

# Sistema de análise de dados em Mastologia

Anita Maria da Rocha Fernandes<sup>1</sup>, Benjamin Grando Moreira<sup>1,2</sup>

<sup>1</sup>Laboratório de inteligência aplicada - CTTMar - Universidade do Vale do Itajaí  
(UNIVALI), Brazil  
anita.fernandes@univali.br

<sup>2</sup>Instituto de Engenharia Biomédica - CTC - Universidade Federal de Santa Catarina  
(UFSC), Brasil  
benjamin@ieb.ufsc.br

## Resumo

O câncer de mama no Brasil cresceu não somente no número de novos casos quanto no número de óbitos. Tornando necessário aprimorar os métodos de análise de dados, para elucidar as causas da origem e/ou do tempo de sobrevivência livre de doença. Neste trabalho, apresenta-se um sistema de análise de dados para câncer de mama que utiliza técnicas estatísticas e de mineração de dados, construído com uma interface web, utilizando a linguagem PHP; o qual armazena as informações em um banco de dados ORACLE. O sistema permite ao especialista através de formulários dinâmicos, a criação e alteração dos dados de entrada. Os dados usados neste trabalho são provenientes do ambulatório da Universidade do Vale do Itajaí.

Palavras-chave: Informática Médica, Mineração de dados

## Abstract

Breast Cancer in Brazil increased not only in new cases but in new dead. So, it is necessary to improve data analysis methods, in order to find the causes that origin and/or the time of life without that illness. This paper presents a data analysis system for breast cancer that uses statistical methods and data mining method, developed with a web interface, using PHP language, which storage information in ORACLE database. This system allows to experts, through dynamic data forms, to create and to change input data. Data used in this work were obtained from Universidade do Vale do Itajaí.

Key-words: Multimedia Systems, Computers Networks, Rules (at least three words, not exceeding one line).

## 1. Introdução

A incidência de patologias mamárias, frequentemente malignas e com altas taxas de mortalidade, levou à Medicina a desenvolver estudos mais detalhados para um melhor conhecimento das alterações apresentadas pelas glândulas mamárias. O consequente aperfeiçoamento das técnicas de tratamento conduziu o desafio para o surgimento de uma especialidade médica: a mastologia [1].

A especialidade de mastologia é uma área ampla e abrangente das glândulas mamárias, entretanto sendo o câncer de mama uma doença que atinge um grande número da população, ela é identificada com o estudo desta doença. A Organização Mundial da Saúde estima que, por ano, ocorram mais de 1.050.000 casos novos de câncer de mama em todo o mundo, o que o torna o câncer mais comum entre as mulheres [2].

Segundo o Instituto do Câncer [3], o número esperado de casos novos de câncer de mama para o Brasil em 2006 é de 48.930, com um risco estimado de

52 casos a cada 100 mil mulheres. Em Santa Catarina, o número de casos novos está estimado em 1.610, fazendo parte dos estados do Brasil com a maior taxa bruta estimada de incidência no intervalo de 45,38 a 96,95 por 100.000 mulheres.

O câncer de mama permanece como o segundo tipo de câncer mais freqüente no mundo e o primeiro entre as mulheres. Alguns fatores de risco, comuns a outros tipos de câncer, também estão associados ao câncer de mama como a obesidade pós-menopausa e exposição à radiação ionizante. Porém, a prevenção primária deste câncer ainda não é totalmente possível, pois fatores de risco ligados à vida reprodutiva da mulher e características genéticas estão envolvidos em sua etiologia. Estudos têm sido realizados no sentido de validar novas estratégias de rastreamento factíveis para países com dificuldades orçamentárias, já que o único método de detecção precoce que, até o momento, mostrou reduzir a mortalidade, foi o rastreamento populacional com mamografia para mulheres com idade entre 50 e 69 anos [3].

Apesar de ser considerado um câncer de relativamente bom prognóstico, se diagnosticado e tratado oportunamente, as taxas de mortalidade por câncer de mama continuam elevadas no Brasil, muito provavelmente porque a doença ainda seja diagnosticada em estádios avançados [3].

