

## Descoberta de padrões para identificação de casos de alto custo em operadoras de planos de saúde

### Discovering standards for identifying high-cost cases in health insurance companies

Leandro Fabian Almeida Escobar  
Universidade Positivo – UP – Curitiba – Brasil  
[leandro.escobar@up.edu.br](mailto:leandro.escobar@up.edu.br)

Marcelo Rosano Dallagassa  
Pontifícia Universidade Católica do Paraná – PUCPR – Curitiba – Brasil  
[marcelo.dallagassa@pucpr.br](mailto:marcelo.dallagassa@pucpr.br)

William Felipe Sokoloski  
Pontifícia Universidade Católica do Paraná – PUCPR – Curitiba – Brasil  
[willian.fsokoloski@gmail.com](mailto:willian.fsokoloski@gmail.com)

Denise Tsunoda  
Universidade Federal do Paraná – UFPR – Curitiba – Brasil  
[dtsunoda@gmail.com](mailto:dtsunoda@gmail.com)

Deborah Ribeiro Carvalho  
Pontifícia Universidade Católica do Paraná – PUCPR – Curitiba – Brasil  
[ribeiro.carvalho@pucpr.br](mailto:ribeiro.carvalho@pucpr.br)

#### Resumo

Apresenta-se a construção de um modelo para a identificação de casos de alto custo, aplicando o processo de *Knowledge Discovery in Databases* apoiados na literatura sobre a necessidade de aplicações sobre técnicas e algoritmos para a identificação de itens de interesse em grande volume de dados para o apoio à tomada de decisões em saúde. Como método propõe-se algoritmos de regra de associação com derivações em relação a medidas de interesses sequencial e temporal. Para validação e teste da técnica aplica-se um experimento sobre 16.355 liberações de procedimentos médicos, tendo sido descobertas 12.992 regras de associação, sendo 12.305 regras sequenciais e 5.554 padrões na medida temporal. O protótipo desenvolvido foi validado por seis especialistas, entre eles profissionais da enfermagem, analistas de sistemas e um médico, que foram capacitados para o seu uso, avaliando os padrões descobertos e a própria ferramenta, então, responderam a um questionário. As respostas foram tabuladas e o resultado foi obtido mediante ao índice de validade de conteúdo. O modelo alcançou uma aceitação geral na ordem de 78% indicando uma aceitação parcial pelos especialistas.

**Palavras-chave:** inteligência artificial, mineração de dados, sistemas de apoio a decisões administrativas, sistemas de identificação de pacientes.

#### Abstract

We present the design of a model for the identification of high cost cases, applying the process of Knowledge Discovery in Databases supported in the literature by the need of applications on techniques and algorithms for the identification of items of interest in high volume of data for the support decision-making in health. As method, association rule algorithms with derivations were proposed in relation to sequential and temporal measures. For validation and testing an experiment is applied on 16,355 releases of medical procedures, and 12,992 rules of association were discovered, with 12,305 sequential rules and 5,554 patterns in temporal measure. The prototype developed was validated by 6 specialists, nursing professionals, system analysts and a physician, who were trained for their use by evaluating the discovered patterns and the tool, and then answered a questionnaire. The responses were tabulated, and the result was obtained using the content validity index. The model reached an overall acceptance of around 78% indicating partial acceptance by the specialists.

**Keywords:** artificial intelligence, data mining, decision support systems management, patient identification systems.

## 1. Introdução

Não apenas na área da saúde, mas também de forma geral, a tecnologia da informação favorece o crescimento do volume de dados armazenados, o que não facilita a sua utilização a partir de estratégias convencionais. Processos que automatizem a extração de informações oportunas passam a ser imprescindíveis, dada a dificuldade em analisar grandes volumes de dados (TRINDADE, 2005; KOBUS, 2006; SANTANA et al., 2006; LOPES, 2007; CARVALHO et al., 2015).

Dessa maneira há uma necessidade urgente da utilização de técnicas analíticas avançadas para prover os especialistas que explorem e detectem itens de interesse em grande volume de dados, combinando o potencial da análise estatística, com a descoberta de padrões de forma ágil, para auxiliar à tomada de decisão em saúde (TRINDADE, 2005; WANG et al., 2018).

Neste contexto, o processo de descoberta de conhecimento em bases de dados (KDD – *Knowledge Discovery in Databases*) (FAYYAD et al., 1996), apresenta notório potencial em ampliar a produção de conhecimento em vários domínios de negócios, inclusive na área da saúde (SHILLABEER, 2005), uma vez que a descoberta de padrões pode levar a identificação de informações relevantes, fornecendo modelos para o planejamento de ações preventivas ou para a tomada de decisão quanto à população (RAMON et al., 2007; CARVALHO et al., 2015; MEYFROIDT et al., 2009; WANG et al., 2018), aumentando a eficácia de tratamentos e políticas de prevenção e promoção à saúde, bem como a eficiência na aplicação de recursos em tais iniciativas (DALLAGASSA, 2009)

Ainda destaca-se que, a identificação de situações de interesse, por meio do processo de KDD em contas médicas de uma operadora de plano de saúde (OPS) permitem avaliar de maneira dinâmica a ocorrência do ciclo de utilização de novas tecnologias em saúde e analisar a sua efetividade e possíveis mudanças de seu comportamento (ARAYA et al., 2016).

No entanto, obstante a existência de estudos científicos quanto ao uso do KDD na saúde, sua adoção no processo cotidiano ainda é incipiente dada a pouca familiaridade dos especialistas quanto à metodologia, também como suas vantagens e dificuldades (MARISCAL et al., 2010; MEYFROIDT et al., 2009). Em geral são adotadas apenas estratégias estatísticas com o objetivo de revelar relações lineares entre os fatores de saúde (CRUZ-RAMÍREZ et al., 2013) ou mesmo relatos comparando as estratégias de mineração de dados, sem priorizar as etapas relativas à interpretação e avaliação de

resultados (BLONBERG, 2010).

Assim, este trabalho tem como objetivo de conceber um modelo e o desenvolvimento de um protótipo que evidencie a identificação antecipada de possíveis casos de alto custo utilizando-se de dados administrativos de processamento de contas médicas, de forma a apoiar o planejamento de ações preventivas e otimizar o processo de auditoria de procedimentos, contribuindo para o ganho de eficiência na aplicação de recursos em organizações em saúde.

O modelo proposto pretende contribuir para a melhora de análise em grande volume de bases de dados, mediante a mineração de dados e pós-processamento de padrões de maneira eficiente e possibilitando o aumento da capacidade em elaborar planos de promoção da saúde e de prevenção e tratamento de doenças, o que se reflete em eficiência para as organizações de saúde e melhor atendimento às necessidades de pacientes e beneficiários.

Tradicionalmente, decidir é optar pela alternativa mais satisfatória entre várias e estabelecer uma série de ações para a realização de tal alternativa, resolvendo, assim, um determinado problema (SHIMIZU, 2010).

SAD – Sistema de Apoio à Decisão – compreende equipamentos, softwares, especialistas e metodologias para assistir os gestores nos seus processos decisórios (SPRAGUE, et al., 1991; TURBAN; ARONSON 2001; TURBAN et al., 2009). Permite acesso interativo aos dados, mediante modelos que ajudem a análise adequada e permitam formular decisões mais eficientes e mais eficazes (TURBAN et al., 2009).

SAD em saúde constitui valiosa contribuição para a acurácia do diagnóstico, da qualidade e da eficiência da identificação de casos para tratamentos específicos (MANGIAMELI et al., 2004; KOBUS, 2006; CARVALHO, et al., 2015; WANG et al., 2018). Tais sistemas facilitam localizar, recuperar e filtrar as informações necessárias para um desempenho mais efetivo, apoiando a integração de informações relevantes, representando, assim, efetivo suporte ao processo decisório, dado o volume de dados a ser manipulado pelos especialistas (PATEL et al., 2013).

Entretanto, o processo de tomada de decisão na área da saúde apresenta desafios adicionais, ou seja, contexto dinâmico, corpo de conhecimento crescente, envolvendo múltiplas variáveis e indivíduos (WANG et al., 2018). Decisões em saúde são constituídas por ciclos decisão-ação onde as situações evoluem e durante as quais decisões e ações se influenciam mutuamente. A decisão não representa apenas um resultado, mas também um componente das atividades desenvolvidas.

