

**OpMM - FERRAMENTA PARA EXTRAÇÃO DE CONHECIMENTO
UTILIZANDO PRIMITIVAS TEMPORAIS: UM ESTUDO DE CASO EM
PACIENTES INTERNADOS EM UTI**

Karla Haryanna Santos Moura

Mestrado em Ciência da Computação - Associação Ampla UERN/UFERSA, 2013.
Especialista pela Faculdade de Natal em Gerência de Projetos em Tecnologia da
Informação. Graduada em Ciência da Computação pela Universidade do Estado do Rio
Grande do Norte, 2007. Email: karlaharyanna@gmail.com

Maximiliano Araújo da Silva Lopes

Mestrado em Ciência da Computação – Associação ampla UERN/UFERSA, 2011.
Especialista em Informática Aplicada pela UERN. Graduado em Ciência da Computação
pela UFRN. Email: profmaxlopes@gmail.com

Ana Maria Guimarães Guerreiro

Mestrado em Engenharia Elétrica pela Universidade Federal do Rio Grande do Norte,
1999. Doutorado em Engenharia Elétrica pela University Of Colorado at Colorado Springs,
2004. Graduada em Engenharia Eletrica pela Universidade Federal do Rio Grande do
Norte, 1997. E-mail: amggguerreiro@gmail.com

Cicília Raquel Maia Leite

Mestrado em Engenharia Elétrica pela Universidade Federal de Campina Grande, 2005.
Doutorado em Engenharia Elétrica e de Computação pela Universidade Federal do Rio
Grande do Norte, 2011. Pós-doutorado no Massachusetts Institute of Technology (MIT) no
Laboratory of Computational Physiology, Harvard-MIT Division of Health Science and
Technology. Email: ciciliam Maia@gmail.com

RESUMO

A quantidade de informações armazenadas em banco de dados está crescendo em uma velocidade maior que a capacidade computacional e humana de interpretar esses dados. Logo, à medida que o uso de sistemas de banco de dados cresce, os usuários exigem funcionalidades adicionais, com a finalidade de facilitar a implementação de aplicações de usuário mais avançadas e mais complexas. A busca por informações contidas em bancos de dados tem acontecido de forma ampla e generalizada e, mais especificamente nos banco de dados médicos, foco deste trabalho, tem sido alvo de muitas pesquisas, como: mineração de dados, segurança, ética e/ou sigilo dos dados, modelo de predição, sistemas de apoio a decisão, utilização de técnicas inteligentes, entre outros. Diante do exposto, o objetivo deste trabalho é o desenvolvimento de uma ferramenta para extração de conhecimento utilizando primitivas temporais, intitulada OpMM (*Open Medical Mining*), baseado nas etapas do *Knowledge Discovery Database* (KDD), a fim de subsidiar profissionais da área médica para auxílio ao seu diagnóstico. A ferramenta está baseada em técnicas e algoritmos de mineração de dados, através do algoritmo de árvore de decisão. O OpMM tem como princípio a transformação dos dados em informações, por meio do conhecimento dos especialistas e dos parâmetros de normalidade dos sinais vitais de pacientes. Para validação da ferramenta utilizou-se um estudo de caso em pacientes internados em UTI através do banco dados público *Multiparameter Intelligent Monitoring in Intensive Clinical* (MIMIC II).

PALAVRAS-CHAVE: Mineração de dados, Unidade de Terapia Intensiva, Descoberta de Conhecimento em Banco de Dados, Sinais Vitais.

**OpMM - TOOL PRIMITIVE EXTRACTION USING TEMPORAL KNOWLEDGE:
A CASE STUDY IN PATIENTS ADMITTED IN ICU**

RESUMO

The amount of information stored in database is growing at a speed greater than the computational and human capacity to interpret these data. Thus, as the use of database systems grows users require additional functionality in order to facilitate the implementation of more advanced and more complex user applications. The search for information contained in databases has been broadly and widespread and, more specifically medical database, focus of this work has been the subject of much research, such as: data mining, security, ethics and/or data confidentiality, prediction model, decision support systems, intelligent techniques among others. Based on the considerations above, the objective of this work is to develop a tool for extracting knowledge using primitive temporal titled OpMM (Open Medical Mining), based on the steps of the Knowledge Discovery Database (KDD), assisting medical professionals to aid medical diagnosis. The tool is based on techniques and data mining algorithms, through the decision tree algorithm. The OpMM has as principle the transformation of data into information through the of the experts' knowledge and normality parameters of the patient vital signs. Case study in patients admitted to the ICU through public data bank Multiparameter Intelligent Monitoring in Intensive Clinical (MIMIC II) was used to validate the tool.

PALAVRAS-CHAVE: Data Mining, Intensive Care Unit, Knowledge Discovery in Database, Vital Signs.

**OpMM - FERRAMENTA PARA EXTRAÇÃO DE CONHECIMENTO
UTILIZANDO PRIMITIVAS TEMPORAIS: UM ESTUDO DE CASO EM
PACIENTES INTERNADOS EM UTI**

INTRODUÇÃO

A tecnologia da rede mundial tem avançado com grande rapidez, permitindo o acesso aos dados de forma transparente. O uso universalizado de leitores magnéticos, aquisição automática de dados, equipamentos médicos e a internet, estão produzindo grande quantidade de dados numa rapidez sem precedentes. E ainda, o custo do poder de processamento dos computadores tem diminuído drasticamente. Somado ao crescimento de dados, as tecnologias de software, sistemas de gerenciamento de banco de dados têm apresentado grandes avanços.

Pensando na área da saúde como um todo, em operações que poderiam extrair melhores observações em relação a certo tipo de doença, e nos processos de automação hospitalar, torna-se imprescindível o apoio de tecnologias no auxílio das tomadas de decisão.

Com o surgimento de tecnologias capazes de automatizar ambientes hospitalares e capturar tais informações, é possível aperfeiçoar processos da área da saúde, sendo, portanto uma forma de minimizar riscos, contribuindo de maneira efetiva para a melhoria qualidade da saúde da população (Tsumoto, 2007).