Assim, o câncer de mama tem duas questões a serem abordadas, a primeira diz respeito à determinação das causas que a originam e a outra se refere à determinação das formas de tratamento para incrementar o tempo de sobrevivência. Entretanto, até hoje os métodos para análise dos dados desta doença tem sido tratados através da epidemiologia utilizando hipótese estatística. Tal modalidade de análise, segundo [4], oferece menos introspecção em dados epidemiológicos do que métodos alternativos que enfatizam a estimação de medidas interpretáveis.

Nesse contexto, a modalidade de análise alternativa pode ser a Descoberta de Conhecimento em Banco de Dados (DCBD) mais conhecida por Knowledge Discovery in Database (KDD). Para a área médica, [5] diz que a DCBD pode fornecer conhecimento novo, como por exemplo, a relação entre algumas doenças e certos perfis profissionais, socioculturais, hábitos pessoais e local de moradia. Essas relações são utilizadas para melhor entendimento das doenças e seus tratamentos.

## 2. Descoberta de Conhecimento

DCBD mais conhecida por Knowledge Discovery in Database (KDD) a qual é uma linha de pesquisa da inteligência artificial cuja metodologia permite analisar grandes bases de dados para extrair o conhecimento implícito contido neles. Esta característica o torna uma ferramenta de suporte para análise de dados em medicina, permitindo examinar as relações de similaridade entre os dados de vários pacientes portadores de uma mesma doença, podendo apontar mais informações sobre as suas causas, a sua forma de disseminação e os métodos de prevenção.

A DCBD não parte de um modelo da população, mas, pretende encontrar o modelo que a população em estudo apresenta, assim sendo, neste trabalho se propõe um sistema que contém tanto métodos estatísticos quanto de mineração de dados para a análise de dados do câncer de mama.

A DCBD foi desenvolvida com a finalidade de encontrar informações úteis em grandes bases de dados, as quais não estão visíveis ao ser humano. [6]. É uma ferramenta que possibilita a análise de grandes conjuntos de base de dados, utilizando métodos aproximados, a qual é uma característica intrínseca desta teoria. Isto permite flexibilidade no raciocínio, a qual é desejável na análise de dados clínicos, os quais quase sempre são imprecisos e incertos.

A metodologia da DCBD baseia-se em dois pilares fundamentais, o armazém de dados e a mineração de dados. De um lado é necessário e fundamental criar um armazém de dados de forma organizada e com suficientes dados sobre o assunto a

analisar e de outro é importante contar com métodos que permitam minerar os dados, assim organizados, para a descoberta das relações existentes em tais dados.

O Processo de DCBD é feito pela definição e domínio do problema que se deseja resolver, fazendo-se uma análise das atividades, com a finalidade de atingir os objetivos propostos para garantir que os resultados obtidos no processo não serão inúteis. O analista de mineração deve estar junto ao especialista da área, tendo como finalidade especificar o problema da melhor forma possível, para chegar-se a resultados positivos e úteis.

O processo de DCBD é um processo de natureza iterativa e interativa cujos passos básicos são [6]:

- Desenvolver um entendimento do domínio da aplicação: o conhecimento prévio e as metas do usuário final;
- Criação de um conjunto de dados e de pré-processamento: operações básicas, tais como remoção de ruído, coleta de informações necessárias ao modelo ou quantidade de ruído, decisão sobre estratégias para manipulação de informações inexistentes;
- Redução e projeção dos dados: encontrar características úteis para representar os dados dependendo da meta da tarefa. Deve-se utilizar redução de dimensionamento ou métodos de transformação para reduzir o número efetivo de variáveis em consideração ou encontrar representações variantes para os dados;
- Mineração de Dados (MD): Escolha da tarefa, decidir se a meta do processo é classificação, associação ou agrupamento; implementação do algoritmo de mineração para a busca dos padrões de interesse;
- Interpretação dos padrões encontrados: se necessário retorno a qualquer um dos passos anteriores para melhorar a busca e;
- Consolidação da descoberta de conhecimento: incorporação desse conhecimento ao desenvolvimento do sistema, ou, simplesmente, documentá-lo e relatá-lo às partes de interesse.

