: 1 de 2 ou 1/2 ou ainda 50 % de chance para qualquer resultado. Refletindo sobre o assunto quando ocorre um evento e, o conceito de sua probabilidade é a medida da frequência em que ocorre.

O metaconceito do modelo da moeda a probabilidade de 0.5 para o evento que a moeda irá cair cara. Porém se for conhecido que no mundo real a moeda está viciada em favor da cara, existe a necessidade de se calibrar o modelo afim de ajusta o modelo real do metamodelo teórica e redefinir a certeza e a proporção da probabilidade em favor do lado favorecido e testado no mundo real. A notação da probabilidade é descrita como:

1.	$p(e \xi)$
----	------------

A leitura de “1.”, é: A probabilidade de “e” dado “ξ”. O símbolo ξ representa o estado de conhecimento de quem provê a probabilidade.

Uma outra interpretação de probabilidade de “e” representa o grau de crença de uma pessoa que o evento e irá ocorrer em um experimento único [RUSSEL, 1995]. A interpretação de probabilidade como frequência em uma série de experimentos repetidos é referida como interpretação objetiva que calibra o modelo [CMMI, 2006].

Ao contrário, a interpretação de probabilidade como grau de crença é chamado de interpretação subjetiva ou bayesiana. Na interpretação bayesiana, a probabilidade sempre dependerá do conhecimento de quem provê a probabilidade. Em um sistema especialista real uma pessoa pode avaliar a probabilidade baseada na informação que ele assume ser verdade.

Para a implementação de sistema especialista precisou-se utilizar mais um conceito da probabilidade, este chamado de probabilidade condicional, PC. No estudo da PC para o



fenômeno do lançamento da moeda viciado, para a probabilidade de que a moeda irá mostrar cara no terceiro lançamento, corrigidas pela constatação do vício descrevê-se:

2.	$p(e2 e1, \xi)$
----	-----------------

A leitura de “2.”, é: A probabilidade do evento dois “e2” dado que o evento um “e1” e dado o conhecimento e a crença da distorção do “ ξ ”.

A outra maneira de expressar a PC é em termos de “variáveis” em vez de “eventos”. Para cada variável é atribuído um valor a partir de uma coleção de estados, na qual cada “estado” corresponde algum “evento”.

Para o estudo das variáveis, sabe-se da estatística que a variável pode ser discreta, tendo um número finito estados, ou pode ser contínua. Por exemplo, uma variável discreta pode ser usada para representar as possíveis saídas de um lançamento de uma moeda, Cara “0” ou Coroa “1”; já o peso da moeda pode ser representar uma variável contínua, Sendo aproximadamente $2,54 \pm 0,01$ gramas.

Sendo que para as variáveis discretas, podemos ter os seguintes eventos: “Cara” e “Coroa”. Para cada evento podemos ter os seguintes estados: “0” e “1”. Conhecendo as variáveis e seus estados podemos estabelecer o seu relacionamento. Para a primeira simulação definiu-se que para o evento: “Cara” = “x” atribui-se um valor inteiro “k” = “0”, “Coroa” = “x” atribui-se um valor inteiro “k” = “1”. Destas proposições pode-se afirmar a seguinte sentença:

3.	$x=k$
----	-------

A leitura de “3.” consiste em “x” = “k”, para que a variável “x” existe um estado discreto “k”.

1.1. Teorema de Bayes

O teorema de Bayes¹ atribui a parte quantitativa e qualitativa uma formação constituída pelo conjunto de probabilidades totais e condicionais, a *priori*, associadas a cada nós e arcos existentes no grafo variáveis contínuas e discretas. A utilização de cada tipo de variável é customizado de acordo com a análise do cenário e necessidade de cada nó individualmente selecionado.

Para facilitar a visualização e implicância de cada escolha do tipo de variável em cada nó dentro do modelo. Um dos grandes incentivos é a utilização do método de representação gráfica proporcionado pelas RB, no qual permite prontamente ver a repercussão da coleta da experiência física apontada pelo analista ao constatar o fenômeno e atribuir um valor “discreto” no modelo da RB. A pronta correção da probabilidade em tempo real permite a realização de simulações e análises de sensibilidade, envolvendo alocação de recurso ou suspensão de atividade e o impacto no produto decorrente desta decisão aplicada a linha de produção.

As referidas distribuições de probabilidade condicionais, que representam o conhecimento adquirido do especialista ou dos dados, devem ser registradas por meio de algum teorema ou corolário matemático. Isso tem sido obtido satisfatoriamente pelo Teorema de Bayes é a base para a parte quantitativa das RB. É importante ao usuário conhecer, ao menos em certo grau, os fundamentos desse teorema, já que isso é importante na fase de validação da RB em um processo de classificação.

