

UNIVERSIDADE FEEVALE

FABRÍCIO HENRIQUE RODRIGUES

EXTRAÇÃO DE MODELOS BAYESIANOS A PARTIR DE
ONTOLOGIAS CONSTRUÍDAS PARA DIAGNÓSTICO
MÉDICO

(Título Provisório)

Anteprojeto de Trabalho de Conclusão

Novo Hamburgo
2011

FABRÍCIO HENRIQUE RODRIGUES

EXTRAÇÃO DE MODELOS BAYESIANOS A PARTIR DE
ONTOLOGIAS CONSTRUÍDAS PARA DIAGNÓSTICO
MÉDICO

(Título Provisório)

Anteprojeto de Trabalho de Conclusão
de Curso, apresentado como requisito
parcial à obtenção do grau de Bacharel
em Sistemas de Informação pela
Universidade Feevale

Orientadora: Marta Rosecler Bez

Coorientadora: Cecília Dias Flores

Novo Hamburgo
2011

RESUMO

O domínio médico tem representado terreno fértil para avanços em informática. Entre eles está o SimDeCS – um sistema educativo para simulação de casos clínicos e avaliação da atuação do aluno sobre eles. Dada a incerteza inerente à área de diagnóstico médico, no SimDeCS o conhecimento do domínio é internamente representado na forma de redes bayesianas. No entanto, pela complexidade dessa representação, a aquisição de conhecimento torna-se custosa. Assim a representação por ontologias tem sido cogitada, por sua maior simplicidade. Entretanto, ela não é completamente adequada por não suportar raciocínio probabilístico. Dessa forma, este trabalho propõe o desenvolvimento de uma técnica para extração de modelos bayesianos a partir de ontologias criadas para diagnóstico médico, de modo a conciliar as duas abordagens, aproveitando as forças de cada uma delas.

Palavras Chave: Inteligência Artificial. Redes Bayesianas. Ontologias. Informática Médica. Informática na Educação.

SUMÁRIO

MOTIVAÇÃO	5
OBJETIVOS	8
METODOLOGIA	9
CRONOGRAMA	12
BIBLIOGRAFIA	13

MOTIVAÇÃO

Tradicionalmente o ensino de medicina tem sido conduzido por uma dinâmica em que os diversos conteúdos são apresentados de forma segmentada e isolada, deixando de evidenciar sua natural inter-relação e necessária articulação quando de sua aplicação a casos clínicos reais e ao cotidiano da profissão em geral. Atualmente, entretanto, diversas iniciativas têm sido tomadas no intuito de corrigir essa falha. Uma delas consiste na utilização de simulações. Simulações provêm um meio para uma abordagem holística dos problemas, possibilitando, estimulando e mesmo requerendo o relacionamento dos diversos conhecimentos necessários para sua resolução.

Dentro desse conjunto de esforços caracterizado por simulações, têm-se o SimDeCS. Esse projeto consiste num ambiente de aprendizagem multiagente que propõe casos clínicos para diagnóstico pelo estudante. O sistema, então, monitora o processo de resolução e avalia a qualidade da solução proposta e do caminho percorrido para sua obtenção. Nele a aprendizagem ocorre por um processo de negociação pedagógica, em que a constatação do aprendizado é dada quando resta estabelecida uma relação de confiança das partes na capacidade do aprendiz de aplicar, corretamente, em ocasiões futuras, o conhecimento construído durante a negociação (FLORES, 2005)

O núcleo do SimDeCs, responsável pelo acompanhamento da interação do estudante e sua avaliação, é baseado no AMPLIA - Ambiente Multiagente Probabilístico Inteligente de Aprendizagem, originalmente desenvolvido para apoiar o desenvolvimento da formulação de hipóteses e raciocínio probabilístico nos estudantes de medicina (FLORES et al, 2005). Ele é constituído por uma comunidade de 3 agentes autônomos e cooperativos, com distintas capacidades e responsabilidades: o Agente Aprendiz, que representa o aluno e monitora sua interação com o sistema; o Agente de Domínio, que representa o conhecimento do especialista, detém os casos clínicos e avalia a qualidade da solução fornecida pelo aluno; e o Agente Mediador, que realiza a negociação pedagógica entre os outros dois agentes, avaliando o desempenho geral do aluno e decidindo pela forma de intervenção mais adequada, quando necessário. Nesse ambiente, o conhecimento do especialista – que, para os fins do sistema, constitui-se, basicamente, das relações entre sintomas e doenças necessárias para a efetivação do diagnóstico em cada situação – é representado por redes bayesianas (i.e. um formalismo para modelar domínios que envolvam incerteza e informação incompleta – o como é o