A MD, na DCBD, é encarregada dos padrões que são extraídos e enumerados a partir dos dados. A descoberta de conhecimento envolve a avaliação e a interpretação dos padrões para se decidir o que constitui do que não constitui conhecimento. Isto, também, inclui a escolha de esquemas adequados, de pré-processamento, de amostragem e de projeções dos dados antes da enumeração [7].

As técnicas e ferramentas de MD permitem analisar a base de dados (BD) e descobrir novo conhecimento que se encontra implícito nela. A MD pode tornar visíveis informações valiosas contidas em um BD, ajudando na análise ou na tomada de decisão ou na validação do conhecimento. Por esse motivo o assunto sobre MD será melhor aprofundado.

### 3. Mineração de dados

MD é o termo utilizado para nomear o processo de análise de conjuntos de dados com o objetivo de encontrar padrões que representem informações úteis e não triviais. Para tanto, utiliza-se de métodos matemáticos, heurísticas e algoritmos. A MD é parte de um processo maior e mais abrangente, o de descoberta de conhecimento em bancos de dados [8].

A MD é considerada a principal etapa do processo de DCBD e são diversas as técnicas utilizadas nesse processo a fim de realizar as tarefas de DCBD. Dentre as principais técnicas estão: Algoritmos Genéticos; Redes Neurais; Árvores de decisão; Estatística; e Raciocínio Baseado em Casos.

[5], define a MD “como o uso de técnicas automáticas de exploração de grandes quantidades de dados de forma a descobrir novos padrões e relações que, devido ao volume de dados, não seriam facilmente descobertos a olho nu pelo ser humano”.

Segundo [9], existem vários algoritmos de MD utilizados para resolver problemas específicos. Os principais são:

- Associação: achar associações em que a presença de um conjunto de itens em uma transação implica em outros itens;
- Classificação ou geração de perfis: desenvolvem perfis de diferentes grupos;
- Padrões sequenciais: identificam tipos de padrões sequenciais em restrições mínimas especificadas pelo usuário; e
- Agrupamento (ou clusterização): segmentam o banco de dados em subconjuntos ou grupos.

Para esse trabalho, o algoritmo de MD utilizado foi o de regras de associação que, para [10], procuram identificar afinidades entre registros de um subconjunto de dados. Essas afinidades/associações são expressas na forma de regras.

Associações são relacionamentos significativos entre itens de dados armazenados e procura encontrar tendências que possam ser usadas para entender e explorar padrões de comportamento dos dados [11].

### 4. Sobre o sistema desenvolvido

O sistema foi desenvolvido no intuito de auxiliar a análise de dados na área de mastologia, utilizando cálculos estatísticos e mineração de dados, pois os cálculos estatísticos são utilizados pelos especialistas da área e a mineração de dados busca descobrir alguma informação intrínseca nos dados.

Os dados foram coletados pelo serviço de mastologia de uma Unidade de Saúde da Família. A coleta de dados foi realizada através do preenchimento de um formulário, em papel, desenvolvido pelo serviço de mastologia e posteriormente digitado no Microsoft Excel. Tais dados foram importados do Excel através

de um aplicativo desenvolvido exclusivamente para esse fim e após um processo de limpeza e transformação foram incorporados à base de dados.

O sistema desenvolvido buscou unificar algumas funções, que eram utilizados pelo serviço de Mastologia, do EpiInfo e do Excel, evitando assim o trabalho de repassar dados de uma ferramenta para outra, o qual pode estar sujeito a erros humanos. A DCBD foi uma opção acrescentada para familiarizar os profissionais da área com essa forma de análise de dados.

Tanto análise quantitativa como qualitativa foram implementadas nos sistema. Ofuncionamento é simples, o sistema seleciona os dados que podem ser analisados bastando o usuário marcar o dado para análise e qual estatística pretende obter. A Figura 1 mostra uma tela para obter estatística de dados quantitativos.