A utilização do teorema “fórmula das probabilidades das causas ou dos antecedentes” descrita em “4.” segundo a literatura específica é matematicamente perfeito, porém há o revés de que com uma má avaliação dos $P(H_i)$ pode tornar sua aplicação discutível [RUSSEL, 1995].

¹ Teorema de Bayes. Esse teorema, herança devida ao estudo das probabilidades por parte do teólogo e matemático Thomas Bayes (1702 - 1761)

4.

$$P(H_i | e) = \frac{P(H_i \cap e)}{P(e)} = \frac{P(e | H_i) \cdot P(H_i)}{\sum_{j=1}^k P(e | H_j) P(H_j)}$$

A leitura de “4.” dada a equação “fórmula das probabilidades das causas ou dos antecedentes”, consiste formalizar o teorema através da atribuição de valores discretos a cada uma das variáveis. Neste caso e sendo que “i” varia de i = “1”, “2”, ..., “k”; e para as variáveis discretas “H” varia de “H” = “H 1”, “H 2”, ..., “H k” para cada nó do grafo dentro de um espaço amostral “S”.

A calibração manual da PC ocorre quando um evento “e” associado ao espaço amostral “S” a probabilidade deste caso particular constitui a hipótese “H i” causa ocorrer, dada a ocorrência do evento “e”. Para o processo de aquisição do conhecimento, os “H’s” são as hipóteses possíveis e mutuamente excludentes do estado de uma variável. Os eventos “e’s” são entendidos como as evidências provocadas pelo estado “H” da variável.

Um resultado do teste uma hipótese para o uso em RB é obtido a partir de uma suposição prévia, ou a verificação, de que dois ou mais eventos “e1”, “e2”, ..., “en” independentes. No momento em que um dado fenômeno ocorre levanta-se uma hipótese “H”, ou seja, o conhecimento sobre a ocorrência do evento “e1” não traz informação adicional sobre a ocorrência de “e2” e vice versa assume a forma seguinte:

5.

$$P(H_i | e_1 \cap \dots \cap e_n) = \frac{P(H_i) \cdot \prod_{k=1}^n P(e_k | H_i)}{P(e_1 \cap \dots \cap e_n)}$$

A leitura de “5.” dada a equação o Teorema de Bayes, sob a hipótese de evidências múltiplas e independentes.

1.1.1 Atualização Bayesiana para Novas Evidências

As RB, estão inseridas em aplicações de sistemas inteligentes e extração de conhecimento, afim de capacitar o controle e planejamento de projeto com o abastecimento de dados atualizados por evidências obtidas da realidade percebidas a partir do mundo real do projeto, corrigindo o modelo proposto *a priori* pela correção *a posteriori*.

Essa capacidade de atualização é proporcionada pelo teorema em tela e, para tal, considerando-se a referida hipótese de independência condicional, que simplifica sobre maneira a atualização bayesiana da rede, basta que seja adaptada a equação “5.”. De fato, considerando-se uma hipótese H i e denotando-se uma sequência de eventos, supostamente independentes em relação à “Hi”, por “e1”, “e2”, ..., “en” = “en”, a ocorrência de uma nova evidência “en+1” pode ser contabilizada da seguinte forma:

6.

$$P(H_i | e^n \cap e_{n+1}) = \frac{P(H_i | e^n) \cdot P(e_{n+1} | e^n \cap H_i)}{P(e_{n+1} | e^n)}$$

A leitura de “6.” dada a equação o Teorema de Bayes, sob a hipótese de evidências múltiplas e independentes para novas evidências evidências obtidas da realidade percebidas a partir do mundo real do projeto, corrigindo o modelo proposto *a priori* pela correção *a posteriori*.

A correção da fórmula descrita em “6.” mostra que ao ser calculada a probabilidade condicional da hipótese “H i”, e dado um conjunto de evidências “en”, isto é, $P(H_i | e^n)$ a probabilidade conjunta das evidências passadas $P(e_1 \cap \dots \cap e_n)$ pode ser desprezada na equação, pois $P(H_i | e^n)$ passa a representar completamente a experiência passada, e sua

atualização para um novo evento “en+1” requer apenas sua multiplicação pela

razão $\frac{P(e_{n+1} | e^* \cap H_i)}{P(e_{n+1} | e^*)}$, a qual contabiliza o impacto da nova evidência.

1.1.2 Classificação e Descrição Bayesiana

Evoluindo a utilização da proposta da aplicação de RB para *data-mining* na descrição e classificação, foi identificado que é necessário que a RB aprenda com determinadas características e parâmetros a partir dos dados, de forma a estar se apte ao mundo real para corrigir o planejamento e o controle produção, PCP, de projetos aplicados a linhas de produção a fim de produzir o conhecimento pela estruturação da em classes ou categorias da probabilidade ligados aos aspectos experimentais.

