



19 Congresso de Iniciação Científica

APLICAÇÃO DA LÓGICA NEBULOSA A MODELO DE CONHECIMENTO DO PROCESSO DE MINERAÇÃO DE DADOS

Autor(es)

ANDRE DE ANDRADE BINDILATTI

Orientador(es)

ANDERSON BERGAMO, ANA ESTELA ANTUNES DA SILVA

Apoio Financeiro

FAPIC/UNIMEP

1. Introdução

Os primeiros esforços realizados em pesquisas sobre inteligência artificial (IA) buscaram por desenvolver modelos computacionais para a resolução de problemas de natureza geral; modelos que fossem capazes de reproduzir o processo de inferência necessário para a busca de soluções para problemas genéricos. O trabalho de Allen Newell e Herbert Simon (1959) com o GPS (General Problem Solver) é um exemplo.

A busca por modelos de propósito geral se mostrou um desafio difícil de alcançar. Sistemas Especialistas (SEs) surgiram então, como uma nova abordagem nas pesquisas em IA, na qual o foco consistia na modelagem do conhecimento específico de um domínio para a resolução de problemas inerentes a este.

Ao consultar um paciente, um médico utiliza seus conhecimentos sobre medicina para produzir um diagnóstico. Neste processo, sua competência e experiência no desempenho desta tarefa, são empregadas. Conhecimento pertinente a outros domínios são, normalmente, desnecessários na realização dessa atividade específica.

Baseados neste princípio, SEs são sistemas de computador cuja estratégia para a resolução de problemas referentes a um dado domínio, consiste na tentativa de modelar o conhecimento e heurísticas pertinentes a este, de modo a emular a competência de um especialista humano da área.

A arquitetura de um SE é dividida em três componentes principais: uma interface de usuário, responsável pela interação do usuário com a competência modelada pelo sistema; uma base de conhecimento, que armazena uma representação do conhecimento sobre o domínio de interesse; e um mecanismo de inferência, que possui um algoritmo responsável por simular o processo de raciocínio, de modo a proporcionar as respostas do sistema.

Com fundamento na lógica nebulosa, SEs podem ser estendidos permitindo o processamento sobre incerteza, imprecisão e por meio do uso de variáveis linguísticas. SEs baseados em lógica nebulosa possuem uma base de conhecimento expressa no formato de regras de inferência nebulosas (fuzzy inference rules); regras do tipo *Se (condição) Então (consequência)*.

Uma variável linguística, assim como seu próprio nome sugere, é uma variável que assume valores linguísticos, isto é, uma variável que pode assumir palavras (rótulos) como valor (SILER; BUCKLEY, 2005).

Introduzida em 1965 pelo Dr. Lotfi A. Zadeh (ZADEH, 1965), a teoria dos conjuntos nebulosos (do inglês, fuzzy sets theory) consiste em uma teoria sobre conjuntos que define o conceito de pertinência de um dado elemento a um conjunto, por meio de **uma função de pertinência** ou **função característica**.

Neste trabalho, a lógica nebulosa foi utilizada para representação de conhecimento em um modelo que tem como finalidade fornecer apoio na decisão de qual das tarefas de mineração empregar sobre um determinado conjunto de dados.

Conforme Han e Kamber (2006), a mineração de dados é uma das etapas do processo de descoberta de conhecimento em bases de dados, e envolve a escolha e aplicação de algoritmos para a extração de conhecimento ou padrões, implícitos em tais fontes de dados.

Existem diversas tarefas de mineração (BERRY; LINOFF, 2004), contudo, este projeto foca na decisão entre as tarefas de associação, classificação e agrupamento. Tarefas de mineração, como mencionado anteriormente, compreendem um conjunto de técnicas e algoritmos para a análise de padrões e extração de conhecimento de conjuntos de dados. Cada tarefa pressupõe um formato sobre os dados de entrada e em geral, produzem como saída, conhecimento expresso de diferentes formas.

2. Objetivos

Este trabalho teve como objetivo específico, realizar a aplicação da lógica nebulosa para a modelagem de uma base de conhecimento. Um projeto anteriormente desenvolvido (BINDILATTI, 2010), realizou a implementação de um mecanismo de inferência e de uma base de conhecimento para a ferramenta Kira (MENDES, 2009).

O presente trabalho procurou aprimorar tal base de conhecimento, por meio da aplicação da lógica nebulosa em sua modelagem. Além disso, a implementação prática da base de conhecimento e de um mecanismo de inferência baseados na lógica nebulosa, também foram objetivos deste projeto.