O ambiente hospitalar requer demanda de tecnologia, desde o simples atendimento em ambulatório até os equipamentos de monitoramentos de pacientes internados em Unidade de Terapia Intensiva (UTI). Através dessas tecnologias abordadas na área médica, é que surge a necessidade de integração de dados. Todas as informações geradas a partir de um paciente em atendimento contribuem para o aperfeiçoamento dos resultados médico.

Assim, a necessidade de inovação tecnológica e conseqüentemente o surgimento de novos produtos nessa área é vital, visto que muitos procedimentos médicos encontram-se realizando de forma manual, aspecto que dificulta o controle dos dados e o gerenciamento das informações, podendo levar a erros graves em relação à vida dos pacientes.

Em dados de qualquer ambiente médico a ferramenta para extração de conhecimento utilizando primitivas temporais, intitulada OpMM, baseado nas etapas do *Knowledge Discovery Database* (KDD), com a finalidade de subsidiar profissionais da área médica para auxílio ao diagnóstico médico. A ferramenta está baseada em técnicas de mineração de dados através do algoritmo de árvore de decisão associada a primitivas temporais.

Os SGBD-TR podem ser vistos como a integração de um SGBD convencional com um sistema em Tempo-Real (STR). Como um SGBD processa transações e garante a integridade lógica dos dados e como um STR garante restrições temporais das transações. Portanto, um SGBD-TR possui como principais características o tratamento de dados com validade temporal e transações com restrições explícitas de tempo de execução (LEITE, 2005). Estas características são úteis para aplicações com tempo crítico que necessitam coletar, modificar e recuperar grandes volumes de dados compartilhados.

Os dados usados em um SGBD-TR devem refletir o estado real do ambiente da aplicação, portanto, a estrutura desses dados indica que os valores gravados são válidos apenas por um determinado intervalo de tempo. Assim, os intervalos de tempo especificam a validade temporal dos dados. Desta forma, um SGBD-TR busca garantir a consistência temporal dos mesmos através da validação desses intervalos.

Além do suporte convencional de um SGBD, um SGBD-TR deve oferecer suporte para o tratamento dos dados e das transações com restrições temporais. Portanto, um SGBD-TR deve fornecer suporte para os usuários especificarem as restrições temporais dos dados, isto é, um intervalo de tempo durante o qual um dado é considerado válido, além de definir os prazos das transações, ou seja, o tempo no qual uma transação deve ser executada. Esses requisitos são bastante complexos, no entanto várias pesquisas estão sendo realizadas no sentido de atendê-los (LEITE, 2005).

Para garantir a integridade do processo de descoberta de conhecimento dos dados, para o KDD é necessário que as etapas, como mostra a Figura 1, se adéquem a base de dados. Assim, Malucelli (2010), enfatiza a sequência de cada etapa e sua devida importância.

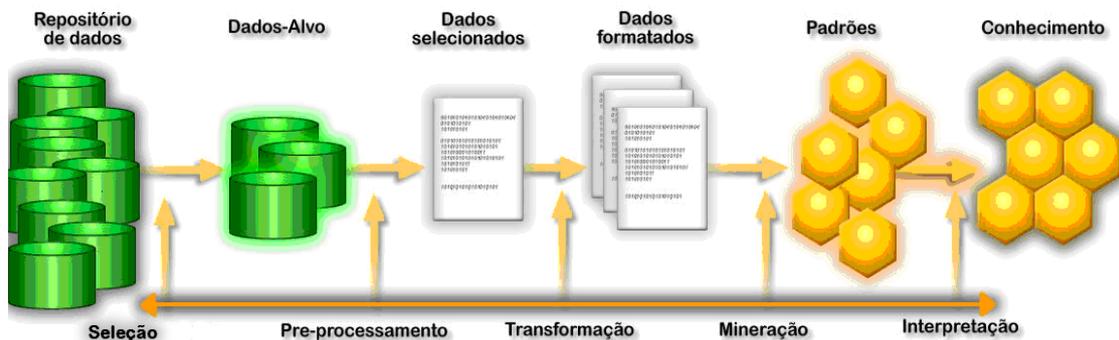


Figura 1: Etapas do Processo KDD. Fonte: Adaptado de Fayyad, 1996.

O KDD é um processo composto pela seleção de dados, pré-processamento, transformação dos dados e estabelecimento de padrões úteis na extração de conhecimento, ou seja, tradução de dados brutos em informações relevantes (VIANA et al., 2010).

MINERAÇÃO DE DADOS

Para a aplicação da mineração de dados ao banco de dados, é necessária toda a adaptação ao processo do KDD, onde as etapas são necessárias para a obtenção do objetivo que são os resultados obtidos através da mineração de dados.

Atualmente, as técnicas de mineração de dados consistem, sobretudo, na análise dos dados após a extração, buscando-se levantar as necessidades reais e hipotéticas de cada usuário. A tecnologia disponível gira em torno de algoritmos avançados, computadores multiprocessados e bancos de dados grandes e poderosos.

O seu processo de desenvolvimento atual é composto de métodos e algoritmos que possibilitam a extração de novos conhecimentos. Entre os seus vários algoritmos, destacam-se alguns que são os mais utilizados: associação, classificação, regressão, clusterização e sumarização (GALVÃO, 2012).

Segundo Elsmari (2011), o resultado da mineração de dados é descobrir um tipo de informação nova, que estabelece através da regra de associação os fatos e padrões sequenciais. Os fatos se aproximam através de registros do banco e árvore de classificação.

ÁRVORES DE CLASSIFICAÇÃO

O algoritmo de classificação tem o objetivo de encontrar uma hierarquia de classes predeterminadas, ligada ao método de aprendizado supervisionado, com o objetivo de

induzir conceitos a partir de atributos que estejam rotulados. Aplicando essa estratégia de modelo podem-se classificar novos dados.