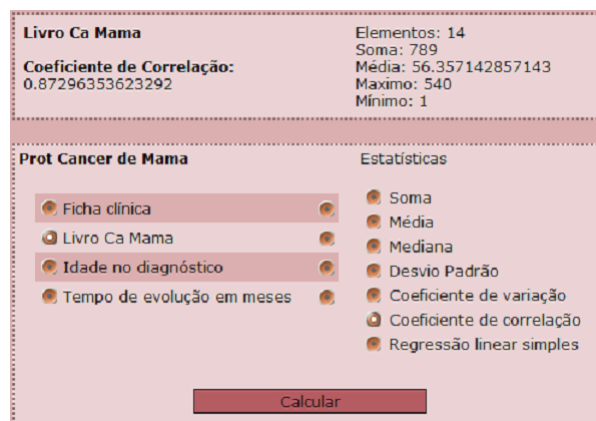


Figura 1 – Tela para obtenção de estatísticas

Para o projeto optou-se por utilizar os cálculos de soma, media, mediana, desviopadrão, coeficiente de variação, coeficiente de correlação e regressão linear simples para análise de dados quantitativos. A abordagem qualitativa foi uma opção proposta com o intuito de ser avaliada sua utilização e para tanto foram implementados os cálculos de Soma, Média, Mediana e Coeficiente de correlação.

A partir do momento que a base de dados foi povoada (150 registros de atendimento), a MD pode ser utilizada. Para tanto, uma série de itens candidatos são geradas e o usuário pode selecionar (através da Figura 2) quais são de interesse de pesquisa. Quanto menor o número de itens candidatos selecionadas, menor o tempo de resposta da aplicação.

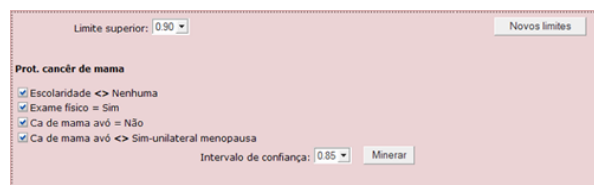


Figura 2 – Seleção de itens candidatos, limites de suporte e intervalos de confiança

O usuário também pode refinar a geração de itens candidatos, especificando maiores limites de

suporte. Também é possível refinar o intervalo de confiança das regras geradas.

Estando definidos os parâmetros desejados, são geradas e mostradas regras descobertas (Figura 3). Estando as regras geradas, o usuário pode analisá-las apenas lendo as associações geradas, ou usar um link formado em cada regra. Toda regra gerada é um link que, ao ser clicando, abre uma página do sistema onde todos cálculos efetuados pelo sistema, bem como gráficos gerados são mostrados. A aplicação reconhece os dados que geram a regra e realiza os cálculos estatísticos e geração de gráficos apenas para esses dados. Com isso o especialista tem a disposição diversas maneiras de avaliar a regra criada. Essa forma de visualizar os dados de uma regra, também podem ser utilizados como um relatório sobre a regra.

Prot Cancer de Mama	
SE (Escolaridade = Nenhuma) ENTÃO (Dor = Não). Conf = 0.86	
SE (Escolaridade = Nenhuma) ENTÃO (Ulceração na mamografia sem outros sinais ou sintomas = Não). Conf = 0.86	
SE (Dor = Não) ENTÃO (Ulceração na mamografia sem outros sinais ou sintomas = Não). Conf = 0.86	

Figura 3 – Regra geradas pela MD

Dentre os cálculos estatísticos realizados estão: soma, média, mediana, desvio padrão, coeficiente de variação, coeficiente de correlação e regressão linear simples, para dados quantitativos, e soma, média, mediana e coeficiente de correlação para dados qualitativos.

#### 4.1. Formulários dinâmicos

Buscando uma melhor adaptabilidade, o sistema possui interface dinâmica, que pode ser configurada de acordo com a necessidade do especialista. Sendo assim, caso se deseje adicionar ou remover algum novo campo de informação, essa tarefa é realizada facilmente pelo próprio especialista, não precisando uma modificação no sistema por parte do programador.

