

**7º Congresso de Pós-Graduação****MODELAGEM DE BASE DE CONHECIMENTO PARA TAREFA DE CLASSIFICAÇÃO EM
MINERAÇÃO DE DADOS****Autor(es)**

LIDIA MARTINS DA SILVA

Orientador(es)

ANA ESTELA ANTUNES DA SILVA

1. Introdução

Um sistema especialista é delineado e concebido para atender a uma aplicação determinada e limitada da experiência humana, capaz de manifestar uma decisão, apoiado em conhecimento demonstrado, a partir de uma base de informações, tal qual um especialista de determinada área do conhecimento humano.

Sistemas Especialistas são desenvolvidos para tratar problemas complexos do mundo real que necessitem da interpretação e da análise de especialistas humanos e que ao mesmo tempo cheguem a conclusões e decisões que o especialista humano chegaria, como estivesse defrontando-se com os mesmos problemas. Sistemas Especialistas utilizam-se de regras de inferência sobre uma grande base de conhecimentos, sobre sintomas e tratamentos possíveis, para que possam identificar o problema e posteriormente oferecer um tratamento adequado, auxiliando à tomada de decisões.

A tecnologia de mineração de dados é formada por um conjunto de ferramentas que, através do uso de algoritmos de aprendizado ou baseado em redes neurais e estatísticas, são capazes de explorar um grande conjunto de dados, extraíndo, destes, conhecimento na forma de hipóteses e de regras. Diariamente as empresas acumulam diversos dados em seus bancos de dados, tornando-os verdadeiros tesouros de informação a cerca dos vários processos e procedimentos das funções da empresa, inclusive com dados e hábitos de seus clientes, suas histórias de sucesso e fracassos. Todos esses dados podem contribuir com a empresa, aludindo tendências e características pertinentes a ela e seu meio ambiente interno e externo, visando uma célere ação de seus gestores. Para a extração do conhecimento é necessário utilizar-se de tarefas de mineração de dados, tais como: associação, agrupamento e classificação. A tarefa de classificação é o processo de encontrar um conjunto de modelos (funções) que descrevem e distinguem classes ou conceitos, com o propósito de utilizar o modelo para predizer a classe de objetos que ainda não foram classificados. O modelo construído baseia-se na análise prévia de um conjunto de dados de amostragem ou dados de treinamento, contendo objetos corretamente classificados, sendo a tarefa mais estudada e a mais utilizada num processo de mineração de dados.

Diante da necessidade de ferramentas de apoio nos processos decisórios surgiu a motivação para fazer este trabalho de modelagem de uma base de conhecimento para mineração de dados, a qual utiliza a tarefa de classificação.

2. Objetivos

O trabalho tem como objetivo a modelagem de uma base de conhecimento para mineração de dados relacionais utilizando a tarefa de classificação.

3. Desenvolvimento

Um sistema especialista deve possuir uma base de conhecimento, formada de fatos, regras e heurísticas sobre o domínio, assim como um especialista humano o faz. Deve ser capaz também de dar sugestões e conselhos aos usuários, como também adquirir novos conhecimentos com essa interatividade. A base de conhecimento é a coleção de informações, as quais supostamente devem agir conforme um especialista humano. Seria, então, a “alma” do sistema especialista. Segundo Nikolopoulos (1997), ela é responsável por estruturar todo o conhecimento sobre o domínio da aplicação. A base de conhecimento é um elemento permanente, mas específico de um sistema especialista. Contém conhecimento, sob a forma de regras de produção, quadros, redes semânticas. Uma das mais comuns é por sentenças do tipo “Se – Então”, contém um somatório de fatos, de heurísticas e de crenças.

Já o princípio da tarefa de classificação é descobrir algum tipo de relacionamento entre os atributos preditivos e o atributo objetivo, de modo a descobrir um conhecimento que possa ser utilizado para prever a classe de uma tupla desconhecida, ou seja, que ainda não possui uma classe definida (FREITAS, 1998).

As técnicas de mineração de dados por classificação utilizam-se de dados sobre o passado, ou seja, conjunto de treinamento para classificar dados futuros (conjunto de execução de um modelo de classificação), dados que não pertencem ao conjunto de treinamento. Em geral, são dados coletados cronologicamente após os dados de treinamento. Uma classificação consiste em prever o valor que um determinado atributo do conjunto de execução assumirá, dado um conjunto de valores dos demais atributos do conjunto de execução. Esse primeiro atributo, chamado atributo de classificação, indica a classe a que cada instância do conjunto de execução pertence. O algoritmo indutor de classificação baseado em um modelo de classificação, é denominado de classificador a partir do conjunto de treinamento.

2.1 Metodologia

A metodologia de desenvolvimento seguiu um ciclo de vida clássico com técnicas voltadas ao desenvolvimento de sistemas especialistas. Sendo assim, foi desenvolvida a fase de aquisição de conhecimento. Essa fase foi realizada com a utilização de técnicas, tais como: Entrevistas com especialistas em mineração de dados; Estudo sobre a tarefa classificação de mineração de dados; Estudo da ferramenta Kira - Ferramenta de apoio ao analista de mineração de dados (MENDES,2009) e Análise de discurso (FERREIRA, 2006; KUIPERS e KASSIER, 1984).

