

OBSERVATÓRIO DE SAÚDE DA AMAZÔNIA - SISTEMA DE SUPORTE À DECISÃO PARA OS GESTORES DE SAÚDE PÚBLICA DA AMAZÔNIA

Ádamo L. de Santana, Carlos Renato L. Francês, João C. W. A. Costa¹, Paulo de Tarso²

Abstract — *In the current state-of-art, it is not more satisfactory to get only the historical information from databases. It is important to make use of information a priori, that is, to make possible to carry out prospections from data of system. This work presents a strategy to investigate the application of public resources in the health area, verifying the impact caused in the index of human development (IDH), in the context of the cities of the State of Pará.*

Resumo — *No atual estado-da-arte, não é mais satisfatório obterem-se apenas as informações históricas de bases de dados. É importante dispor de informações a priori, isto é, que possibilitem realizarem-se prospecções do sistema. Este trabalho apresenta uma estratégia para investigar a aplicação de recursos públicos na área de saúde, verificando o impacto causado no índice de desenvolvimento humano (IDH), no contexto dos municípios do Estado do Pará.*

Palavras-chave: *Data Mining, IDH, Redes Bayesianas, Saúde Pública.*

1. INTRODUÇÃO

Pela exigüidade de recursos financeiros disponíveis nos países de terceiro-mundo, é imprescindível que esses poucos recursos sejam aplicados em projetos que efetivamente possam gerar impacto na melhoria das condições de vida da população. Desta forma, torna importante a capacidade de obter-se um conjunto de informações privilegiadas, através das quais seja possível traçar estratégias de políticas públicas e, eventualmente, redirecionar esforços, corrigir rumos e realizar ações preventivas.

Uma abordagem possível é prover sistemas de suporte à decisão par os gestores do setor de saúde, combinando a capacidade estatísticas/probabilística com o know-how dos especialistas do domínio da saúde. Obviamente, a avaliação deverá ser realizada em função de um ou mais índices entendidos como plausíveis, para a verificação da eficácia das informações obtidas.

Este trabalho apresenta uma estratégia implementada para uma instituição recente da Fundação Nacional de Saúde, chamada de “Observatório de Saúde da Amazônia - OSA” (Seção Pará), cuja idéia é realizar a integração de diversas informações existentes em bases de dados isoladas,

que contemplem informações sobre a Região Amazônica, utilizando-se a estrutura de comunicação e de geo-referenciamento do Sistema de Proteção da Amazônia (SIPAM), antigo Sistema de Vigilância da Amazônia (SIVAM).

Este trabalho apresenta um estudo pretendido pelo Observatório, abordando o impacto do bom/mal uso de recursos públicos nos índices/contas sociais, dentre os quais, destaca-se o Índice de Desenvolvimento Humano (IDH), dentro dos municípios do Estado do Pará. Para realização dessa tarefa faz-se uso da abordagem de “descoberta de conhecimento” baseada em redes bayesianas.

Desta forma, este artigo apresenta a seguinte estrutura: a Seção 2 é apresentado o processo de extração de conhecimento (KDD - Knowledge Discovery in Database), com ênfase em redes bayesianas. Na Seção 3 é apresentada a metodologia utilizada para a análise do estudo de caso; apontando também as particularidades da infra-estrutura do Observatório de Saúde da Amazônia. Na seção 4 é apresentada a aplicação do processo KDD na área de Saúde Pública, incluindo uma análise particular sobre as informações referentes ao estado do Pará. A Seção 5 apresenta as considerações finais do trabalho.

2. EXTRAÇÃO DE CONHECIMENTO DE BASES DE DADOS

O processo de Extração de Conhecimento de Bases de Dados (KDD - Knowledge Discovery in Database) [1] desponta como uma tecnologia capaz de cooperar amplamente na busca do conhecimento embutido nos dados. Desta forma, o principal objetivo deste processo é encontrar padrões válidos e potencialmente úteis nos dados. Como ressalta [2], as tecnologias computacionais como KDD estão sendo cada vez mais utilizadas para auxiliar analistas em seus trabalho de investigação e análise de enormes conjuntos de dados, a fim de se extrair conhecimento.

A extração de conhecimento a partir de dados pode ser entendida como um processo contendo, pelo menos, os seguintes passos: compreensão do domínio da aplicação, seleção e preparação dos dados, data mining, avaliação do conhecimento extraído e consolidação e utilização do conhecimento extraído. Sendo a etapa de data mining considerada o núcleo do processo de KDD, em razão ser nela que são efetivamente aplicados os métodos e algoritmos

¹ Ádamo Lima de Santana, Carlos Renato Lisboa Francês. João Crisóstomo Weyl Albuquerque Costa, Laboratório de Planejamento de Redes de Alto Desempenho (LPRAD), Universidade Federal do Pará, Caixa Postal 479 - 66.075-110, Belém, PA, Brasil. {adamo, rfrances, jweyl}@ufpa.br.

