

Trabalho 3.1: Algoritmo para classificação de patologias da coluna vertebral.

Até meados de 2005, a escassez de trabalhos voltados ao diagnóstico automático de patologias da coluna vertebral se devia em grande parte à ausência de atributos numéricos que descrevessem quantitativamente as patologias de interesse para o campo da Ortopedia, de modo a gerar um conjunto de dados adequado para o projeto de classificadores de padrões. No entanto, há poucos anos (2011, mais precisamente), um grupo de especialistas em Ortopedia e Biomecânica definiu um conjunto de informações (atributos) biomecânicas relacionadas a dores e deformidades da coluna vertebral.

O problema-alvo deste trabalho concentra-se na área de Ortopedia, e mais especificamente no diagnóstico de patologias da coluna vertebral, a saber, hérnia de disco e espondilolistese. O arquivo deste trabalho contém dados extraídos de 310 pacientes, a partir de radiografias panorâmicas sagitais em formato de 30 × 90 cm. Destes, 100 indivíduos são voluntários que não possuem patologias na coluna, doravante chamados de normais. Os dados restantes são obtidos a partir de radiografias de pacientes operados de hérnias de disco (60 indivíduos) ou espondilolistese (150 indivíduos). Cada um dos 310 pacientes é descrito por seis atributos biomecânicos:

- ângulo de incidência pélvica;
- ângulo de versão pélvica;
- declive sacral;
- ângulo de lordose;
- raio pélvico; e
- grau de deslizamento.

Exposto isso, este tema de trabalho compreende o desenvolvimento de um sistema de classificação de indivíduos em: normal (sem patologia), hérnias de disco ou espondilolistese. O sistema deve receber 6 atributos biomecânicos de um paciente e informar em qual condição ele se encontra. Para inferir esta condição, o sistema utilizará o algoritmo dos k vizinhos mais próximos. O algoritmo procura em um conjunto de registros (outros pacientes) os k pacientes mais similares que este paciente em questão. Ao extrair estes k pacientes, o algoritmo computa a moda (moda é o valor que mais se repete em um conjunto/lista/vetor) e infere o diagnóstico com base nessa informação de moda.

Primeiramente, o algoritmo deve perguntar os atributos biomecânicos de um novo paciente e localizar os k outros pacientes mais similares. Para recuperar um paciente similar basta computar a distância euclidiana entre os atributos biomecânicos do paciente em questão e os pacientes já registrados. Quando menor a distância, mais similares os pacientes são. Ao encontrar os k outros pacientes mais similares, o algoritmo computa a moda da doença, ou seja, a doença que mais se repetir é o diagnóstico.

$$\text{Distância euclidiana} = d(A, B) = \sqrt{\sum_{i=0}^5 (A[i] - B[i])^2}.$$

Trabalho 3.2: Algoritmo para classificação de flores íris.

O conjunto de dados de flor Iris ou o conjunto de dados Iris de Fisher é um conjunto de dados introduzido pelo estatístico e biólogo britânico Ronald Fisher em seu artigo de 1936. O conjunto de dados contém 3 classes de 50 amostras cada, em que cada classe se refere a um tipo de planta da íris, a saber, setosa, versicolor e virgínica. Cada uma das 150 amostras é descrito por quatro atributos:

- comprimento da sépala (em cm);
- largura da sépala (em cm);
- comprimento da pétala (em cm); e
- largura da pétala (em cm).

Exposto isso, este tema de trabalho compreende o desenvolvimento de um sistema de classificação de florzinhas em: setosas, versicolores ou virgínicas. O sistema deve receber 4 atributos de uma íris qualquer e informar em qual das classes acima essa amostra se encontra.

OBSERVAÇÃO: Em ambos trabalhos



Iris Versicolor



Iris Virginica



Iris Setosa