A decisão em saúde pode ser dividida em decisão clínica curativa, preventiva e decisão para aplicação de recursos (MANGIAMELI et al., 2004; KOBUS, 2006; SANTANA et al., 2006; CARVALHO, et al., 2015, NÉRI et al., 2011).

A decisão clínica curativa está relacionada com diagnóstico, a determinação do tratamento a ser habilitado e com a segurança do paciente durante tal tratamento. Busca-se a resposta necessária uma vez que a doença está instalada e identificada. Para tanto, o médico coleta dados do paciente de maneira a formular um diagnóstico e definir um conjunto de medidas curativas (MANGIAMELI et al., 2004; KOBUS, 2006; NÉRI et al., 2011; LIN et al., 2017).

A decisão preventiva trata do mapeamento da situação sanitária, indicando a prevalência de doenças e de riscos à saúde atual e futura de indivíduos ou de uma determinada população, servindo como base para a elaboração de estratégias de prevenção e promoção da saúde (KOBUS, 2006; CARVALHO et al., 2015) e reduzindo falhas e antecipando a intervenção dos casos (LIN et al., 2017).

A decisão para aplicação de recursos trata do problema de escassez de recursos, uma vez que o agravamento das doenças relativas ao trabalho e ao envelhecimento da população requerem mais eficiência, exigindo que as organizações de saúde conheçam profundamente os ciclos de atendimento, prevenção de doenças, diagnósticos e

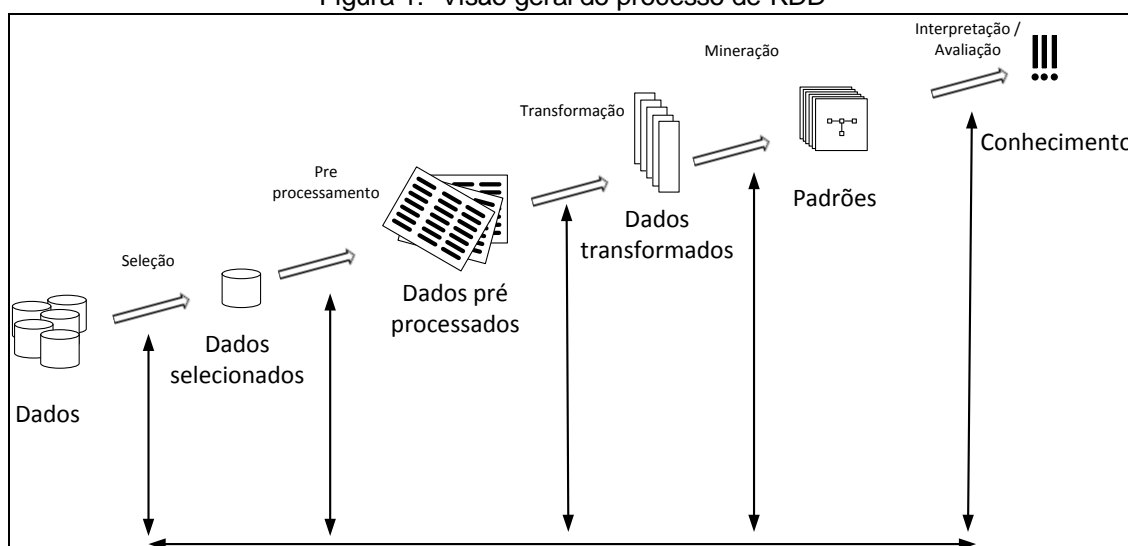
gerenciamento de casos como bases para o aumento da qualidade no atendimento e a redução dos custos (MANGIAMELI et al., 2004; KOBUS, 2006; SANTANA et al., 2006; CARVALHO et al., 2015).

Compreender como a tomada de decisão em saúde se dá, suas características e os domínios nos quais os diferentes tipos estão inseridos permite determinar quais estratégias podem ser adotadas para a elaboração de sistemas.

O modelo proposto é aplicado à identificação de beneficiário de alto custo, contribuindo para a eficiência tanto do processo de reconhecimento de casos, quanto para as ações decorrentes. Assim é dirigido a uma melhor aplicação dos recursos de saúde e qualidade do atendimento para o paciente e para tanto se faz necessário aplicar os conceitos que possam solucionar a questão, entre eles o KDD.

O KDD é um processo iterativo, iterativo e não trivial de descoberta de informações, mediante a identificação de padrões válidos, novos, potencialmente úteis e compreensíveis (FAYYAD et al., 1996), conforme as etapas demonstradas na Figura 1.

Figura 1: Visão geral do processo de KDD



Fonte: Fayyad et al. (1996)

Também é possível entender o KDD como um conjunto de atividades organizadas em três grandes etapas: pré-processamento, mineração e pós-processamento.

O pré-processamento visa realizar a limpeza dos ruídos, preenchimento de dados em branco, remoção de inconsistências, tratamento de dados ausentes, limpeza, integração, redução ou a organização dos dados por conta de relações cronológicas (HAN; KAMBER, 2006).

A mineração de dados compreende tarefas que buscam descrever os padrões ou prever uma classe previamente identificada (CARVALHO, 2005; HAN; KAMBER, 2006; BELLAZZI; ZUPAN, 2008; MILOVIC; MILOVIC, 2012).

Entre as alternativas de predição se destaca a classificação, que obtém modelos ou funções que preveem um objetivo (considerado atributo meta ou classe) a partir dos dados e que determina a qual classe, dentre várias, uma nova instância pertence (HAN; KAMBER, 2006; MILOVIC; MILOVIC, 2012).

Nas tarefas descritivas tem-se a descoberta de regras de associação e o agrupamento. O agrupamento utiliza-se de medidas de similaridade de maneira a reorganizar o espaço de representação dos dados e, desta forma, induzindo diferentes classes. O resultado é um conjunto finito de grupos que ocupam o mesmo espaço de classificação (CHEN et al., 1996; CARVALHO, 2005).

A associação é uma tarefa de obtenção de padrões descritivos denominados regras de associação que demonstram relações frequentes entre os itens de dados,

sendo representadas na forma de regras do tipo se X então Y ( $X \rightarrow Y$ ). Sendo X e Y conjuntos de itens de dados do que atendem a seguinte propriedade  $X \cap Y = \emptyset$ . X representa o antecedente da regra e Y o conseqüente da regra. Para cada regra de associação obtida, são determinadas medidas denominadas suporte e confiança (AGRAWAL; SRIKANT, 1994).

A presente pesquisa adota o seguinte suporte (Expressão 1) (BORGELT, 2010).

$$\text{suporte} = \frac{X}{T} \quad (1)$$

A confiança (Expressão 2) é representada por (AGRAWAL e SRIKANT, 1994):

$$\text{confiança} = \frac{XY}{X} \quad (2)$$

A interpretação das regras de associação, bem como o interesse por valores de suporte e confiança maiores ou menores, deve considerar a situação problema que originou o uso da mineração de dados. Não existe uma regra geral que remeta que valores de suporte e confiança devam ser altos ou baixos.

Também pode ocorrer a necessidade de pós-processar os padrões obtidos originalmente, depurando-os ou sintetizando-os (CARVALHO et al., 2012) ou descobrindo exceções aos padrões obtidos (HUSSAIN et al., 2000).

Roddick e Spiliopoulou (1999) discutem a ordenação de eventos em cadeias causais, identificando padrões em que o evento X leva ao evento Y, se aproximando ao que se denomina mineração temporal, em que são descobertos padrões ordenados na linha do tempo.

Gomes e Carvalho (2011) propõem um método híbrido para a extração de regras sequenciais, tendo desenvolvido o algoritmo *Chrono\_Assoc* (GOMES; CARVALHO, 2011), que trata como entrada as regras de associação descobertas pelo algoritmo *Apriori* (AGRAWAL e SRIKANT, 1994) e determina, então, novas medidas de suporte e confiança cronológicas.

Segundo Gomes e Carvalho (2011) o suporte cronológico trata a frequência de ocorrências do antecedente e do conseqüente nesta ordem *Is*, em relação ao número total de transações *T* na base (Expressão 3). A confiança cronológica representa a frequência dos itens da regra tomados em sequência *Is* e em relação ao conjunto total de regras que possuem tais itens (Expressão 4).

$$\text{SuporteCronológico} = \frac{Is}{T} \quad (3)$$

$$\text{ConfiançaCronológica} = \frac{Is}{I} \quad (4)$$

- *Is* é o número de registros nos quais ocorrem o antecedente e conseqüente, nessa ordem;
- *T* é o total de transações na base de dados;
- *I* é o total de transações na base que possuem tanto o antecedente quanto o conseqüente, em qualquer ordem.

