

Practicum in Artificial Intelligence: Optical Character Recognition via Neural Networks

Bryan Cuccioli (blc72@cornell.edu), Renato Amez (ra374@cornell.edu)

October 15, 2012

We propose the design, implementation, training, and evaluation of a feed-forward neural network adapted to the task of optical character recognition (OCR). We define a precise metric for measuring the success of such a network on real-world data. We experimentally measure the efficacy of the neural network under variation of certain parameters of its construction and seek to maximize a precise quantity expressing the success of the neural network versus the difficulty in constructing and training it.

1 Description

We propose to implement a system for optical character recognition using a feed-forward neural network with backpropagation learning. We will prepare a set of training data consisting of individual letters and words along with the desired output for each input item. Characters included in the training data will be both handwritten and rendered by computer in both serif and sans-serif font faces. Images of words will first be segmented into individual characters; images of handwritten characters will be represented in the system as two-dimensional arrays of pixels normalized to a common size. Inference rules will be applied in the case of whole words to make educated guesses about difficult-to-recognize characters.

For data set Ψ , and character $x \in \Psi$, we define

$$\delta(x) = \begin{cases} 1 & \text{character correctly recognized} \\ 0 & \text{otherwise} \end{cases}. \quad (1)$$

We then define the **success of the neural network** \mathcal{S} on data set Ψ as

$$\mathcal{S} = \sqrt{\sum_{a \in \Psi} \delta(a)}. \quad (2)$$

This definition is chosen so as to more heavily weight improvements to the neural network once the network is already performing well, under the guise that e.g. 70% recognition is not particularly better than 40% recognition.

Training of the network will be complete once \mathcal{S} falls above a pre-determined threshold α . For the purposes of experimentation, α will be varied and the success of the neural network in interpreting the real data will be measured as a function of α .

2 Evaluation

Evaluation of the neural network will be performed using data distinct from the training data. Samples rendered by computer and printed out as well as samples handwritten by multiple third parties will be used. We will evaluate the system varying parameters such as

1. the training error threshold, α ;
2. the number of layers in the neural network, ℓ ;
3. the size σ and variety of the training data.

For experimentation purposes, the success of the neural network in performing OCR will be measured as a function of each of these parameters separately. We will experimentally determine the values of these parameters that maximize the *training efficiency*

$$\mathcal{E} = \frac{g^2}{\alpha \cdot \ell \cdot \sigma}. \quad (3)$$

3 Timeline

We propose the following timeline for the implementation and evaluation of our system:

- *October 15*: Submission of project proposal;
- *October 26*: Design of a low-level outline of the architecture of our neural network, data storage, and test harness;
- *November 15*: Completion of implementing a basic neural network, system for storing training data and real data, and test harness for evaluating the efficiency of our neural network;
- *November 26*: Completion of gathering data based on the performance of the neural network on separate data under variation of the above parameters.
- *November 29*: Completion of final presentation and final project report.