

# **DESCOBERTA DE CONHECIMENTO CONSIDERANDO ASPECTOS TEMPORAIS**

Anelise de Macedo Lucas (anelisel@inf.ufrgs.br)\*

Nara Martini Bigolin (bigolin@inf.ufrgs.br)\*\*

Luís Otavio Alvares (alvares@inf.ufrgs.br)\*

\*Universidade Federal do Rio Grande do Sul - Instituto de Informática - Av. Bento Gonçalves 9500  
Porto Alegre, RS Brasil, 91501-970, Tel. ++5551-3316-6828

\*\*Universidade Luterana do Brasil- Faculdade de Informática- Rua Miguel Tostes,101  
Canoas, RS, Brasil, 92.420-280 Tel.++5551- 4774000

## **RESUMO**

A descoberta de conhecimento em banco de dados é uma área nova, mas que está sendo cada vez mais desenvolvida e utilizada em escala empresarial. Este trabalho aborda uma parte dessa área, ainda pouco explorada, que considera os aspectos temporais. Inicialmente serão definidos os principais conceitos relativos a descoberta de conhecimento considerando aspectos temporais. Em seguida é apresentado as etapas do processo de descoberta de conhecimento e finalmente a abordagem é validada com um experimento utilizando dados da Secretaria da Saúde do estado do Rio Grande do Sul.

## **Palavras-chave**

Mineração de Dados. Descoberta de Conhecimento em Banco de Dados. Inteligência Artificial. Banco de Dados. Aspectos Temporais. Mineração de Dados Temporais.

## **1 Introdução**

Atualmente, observa-se que o crescimento do volume e do número de bancos de dados existentes excedem a capacidade humana de analisar seus dados com simples consultas ou pesquisas e, desta forma, não podem analisar o seu conteúdo quanto a algum conhecimento implícito importante existente nos dados. Desta forma, devido ao crescimento do volume de informações que as corporações manipulam, gera-se a necessidade urgente de técnicas e ferramentas que transformem dados em conhecimento útil de forma inteligente e automática. A

solução para esta necessidade das organizações de obterem conhecimento a partir de grandes volumes de dados está na utilização de técnicas de mineração de dados para extrair as informações implícitas existentes nos bancos de dados destas organizações.

A partir disto, o processo de Descoberta de Conhecimento em Banco de Dados (DCBD) objetiva obter conhecimento a partir de bancos de dados (BD). Este processo é composto por três etapas sucessivas: pré-processamento, mineração e pós-processamento dos dados [7]. A etapa da mineração dos dados (MD) tem recebido um significativo interesse por parte de diversas áreas como, por exemplo, *marketing*, finança, transporte, agricultura e saúde entre outras. Esta tendência se deve ao fato desta etapa permitir a extração de informações implícitas dos BD, visando descobrir modelos de representação do conhecimento [1; 2; 3]. Adicionalmente, o aumento do interesse nos aspectos temporais dos dados é observado pelo número crescente de protótipos de sistemas que estão utilizando a compreensão temporal para explicar as relações comportamentais associadas à natureza da variação do tempo implícita no universo. Desta forma, este artigo procura apresentar um estudo sobre a mineração de dados temporais (MDT) aplicado à área da saúde.

A motivação para este trabalho está no grande volume de dados que é coletado a cada dia no decorrer do tempo de um evento<sup>1</sup>. Estes dados se constituem em valiosas fontes de informação que podem ser analisadas em uma seqüência de determinados eventos, ou conjuntos de eventos relacionados a particularidades temporais. Estes tipos de análises podem ser muito úteis para derivar a informação implícita dos dados crus, e para predizer o comportamento futuro de um processo monitorado contribuindo para o apoio a decisões e para o enriquecendo das áreas envolvidas na tarefa de MDT.

O presente documento estrutura-se da seguinte maneira: na próxima seção, são descritos os conceitos e definições da MDT; na terceira seção, é apresentado o processo de descoberta temporal; na quarta seção, é relatado o experimento da aplicação da MDT sobre um caso real; na quinta seção, são apresentadas as considerações resultantes deste trabalho e suas contribuições futuras.