Desta forma, uma regra é considerada sequencial quando seu suporte cronológico e sua confiança cronológica são superiores a 0 (zero).

Sokoloski et al. (2014) desenvolvem o algoritmo *Assoctemp*, para pós-processar regras de associação avaliando a sequência e os intervalos de tempo relativos à ocorrência de eventos. Compõem a o resultado do processamento medidas de suporte e confiança temporais, atribuídas às regras descobertas, demonstrando a frequência dos

eventos dentro da janela de tempo de interesse.

O suporte temporal<sub>1</sub> (Expressão 5) demonstra a proporção de transações que contém o antecedente e o consequente, respectivamente,  $I_s$  em relação ao número de transações  $T$  na base. Esta métrica retrata a frequência de transações às quais o padrão temporal se aplica.

$$\text{SuporteTemporal}_1 = \frac{I_s}{T} \quad (5)$$

- $I_s$  é o número de registros nos quais ocorrem o antecedente e consequente, nessa ordem);
- $T$  é o total de transações na base de dados.

A confiança temporal<sub>1</sub> (Expressão 6) representa a proporção de transações que possuem o antecedente e o consequente dentro da janela temporal de interesse  $I_j$ , respectivamente, em relação ao número transações que contém o antecedente e o consequente, respectivamente,  $I_s$ .

$$\text{ConfiançaTemporal}_1 = \frac{I_j}{I_s} \quad (6)$$

- $I_j$  é o número de registros nos quais o antecedente e consequente dentro do intervalo de tempo (janela temporal);
- $I_s$  é o número de registros nos quais ocorrem o antecedente e consequente, nessa ordem).

O suporte temporal<sub>2</sub> (Expressão 7) representa a razão entre o número de aparições de sequências – antecedente e consequente, respectivamente  $I_{as}$  e o total de transações  $T$  na base. Esta métrica retrata a densidade do padrão temporal em relação às transações, dado que uma transação pode conter a mesma sequência de eventos mais de uma vez.

$$\text{SuporteTemporal}_2 = \frac{I_{as}}{T} \quad (7)$$

- $I_{as}$  é o número de aparições da sequência nas transações em quais ocorrem o antecedente e consequente;
- $T$  é o total de transações na base de dados.

A confiança temporal<sub>2</sub> (Expressão 8) representa a razão entre o número de aparições de sequências dentro da janela temporal de interesse  $I_{sj}$  e o número de aparições de sequências  $I_{as}$ .

$$\text{ConfiançaTemporal}_2 = \frac{I_{sj}}{I_{as}} \quad (8)$$

- $I_{sj}$  é o número de aparições das sequencias dentro da janela de tempo dentro das transações;
- $I_{as}$  é o total de aparições da sequência na base de dados.

O pós-processamento permite, assim, facilitar a avaliação e obter outros padrões a partir dos padrões originalmente obtidos, aumentando as chances de identificar aqueles que venham a ser úteis e agreguem conhecimento para o especialista (CARVALHO et al., 2012).

A interpretação dos resultados da mineração de dados garante que conhecimentos úteis sejam derivados tanto dos dados quanto dos padrões encontrados neles (FAYYAD et al., 1996), o que requer recursos de visualização e navegação, a fim de aumentar sua inteligibilidade. Deve-se permitir e facilitar que o especialista experimente os resultados

obtidos frente ao seu contexto (MILOVIC; MILOVIC, 2012).

## 2. Método

Trata-se de uma pesquisa qualitativa, descritiva, aplicada e de raciocínio indutivo para propor um modelo baseado em KDD, que permita a descoberta e visualização dos padrões descobertos de uma tarefa aplicada na gestão da saúde.

As necessidades específicas dos especialistas foram levantadas preliminarmente, mediante regras de associação descobertas com o algoritmo *Apriori* (AGRAWAL; SRIKANT, 1994), pós-processando padrões sequenciais, com o uso do *Chrono\_Assoc* (GOMES; CARVALHO, 2011) e dos padrões de janelas de tempo com o do *Assoctemp* (SOKOLOSKI et al., 2014) no intuito de obter a descrição dos padrões de ocorrência de procedimentos em beneficiários de alto custo.

Para tal experimento preliminar, utilizou-se uma base de dados contendo 55.722 liberações de procedimentos realizados entre janeiro de 2006 e janeiro de 2014 relativas a 248 beneficiários. Foram obtidas 13.649 regras de associação, pós-processadas, em conjunto com os dados originais, mediante o algoritmo *Chrono\_Assoc* (GOMES; CARVALHO, 2011) e com o *AssocTemp* (SOKOLOSKI et al., 2014).

Para viabilizar a análise de regras com potencial interesse, o especialista indicou oito procedimentos de interesse (Quadro 1), que representam tratamentos de alto custo, servindo como filtros para as análises dos padrões obtidos.

Os dados coletados foram sistematizados mediante regras de associação, indicando quais aspectos citados nos estudos incluídos são relacionados entre si, e pós-processados para obtenção de regras de exceção, de forma a indicar quais aspectos devem ser considerados no conjunto de requisitos para o desenvolvimento de soluções de mineração de dados para a saúde.

Quadro 1: Frequência absoluta dos procedimentos de interesse indicados por especialista

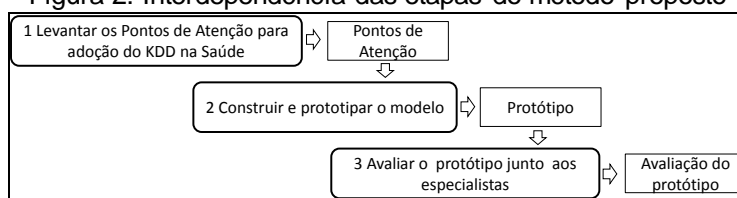
Procedimentos	Total
CONSULTA EM PRONTO SOCORRO	2207
ACIDO LACTICO (LACTATO)	60
ANTIGENO CARCINOEMBRIOGENICO (CEA), DOSAGEM	164
GASOMETRIA (PH, PCO2, SA, O2, EXCESSO BASE)	10
HERCEPTIN - 440mg	453
MARCADORES TUMORAIS (CA 19.9, CA 125, CA 72-4, CA 15-3, ETC.) CADA, DOSAGEM	611
PROTEINA C, DOSAGEM	362
TORAX - 1 INCIDENCIA	237

Fonte: Os autores (2018)

Organizou-se, então, um modelo para leitura, processamento, pós-processamento, armazenamento e navegação em padrões relacionados a eventos de alto custo em saúde, a partir do qual se desenvolveu um protótipo, de maneira a demonstrar o processo para obtenção e uso dos padrões e, ainda, permitir a avaliação do modelo quanto à sua capacidade para o uso na rotina em saúde. Finalizou-se com a avaliação do modelo proposto e sistematização dos resultados obtidos.

As etapas metodológicas desta pesquisa estão conectadas criando uma relação de interdependência baseada nos seus resultados preliminares, conforme exibido na Figura 2.

Figura 2: Interdependência das etapas do método proposto



Fonte: Os autores (2018)

O critério de seleção dos pontos de atenção foi a partir do foco do modelo que é a identificação dos padrões de eventos de alto custo. O quadro 2 contém os pontos de atenção relacionados com o foco do modelo.

Quadro 2: Frequência dos pontos de atenção associados com auxílio ao planejamento em saúde

Ponto de Atenção	Frequência absoluta	Frequência relativa
Descrição de eventos ocorridos	41	72%
Avaliação da qualidade dos dados	24	42%
Fornecimento de subsídios para explicações causais	20	35%
Avaliação (subjetiva ou objetiva) da relevância dos padrões encontrados	11	19%
Representação visual dos resultados	7	12%

Fonte: Adaptado de Carvalho, Escobar e Tsunoda (2014)

Assim, os requisitos foram divididos em dois grupos. Os pontos de atenção, oriundos do trabalho de Carvalho et al. (2014) e requisitos funcionais, derivados diretamente dos especialistas envolvidos (Quadro 3).

