

# Reconhecimento de Escrita Cursiva

Carlos Rodrigo Cordeiro Garcia

*Deinfo, Departamento de Estatística e Informática  
UFRPE, Universidade Federal Rural de Pernambuco  
carlosrodrigocrcg@gmail.com*

Giovanni Paolo dos Santos Carvalho

*Deinfo, Departamento de Estatística e Informática  
UFRPE, Universidade Federal Rural de Pernambuco  
giovanni1995@gmail.com*

**Abstract**—In this paper are discussed a few approaches to handwritten characters recognition using neural networks based on backpropagation and recurrent networks. The studies are based upon the MNIST digits database - a dataset containing handwritten numbers from 0 to 9 with 60,000 training samples and 10,000 test samples. An overviewing of the features of the database is also included, overviewing properties such as attributes distribution, means, standard deviations and others alike. An off-line handwritten recognition system proposal is discussed along with methods to improve its classification. Image processing and feature extraction are one of the proposed methods for the classification of the characters. The proposed recognition system may have various applications, such as ZIP code recognition in letters or number recognition from bank accounts and paychecks.

**Keywords**—*Handwriting character recognition; Image processing; off-line; Neural networks; MNIST;*

## I. INTRODUÇÃO

Reconhecimento de escrita é a capacidade de um computador receber e interpretar uma entrada manuscrita inteligível a partir de fontes tais como documentos em papel, imagens, telas sensíveis ao toque e outros dispositivos. Este problema tem sido uma das áreas de pesquisa mais fascinantes e desafiadores em campo de processamento de imagens e reconhecimento de padrões nos últimos anos.[1]

Em geral, podemos classificar o problema de reconhecimento de escrita em dois tipos: off-line e on-line. No método off-line, normalmente, a escrita é capturada por algum tipo de scanner e ela é disponibilizada para o classificador como uma imagem, já no on-line os movimentos de uma caneta podem ser sentidos por algum tipo de sensor para detectar-se o que está sendo escrito, como por exemplo, usando telas de computador sensíveis à toques e canetas.[2] A abordagem on-line tem se mostrado superior ao método off-line no reconhecimento de escrita, porém os sistemas off-line também mostraram resultados bastante satisfatórios principalmente com a utilização de redes neurais e redes recorrentes. Várias aplicações como processamento de cheques bancários, reconhecimento de CEP e leitura de documentos escritos à mão precisão de um sistema que consiga reconhecer escrita em um documento normal. Dessa forma reconhecimento de escrita off-line continua sendo uma área de pesquisa bastante relevante para pesquisadores que possam explorar novas técnicas e melhorar tanto na acurácia, generalização e complexidade desses sistemas.

Em [2] são consideradas algumas abordagens para aprendizado e classificação online e offline para reconhecimento de escrita cursiva. Em [3], uma aplicação de um MLP visando performance ótima é articulada. Vale também ressaltar o trabalho em [8] que apresenta os melhores métodos até agora de reconhecimento de caracteres off-line. Por fim o trabalho apresentado em [9] que é um artigo em forma de overview da história e dos métodos para reconhecimento de escrita em geral.

## II. BASE DE DADOS

A base de dados MNIST é uma das bases mais famosas para o treinamento e teste de classificadores para o problema de reconhecimento de escrita, originalmente criada por Yann LeCun, Courant Institute, NYU com contribuição de Corinna Cortes, Google Labs, New York e Christopher J.C. Burges, Microsoft Research, Redmond. A base de dados MNIST contém um conjunto de treinamento com 60,000 exemplos, e um conjunto de teste com 10,000 exemplos de dígitos entre 0 e 9. Este conjunto é um subconjunto de uma base de dados maior que é disponibilizado pelo NIST(National Institute of Standards and Technology) e pode ser encontrado em[10]. O conjunto de treino contém exemplos de cerca de 250 escritores, o conjunto de escritores da base de teste e de treinamento conjunto são disjuntos. Cada dígito é normalizado e centralizado numa imagem em níveis de cinza com tamanho  $28 \times 28$ , resultando numa imagem de 784 pixels, ou seja, existem 784 atributos ao total nessa base de dados que variam do pixel 1 ao 785 como características. Alguns exemplos são mostrados na Figura 1.