## **2 Mineração de Dados Temporais**

### **2.1 Conceitos e Definições**

A evolução da MD tem relação direta com a sua aplicação para diversos domínios. Adiciona-se a isto, o fato que estes domínios sofrem influência temporal para a compreensão de seus dados. A atividade de extração do conhecimento que leva em consideração o aspecto temporal dos dados chama-se de mineração de dados temporais (MDT).

A MDT constitui-se na extração de conhecimento novos, úteis e implícitos entre seqüências e sub-seqüências de eventos [3]. A figura 1 apresenta a MDT dentro do processo de descoberta temporal. A entrada de dados do MDT são seqüências temporais e o resultado do processo são regras temporais que interpretados transformam-se em padrões temporais.

---

<sup>1</sup> Evento é uma ocorrência importante em um BD.

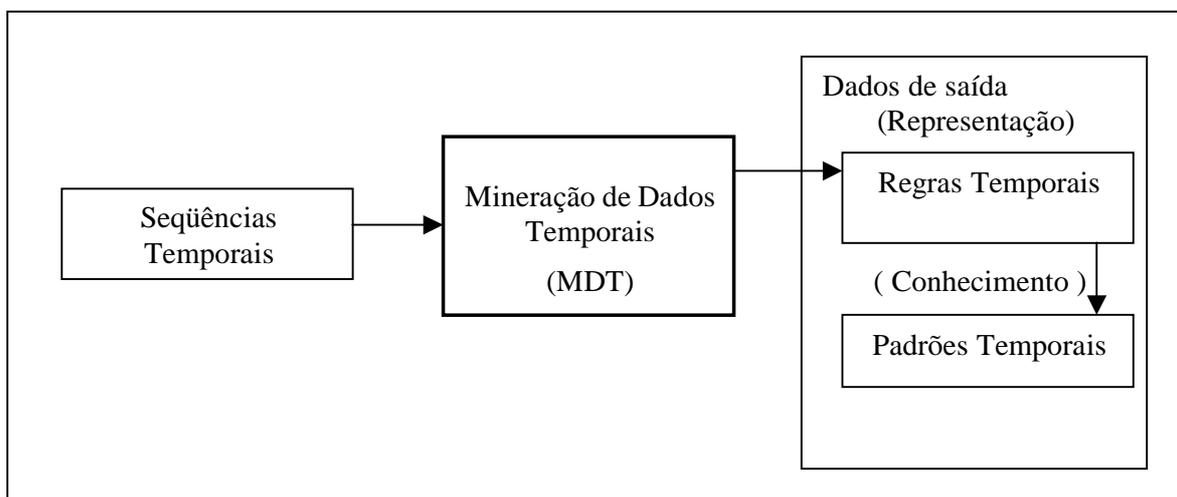


FIGURA 1: A MDT dentro do processo de descoberta temporal.

A *seqüência temporal* pode ser expressa como uma lista de eventos ordenados de acordo com o tempo dos eventos, sendo que cada evento ( $E_n$ ) corresponde a um conjunto de elementos [3]. Desta forma, a seqüência temporal = {conjunto de elementos ( $E_1$ ), conjunto de elementos ( $E_2$ ), conjunto de elementos ( $E_3$ ),...conjunto de elementos ( $E_n$ )}. O problema existente está em descobrir todas as seqüências dos dados para os registros do BD e neste conjunto de seqüências procurar por um padrão freqüente nos dados[2].

Observa-se, que o conjunto de elementos dos eventos é um conjunto de elementos que existem, ou seja, elementos não-vazios e não-nulos. A partir disto, têm-se que uma seqüência corresponde a uma lista de elementos ordenados. Sem perda de generalização, assume-se que o conjunto de elementos é mapeado para um conjunto de números inteiros. Conforme a teoria dos conjuntos, um dado conjunto A é subconjunto de B, se e somente se, todo elemento de A pertence também a B, representado por:  $A \subseteq B$ . Assim, uma seqüência\_a = { $a_1, a_2, a_3, \dots, a_n$ } está contida em outra seqüência\_b = { $b_1, b_2, b_3, \dots, b_n$ } se, na seqüência\_b existirem elementos inteiros, tais que,  $e_1 < e_2 < e_3 < \dots < e_n$ , de forma que,  $a_{e_1} \subseteq b_{e_1}, a_{e_2} \subseteq b_{e_2}, a_{e_3} \subseteq b_{e_3}, \dots, a_{e_n} \subseteq b_{e_n}$  [3].

