

Reconhecimento Hierárquico de Caracteres

Breno Flesch Franco e Ricardo Alexandre Bastos

MAC 499 - Trabalho de Formatura Supervisionado

17 de novembro de 2009

- 1 Reconhecimento
 - Definição
- 2 Classificação Hierárquica
 - BHC
- 3 Implementação
- 4 Classificação
- 5 Testes
- 6 Agradecimentos
- 7 Perguntas

Reconhecimento de Caracteres

Reconhecimento de caracteres manuscritos

É o processo de conversão de um caractere manuscrito em um símbolo.



torna-se

\alpha

Reconhecimento de Caracteres Manuscritos

Símbolos e características

Representação de um símbolo

Um conjunto de pontos do R^3 , sem tamanho fixo

Características

Definem uma dimensão fixa para toda representação de caractere

- diminuem a dimensão do problema
- perda de informação

Por isso é importante escolher boas características

Reconhecimento de Caracteres Manuscritos

Dificuldades inerentes

Grande Quantidade de Classes

Muitas classes \Rightarrow alto custo de treinamento, grande chance de erro

Falta de Contexto

A liberdade para escrita de fórmulas matemáticas não permite muitas inferências

Fator Humano

É muito raro que dois caracteres sejam idênticos, mesmo quando escritos pela mesma pessoa.

Classificação Hierárquica

- É uma abordagem natural
- Utiliza uma hierarquia de subconjuntos das características
- Classificação se torna mais simples

Classificador Binário Hierárquico

Utiliza uma árvore binária, onde:

Nós internos

- conjunto de Classes
- um extrator de características
- um classificador

Folhas

Cada uma representa uma classe

Classificador Binário Hierárquico

Diferenças entre abordagens

Bottom-Up

- *Clustering*: união de classes semelhantes
- Comportamento quadrático em relação ao número de classes

Top-Down

- Divisão em Módulos: Divisão em 2 conjuntos semelhantes
- Comportamento linear em relação ao número de classes

Classificador Binário Hierárquico Top-Down

Divisão do Conjunto de Classes

Associação das classes aos subconjuntos

Todas as classes são associadas aos subconjuntos, com exceção de uma, que é associada a apenas um deles

Encontrar uma ponderação das características para a divisão

Usa-se um discriminante para encontrar a reta que melhor divide essas características

- Atualização dos associações

Classificador Binário Hierárquico Top-Down (cont.)

Divisão do Conjunto de Classes (cont.)

Cálculo da entropia

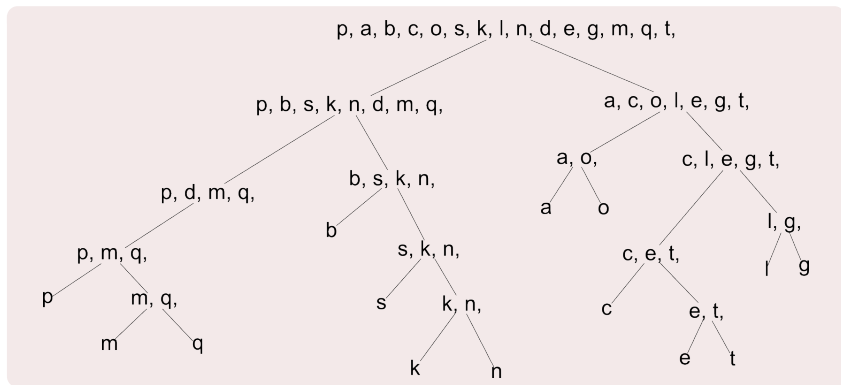
- entropia de uma classe é uma medida de quão dividida essa classe está entre os subconjuntos
- queremos reduzi-la a um valor pequeno
- Separação das classes em dois subconjuntos
- Recursão

Classificação de uma amostra

Em cada nó:

- É calculado o ponto gerado pelas características de acordo com a ponderação definida para o nó
- A amostra é destinada a um dos filhos do nó, de acordo com o resultado
- Ao chegar a uma folha, a amostra é classificada

Testes



Testes

Alguns resultados de testes preliminares:

- 95% de acerto com 15 classes, 20 amostras/classe, 4 testes/classe
- 35% de acerto com 35 classes, 18 amostras/classe, 2 testes/classe, escritores diferentes

Conclusão

O desempenho do classificador ainda deixa a desejar no caso geral, mas demonstra ter um desempenho promissor ao ser treinado com uma boa quantidade de dados do mesmo usuário.

Agradecimentos

Agradecemos a Cristiano Perez Garcia, pela cessão de seu conjunto de dados, e a Nina Sumiko Tomita Hirata, Ricardo Sider e Bruno Yoiti Ozahata pela cooperação.

Perguntas

Perguntas?