

Predição de prescrição médica eletrônica: análise de um sistema especialista

Édson das Neves Oliveira
UNILASALLE, Canoas/RS
edymoral@gmail.com

Patrícia Kayser Vargas Mangan
UNILASALLE, Canoas/RS
kayser@unilasalle.edu.br

Resumo

A Informática na Saúde é uma área emergente na Ciência da Computação, que se utiliza de técnicas e metodologias de outras áreas como Inteligência Artificial e Banco de Dados. Neste contexto, alguns trabalhos vêm sendo realizados buscando auxiliar o médico no diagnóstico e/ou prescrição de tratamentos. Este trabalho tem como objetivo apresentar e validar um modelo para predição de prescrição eletrônica, automatizando o processo de construção de prescrição e, conseqüentemente, padronizando o uso de medicações, dosagens e frequências. Para o desenvolvimento do modelo, utiliza-se técnicas de Sistemas de Raciocínio Baseados em Casos (RBC), Sistemas Especialistas (SE) e Data Mining, resultando em uma solução híbrida. O modelo foi validado utilizando-se uma base de casos reais, concedidos através da aprovação pelo Comitê de Ética e Pesquisa da instituição estudada. Como resultado, dependendo do Código Internacional da Doença (CID) obtém-se resultados na faixa dos 85% de similaridade em relação ao tratamento real construído pelo usuário especialista e, em outros casos consegue-se apenas similaridades na faixa dos 20%. Dessa forma, investiga-se o porquê da diferença, analisando e aperfeiçoando o modelo proposto para adaptar-se às características das bases de dados reais.

Abstract

Health Informatics is an emergent area in Computer Science, which uses techniques such as Artificial Intelligence and Data Bases. In this context, some efforts are being done in order to help medical staff on diagnosis and treatment prescription. This work, has the goal to present and validate an electronic prescription prediction model, automating the process of building a prescription and consequently standardizing the use of drugs, doses, and frequency. The development of this model requires techniques of Case-Based Reasoning (CBR), Expert Systems (ES) and Data Mining, providing a hybrid solution model. This model was evaluated with a real cases base, granted by the approval of the Ethics and Research Commission of the target institution. The results, considering the International Disease Code, indicate around 85% of similarity concerning the system results and the real treatment proposed by an expert. For some cases, it was obtained similarities around 20%,

which motivated further analyses to allow the model improvement.

1. Introdução

A informática na saúde é considerada uma área emergente em sistemas de informação, possuindo grande demanda de desenvolvimento de novas soluções tanto em hardware quanto em software. A informática na saúde é um campo de rápido desenvolvimento científico que lida com armazenamento, recuperação e uso da informação, dados e conhecimentos biomédicos para a resolução de problemas e tomada de decisão [1]. Os módulos de Prescrição Eletrônica existentes em sistemas de Prontuário Eletrônico do Paciente (PEP) podem ser aprimorados com uso de técnicas da Inteligência Artificial. Nesse trabalho utilizou-se como estudo de caso um sistema de PEP utilizado na Santa Casa de Misericórdia de Porto Alegre (ISCOMPA).

A Prescrição Médica é uma relação terapêutica importante entre o médico e o paciente, representando o resultado da capacidade diagnóstica e terapêutica do médico, sendo o documento onde estão dispostos todos os dados necessários para a equipe assistencial dar andamento no tratamento do paciente [2]. Dependendo da gravidade do estado clínico do paciente, prescrições extensas podem ser geradas as quais, conseqüentemente, têm sua construção trabalhosa, com chances de erros em dosagens e interações medicamentosas. As medicações são a causa mais freqüente de geração de eventos adversos, aumentando o número de internações hospitalares e o tempo de internação, assim como gerando custos extras [3].

Considerando-se a complexidade da definição de certas prescrições e os problemas que podem ocorrer, é possível afirmar que uma prescrição médica não pode ser baseada em regras simples, nem em informações padrão a diversos tipos fisiológicos de pacientes. Assim, um sistema computacional para auxílio a prescrição precisa contemplar o tratamento dos dados de entrada de forma dinâmica, moldado de acordo com cada novo caso e considerando-se a experiência médica para casos solucionados anteriormente. Portanto, para utilizar-se a base de experiência médica necessita-se avaliar, modelar e desenvolver uma metodologia que seja possível extrair as informações necessárias para a construção de uma proposta de tratamento. Sendo assim, o problema de pesquisa abordado neste trabalho visa elucidar qual seria a melhor ou melhores técnicas que, combinadas, poderiam retornar resultados

(proposta de tratamento) com um alto índice de eficácia.

Por conseguinte, o principal objetivo desse trabalho é avaliar a combinação de técnicas da Inteligência Artificial (IA), com a criação de métricas para comparação e extração de informações importantes para a construção de uma prescrição de tratamento considerando-se as características epidemiológicas do paciente. Ainda busca-se nesse estudo um alto grau de similaridade entre a proposta de tratamento construída pelo sistema, comparada ao tratamento construído pelo profissional de saúde. Este trabalho descreve o desenvolvimento de um modelo visando atingir os objetivos propostos, sendo que o restante do texto está organizado da seguinte forma. Na Seção 2, discutem-se os trabalhos relacionados ao uso de sistemas RBC e SE na área de Ciências da Saúde. A Seção 3 apresenta as principais técnicas utilizadas para desenvolvimento do modelo proposto. Na Seção 4, o modelo proposto é apresentado, com as suas respectivas etapas. Na seção 5 discutem-se os resultados obtidos e as perspectivas futuras para o presente trabalho. Finalmente, a Seção 6 apresenta as conclusões e as propostas de trabalhos futuros.