A partir de então, observa-se que a seqüência temporal caracteriza-se por apresentar três componentes fundamentais que são: duração, simultaneidade e intervalo temporal nas relações entre os eventos.

A duração refere-se ao período total de tempo observado para a descoberta, como, por exemplo, o experimento efetuado com as internações hospitalares do ano de 2000.

A simultaneidade trata de eventos que ocorrem em um espaço temporal específico e são analisados como ocorrências simultâneas. Um exemplo para esta característica é verificada ao observar-se que, um paciente com dor no braço esquerdo pouco tempo depois tem um ataque cardíaco. Estes dois eventos são tratados como ocorrências simultâneas em análises de seqüência.

O intervalo temporal diz respeito ao tempo decorrido entre os eventos de uma seqüência. Por exemplo, a internação de um paciente com Colecistite Aguda<sup>2</sup> é seguida por uma internação para Colectomia<sup>3</sup> após três meses. Estes dois eventos consecutivos ocorrem em uma seqüência com um determinado intervalo.

A **MDT** têm por finalidade encontrar padrões previsíveis de comportamento ao longo de um período de tempo. Através da utilização da MDT pode-se descobrir que determinado procedimento sempre precede outro, ou que determinado procedimento só pode ser repetido após um intervalo mínimo de tempo atendendo a uma periodicidade. Assim, um certo comportamento estará mais suscetível de produzir um outro comportamento ou uma seqüência de comportamentos em um determinado período de tempo.

Pode-se discernir o estudo do aspecto temporal da MDT em duas direções conforme o tipo de dado e a descoberta que se pretende efetuar.

Uma direção utiliza dados discretos ou categorizados e a extração do conhecimento efetuada é realizada sobre uma seqüência temporal. Um exemplo de uma seqüência temporal é representado na definição de uma seqüência de doenças diagnosticadas como, por exemplo, {Malária, Leptospirose, Meningite}. O modelo de conhecimento extraído a partir de um padrão seqüencial é representado por uma regra temporal. Uma regra temporal indica que a ocorrência de um evento é seguida pela ocorrência de outro evento.

A outra direção, diz respeito aos dados dispostos em uma seqüência contínua de elementos de valores reais que é conhecida por uma série temporal. Estes valores são medidos nos mesmos intervalos temporais. Um exemplo de uma série temporal são as médias de temperatura atmosférica de uma determinada região = {23.9, 24.4, 26.1, 26.2, 26.3, 26.5}. Ela efetua a descoberta de padrões similares entre seqüências temporais semelhantes ou diferentes. A descoberta de padrões similares também é conhecida por análise de tendências. A área da MDT que estuda a identificação de seqüências similares<sup>4</sup> processa-se através de padrões pré-especificados, como os utilizados no processamento de sinais ou *strings*.

Acrescenta-se a isto, que ambos os aspectos temporais, mas sobretudo, as séries temporais estão mais sujeitas ao aumento do volume dos dados, e por esta razão, ao problema da alta-dimensionalidade dos dados. Para solucionar este problema, a MDT necessita que todos os dados estejam amostrados e convertidos em uma seqüência de dados discretos. Desta maneira, é escolhida uma certa granularidade temporal<sup>5</sup> mínima, isto permite diminuir a quantidade dos dados através da conversão de toda a massa de dados para esta granularidade. Assim, os padrões temporais são discretizados e designados de seqüências temporais.

O problema da descoberta em seqüências temporais consiste em encontrar dentre todos os padrões temporais que atendam ao percentual mínimo de suporte para ser apresentado pela seqüência temporal, aquela seqüência candidata que corresponda a seqüência máxima. A seqüência máxima é uma seqüência que está contida em outra seqüência [2].

Os **padrões temporais** são as relações de dependência em dados dos BD onde existe uma associação temporal nos eventos em uma ordem determinada. Esta associação temporal nos eventos

---

<sup>2</sup> Colecistite Aguda é a inflamação na vesícula biliar.