² Paulo de Tarso, Fundação Nacional de Saúde, Av. Visconde de Souza Franco, 616, Reduto, 66.053-000, Belém, PA, Brasil. pptarso@funasa-pa.gov.br.

Environmental and Health World Congress 2006

que possibilitam a extração de conhecimentos a partir da base de dados.

Neste trabalho foi utilizado para a etapa de data mining o algoritmo de inteligência computacional conhecido como redes bayesianas.

A escolha das redes bayesianas para a construção do Sistema apresentado neste trabalho deveu-se principalmente a semântica deste modelo, que permite verificar a correlação entre os atributos e, que facilita, dada a inerente representação causal dessas redes, o entendimento e o processo de tomada de decisão, por parte dos usuários desses modelos [3].

As redes bayesianas podem ser entendidas como modelos que codificam os relacionamentos probabilísticos entre as variáveis que representam um determinado domínio [4]. Esses modelos possuem como componentes uma estrutura qualitativa, representando as dependências entre os nós, e quantitativa (tabelas de probabilidades condicionais desses nós), avaliando, em termos probabilísticos, essas dependências [3][5]. Juntos, esses componentes propiciam uma representação eficiente da distribuição de probabilidade conjunta do conjunto de variáveis X de um determinado domínio [6]. A distribuição conjunta é dada por (1):

$$P(X) = \prod_{i=1}^n P(X_i | Pa_i) \quad (1)$$

sendo Pa_i os nós-pais da variável X_i

Dessa forma, é possível efetuar análises de causa e efeito a partir do cálculo da distribuição de probabilidade condicional de um conjunto de variáveis, dado o valor de um conjunto de variáveis de evidências.

3. INFRA-ESTRUTURA E METODOLOGIA UTILIZADA

O Observatório de Saúde da Amazônia está utilizando duas abordagens para prover informações e extrair conhecimento sobre parâmetros de saúde da região amazônica:

- Utilização das bases de dados já disponíveis nas diversas instituições integrantes do projeto: ADA/SUDAM (Agência de Desenvolvimento da Amazônia/Superintendência de Desenvolvimento da Amazônia), SIPAM, Museu Emílio Goeldi, FUNASA (Fundação Nacional de Saúde), DATASUS (Departamento de Informática do Sistema Único de Saúde), SESPA (Secretaria Executiva de Saúde Pública do Pará), por exemplo. Nesta estratégia, atualmente, está sendo realizado o processo de convergência das diversas bases citadas para uma base única, na qual será aplicado o processo de data mining;
- Aquisição de novos dados, utilizando-se a infraestrutura de comunicação disponível nas diversas instituições participantes.

O SIPAM tem uma infra-estrutura comum e integrada

de meios técnicos destinados à aquisição e tratamento de dados e para a visualização e difusão de imagens, mapas, previsões e outras informações. Esses meios abrangem o sensoriamento remoto, a monitoração ambiental e meteorológica, a exploração de comunicações, a vigilância por radares, recursos computacionais e meios de telecomunicações.

O SIPAM conta, em sua atual estrutura, com os seguintes subsistemas: Subsistema de Aquisição de Dados; Subsistema de Tratamento e Visualização de Dados; Subsistema de Telecomunicações; Subsistema de Suporte de Transmissão; Subsistema de Auxílio à Navegação Aérea. Dentre os quais, o Observatório utiliza com maior intensidade os subsistemas Aquisição de Dados e de Telecomunicações. Sucintamente, o subsistema de Aquisição de Dados é composto pelos recursos técnicos necessários para obter os dados necessários à geração de informações atualizadas e confiáveis, utilizando os seguintes recursos: Sensoriamento Remoto por Satélite; Sensoriamento Aéreo; Sensores Meteorológicos; Plataformas de Coleta de Dados; Radiodeterminação; Rede de Exploração de Comunicações; Rede de Detecção Radar. Já o Subsistema de Telecomunicações do SIVAM é constituído por Redes e Serviços que permitem a veiculação de vários tipos de mensagens operacionais, técnicas e administrativas entre os diversos Órgãos Usuários, dentre os quais se inclui o OSA. Este subsistema utiliza os meios do Suporte de Transmissão para interligação dos usuários, compondo suas Redes e Serviços.

