ON CLASSIFICATION OF 17TH CENTURY FONTS USING NEURAL NETWORKS

Tudor Bumbu

Vladimir Andrunachievici Institute of Mathematics and Computer Science,
State University of Tiraspol, Technical University of Moldova,
Chişinău, Republic of Moldova

`tudor.bumbu@math.md, bumbutudor10@gmail.com`

This paper represents a solution to the problem of identifying the fonts in the books printed in the 17th century with characters from the Romanian Cyrillic alphabet. At the same time, this paper is an extension of the paper presented in [1].

Considering the fact that in Wallachia (Romanian: Ţara Românească) there was a diversification of printing styles (some printing houses borrowed the Slavonic printing, others created their own printing based on the Cyrillic alphabet), some documents from the 17th century require particular approaches to processing, namely optical character recognition (OCR).

The problem of identifying the font in a document printed in the 17th century can be formulated as follows: *Given a document X from 17th century printed in Cyrillic Romanian and a set of N OCR models trained on documents involved in this period. Choose the most appropriate model M from set N for document X.*

A trivial solution would be to recognize a sample (a page snippet) from document *X* using all models in set *N* and based on the results, choose the model that offers the highest accuracy (best result). This solution is easy to implement, but the time complexity is too big, as we have to load each model separately. Model upload time and sample recognition can exceed 2 minutes depending on page size. If we have 5 different models, we'll have to wait for approx. 10 minutes to find the right model.

The second and the proposed solution would be to train a neural network to classify snippets from document *X* based on samples from several Romanian documents printed in the years 1630-1680 at different printing houses in Wallachia. A neural network will learn a data set consisting of tuples *(image character, class (Font1, Font2, Font3, ..., FontN))* in order to be able to further classify new samples.

Therefore, we train a neural network to classify between 2 distinct fonts used by two printing houses: *1. Cetatea Belgradului din Ardeal; 2. Casa Sf. Mitropolii din Iaşi;* from the 17 century.