<sup>3</sup> Colectomia é o procedimento médico de retirada da vesícula biliar.

<sup>4</sup> Seqüências similares são ocorrências de seqüências semelhantes em um BD.

<sup>5</sup> Granularidade temporal consiste na duração do tempo de um evento como, por exemplo, horas ou dias.

mostra que os dados temporais podem estar ordenados em uma ou mais dimensões do tempo, pois a ordenação destes dados relacionados leva à formação de novas seqüências temporais. Desta forma, múltiplas dimensões do tempo são admissíveis para um mesmo evento, caso um sistema de MDT acomode diversas linhas de tempo<sup>6</sup> apresentando informações armazenadas conforme o tempo de validade, o tempo de transação ou o tempo de decisão [6].

## 2.2 O Processo de Descoberta Temporal

O processo para a descoberta temporal, conforme a figura 2 a seguir, está dividido em três fases que são: pré-processamento, MDT e pós-processamento.

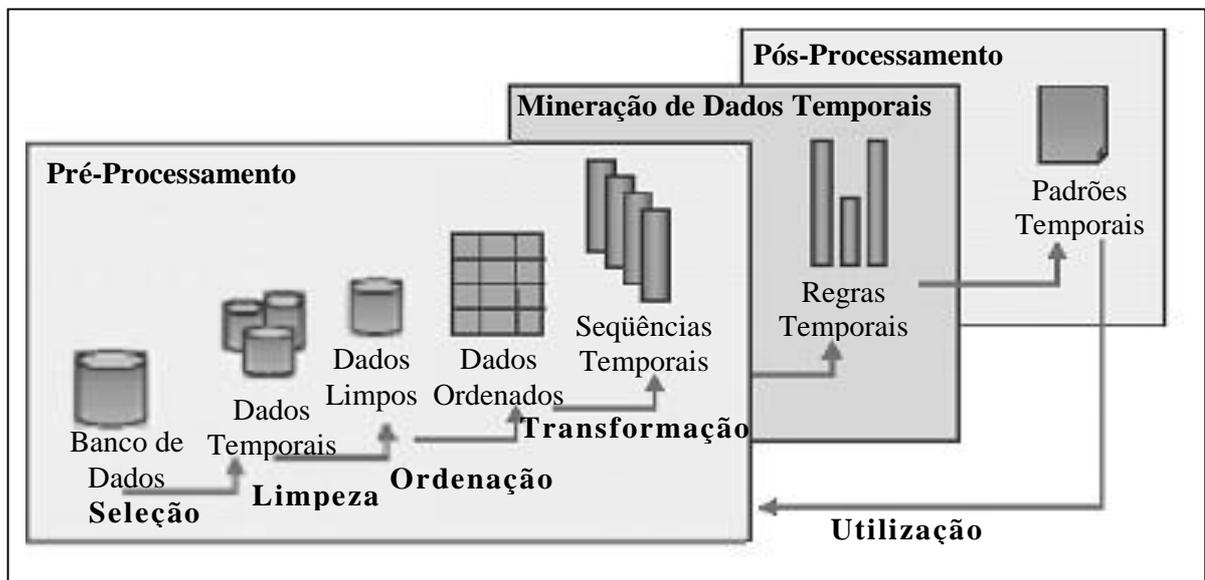


FIGURA 2 – Etapas do Processo de Descoberta do Conhecimento.  
Fonte: FAYYAD, 1996 p. 10 com adaptações.

A fase do pré-processamento que utiliza 80% do esforço no processo é composta pelas etapas: seleção, limpeza, ordenação e transformação.

Na seleção, conforme a tabela 1, um BD (D) original é selecionado. São selecionados os campos: identificação, associação temporal e valor.

<sup>6</sup> Linha de tempo é uma seqüência discreta de pontos consecutivos.