Witten e Frank descrevem que os atributos são identificados quanto a sua importância, a árvore de decisão facilita para compreensão de qual atributo influencia mais em seus resultados (WITTEN e FRANK, 2005).

A árvore de decisão representa a descrição de cada classe de atributo através das regras de classificação, com a representação gráfica de uma árvore. A Figura 2, de Elsmari (2011) representa uma árvore de classificação, para aplicação especificamente de cartão de crédito, com o atributo casado, sendo o atributo mais abrangente, e especificamente cada ramificação restringe a uma situação.

A árvore de decisão apresenta um conjunto de dados, onde todas as amostras estão representadas na raiz da árvore. Tais amostras são classificadas com sub-rotinas baseadas nos atributos que forem selecionados. A árvore de decisão representa a descrição de cada classe de atributo através das regras de classificação, com a representação gráfica de uma árvore.

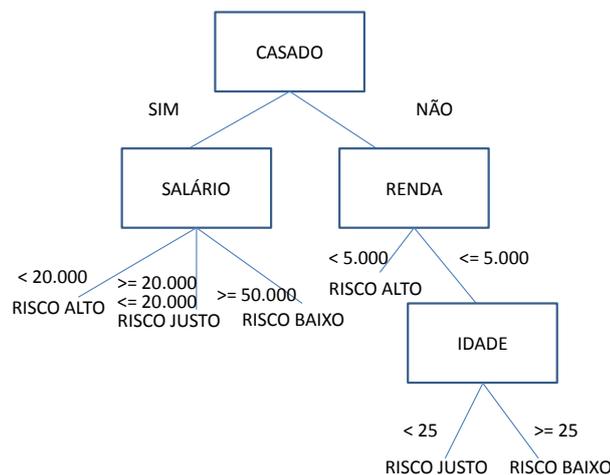


Figura 2: Árvore de decisão (ELSMARI, 2011, p. 710).

BASE DE DADOS MÉDICA

A maior parte das tarefas clínicas requer a captação de numerosos e múltiplos tipos de dados de pacientes, frequentemente através de meios eletrônicos, sendo que para fazer diagnósticos ou tomar decisões terapêuticas é necessário uma interpretação destes dados. A maioria dos dados armazenados inclui uma validade temporal em que estes são considerados válidos.

As tendências temporais e os padrões dos dados clínicos adicionam introspecções significativas à análise estática. Assim, é desejável criar automaticamente abstrações (curtas, informativas e interpretativas) de dados clínicos em tempo-real e poder responder a perguntas sobre tais abstrações. Ao fornecer esta capacidade de abstrair dados com

restrição de tempo, consegue-se ajudar o médico, criando uma ferramenta de apoio à decisão automatizada (BOAZ e SHAHAR, 2005).

Podem-se aplicar as técnicas KDD em diversas bases de dados, no entanto o foco desse trabalho se concentra em base de dados médica, envolvendo todo o processo de informatização hospitalar de pacientes internados que estão sendo monitorados de alguma forma, seja monitoramento manual, automático, ou através de exames médicos.

A base de dados médica utilizada, para validação deste trabalho, é a *Multiparameter Intelligent Monitoring in Intensive Clinical* (MIMIC II), que é um projeto criado em outubro de 2003, a partir do *Bioengineering Research Partnership* (BRP) com a parceria do *National Institute of Biomedical Imaging and Bioengineering* (NIBIB). O MIMIC II é uma base de dados pública, disponível na internet através do domínio da Physionet, que só libera o acesso após curso de proteção humana, através do *National Institutes of Health* (NIH). A base possui coleta dos dados de pacientes internados em Unidade de Terapia Intensiva (UTI) no *Beth Boston's Israel Deaconess Medical Center* (BIDMC).

O banco de dados MIMIC II contém informações do leito de UTI do paciente, como também informações arquivadas de interesse do hospital. Os dados foram coletados entre o período de 2001 e 2008 de pacientes em leitos de UTI, seja clínico, cirúrgico, coronariano e neonatal. O MIMIC II é dividido em dois tipos de informações, clínicas e em dados gráficos, onde os dados clínicos são toda e qualquer informação fisiológica do paciente e os dados em sinais de alta resolução, de medições fisiológicas minuto a minuto, facilitando comparações de dados clínicos com dados gráficos.

Para identificação dos dados alguns equipamentos de monitoramento, que se pode aplicar ao paciente, onde os dados capturados são armazenados em banco, como: termômetro, oxímetro de pulso, eletrocardiográfico com frequência cardíaca e medida intermitente de pressão arterial, monitor de pressão arterial, capnógrafo, monitor cardíaco, máscara e cateter de oxigênio, cateter central, tubo ortotraqueal e ventilador mecânico. Ainda como formas de monitoramento existem as observações médicas que são informadas manualmente, os registros de exames laboratoriais, os campos com exames de imagem dentre outros diversos registros.

MATERIAIS E MÉTODOS

A aplicação foi concebida com intuito de auxiliar o desenvolvimento de aplicações com fluxo de informação contínua e que necessitam de um acompanhamento e extração de conhecimento dos dados. Com a finalidade de uma aplicação de iniciativa no auxílio da tomada de decisão médica, em dados médicos genéricos, ou seja, independente do problema médico o comportamento do OpMM será o mesmo.

Para o desenvolvimento da aplicação de integração foi utilizada uma base de dados médica MIMIC II, onde a aplicação oferece a funcionalidade de consultas como um SGBD, através da conexão com o banco de dados, onde estão armazenadas as informações de interesse, os registros de cadastro de pacientes, são extraídos do servidor de dados, através de um banco de dados, conforme Figura 3.