Além da infra-estrutura já instituída no SIPAM, há uma série de outras estruturas de coleta e comunicação de dados, disponíveis em localidades não assistidas pelo SIPAM ou atuando de forma complementar, que também estão sendo utilizadas no OSA, podendo-se citar: a rede de acesso do DATASUS, com diversos pontos de presença no Pará e na Amazônia; e a rede da SESPA, com dimensão estadual, mas que possui uma grande capilaridade.

Para as análises do estudo de caso, foi realizada, inicialmente, a aquisição das bases de dados de informações, disponibilizadas através do Observatório de Saúde da Amazônia. A partir daí é realizado o processo de convergência das diversas bases para uma base única, na qual será aplicado o processo de mineração de dados (Data Mining). A criação da base única mencionada é realizada juntamente com um especialista do domínio na área de saúde, e será composta apenas por atributos considerados relevantes às análises a serem aplicadas.

De posse das informações já trabalhadas é iniciada a etapa de extração de conhecimento, onde são aplicadas técnicas de inteligência artificial. Neste trabalho, para a etapa de Data Mining foi aplicada a técnica de redes bayesianas, permitindo que fossem calculadas as probabilidades futuras das variáveis que formam o modelo (atributos existentes na base de dados), se valendo dos

Environmental and Health World Congress 2006

valores atingidos inicialmente pelas mesmas e das relações de dependências existentes entre elas.

Após o término da etapa de Data Mining são então realizados os devidos estudos e análises.

4. APLICAÇÃO DO PROCESSO DE KDD

Nesta seção, serão aplicadas as etapas do processo KDD sobre a base de dados.

4.1. Seleção e Preparação dos Dados

Em razão da complexidade dos sistemas gerenciados pelo OSA, não há a disponibilidade de uma base de dados completa, com todos os valores e atributos específicos e relevantes para a realização das análises, dessa forma foi necessário através de um longo pré-processamento formar uma. Inicialmente verificando quais os dados relevantes para o problema, e em seguida buscando-os e, quando necessário, os adaptando ou modificando.

Nas análises feitas, trabalhadas para cada município do estado do Pará, foram verificados um total de 8 atributos, obtidos de acordo com as suas relações com o atributo principal da análise, o IDH Municipal, como apresentados a seguir:

- IDH-M, atributo correspondendo no caso ao Índice de Desenvolvimento Humano Municipal de cada município do estado do Pará. Juntamente com o Idh-M foram pegos outros atributos correlatos: IDHM-L, IDHM-E, IDHM-R;
- IDHM-L, atributo correspondente ao Índice de longevidade existente em cada município do estado do Pará;
- IDHM-E, atributo correspondente ao Índice de educação existente em cada município do estado do Pará;
- IDHM-R, atributo correspondente ao Índice de renda existente em cada município do estado do Pará;
- Renda per capita, correspondente à Renda per capita verificada em cada município do estado do Pará;
- População, contendo o número de habitantes de cada município do estado do Pará;
- Investimentos, contendo a identificação de cada investimento aplicado em cada município do estado do Pará na área da saúde. Os investimentos existentes foram discretizados em 6 faixas de acordo com o tipo de aplicação destinado ao gasto:
 1. capacitação e cursos: referente à gastos com treinamento, e capacitação e especialização de pessoal na área da saúde;
 2. apoio e assistência: referente à gastos com suporte de comunidades e projetos vigentes;
 3. programas, projetos e eventos: gastos com implantação de novos programas, projetos e eventos relacionados com a área da saúde;
 4. reforma, ampliação e aquisições de equipamento: gastos com manutenção e melhoria de instalações existentes, assim como compra de novos equipamentos;
 5. construções de centros e postos de saúde: gastos com a construção de novos centros, postos e unidades de saúde;
 6. none: utilizado para referenciar os municípios que não foram beneficiados por investimento algum.

- Valor, contendo o valor gasto com cada um dos investimentos feitos.

De posse dos bancos de dados compostos apenas dos dados relevantes à análise, iniciou-se a etapa de Data Mining.

4.2. Data Mining

Analisando-se os atributos (verificados como sendo todos contínuos menos o atributo Investimentos, que é discreto) e suas dependências com os demais foi montada a rede bayesiana apresentada a seguir.