TABELA 1 - BD original (D).

| ORDENAÇÃO | IDENTIFICAÇÃO | ASSOCIAÇÃO<br>TEMPORAL | VALOR |
|-----------|---------------|------------------------|-------|
| 5         | 2             | 04/01/02               | Z     |
| 24        | 1             | 01/01/01               | B     |
| 963       | 1             | 04/01/02               | F     |
| 1000      | 1             | 02/01/02               | C     |
| 6897      | 2             | 03/01/02               | C     |
| 6989      | 2             | 01/01/02               | F     |
| 11030     | 1             | 01/01/02               | A     |
| 50974     | 2             | 02/01/02               | X     |
| 78451     | 2             | 01/01/02               | A     |
| 99999     | 1             | 03/01/02               | D     |

A etapa da limpeza objetiva aumentar a qualidade dos dados selecionados tratando as irregularidades existentes em BD e garantindo a consistência dos dados, como, por exemplo, promovendo o tratamento dos valores nulos.

Na ordenação, os campos identificação e associação temporal são escolhidos como chaves: primária e secundária respectivamente. Este procedimento, implicitamente, converte o BD original em um BD de identificadores de seqüências (L) observado na tabela 2.

TABELA 2 - BD de identificadores de seqüências (L).

| ORDENAÇÃO | IDENTIFICAÇÃO | ASSOCIAÇÃO<br>TEMPORAL | VALOR |
|-----------|---------------|------------------------|-------|
| 24        | 1             | 01/01/01               | B     |
| 11030     | 1             | 01/01/02               | A     |
| 1000      | 1             | 02/01/02               | C     |
| 99999     | 1             | 03/01/02               | D     |
| 963       | 1             | 04/01/02               | F     |
| 6989      | 2             | 01/01/02               | F     |
| 78451     | 2             | 01/01/02               | A     |
| 50974     | 2             | 02/01/02               | X     |
| 6897      | 2             | 03/01/02               | C     |
| 5         | 2             | 04/01/02               | Z     |

Após a ordenação dos dados temporais, identificam-se todos os elementos (l) do BD (L) conforme a tabela 3. Nesta fase, simultaneamente, encontram-se os conjuntos de todos os elementos de maiores seqüências expressos por  $\{ \langle l \rangle \mid l \in L \}$ . Este mapeamento é necessário para a identificação e comparação de  $I_1 = \langle B, A, C, D, F \rangle$  e  $I_2 = \langle F, A, X, C, Z \rangle$ .

TABELA 3 - Identificação das seqüências do BD.

| IDENTIFICAÇÃO | VALOR                           |
|---------------|---------------------------------|
| 1             | $\langle B, A, C, D, F \rangle$ |
| 2             | $\langle F, A, X, C, Z \rangle$ |

Na transformação, determinam-se os maiores agrupamentos contidos no BD de identificadores de seqüências. Cada identificador de seqüência pode ser transformado em uma representação alternativa discretizada conforme a tabela 4. Um identificador de seqüência é representado por uma lista de valores expressos por  $\{I_1, I_2, I_3, \dots, I_n\}$ . A transformação efetuada sobre o BD (D) é representada por  $D_T$ .

TABELA 4 - Transformação das seqüências do BD.

| Seqüências  | $D_i$ |
|-------------|-------|
| <B,A,C,D,F> | <A,C> |
| <F,A,X,C,Z> | <A,C> |

A fase da MDT utiliza o conjunto de valores para encontrar as seqüências candidatas à seqüência máxima conforme a tabela 5.

TABELA 3 – Seqüência máxima extraída do BD.

|  |
|--|
| Padrão seqüencial temporal<br>com suporte = 100% |
| <b>(A,C)</b>                                     |

A fase do pós-processamento realiza a análise das regras temporais obtidas na MDT caracterizando-as em: conhecidas; novas; desconhecidas, interessantes e úteis; desconhecidas, desinteressantes e inúteis por não agregarem conhecimento algum a base de conhecimento.

### 3 Experimento

A realização deste trabalho constituiu-se em um experimento sobre o BD da Secretaria Estadual da Saúde do Rio Grande do Sul (SES). Este experimento objetiva descobrir relacionamentos expressos sob a forma de regras temporais. A partir disto, o modelo de conhecimento obtido descreve propriedades que representam ocorrências freqüentes em padrões de relacionamentos dos dados.

