

369

OBTENÇÃO DE ELEMENTOS PARA UM CLASSIFICADOR BAYESIANO DE TEXTOS.

Paulo Schreiner, Leonardo Rech Dalpiaz, Manuel Menezes de Oliveira Neto (orient.) (UFRGS).

A classificação de textos é uma tarefa extremamente importante no mundo sobrecarregado de informações em que vivemos, principalmente devido ao grande volume de dados disponibilizados através da internet. Para a maioria dos casos, o uso de métodos baseado em palavras-chave não é suficiente para identificar um tipo de texto. Diante da dificuldade de criar um classificador de textos cognitivos, resta a alternativa de utilização de uma classificação estatística, mais especificamente bayesiana. As grandes vantagens desta abordagem são o baixo custo computacional, sua eficiência em comparação com métodos baseados em palavras-chaves e, principalmente, sua adaptabilidade. Um classificador bayesiano necessita ser treinado, gerando uma banco de dados com elementos que aparecem em cada categoria de texto. Estes elementos podem ser diversos, não necessariamente palavras individuais. Neste trabalho, implementamos um destes classificadores, e fizemos uma análise da influência que diversos métodos de obtenção de elementos possuem no índice de acertos, requerimento computacional e espaço ocupado para diferentes tipos de texto, e com diferentes níveis de treinamento.