OpMM

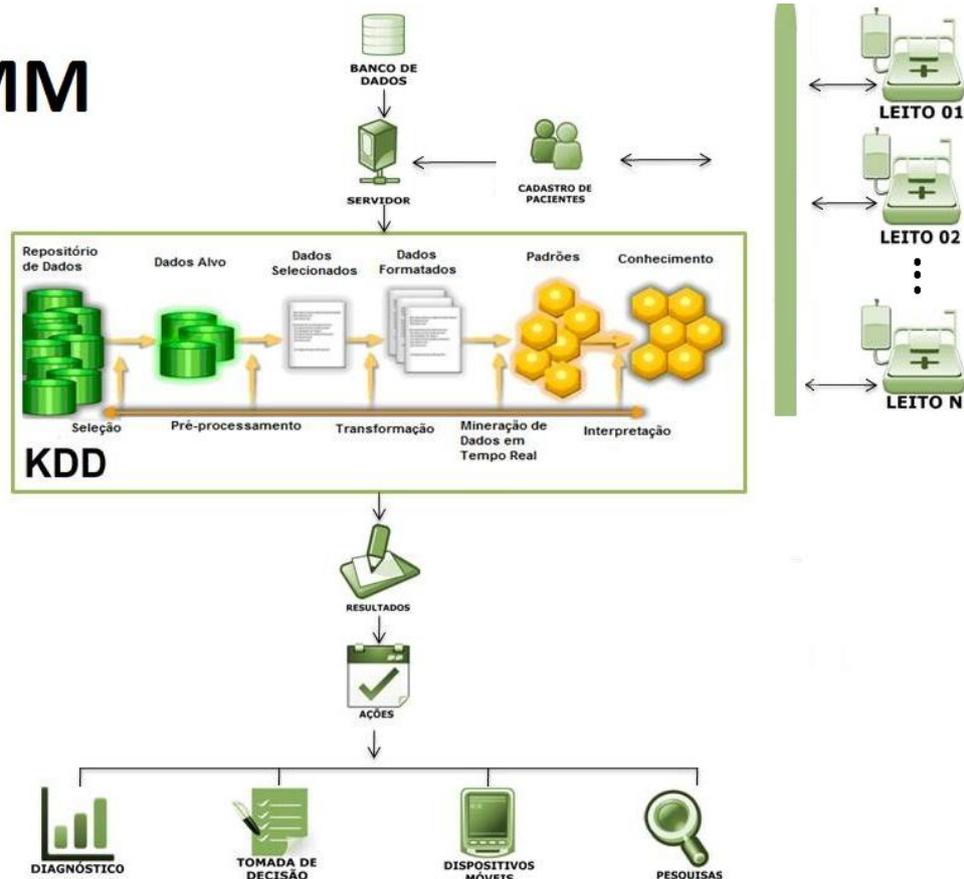


Figura 3: Mineração de Dados Médico em Tempo-Real (MOURA, 2013).

O processo do KDD está totalmente integrado, as coletas contínuas dos dados, como foram apresentadas na Figura 3. O Banco de Dados com o armazenamento dos dados do paciente de cada leito direcionado para o servidor. Assim, os padrões são identificados gerando conhecimento válido no processo do KDD, da base médica em análise, podendo ser disseminados através de diagnóstico, tomada de decisão, alerta de mensagens dos dispositivos móveis e computadores e pesquisas.

A aplicabilidade do OpMM, no ambiente médico, enfatiza a possibilidade de compreensão de adequação a qualquer base de dados da área, tornando a extração de conhecimento em um grande volume de dados, descobrindo novas correlações, padrões e tendências entre as informações de um paciente.

O fluxo de dados na arquitetura é adaptado à base na qual esteja instalado, independente de qual tipo de dados médico seja analisado. Os dados gerados periodicamente são pré-processados, com a aplicação da técnica de mineração e a obtenção de resultados são permitidas munidas das aplicações das ações de interesse, conforme ilustrado na Figura 4.

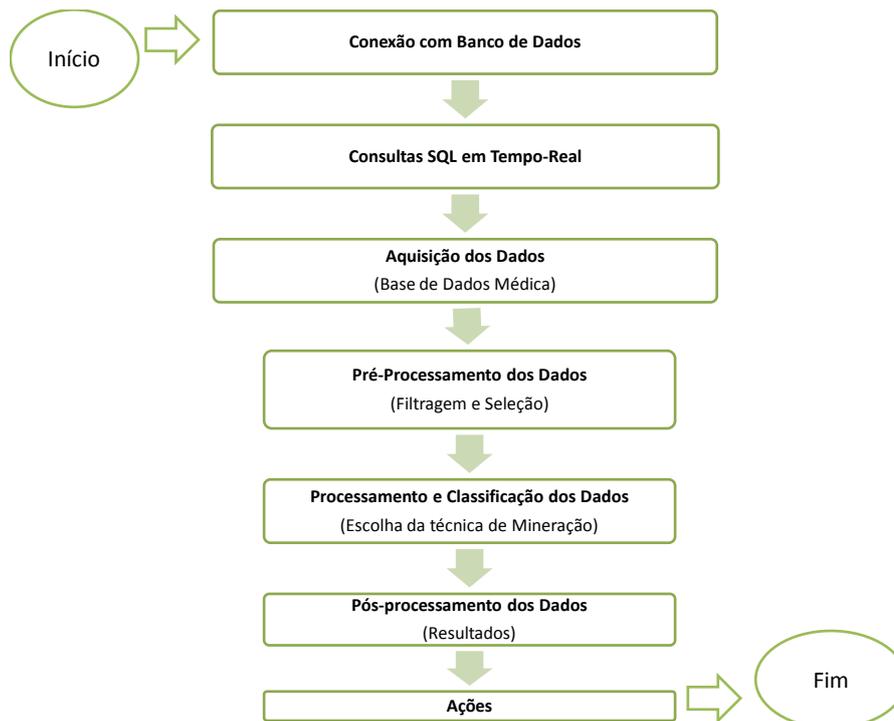


Figura 4: Fluxograma do OpMM (MOURA, 2013).