O BD da SES armazena as Autorizações de Internação Hospitalar (AIH). A estrutura interna das AIH contém informações relativas a identificação do paciente como, por exemplo, o Cadastro Nacional de Pessoas Físicas, e as características da internação hospitalar efetuada apresentando a data da internação e os procedimentos médicos realizados.

#### 3.1 Pré-processamento

O entendimento do domínio do problema compreendeu várias entrevistas com auditores médicos, especialistas da SES. As atividades efetuadas nesta fase objetivaram a aquisição do conhecimento onde processou-se a imersão do analista na área da qual efetuou-se a MDT.

Esta fase contribui para a identificação dos objetivos do processo do ponto de vista de quem utilizará o resultado da descoberta. Estes objetivos se distinguem em: verificação e descoberta. Na verificação, o sistema está limitado em verificar as hipóteses do usuário. No descobrimento, o sistema descobre novos padrões. Assim, este experimento procura apresentar tradicionais comportamentos dos tratamentos nas AIH, detectar situações anômalas, bem como, acompanhar a evolução das doenças. Para isto, a fase do pré-processamento compreendeu as etapas de seleção, limpeza, ordenação e transformação dos dados.

Na fase da seleção dos dados, foram escolhidos os dados do movimento das AIH de 2000 porque estes dados apresentam: um atributo que compreende diversos eventos de identificação como, o Cadastro Nacional de Pessoas Físicas; um campo que compreende vários valores como, o atributo data de internação do paciente e um atributo valor que representa o procedimento realizado pelo paciente.

A etapa da limpeza objetivou aumentar a qualidade dos dados selecionados. A princípio, utilizou-se o cadastro nacional de pessoas físicas com atributo principal. Entretanto, observou-se que o arquivo de movimento das AIH apresentava muitos registros que não dispunham desta informação ou continham a mesma informação para mais de um paciente. Isto impossibilita acompanhar o tratamento de um paciente. A solução adotada constituiu-se da exclusão dos registros que apresentavam estas irregularidades. Entretanto, observou-se que o volume do BD resultante foi significativamente reduzido. Esta significativa redução do volume dos dados de entrada refletiu na qualidade das regras temporais extraídas. Desta maneira, o problema da grande perda de informação devido a escolha impraticável do atributo principal foi tratada com a mudança da seleção do atributo principal para o nome do paciente. Este campo foi concatenado com a data de nascimento do paciente para evitar problemas de homonímia. Assim, através da escolha do nome do paciente<sup>7</sup> como atributo principal é possível resolver o problema de acompanhamento do tratamento dos pacientes. Todavia, pacientes idênticos com variação na digitação do nome constituem-se em pacientes diferentes. Adicionalmente, observou-se que existem AIH com o mesmo número, em diferentes períodos, devido as reinternações e transferências de pacientes de outros hospitais que geram duplas ou triplas de registros de um mesmo paciente. A solução adotada foi a exclusão dos registros duplicados.

Após a limpeza dos atributos pacientes, internações e procedimentos realizados, foi efetuada a ordenação das seqüências existentes. Para tanto, o arquivo foi ordenado pelos pacientes seguido das suas internações.

A seguir, os dados ordenados são transformados em seqüências temporais. Esta etapa foi efetuada na ferramenta *Intelligent Miner* da IBM [8], devido ao grande volume de dados a ser transformados e por esta ser uma tarefa totalmente automatizada nesta ferramenta.

### 3.2 MDT

A fase da MDT também foi efetuada na ferramenta *Intelligent Miner* da IBM. Nesta fase, foram utilizados os mapeamentos de nomes para atribuir nomes mais significativos aos valores do atributo elementar. Para tanto, os arquivos de descrição dos procedimentos realizados e o Cadastro Internacional de Doenças (CID) foram utilizados para tornar mais representativo o conhecimento extraído a partir dos procedimentos realizados. O *Intelligent Miner* da IBM implementa o algoritmo *Apriori* [2] para a MDT.

---

<sup>7</sup> Utilizou-se o nome do paciente transformado por um algoritmo de criptografia para evitar sua identificação real.

O tempo decorrido durante a fase de MDT para processar os 568.833 registros foi de cinquenta e cinco segundos.

### **3.3 Pós-processamento**

Na fase do pós-processamento, foram extraídos os resultados das descobertas efetuadas sobre os dados reais. O conhecimento obtido foi representado na forma de regras temporais.