Os formulários dinâmicos facilitaram a criação dos formulários eletrônicos baseados nos formulários em papel, possibilitando uma maior flexibilidade do sistema e a rápida alteração dos formulários. Os formulários dinâmicos independem da especialidade com a qual será utilizado, portanto eles poderão vir a ser utilizados para as outras especialidades do ambulatório. A Figura 4 ilustra uma tela para criação de formulário dinâmicos.

The interface shows a list of fields to be added to a form, each with an 'Excluir' button. The fields are: Ficha clínica, Livro Ca Mama, Idade no diagnóstico, Data do diagnóstico, Data 1ª consulta, Escolaridade (set to 'Nenhuma'), Nódulo (set to 'Não'), and Tempo de evolução em meses. A legend on the right lists the data types: List, Checkbox, Radio Button, Text, and Data. A 'Confirmar criação do Formulário' button is at the bottom.

Figura 4 – Criação de formulários dinâmicos

Os nomes que identificam os tipos de dados (*list*, *checkbox*, *radio button* e *text*) são dados pelo tipo de dados do formulário HTML. Data é um campo que não existe nos formulários HTML. Para o tipo 'data' utilizou-se o campo 'text' e a inserção é controlada com uma função JavaScript.

A partir das tabelas que armazenam a estrutura dos formulários é gerada a base de dados onde serão armazenados os registros. Para prover um maior desempenho no momento de minerar os dados, os dados são inseridos na base transformados. Dados do tipo list, checkbox e radio button muitas vezes são compostos por palavras, mas na base de dados eles são transformados em números antes da inserção. O campo data é transformado para data Juliana sendo armazenado como dado numérico.

#### 5. Metodologia

Buscando uma melhor adaptabilidade, o sistema possui a geração de formulários de forma dinâmica, em que os formulários podem ser configurados de acordo com a necessidade do especialista. Sendo assim, caso se deseje adicionar ou remover algum novo campo de informação, essa tarefa é realizada facilmente pelo próprio especialista, não precisando uma modificação no sistema por parte do programador. A tela de criação dos formulários é exemplificada pela Figura 5.

This is a duplicate of the screenshot in Figure 4, showing the 'Prot Cancer de Mama' form creation interface with various input fields and a legend.

Figura 5 – Tela de criação dos formulários

Os formulários dinâmicos facilitaram a criação dos formulários eletrônicos baseados nos formulários

em papel, possibilitando uma maior flexibilidade do sistema e a rápida alteração dos formulários. Novos formulários também podem ser criados não necessitando da intervenção de um programador para manipulação de formulários.

O banco de dados utilizado é o Oracle devido ao fato dos sistemas sendo desenvolvidos para o ambulatório estarem utilizando esse BD, mantendo-se assim a homogeneidade de tecnologias. O mesmo motivo se deu na escolha da linguagem de programação PHP. O sistema ainda utilizará a linguagem JavaScript para interação com o usuário, por essa ser a linguagem de programação comum nos navegadores de internet e por isso sua escolha se fundamenta na maior compatibilidade dos navegadores com o sistema proposto.

Embora o banco de dados utilizado seja o Oracle, o sistema trabalha com uma classe genérica de comunicação com bancos de dados, isso permite que outros bancos de dados possam ser utilizados sem grandes alterações no sistema. ADOdb é uma classe aberta de comunicação com o banco de dados. Ela suporta os principais bancos de dados: MySQL, PostgreSQL, Interbase, Firebird, Informix, Oracle, MSSQL, Foxpro, Access, ADO, Sybase, FrontBase, DB2, SQLite, LDAP e ODBTP. Com a utilização dessa classe, busca-se uma maior adaptabilidade do sistema a outras instituições sem maiores complicações.

Outra classe utilizada no trabalho foi a PHPlot, para geração de gráficos. A versão utilizada dessa classe suporta os gráficos de barra, linha, área e ponto. Uma mostra de um gráfico de barra gerado é vista na Figura 6.