2. Trabalhos Relacionados

Embora os conceitos de RBC e SE sejam propostas da década de 90, muitos trabalhos recentes vêm sendo feitos na área da saúde. *Begum et al.* [4] apresentam uma pesquisa com os últimos sistemas desenvolvidos na área da saúde com técnicas de Inteligência Artificial (I.A.). Como resultado, dos 34 sistemas estudados apenas 12 têm como um de seus objetivos planejamentos terapêuticos, sendo que desses 12, apenas três são sistemas inteiramente dedicados para esse objetivo. Dessa forma o estudo apresenta uma tendência atual de desenvolvimento de sistemas que contenham uma proposta múltipla, de acordo com o gráfico reproduzido na Figura 1.

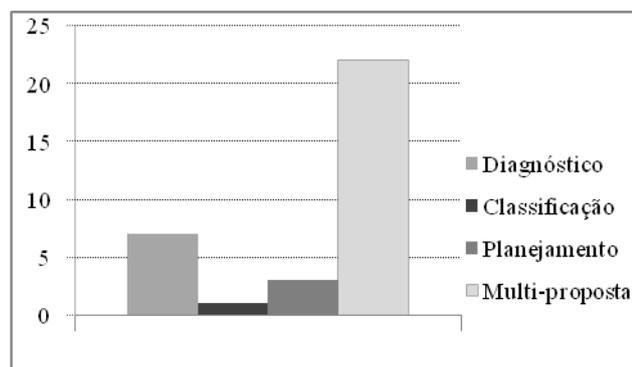


Figura 1. Gráfico com a tendência atual de desenvolvimento de sistemas RBC conforme [4].

O trabalho aqui proposto pressupõe que o diagnóstico do paciente já esteja definido, possuindo seu foco no planejamento da terapia para o tratamento

da doença. Deste modo, buscou-se na literatura sistemas nesse contexto, dos quais dois merecem destaque.

No trabalho “*A case-based multi-modal clinical system for stress management*” [5], é apresentado um sistema, cuja proposta é uma abordagem híbrida para apoio a decisão clínica no gerenciamento de *stress*. Utiliza técnicas da Inteligência Artificial (I.A.) para o diagnóstico e a apresentação de recomendações para o controle do nível de *stress* [5]. Utiliza, também, RCB como técnica base devido ao seu domínio não ser totalmente conhecido e a alta complexidade de um diagnóstico psiquiátrico. Dessa forma, inviabiliza qualquer construção de regras pré-definidas. A técnica de RBC é utilizada para o diagnóstico e para as recomendações, mas quando possuem poucos casos similares utiliza-se lógica *Fuzzy* para gerar casos para confrontar com o atual.

Esse sistema, diferentemente do trabalho proposto, auxilia no diagnóstico e apresenta algumas recomendações pré-definidas na base de conhecimento para o nível de *stress* detectado, não considerando características epidemiológicas do paciente, sendo restrito apenas para diagnóstico de *stress*.

O segundo trabalho, “*Case-Based Decision Support for Patients with Type 1 Diabetes on Insulin Pump Therapy*” [6], apresenta um sistema que gerencia diariamente a dose de insulina necessária para pacientes com Diabetes do tipo 1 com bomba de infusão de insulina. O paciente informa diariamente via Web o seu nível de glicose atual. Com o nível de glicose no sangue e sabendo dos hábitos do paciente, o sistema procura em sua base de conhecimento características similares para aquela situação e oferece ao médico opções de doses terapêuticas utilizadas em situações similares, para possível recomendação para o paciente [6].

Esse sistema propõe alterações na dosagem de insulina a partir de um nível de glicose e de eventos ou situação reportadas pelo paciente, diferentemente do sistema proposto que tem por objetivo apresentar uma proposta completa de tratamento de acordo com o diagnóstico informado pelo médico envolvendo além de medicamentos, procedimentos a serem realizados pela equipe assistencial.

Pesquisando-se o estado da arte, não se encontrou sistemas totalmente similares à proposta apresentada, apenas sistemas que possuem aspectos do seu funcionamento, aplicação de técnicas ou alguns objetivos similares ao trabalho proposto. Este trabalho também destaca-se pelo uso de dados reais obtidos na rotina de um hospital, o que permite a visão de diversos especialistas.

3. Raciocínio Baseado em Casos (RBC) e Sistemas Especialistas (SE)

Os sistemas de **Raciocínio Baseado em Casos** (RBC) são inspirados em um modelo cognitivo da

forma de como os seres humanos resolvem os seus problemas utilizando experiências passadas [5].

Os sistemas RBC foram criados utilizando-se os princípios da Memória Dinâmica, que define que o processo de aprendizagem ocorre quando existe diferença entre o que é esperado acontecer do que realmente acontece em determinada situação [7]. Como característica principal, indica-se o aprendizado automatizado, pois poderão utilizar um novo caso para resolver futuros casos [8]. Um sistema RBC deve apresentar suas etapas dentro de quatro processos [8]: Recuperação, Reutilização, Revisão e Retenção.

Durante a Recuperação, o caso atual é comparado com os casos existentes na base de conhecimento, sendo que utilizando-se o conhecimento do domínio é possível saber o quanto similar é o novo caso com o anterior. Após a seleção do caso anterior mais adequado para resolução do problema atual, o caso atual é armazenado para possível uso no futuro.