Todo o processo do fluxograma ocorre desde a conexão com o banco de dados a ser ajustado a arquitetura do OpMM, com a consulta SQL em tempo real já é possível a restrição aos dados e ao intervalo de tempo, para as informações de interesse que é a aquisição dos dados. Na etapa de pré-processamento dos dados identifica-se os registros inconsistentes ou inválidos. No processamento e classificação dos dados é feito o processo de mineração de dados e na etapa do pós-processamento dos dados já acontece os resultados da mineração de dados. Finalizando o fluxograma com as ações que o usuário irá seguir com os resultados da mineração.

DESCRIÇÃO GERAL DE INTERFACE

A interface descrita a seguir dá suporte à elaboração de consultas utilizando a sintaxe da Linguagem de Consulta em Tempo-Real, permitindo a manipulação de dados armazenados em um SGBD convencional, no entanto, tal manipulação obedece às restrições de tempo. O OpMM disponibiliza duas interfaces: uma interface para realizar conexão com o banco de dados e outra interface para declarar consultas.

O OpMM foi implementado utilizando a linguagem de programação Java, com a ferramenta de desenvolvimento Eclipse, versão *Indigo*. A linguagem Java disponibiliza um pacote `java.util` que contém uma Classe denominada *Timer* e uma Classe denominada *TimerTask*. A Classe *Timer* disponibiliza um conjunto de primitivas de tempo e métodos para o agendamento de tarefas. O método `schedule()`, por exemplo, pode receber até três parâmetros: uma tarefa, uma data para a tarefa iniciar e a frequência que a mesma deve ser realizada.

A interface foi desenvolvida através da linguagem Java, que disponibiliza um conjunto de características fundamentais para sua implementação, tais como: portabilidade, segurança, robustez, bibliotecas para acesso a banco de dados e pacotes que disponibilizam primitivas de tempo-real (Eckel, 2002). O ambiente de desenvolvimento utilizado foi o Eclipse (Eclipse, 2012). O SGBD utilizado foi o PostgreSQL (PostgreSQL, 2012).

INTERFACE DE CONEXÃO

Pela conexão definida no ambiente Java (Java, 2012). A comunicação é feita, através da interface desenvolvida, o usuário passa como parâmetros o seu nome (*User*), uma senha (*Password*) e o nome de um banco de dados válido (*Database*). Caso os parâmetros sejam validados, a conexão com o banco de dados é estabelecida e uma mensagem de conexão estabelecida é exibida no campo Output da interface. Caso contrário uma mensagem de erro será exibida no mesmo campo, conforme Figura 5.

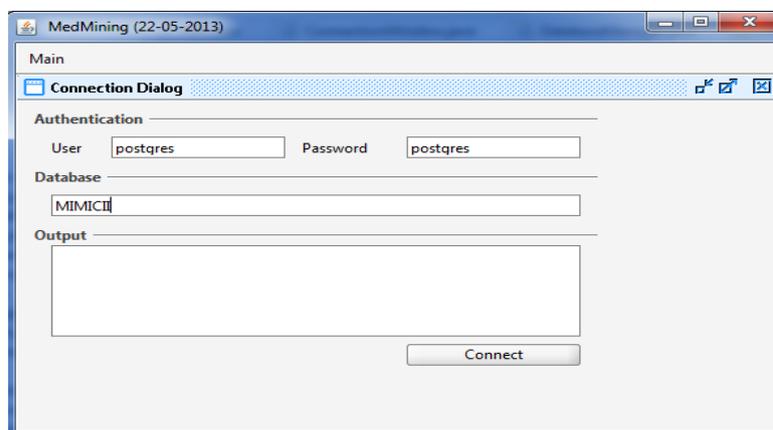


Figura 5: Interface de Conexão.

INTERFACE DE CONSULTA

A interface de consulta segue a mesma tela de visualização de um SGBD, com a implementação das consultas em tempo-real e a execução, conforme Figura 6.

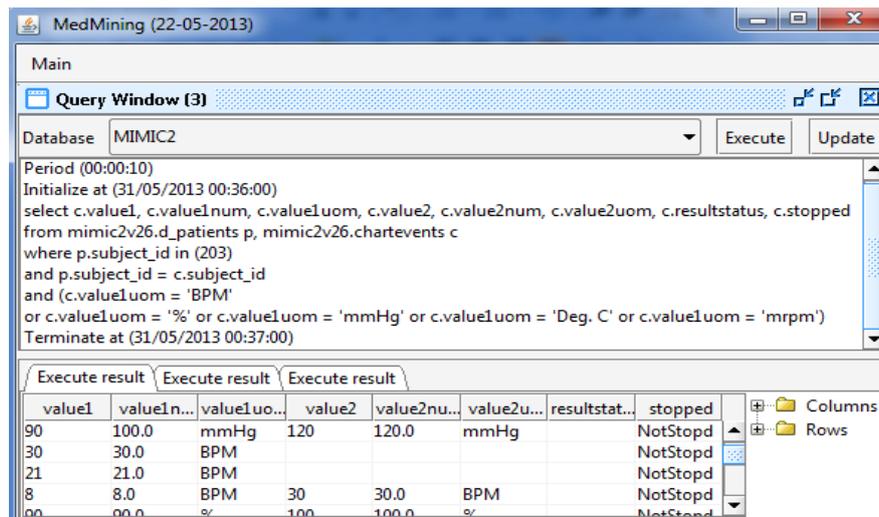


Figura 6: Interface de Consulta.

Os resultados da consulta são apresentados na mesma tela da consulta, ainda sem a aplicação do algoritmo. Porém, logo que o algoritmo finaliza o processo, a tela de aplicação do algoritmo será apresentada conforme Figura 7, com uma tela de mineração de dados, dividida em classificação de dados e a árvore de decisão.

