

Definição de pesos aos atributos em uma base de dados sem domínio definido no Raciocínio Baseado em Casos

Jonathan William Hrenechen^{b,c}, André Kosciński^{a,b}, André Pinz Borges^{a,b}

^a*Departamento Acadêmico de Informática, Pós-Graduação em Ciência da Computação, Av. Monteiro Lobato, Km 4, s/n, Ponta Grossa - Paraná - Brasil*

^b*Universidade Tecnológica Federal do Paraná, Av. Monteiro Lobato, Km 4, s/n, Ponta Grossa - Paraná - Brasil*

^c*Autor para correspondência: hrenechen@gmail.com*

Palavras-chaves: RBC, peso de atributos, data-driven

O método de Raciocínio Baseado em casos (RBC), utiliza a experiência passada para solucionar um novo problema, buscando a solução mais próxima dentro de uma base de casos. Cada caso é composto por um conjunto de atributos e uma categoria. O método de RBC depende de um especialista no domínio que vai ser aplicado para avaliar quais são as informações relevantes para gerar uma solução, neste quesito a abordagem ‘data-driven’, explora a falta deste especialista ou mesmo um especialista com pouca competência no domínio apresentado.

Na abordagem data-driven é criado um placar com a pontuação dos atributos, em que apenas uma porcentagem delas será utilizada. O placar é ordenado por relevância e para isso podem-ser aplicados métodos como Information Gain, Gain Ratio, Gini, Chi2, ReliefF.

O cálculo de pontuação é realizado dentro do intervalo [0, 1], sendo que cada dimensão (atributo) da base de dados pode receber um peso diferente. Mudanças nos pesos resultam em uma nova base de dados, sua ordenação é relacionada ao peso de cada atributo, e a porcentagem de utilização irá excluir os atributos com menor pontuação, exceto nos casos que 100% da base seja utilizada.