Para montagem e aquisição de conhecimento existem muitas técnicas que vão desde técnicas manuais até totalmente automatizadas [9]. Nesse trabalho utiliza-se uma técnica semi-automática através da mineração de dados para a aquisição de conhecimento, aproveitando-se o conhecimento prévio da base de informações.

Em sistemas RBC, o conhecimento é organizado em estruturas denominadas casos. Cada caso representa uma peça de conhecimento contextualizado representando um episódio ou problema que foi resolvido anteriormente [10]. Um caso contém normalmente um problema, uma solução, e seus resultados [4].

A organização dos casos é fundamental para obter-se um bom desempenho em sistemas RBC. Para o aproveitamento de experiências anteriores é necessário possuir-se uma memória com experiências eficazes, possuindo experiências específicas categorizadas e indexadas [7]. Como índice dos casos armazenados pelo sistema, utiliza-se o Código Internacional de Doença(CID) de cada paciente.

Sistemas que solucionam problemas que poderiam ser solucionados por um especialista humano são classificados como **Sistemas Especialistas** (SE). Os SEs utilizam uma grande base de conhecimentos montada de maneira mais eficiente possível através de especialistas humanos [11]. Com o conhecimento do domínio apropriado, o sistema especialista é possível igualar ou superar o desempenho de peritos humanos em tarefas estritamente definidas [12].

Como em sistemas RBC os sistemas especialistas (SE) possuem a sua fase de aquisição de conhecimento, um dos principais problemas, que é a disponibilidade do especialista para se extrair o conhecimento do negócio [9]. Por isso, é possível obter-se o conhecimento a partir dos seguintes métodos [13]:

- Pesquisas na literatura: obtém-se um conhecimento inicial sobre o processo;
- Entrevistas com usuários especialistas: o engenheiro de conhecimento realiza

entrevistas, para posteriormente ser analisá-las em busca de requisitos necessários;

- Acompanhamento: acompanha o raciocínio do especialista em casos reais. Dessa forma, é pesquisada a forma como os casos são resolvidos por esse especialista;
- Processos semi-automáticos: utilizam-se ferramentas computacionais para a extração do conhecimento de alguma base de informações existente.

Nesse trabalho utilizaram-se processos semi-automáticos para a extração de conhecimento, com o emprego de técnicas de *Data Mining* para a manipulação do grande volume de informações.

4. Modelo de Prescrição Automática

Com o estudo de técnicas e ferramentas de sistemas RBC em conjunto com o conhecimento adquirido com a literatura e experiência prévia, apresenta-se um modelo computacional de prescrição médica adequada para o tratamento de um paciente previamente diagnosticado, considerando seu estado clínico e suas características epidemiológicas. Como requisito para o início da avaliação das técnicas, necessita-se explicar o processo de aquisição do conhecimento para posterior construção da base de casos. Como já citado na introdução deste artigo, utilizou-se uma base de casos reais, concedidos através da aprovação pelo Comitê de Ética e Pesquisa da Santa Casa de Misericórdia de Porto Alegre (ISCOMPA), permitindo ter como estudo de caso os dados do sistema de Prontuário Eletrônico do Paciente (PEP) utilizado na ISCOMPA. Esta base de dados foi utilizada tanto para a construção da base de conhecimento quanto para a avaliação do sistema.

4.1. A Aquisição de Conhecimento

Para a aquisição do conhecimento utiliza-se informações das três etapas que compõe a internação do paciente: admissão, tratamento e alta. Durante a admissão do paciente utilizam-se as seguintes informações: Sexo, Data de nascimento, Peso, Altura e classificação de Índice de Massa Corpórea (IMC).

Com as informações epidemiológicas dos pacientes necessita-se também das informações referentes ao tratamento que o paciente foi submetido, ou seja, das informações geradas durante a criação e/ou alteração da prescrição médica:

- Item prescrito: medicação, dieta ou procedimento;
- Dose: valor numérico correspondente a quantidade de medicação;
- Dose Unidade: unidade de dosagem, por exemplo: miligramas, comprimidos, etc.;
- Via: Via de administração da medicação, por exemplo: oral, endovenosa, etc.;

- Frequência: frequência da medicação, por exemplo: 2 vezes ao dia, de 2 em 2 horas, etc.

Finalizando a fase de aquisição de conhecimento, necessita-se da informação mais importante para a indexação dos casos, o CID da alta, devido a sua inclusão no prontuário ser feita após todo o desfecho da internação do paciente estar concluído. Para fins de análise de resultados, a data de alta também será considerada para diferenciar qual o ano que o caso ocorreu. Dessa forma, as seguintes informações serão utilizadas: CID da alta e a data de alta.

4.2. Extração de Informações

Nesse modelo utiliza-se a aquisição de conhecimento semi-automática, através da técnica de *Data Mining*, para buscar a melhor amostragem e a representação correta dos dados pesquisados. Considerando-se o problema de disponibilidade do especialista, através de pesquisa na literatura e do conhecimento adquirido durante o desenvolvimento do sistema PEP principal, obteve-se conhecimento suficiente para a execução da fase de clusterização do *Data Mining*. Logo após, executa-se a fase de classificação, que obtém como resultado um modelo de informações agrupadas conforme a necessidade do estudo proposto. Esse modelo utiliza cubos de informações que, por definição, são tabelas desnormalizadas que possibilitam múltiplas visões de uma mesma estrutura de dados com independência total da base de dados do sistema principal.

Com a metodologia de extração de informações definida, necessita-se representar as mesmas de forma correta, para iniciar-se a construção da base de conhecimento.