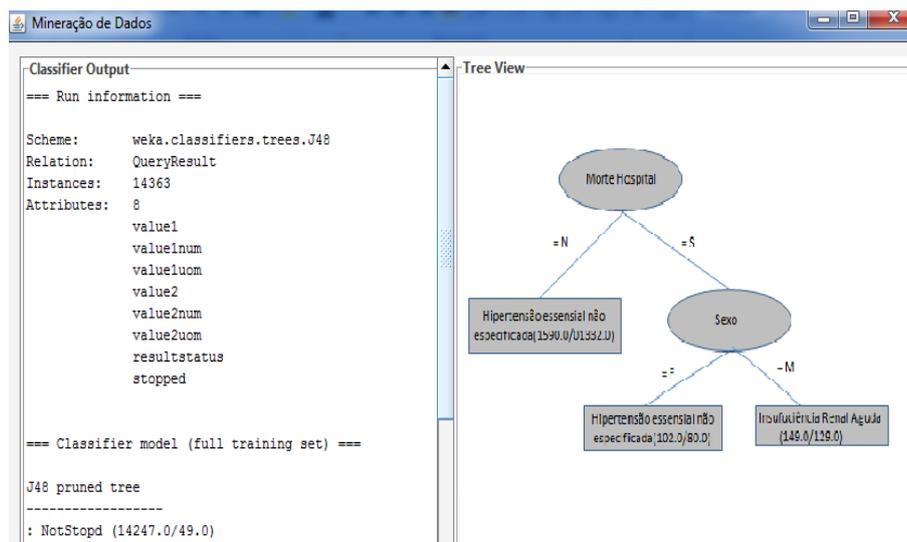


Figura 7: Mineração dos dados.

No entanto, as informações de interesse para o monitoramento dos pacientes, podem ser armazenadas através da automação. O monitoramento é contínuo e direto, ou seja, para cada paciente, um monitor de cabeceira e/ou transporte é disponibilizado.

Para o médico as hipóteses que favorecem a admissão e conhecimento do estado de cada paciente, seria o índice apontado pelos sinais vitais. Assim, para a consulta (1), procurou-se na base de dados os registros relacionados com Batimentos cardíacos Por Minuto (BPM), nível de oxigenação sanguínea, Pressão Arterial (PA), temperatura e Movimentos Respiratórios Por Minuto (MRPM).

Para demonstração da execução das consultas em tempo-real, foi selecionado o paciente com *subject_id* '203' e os sinais vitais do mesmo, em um intervalo de dados que inicialize

na data e hora relacionada (25/05/2013 14:00:00) e termine no (25/05/2013 18:00:00), alguns exemplos de consulta serão apresentados abaixo:

Consulta (1):

```
INITIALIZE AT (25/05/2013 14:00:00)
PERIOD (00:00:10)
select c.value1, c.value1num, c.value1uom, c.value2,
       c.value2num, c.value2uom, c.resultstatus, c.stopped
from mimic2v26.d_patients p, mimic2v26.charthevents c
where p.subject_id in (203)
and p.subject_id = c.subject_id
and (c.value1uom = 'BPM' or c.value1uom = '§'
or c.value1uom = 'mmHg'
or c.value1uom = 'Deg. C'
or c.value1uom = 'mrpm')
TERMINATE AT (25/05/2013 18:00:00)
```

A Consulta (1) recupera do paciente '203' (p.subject_id in (203)) os dados de monitoramento (c.value1, c.value1num, c.value1uom, c.value2, c.value2num,), onde armazena os valores referentes aos sinais vitais, dentro do período de 04 horas (INITIALIZE AT (25/05/2013 14:00:00)), (TERMINATE AT (25/05/2013 18:00:00) e a cada 10 (PERIOD (00:00:10)) segundos. Com a Consulta (1), pode-se observar o comportamento do paciente e suas intercorrências, dentro do intervalo de tempo citado.

Aplicando o algoritmo supervisionado de classificação por árvore de decisão J48, obteve-se padrões de resultados para o paciente '203', como Pressão Arterial normal, com febre, apresentou Frequência Respiratória baixa.

A Consulta (2) faz observações como sexo, morte ou não no hospital, idade, peso, nos registros da base, sem a inserção de restrição temporal, apenas por finalidade de conhecimento dos dados médico. Preparando os dados para execução, identificando apenas alguns atributos, dentre vários existentes, como sexo, Classificação Internacional de Doenças (CID), morte no hospital, estado civil, etnia, convênio, religião, tipo de admissão, fonte de admissão, idade, peso e altura.

Consulta (2):

```
INITIALIZE AT (26/05/2013 14:00:00)
PERIOD (00:00:10)
select o.subject_id, o.medication, o.status, o.frequency, i.description, d.admission_type_descr,
       t.icustay_age_group as idade, p.sex as sexo, p.hospital_expire_flg as morte_hospital
from mimic2v26.poe_order o, mimic2v26.icd9 i, mimic2v26.demographic_detail d, mimic2v26.d_patients p, mimic2v26.icustay_detail t
where o.subject_id = p.subject_id and o.subject_id = i.subject_id and o.subject_id = d.subject_id
and d.subject_id = t.subject_id and o.subject_id = t.subject_id and p.subject_id = t.subject_id
and i.subject_id = t.subject_id and d.subject_id = p.subject_id and i.subject_id = d.subject_id
and i.subject_id = p.subject_id and p.hospital_expire_flg = 'Y'
and i.code in ('V29.0', 'V05.3', '584.9', '599.0', '285.9', '401.9', '414.01', '250.00', '428.0', '427.31', '486.0', '272.0', '518.81')
TERMINATE AT (26/05/2013 18:00:00)
```

A Consulta (2) recupera os pacientes (o.subject_id), por medicação (o.medication), tipo de admissão (d.admission_type_descr), idade (t.icustay_age_group), sexo (p.sex) e se foi a óbito (p.hospital_expire_flg) com restrições algumas doenças identificadas por CID ('V29.0','V05.3','584.9','599.0','285.9','401.9','414.01','250.00','428.0','427.31','486.0','272.0','518.81'). Essa consulta foi realizada baseada na hipótese de relacionamento entre as doenças do paciente, sexo e idade.