caso da área de diagnóstico médico –, de cunho probabilístico e baseado no teorema de Bayes, representado por grafos acíclicos em que os nós constituem variáveis aleatórias captadas do domínio e as arestas, suas relações (PEARL, 1993).

Diferentemente do uso inicial do AMPLIA – em que a negociação pedagógica dava-se, em linhas gerais, pela submissão, pelo aluno ao sistema, de uma rede bayesiana para diagnóstico no caso em questão e sua comparação com aquela fornecida pelo especialista para o mesmo caso – a negociação no SimDeCS ocorre pela atuação médica do aluno sobre um caso clínico. Para isso, casos clínicos são formulados e inseridos na base de casos do sistema por professores, com o uso de uma linguagem de domínio específico de alto nível – a VR-Med (MOSSMANN, 2010). Com essa linguagem – de paradigma visual – é possível definir a conduta esperada do médico na abordagem do problema – desde a anamnese até o tratamento dado ao paciente. A VR-Med também possibilita resgatar do AMPLIA os fragmentos de conhecimento necessários para o diagnóstico na situação proposta, articulando-os de forma apropriada para a inserção no caso clínico.

Entretanto, apesar de sua estrutura consistente com os fins pedagógicos a que se propõe, o SimDeCs sofre de uma limitação prática: a aquisição de conhecimento especialista. Tendo em vista que a representação do conhecimento do domínio é feita com a utilização de redes bayesianas – um formalismo matemático-computacional, não recorrente na área médica – o processo de inserção do conhecimento clínico no sistema torna-se sensivelmente custoso ao especialista – que geralmente não tem a proficiência em redes bayesianas entre suas competências. Essa circunstância pode comprometer a extensão da base de conhecimento do sistema, limitando sua aplicação e eficácia.

Com esse problema em mente, foi proposta a alteração na maneira de retratar o conhecimento para inserção no sistema. A alternativa cogitada foi o uso de ontologias – uma especificação formal e explícita de conhecimento (GRUBER, 1993), descrevendo conceitos e suas relações dentro de um dado domínio. Essa forma de representação do conhecimento revela-se pertinente por três razões principais: i) tem grande poder de expressão, ii) está mais próximo da linguagem do especialista médico e iii) é largamente utilizada para modelagem de conhecimento médico, como demonstrado em [OBOFOUNDRY.ORG, 2011].

Contudo, ao passo que supera a limitação anteriormente apresentada, o uso de ontologias introduz uma nova limitação: o determinismo do diagnóstico. Como o mecanismo de inferência sobre ontologias é fortemente associado à lógica de primeira ordem, diagnósticos feitos com base nelas são, em geral, de caráter determinístico (RUSSEL, NORVIG, 2004), exigindo completude de informação – o que nem sempre é realidade cotidiano médico. Além disso, o uso da lógica de primeira ordem para lidar com domínios como o de diagnóstico médico falha por três principais razões: i) Custo elevado: o trabalho necessário para listar todos os antecedentes e consequentes para assegurar uma regra sem exceções e para o posterior uso de tais regras pode tornar a tarefa inviável; ii) Ignorância teórica: não se possui o conhecimento completo do domínio; iii) Ignorância prática: ainda que todas as regras necessárias fossem conhecidas, a informação completa sobre um caso específico pode não estar acessível (RUSSEL, NORVIG, 2004).

Dessa forma, tem-se um compromisso entre a facilidade de representação do conhecimento e a robustez do diagnóstico que a representação pode proporcionar. De um lado, as ontologias simplificam a inserção de conhecimento no sistema, limitando, porém, sua capacidade diagnóstica; de outro, as redes bayesianas possibilitam diagnósticos em situações de incerteza, todavia tornando sua aquisição mais custosa.