4.3. Representação do conhecimento

As informações contidas em diversos processos que compõe uma internação são utilizadas para a representação do conhecimento. Com isso, considera-se como um caso, a internação de um paciente com tratamento que lhe foi submetido contendo todo o conjunto de prescrições, com suas respectivas medicações administradas

Neste trabalho, cada caso é representado por um objeto Java do tipo Value Object (VO) descrito como uma classe composta por uma estrutura simples, somente de atributos, sendo utilizada para encapsular os dados que transitam entre as camadas de apresentação e de negócios da aplicação [15]. Como modelagem dos objetos tem-se o diagrama de classes da Figura 2.

A Figura 2 demonstra que, para cada caso, são incluídos os atributos fisiológicos do paciente, bem como seus respectivos itens utilizados no tratamento. Os itens podem ser tanto procedimentos quanto medicamentos. Um exemplo de procedimento constante no PEP seria “uso de colchão...” e de medicamento seria “2 comprimidos de paracetamol

250 mg”. Tais informações constantes são analisadas e transformadas na estrutura da Figura 2.

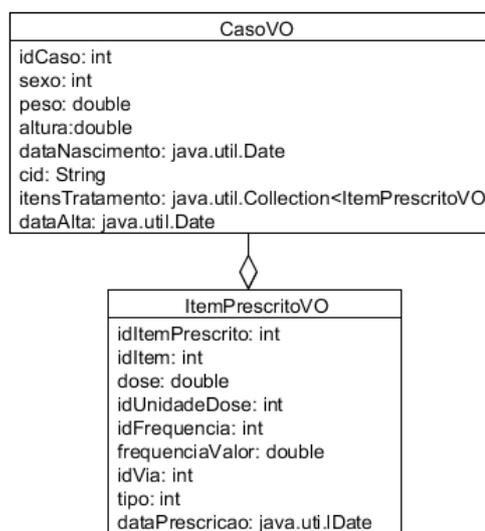


Figura 2. Detalhe do Diagrama de classes do modelo, indicando as classes utilizadas para representação (armazenamento) dos casos.

4.4. Base de Conhecimento

Uma das preocupações durante a concepção do sistema, é que a solução adotada fosse a menos intrusiva possível no sistema em produção. Aspectos de segurança dos dados e de desempenho pesaram na decisão de realizar o armazenamento dos casos em uma base isolada do sistema principal. Essa separação permite afirmar que não há influência no desempenho do sistema principal. Além disso, por ser criada uma estrutura de dados integrada com a aplicação que fica em memória principal, o desempenho do sistema de prescrição é melhor do que se utilizasse a base original. Além disso, por não importar nesta cópia os dados pessoais dos pacientes, também garante privacidade dos dados médicos.

Outro aspecto é a portabilidade das informações, já que esta cópia é feita de forma independente de um sistema gerenciador de banco de dados. Para obter-se o máximo de independência e a possibilidade de gravar o objeto Java com o seu estado atual, optou-se por não utilizar banco de dados relacional. Por isso, quando inicia-se o sistema proposto a base de conhecimento é colocada em memória, através de leitura de arquivos gravados em disco.

Ainda para obter-se um bom desempenho na recuperação, de acordo com Tonidandel [16], a base de conhecimento poderia ser desenvolvida como ponteiros, identificando-se os casos de acordo com algumas características principais. Por isso, optou-se por criar um sistema de indexação (análogo ao uso de ponteiros em uma linguagem imperativa) que permite um acesso mais rápido aos dados, conforme será discutido nas seções posteriores.

4.5. Indexação

Para a obtenção de um desempenho favorável em um sistema baseado em casos, é fundamental a construção de uma estrutura indexada, que atenda com precisão e velocidade as diversas situações que o sistema será exposto [16]. Portanto, possuindo-se um processo de indexação com uma modelagem eficaz, diminui-se consideravelmente a complexidade dos algoritmos utilizados no cruzamento e recuperação dos casos.

Nesse trabalho os casos são indexados a partir do código da doença do paciente, portanto, cada caso tem como índice principal o Código Internacional de Doenças (CID), que foi atribuído no processo de alta do paciente. Este esquema de mapeamento é ilustrado na Figura 3.

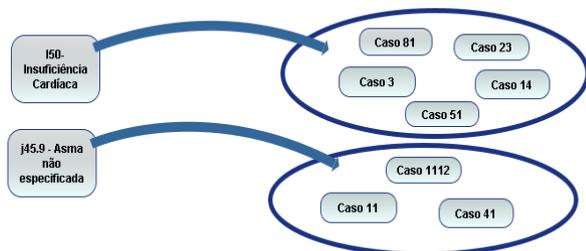


Figura 3. Indexação de casos por CID.

Observa-se que para cada CID pode-se ter vários casos distintos, ou seja, no momento da recuperação, os casos já estarão sendo filtrados de maneira eficiente, reduzindo consideravelmente o tempo de recuperação. Logo após a recuperação dos casos pelo CID informado, inicia-se o processo que leva em consideração os dados epidemiológicos dos pacientes utilizando os critérios de similaridade definidos. É importante também notar que a Figura 3 evidencia que cada caso possui um único CID principal, mesmo que existam outras doenças associadas.

4.6. Recuperação de Casos

Com a montagem da base de conhecimento indexada e a definição dos critérios de similaridade é possível a definição da estratégia a ser utilizada para a recuperação de casos. Esse processo tem seu início com a entrada do novo caso e seu término após a apresentação de possíveis alternativas de resolução para o novo caso.