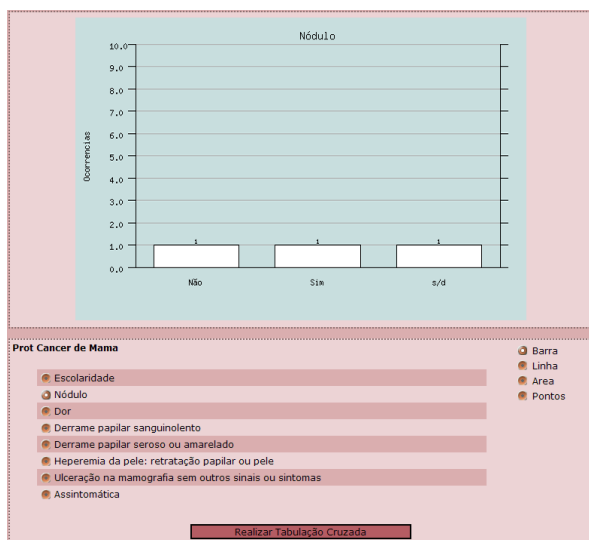


Figura 6 – Tela de geração de gráficos

Para a geração de gráficos, somente dados que forem qualitativos estão disponíveis para utilização. Basta o especialista escolher qual dado deseja analisar e com qual tipo de gráfico.

## 5.1. O algoritmo de regras de associação

O sistema cria regras de associação existente nos dados armazenadas a partir dos formulários cadastrados. O primeiro passo nessa técnica consiste em encontrar os itens que possam virar regras (itens candidatos). O algoritmo é responsável por descobrir o conjunto de itens com maior frequência na base de dados.

A segunda etapa do algoritmo é responsável por realizar o cruzamento dos itens selecionados na primeira etapa. Através dos cruzamentos espera-se que regras possam se formadas. Um cruzamento gera regra quando o resultado do cruzamento é superior ao do fator de confiança pedido solicitado pelo especialista. Nessa etapa, o algoritmo é responsável por realizar o cruzamento dos itens de maior frequência e gerar uma regra do tipo “se (condição) então (causa) com intervalo de confiança de valor x.” (regra do tipo 1).

Também é possível descobrir regras do tipo “se (condição1) e (condição2) então (causa) com intervalo de confiança de valor x.” (regra do tipo 2). Para criação de regras desse tipo, alguns controles adicionais foram necessários para evitar que dados que já haviam se cruzados voltassem a se cruzar gerando processamento desnecessário.

Regras do tipo 1 são denominadas, por [12], de regras de associação transacionais, enquanto as que envolvem múltiplos atributos (regra do tipo 2) recebem uma denominação de regra de associação multidimensional.

[12] também introduzem o conceito de regra de associação híbrida que é um tipo especial de regra multidimensional onde uma das dimensões pode aparecer repetidas vezes no "corpo" da regra. Esse tipo de regra também é possível de ser gerada pelo algoritmo, embora nenhuma tenha sido gerada.

## 6. Melhorias propostas

Buscando melhorar a utilização e manutenção do sistema, estão sendo avaliadas melhorias em 3 aspectos: bla bla bla, balbalbla

### 6.1. Melhorias na criação dos formulários

A criação de formulários dinâmicos embora esteja funcional, pode ser melhorada, aumentando o controle sobre os formulários. Dentre as melhorias sendo avaliadas estão:

- Possibilitar a troca da seqüência que o formulário segue. Atualmente a seqüência seguida é a da ordem de entrada dos dados;
- Possibilitar a modificação dos valores de entrada dos dados. Atualmente ao criar uma list, checkbox ou radio button, não é possível alterar, incluir e excluir opções, necessitando o dado ser excluído do formulário e reentrado; e
- Criar divisões dentro de um mesmo formulário. Isso possibilitaria agrupar dados

de um formulário, aumentando a facilidade na visualização dos dados.