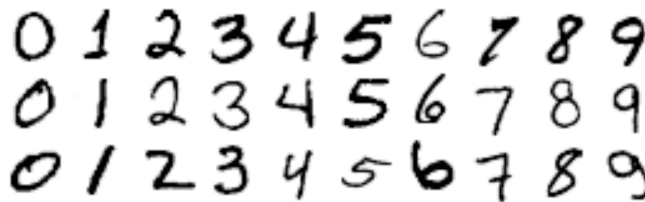


Figura 1: base de dados MNIST.

Para conveniência de manipulação dos dados também utilizamos a base localizada em[11]. Que consiste na mesma base MNIST em formato .arff, dessa forma usamos esta base

para uma leitura mais simples e prática dos dados com a finalidade de, por exemplo, fazer seleção de atributos.

III. IMPLEMENTAÇÃO

// TODO  
Será implementado na VA2.

IV. ANÁLISE DOS DADOS

Parte da análise dos dados é feita utilizando-se representações gráficas que correspondem à distribuição dos dados a serem usados pelo algoritmo. Base de dados que contenham padrões com poucas características são mais fáceis de serem visualizados e dessa forma verificados. Assim, se tivéssemos uma base de dados com 2 atributos poderíamos lê-los num gráfico 2D onde o eixo X representa o dado 1 e o eixo Y representa o dado 2, podemos assim aumentar a dimensionalidade do problema e continuar a representá-lo num gráfico, por exemplo, para 3 atributos teríamos um espaço com 3 dimensões, onde cada vértice representaria 1 dos dados do problema. Porém o quanto mais a dimensionalidade cresce mais difícil fica para poder aplicar essa técnica de visualização. Uma solução para esse problema é tentar representar dados de uma dimensão alta em um espaço menor. Normalmente tentamos utilizar uma visualização num espaço de 2 dimensões. Para alcançar esse resultado desejado utilizaremos redução não-linear de dimensionalidade (nonlinear dimensionality reduction), ou manifold learning[4]. Vamos utilizar uma técnica chamada t-SNE que visualiza essas alta dimensões mapeando cada datapoint para um gráfico com duas ou três dimensões[5].

O resultado da aplicação do t-SNE na base de dados MNIST é mostrado na Figura 2.

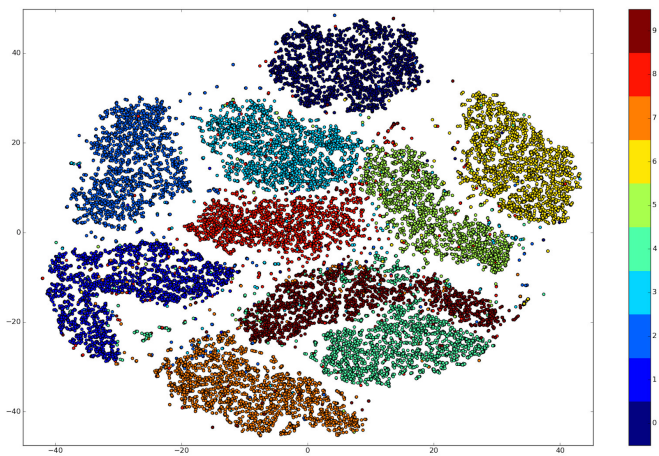


Figura 2: base MNIST representada em 2D

Dados como média e desvio padrão também são fornecidos na tabela 1, vale lembrar que como a imagem tem 784 pixels em tons de cinza, cada pixel representa uma característica que tem valores mínimos e máximos de 0 e 255, respectivamente.

Pelo fato de termos 784 atributos apenas os 10 mais relevantes, selecionados usando um dos métodos de seleção de atributos demonstrados no próximo item, de toda a base de treinamento estarão na tabela 1.