Para a construção desse processo, dividiu-se o mesmo em etapas distintas, cada uma com a necessidade de dados de entrada e seu respectivo retorno que se utiliza na próxima etapa dependente. A definição de cada etapa foi de acordo com sua respectiva responsabilidade, sendo divididas nos seguintes itens: pesquisa CID, similaridade epidemiológica, similaridade tratamento, ranking itens e apresentação da proposta de tratamento.

Pesquisa CID: Nessa etapa recebe-se como informação de entrada o CID que o novo paciente foi classificado. Após receber o código, o algoritmo de pesquisa acessa a estrutura de casos.

Similaridade Epidemiológica: Para início dessa etapa, necessita-se possuir duas informações: dados epidemiológicos do novo caso e os casos recuperados com o mesmo CID. Com essas informações de entrada os casos anteriores são confrontados em busca de alguma similaridade entre as informações. Todos os casos têm sua distância entre os atributos calculada, resultando a distância entre os casos. Para cada atributo se obtém o valor da distância utilizando-se a Equação 1:

$$d = \sqrt{(p * |i|)} \quad (1)$$

A variável p representa o valor definido para o peso de cada atributo. Esse peso é multiplicado pela diferença absoluta entre o valor do atributo do caso atual com o anterior representado pela variável i . Através da iteração por todos os atributos do caso anterior, obtém-se a distância total conforme a Equação 2:

$$\begin{aligned} dt = & \text{distância idade} \\ & + \text{distância de peso} \\ & + \text{distância da altura} \\ & + \text{distância sexo} \end{aligned} \quad (2)$$

Com o resultado da distância total, compara-se o resultado com o chamado indicador de similaridade. Esse indicador é um valor numérico que é definido de acordo com o grau de precisão que se necessita dos resultados, ou seja, quanto menor esse valor, menor deve ser a diferença entre os casos.

Similaridade Tratamento e Ranking: A busca pela similaridade do tratamento é realizada de forma semelhante da similaridade dos atributos epidemiológicos. Inicia-se tendo como entrada os casos similares obtidos na etapa anterior. Diferente da busca da similaridade epidemiológica entre os casos, nessa etapa os casos selecionados são comparados entre si, em busca de itens similares em seus tratamentos. Durante a comparação é registrado quantas vezes o item já foi utilizado nos tratamentos pesquisados.

Apresentação da proposta: Nessa etapa final, apresenta-se uma tela com os itens propostos com sua respectiva descrição e seu número de utilização. Essa pontuação é necessária em decorrência da lista final possuir, por exemplo, duas medicações com algumas pequenas variações de dosagem. Nesse caso, o sistema apresenta as duas soluções na lista de itens da proposta de tratamento, para que o usuário final decida qual item deve ser utilizado.

5. Análise dos Resultados

Todos os trabalhos relacionados apresentados demonstram como resultado final os casos similares

encontrados. Dessa forma, o usuário especialista precisa selecionar o caso mais adequado para utilizá-lo. Já o sistema aqui proposto retorna como resultado uma lista com a combinação de todos os itens dos tratamentos dos casos retornados como similares. Assim, para o usuário apresenta-se uma lista com os itens propostos pelo sistema. Em contato com especialistas, tal apresentação, por ser mais flexível e detalhada, apresenta mais potencial para adoção. A Figura 4 apresenta um exemplo de saída para o usuário.

Selecionar	Itens de Prescrição	Relevância
<input type="checkbox"/>	Medicamento Não Padronizado, 1 comp Oral, às 8 horas	14
<input type="checkbox"/>	DIPIRONA SÓDICA (500mg/ml amp 2ml), 1 amp Endovenosa, 6/6 horas	8
<input type="checkbox"/>	Cama com colchão piramidal, 1 null Tópica, contínua	8
<input type="checkbox"/>	CLORETO DE SÓDIO 0.9% (fr 250ml), 3 ml Inalatória, 4/4 horas	7
<input type="checkbox"/>	AZITROMICINA(500mg Comprimido), 1 comp Oral, 1x ao dia	6

Figura 4. Tela com proposta de prescrição.

Como apresentado na Figura 4, os itens do tratamento proposto possuem à sua direita o indicador de relevância, ou seja, quantas vezes apareceram nos tratamentos considerados similares pelo sistema.. Assim, a proposta é montada considerando-se o dia de internação do caso atual, pois devido a evolução do estado clínico do paciente, procedimentos e medicações podem variar de um dia para o outro.

Como o sistema retorna apenas a lista de itens do tratamento, a avaliação quanto a sua real efetividade fica mais difícil, devido à combinação dos resultados das diversas etapas que são executadas desde a busca dos casos pelo CID até a sua apresentação na tela. Por isso, a sua avaliação deveria ser feita por um usuário especialista no momento de criar uma nova prescrição para um paciente. Como o problema de falta de disponibilidade do usuário especialista médico é recorrente na maioria dos Sistemas Especialistas, optou-se por validar os resultados utilizando-se uma amostragem específica da própria base de conhecimento.

5.1. Montagem avaliação

Para verificar-se o quanto que a proposta do sistema é efetiva, necessita-se obter o grau de similaridade entre a proposta de prescrição e uma prescrição real construída pelo usuário especialista. O grau de similaridade é um percentual que varia de acordo com o CID estudado sendo que, quanto mais itens em comum entre a prescrição proposta e a real, mais próximo dos 100% de semelhança entre as prescrições.

A base de conhecimento foi construída utilizando-se casos com CID mais comuns atendidos pelo ISCMPA.

