Relatório do EP de IA

Grupo: Beatriz, Luciana, Murilo

Linguagem de programação utilizada: Java 1.7

Parte 1:

Introdução

In decision tree learning, ID3 (Iterative Dichotomiser 3) is an algorithm used to generate a decision tree invented by Ross Quinlan.[1] ID3 is the precursor to the C4.5 algorithm.

Artigo ex. 1 -> <http://www.cise.ufl.edu/~ddd/cap6635/Fall-97/Short-papers/2.htm>

Artigo ex. 2 -> <http://www.cis.temple.edu/~ingargio/cis587/readings/id3-c45.html>

Implementação do algoritmo ID3.

* A construção do algoritmo foi feita na linguagem Java. Numa primeira leitura do Cp3 do Mitchell, identificamos a necessidade de criar um arquivo com os dados da tabela 3.2, página 59. Tal arquivo é um arquivo de texto simples do Windows, com os exemplos separados por linhas e os valores dos atributos por vírgulas, conforme abaixo:

ensolarado, quente, alta, fraco, não

ensolarado, quente, alta, forte, não

nublado, quente, alta, fraco, sim

chuva, moderada, alta, fraco, sim

chuva, fria, normal, fraco, sim

chuva, fria, normal, forte, não

nublado, fria, normal, forte, sim

ensolarado, moderada, alta, fraco, não

ensolarado, fria, normal, fraco, sim

chuva, moderada, normal, fraco, sim

ensolarado, moderada, normal, forte, sim

nublado, moderada, alta, forte, sim

nublado, quente, normal, fraco, sim

* Uma das maiores dificuldades foi a decisão de como definir as classe e objetos java que atendessem aos objetivos. Nesse sentido, desenvolvemos uma primeira abordagem incorreta na qual cada atributo armazenava seu ganho e uma lista contendo os valores para cada exemplo.
* A abordagem que obteve sucesso foi mais coerente, definindo as seguintes classes:
  + Leitor: responsável pela leitura do arquivo e construção do Conjunto. Descrever melhor.
  + Conjunto: representação da tabela playtennis, contendo os exemplos (objetos da classe Registro), a lista de Atributos e uma série de métodos úteis, tais como contar o número de exemplos de determinada classe, realizar partição (dado uma valor de atributo), valor mais comum de um dado Atributo
  + Registro sabe sua classe, seu numero de atributos e quais os valores desses atributos.
  + Atributo: sabe seu nome e contém uma lista de valores possíveis que pode assumir. Exemplo:
* DESCREVER CLASSES E MÉTODOS
* Para o cálculo da entropia houve dificuldade inicialmente para ajustar os double e arredondamentos, mas depois bastou seguir a referência. O java não fornece cálculo deireto para log2, então para obter log2(n), calculamos log10(2) = .301 e dividimos pelo log10(n). É impossível calcular também, no pc, log10(0), assim seguimos mitchell (página XX)
* O ganho também foi desenvolvido facilmente com a somatória sendo realizada através de um loop for.
* Tanto entropia quanto ganho foram testados utilizando jUnit com os dados fornecidos pelo livro e slides da disciplina.
* Enfrentamos alguns problemas de acesso e cópia das listas em Java, principalmente ao criar partições.
* O algoritmo ID3 em si, uma vez que as classe e métodos estavam bem definidos, foi facilmente implementado seguindo o exemplo da página 56. [Mostrar o código Java e explicar]
* Na primeira vez, montamos o algoritmo em cima da JTree, que é uma classe da API java que permite a especificação de estruturas de pastas, conforme a figura abaixo.
* Decidimos exibir a árvore

ADULT foi alterado para incluir uma linha de cabeçalho.

Método para discretização: o mesmo usado no WEKA: <http://trs-new.jpl.nasa.gov/dspace/handle/2014/35171>