Média e Desvio Padrão para os 5 atributos mais relevantes		
Atributo	Média	Desvio Padrão
Pixel 379	113.7	114.3
Pixel 407	132.6	113.9
Pixel 351	90.2	109.3
Pixel 435	138.1	112.5
Pixel 410	129.8	113.3

Tabela 1: Média e Desvio Padrão

V. SELEÇÃO DE ATRIBUTOS

O método de seleção de atributos é um dos principais fatores para obter-se uma boa taxa de classificação e um classificador robusto para reconhecimento de caracteres, sem mencionar a redução de poder computacional que é alcançado quando pode-se extrair características de uma imagem e deixá-las com a mesma representação já demonstrada. Existem ainda casos em que a classificação melhora substancialmente depois da seleção, provando ser um passo extremamente importante para o reconhecimento de caracteres.

Foram utilizados 3 métricas de seleção de atributos: ganho de informação, razão de ganho e correlação. Os resultados obtidos foram bastante parecidos entre essas 3 abordagens, principalmente entre ganho de informação e correlação. Os atributos que tiveram as maiores taxas de relevância, ou seja, aqueles que tem mais informação sobre essa base de dados foram os pixel que estavam na região central da imagem. Analogamente os atributos com menor importância foram os primeiros, os laterais e últimos pixels, ou seja, atributos que estavam localizados nas extremidades da imagem.

O que faz sentido e comprovou aquilo que nós criamos como modelo para seleção. Os pixels centrais contêm mais informação, dado que esta é a região onde há mais presença de dados pelo fato da base ser centralizada e pela natureza da escrita de uma pessoa que geralmente preenche espaços centrais de uma determinada região. Já as regiões das extremidades, usualmente, não contêm muitas informações para um determinado caractere pois essa região tende a não ser preenchida quando estamos escrevendo algo num documento.

VI. DESCRIÇÃO DOS ALGORITMOS DE AM A SEREM UTILIZADOS

A princípio, os principais algoritmos para classificação são o Multilayer Perceptron com backpropagation e Redes recorrentes. Entre os principais motivos que levaram à escolha destes, estão o fato de redes neurais baseadas em backpropagation com pelo menos duas camadas e um número suficiente de nós ocultos serem comprovadamente aproximadores universais [6][7]. Elas não requerem conhecimento algum a respeito de qualquer relação entre

atributos e saída e nesta área de reconhecimento de caracteres já vêm sendo usadas com altíssimas taxas de sucesso. Além disso, se adequam bem ao treinamento offline que será aplicado, não havendo necessidade de adaptação constante do modelo, mas sim apenas a formulação de um modelo com boa generalização para o conjunto utilizado.

## VII. TRABALHOS FUTUROS

Visa-se futuramente expandir o reconhecimento para um grupo mais abrangente de caracteres e não apenas números. Também há uma possibilidade de reconhecer letras em frases e não apenas caracteres avulsos, utilizando métodos de segmentação para poder separar uma palavra em letras que possam ser reconhecidas pelo algoritmo a ser implementado.

## REFERÊNCIAS

1. S. Mori, C.Y. Suen & K. Kamamoto, "Historical review of OCR research and development," Proc. of IEEE, vol. 80, pp. 1029-1058, July 1992.
2. Réjean Plamondon & Sargur N. Srihari "On-line and off-line handwriting recognition: a comprehensive survey." IEEE Trans Pattern Anal Mach Intell.
3. Anita Pal & Dayashankar Singh, "Handwritten English Character Recognition Using Neural," Network International Journal of Computer Science & Communication.
4. Michael Gashler, Dan Ventura, e Tony Martinez, "Manifold Learning by Graduated Optimization" IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS—PART B: CYBERNETICS, VOL. 41, NO. 6, DECEMBER 2011.
5. Laurens van der Maaten, Geoffrey Hinton "Visualizing Data using t-SNE" Journal of Machine Learning Research 9 (2008) 2579-2605.
6. G. Cybenko, "Approximation by superpositions of a sigmoidal function," Math. Contr., Signals, Syst., vol. 2, pp. 303-314, 1989.
7. K. Hornik, M. Stinchcombe and H. White (1989). Multilayer feedforward networks are universal approximators. Neural Networks, 2, 359-366.
8. B. Verma, M. Blumenstein & S. Kulkarni, "Recent achievements in off-line handwriting recognition system"
9. Homayoon S.M. Beigi, "An Overview of Handwriting Recognition"
10. MNIST dataset, <http://yann.lecun.com/exdb/mnist/>
11. MNIST formato ARFF, <http://axon.cs.byu.edu/data/mnist/>