Os casos considerados para a montagem da base de conhecimento pertencem ao período de janeiro de 2004 até abril de 2010. Nesse estudo buscou-se um período mais atualizado de informações devido à atualização frequente em tratamentos. Para manter a integridade das informações alguns critérios essenciais devem ser atendidos, tais como os expressos no critério 1.

Critério 1: *todas as informações obrigatórias (peso, altura, sexo e data de nascimento) devem estar preenchidas.*

As informações obrigatórias para o modelo proposto são de vital importância no processo de comparação da similaridade. Por isso, criaram-se alguns filtros nas consultas utilizadas durante a extração das informações da base do sistema principal. Através desses filtros, diversos casos não podem ser considerados para a construção da base de conhecimento, diminuindo assim o número total como é apresentado na Tabela 1.

Tabela 1. Total de casos por CID.

CID	Total	Completo
N18.- Insuficiência renal crônica	2847	406
I50-Insuficiência Cardíaca	1848	276
J45.9-Asma não especificada	2727	263
J18.0-Broncopneumonia	4786	93

Critério 2: *utilização somente de informações consistentes.*

Durante a etapa de testes preliminares, encontrou-se algumas inconsistências nas informações, tais como:

- Data de nascimento: ano da nascimento superior a data atual;
- Sexo: informação como indeterminado;
- Peso: com valor igual a zero ou negativo;
- Altura: com valor igual a zero ou negativo.

Estas informações são de vital importância no processo de comparação da similaridade. Dessa forma, também verificou-se a necessidade de criação de filtros para tratar esses tipos de inconsistências. Assim, o número de casos da base de conhecimento sofreu algumas alterações como apresentado na Tabela 2.

Tabela 2. Total de casos após filtragem

CID	Casos antes dos filtros	Casos após a filtragem	Percentual de redução
N18.- Insuficiência renal crônica	406	283	32,01%
I50-Insuficiência Cardíaca	276	222	19,56%
J45.9-Asma não especificada	263	160	39,16%
J18.0-Broncopneumonia	93	64	31,18%
Total	1038	729	29%

Critério 3: *alguma padronização no tratamento, ou seja, CID referentes a doenças com tratamento relativamente padronizado.*

A avaliação eficaz dos resultados depende de alguma padronização nos tratamentos referentes aos CID selecionados. Não seria possível avaliar o modelo proposto possuindo-se na base de conhecimento tratamentos para pacientes similares com quase nenhuma semelhança. Isso geralmente ocorre com doenças novas ou com pouca ocorrência. A relação de CID com alguma padronização no tratamento foi obtida através de conhecimento prévio adquirido no desenvolvimento do sistema em uso pela ISCMPA.

5.2. Período de amostragem

Para avaliar a solução proposta dividiu-se a base de conhecimento em período de aprendizagem e período de amostragem de avaliação, como é demonstrado na Tabela 3.

Tabela 3. Período de amostragem.

2004	2005	2006	2007	2008	2009	2010
Período de aprendizagem					Novos Casos	

De acordo com a Tabela 3, o período utilizado para aprendizagem e a montagem da proposta de prescrição ocorre de 2004 até dezembro de 2008. Assim, o período utilizado pela amostra de avaliação é composto pelo ano de 2009 e 2010. Com a definição desses períodos de amostras é possível iniciar-se o processo de avaliação.

5.3. Processo de avaliação

Como visto na seção de critérios para montagem da amostragem, alguns filtros foram executados durante o processo de construção da base de conhecimento para garantir a integridade das informações. Dessa forma, obtém como resultado uma base consistente, mas com menor quantidade de casos. Por isso, necessita-se selecionar quais os casos utilizados para comparação, dando preferência aos que possuem maior quantidade similares na base de conhecimento.

Tabela 4. Maior incidência de acordo com idade.

Idade	I50		J45.9		N18		J18	
	M.	F.	M.	F.	M.	F.	M.	F.
0 - 9	1	3	68	44	11	15	4	2
10-19	0	1	10	20	32	17	0	1
20-39	1	4	1	0	32	30	1	1
40-59	18	10	0	3	42	27	4	2
60-79	55	35	0	8	25	20	14	9
>= 80	28	43	1	1	6	5	5	9
Total	103	96	80	76	148	114	28	24

Como resultado, a base de conhecimento possui um total de 729 casos distintos, divididos em quatro tipos de CID. Para fins de melhor apresentação, dividiu-se em três tabelas, classificando-os de acordo com suas respectivas características.

Tabela 5. Maior incidência de acordo com peso.

Peso	I50		J45.9		N18		J18	
	M.	F.	M.	F.	M.	F.	M.	F.
0-19	1	2	66	40	24	21	4	3
0-39	0	0	10	12	20	20	1	2
40-59	10	33	1	15	29	33	7	11
60-79	45	38	3	5	52	32	11	6
80-99	41	17	0	3	20	7	4	1
>= 100	6	6	0	1	3	1	1	1
Total	103	96	80	76	148	114	28	24

Tabela 6. Maior incidência de acordo com altura.

Altura	I50		J45.9		N18		J18	
	M.	F.	M.	F.	M.	F.	M.	F.
<= 99	1	3	50	21	15	14	3	2
100 – 149	0	6	26	34	37	34	1	2
150-169	43	81	4	21	46	63	11	19
170-189	59	6	0	0	50	3	13	1
> 190	0	0	0	0	0	0	0	0
Total	103	96	80	76	148	114	28	24

De acordo com as Tabelas 4, 5 e 6 verifica-se que dependendo do CID ocorre incidência maior em determinadas faixas de valores dos atributos que compõe o caso. Por exemplo, o CID I50 ocorre com maior incidência em pacientes do sexo masculino com idade entre 60 e 79 anos com 65 casos (valor entre parênteses). Assim como, o CID J45.9 ocorre mais em pacientes com idade de 0 a 9 anos.