Diversas outras funcionalidades são elaboradas para a criação de formulários dinâmico, mas cada nova funcionalidade aumenta a complexidade na criação do formulário. Como o interesse é possibilitar ao especialista criar seus formulários, as novas funcionalidades devem ser criteriosamente selecionadas para não desmotivar o especialista na criação de formulários.

## 6.2. Editor de cálculos estatísticos

O editor de cálculos estatísticos possibilitará maior independência do sistema. Novos cálculos estatísticos poderão ser adicionados sem a necessidade de modificações no sistema e sem a necessidade de presença de um programador para criação de novas estatísticas, podendo o especialista criá-las no momento que necessitar.

Os códigos para realização dos cálculos ficarão armazenados no banco de dados e executados assim que forem chamados. A tarefa de criar a forma como os cálculos serão realizados é complicada e o editor deve facilitar o máximo possível sua criação.

## 6.3. Novos algoritmos de mineração

As próximas tarefas e serem implementadas serão as de classificação e agrupamento. Essas tarefas podem retornar outras relações nos dados que podem se mostrar melhor dos que as apresentadas pela associação.

A classificação cria automaticamente um modelo a partir de um conjunto inicial (treinamento) de registros. O modelo é composto de padrões, essencialmente generalizações em relação aos registros, os quais são usados para diferenciar as classes. Uma vez obtido o modelo, este é usado para classificar automaticamente os demais registros [11].

Agrupamento ou Clustering é uma tarefa que tem como objetivo segmentar os dados formando grupos homogêneos. Para [12], ela é utilizada para a identificação de grupos de registros relacionados, os quais podem representar classes potenciais, que podem ser usadas como ponto de partida para a exploração de outros relacionamentos. Para [13], agrupamento é uma tarefa descritiva comum, onde o objetivo é identificar um conjunto finito de categoriais ou agrupamentos para descrever os dados.

## 7. Discussão e Conclusões

Dos diversos sistemas de análise de dados na área da saúde, a maioria realiza cálculos estatísticos e possuem o inconveniente de serem complicados de se utilizar. O sistema proposto e implementado neste trabalho apresenta uma maior facilidade na inserção dos dados no sistema; efetua cálculos estatísticos e inclui a

possibilidade de análise com a utilização de mineração de dados.

Embora não existam particularidades em desenvolver uma MD para a saúde, esse pode pedir que dados possuam mais segurança, exigindo a utilização de criptografia, mas essa opção é mais recomendada quando o DW é utilizado para prontuário eletrônico. Por se tratar de um sistema que realiza apenas análise estatística e não é utilizado como prontuário eletrônico, o sistema não se preocupou com a segurança de acesso aos dados, apenas na confiabilidade nos cálculos efetuados.

Os formulários dinâmicos facilitaram a criação dos formulários eletrônicos baseados nos formulários em papel, possibilitando uma maior flexibilidade do sistema e a rápida alteração dos formulários. Os formulários dinâmicos independem da especialidade com a qual será utilizado, portanto eles poderão ser utilizados para as outras especialidades do ambulatório.

A principal dificuldade do trabalho deu-se no desenvolvimento do formulário dinâmico. Dado que estes formulários precisavam de uma estrutura que permitisse visualizar os dados de forma clara e ao mesmo tempo gerasse um DW dentro da padronização.

O algoritmo de mineração implementado utiliza regras de associação. O desempenho do algoritmo é inversamente proporcional ao número de requisições realizadas ao BD. Assim, procurou-se minimizar esses acessos melhorando o código e procurando condições de diminuir o número de testes na procura de informação.

De modo geral, pode-se dizer que o sistema revoluciona o processo de análise de dados no serviço de Mastologia do ambulatório onde foi implantado, já que a análise se restringia apenas à realização de cálculos estatísticos a partir de alguma hipótese levantada. Espera-se que o sistema seja aceito e que os especialistas da Mastologia incorporem a MD como ferramenta de análise de dados.

As principais recomendações para aprimoramento deste sistema referem-se aos formulários e às ferramentas de análise de dados. Embora a criação de formulários dinâmicos e as ferramentas de análise sejam funcionais e de bom desempenho respectivamente, poderiam ser acrescentadas melhoras, tais como: possibilitar a troca da seqüência que o formulário segue. Atualmente a seqüência seguida é a da ordem de entrada dos dados; possibilitar a modificação dos valores de entrada dos dados.