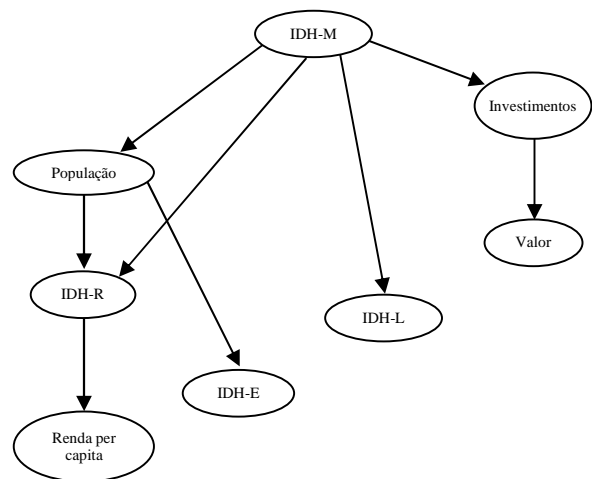


FIGURA. 1

REDE BAYESIANA CRIADA A PARTIR DA BASE DE DADOS

A partir das tabelas de probabilidade condicionais e das configurações de dependência são realizadas as inferências de modo a verificar o comportamento futuro dos dados; é então, dada a análise a ser feita ou conhecimento a ser obtido, evidenciada determinada distribuição do atributo, ou seja, é admitida a ocorrência de tal distribuição ou evento em detrimento das outras. A ocorrência desse evento é então propagada para toda a rede, sendo refletida nas distribuições dos demais atributos, agora atualizadas de modo que as distribuições de cada atributo representem agora a sua probabilidade de ocorrência dado que o evento evidenciado em questão ocorreu.

Environmental and Health World Congress 2006

Face às características exploratórias das redes bayesianas, que permitem que sejam realizadas quaisquer inferências, torna-se possível efetuar uma série de análises de dependência entre os atributos, os quais serão abordados na seção seguinte.

4.3. Avaliação do Conhecimento Extraído

De posse da rede bayesiana criada, foram iniciadas as análises e inferências sobre a mesma; essa seção aponta alguns dos resultados obtidos.

Inicialmente pode ser verificado que a maior parte dos recursos investidos destinam-se à reforma, ampliação e aquisições de equipamento, seguido pelas construções de centros e postos de saúde, deixando as demais faixas com uma quantidade de recursos irregular.

Verificou-se também que um aumento no capital investido nos municípios impulsionaria o seu IDH Municipal em pelo menos 15%, em especial no que diz respeito à longevidade.

Foi também verificada a evidencia de uma contínua maior aplicação dos investimentos na ampliação e aquisições de equipamento, ou nas construções de centros e postos de saúde (atualmente os pontos mais investidos), e que isso não causaria um impacto muito grande na melhoria do IDH Municipal, muito embora o investimento em construções de centros e postos de saúde apresentasse uma maior melhoria quando comparado com a ampliação e aquisições de equipamentos. No entanto, um maior investimento nos demais pontos, como uma maior assistência às comunidades e projetos em funcionamento, assim como incentivo e implantação de novos programas e projetos, até uma melhor capacitação do pessoal e funcionários existentes, impulsionaria o IDH Municipal em pelo menos 50%. Assim, é possível comprovar que o investimento em projetos que tenham continuidade, muito mais do que em reformas e compra de equipamentos pode gerar um bem-estar muito maior à população dos municípios estudados.

Em um outro momento, foi possível também realizar as seguintes análises sobre os demais atributos existentes.

Pôde ser visto que, analogamente ao que foi verificado na análise com o IDH-M, um maior investimento feito nos pontos de capacitação e cursos, apoio e assistência das comunidades e, na implantação de novos projetos elevaria também a renda per capita do município, ao contrário do que aconteceria se a tendência de investimentos atual prosseguisse a mesma, ou seja, um maior número de investimentos em aquisições de equipamento e construções de centros e postos de saúde.

No que diz respeito à população pôde ser visto que nos municípios que se localizam na menor faixa populacional estabelecida (de 3.780 a 19.958 habitantes) os investimentos realizados foram quase que totais em apenas aquisições de equipamento e construções de centros e postos de saúde,