Quadro 3: Requisitos funcionais em função aos grupos do estudo

Grupo	Requisitos
Grupo 1: Pontos de Atenção	PA01 - Desenvolvimento de funcionalidade específica
	PA02 - Representação visual dos resultados
	PA03 - Descrição de eventos ocorridos
	PA04 - Fornecimento de subsídios para explicações causais
Grupo 2 – Requisitos funcionais	RF01 - Carregar dados desnormalizados com os registos de eventos oriundos das operações da empresa.
	RF02 – Mapear os atributos que representam os beneficiários, procedimentos e as datas de ocorrência.
	RF03 - Gerar lista classificada de acordo com a frequência dos eventos da base de dados.
	RF04 - Obter padrões de associação, cronologia e janelas de tempo.
	RF05 - Navegar nos padrões obtidos <ul style="list-style-type: none"> <li>- filtrar regras de associação que possuem o evento selecionado no antecedente, no conseqüente ou ambos;</li> <li>- exibir todas as regras ou somente implicações;</li> <li>- exibir os eventos concorrentes com o evento selecionado.</li> </ul>
	RF06 - Exibir os dados relativos aos beneficiários, filtrados



de acordo com os padrões selecionados.

Fonte: Os autores (2018)

O modelo proposto foi avaliado tendo em vista o seu potencial para uso na rotina em saúde, mediante um instrumento de avaliação aplicação durante uma sessão conjunta com os especialistas. A avaliação abordou quatro construtos e 13 questões, refletindo as dimensões do modelo a serem avaliadas pelos especialistas envolvidos (Quadro 4).

Quadro 2: Construtos e respectivas questões para a avaliação do modelo proposto

Construto	Questões
Satisfação do especialista quanto ao modelo	Q1. O sistema mostra informações claramente e é fácil compreendê-las
	Q2. Aprendi rapidamente a utilizar a interface de navegação do sistema
	Q3. Utilizaria o sistema em minha rotina de trabalho
Apresentação dos padrões e informações	Q4. A frequência dos eventos é útil para a análise dos padrões
	Q5. Os padrões apresentados são claros e facilmente compreendidos
	Q6. Os dados apresentados sobre os beneficiários são úteis e relevantes para meu trabalho
	Q7. A disposição dos padrões e informações é adequada para o uso em minha rotina
Capacidade de descrever os eventos	Q8. Os padrões apresentados esclarecem como os eventos estão relacionados.
	Q9. As janelas de tempo permitem compreender como os eventos estão relacionados.
	Q10. Compreendi claramente os padrões apresentados e como eles descrevem os eventos ocorridos.
Perfil dos entrevistados	Q11. Sua área de formação/especialidade
	Q12. Há quanto tempo atua com auditoria?
	Q13. Utiliza ferramentas estatísticas para a análise de procedimentos a serem auditados?

Fonte: Os autores (2018)

As repostas possíveis consideraram a escala psicométrica de Likert (1932), com três possibilidades variando desde a negação total até a aceitação da informação, no qual cada resposta recebe um peso relativo à sua posição na escala de 0 a 2, onde:

- Discordo, peso = 0;
- Sou Indiferente, peso = 1;
- Concordo, peso = 2.

O formulário do perfil dos entrevistados, contou com respostas diversas à escala Likert (1932), conforme exposto no Quadro 5.

Quadro 3: Chaves de respostas para as questões relativas ao perfil dos entrevistados

Questão	Chave de respostas
Q11. Área de formação/especialidade	Resposta aberta.
Q12. Há quanto anos atua com auditoria?	____ anos.
Q13. Utiliza ferramentas estatísticas para a análise de procedimentos a serem auditados?	( ) Sim ( ) Não Se respondeu não, descreva como realiza a análise de procedimentos.

Fonte: Os autores (2018)

A aplicação da avaliação se deu mediante a experimentação do modelo construído e a aplicação de formulário de pesquisa em sessão coletiva, envolvendo seis especialistas em auditoria de procedimentos de operadora de saúde, que experimentaram o protótipo e os padrões obtidos por ele e, então, responderam ao questionário.

A avaliação da aceitação do modelo proposto foi realizada mediante a tabulação das respostas e da determinação do índice de validade de conteúdo (IVC), proposto por Waltz et al (1991, apud DALLAGASSA, 2009, p. 58) conforme o seguinte processo:

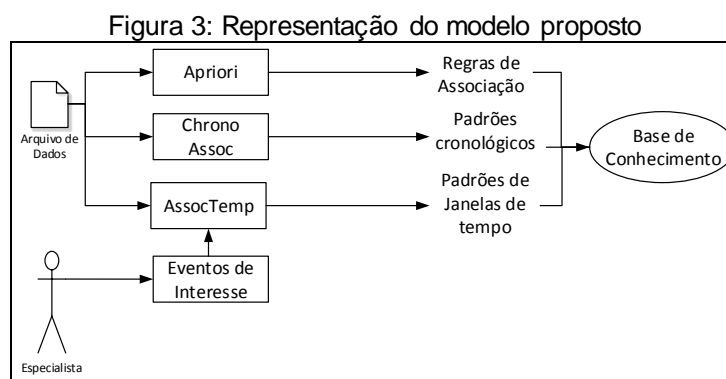
- Somar os pesos obtidos em cada construto, com base nas respostas dos especialistas;
- O total máximo possível para cada construto é obtido somando os máximos

- pesos de cada resposta relativa ao construto em questão;
- c) O IVC é obtido dividindo-se o total obtido pelo total máximo possível, multiplicado por 100 (Expressão 9).

$$IVC_{Construto} = \left( \frac{Total\ obtido_{Construto}}{Total\ máximo\ possível_{Construto}} \right) 100 \quad (9)$$

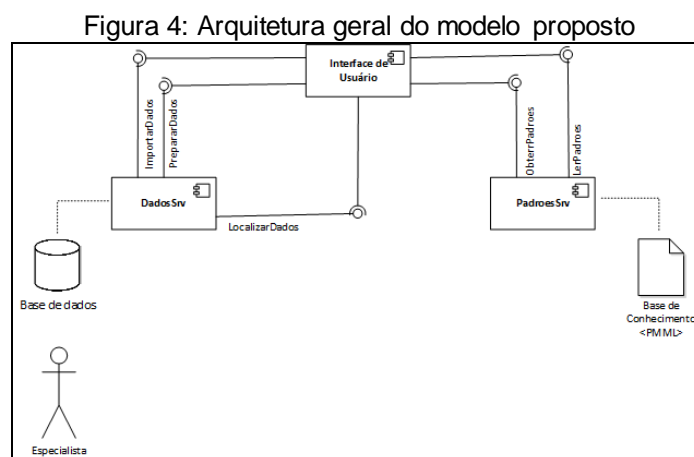
Para fins de avaliação, valores obtidos abaixo de 70% foram considerados como não aceitos, entre 70% e 80% foram considerados como parcialmente aceitos e acima de 80% foram considerados totalmente aceitos.

Um arquivo de dados, contendo os eventos e as datas de liberação para cada beneficiário da operadora, serve de entrada para o modelo, que obtém os padrões de regras de associação com o algoritmo *Apriori* (AGRAWAL; SRIKANT, 1994) e implementado no programa de mesmo nome – *Apriori.exe* – (BORGELT, 2010), os padrões cronológicos com o *Chrono\_Assoc* (GOMES; CARVALHO, 2011). O algoritmo *AssocTemp* (SOKOLOSKI et al., 2014) é acionado, utilizando tanto o arquivo de dados quanto os eventos de interesse. Os padrões são gravados na base de conhecimento do modelo e ficam disponíveis para análise pelos especialistas (Figura 3).



Fonte: Os autores (2018)

Estão definidos dois serviços de alto nível, dados e padrões cujas responsabilidades são acesso aos dados, tanto para leitura quanto para gravação, carga e acesso aos padrões obtidos, respectivamente. Além dos dois serviços de alto nível, o modelo conta com a interface de usuário e com um banco de dados, para o armazenamento dos dados de entrada, bem como os padrões e demais parâmetros para funcionamento (Figura 4).