Para fazer sentido a Inteligência Artificial, IA, representada pela RB há a necessidade da aprendizagem através do levantamento das informações necessárias para uma fase estabelecida pela estruturação de um processo devidamente definido [CMMI, 2006} *a priori* associadas aos estados das variáveis aleatórias envolvidas no domínio. Na figura 01 apresentou-se uma Grafo que foi desenvolvido para atender as necessidade simultâneas da Teroria dos Grafos do GQM afim de ser pré-requisito de uma RB com obrigatoriedade de uma estrutura qualitativa e uma parte quantitativa. Este *construct* visa possibilitar a implementação do conhecimento e futuro aprendizado induzido pelos dados fenomenologicamente obtidos. Identificado o modelo pesquisou-se um ferramenta que permiti-se o teste do conceito, sendo que a ferramenta identificada foi a Netica.

O processo de aprendizado é calibrado pela ferramenta que ao se coletar um dado da prática esta tem a opção de estatisticamente poderar e corrigir o modelo pela a realizada se for requerido este fato pela analista. Para que está hipótese seja verificável foi implementado o modelo GQM do Quadro 01 na ferramenta Netica para teste o conceito do aprendizado na estrutura gráfico. O aprendizado numérico atualmente é considerado relativamente simples, enquanto que o aprendizado da estrutura gráfica, que envolve a definição e orientação das relações de dependência condicional, no qual o seu estabelecimento pode ser bastante complexo, pois se houver hipoteticamente dez variáveis em uma base de dados, e em suas relações contiver um erro pelo menos relativo ao modelo, construção, limitação da ferramenta ou pela incapacidade de representar o mundo real, a capacidade de prognóstico do modelo fica comprometida no sentido de completude.

O comprometimento do modelo até pode ser calculado e corrigido. Como estratégia de obtenção de uma RB para a pesquisa bayesiana utilizou-se a abordagem ingênua para o aprendizado da estrutura. O enfoque ingênuo assume a mencionada hipótese simplificadora de independência condicional entre as variáveis. De forma a aplicar esse conceito ao GQM, sejam “K”, “X”, “Y”, e “Z” compõem quatro variáveis aleatórias discretas. Dizemos que “K” é condicionalmente independente de “X” e “Y”, sendo que dado “Z” a distribuição das probabilidades de “K” e “X” são independentes do valor de “Y”, dado o valor de “Z”.

Esta suposição GQM implica, na prática, que os atributos para as variáveis em questão são condicionalmente independentes, dada uma determinada classe “Ci,” resultando que a probabilidade de se observar o conjunto de atributos: “e1”,..., “en” é dada pela equação “7.”:

7.

$$P(e_1 \cap \dots \cap e_n | C_i) = \prod_{k=1}^n P(e_k | C_i)$$

A leitura de “7.” Suposição de GQM aplicado a classi C_i para quatro variáveis.

Com isso, as probabilidades $P(e_1 | C_i), \dots, P(e_n | C_i)$ podem ser aprendidas facilmente a partir de uma amostra de estimação e, utilizando-se das equações “4.” e “5.”, vai-se atribuir ao caso ou indivíduo a classe de maior probabilidade posterior, o que, na prática, implica atribuir-lhe a classe de maior valor de:

8.

$$P(C_i) \cdot \prod_{k=1}^n P(e_k | C_i)$$

A leitura de “8.” Fórmula para a classe de maior valor para o GQM aplicado a classi C_i para quatro variáveis.

Estudos comparativos informados na literatura dão conta de que o classificador ingênuo, para certos domínios, é comparável em termos de resultados com as redes neurais e as árvores de decisão. Contudo, há ocasiões em que a rigidez da hipótese de independência condicional não funciona a contento, e ela tem de ser relaxada, requisitando um aprendizado generalizado, de forma a produzir resultados mais próximos da realidade.

O Treinamento de RB eleva bastante o grau de complexidade do aprendizado de estruturas, pois todas as relações de dependência/independência passíveis de ocorrerem a partir de uma distribuição de probabilidade conjunta “P”, são, a princípio, passíveis de superposição no grafo, e, portanto devem ser investigadas. Ao cair na aprendizagem generalizada identifica-se a inserção do modelo em um campo em aberto de pesquisa, no qual requer exaustivos estudos e pesquisas envolvendo elevado nível de complexidade e de processamento computacional. A partir das noções aqui apresentadas na aplicações de RB objetivando efetuar um aprendizado automático de estrutura e quantitativo, pelo processo ingênuo, a partir de uma amostra oriundo de um processo formal de uma fase de teste de um projeto com coleta de dados dos fenômenos realizados. Uma das regras mais comuns é o teorema de bayes, definido por:

9.

$$p(X | Y, \xi) = \frac{p(Y | X, \xi)}{p(Y, \xi)} p(X | \xi) \text{ , para } p(Y, \xi) > 0$$

A leitura de “9.” No qual, $p(x|\xi)$ é a distribuição da probabilidade de X antes de conhecer Y, e $p(X|Y, \xi)$ é a distribuição de probabilidade de X depois de conhecer Y.

