



Sistema de recomendação de filmes com indução de atributo

Danilo Lopes do Carmo

Pontifícia Universidade Católica do Rio de Janeiro
Rua Marquês de São Vicente, 225 - Gávea, Rio de Janeiro - RJ
danlcarmo@live.com

Marco Antônio da Cunha Ferreira

Pontifícia Universidade Católica do Rio de Janeiro
Rua Marquês de São Vicente, 225 - Gávea, Rio de Janeiro - RJ
macfe@ele.puc-rio.br

Juliana Christina Carvalho de Araújo

Pontifícia Universidade Católica do Rio de Janeiro
Rua Marquês de São Vicente, 225 - Gávea, Rio de Janeiro - RJ
juliana@ele.puc-rio.br

Henrique Helfer Hoeltgebaum

Pontifícia Universidade Católica do Rio de Janeiro
Rua Marquês de São Vicente, 225 - Gávea, Rio de Janeiro - RJ
hhhelfer@hotmail.com

Reinaldo Castro Souza

Pontifícia Universidade Católica do Rio de Janeiro
Rua Marquês de São Vicente, 225 - Gávea, Rio de Janeiro - RJ
reinaldo@ele.puc-rio.br

RESUMO

O presente trabalho propõe o uso de indução de atributo para calibração de um sistema de recomendação que vai sugerir um determinado filme para um usuário do serviço baseado nas preferências do mesmo e nas informações dos filmes. Na etapa de pré-processamento, por meio de indução de atributos por *Fuzzy C-means* são geradas bases com três, cinco e sete *clusters fuzzy* para, posteriormente, na etapa de processamento serem aplicados modelos e técnicas de Regressão Linear Simples, Regressão Linear Bayesiana, *Boosted Decision Tree*, Redes Neurais e Regressão de Poisson. Com isso, são efetuados mapeamento inter-modelos (escolha da base de dados e melhor modelo), mapeamento intra-modelo (otimização de parâmetros do melhor modelo) e teste (aplicação do melhor modelo com parâmetros utilizados ao conjunto de teste), sendo que, a fim de comparação de resultados opta-se pela métrica *Root Mean Square Error*. No mapeamento inter-modelos é possível observar que o aumento do número de *clusters fuzzy* traz melhorias a todos os modelos, sendo a base com sete *clusters* aquela que obteve melhor resultado. Assim, a partir desta cria-se uma base B_U por meio da indução de atributos utilizando-se do peso de voto do usuário, e uma base B_{UF} , por meio da indução de atributos utilizando-se do peso de voto dos usuários e do peso de voto nos filmes. Com o aumento no número de atributos notou-se redução considerável do *Root Mean Square Error*, assim como a utilização do modelo *Boosted Decision Tree*. Por fim, diferentes configurações do modelo são testados na base B_{UF} a fim de se realizar uma análise de sensibilidade e, de forma a verificar os resultados, a melhor configuração é aplicada ao conjunto teste para consolidar os resultados. Todos os tipos de indução de atributos testados e os modelos aplicados são analisados para, por fim, o melhor resultado obtido ser comparado com a literatura especializada, de modo a confrontar a taxa de erro do modelo proposto com a de métodos consagrados.

PALAVRAS-CHAVE. Sistemas de recomendação, *Boosted Decision Tree*, *Fuzzy C-means*.