

Notas sobre o desenvolvimento de uma rede neural para classificar caracteres escritos à mão.

Motivado pelo desafio:

“Your company is developing a software to digitize handwritten text. Your team has already developed code to extract the image of each one of the characters in a given text. You are given the task of developing a machine learning model capable of reliably translating those images into digital characters. After some research, you find the [EMNIST dataset](#), which seems perfect for the task.”

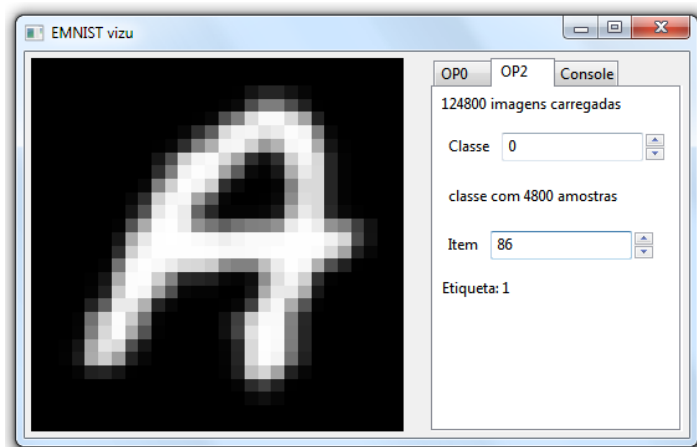
proposto pela DeepX.

O banco de dados EMNIST fornece milhares de itens com associação entre a representação visual de um caractere escrito à mão e uma etiqueta identificando-o.

Tomando a liberdade de supor que a empresa do enunciado pré-processa os caracteres para formato semelhante, o desafio foi alinhado ao desenvolvimento de uma (mini) biblioteca para treino de rede neural fortemente conectada. Com esta biblioteca, foi treinada uma rede que, aplicada ao banco de teste balanceado do EMNIST, obteve taxa de acerto de 87% para letras e 96% para dígitos numéricos.

Interpretação dos Bancos

Como os arquivos fornecidos no EMNIST não são *legíveis para humanos*, o primeiro passo foi desenvolver um programa gráfico auxiliar (com wxWidgets) para visualizá-los:



A representação visual é um bitmap de 28x28 pixels em escala de cinza com um byte por pixel, e a etiqueta é um byte que associa o bitmap à classe correta (*que letra representa*).

Completa a leitura dos arquivos, decidiu-se que a rede deveria mapear a intensidade dos pixels de entrada para uma resposta indicando qual caractere mais provável. Começou-se o desenvolvimento considerando-se as amostras de caracteres que representam números (por ser desafio mais simples, apenas dez classes), e depois a rede foi adaptada para lidar com o banco de caracteres balanceados (que combina em mesma classe letras que são semelhantes quando maiúsculas e minúsculas), considerando as letras.

Objetivo

Foi tomada a liberdade de usar do desafio para estudar o tema e implementar o algoritmo da rede neural explicitamente. Para isso foi usado o C++. É provável que em desenvolvimento de produto real fossem usadas bibliotecas pré-existentes para treinamento de rede neural convolucional, que atingem taxa de acerto superior a 99% para este tipo de tarefa.

As justificativas matemáticas para os algoritmos utilizados estão no relatório técnico.

Desenvolvimento

Pesquisa inicial indicou que, para esta tarefa específica, consegue-se boa taxa de acerto com um tipo de rede neural treinada por *backpropagation*. No tipo de rede considerada, a saída de cada camada (chamada de vetor ***a***) é uma função de ativação aplicada à ponderação da entrada. As funções de ativação implementadas são sigmoide e *softmax*:

$$\mathbf{a}(\mathbf{x}) = f_{\text{ativação}}(\mathbf{W} \times \mathbf{x} + \mathbf{b})$$

A entrada da primeira camada é um vetor contendo a intensidade de cada pixel, e a saída da última camada é um vetor cujo maior componente indica a classe a que a rede associou a entrada.

Depois de treinadas as redes neurais, os coeficientes foram salvos em um arquivo, que é utilizado para prever caracteres em um programa exemplo.

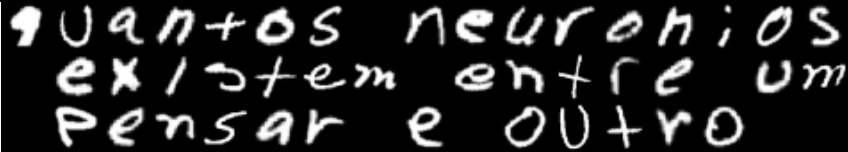
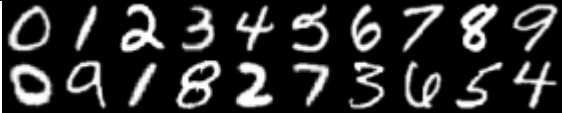
Os protótipos expostos das funções desenvolvidas para uso da rede são:

```
char identificar_letra ( sistema_neural S, bm_ref_t img )  
char identificar_digito (sistema_neural S, bm_ref_t img )
```

Cada função devolve o caractere mais provável para a imagem fornecida (o segundo argumento, de tipo *bm_ref_t*, é um array de 28 por 28 bytes). A primeira função espera uma rede treinada pra identificar letras, a segunda espera uma rede treinada para identificar números.

Resultado

Do programa resultante de *exemplo.cpp*, que varre grades de um arquivo *.bmp*:

Entrada	Saída
	qu q ntos neuron j os existem q ntre um pensar e outro
	0123456789 2 918273654

Os caracteres foram sorteados do banco do EMNIST para formar as imagens. Conforme mencionado anteriormente, redes convolucionais são melhores para esta tarefa, e existem bibliotecas pré-disponíveis e abertas para isto. A criação de uma (mini) biblioteca para redes fortemente conectadas foi apenas para fins didáticos de interesse próprio.