

Relatório Lista de Adjacências TEG

Alunos: Pedro Vargas e Gustavo Felipe

Resumo do algoritmo:

O algoritmo permite ler o arquivo csv, utilizando um parser que ignora todas os dados não-numéricos existentes no arquivo, já que este se encontra organizado (Os vértices de 1 até 50 representam Setosas, os vértices de 51 até 100 representam Versicolors, os vértices de 101 150 representam Virginicas). A partir dessa leitura, calculamos e montamos uma matriz com todas as distâncias entre cada par de vértice. A normalização é feita e, por fim, criamos e organizamos um arquivo txt, representando a lista de adjacências. Feito isso, é possível importar para uma estrutura de dados o grafo representado no arquivo de texto.

Funções:

cria - Pedro Vargas:

Esta função é responsável por criar a estrutura de dados a partir de um arquivo txt existente. Dado qualquer par de vértices v1 v2, que representa uma aresta, a função adiciona na posição v1-1 do vetor de vértices uma aresta com o label de v2 e adiciona, na posição v2-1, uma aresta com o label de v1. O grau de cada vértice é incrementado em uma unidade.

destrói - Pedro Vargas:

Esta função libera todo espaço utilizado pela estrutura de dados.

printar_grafo - Pedro Vargas:

Printa a lista de adjacências no terminal.

cria_tabela - Gustavo Felipe (lógica da criação da matriz) e Pedro Vargas (Parser para ler csv):

Lê o arquivo csv, calculando, para cada vértice, a distância em relação a todos os outros vértices, criando assim uma matriz de distâncias. Enquanto esse cálculo é feito, o valor da maior e menor distância é salvo, para, posteriormente, ser feita a normalização das distâncias.

normaliza - Gustavo Felipe:

Realiza a normalização da matriz de distâncias.

fazerTxt - Gustavo Felipe:

Um arquivo contendo todo par de linha e coluna que apresenta distância menor ou igual a 0.3 (considerando normalização prévia) é criado.

Calcula a acurácia - Gustavo Felipe e Pedro Vargas:

A partir do arquivo grafo.txt, gerado a partir de um limiar de distância normalizada, a função calcula a acurácia do algoritmo ao tentar separar os vértices em Setosas e Não setosas, utilizando o conceito de uma matriz de confusão 2x2.

Observações sobre os resultados obtidos:

Ao montar a lista de adjacências, é possível observar que, em geral, plantas de uma determinada variedade se ligam, majoritariamente, a outras plantas da mesma variedade. Esta divisão, porém, é apenas observada se compararmos os vértices correspondentes a Setosas (1 até 50) com as outras duas variedades, isto é, não houve uma boa divisão entre Versicolors e Virgínicas. Isso pode ser visto tomando um vértice de número inferior a 50 e observando que este se liga, em sua maioria, a outros vértices do mesmo intervalo. Se tomarmos um vértice do intervalo 51 até 100, correspondente a Versicolors, este se mostra adjacente a um número elevado de vértices do intervalo superior (101 até 150) - correspondente a Virgínicas - demonstrando que a separação entre ambas variedades não foi bem contemplada no grafo .

Exemplo:

v[21]-1-2-3-4-5-6-7-8-9-10-11-12-13-14-15-16-17-18-19-20-22-23-24-25-26-27-28-29-30-31-32-33-34-35-36-37-38-39-40-41-42-43-44-45-46-47-48-49-50-58-99

Variedade se mostra isolada em relação às outras duas variedades.

v[57]-51-52-53-54-55-56-59-60-62-63-64-65-66-67-68-69-70-71-72-73-74-75-76-77-78-79-80-81-82-83-84-85-86-87-88-89-90-91-92-93-95-96-97-98-100-101-102-103-104-105-107-108-109-110-111-112-113-114-115-116-117-120-121-122-124-125-126-127-128-129-130-131-133-134-135-136-137-138-139-140-141-142-143-144-145-146-147-148-149-150

Variedade não se mostra adequadamente isolada, muitos vértices do intervalo superior (Virgínicas, 101 até 150) se mostrando adjacentes ao vértice analisado (Versicolor, 51 até 100)