Estas distribuições são algumas vezes chamadas respectivamente *a priori* de X e *a posteriori* de X. Este teorema auxilia na alteração na probabilidade a partir do surgimento de uma nova evidência.

2. As Redes Bayesianas restringidas pelo GQM de teste

O universo delimitado pelo GQM definiu o conjunto de variáveis discreta as serem modeladas. A RB modelada pelo GQM definiu as relações entre variáveis de interesse do estudo dentro do domínio de PCP de projetos da linhas de produção de software. A representação consiste de um conjunto de distribuições e de probabilidades locais, combinado-as em uma estrutura de árvores que permita um conjunto de asserções de independência condicional para a construção de uma distribuição global a partir das distribuições locais.

A seguir segue a delimitação das medidas a modeladas com base no Quadro 01, e pelas regras de “2.” e “3.” Retirou-se as regras de inferência para a realização do Quadro 02. Para ilustrar a representação considerar-se a o domínio de uma inspeção da fase de teste de um componente de software embarcado e de tempo real. O primeiro passo para construir uma rede bayesiana é decidir quais são as variáveis e estados do modelo. Utilizou-se a definição GQM para resolver o domínio, e o resultado da inferência estudada foi: “Assegurar que todos os defeitos são corrigidos antes do software ser liberado para uso”. O segundo passo foi para

cada uma das perguntas do GQM definidas estabelecer uma árvore em forma de grafo e aplicar os fundamentos do Teorema da Rede Bayesiana para a criação do modelo, nesta etapa foi resolvida e os resultados são apresentados no Quadro 02.

Quadro 02 – Regras de Inferências GQM - Grafo - Bayesina.

- 1-1 - Qual número de defeitos Sistema?** com os estados Péssimo, Ruim, Razoavel e Bom;
- 1-1-1 – Quantitativo: com os estados Péssimo, Ruim, Razoavel e Bom;
- 1-1-1-1 – Entradas de Dados 01: com os estados 0 to 1, 1 to 2, 3 to 4, 5 to 6, 7 to 8, 9 to 10.
- 1-1-2 – Qualitativo: com os estados Péssimo, Ruim, Razoavel e Bom;
- 1-1-2-1 – Entradas de Dados 02: com os estados 0 to 1, 1 to 2, 3 to 4, 5 to 6, 7 to 8, 9 to 10.
- 1-2 - Quantos defeitos Unitário, Integrado, Sistema tem-se atualmente?** com os estados Péssimo, Ruim, Razoavel e Bom;
- 1-2-1 – Quantitativo: com os estados Péssimo, Ruim, Razoavel e Bom;
- 1-2-1-1 – Entradas de Dados 01: com os estados 0 to 1, 1 to 2, 3 to 4, 5 to 6, 7 to 8, 9 to 10.
- 1-2-2 – Qualitativo: com os estados Péssimo, Ruim, Razoavel e Bom;
- 1-2-2-1 – Entradas de Dados 02: com os estados 0 to 1, 1 to 2, 3 to 4, 5 to 6, 7 to 8, 9 to 10.
- 1-3 - Qual a cobertura Planejado e Realizado dos testes?** com os estados Péssimo, Ruim, Razoavel e Bom;
- 1-3-1 – Quantitativo: com os estados Péssimo, Ruim, Razoavel e Bom;
- 1-3-1-1 – Entradas de Dados 01: com os estados 0 to 1, 1 to 2, 3 to 4, 5 to 6, 7 to 8, 9 to 10.
- 1-3-2 – Qualitativo: com os estados Péssimo, Ruim, Razoavel e Bom;
- 1-3-2-1 – Entradas de Dados 02: com os estados 0 to 1, 1 to 2, 3 to 4, 5 to 6, 7 to 8, 9 to 10.
- 1-4 - Qual a número Planejado e Realizado de requisitos testados?** com os estados Péssimo, Ruim, Razoavel e Bom;
- 1-4-1 – Quantitativo: com os estados Péssimo, Ruim, Razoavel e Bom;
- 1-4-1-1 – Entradas de Dados 01: com os estados 0 to 1, 1 to 2, 3 to 4, 5 to 6, 7 to 8, 9 to 10.
- 1-4-2 – Qualitativo: com os estados Péssimo, Ruim, Razoavel e Bom;
- 1-4-2-1 – Entradas de Dados 02: com os estados 0 to 1, 1 to 2, 3 to 4, 5 to 6, 7 to 8, 9 to 10.

