

# Seleção de Atributos

**A Seleção de Atributos é uma etapa da fase de pré-processamento do processo de Descoberta de Conhecimento em Banco de Dados. Como o próprio nome já diz, o objetivo é escolher um subconjunto de atributos (também conhecidos como variáveis) ou criar outros atributos que substituam um conjunto deles a fim de reduzir a dimensão do banco de dados. Com essa redução de dimensão, reduz-se a complexidade do banco de dados e assim o tempo de processamento para extrair dele algum conhecimento. Além disso, atributos desnecessários podem causar ruído no resultado final e isto pode ser evitado com a aplicação de técnicas de Seleção de Atributos.**

Primeiramente é necessário que o conjunto global de atributos seja submetido a algum algoritmo que busque um subconjunto de atributos relevantes como, por exemplo, o método Best-First (*Heuristics: Intelligent Search Strategies for Computer Problem Solving*. Addison-Wesley, 1984. p. 48.). Em seguida avalia-se a qualidade do subconjunto encontrado através de algum critério de avaliação (precisão, consistência, dependência).

Há duas abordagens principais para se realizar a Seleção de Atributos: filtro e wrapper (Kohavi R, John GH (1997). "Wrappers for feature subset selection". *Artificial Intelligence* **97** (1-2): pp. 273-324.).

Na primeira abordagem os atributos são ignorados independentemente dos valores contidos nos registros do conjunto de teste, ou seja, apenas as propriedades dos atributos são consideradas para classificá-la como útil ou não; neste caso os atributos são todos selecionados independente da execução do algoritmo de aprendizado. A seleção dos atributos ocorre de maneira estática.

Na segunda abordagem (wrapper), a seleção dos atributos também é externa ao algoritmo de aprendizado, porém ocorre a cada iteração. Novos subconjuntos de atributos são escolhidos e avaliados dinamicamente enquanto a mineração ocorre até que não seja viável realizar mais melhorias na escolha dos atributos, segundo algum critério de avaliação. Neste caso os atributos ou conjuntos de atributos são avaliados considerando os valores encontrados nos registros do conjunto de teste. Uma porcentagem de acertos nas informações geradas pelo algoritmo determina quão bom ele está ou pode melhorar e isto é utilizado como critério de parada do processo. Esta abordagem avalia os atributos usando estimativas de precisão providas por algoritmos de aprendizado pré-determinados.

A etapa de Seleção de Atributos como pode ser vista, está presente apenas no início do processo de Mineração de Dados, porém sua aplicação requer grande responsabilidade, pois toma a maior parte do processo todo de mineração e consome muito processamento daí sua importância. Tomar decisões com auxílio dos computadores é extremamente desejável, isto faz com que se impulse o crescimento desta área ainda jovem cada vez mais, com o descobrimento de algoritmos novos e aprimoramento dos já usuais. Em especial a Seleção de Atributos, por ser tão importante merece destaque na pesquisa, já que ainda não se tem uma receita pronta para a melhor combinação de técnicas. Fica, portanto aqui estabelecida a “vírgula” na linha do tempo da pesquisa sobre Seleção de Atributos e ao caro leitor pesquisar o desafio de encontrarmos juntos um “ponto” que seja o mais eficiente possível.