Atualmente ao criar uma list, checkbox ou radio button, não é possível alterar, incluir e excluir opções, necessitando o dado ser excluído do formulário e reentrado; criar divisões dentro de um mesmo formulário. Isso possibilitaria agrupar dados de um formulário, aumentando a facilidade na visualização dos dados; implementar cálculos estatísticos que permitam ao usuário a avaliação de outros indicadores.

Também é interessante implementar outros algoritmos de mineração e avaliar o desempenho para a escolha do mais representativo. Mesmo o algoritmo de

regras de associação sendo adequado a necessidade, a possibilidade de criar outros formulários até mesmo para outras áreas, faz com que as regras de associação não sejam aplicadas a análises dessas outras áreas.

Embora nenhuma relação significativa nova tenha sido encontrada, esse resultado já era esperado já que o formulário implementado já tinha sido analisado diversas vezes. Sendo assim somente foi a utilização funcionou apenas para análise estatística dos dados. Mas como o sistema permite a criação de outros formulários, esperasse um novos dados para pesquisa.

## Referências

- [1] Montoro, A. F.(1984). “Mastologia”. Editora Sarvier.
- [2] Ministério da Saúde, Secretaria de Atenção à Saúde – Instituto Nacional de Câncer. “Controle de câncer de mama: Documento de consenso”, [<http://www.inca.gov.br/publicacoes/Consensointeg ra.pdf>], fev 2006.
- [3] INCA (2006) Ministério da Saúde – Secretaria de Atenção à saúde - Instituto Nacional do Câncer – Coordenação de Prevenção e Vigilância. “Estimativa 2006: Incidência de câncer no Brasil”, Rio de Janeiro: INCA, 2005.
- [4] Rotman, K.J. (1998). “Modern Epidemiology”. Maple Press.
- [5] Carvalho, L. A. V. (2001). “Datamining: a Mineração de Dados no Marketing, Medicina, Economia, Engenharia e Administração”. Editora Érica, São Paulo, Brasil.
- [6] Fayyad, U. M. et al. (1996) “From data mining to knowledge discovery: an overview”. In: *Advances in Knowledge discovery and data mining*. Menlo Park.
- [7] Siebes, A. Homogeneous contain no surprises: Inferring Risk-profiles from Large Databases. Amsterdam: 1994. p. 11.
- [8] Pitoni R. M. Mineração de regras de associação nos canais de informação do direto. Trabalho de Conclusão (Graduação)–Curso de ciência da computação, Instituto de informática, Universidade Federal do Rio Grande do Sul, Porto Alegre, 1998.
- [9] Kurike, R. Z.; Dias, L. R, “Grupo de Sistemas Inteligentes - Mineração de Dados. Departamento de Informática, Universidade Estadual de Maringá”, [[http://www.din.uem.br/ia/a\\_multid/mineracao/](http://www.din.uem.br/ia/a_multid/mineracao/)]. jun 2004.
- [10] Baptista J., Carvalho D. R. Data mining como apoio à decisão em projetos públicos. In: CONGRESSO BRASILEIRO DE COMPUTAÇÃO, 3., 2003, Itajaí. Anais... Itajaí: Univali-CTTMar, 2003.
- [11] Campos, M. L.; Rocha Filho A. V. Data warehouse. In: MEDEIROS, C. M. B. (Ed.). JORNADA DE ATUALIZAÇÃO EM INFORMÁTICA, 16., 1997, Brasília. Anais... Brasília: UnB, 1997. p. 221-262.
- [12] Han, J. e Kamber, M. (2001). *DataMining: Concepts and Techniques*. Morgan Kaufmann Publishers.
- [13] Ávila B. C. “Data mining”. In: Escola de informática da SBC-regional sul, 6., 1998, Blumenau. *Anais...* Curitiba: Champagnat, 1998. p.87-106.