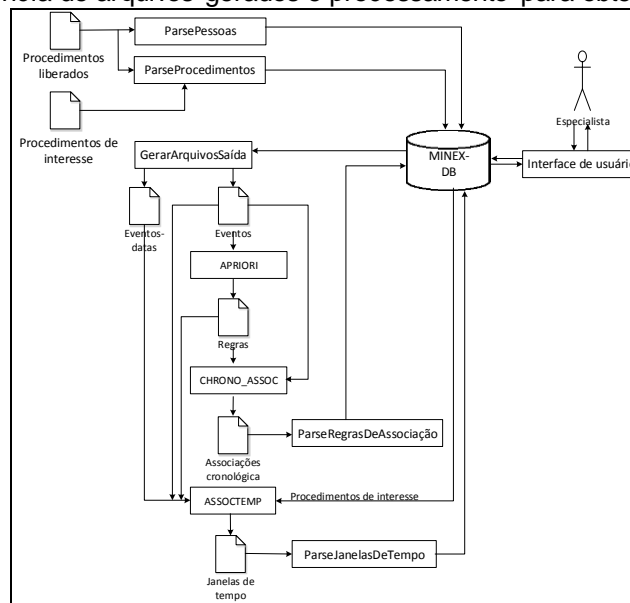


Fonte: Os autores (2018)

A partir da importação inicial dos arquivos de dados, o modelo automatiza a obtenção das regras de associação com o *Apriori* (AGRAWAL; SRIKANT, 1994), pós-

processa os padrões obtendo regras sequenciais com o *Chrono\_Assoc* (GOMES; CARVALHO, 2011) e, também obtém as janelas de tempo com o *AssocTemp* (SOKOLOSKI et al., 2014), mediante a geração de arquivos específicos e a execução de cada algoritmo, conforme demonstrado (Figura 5).

Figura 5: Sequência de arquivos gerados e processamento para obtenção dos padrões



Fonte: Os autores (2018)

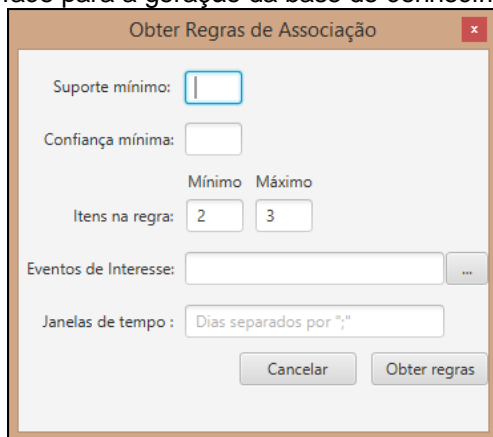
Para a obtenção dos padrões, o modelo gera dois arquivos de saída a partir dos dados importados para sua base. O arquivo eventos contém os procedimentos liberados para os beneficiários e é utilizado como entrada para o *Apriori* (AGRAWAL; SRIKANT, 1994). O arquivo eventos-datas contém os mesmos procedimentos, mas acompanhados das datas de realização, de forma a servir de entrada para o *AssocTemp* (SOKOLOSKI et al., 2014).

Os arquivos associações cronológicas e janelas de tempo são lidos e, então, as regras de associação, padrões de sequência e padrões de janelas de tempo são gravados na base de dados para uso posterior. A interação do usuário com o modelo proposto se dá mediante um protótipo desenvolvido e em três momentos. Primeiramente, o arquivo com os procedimentos deve ser importado para a base de dados do modelo. Então, os padrões devem ser descobertos, momento em que a base de conhecimento é gravada. Após a gravação da base de conhecimento, os especialistas podem navegar nos padrões a fim de explorá-lo.

A interface de usuário do modelo tem ênfase na facilidade de navegação, de maneira que seu uso seja simples e contribua para a eficiência das análises dos padrões. Assim, o modelo possui três janelas. Uma para a obtenção das regras de associação e para pós-processamento dos padrões e uma janela para análise dos padrões e uma que exibe detalhes a ficha médica com os procedimentos realizados por um dado beneficiário.

Os padrões são obtidos mediante a tela “Obter Regras de Associação” (Figura 6), na qual o usuário indica o suporte e confiança mínima; o número mínimo e máximo de itens nas regras para o *Apriori* (AGRAWAL; SRIKANT, 1994); o caminho do arquivo contendo os eventos de interesse e os intervalos de tempo para o *AssocTemp* (SOKOLOSKI et al., 2014).

Figura 6: Interface para a geração da base de conhecimento do modelo

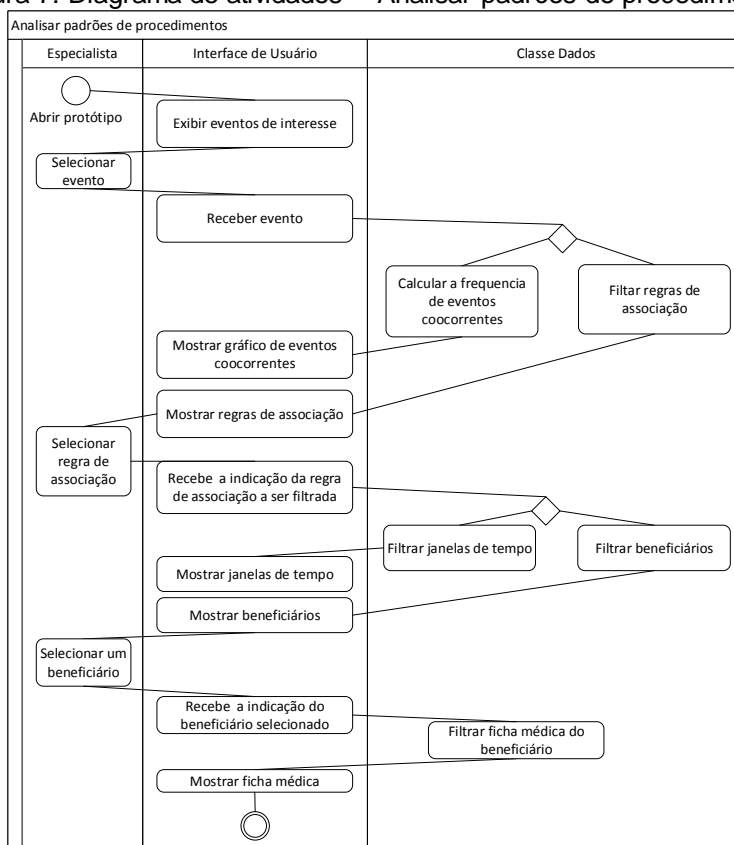


Fonte: Os autores (2018)

Ao clicar no botão “Obter regras”, o modelo gera os arquivos de entrada para os algoritmos *Apriori* (AGRAWAL; SRIKANT, 1994), *Chrono\_Assoc* (GOMES; CARVALHO, 2011) e *AssocTemp* (SOKOLOSKI et al., 2014), inicia cada um dos algoritmos e grava os padrões obtidos na base de conhecimento, disponibilizando-os para análise. É importante citar que esta operação pode ser realizada uma só vez ou então, cada vez que o especialista deseje renovar a base de conhecimento.

A análise dos padrões, por sua vez, se dá mediante uma sequência de interações entre o especialista e o modelo, descritas no diagrama de atividades da Figura 7.

Figura 7: Diagrama de atividades – Analisar padrões de procedimentos



Fonte: Os autores (2018)

A interface para análise dos padrões permite ao especialista selecionar o procedimento que deseja analisar, apresenta a frequência dos procedimentos que ocorrem em conjunto com a seleção, as regras de associação, padrões de sequência e os intervalos de tempo, bem como os beneficiários que estão cobertos pela regra de associação selecionada pelo especialista (Figura 8).

Adicionalmente, o especialista pode configurar como o modelo se comporta, filtrando regras nas quais o procedimento selecionado esteja presente no antecedente ou no conseqüente, exibindo somente as regras que representam implicações, em que o antecedente vem antes do conseqüente, exibir as regras que não possuem o procedimento selecionado ou definindo se o modelo deve exibir os dados filtrados automaticamente ou se o especialista deseja aplicar os filtros manualmente.

Nessa interface, o especialista pode verificar a frequência dos eventos encontrados na base e quais são os eventos que ocorrem em conjunto com o evento selecionado concorrentes. Também é possível filtrar as regras de associação e os padrões cronológicos, de maneira a melhor evidenciar os padrões obtidos. Os dados dos beneficiários envolvidos nos eventos são exibidos à medida que os padrões são selecionados, detalhando as informações que atendem a tais padrões.

O especialista pode selecionar um dos beneficiários cobertos pela regra de associação selecionada e exibir os detalhes relativos aos procedimentos realizados.