#### **3.3.1 Resultados obtidos**

Foram descobertas várias regras temporais para os procedimentos realizados. Todavia, a seguir, são apresentadas três regras temporais obtidas devido a sua importância de interpretação:

A primeira regra temporal extraída apresenta que o procedimento realizado de Doença Pulmonar Obstrutiva Crônica é seguido pelo mesmo procedimento realizado com suporte de 1,103% entre os 459.189 pacientes encontrados no BD. Isto equivale a 5.064 casos.

A interpretação da primeira regra temporal está diretamente relacionada ao caráter crônico da patologia que faz com que um paciente sofra seguidas internações com o mesmo procedimento realizado. Adicionalmente, observa-se que esta regra temporal apresenta o procedimento realizado mais freqüente no Estado do RS de acordo com os auditores da SES.

A segunda regra temporal apresenta Colecistite Aguda (inflamação na Vesícula Biliar) seguida de Colisistectomia (retirada da Vesícula biliar) com suporte de 0,088% entre os 459.189 pacientes do BD da SES. Isto equivale a 404 casos.

A interpretação para a segunda regra temporal extraída consiste em um padrão temporal de comportamento para o tratamento da doença de Colecistite Aguda. Ela constitui-se em um indicativo para os gestores da SES. Uma doença como esta, assim que diagnosticada, deve ser encaminhada a um procedimento de cirurgia eletiva de Colisistectomia, pois a seguida internação só acarreta sofrimento desnecessário ao paciente e ônus ao sistema.

A terceira regra temporal apresenta o procedimento realizado de Colecistite Aguda seguido do mesmo procedimento realizado com suporte de 0,031% entre os 459.189 paciente do BD da SES. Isto equivale a 142 casos investigados.

Na interpretação desta terceira regra temporal extraída, observa-se que esta regra temporal tem um comportamento diferente do padrão temporal já encontrado na segunda regra temporal e deve ser analisada pelos analistas da SES.

#### **3.3.2 Verificação e Validação**

A seguir, são apresentadas a análise segundo os especialistas da SES e a análise pessoal a partir do conhecimento extraído.

De acordo com os especialistas da SES, mais de 70 % das internações efetuadas dizem respeito ao baixo nível sócio-econômico dos pacientes, falta de políticas de prevenção de doenças, insuficiência de atenção básica, deficiência no acompanhamento dos pacientes no posto de saúde e carência de saúde básica da população.

A tabela 4 apresenta a análise dos padrões temporais extraídos segunda a interpretação de auditores de S.E.Se classifica o conhecimento descoberto em já conhecido ou desconhecido.

TABELA 4 - Verificação pelos auditores da SES do conhecimento extraído

| Comportamento dos procedimento realizados pelos pacientes nas AIH de 2000.  | Verificação do conhecimento |              |
|---|-----------------------------|--------------|
|   | Conhecido                   | Desconhecido |
| Um procedimento realizado de Doença Pulmonar Obstrutiva Crônica é seguido pelo mesmo procedimento até nove (09) internações em 82 casos analisados. | X                           |              |
| Um procedimento realizado de Insuficiência Cardíaca é seguido pelo procedimento de Acidente Vaso-Cerebral Agudo em 280 casos analisados.            |                             | X            |
| Um procedimento realizado de Pneumonia em Lactente é seguido pelo procedimento realizado de Entero Infecções em Lactente em 234 casos analisados.   | X                           |              |

Observam-se nas regras extraídas o crescimento das doenças do aparelho respiratório e cardiovasculares.

#### 4 Conclusão

Este trabalho objetivou aplicar as técnicas de MDT em uma área de grande importância como a saúde. No decorrer do experimento efetuado, procedeu-se a adaptação da técnica dos padrões temporais para a aplicação do movimento das AIH onde obteve-se a extração implícita, não trivial, nova e potencialmente utilizável da informação dos dados considerando seus aspectos temporais. Um exemplo disto, são os relacionamentos extraídos dos dados nas internações de pacientes e os procedimento médicos efetuados. Estes relacionamentos apresentam o mundo real armazenado em seus dados e representam um conhecimento valioso acerca do BD.