Fonte: Autores

O terceiro passo foi o processo de implementação da modelagem para construir uma rede bayesiana através de uma ferramenta especialista, direcionada para um grafo acíclico para manter as características de inserção de valores para qualquer nó do grafo respeitando a teoria da independência condicional. Ao implementar todas estas restrições temos uma estrutura da rede bayesiana para novas evidências equação “6.”. O passo final foi definir as distribuições locais uma para cada estado. O resultado pode ser visto na figura:

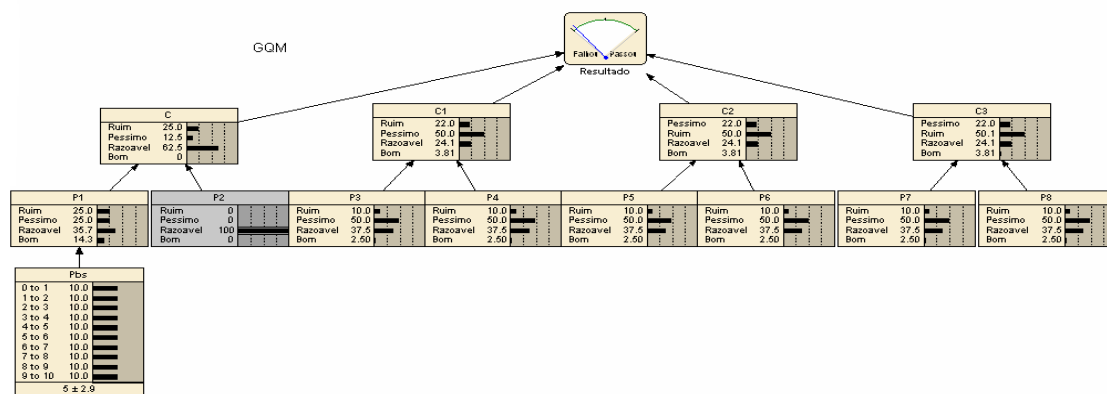


Figura 02 - Exemplo de Redes Bayesianas. Fonte: Autores

3. Aplicações

A aplicação desta RB foi modelada para as necessidades de controle e planejamento de produção de software para a aquisição de conhecimento do comportamento da linha de produção; e a consequência natural deste fato é o seu melhor dimensionamento das atividades nos próximos processos. As aquisições destes conhecimentos têm uma gama aplicação, como: aplicado engenharia de produção de software, gerenciamento de software, pontos de função, etc.

A aplicação resolveu a questão de aderência de relacionamento entre dos indicadores de GQM com os grafos acíclicos da rede bayesina para determinar um indicador de tomada de decisão sobre a qualidade do produtos testado durante a fase de homologação na fabricação de software. O estudo de caso de um componente de software embarcado e de tempo real

produzido em C++, no qual tinha a suspeita de se ter defeitos após a fase de testes. A técnica então foi utilizada na fase de teste de aceite do usuário, *Users Accept and Test – UAT*.

O software Netica também possibilitou a implementação da teoria da decisão, com isso inclui nas redes bayesianas os nós de decisão, ou seja as ações a serem tomadas, com isso foi possível determinar se o planejamento e a exceção está dentro de um padrão aceitável. O modelo derivado de GQM inclui mais de 70 casos de testes distribuídos nos 15 nós bayesianos. A ferramenta possibilita a inserção um dados ou variável discreta em um nó específico quando um fenômeno ocorre.

4. Resultados obtidos

4.1. Fase Inicial

Da fase inicial foi feito um estudo teórico sobre GQM e redes bayesianas para entender seus conceitos básicos e também a sua aplicação prática, principalmente no campo da confiabilidade. Após feito esse estudo teórico, foi realizado então uma análise das ferramentas existentes para a inferência de redes bayesianas, que permitissem o teste da hipótese da restrição do escopo pela inferência da GQM.

4.2. Resultado

O resultado da pesquisa foi a calibração da tabela com as inferência de probabilidade da redes bayesianas na linha de produção pelo software Netica. A ferramenta tem um grau de usabilidade que permite a visualização tanto da rede através do grafo assomo como também da tabela de probabilidade e decisão. Com relação as inferências são calculadas rapidamente pelo departamento de engenharia de produção da empresa estudada [TCS, 2003].