Apesar dessa aparente incompatibilidade, encontram-se em (BUCCI, SANDRUCCI, VICARIO, 2011) relatos de esforços no sentido de conciliar as abordagens conflitantes. Segundo eles, com algumas extensões na representação ontológica do conhecimento para incluir informações de caráter probabilístico, é possível o desenvolvimento de algoritmos que construam modelos bayesianos a partir de ontologias. Revelando-se viáveis, técnicas dessa natureza caracterizam uma solução às limitações atualmente enfrentadas pelo SimDeCS. Assim, o desenvolvimento de uma estrutura para extração automática de redes bayesianas a partir de conhecimento formatado em ontologias qualifica-se como uma contribuição relevante para o projeto e, conseqüentemente, para a área de pesquisa, razão pela qual esse objetivo é perseguido neste trabalho.

OBJETIVOS

Objetivo geral

Desenvolver uma técnica para extração automática de redes bayesianas a partir de conhecimento representado na forma de ontologias para utilização em diagnóstico médico

Objetivos específicos

- Pesquisar o estado-da-arte no estudo de redes bayesianas e ontologias.
- Estudar o estado-da-arte da aplicação de redes bayesiana e ontologias à área médica.
- Explorar técnicas relacionadas à proposta do presente trabalho.
- Explorar as ferramentas com que a técnica proposta terá de interagir.
- Projetar os algoritmos para a extração de redes bayesianas a partir de ontologias.
- Implementar os algoritmos projetados.
- Validar a técnica proposta.

METODOLOGIA

A figura a seguir apresenta um resumo da metodologia que guiará este trabalho (com os atributos que a caracterizam em vermelho).

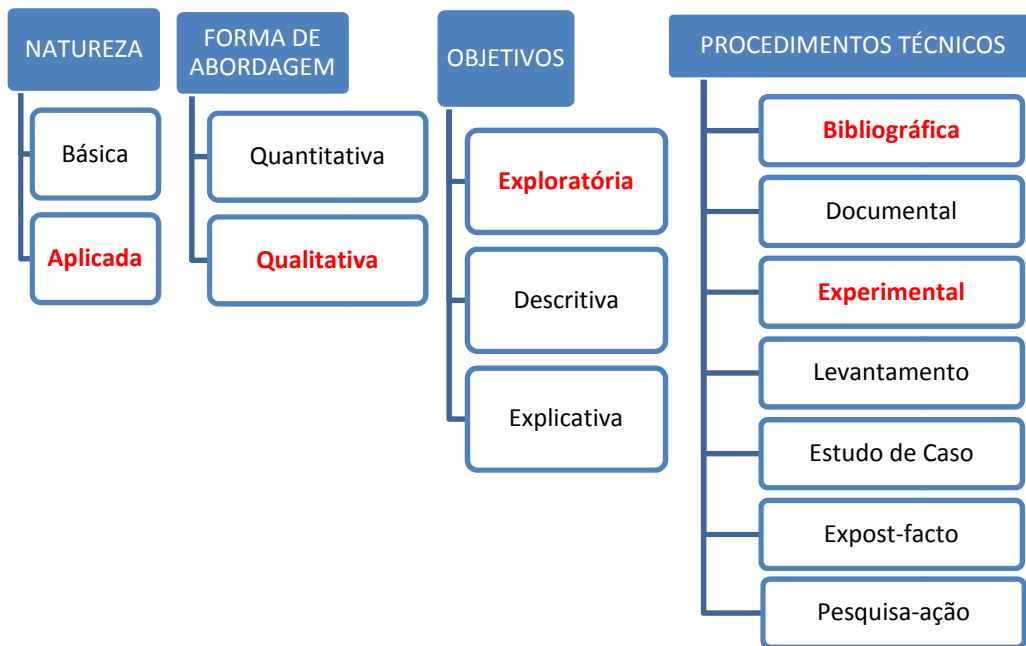


Figura : Classificação da pesquisa (adaptado de BEZ, 2011)

O presente trabalho caracteriza-se como pesquisa aplicada, dado que seu produto (i.e. uma técnica para extração de redes bayesianas a partir de ontologias para diagnóstico médico) visa à solução de um problema de ordem prática (i.e. a tradução de uma representação de conhecimento acessível ao especialista médico para um formato apropriado para uso pelo SimDeCS para avaliação de diagnósticos propostos). Nele serão articulados conhecimentos básicos, já estabelecidos na área de Ciências da Computação, para uso em um caso real, validando e demonstrando a extensão de sua aplicação, sem o objetivo de formular novos conceitos ou mudar algum paradigma.