Os objetivos genéricos foram:

- O estudo de sistemas especialistas;
- O estudo da lógica nebulosa como meio para representação do conhecimento;
- O estudo das principais tarefas de mineração de dados;
- O estudo de mecanismos de inferência baseados na lógica nebulosa;

3. Desenvolvimento

Para o desenvolvimento deste projeto, um levantamento bibliográfico sobre a lógica nebulosa foi realizado com base na literatura disponível. Com os fundamentos teóricos pesquisados, a lógica nebulosa foi utilizada como ferramenta para fornecer uma nova modelagem para a base de conhecimento desenvolvida em um projeto anterior (BINDILATTI, 2010).

Regras linguísticas do tipo *Se-Então* foram utilizadas para a representação do conhecimento. Logo, as variáveis ou fatos relacionados à base de conhecimento foram definidos como variáveis linguísticas e assim, seus respectivos valores são termos linguísticos cujo significado é atribuído por conjuntos nebulosos.

Funções de pertinência triangulares e trapezoidais foram utilizadas para a modelagem dos termos linguísticos. Testes e experimentos foram efetuados a fim de encontrar definições adequadas para as funções de pertinência utilizadas para representação dos termos linguísticos, e das regras da base de conhecimento.

O modelo de conhecimento foi validado com base em dois contextos de aplicação fornecidos por especialistas do domínio. As respostas fornecidas pelo sistema estão de acordo com as respostas dos especialistas do domínio. Com isso, constatou-se que o modelo

de conhecimento se adéqua para o comprimento dos objetivos estabelecidos para este trabalho.

As entradas de dados são feitas por meio da resposta a um questionário. O usuário que interage com a base de conhecimento deve atribuir valores numéricos no intervalo fechado $[0,10]$, de acordo com o quanto o conceito levantado pela questão, se aplica ao contexto do problema do usuário.

Três variáveis linguísticas, cada uma referente a uma das tarefas de mineração de dados, associação, classificação e agrupamento, representam as saídas do sistema. Após o processo de “defuzzyficação”, realizado com o uso do cálculo do centro de massa, saídas numéricas são produzidas, indicando o quanto favorável é a aplicação de cada uma das tarefas.

Um programa foi implementado para permitir testes e a avaliação da base de conhecimento. Um *framework* de classes foi desenvolvido, fornecendo classes para a representação de termos linguísticos, regras de inferência, conjuntos nebulosos e funções de pertinência, além da implementação de um mecanismo de inferência baseado no método Mamdani.

Tal *framework* de classes foi escrito utilizando a linguagem C# e pode ser reutilizado em futuros projetos, podendo ser facilmente estendido para se adaptar a novas necessidades ou possivelmente, para ser utilizado em outros domínios de conhecimento.

4. Resultado e Discussão

A fim de atingir os objetivos deste projeto, um estudo sobre a lógica nebulosa foi realizado com o intuito de aplicá-la na representação de conhecimento de uma área específica. A representação de conhecimento consiste na adoção de uma forma sistemática e estruturada que possibilite que o conhecimento necessário para resolução de determinados problemas, possa ser armazenado e processado em sistemas computacionais.

Esforços na modelagem de uma base de conhecimento foram realizados anteriormente em um projeto de iniciação científica (BINDILATTI, 2010). A referida base de conhecimento possui como domínio de interesse, o conhecimento sobre tarefa de mineração de dados.

A finalidade de tal base de conhecimento é oferecer apoio à decisão entre qual das tarefas de mineração de dados, melhor serve aos propósitos daquele que conduz o processo de mineração. Para tal, o conhecimento foi modelado na forma de regras de produção.

A modelagem de conhecimento envolve a representação de heurísticas. O conhecimento de um especialista sobre seu domínio de competência é, na maior parte do tempo, oriundo da combinação de conhecimentos a cerca do problema e conhecimento obtido de experiências anteriores no desempenho de tarefas relacionadas.

Ao tentar descrever o conhecimento implícito na mente de um especialista, o uso de termos linguísticos é bastante comum, tal como na seguinte expressão: “quando a temperatura for **alta**, ativar o sistema de arrefecimento”. O uso de termos imprecisos é frequentemente constatado.

Neste aspecto, a lógica nebulosa fornece uma ferramenta com uma boa fundamentação teórica, que permite uma maneira formal de descrição para termos vagos e expressões que possuem graus parciais de confiança, normalmente inerentes aos domínios de interesse modelados durante o processo de engenharia de conhecimento. Portanto, o uso da lógica nebulosa na representação de conhecimento é justificável.

