

Reconhecimento de Escrita Cursiva *Offline* Utilizando um Modelo Composto por MDRNN-RC

Maxwell Francisco, Felipe Gouveia, Byron Bezerra, Mêuser Valença

Escola Politécnica de Pernambuco

Universidade de Pernambuco

mqf@ecomp.poli.br, fmg@ecomp.poli.br, byronleite@ecomp.poli.br, meuser@ecomp.poli.br

Resumo — Este trabalho tem como objetivo propor um sistema composto por MDRNN-RC (*Multi-dimensional Recurrent Neural Network – Reservoir Computing*) com a finalidade de reconhecimento de escrita cursiva *off-line* e explorar posteriormente os resultados, pois essas técnicas quando combinadas podem resultar em melhorias, visto que são robustas e notadamente referenciadas na literatura para esse tipo de tarefa.

Palavras-chave — Reconhecimento de Caracteres Manuscritos; Reconhecimento de Padrões; Redes Neurais Artificiais; Redes Recorrentes; Reservoir Computing;

I. INTRODUÇÃO

A evolução dos métodos de representação e armazenamento propiciou que muitas empresas pudessem fazer um maior uso desses recursos. Contudo, muitos dados ainda são inseridos manualmente, tais como formulários, cheques, notas fiscais, boletos, dentre outros. Desta forma, há necessidade de utilizar um sistema que automatize esse processo, diminuindo assim custos e falhas provenientes da interferência humana [1]. A área de visão computacional trata da extração de informações das imagens, identificação e classificação dos objetos presentes nestas [2]. Inserido nesse contexto, o reconhecimento de caracteres e palavras, exigem o uso de algoritmos sofisticados em Processamento Digital de Imagem (PDI), aprendizado de máquina e classificação, que o presente trabalho propõe explorar, tais como MDRNN-LSTM combinado com RC, propostos respectivamente por Graves [3] e Verstraeten [4].

II. OBJETIVOS

O reconhecimento automático de escrita é uma tarefa desafiadora e tem sido objeto de pesquisa há mais de 50 anos [5]. Este processo pode ser dividido em dois problemas de classificação: *on-line* e *off-line*. No primeiro caso, o reconhecimento é realizado em tempo real, ou seja, à medida que o usuário escreve, seus movimentos são capturados e armazenados, para que estas informações auxiliem na classificação do caractere [1]. No segundo caso, o conteúdo a ser reconhecido está disponível em formato digital e é capturado por equipamentos tais como scanners e câmeras, não possuindo quaisquer informações temporais. De acordo com Bunke [6], a falta de informação temporal torna o processo de reconhecimento *off-line* ainda mais difícil.

O objetivo geral deste trabalho é avaliar os resultados da arquitetura MDRNN-LSTM combinado com o método RC, no reconhecimento de caracteres isolados, em seu formato de imagem digital crua, ou seja, diretamente a partir da matriz de pixels.

III. RECONHECIMENTO DE CARACTERES ESCRITOS À MÃO

Este tipo de reconhecimento consiste na classificação de caracteres através de imagens digitais, que estão separadas em um total de 36 classes, sendo 10 no caso dos dígitos arábicos e 26 para o alfabeto ocidental [7].

Camastra34D, um algoritmo que foi proposto em 2006 por Francesco Camastra [8], define 34 características para apenas um caractere. O algoritmo consiste na extração de características locais e globais da imagem, sendo as locais extraídas de sub-imagens da imagem original e as globais provenientes de informações sobre o formato geral do caractere, tal como a razão entre largura e altura. Os experimentos realizados por Camastra com este algoritmo, foram executados na base pública C-Cube.

Cruz *et al.* [9] apresentaram em 2010 um sistema para reconhecimento de caracteres manuscritos *off-line*, usando 6 técnicas distintas para extração de características. O mesmo utilizava uma técnica seletora de características independentes e complementares a fim de destacar ainda mais as classes mais divergentes. A base utilizada foi a mesma que Camastra [8] e o autor utilizou um classificador MLP.

IV. MDRNN-LSTM

A arquitetura proposta por Graves [3], produziu um dos melhores resultados na competição de reconhecimento de escrita na *International Conference on Documents Analysis and Recognition – ICDAR*, realizada em 2009. É estruturada em uma rede neural multidimensional recorrente, em camadas organizadas hierarquicamente, em conjunto com um método *Long Short Term Memory* (LSTM) e uma camada de saída que utiliza a classificação CTC - *Connectionist Temporal Classification* [10].

