

INVESTIGAÇÃO DE TÉCNICAS EM *DATA MINING* PARA A VALIDAÇÃO DE INFORMAÇÃO EM BASES DE DADOS DE SELEÇÃO DE MATERIAIS.

Ary Fagundes Bressane Neto, João Fernando Marar, Leonardo Ramalho de Oliveira, Sérgio Henrique Prado Scolari – Ciência da Computação – Bacharelado em Ciência da Computação – Departamento de Computação – Faculdade de Ciências – Campus de Bauru

Informações sobre materiais e processos de fabricação estão disponíveis em diversas fontes, com diferentes conteúdos e interfaces. Existem recursos gratuitos disponíveis na Internet, e sistemas pagos, além dos tradicionais *handbooks* e do material publicitário disponibilizado por fornecedores de materiais. Entretanto, algumas fontes não acompanham o dinamismo do desenvolvimento de novos materiais. Os novos conceitos de desenvolvimento de *software* onde usuários colaboram diretamente com o conteúdo dos sites têm apontado para uma solução ao problema da falta de agilidade dos sistemas tradicionais. Dentro desses novos conceitos, o controle de usuários mostra-se de grande relevância para manter a organização dos dados e evitar um vandalismo junto às bases de dados. Esse controle pode ser feito manualmente em bases pequenas, mas se torna impraticável nas bases de grande porte.

Neste sentido, este trabalho tem por objetivo investigar técnicas de *Data Mining* para a validação e auxílio de usuários de um sistema digital de informações sobre materiais e processos de fabricação adequado à atividade de *Design* no Brasil.

A definição de *Data Mining* aceita pela maioria dos pesquisadores da área foi elaborada por Fayyad, Piatetsky-Shapiro, & Smyth como sendo: “Extração de conhecimento de bases de dados é o processo de identificação de padrões válidos, novos, potencialmente úteis e compreensíveis embutidos nos dados.” [1].

Esse processo é comumente dividido em várias etapas, existindo ainda diversas abordagens para essa divisão. Para este trabalho, foi considerada a divisão do processo em cinco etapas, conforme apresentadas na figura 1 [3] e descritas na tabela 1 [2, 4].

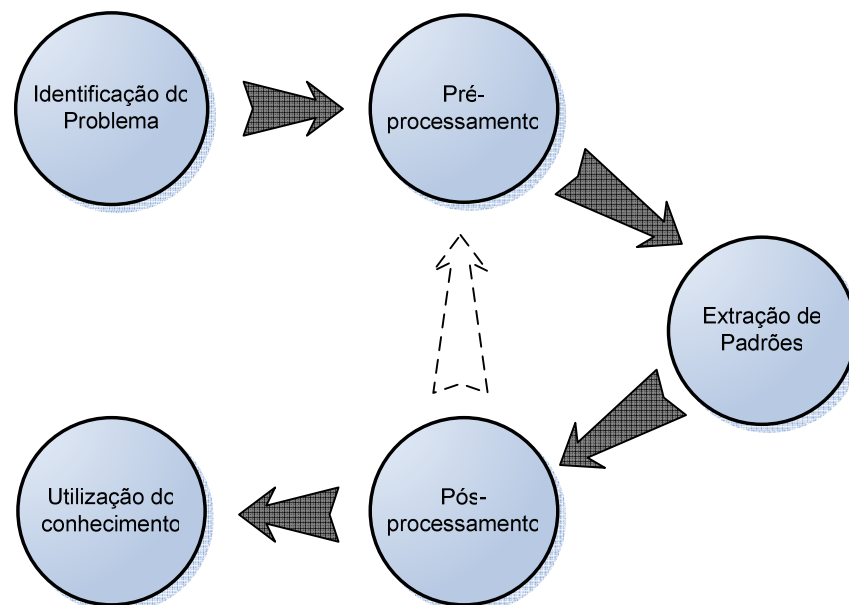


Figura 1: Etapas do processo de *Data Mining*

Etapa	Definição
Identificação do problema	Realização de um estudo inicial para identificação do domínio do problema.
Pré-processamento	Seleção dos dados mais significativos.
Extração de padrões	Algoritmos para extração dos padrões embutidos nos dados.
Pós-processamento	Avaliação dos padrões extraídos.
Utilização do conhecimento	Aplicação dos resultados obtidos.

Tabela 1: Descrição das etapas do processo de *Data Mining*

O processo de Data Mining, quando aplicado no contexto deste trabalho, apresenta na fase de pré-processamento a seleção dos atributos mais significativos de cada material ou processo, baseada na análise de um especialista. A etapa de extração de padrões é composta por dois elementos: escolha da tarefa, que analisa os objetivos, e escolha do algoritmo, que é dependente da linguagem de representação dos padrões. A tarefa selecionada para a solução do problema da validação da informação foi a atividade de predição (ou Mineração de Dados Preditivo), que generaliza exemplos ou experiências passadas por meio de respostas conhecidas em uma linguagem capaz de identificar a classe de um novo exemplo. O algoritmo escolhido foi a utilização de redes neurais artificiais, pois oferecem uma abordagem mais adequada para o resolução de problemas deste gênero. Uma rede neural artificial pode ser treinada a partir de grandes quantidades de dados. Por ser um sistema adaptativo, o aumento dos dados fornecidos melhora a capacidade de reconhecimento da rede, diminuindo assim a taxa de erro. Após o treinamento da rede, os resultados são apresentados a um especialista que indica se há a necessidade de realizar ajustes nos parâmetros da rede. O resultado final indicará se as informações fornecidas pelo usuário são consistentes

Esse estudo inicial apresentou a análise de técnicas de Inteligência Artificial, *Data Mining* e a análise de procedimentos para a validação preditiva de dados em um sistema digital de informações sobre materiais e processos de fabricação. O trabalho mostra caminhos para o desenvolvimento de sistemas autônomos inteligentes que podem fazer parte de um sistema digital colaborativo.

Referências Bibliográficas

- [1] FAYYAD, U., PIATETSKY-SHAPIRA, G., SMYTH, P. **From data mining to knowledge discovery: an overview**. Em *Advances in Knowledge Discovery & Data Mining*, 1996
- [2] HAN, J., KAMBER, M. **Data Mining: Concepts and Techniques**, Morgan Kaufmann, 2000.
- [3] REZENDE, S. O. **Sistemas Inteligentes: fundamentos e aplicações**, Manole, 2003
- [4] WITTEN, I. H., FRANK, E. **Data Mining: Practical Machine Learning Tools and Techniques**, 2 ed., Morgan Kaufmann, 2005.

Bolsa: CNPq