Um conjunto de regras de inferência foi elaborado como parte dos resultados obtidos neste trabalho. Quinze regras foram propostas, uma amostra com três destas regras é fornecida a seguir:

- **Se** (transações = muito provável) e (Marketing Basket = provável) e (Decisão imediata = provável) e (frequência entre itens = muito provável) **Então** Associação = muito alto;
- **Se** (atributo alvo = parcialmente provável) e (padrões ocultos = parcialmente provável) **Então** Classificação = médio;
- **Se** (atributo alvo = provável) e (semelhança entre os dados = pouco provável) e (atributos numéricos = pouco provável) **Então** Agrupamento = Baixo;

Um programa foi então implementado para a realização de testes e experimentos. A Figura 1 ilustra a tela de entrada e saída de dados do programa desenvolvido; e a Figura 2 exibe a tela de visualização dos conjuntos nebulosos utilizados para a definição dos termos linguísticos utilizados.

Para tal implementação, um *framework* de classes para a representação de conjuntos nebulosos, funções de pertinência, regras e mecanismo de inferência, foi projetado. A Figura 3 exibe um diagrama de classes que ilustra as principais classes do *framework*.

Como pode ser observado no diagrama da Figura 3, as classes do *framework* representam conceitos relacionados a sistemas de inferência baseados na lógica nebulosa. Conceitos como funções de pertinência, métodos de inferência e “defuzzificação” são representados como classes de interface; regras de inferência e conjuntos nebulosos são representadas como classes abstratas. Essa abordagem no projeto do *framework* é o que permite sua fácil extensão; no caso de necessidades futuras, para implementar novos métodos de inferência, funções de pertinência e etc., basta criar novas classes concretas que implementem as referidas interfaces e classes abstratas.

A validação da base de conhecimento foi realizada comparando as respostas obtidas pela ferramenta implementada, com as opiniões de especialistas do domínio sobre dois contextos de aplicação. Três especialistas, que estão vinculados a projetos cujo tema está relacionado com mineração de dados, foram consultados.

Dois contextos de aplicação foram fornecidos pelos especialistas. Sobre perspectiva desses contextos, os especialistas responderam às questões e sugeriram quais tarefas de mineração de dados eram esperadas como resposta. As respostas dos especialistas foram semelhantes às respostas fornecidas pelo programa desenvolvido.

Infelizmente, não foi possível encontrar um conjunto de dados grande o suficiente para a elaboração de um processo de validação mais completo, o que caracterizou uma das maiores dificuldades encontradas no desenvolvimento deste projeto. Apenas dois contextos de aplicação puderam ser obtidos por meio da colaboração de especialistas do domínio.

5. Considerações Finais

Um sistema de inferência baseado em lógica nebulosa atua de forma semelhante a uma função matemática, mapeando um conjunto de variáveis de entrada em um conjunto de valores de saída.

Encontrar modelos matemáticos para prever o comportamento de sistemas complexos se trata de uma tarefa de grande dificuldade. Representar o conhecimento de domínio por meio de regras do tipo *Se-Então*, por outro lado, consiste em uma tarefa mais simples e intuitiva.

Com o uso da lógica nebulosa foi possível modelar uma base de conhecimento para apoio na decisão sobre qual tarefa de mineração de dados aplicar sobre um dado problema.

Um *framework* de classes foi implementado para a realização do processo de inferência e representação das regras, conjuntos nebulosos, funções de pertinência, termos linguísticos e etc. Esse *framework* foi projetado para ser facilmente estendido, permitindo sua adaptação a futuras necessidades ou aplicação a outros domínios de conhecimento.

O objetivo deste modelo de conhecimento é permitir que usuários que não detém conhecimento sobre as várias tarefas de mineração, possam mesmo assim, usufruir das vantagens dessa tecnologia. Outro ponto a salientar, é que com o uso da lógica nebulosa na representação de conhecimento, as entradas de dados se tornam mais flexíveis, fazendo com que as respostas ao questionário sobre o problema sejam mais intuitivas.

Assim sendo, a aplicação da lógica nebulosa na modelagem de conhecimento, mostrou-se uma interessante abordagem para a implementação de sistemas para apoio em decisões complexas ou pouco estruturadas. Contudo, a falta de metodologias bem estabelecidas para apoio ao processo de aquisição de conhecimento, e para o processo de validação da base de conhecimento, agregam dificuldades na prática dessa abordagem.

Oportunamente, trabalhos futuros poderiam buscar por estender os resultados obtidos neste projeto, por meio da pesquisa de arquiteturas híbridas de sistemas inteligentes. Tais arquiteturas possuem a vantagem de realizar uma combinação sinérgica de diversas técnicas utilizadas na implementação de sistemas inteligentes, tais como RNAs, algoritmos genéticos e sistemas baseados em