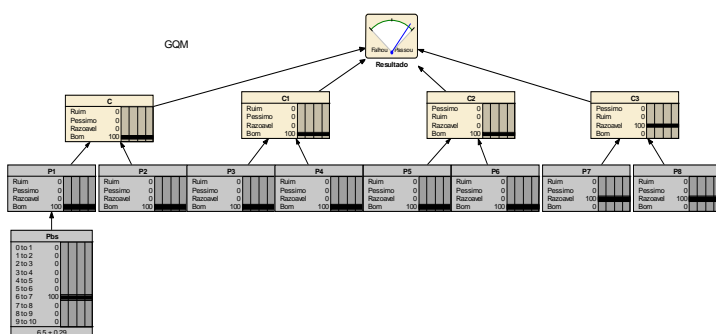


Figura 03. Interface com os Resultados do Netica. Fonte Autores

P1	P2	Ruim	Pessimo	Razoavel	Bom
Ruim	Ruim	100,00	0,000	0,000	0,000
Ruim	Pessimo	75,000	25,000	0,000	0,000
Ruim	Razoavel	75,000	0,000	25,000	0,000
Ruim	Bom	75,000	0,000	25,000	0,000
Pessimo	Ruim	75,000	25,000	0,000	0,000
Pessimo	Pessimo	100,00	0,000	0,000	0,000
Pessimo	Razoavel	25,000	50,000	25,000	0,000
Pessimo	Bom	15,000	75,000	10,000	0,000
Razoavel	Ruim	0,000	0,000	100,00	0,000
Razoavel	Pessimo	0,000	75,000	25,000	0,000
Razoavel	Razoavel	0,000	0,000	100,00	0,000
Razoavel	Bom	0,000	0,000	100,00	0,000
Bom	Ruim	75,000	0,000	25,000	0,000
Bom	Pessimo	0,000	25,000	75,000	0,000
Bom	Razoavel	0,000	0,000	100,00	0,000
Bom	Bom	0,000	0,000	0,000	100,00

Figura 04. Tela com a tabela de definição das Probabilidades Condicionais de um Nó. Fonte Autores

Estes modelos foram submetidos as mesmas evidências e foram constatadas pelo processo de Verificação e Validação [CMMi, 2006]. Logicamente o modelo tem uma forma de lidar com o conhecimento de forma distinta, porém os resultados obtidos foram, de certa forma, bastante aproximados. As evidências que obteve-se os resultados respectivos de forma que foram encontrados “3” três erros em testes unitários com alta severidade no produto final, e “0” zero erros nos teste de sistemas e “0” zero erros na fase de teste integrados em dois ciclos de teste de homologação. Este artigo apresentou a implementação de uma RB baseada em um processo especialista para sistema de tempo real.

Na Figura 05 descrevê-se os valores linguísticos apresentados como evidência, e após o processo de inferência GQM com a rede bayesiana o resultado obtido foi “3 “ erros com severidade alta em “70” casos possíveis sendo uma taxa de erro com um indicador de risco de 4,29 % possibilitando a intervenção e correção dos erros encontrados.