Através dessas conclusões, selecionou-se as características com maior incidência de pacientes, considerando-se o CID correspondente. Dessa forma, consideram-se essas características para selecionar os casos de 2009-2010 como novos casos para a análise do sistema.

5.4. Execução da Avaliação

Para a avaliação da similaridade da prescrição proposta com a construída pelo especialista, utiliza-se uma tela que compara as duas prescrições (sistema x especialista), sinalizando os itens em comum entre ambas e o percentual de similaridade. Esse percentual é calculado com um simples regra de três em que os 100% são todos os itens da prescrição construída pelo especialista.

5.5. Resultados da avaliação

O processo de avaliação foi realizado de forma manual, como citado anteriormente, utilizando-se o grupo de casos de 2009-2010 selecionados na etapa de análise e seleção de novos casos para avaliação. Para cada CID o conjunto de casos foi inserido no sistema e a proposta correspondente foi confrontada com a proposta feita pelo especialista. Como exemplo, a Tabela 7 demonstra os resultados preliminares para o CID J45.9.

Tabela 7. Similaridade dos casos do CID 45.9.

Sexo	Idade	Peso (kg)	Alt. (cm)	Casos	Sim(%)	Itens
Masc	6	15	92	17	50%	14
Masc	5	11	90	19	87%	8
Fem.	4	17	98	16	48%	21
Fem.	2	10	85	3	62%	16

De acordo com os resultados apresentado pela Tabela 7, verificou-se um baixo índice de similaridade entre ambas. Assim, como a maioria dos sistemas RBC, apresentou-se esse comportamento devido ao baixo número de casos similares na base de conhecimento. Além do baixo número de casos, considerado o principal fator de baixas taxas de similaridade, outros fatores contribuíram com menor proporção ao baixo índice, tais como: dosagens com padrões diferentes e frequências com padrões diferentes.

A diferença entre dosagens ocorre principalmente com medicamentos. Uma medicação pode ser prescrita utilizando-se uma unidade de apresentação ou uma unidade de concentração. A unidade de apresentação é referente ao formato físico da medicação (comprimidos, ampolas, cápsulas etc.). A unidade de concentração refere-se à concentração da medicação a ser administrada para o paciente, são consideradas unidades básicas de medida (miligramas, microgramas, etc.). Dessa forma, ocorrem situações como a Tabela 8.

Tabela 8. Exemplo medicações dosagem.

Medicamento	Exemplo 1	Exemplo 2
Paracetamol (250mg)	2 comprimidos	500 mg
Heparina (5000 ui amp)	5000 ui	1 ampola

Como apresentado na Tabela 8, o medicamento Paracetamol pode ser prescrito de duas formas diferentes, mas para fins terapêuticos é o mesmo resultado. Assim como as medicações, as frequências dos itens de prescrição podem ser informados de

diferentes formas, algumas tem o mesmo resultado final e outros são diferentes no sentido terapêutico. A tabela 9 apresenta alguns exemplos.

Tabela 9. Exemplos de frequências.

Frequência	Exemplo 1	Exemplo 2
Vezes ao dia	2 vezes ao dia	De 12 em 12 horas
Vezes ao dia	3 vezes ao dia	De 8 em 8 horas

A Tabela 9 demonstra o exemplo de frequência vezes ao dia. Dependendo da medicação o seu efeito é prejudicado se não for administrada em períodos regulares de tempo, ou seja, nesse caso utiliza-se frequências do tipo “de x em x horas”. E em casos que não precise um rigor na sua execução são utilizadas frequências do tipo x vezes ao dia. Entretanto, o sistema de prontuário eletrônico da ISCMPA possui uma funcionalidade denominada como aprazamento automático.

Essa funcionalidade permite ao usuário médico, prescrever um item com a frequência “vezes ao dia” e o sistema gera automaticamente os horários para o item considerando o horário atual. Dessa forma, pode-se considerar similar um item prescrito com uma frequência “vezes ao dia” com uma “de x em x horas”. Após a breve reflexão sobre alguns possíveis itens que influenciam a similaridade entre as prescrições, avaliou-se as prescrições com opções de considerar as dosagens e frequências, ou simplesmente, descartar comparando apenas o item e sua respectiva via de administração.

Tabela 10. Resultados sem considerar dosagem e frequências.

	I50	J45.9	N18	J18
Masculino	25%	68,5	21%	20%
Feminino	40%	55%	10%	Sem casos

Tabela 11. Resultados considerando-se dosagem e frequências.

	I50	J45.9	N18	J18
Masculino	43,5%	92,5%	50%	28%
Feminino	80,5%	80,5%	40%	Sem casos

Com a execução da avaliação caso a caso possibilitou-se a apresentação final dos resultados globais. Essa apresentação organiza-se o índice de similaridade de acordo com o CID e o sexo correspondente dos casos avaliados. Um comparativo é apresentado através das tabelas 10 e 11, pois a Tabela 10 apresenta os resultados considerando dose e frequência enquanto, a

Tabela 11 apresenta sem considerá-los.