A segunda fase da metodologia consiste da modelagem do conhecimento. Essa modelagem inicialmente será constituída das seguintes técnicas: Rede Semântica: para representação do conhecimento; Regras de Produção: para representação do conhecimento; Diagrama de Atividades: Para representar o fluxo de atividades da tarefa de classificação.

4. Resultado e Discussão

Apresentamos a seguir os resultados obtidos, tais como: rede semântica, regras de produção e diagrama de atividades.

Rede Semântica

Um dos pontos mais importante no projeto de um sistema especialista, é a escolha do modelo de representação de conhecimento. A linguagem incorporada ao modelo escolhido deve ser muito expressiva para permitir a representação do conhecimento a respeito do domínio escolhido de maneira completa, e sendo eficiente (FLORES, 2003). Dentre os métodos de representação, utilizou-se neste trabalho a representação do conhecimento através da rede semântica, formada por um conjunto de nodos conectados por um conjunto de arcos. Os nodos geralmente representam objetos e os arcos, relações binárias entre esses objetos. Mas os nodos podem também ser utilizados para representar predicados, classes, palavras de uma linguagem, entre outras possíveis interpretações, depende apenas do sistema de redes semânticas em questão.

A Figura 1 mostra a rede semântica que representa o domínio de conhecimento da mineração de dados utilizando a tarefa de classificação.

Regras de Produção

Regras de produção são regras no formato de SE (Condição) ENTÃO (Ação). Pode-se ainda relacionar várias condições através de conectivos lógicos (E, OU e NÃO) para a criação de condições mais complexas.

Um sistema baseado em regras segundo Rich & Knight (1993) consiste em um conjunto de regras no qual se determina a aplicabilidade da regra e a ação a ser realizada se a regra for aplicada.

As regras de produção a seguir apresentadas ainda são bastante genéricas e serão mais específicas durante o projeto e implementação do conhecimento.

R1: Se problema = definido

Então definir objetivo.

R2: Se objetivo = definido
Então elaborar plano preliminar.

R3: Se base de dados = escolhido
Então selecionar Tabela.

R4: Se tabela= selecionado
Então selecionar atributos.

R5: Se atributo = selecionado
Então definir atributo meta.

R6: Se problema é de Classificação
Então escolher Tarefa de Classificação.

R7: Se Tarefa = Classificação
Então escolher Algoritmo de Classificação.

R8: Se algoritmo de classificação = definido
Então parametrizar algoritmo.

R9: Se Algoritmo = parametrizado
Então aplicar algoritmo de classificação.

R10: Se algoritmo = aplicado
Então gerar modelo descritivo.

R11: Se algoritmo = modelo descritivo gerado
Então avaliar e interpretar as classes.

R12: Se classe avaliadas = ótimas
Então gravar arquivos.
Então gerar relatório.
Então apresentar resultados.

R13: Se classe = ruim
Então aplicar algoritmo de classificação.

Diagrama de Atividades

O objetivo do diagrama de atividades é mostrar o fluxo de atividades em um único processo. O diagrama mostra como uma atividade depende uma da outra. A Figura 2 mostra o Diagrama de Atividades da tarefa de classificação.

Figura 2. Diagrama de atividades da tarefa classificação.

5. Considerações Finais

Este trabalho teve como principal objetivo apresentar a modelagem da base de conhecimento utilizando tarefa de classificação. Para a realização deste foram estudados conceitos sobre sistemas especialistas, base de conhecimento, mineração de dados, tarefa de classificação e regras de produção.

Dentro do objetivo proposto, os resultados alcançados foram a modelagem da representação do conhecimento através de uma rede semântica mostrando o processo de mineração de dados utilizando a tarefa de classificação, regras de produção e diagrama de atividades da tarefa de classificação.

Como trabalho futuro deverá ser refinada e implementada a base de conhecimento com os conceitos referentes à tarefa de

classificação pra mineração de dados.

Referências Bibliográficas

FERREIRA, M. C. L. Análise do Discurso no Brasil: notas à sua história. (20/06/2006) Disponível em: . Acesso em: 12 Mai. 2009.

FLORES, C. D. Fundamentos dos Sistemas Especialistas. In: BARONE, D. A. C. (Ed.). Sociedades Artificiais: a nova fronteira da inteligência nas máquinas. Porto Alegre: Bookman, 2003. p.332.

FREITAS, A. A., Lavington, S. H. Mining Very Large Databases With Parallel Processing. Kluwer Academic Publishers. 1998.

HAN, J.; KAMBER, M. Data Mining - Concepts and Techniques. 2 ed. Nova York: Morgan Kaufmann, 2006.

KUIPERS, B.; KASSIER, J. P. Knowledge Acquisition by Analysis of Verbatim Protocols. This article appeared in A. Kidd (Ed.), Knowledge Acquisition for Expert Systems. New York: Plenum,1987. Cognitive Science 8: 363-385,1984.

MENDES, E. F. Automatização da Técnica de Mineração de Dados Auxiliado por Guias. Dissertação de Mestrado. Universidade Metodista de Piracicaba. Piracicaba. São Paulo, 2009.

NIKOLOPOULOS, C. Expert systems: introduction to first and second generation and hybrid knowledge based systems. New York: Marcel Dekker, Marcel Dekker Inc, 1997. 331 p.

RICH, E.; KNIGHT, K.. Artificial Intelligence. McGraw-Hill, São Paulo. 1993.

Anexos