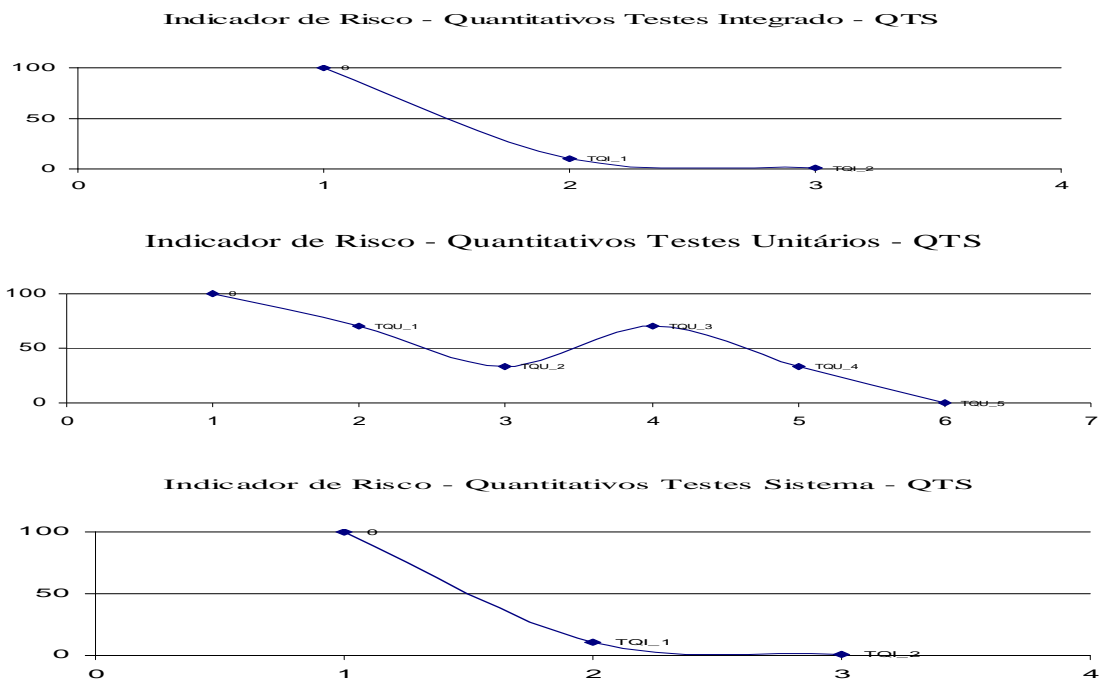


Figura 05 – Evolução da probabilidade ao longo do tempo, nas respectivas dimensões de testes: Unitário, Integrado e de Sistema. Fonte Autores

O Figura 05 apresenta o controle da probabilidade de se encontrar defeitos no componente de software de computador desenvolvido em laboratório, com o monitoramento do teste e de da probabilidade de encontrar erro ao longo do tempo na fase de UAT. Os dados forma coletados de hora em hora através das atividade de inspeções de qualidade, sendo que estas atividades são rotineiras [CMMI, 2006]. A tendência da probabilidade de se encontrar erro tem que tender a “0” zero até a fase de entrega e homologação a ser realizada pelo cliente. Foi constatada está tendência com coleta de dados, assim como os defeitos encontrados ao longo da fase de teste de aceite de usuário e na fase de *postmortem*.

5. Discussão e Conclusões

Os resultados, apesar de preliminares, remetem a uma corroboração entre o planejamento e o realizados pelas técnicas apresentadas por Vasili, o GQM, e por Bayes, a RB, compondo assim através destas técnica de múltiplas hipóteses e múltiplas evidências. Algumas divergências ocorreram nas fase de construção, provavelmente causadas pela falta de refinamento dos modelos para reproduzirem a mesma realidade.

Caso considere-se que o modelo no qual grande parte dos erros de testes estejam acertadamente representadas, um sistema de apoio a decisão de projeto de software baseado em GQM e Bayes poderia ser uma eficiente ferramenta no processo verificação e validação da fase de teste de contrução de software. A fundamentação deste sistema de controle viabilizará a construção de sistemas especialista de IA, pois constituem uma estrutura estável de reproduzível de conhecimento sobre o comportamenta de uma linha de produção de célula de manufatura. A utilização de ferramenta especialistas como o Netica facilita a implementação da teoria bayesina, e trouxe a possibilidade de entrar em contato com conceitos não muito vistos em computação, principalmente conceitos de definição de critérios integrados a gestão como o GQM aplicados a questões de confiabilidade de teste com a teoria de grafos em redes bayesianas de evidências múltiplas e independentes.

6. Referências

- [BARRETO, 2002] BARRETO, Alexandre; VIEIRA, Renato C.; NASSAR, Silvia M. Redes Bayesianas e Produção de Conhecimento: uma abordagem de data-mining em dados de um concurso vestibular. In: 35º