6. Conclusão

Ao longo desse trabalho, estudaram-se três técnicas da IA que combinadas resultaram no sistema híbrido apresentado. Esse modelo teve grandes desafios, principalmente a manipulação uma grande massa de dados, dos quais nem todos completos ou corretos. Atualmente, a ISCMPA possui mais de 1200 leitos divididos em cinco hospitais que geram dados diários que alimentam a base consultada. Com a aprovação do CEP, a solução começou a ser avaliada com informações reais de pacientes que obtiveram alta da instituição. O modelo na etapa de avaliação apresentou limitações. Como todo sistema RBC, quanto menor os casos similares pior era o índice de similaridade alcançado entre a prescrição proposta com a do especialista. Esse problema ocorreu devido a filtragem necessária para garantir o mínimo de informações obrigatórias para o processo de comparações entre os casos.

Como contribuição, destaca-se a pesquisa direcionada por CID, sendo eficaz na proposta de tratamentos já padronizados, sem a necessidade do médico criar novamente uma prescrição necessitando apenas, alguns ajustes para adaptá-la as necessidades clínicas de outro paciente.

Ainda destaca-se que não se encontrou proposta similar ao modelo proposto, que tem como principal diferencial, apresentar a prescrição pronta, sendo apenas necessário o usuário selecionar os itens que são relevantes para o novo caso.

Através da análise da base de casos, estabeleceram-se algumas informações relevantes sobre a incidência de doenças de acordo com gênero, idade, peso e altura dos sujeitos. Estas informações nortearam algumas das decisões do processo de avaliação do SE. Cabe ressaltar, no entanto, que tais dados não podem ser extrapolados para uso como resultado científico. Embora sejam indicadores relevantes para a comunidade da área de saúde, não foi resultado de estudos sistemáticos ou com amostras que permitam dados estatisticamente significativos.

Um interessante trabalho futuro seria aplicar parte da metodologia de tratamento de dados desenvolvidos para auxiliar equipes da área de saúde a obter dados para análise em períodos e amostras maiores e que permitam extrair conhecimento sobre a influência das características epidemiológicas do paciente.

Como outro trabalho futuro, o modelo pode ser aprimorado para considerar outras características do estado clínico do paciente, histórico de doenças familiares, hábitos etc. O modelo também pode ser aprimorado para conversão de doses e frequências diferentes, aumentando a precisão do modelo.

Ainda como trabalho futuro, o modelo proposto

pode ser utilizado para fins educacionais no treinamento de estudantes de medicina, avaliando as diferentes condutas que podem ser adotadas de acordo com o estado clínico geral do paciente.

Referências

- [1] M. S. Blois, E. H. Shortliffe, "The Computer Meets Medicine: Emergence of a Discipline", Addison-Wesley Publishing, 1990.
- [2] L. L. Brunton, J. S. Lazo, K. L. Parker, "Goodman & Gilman as Bases Farmacológicas da Terapêutica", McGraw-Hill, 11ª Ed., 2007.
- [3] T. Morimoto.; T. K. Gandhi, A. C. Seger, et al, "Adverse drug events and medication errors: detection and classification methods", *Qual Saf Health Care*, v. 13, p. 306-314, 2004.
- [4] S. Begum, M. U. Ahmed, P. Funk, X. Ning and M. Folke, "Case-Based Reasoning Systems in the Health Sciences: A Survey Recent Trends and Developments" *IEEE Trans. On Systems, Man and Cybernetics- Part C: Applications e Reviews*, vol. 40, no.6, pp. 1-15, Set. 2010.
- [5] M. U. Ahmed, S. Begum, P. Funk, and X. Ning, "A multi-modal case-based system for clinical diagnosis and treatment in stress management," presented at 7th Workshop on Case-Based Reason. Health Sci., Seattle, Washington, 2009.
- [6] C. Marling, J. Shubrook, and F. Schwartz, "Case-based decision support or patients with type 1 diabetes on insulin pump therapy", in *Proc. 9th Eur. Conf. Case-Based Reason: (ECCBR)*. Berlin, Germany. Springer - Verlag, 2008, pp. 325-339.
- [7] R. Schank, "Dynamic Memory Revisited", Cambridge University Press., 1999.
- [8] A. Aamodt and E. Plaza, "Case-based reasoning: Foundational issues, methodological variations, and system approaches", *Artif. Intell. Commun.*, vol. 7, 1994.
- [9] S. Rezende, "Sistemas Inteligentes: fundamentos e aplicações", Manoela, 2003.
- [10] V. Wangenheim, "Raciocínio Baseado em Casos: fundamentos, técnicas e desenvolvimento de aplicações", Manole, 2009.
- [11] E. Rich and K. Knight, "Inteligência Artificial", Makron Books, vol. 2, 1994.
- [12] S. Russell and P. Norvig, "Artificial Intelligence: A modern Approach", Prendice Hall, 2009.
- [13] W. Costa and S. Silva, "Aquisição de conhecimento: o grande desafio de sistemas especialistas" *Revista HOLOS*, vol. 2, no.4, pp. 37-46, 2005.
- [14] L. Coelho and V. Mariani, "Sistema Híbrido Neuro-evolutivo aplicado ao controle de processo multivariável", *Revista SBA*, vol. 17, no.1, 2006.
- [15] D. Alur, J. Crupi, D. Mals, "J2EE Patterns: Best practices and design strategies", Sun Microsystems Press, 2003.
- [16] F. Tonidandel, "Desenvolvimento e implementação de um sistema de planejamento baseado em casos", Dissertação de mestrado, Universidade Federal de São Paulo - USP, 2003.