Figura 8: Interface do usuário do modelo

Fonte: O autor (2018)

### 3. Resultados

Com o intuito de avaliar o modelo proposto, a base de conhecimento foi previamente carregada, permitindo que especialistas naveguem nos padrões. Para tanto, foram importados 16.355 registros de liberações de procedimentos relativos a 127 beneficiários e contendo 2.083 procedimentos realizados entre agosto de 2008 e outubro de 2014.

A base de conhecimento foi carregada com 12.992 regras de associação com suporte mínimo 5%, de maneira que associações menos frequentes sejam obtidas, e confiança mínima de 80%, buscando regras cuja associação sejam fortes e, portanto,

relevantes.

O protótipo obteve 12.305 regras de associação sequenciais, restando 687 regras de associação não sequenciais. O especialista indicou seis procedimentos de interesse para obtenção das janelas de tempo (Quadro 6), levando em conta intervalos de 90, 120 e 150 dias, de acordo com o processo específico já adotado na organização, resultando em 5.554 padrões de janelas de tempo.

Quadro 4: Procedimentos de interesse indicado pelo especialista

ANTÍGENO CARCINOGENICO (CEA), DOSAGEM
GASOMETRIA (PH, PCO2, SA, O2, EXCESSO BASE)
RX – 1 INCIDENCIA
MARCADORES TUMORAIS (CA 19.9, CA 125, CA 72-4, CA 15-3), DOSAGEM
CONSULTA EM PRONTO SOCORRO
ÁCIDO LÁTICO, DOSAGEM

Fonte: Os autores (2018)

Para exemplificar os padrões obtidos pelo modelo, enfatiza-se a leitura que pode ser realizada pelo especialista durante a navegação nas regras de associação, nas seqüências e nos intervalos de tempo.

No Quadro 7 são exemplificadas duas regras sequenciais, cujo suporte e confiança cronológicos são maiores que 0 (zero) e duas regras de associação não sequenciais, cujos suporte e confiança cronológicos são 0 (zero).

Quadro 7: Exemplos de regras de associação sequenciais e não sequenciais obtidas pelo modelo

Antecedente		APRIORI	Suporte		SEQUENCIAL
DOSAGEM DE COLESTEROL E GAMA-GLUTAMIL			6,3%	Confiança	
Consequente	ÁCIDO LÁTICO	CHRONO-ASSOC	Suporte cronológico	2,4%	SEQUENCIAL
			Confiança cronológica	42,9%	
Antecedente	LUVA CIRÚRGICA E GASOMETRIA	APRIORI	Suporte	6,3%	NÃO SEQUENCIAL
			Confiança	85,5%	
Consequente	ÁCIDO LÁTICO	CHRONO-ASSOC	Suporte cronológico	0,0%	NÃO SEQUENCIAL
			Confiança cronológica	0,0%	
Antecedente	ZOFRAN E DOSAGEM DE FOSFATASE ALCALINA	APRIORI	Suporte	5,5%	SEQUENCIAL
			Confiança	85,7%	
Consequente	ANTÍGENO CARCINOGENICO	CHRONO-ASSOC	Suporte cronológico	0,8%	SEQUENCIAL
			Confiança cronológica	16,7%	
Antecedente	HEPAMAX E GAMA-GLUTAMIL TRANDErase	APRIORI	Suporte	5,5%	NÃO SEQUENCIAL
			Confiança	87,5%	
Consequente	ANTÍGENO CARCINOGENICO	CHRONO-ASSOC	Suporte cronológico	0,0%	NÃO SEQUENCIAL
			Confiança cronológica	0,0%	

Fonte: Os autores (2018)

A leitura dos padrões obtidos segue uma seqüência específica de interpretações, cujos exemplos são dados a seguir: a regra de associação “Se DOSAGEM DE TRANSAMINASE OXALACETICA e GASOMETRIA então ÁCIDO LÁTICO” como exemplo, interpreta-se (Quadro 8):

- a) Pelo suporte e confiança cronológicos: 393 (suporte cronológico = 2,4%) beneficiários realizaram DOSAGEM DE TRANSAMINASE OXALACETICA e GASOMETRIA antes e, destes, 131 (confiança cronológica = 33,3%) realizaram ACIDO LÁTICO na sequência;
- b) Pelos intervalos de tempo: Prevalece até 90 dias entre a ocorrência de “DOSAGEM DE TRANSAMINASE OXALACETICA e GASOMETRIA” e “ÁCIDO LÁTICO” (confiança no intervalo = 100%).

Quadro 8: Padrões descobertos para a regra de associação

Antecedente	DOSAGEM DE TRANSAMINASE OXALÉTICA E GASOMETRIA	ASSOCIAÇÃO		Suporte	7,9%
				Confiança	90,0%
		SEQUÊNCIA		Suporte cronológico	2,4%
				Confiança cronológica	33,3%
Consequente	ÁCIDO LÁTICO	JANELAS DE TEMPO	90 DIAS	Suporte	3,9%
				Confiança	100,0%
			120 DIAS	Suporte	3,9%
				Confiança	100,0%
			150 DIAS	Suporte	3,9%
	Confiança	100,0%			

Fonte: Os autores (2018)

A regra de associação “Se CULTURA AUTOMATIZADA e DOSAGEM DE FOSFATASE ALCALINA então ÁCIDO LÁTICO” como exemplo, interpreta-se (Quadro 9):

- a) De acordo com suporte e confiança cronológicos: 131 (suporte cronológico = 0,8%) beneficiários realizaram CULTURA AUTOMATIZADA e DOSAGEM DE FOSFATASE ALCALINA antes e, destes, 19 (confiança cronológica = 14,3%) realizaram ACIDO LÁTICO na sequência;
- b) Pelos intervalos de tempo: Prevalece até 150 dias entre a ocorrência de “CULTURA AUTOMATIZADA e DOSAGEM DE FOSFATASE ALCALINA” e “ÁCIDO LÁTICO” (confiança no intervalo = 80%).

Quadro 9: Padrões descobertos para a regra de associação

Antecedente	CULTURA AUTOMATIZADA E DOSAGEM DE FOSFATASE ALCALINA	ASSOCIAÇÃO		Suporte	6,3%
				Confiança	85,5%
		SEQUÊNCIA		Suporte cronológico	0,8%
				Confiança cronológica	14,3%
Consequente	ÁCIDO LÁTICO	JANELAS DE TEMPO	90 DIAS	Suporte	3,9%
				Confiança	80,0%
			120 DIAS	Suporte	3,9%
				Confiança	80,0%
			150 DIAS	Suporte	3,9%
	Confiança	80,0%			

Fonte: Os autores (2018)

A regra de associação “Se MARCADORES TUMORAIS e DOSAGEM DE FOSFATASE ALCALINA então ANTÍGENO CARCINOEMBRIOGÊNICO” como exemplo, interpreta-se (Quadro 10):

- a) Pelo suporte e confiança cronológicos: 393 (suporte cronológico = 2,4%) beneficiários realizaram MARCADORES TUMORAIS e DOSAGEM DE FOSFATASE ALCALINA antes e, destes, 147 (confiança cronológica = 14,3%) realizaram ANTÍGENO CARCINOEMBRIOGÊNICO na sequência;
- b) Pelos intervalos de tempo: Prevalece até 90 dias entre a ocorrência de “MARCADORES TUMORAIS e DOSAGEM DE FOSFATASE ALCALINA” e

“ANTÍGENO CARCINOEMBRIOGÊNICO” (confiança no intervalo = 100%).

Quadro 10: Padrões descobertos para a regra de associação

Antecedente	MARCADORES TUMORAIS E DOSAGEM DE FOSFATASE ALCALINA	ASSOCIAÇÃO		Suporte	7,1%
				Confiança	88,9%
		SEQUÊNCIA		Suporte cronológico	2,4%
				Confiança cronológica	37,5%
Antecedente	ANTÍGENO CARCINOEMBRIOGÊNICO	JANELAS DE TEMPO	90 DIAS	Suporte	3,9%
			120 DIAS	Confiança	100,0%
				Suporte	3,9%
			150 DIAS	Confiança	100,0%
		150 DIAS	Suporte	3,9%	
Confiança	100,0%				

Fonte: Os autores (2018)

A regra de associação “Se DOSAGEM DE UREIA e ZOFRAM então ANTÍGENO CARCINOEMBRIOGÊNICO” como exemplo, interpreta-se (Quadro 11):

- Pelo suporte e confiança cronológicos: 131 (suporte cronológico = 0,8%) beneficiários realizaram DOSAGEM DE UREIA e ZOFRAM antes e, destes, 22 (confiança cronológica = 16,7%) realizaram ANTÍGENO CARCINOEMBRIOGÊNICO na sequência;
- Pelos intervalos de tempo: Prevalece até 90 dias entre a ocorrência de “DOSAGEM DE UREIA e ZOFRAM” e “ANTÍGENO CARCINOEMBRIOGÊNICO” (confiança no intervalo = 75%), sendo que o intervalo máximo para estas sequências é de 120 dias (confiança no intervalo de 120 dias).