As aplicações do filtro de classificação com atributos de sexo e CID identificaram visualmente, o cruzamento das informações, que geraram os resultados de doenças com maiores incidência nos 10.817 pacientes que foram admitidos pelo hospital, sendo que, 292 apresentaram registros de hipertensão, 216 de aterosclerose coronária e 146 de fibrilação atrial, onde pode ser observado que 800 do sexo feminino e 1041 do sexo masculino, conforme apresentado na Figura 8.

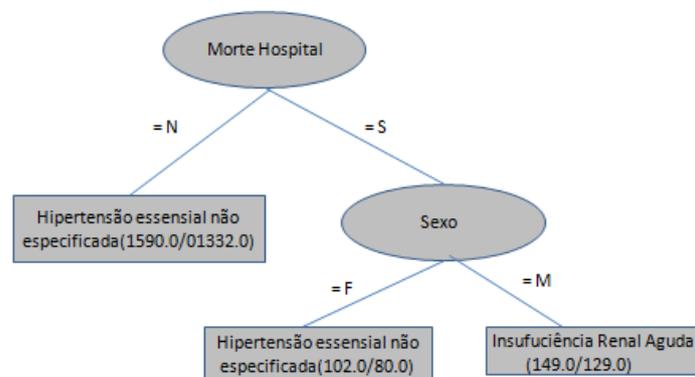


Figura 8: Árvore de Decisão.

A partir de observações realizadas na Consulta (2) após análise de pós-processamento, outra consulta SQL, Consulta (3), foi desenvolvida com base no relato de doenças específicas, analisou-se informações com base em faixa etária observando as doenças como: infarto, arritmias cardíacas, hipertensão, diabetes, linfoma, aids, dentre outros.

Consulta (3):

```

INITIALIZE AT (27/05/2013 14:00:00)
PERIOD (00:00:10)
select p.subject_id as prontuario, p.sex as sex, p.hospital_expire_flg as death_hospital,
t.icustay_age_group as age_group, t.height as height, t.weight_first as weight,
dischstatus as status, congestive_heart_failure, cardiac_arrhythmias, valvular disease,
pulmonary_circulation, peripheral_vascular, hipertension, paralysis, other_neurological,
chronic_pulmonary, diabetes_uncomplicated, diabetes_complicated, hypothyroidism, renal_failure,
liver_disease, peptic_ulcer, aids, lymphoma, metastatic_cancer, solid_tumor, rheumatoid_arthritis
from mimic2v26.icustay_detail t, mimic2v26.d_patients p, mimic2v26.demographic_detail d,
mimic2v26.censusevents v, mimic2v26.comorbidity_scores s
where d.subject_id = p.subject_id and t.subject_id = d.subject_id and p.subject_id = t.subject_id
and v.subject_id = t.subject_id and p.subject_id = v.subject_id and d.subject_id = v.subject_id
and t.subject_id = s.subject_id and p.subject_id = s.subject_id and v.subject_id = s.subject_id
and d.subject_id = s.subject_id and p.subject_id = d.subject_id
TERMINATE AT (27/05/2013 14:00:00)
    
```

A Consulta (3) recupera as mesmas informações da Consulta (2), aborda-se precisamente a altura (t.height), peso (t.weight_first) junto as doenças complexas em atendimento hospitalar, como por exemplo arritmia cardíaca (cardiac_arrhythmias). Na expectativa de relacionar peso e altura, tipos de admissão e quantificar os tipos de atendimentos relacionadas com determinadas doenças apontadas na Consulta (3).

Abordando os atributos, conforme Consulta (2) de sexo, morte no hospital e CID. Onde foi possível gerar regras do tipo:

- Pacientes que morreram no hospital, do sexo feminino, tiveram como maior índice o diagnóstico de hipertensão.
- Pacientes que morreram no hospital, do sexo masculino, tiveram como maior índice o diagnóstico de doença renal aguda.

Uma avaliação feita em pacientes com linfoma foi possível identificar a predominância de outras doenças que os acometem, como vários tipos de diabetes, casos de hipertensão, infarto agudo do miocárdio e gastroparesia, que é uma diminuição da força de contração da musculatura do estômago.

Com a aplicação da técnica de árvore de decisão, vários resultados foram gerados, de acordo com os atributos de sexo, fonte de admissão e morte no hospital, como apresentado na Tabela 1:

Tabela 1 – Resultados da Mineração

ATRIBUTOS	RESULTADOS
Sinais Vitais	<ul style="list-style-type: none"> • Paciente '203' com PA normal, com febre, apresentou FR baixa. • Paciente '203' com proximidade de temperatura elevada apresentou FR baixa. • FC baixa, a temperatura e a FR baixa. • PA Normal, FC normal, apresentaram febre e FR baixa.
Sexo	<ul style="list-style-type: none"> • Prevaecem 6.398 são pacientes do sexo masculino.
Fonte de Admissão	<ul style="list-style-type: none"> • A maior admissão é na sala de emergência com 5.853, transferência de pacientes com 1.848, somente como urgência e emergência, possuindo maior incidência em pacientes do sexo masculino e casados. • Os 1.106 recém-nascidos foram admitidos pelo hospital como encaminhamento normal e prematuro. • Para o tipo de admissão eletiva tiveram somente fonte

	de admissão normal sem referência específica.
Morte no Hospital	<ul style="list-style-type: none"> • Dos 10.817 pacientes admitidos 251 foram a óbito com maioria do sexo masculino, casado e admitidos em urgência.

Os resultados do OpMM apresentados foram executados através da interface implementada neste trabalho, permitindo a inserção de consultas com interesse, com a visualização de novos resultados através do algoritmo em questão.

DISCUSSÃO

Este processamento foi realizado através do conhecimento de especialistas médicos que ajudam a gerar hipóteses com o uso de algoritmos de mineração de dados. Logo, com as informações extraídas é possível auxiliar e apoiar a equipe médica na transformação de ações em diagnósticos, com necessidade de avaliações precisas e automáticas.