apontando também o posicionamento dos seus índices de desenvolvimento humano (IDH-M, IDH-E, IDH-R e IDH-L) longe das melhores faixas; em municípios com populações nas faixas de 19.958 a 41.817 e de 41.817 a 262.538 habitantes é verificada uma mudança não muito grande com relação aos tipos de investimentos mais realizados (ainda em aquisições de equipamento e construções de centros e postos de saúde) quando comparadas com a faixa populacional anterior (de 3.780 a 19.958 habitantes), quanto aos seus índices de desenvolvimento humano, a faixa de 19.958 a 41.817 habitantes também não apresentou uma grande mudança, no entanto, os municípios na faixa de 41.817 a 262.538 habitantes possuiriam um IDH-R e principalmente um IDH-E maior, mas um IDH-L menor; por sua vez, os municípios que se apresentassem na faixa mais populosa (de 262.538 a 1.280.614 habitantes) possuiriam índices de desenvolvimento muito maiores e, quanto aos investimentos, uma ainda maior quantidade em aquisições de equipamento e construções de centros e postos de saúde, mas agora também um maior investimento na implantação de novos programas e projetos; é também nessa faixa de municípios mais populosos que estão localizados os maiores potenciais dos índices IDH-E, IDH-L e IDH-R, e onde também são aplicados os investimentos de maior valor.

Pôde ser constatado também que dentre os índices de desenvolvimento humano analisados, o índice de educação (IDH-E) é o índice que influi de maneira mais ativa no IDH-M, seguido pelo IDH-R e IDH-L. Da mesma forma, por se tratarem de atributos que constituem o índice de desenvolvimento humano municipal, eventos que causariam um aumento no atributo IDH-M implicariam também em um aumento nos atributos IDH-L, IDH-R e IDH-E.

Pode ser entendido, ao verificar a relação e o efeito que os índices de desenvolvimento humano IDH-L, IDH-E e IDH-R exercem entre si, que ao ser inferida a existência de um nível maior de ocorrência de cada um dos três índices, as seguintes tendências emergem:

- A inferência de um maior nível do IDH-R, ou seja, uma maior renda, ocasionaria em um maior aumento no índice de longevidade (IDH-L) e no índice de educação (IDH-E), respectivamente;
- Já um maior índice de educação, ocasionaria um aumento no índice de longevidade, e em seguida no de renda;
- E por fim, a ocorrência de um maior nível do índice de longevidade implicaria em um aumento primeiramente no índice de educação, seguido do índice de renda.

Além disso, como ponto principal, pôde ser visto que o atributo Valor não influencia diretamente no atributo IDH-M, mas sim o atributo Investimentos; ou seja, no que diz respeito ao índice de desenvolvimento humano municipal, o que realmente implicaria em sua melhora não é o valor aplicado, mas sim em que tipo de investimento ele seria aplicado.

Environmental and Health World Congress 2006

5. CONSIDERAÇÕES FINAIS

Os estudos realizados neste trabalho apresentaram resultados que retratavam o sistema e a realidade atual existente nos municípios do estado do Pará, no que diz respeito aos aspectos de saúde da população e da atuação e trabalho investido pelo governo.

O estado atual pode ser retratado de maneira confiável através dos dados e informações existentes nas bases de dados utilizadas, o que proveu meios para que além dessa visão inicial pudessem também ser feitas prospecções sobre o comportamento desses dados. A possibilidade de se realizar tais prospecções são valiosas e imprescindíveis no processo de tomada de decisões, permitindo uma melhor visão sobre os impactos que cada ação em particular pode ter com o resto do sistema, ou que outros pontos ela viria a afetar e com que intensidade; permitindo assim uma melhor visualização do modelo trabalhado, assim como a importância de cada elemento dele, seja de maneira singular ou em conjunto, conseqüentemente provendo meios para uma escolha mais consciente das atitudes a tomar.

A eficiente aplicação dos métodos executados neste trabalho, assim como seus resultados, trabalhando unicamente com os municípios paraenses, vem a validar a sua utilização para o trabalho do Observatório de Saúde da Amazônia não somente com o estado do Pará, mas também com todo o restante da região amazônica.

REFERÊNCIAS BIBLIOGRÁFICAS

- [1] Fayyad, U., Piatetsky-Shapiro, G., Smyth, P., "The KDD Process for Extracting Useful Knowledge from Volumes of Data", *Communication of the ACM*, vol. 39, nº 11, 1996, p. 27-34.
- [2] Turban, E., Aronson, J. E., "Decision Support Systems and Intelligent Systems", 6ª Ed. Prentice-Hall, 2001.
- [3] Korb, K. B., Nicholson, A. E., "Bayesian Artificial Intelligence". CRC PRESS, 2003.
- [4] Russell, S., Norvig, P., "Artificial Intelligence - A Modern Approach", Prentice-Hall, 2003.
- [5] Chen, Z., "Data Mining and Uncertain Reasoning - an Integrated Approach", John Wiley Professional, 2001.
- [6] Pearl, J., "Probabilistic Reasoning in Intelligent System", Morgan Kaufmann Publishers, 1988.