Quadro 11: Padrões descobertos para a regra de associação

Antecedente	DOSAGEM DE URÉIA E ZOFRAM	ASSOCIAÇÃO		Suporte	5,5%
				Confiança	85,7%
		SEQUÊNCIA		Suporte cronológico	0,8%
				Confiança cronológica	16,7%
Consequente	ANTÍGENO CARCINOEMBRIOGÊNICO	JANELAS DE TEMPO	90 DIAS	Suporte	3,1%
			120 DIAS	Confiança	75,0%
				Suporte	3,1%
			150 DIAS	Confiança	100,0%
		150 DIAS	Suporte	3,1%	
Confiança	100,0%				

Fonte: Os autores (2018)

Das dez questões avaliativas, três não foram aceitas, uma foi aceita parcialmente e seis foram totalmente aceitas (Quadro 12)

Quadro 12: Pontuação obtida nas questões da avaliação do modelo proposto

Questão	Pontos
Utilizaria o sistema em minha rotina de trabalho	100%
Os dados apresentados sobre os beneficiários são úteis e relevantes para o meu trabalho	92%
A frequência dos eventos é útil para a análise dos padrões	92%
O sistema mostra informações claramente e é fácil de compreendê-las	92%
Os padrões apresentados esclarecem como os eventos estão relacionados	83%
A disposição dos padrões e informação é adequada para o uso	83%



em minha rotina	
Os padrões apresentados são claros e facilmente compreendidos	75%
Compreendi claramente os padrões apresentados e como eles descrever os eventos	67%
Apreendi rapidamente a utilizar a interface de navegação do sistema	58%
As janelas de tempo permitem compreender como os eventos estão relacionados	42%

Fonte: Os autores (2018)

As questões que obtiveram as menores pontuações foram “As janelas de tempo permitem compreender como os eventos estão relacionados”, 42%; “Apreendi rapidamente a utilizar a interface de navegação do sistema”, 58% e “Compreendi claramente os padrões apresentados e como eles descrevem os eventos ocorridos”, 67%, sendo avaliadas como “não validadas” pelos especialistas.

Para essas questões, os especialistas solicitaram a inclusão dos valores nominais representados pelos padrões. Isto sugere a necessidade de relacionar as quantidades de beneficiários às quais os padrões se referem, o que facilitaria a aceitação do modelo no tocante aos quesitos relacionados com as questões que não foram validadas.

A questão “Os padrões apresentados são claros e facilmente compreendidos” obteve 75% da pontuação, sendo avaliada como parcialmente validada. Neste ponto, os especialistas também solicitaram a inclusão dos valores nominais aos quais os percentuais dos padrões se referem. Os especialistas também declararam dificuldades em interpretar os intervalos de tempo, solicitando que os intervalos não se sobreponham, mas sejam expressos como “até 90 dias”, “de 91 a 120 dias” e “de 121 a 150 dias”, demonstrando os intervalos de forma mais clara e intuitiva e, por isso, não exigindo explicações adicionais para seu entendimento. O uso dos códigos dos procedimentos no lugar de seus títulos também foi criticado pelos especialistas.

Embora as questões “A disposição dos padrões e informações é adequada para o uso em minha rotina” e “Os padrões apresentados esclarecem como os eventos estão relacionados” tenham obtido 83% dos pontos, e por isso, consideradas validadas, o fato de estarem próximas a 80% traz a discussão quanto ao design de interação do protótipo desenvolvido para o usuário, que pode ter sido a causa das notas intermediárias.

Partindo desses resultados, é possível considerar que o modelo não foi capaz de atender com qualidade ao ponto de atenção “descrição de eventos ocorridos”, especialmente quando se trata da facilidade de interpretação dos padrões. O detalhamento das informações conforme solicitado pelos especialistas e a otimização da navegação, de forma que o especialista perceba melhor onde deve selecionar os elementos da interface gráfica podem compensar a baixa aceitação relativa a estes aspectos.

Por outro lado, as questões que obtiveram as melhores pontuações e, por isso, validadas, foram: “O sistema mostra informações claramente e é fácil compreendê-las”, 92%; “A frequência dos eventos é útil para a análise dos padrões”, 92%; “Os dados apresentados sobre os beneficiários são úteis e relevantes para meu trabalho”, 92% mostram que o modelo foi capaz de atender ao ponto de atenção “representação visual dos resultados”, pois os especialistas declararam facilidade em compreender as informações exibidas.

Muito embora a pontuação total do modelo proposto tenha sido 78%, determinando uma aceitação parcial, a questão “Utilizaria o sistema em minha rotina de trabalho”, obteve 100% dos pontos e mostra que a combinação dos pontos de atenção selecionados para o modelo, juntamente com a implementação dos requisitos levantados, reforçou o potencial de utilidade do modelo proposto, confirmando sua capacidade de clareza, a relevância dos padrões e a clareza das informações. Finalmente, todos os especialistas

entrevistados utilizariam o modelo proposto em suas rotinas de trabalho, confirmando a hipótese de que a combinação de diferentes técnicas de pós-processamento dos padrões mediante uma estrutura de navegação simples e de fácil uso reforçam a possibilidade de adoção da Mineração de Dados na rotina da gestão em Saúde.

#### 4. Discussão

A estratégia geral para aplicação de instrumentos que possibilitem a utilização de análise de maneira mais ágil em grande quantidade de dados seja efetivamente utilizada na rotina em saúde conta com a incorporação de dois grupos de requisitos. Assim, as necessidades, descritas como requisitos funcionais foram reforçadas pela incorporação de pontos de atenção obtidos a partir da literatura mediante um modelo de pós-processamento de padrões descobertos e avaliado por especialistas.

O modelo apresenta pontos a melhorar no que tange à compreensão dos padrões, especialmente das janelas de tempo, obtidas com o *AssocTemp* (SOKOLOSKI et al., 2014), cuja representação ainda requer uma estratégia visual mais intuitiva, o que acabou por prejudicar o rápido aprendizado para uso.

Entretanto, mesmo tendo criticado a capacidade do modelo de descrever os eventos ocorridos, os especialistas se mostraram satisfeitos com a proposta, em especial com a apresentação dos padrões e das informações. Destacando-se o fato de que todos os especialistas envolvidos na avaliação declararam que utilizariam o modelo proposto em sua rotina de trabalho, o que define o sucesso da proposta, dado que o problema de pesquisa gira em torno do fomento do uso do KDD na rotina em saúde.

#### 5. Conclusão

O desenvolvimento de um sistema para suportar e facilitar a obtenção dos padrões e para auxiliar na sua exploração permitiu a aplicação dos pontos de atenção “desenvolvimento de funcionalidade específica” e “representação visual dos resultados”. Por sua vez, a interface de usuário abstraiu a complexidade relativa à filtragem das regras de associação e análise das sequências e dos intervalos de tempo, mantendo o foco dos especialistas na observação dos padrões e permitindo a rápida identificação de beneficiários cobertos pelos padrões obtidos, contribuindo para a eficiência do processo de sua identificação.

A proposta de combinação da automação da obtenção de padrões com uma base de conhecimento explorada a partir de uma interface simples e intuitiva para os usuários contribuiu para o ganho de eficiência no uso do KDD na saúde, uma vez que o modelo proposto assume as atividades de transformação, mineração e apresentação dos padrões, cabendo à operadora de plano de saúde selecionar e preparar os dados e então, explorar os padrões obtidos.

#### Referências

AGRAWAL, R.; SRIKANT R. **Fast algorithms for mining association rules in large databases**. 20th International Conference on Very Large Data Bases. p 478-499, 1994.

BELLAZZI, R.; ZUPAN, B. Predictive data mining in clinical medicine: Current issues and guidelines. **International Journal of Medical Informatics**, n. 77, p. 81–97, 2008.

