

Avaliação da LSTM Multidimensional para o Reconhecimento de Textos Manuscritos em Imagens

Dayvid Welles de Castro Oliveira, Universidade de Pernambuco (dwco@ecomp.poli.br)

Byron Leite Dantas Bezerra, Universidade de Pernambuco (byronleite@ecomp.poli.br)

Mêuser Jorge Silva Valença, Universidade de Pernambuco (meuser@ecomp.poli.br)

O reconhecimento de textos manuscritos trata-se de um campo de pesquisa ativo e amplamente investigado. Ainda assim, o desenvolvimento de sistemas capazes de operar em um cenário offline e irrestrito, no qual há disponível somente a imagem de uma longa sequência de texto manuscrito, permanece como um problema em aberto (BLUCHE, 2015). Tradicionalmente, foram usadas técnicas de aprendizagem de máquina baseadas no Modelo Escondido de Markov. Todavia, essa abordagem apresenta uma habilidade limitada em modelar dependências de longo prazo. Para evitar este problema, as Redes Neurais Recorrentes, em particular as redes LSTM, vêm sendo aplicadas de forma sucedida nessa atividade, devido à sua capacidade em acessar informações contextuais de longo alcance (BLUCHE, 2015). Assim, pode-se destacar o êxito do modelo LSTM Multidimensional (MDLSTM) associado à função objetivo CTC, abordagem vencedora de várias competições de reconhecimento de manuscritos e que atualmente constitui o estado da arte na área (VOIGTLAENDER; DOETSCH; NEY, 2016). Partindo desse modelo, Voigtlaender, Doetsch e Ney (2016) propuseram uma nova arquitetura MDLSTM cujos resultados atingidos são os melhores até então conhecidos em duas das bases de dados mais utilizadas no campo: IAM e RIMES. Ainda assim, como indicam os autores, esses experimentos podem ser considerados como ponto de partida para que novos estudos contribuam para a evolução do novo modelo. Portanto, o objetivo dessa pesquisa é expandir as investigações da MDLSTM proposta ao conduzir novas análises em relação a sua topologia em busca de um ponto de equilíbrio ótimo entre a largura das camadas e a profundidade do modelo. Como no trabalho original somente foram avaliadas camadas MDLSTM e convolucionais de forma alternada, nesse projeto é incorporado o uso dessas camadas de forma consecutiva – inspirado pela filosofia das LSTMs convolucionais. Além disso, propõe-se a adição de uma camada extra de sub-amostragem para intensificar a redução do volume de ativações. Até o momento, o desempenho do modelo foi verificado na base de dados IAM e ao comparar o resultado obtido ao usar a configuração original do trabalho de Voigtlaender, Doetsch e Ney (2016) com uma configuração incorporando as modificações propostas, foram registradas, respectivamente: taxas de erros de caracteres (CER) de 6.64% e 6.39%, taxas de erros de palavras (WER) de 23.99% e 22.82% e tempo de treinamento médio de 96 e 49 minutos por época. Deste modo, observou-se uma diminuição nas taxas de erros e no tempo de execução, o que pode indicar a configuração proposta como uma abordagem promissora no reconhecimento de manuscritos.

Palavras-chave: *Reconhecimento de Manuscritos; Redes Neurais Recorrentes; MDLSTM*

Referências

BLUCHE, Théodore. **Deep Neural Networks for Large Vocabulary Handwritten Text Recognition**. 2015. Tese de Doutorado. Université Paris Sud-Paris XI.



MOSTRA POLI 2017



VOIGTLAENDER, Paul; DOETSCH, Patrick; NEY, Hermann. Handwriting recognition with large multidimensional long short-term memory recurrent neural networks. In: **Frontiers in Handwriting Recognition (ICFHR), 2016 15th International Conference on**. IEEE, 2016. p. 228-233.

.