O projeto será abordado de forma qualitativa, pois a avaliação do trabalho será feita com base na conformidade da solução com o problema proposto.

Os objetivos deste trabalho permitem enquadrá-lo no conceito de pesquisa exploratória. Nele não se lidará com um fenômeno já conhecido – seja decrevendo sua dinâmica, seja buscando as razões para sua ocorrência ou para sua ordem subjacente –, mas sim uma nova solução será proposta e averiguada quanto à adequação ao contexto

em que ela está inserida (i.e. o problema a ser resolvido). Essa estrutura justifica o caráter exploratório da pesquisa, tendo em vista que se buscará ampliar a fronteira conhecida da aplicação das técnicas abordadas.

Quanto aos procedimentos técnicos, considera-se o trabalho como bibliográfico e experimental.

Sua caracterização como bibliográfico advém da necessidade de realização de pesquisa nas publicações da área para apropriar-se do estado-da-arte do estudo de ontologias, de redes bayesianas e dos esforços para conversão entre as duas formas de representação, a fim de identificar tanto as melhores práticas quanto as limitações do que já foi produzido. Com isso visa-se a, numa recombinação das melhores práticas e/ou esforço para superar as limitações, poder atingir algum avanço no limiar do campo de estudo – necessário numa pesquisa exploratória – e assim contribuir com a área de estudo.

Ele também classifica-se como experimental em razão, principalmente, do seu método de validação. Baseando-se na pesquisa bibliográfica feita, será proposta uma nova solução para um problema prático específico, que posteriormente terá de ser validada. Entretanto, não parece adequado que tal procedimento seja feito por meios puramente teóricos. Isso resulta do fato de que, tratando-se de uma aplicação prática – a cujo escopo certamente pertencem variáveis e relações não trivialmente formalizáveis – provavelmente não seja factível determinar sua validade com o uso exclusivo de operações lógico-formais. Para esses casos, faz-se necessária uma prova empírica. Pelo caráter abstrato que é próprio dos processos e algoritmos, não há como realizar prova empírica por observação direta de um objeto. Assim, torna-se necessário induzir a ocorrência do fenômeno para trabalhar com suas variáveis e observar seu comportamento – o que é exatamente a finalidade da pesquisa experimental [PRODANOV, FREITAS, 2009]. Dessa forma, será proposta uma técnica para extração automática de redes bayesianas a partir de ontologias, a qual será implementada e então confrontada com o problema identificado – o que constitui o experimento a ser realizado.

Levando em conta esses aspectos, a metodologia que norteará o trabalho será, em linhas gerais:

- Identificação do problema prático a ser resolvido
- Aprofundamento bibliográfico na área de interesse

- Articulação do conhecimento já existente numa solução nova a ser aplicada a um novo contexto
- Implementação da solução (construção do experimento)
- Validação da solução em termos qualitativos, por via experimental.

CRONOGRAMA

Trabalho de Conclusão I

Etapa	Meses			
	Ago	Set	Out	Nov
Escrita do anteprojeto.				
Revisão do anteprojeto.				
Entrega do anteprojeto				
Estudo do estado-da-arte em redes bayesianas e ontologias				
Estudo de aplicações de redes bayesianas e ontologias na área médica				
Estudo do contexto em que a técnica a ser desenvolvida estará situada (aplicações relacionadas, ...)				
Delineamento da técnica a ser proposta				
Redação do TCC I.				
Revisão do TCC I.				
Entrega do TCC I.				