- Reunião Regional da ABE/SOBRAPO, Florianópolis-SC, 2003 e V Escola Regional da Sociedade Brasileira de Computação (SBC) - Centro-Oeste, 2002.
- [BASILI, 1988] BASILI and ROMBACH'S, Goal-Question-Metrics Paradigm, IEEE Transactions on Software Engineering, 1988.
- [BIGGS, 1976] N. L. Biggs, E. K. Lloyd, and R. J. Wilson. Graph Theory 1736–1936.[História da Teoria dos Grafos, com reprodução de artigos clássicos]. Claredon Press, Oxford, 1976.
- [CMMI, 2006] CMMI® for Development (CMMI-DEV) Version 1.2, CMU/SEI-2006-TR-008, Carnegie Mellon University. Disponível no site: <http://www.sei.cmu.edu/publicatons/documents /06.reports /06tr008.html>, Acessado em Agosto de 2006.
- [CUNHA, 2008] CUNHA A.M, “CE-245 Tecnologias da Informação”, ITA – Instituto de Tecnologia da Aeronáutica. Disponível em : <http://www.ita.br/~cunha>, Acesso em: Maio de 2008.
- [LAUNDON, 2007] LAUNDON Kenneth C. & Jane P. Laudon. “Sistemas de Informação Gerenciais” Pearson Prentice Hall – 7ª Edição, 2007.
- [MONTINI, 2005] Montini, Denis Ávila, Modelo de indicadores de risco para o orçamento de componentes de software para célula de manufatura. / Denis Ávila Montini.360p.Dissertação (Mestrado) em Engenharia de Produção – Universidade Paulista - (2005). Brasileira. Simpósio, Hotel Serra Azul - Gramado, RS, 2005.
- [MONTINI, 2006a] Montini, Denis Ávila; Albuquerque, Antonio Roberto Pereira Leite de; Nascimento, Marcos Ribeiro Do. Strategy for the use of the model of personal software process for the establishment of measurement and analysis for obtain CMMI level 2 in a study of case of a brazilian software factory. In: ASEE - 5th ASEE Global Colloquium on Engineering Education, Rio de Janeiro, 2006.
- [MONTINI, 2006b] Montini, B Denis Ávila; Spinola, Mauro de Mesquita; Sacomano, José Benedito; Nascimento, Marcos Ribeiro do; Battaglia, Danilo. Application of model PSP manual and supported by tool in a study of case of brazilian plant of software. Revista produção on line, Florianópolis - SC - Brasil, 2006.
- [MONTINI, 2006c] Montini, Denis Ávila; Nascimento, Marcos Ribeiro do; Battaglia, Danilo; Andrade, Dalton Cernach. Metodologia de Gestão de Vulnerabilidades para Fábrica de Software. In: WCCSETE'2006 – World Congress On Computer Science, Engineering and Technology Education, 2006, Itanhaém, São Paulo, Brasil, 2006.
- [MONTINI, 2006d] Montini, Denis Ávila; Sekhar, Chandra; Negry, Tatiana Tavares; Nascimento, Marcos; Battaglia, Danilo. Strategy for the use of the model of personal software process for the establishment of measurement and analysis for obtain CMMI level 2 in a study of case of a brazilian software factory. In: Montivideu: TACTICS Iberoamerica 2006.
- [MONTINI, 2007] Montini, Denis Ávila; Moreira, Gabriel de Souza; Vieira, Luiz Alberto; Battaglia, Danilo; Gnatiuc, Carlos Eduardo; Cunha, Adilson Marques da; In: TACTICS Iberoamerica 2007. Estudo de caso de uma estratégia de integração de middleware para um serviço SOA de gerenciamento e controle de fábrica de software: TCS – Tata Consultancy Services – Intranet website de Base de Conhecimento Corporativa KnowMax: TACTICS Iberoamerica 2007, Brasil, São Paulo, SP 25-26/Outubro/2007.
- [MONTINI, 2008] Montini, Denis Ávila; Moreira, Gabriel de Souza; Vieira, Luiz Alberto; Battaglia, Danilo; Gnatiuc, Carlos Eduardo; Cunha, Adilson Marques da; In: Global TACTiCS 4th Global Conference, Study of case of a strategy of middleware integration for SOA service of administration and control of factory of software: TCS – Tata Consultancy Services – Intranet website Corporative Knowledge Data Base KnowMax: Global TACTiCS 4th Global Conference, Índia, Kerala, Thiruvananthapuram. 16-19/janeiro/2008.
- [NEAPOLITAN, 1990] Richard. Probabilistic Reasoning in Expert Systems. John Wiley & Sons, Inc. New York, 1990.
- [NETICA, 1990] Netica 4.02, Norsys and Netica are trademark of Norsys Software Corp. Copyright © 1990-2007.
- [O'BRIEN, 2006] O'BRIEN James. “Sistemas de Informação”. Editora Saraiva – 2ª Edição, 2006.
- [RUSSEL, 1995] RUSSEL, Stuart; NORVIG, Peter. Artificial Intelligence - A Modern Approach. Prentice-Hall, 1995.
- [SILVA, 2002] SILVA, W. T., e LADEIRA, M. Mineração de Dados com Redes Bayesianas. XXI JAI – Jornada de Atualização em Informática. Anais do XXII Congresso da Sociedade Brasileira de Computação, vol. 2, pg. 235-286, 2002.
- [SIMÕES,2001] SIMÕES, Priscyla Waleska Targino de Azevedo; NASSAR, Silvia Modesto; PIRES, Maria Marlene de Souza. Sistema de Apoio na Avaliação da Falência do Crescimento Infantil. In: CONGRESSO BRASILEIRO DE COMPUTAÇÃO, Workshop de Informática Aplicada à Saúde, 2001.
- [STEIN,2000] STEIN, Carlos Efraim. Sistema Especialista Probabilístico: Base de Conhecimento Dinâmica. 2000. Dissertação (Mestrado em Ciências da Computação) - Programa de Pós-Graduação em Ciências da Computação, UFSC, Florianópolis.
- [TCS, 2003] Tata Consultancy Services, transnacional Indiana no segmento de consultoria e Fábrica de desenvolvimtno de software e Serviços com sede na Índia e atuação no Brasil desde de 2003.
- [PEARL, 1988] PEARL, J.- Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. Morgan Kaufmann. 2ª Edição. California - 1988.