

Analysis of neural networks efficiency for determining positions of corrupted bytes

Vladislav Lupanov

velupanov@gmail.com

Sergey Slastnikov

sslastnikov@gmail.com

National Research University Higher School of Economics

Abstract

A lot of files and data, in general, are transferred throughout the networks. But the data may be corrupted by intrusions or package loss so, the executable files may be marked as non-executable and violate the local network policy. Thus, it's necessary to detect such files. In this paper, we present a novel method for detecting broken bytes of a file, so the corrupted files may be detected. Also, the positions of wrong bytes might be helpful in restoring the original file content. This work is devoted to study of modern neural network models applied to detect corrupted bytes of a file problem. Since recurrent neural networks (RNNs) seem to be well suited for such tasks, the main tasks of this work are to analyze the efficiency of popular state-of-the-art RNNs solving the problem mentioned above and to compare results of different models. We use data consisting of the most popular file types collected from the Internet and manually randomly added noise to that data to test our models. An experiment on this data demonstrates the advantages and disadvantages of the considered models.