As contribuições do OpMM é favorável para a sua utilização em qualquer ambiente que necessite de exploração de dados, facilitando a interação com o usuário final a conexão em qualquer SGBD. Assim, as tecnologias utilizadas em todo o processo proposto, como o desenvolvimento de uma ferramenta com aplicação específica para a descoberta de conhecimento na área médica buscando atender às exigências de monitoramento da automação hospitalar.

Durante o desenvolvimento deste trabalho, foram identificados alguns aspectos que podem ser explorados ou estendidos a partir do estudo apresentado. A seguir, destacam-se alguns tópicos de pesquisa que precisam ser realizados:

- A importância da análise exploratória dos dados e validação de outros algoritmos, com a aplicação destas a um problema médico real;
- Desenvolvimento de um sistema *on-line* que permita à equipe médica e aos familiares terem acesso ao quadro clínico do paciente, mediante normas estabelecidas;
- Validação da ferramenta em um cenário e ambiente real, confrontando em tempo-real a geração dos alarmes e a recepção das mensagens;
- Utilização da ferramenta para as aplicações de dispositivos móveis de comunicação;
- Integração de vários módulos inteligentes que possam dar um diagnóstico mais complexo, utilizando histórico, antecedentes familiares e novos parâmetros específicos para cada módulo;
- Aplicação em outro cenário de monitoramento e envio de alertas.

Ressalta-se que a ferramenta OpMM é de grande importância para a base de dados que necessitem da extração de informações com restrição temporal, o que amplia sua contribuição. Assim, com os resultados da mineração de dados com restrição temporal, no momento que se descobre novas informações o apoio à decisão será imediato. Diante das contribuições expostas, foram identificados importantes impactos de natureza científica e tecnológica na automação hospitalar.

REFERÊNCIAS BIBLIOGRÁFICAS

1. Boaz, D., Shahr, Y. A framework for distributed mediation of temporal-abstraction queries to clinical databases. *Journal - Artificial Intelligence in Medicine*. ISSN 0933-3657, 2005.
2. Carvalho, J. S. Verificação Automatizada de Sistemas de Tempo Real Críticos. Dissertação de Mestrado. Universidade da Beira Interior, 2009.
3. Eckel, B. *Thinking in Java*. 3rd. ed. 2002. Available in web, <http://www.bruceeckel.com>. Acesso em 02 maio, 2013.
4. Eclipse. [Available online at: <http://www.eclipse.org/downloads>]. Acesso em 20 out 2012.
5. Elmasri, R.; Navathe, S. B. *Sistemas de Banco de Dados*. 6º. ed. São Paulo: Pearson Addison Wesley, 2011.
6. Fayyad, U. M., Piatetsky-Shapiro, G.; Smyth, P. e Uthurusamy, R., *From Data Mining to Knowledge Discovery in Database*. Cambridge, American Association for Artificial Intelligence/MIT Press, 1996.
7. Fernandes, Y. Y. M. P. Técnica de Controle de Concorrência Semântico para Sistemas de Gerenciamento de Banco de Dados em Tempo-Real. Dissertação de Mestrado, UFCG. Campina Grande, 2005.
8. Galvão, N. D.; Marin, H. de F. Técnica de Mineração de Dados: uma revisão da literatura. *Actapaul. enferm.*, São Paulo, v. 22, n. 5, Oct. 2009. http://www.scielo.br/scielo.php?script=sci_arttext&pid=S010321002009000500014&lng=en&nrm=iso>. Acesso em 12 Jan. 2012.
9. Java. [Available online at: http://www.java.com/pt_BR/download/faq/whatis_java.xml]. Acesso em 20 dez. 2012.
10. Leite, C.R.M., "Linguagem de Consulta para aplicações em Tempo-Real." Dissertação de Mestrado, Campina Grande, PB, 2005.
11. Malucelli, Andreia et al . Classificação de microáreas de risco com uso de Mineração de Dados. *Rev. Saúde Pública*, São Paulo, v.44, n.2, Apr. 2010. Disponível em:<http://www.scielosp.org/scielo.php?script=sci_arttext&pid=S0034-89102010000200009&lng=en&nrm=iso>. Acesso em 16 jan 2012.
12. Minnie, D; Srinivasan S., "Application of Knowledge Discovery in Database to Blood Cell Counter Data to Improve Quality Control in Clinical Pathology," 2011 Sixth International Conference on Bio-Inspired Computing: Theories and Applications, pp.338-342, 2011.
13. Moura, K.H. S. OpMM – Ferramenta para extração de conhecimento utilizando primitivas temporais: Um estudo de caso em pacientes internados em UTI. Dissertação de Mestrado, Mossoró, RN, 2013.
14. Neto, P. R. et al. Uma aplicação de bancos de dados em tempo-real para redes de sensores. VI Workshop de Tempo Real (WTR) - SBRC, p. 45_52, 2004.
15. Oliveira, S. P.; Garcia, A. C. P. Variáveis e indicadores para análise de recursos humanos em saúde no Brasil – Rio de Janeiro: ENSP/FIOCRUZ. 2006.
16. PhysioNet/MIMIC II - Multi-parameter Intelligent Monitoring for Intensive Care (MIMIC II) Database [Available online at: <http://www.physionet.org/physiobank/database/mimic2db/>] Acesso em 12 dez. 2011.

17. PostgreSQL. [Available online at: <http://www.postgresql.org.br/downloads>]. Acesso em 20 dez 2012.
18. Tsumoto, S.; Yokoyama, S.; Matsuoka, K., "Mining Risk Information in Hospital Information Systems as Risk Mining," *Complex Medical Engineering, 2007. CME 2007. IEEE/ICME International Conference on*, vol., no., pp.1917,1921, 23-27 May 2007 doi: 10.1109/ICCME.2007.4382082.
19. Witten, I. H. & Frank, E. *Data Mining: Practical Machine Learning Tools and Techniques*, second edition, Elsevier: San Francisco, ISBN 0-12-088407-0, 2005.