Analisando-se os dados disponíveis, verificou-se que os algoritmos de MDT puderam encontrar padrões passíveis de serem utilizados para predizer quais procedimentos médicos oferecem seqüência nas AIH, e qual grau de incerteza devem ser considerados para autorizar a internação nos hospitais.

Observou-se que a principal característica da MDT é a presença de um domínio temporal dinâmico onde os dados são atualizados em BD regulares. Desta maneira, foi interessante examinar a forma como, tanto os dados, quanto o conhecimento foram derivados através das mudanças do tempo. Verificou-se que para descobrir padrões de seqüências temporais sobre grandes BD, os algoritmos propostos geraram uma seqüência dos dados para os registros do BD e procuram, neste conjunto das seqüências, por um padrão seqüencial freqüente. Assim, a descoberta de tendências, ciclos e padrões nos BD e a sua utilização em análises históricas e na previsão de dados futuros é identificável quando se detecta padrões seqüenciais, tendências aparentes, processos que realçam a troca, e a suas periodicidades nas informações do BD.

Ao longo deste trabalho, pôde-se perceber que as técnicas de DCBD realmente são excelentes para descobrir conhecimento novo e oculto em grandes BD. Durante a etapa de pré-processamento no processo de DCBD se confirmou que ela realmente demanda o maior tempo de um trabalho de DCBD. Acrescenta-se que um tópico que deve ser alvo de estudos futuros dentro da DCBD constitui-se em estudar e pesquisar um método de validação de regras de conhecimento descoberto, no pós-processamento, pois esta etapa é de fundamental importância e muito pouco foi encontrado, sobre a mesma, na bibliografia.

De tudo o que se expôs acima, conclui-se que a MDT sempre tem como um de seus resultados, novas perguntas. Observa-se que o tomador de decisão precisa ter acesso aos recursos da exploração de dados para atender a esta necessidade de iteração. Este trabalho não se justifica apenas pela aplicação de técnicas de aprendizado para análise de dados sobre o aspecto temporal, mas a organização em questão necessita aprender automaticamente por si própria. Dentre os problemas que perturbam e fascinam a sociedade da informação estão a superabundância de dados e como gerenciá-los. Assim, as organizações deverão encontrar o seu próprio caminho para julgar e descobrir conhecimento, e a atividade da MDT prova ser uma parte crucialmente importante neste processo.

## **5 Referências Bibliográficas**

- [1] AGRAWAL, R. et al. Fast Algorithms for Mining Association Rules. VLDB (490-501). Santiago, Chile, 1994.
- [2] AGRAWAL, R. and SRIKANT, R. Mining Sequential Patterns. Proc. of the Int'l Conference on Data Engineering (ICDE), Taipei, Taiwan, March 1995. Expanded version appears as "IBM Research Report RJ9910". October 1994.
- [3] AGRAWAL, R. et al. Mining Sequential Patterns: Generalizations and Performance Improvements. IBM Almaden Research Center. CA: San Jose. 1996.
- [4] ALVARES, Luís Otávio Campos. Ferramentas de Inteligência Artificial. Porto Alegre: Instituto de Informática da UFRGS, 1999. Apostila da disciplina.
- [5] BIGOLIN, Nara M. Data Mining: Conceitos e Técnicas. Escola de Informática da SBC Sul. Maio, 2000. P. 233-249.
- [6] EDELWEISS, N.; PALAZZO, J.M.O. Modelagem de Aspectos Temporais de Sistemas de Informações. Recife: IX Escola de Computação, 1994.
- [7] FAYYAD, U.M., PIATETSKY-SHAPIRO, G., and SMYTH, P. From Data Mining To Knowledge Discovery: An Overview. In Advances In Knowledge Discovery And Data Mining, eds. U.M. Fayyad, G. Piatetsky-Shapiro, P. Smyth, and R. Uthurusamy, AAAI Press/The MIT Press, Menlo Park, CA., 1996, pp. 1-34.
- [8] IBM DB2. Utilizando o Intelligent Miner for Data. Versão 6. Release 1. Edição S517-6338-00. © Copyright International Business Machines Corporation 1996,1999.