Trabalho de Conclusão II

Etapa	Meses			
	Mar	Abr	Mai	Jun
Aprofundamento teórico das particularidades da técnica a ser proposta				
Desenvolvimento, implementação e testes dos algoritmos básicos para viabilização da técnica				
Especificação detalhada e proposição da técnica				
Projeto da técnica proposta				
Implementação da técnica proposta				
Testes				
Aplicação ao problema e validação				
Redação do TCC II.				
Revisão do TCC II				
Entrega do TCC II.				
Apresentação dos resultados à banca avaliadora.				

BIBLIOGRAFIA

- BUCCI, G.; SANDRUCCI, V; VICARIO, E. *Ontologies and Bayesian Networks in Medical Diagnosis*. Em: **Proceedings of the 44th Hawaii International Conference on System Sciences**, 2011.
- DA COSTA, P. C. G.; LASKEY, K. B.; LASKEY, K. J.; Pr-owl: *A bayesian ontology language for the semantic web*. Em **ISWC-URSW**, 2005, p. 23-33.
- DEVITT, A.; DANEV, B.; MATUSIKOVA, K.; *Constructing bayesian networks automatically using ontologies*. **Applied Ontology**, IOS Press, 2006, p. 1-11.
- DING, Z.; PENG. Y; PAN, R.; *A Bayesian approach to ontology mapping*. Em **International Semantic Web Conference**, 2005, p. 563-577.
- FENZ, S.; TJOA, A. M.; HUDEC, M.; *Ontology-based generation of bayesian networks*. Em **International Conference on Complex, Intelligent and Software Intensive Systems**, 2009.
- FLORES, C. D. **Negociação Pedagógica Aplicada a um Ambiente Multiagente de Aprendizagem Colaborativa**. 2005. 121 f. Tese (Doutorado em Ciência da Computação) – Instituto de Informática, Universidade Federal do Rio Grande do Sul, UFRGS, Porto Alegre, 2005.
- FLORES, C. D.; SEIXAS, L.; GLUZ, J. C.; VICARI, R. M.; PATRICIO, D. I.; GIACOMEL, F.; GONÇALVES, L. *AMPLIA Learning Environment Architecture*. In: 2nd International Conference on Machine Intelligence, ACIDCA-ICMI'2005, 2005, Tozeur. **Proceedings of International Conference on Machine Intelligence**, 2005. v. 1. p. 501-506.
- GRUBER, T. R. *A Translation Approach to Portable Ontology Specifications*. Em: **Knowledge Acquisition**, 5(2): 1993, p.199-220.
- HELSPER, E. M.; VAN DER GAAG, L. C.; *Building Bayesian networks through ontologies*. Em **Proceedings of the 15th European Conference on Artificial Intelligence, ECAI 2002**, Lyon, França, 2002, p. 680-684.
- MOSSMANN, J. B., MARONI, V.; DAHMER, A.; FLORES, C. D.; PINHO, M. *VRMED: Linguagem de Domínio Específico para Ambientes Virtuais Aplicados à Educação Médica*. In: XII Congresso Brasileiro de Informática em Saúde 2010. Porto de Galinhas/PE. Anais. 2010. CD-ROM.
- OBOFOUNDRY.ORG; *The Open Biological and Biomedical Ontologies*. **Berkeley Bioinformatics Open Source Project**. Disponível em: <http://www.obofoundry.org/>. Acesso em: 07/09/11
- PEARL, J. *Belief networks revisited*. **Artificial Intelligence**. 1993, v.59, p.49-56.

PRODANOV, C. C.; FREITAS, E. C.; **Metodologia do Trabalho Científico – Métodos e Técnicas da Pesquisa e do Trabalho Acadêmico**. Editora Feevale, Novo Hamburgo, 2009.

RUSSELL, S.; NORVIG, P. **Inteligência Artificial** (tradução da segunda edição). Editora Elsevier, Rio de Janeiro, 2004.

ZHENG, H. T.; KANG, B. Y.; KIM, H. G.; *An ontology-based bayesian network approach for representing uncertainty in clinical practice guidelines*. Em **6th International Semantic Web Conference, ISWC 2007**, Workshop on Uncertainty Reasoning for the Semantic Web, 2007.