

Trabalho Prático

— Aprendizado Indutivo —

Objetivo

O objetivo do trabalho é implementar e experimentar a resolução de problemas com aprendizado por exemplos, em particular, em aplicações de data-mining.

Relatório

Conforme cronograma da disciplina, deve ser entregue um relatório descrevendo o pre-processamento dos dados e os resultados da avaliação das técnicas escolhidas.

Problema

O problema é *classificar* páginas web de conteúdo acadêmico como sendo dos seguintes tipos:

- de uma universidade
- de um departamento
- de um membro da faculdade
- de um projeto
- de um curso
- de um estudante
- ou outro tipo

Para isso iremos treinar um agente a classificar páginas web a partir de exemplos já classificados por um especialista humano. Os exemplos de páginas podem ser obtidos em <http://www.cs.cmu.edu/afs/cs.cmu.edu/project/theo-20/www/data/webkb-data.gtar.gz> e estão organizados em diretórios que indicam a classificação. Por exemplo, a pasta “student/cornell” contém páginas de alunos da universidade Cornell.

A partir destes exemplos, avalie a qualidade de pelo menos duas técnicas/algoritmos de classificação. Para isso reserve algumas exemplos para avaliação.

O trabalho pode ser feito em equipes de até quatro alunos/as.