

# Tarefa 2: Revisão da Tarefa 1

Pedro Paulo Vezzà Campos - 7538743

25 de agosto de 2013

## 1 Definição do problema

Alunos aspirantes a designer aprendem nos primeiros semestres da graduação o conceito de *design com tipos*. São apresentadas as principais famílias tipográficas, tais como serifadas, sem serifa, monoespaçadas etc. Ainda, são estudadas algumas fontes famosas, tais como Garamond, Helvetica, Bodoni, etc.

Uma dificuldade recorrente de alunos é de identificar uma fonte em um dado texto. Tal problema se resume fundamentalmente a um problema de classificação como visto em sala de aula. O usuário forneceria uma série de características da fonte que observa e o sistema o apresentaria as fontes que melhor se encaixam nessas características.

Note que nem sempre o usuário pode consultar o conjunto completo de caracteres da fonte de interesse para fornecer suas características. Um exemplo clássico desse problema é quando o usuário vê a fonte buscada em um logotipo.

## 2 Caracterização dos dados (como representar?)

Cada fonte do banco de dados é caracterizada como sendo uma tupla de *features*. Algumas *features* e seus valores possíveis são apresentados abaixo:

**A fonte possui serifa?** {Sim, Não}

**Como é o formato da cauda da letra Q?** {Cruza o círculo, Toca o círculo, Fica abaixo do círculo e está separada, Está na parte interna do círculo, Faz parte de uma forma aberta}

**Qual a posição da base da letra J em relação à linha de base?** {Na linha de base, Abaixo da linha de base}

**Qual o estilo da linha do caractere \$?** {Uma linha cruzando o S, duas linhas cruzando o S, Uma linha que não cruza o S, Duas linhas que não cruzam o S}

**Como é o cruzamento da linha horizontal com a vertical do P?** {Há um espaço, uma encosta na outra, uma cruza a outra com excesso}

### 3 O que/como seria uma solução (computacional) boa do seu problema?

O sistema não é de missão crítica e não possui a obrigação de conhecer todas as fontes criadas até o momento. Apenas a resposta mais provável das fontes disponíveis no banco de dados já pode ser uma boa indicação ao usuário de qual é a fonte desconhecida.

### 4 Forma de treinamento

Para este problema de classificação, cabe uma abordagem **supervisionada**, de forma a garantir a homogeneidade e qualidade das instâncias de treinamento para o classificador.

### 5 Quais informações de contexto podem ser úteis para o reconhecimento?

Basicamente cada letra, número e símbolo pode ter características determinadas para possivelmente facilitar a busca pela fonte correta.

Ainda, algumas *features* não são relevantes dependendo de certas “macro-características”. Por exemplo, o formato da serifa de uma fonte não é relevante para uma fonte monoespaçada ou sem serifa.

### 6 Você vê desafios? Quais?

1. Determinar quais features utilizar. Como dito anteriormente, cada caractere pode se transformar em uma feature distinta para determinar uma fonte.
2. Produzir um conjunto de treinamento satisfatório para fornecer como entrada ao classificador.

### 7 Ousaria desenhar os passos para chegar à solução desejada? Quais seriam esses passos?

A solução para o problema consiste primordialmente na escolha correta de quais características são relevantes para uma boa classificação, e assim, evitando o problema da “maldição da dimensionalidade” e em seguida a escolha de um classificador que melhor se encaixe no problema proposto.