ARAYA, A. M. P.; CHUPEL, J. S.; CARVALHO, D.R.; DALLAGASSA, M.R.; IOSHII, O.S. Sistema Inteligente para apoio em Auditoria de Contas Médicas. **Journal of Health Informatics**, v. 8, n. 2, 2016.

BLONBERG, L. C. Gestão de métricas e indicadores de doenças em saúde **bucal suportado por um ambiente de descoberta de conhecimento em banco de dados**. 107f Dissertação (Mestrado em Ciência da Computação). Pontifícia Universidade Católica do Rio Grande do Sul, 2010.

BORGELT, C. **Apriori** - Association rule induction. 2010. Disponível em: <<http://www.borgelt.net/apriori.html>>. Acesso em: 22 jul. 2014.

CARVALHO, D. R.; DALLAGASSA, M. R.; DA SILVA, S. H. Uso de Técnicas de Mineração de Dados para a Identificação Automática de Beneficiários Propensos ao Diabetes Mellitus Tipo 2. **Informação & Informação**, n. 30, v. 20, 2015.

CARVALHO, D. R.; ESCOBAR, L. F. A.; TSUNODA, D. F. Pontos de Atenção para o uso da mineração de dados da saúde. **Informação & Informação**, v. 19, n. 1, p. 249-272, fev. 2014.

CARVALHO, D. R. **Árvore de Decisão / Algoritmo Genético para tratar o problema de pequenos disjuntos em classificação de Dados**. 173f. Tese (Doutorado em Ciências). Universidade Federal do Rio de Janeiro. 2005.

CARVALHO, D. R.; MOSER, A. D.; SILVA, V. A.; DALLAGASSA, M. R. Mineração de dados aplicada à fisioterapia. **Fisioterapia e Movimento**, v. 25, n. 3, p. 595-605, 2012.

CHEN, M.S.; HAN, J.; YU, P. S. Data Mining: An overview from database perspective. **IEEE Transactions on Knowledge and Data Engineering**, v. 8, n. 6, p. 866-883, 1996.

CRUZ-RAMÍREZ, M.; HERVÁS-MARTÍNEZ, C.; GUTIÉRREZ, P. A.; PÉREZ-ORTIZ, M.; BRICEÑO, J.; DE LA MATA, M. Memetic 86 41 Pareto differential evolutionary neural network used to solve an unbalanced liver transplantation problem. **Soft Computing**, v. 17, n.2 p. 275-284, 2013

DALLAGASSA, M. R. **Concepção de uma metodologia para identificação de beneficiário com indicativos de diabetes mellitus tipo 2**. 105 f. Dissertação (Mestrado em Tecnologia e Saúde). Pontifícia Universidade Católica do Paraná. 2009.

FAYYAD, U.; PIATETSKI-SHAPIRO, G.; SMYTH, P. The KDD process for extracting useful knowledge from volumes of data. **Communications of the ACM**, v. 39, n. 11, p. 27-34, 1996.

GOMES, H. M.; CARVALHO, D. R. A hybrid data mining method: exploring sequential indicators over association rules. **Iberoamerican Journal of Applied Computing**, v.1, n. 4, 2011.

HAN, J.; KAMBER, M. **Data mining: Concepts and techniques**. Elsevier. San Francisco. 2006

HUSSAIN, F.; LIU, H.; SUZUKI, E.; LU, H. **Exception rule mining with a relative interestingness measure**. PAKDD-2000, v.1805, p. 86-96. 2000.

KOBUS, L. S. G. **Identificação de usuários com doenças cardiovasculares elegíveis para programas de gerenciamento de caso – aplicação da descoberta de conhecimento em bases de dados a uma organização de saúde do tipo auto-gestão.** 146 f. Dissertação (Mestrado em Tecnologia em Saúde). Pontifícia Universidade Católica do Paraná. 2006

LIKERT, R. **A Technique for the measurement of attitudes.** *Achives of Psychology*. University of New York. n.140, p. 5-54, 1932.

Lin, Y.; Chen, H.; Brown, R. A.; Li, S.; Yang, H. Healthcare predictive analytics for risk profiling in chronic care: A bayesian multitask learning approach. **MIS Quarterly**, v. 41, n. 2, 2017.

LOPES, L. **Aprendizagem de máquina baseada na combinação de classificadores em bases de dados da área da Saúde.** 117 f. Dissertação (Mestrado em Tecnologia em Saúde). Pontifícia Universidade Católica do Paraná. Curitiba. 2007.

MANGIAMELI, P.; WEST, D.; RAMPAL, R. Model selection for medical diagnosis decision support systems. **Decision Support Systems**, v. 36, p. 247-259, 2004.

MARISCAL, G.; MARBÁN, O.; FERNANDEZ, C. A survey of data mining and knowledge discovery process models and methodologies. **The Knowledge Engineering Review**, v.25, n. 2, p. 137-166, 2010.

MEYFROIDT G.; GÜIZA, F.; RAMON, J.; BRUYNOOGHE, M. Machine learning techniques to examine large patient databases. **Best Practice & Research Clinical Anaesthesiology**, n. 23, p. 127-143, 2009.

MILANI, C. S.; CARVALHO, D. R. Pós-Processamento em KDD. **Revista de Engenharia e Tecnologia**, v. 5, p. 151-162, 2013.

MILOVIC, B.; MILOVIC, M. Prediction and decision making in health care using data mining. **International Journal of Public Health Science**, v.1, n. 2, p. 69-78, 2012.

NÉRI, E. D. R.; GADÊLHA, P. G. C.; MAIA, S. G.; PEREIRA, A. G. S. Erros de prescrição de medicamentos em um hospital brasileiro. **Revista Associação Médica Brasileira**, v. 57, n. 3, p. 306-314, 2011.

PATEL, V. L.; KAUFMAN, D. R.; KANNAMPALLIL, T. G. Diagnostic reasoning and decision making in the context of health information technology. **Reviews of Human Factors and Ergonomics**, n. 8, p. 149-193, 2013.

RAMON J., FIERENS D.; GÜIZA F.; MEYFROIDT G.; BLOCKEEL H.; BRUYNOOGHE M.; VAN DEN BERGHE, G. Mining data from intensive care patients. **Advanced Engineering Informatics**, n.21, p243–256, 2007.

RODDICK, John F.; SPILIOPOULOU, Myra. A bibliography of temporal, spatial and spatiotemporal data mining research. **Proc. SIGKDD Explorations**, v. 1, n. 1, p. 34-38. 1999.

SANTANA, Á. L.; FRANCÊS, C. R. L.; COSTA, J. C. W. A.; DE TARSO, P. **Observatório de saúde da Amazônia** - Sistema de suporte à decisão para os gestores de saúde pública da Amazônia. Environmental and Health World Congress. 2006.

SILBERSCHATZ, A.; TUZHILIN, A. What makes patterns interesting in knowledge discovery systems. **IEEE Transactions on Knowledge and Data Engineering**, v. 8, p. 970-974, 1996.

SHILLABEER, A. R. J. **Reconceptualising interestingness metrics for medical data mining**. Workshop on Health Data Mining. School of Informatics and Engineering, Flinders University of South Australia, 2005.

SHIMIZU, T., **Decisão nas organizações**. 3ª Edição. Editora Atlas, São Paulo. 2010.

SOKOLOSKI, W. F., CARVALHO, D. R., DALLAGASSA, M. R. **Regra de associação temporal**. XIV Congresso Brasileiro em Informática em Saúde – CBIS, Santos, Brasil, 2014.

SPRAGUE J.R., RALPH H.; WATSON, H. J. **Sistemas de apoio à decisão: Colocando a teoria em prática**. Rio de Janeiro. Campus. 1991.

TRINDADE, C. M. **Identificação do comportamento das hepatites virais a partir da exploração de bases de dados de saúde pública**. 141 f. Dissertação (Mestrado em Tecnologia em Saúde). Pontifícia Universidade Católica do Paraná. 2005.

TURBAN, E.; ARONSON, J. E. **Decision support systems and intelligent systems** Upper Sadle River. Prentice Hall, 2001.

TURBAN, E.; SHARDA, R.; ARONSON, J. E.; KING, D. **Business Intelligence: Um enfoque gerencial para a inteligência do Mercado**. Porto Alegre. Bookmann. 2009.

WANG, Y.; WANG, W. Y. C.; CEGIELSKI C.G. An integrated big data analytics-enabled transformation model: Application to health care. **Information & Management**, v. 55, n. 1, p. 64-79, 2018.