A multidimensionalidade de uma MDRNN pode ser considerada uma extensão das conexões de uma Rede Neural Artificial Recorrente (RNNs) [11]. Um problema conhecido como dissipação de gradiente, inerentes das RNN's, torna difícil o aprendizado para tarefas que contenham mais que dez

atrasos no tempo (*timesteps*) entre a apresentação de entradas e eventos esperados, levando ao “esquecimento” das primeiras entradas, fez com que Graves [12] propusesse a combinação com a rede LSTM.

De acordo com Hochreiter [10], O modelo LSTM consiste em um conjunto de sub-redes conectadas com processamentos recorrentes, nomeada de blocos de memória, que podem ser consideradas uma metáfora para chips de memória. As operações básicas dessa arquitetura são de escrita, leitura e de limpeza (*reset*) dos circuitos. Com isso, o chaveamento das portas permitem preservar a informação do bloco por um maior tempo, consequentemente minimizando o problema da dissipação de gradiente.

V. RESERVOIR COMPUTING

Reservoir Computing é um tipo de RNR na qual apresenta sua estrutura básica é formada por dois componentes: uma topologia não-linear de elementos de processamento, conhecida como *Reservoir*, e uma camada de saída linear, chamada de *readout*. Os estados do *Reservoir* são chamados de *echo states* e eles são os responsáveis pela memória dos padrões de entrada [13].

Esta abordagem foi difundida por Verstraeten [4] na última década e vem sendo adotada como uma nova proposta para modelagem e treinamento de RNRs, pois unifica duas técnicas bem consolidadas na literatura, propostas por Wolfgang Maass com o nome *Liquid State Machine* (LSM) e por Herbert Jaeger com o nome *Echo State Networks* (ESN).

Diferentemente das abordagens tradicionais de RNR, em *Reservoir Computing* não há necessidade de treinamento dos pesos da camada de entrada e nem dos pesos internos da rede (*reservoir*). Apenas os pesos da camada de saída (*readout*) são treinados. Contudo, existe a necessidade de ajuste dos parâmetros da topologia da rede, para que seja adequado a resolução de um determinado problema. [4]

VI. MODELO MDRNN-RC

O modelo MDRNN-RC proposto neste trabalho visa unificar dois métodos bem consolidados e reconhecidos no estado da arte, que tem como princípio básico realizar tarefas de classificação, mas com objetivos de uso distintos. Sendo assim, o sistema será dividido em duas etapas, onde apenas a primeira ocorre impreterivelmente.

Na primeira etapa, que corresponde ao classificador MDRNN com a aplicação do método LSTM, é submetida uma imagem digital, que é então processada e direcionada para uma camada de saída CTC, que foi desenvolvida por Graves, para executar o rotulamento de sequência com uma rede neural, com a finalidade de exibir a possibilidade percentual desta imagem fazer parte de qualquer uma das classes possíveis.

Os resultados da primeira etapa são analisados, caso a margem de erro seja superior a um dado valor percentual, então o melhor par de resultados, ou seja, as duas classes

com maior probabilidade de acerto, servirão como entrada para a etapa seguinte.

Na segunda etapa, a imagem original é então submetida a um algoritmo extrator de características. Neste trabalho será utilizado o modelo proposto por Camastra [8], que consiste em agregar indicadores locais e globais da imagem de entrada, afim de destacar as classes com alto grau de similaridade. De posse das duas classes com maior probabilidade e suas características da imagem extraídas, essas informações serão submetidas a um classificador binário (RC), que será responsável por determinar qual caractere representa a imagem de entrada.

A Figura 1 representa o fluxograma do modelo.

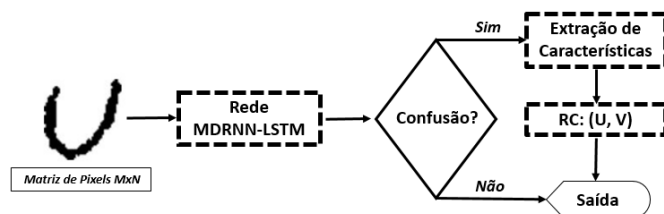


Figura 1 – Fluxograma do Modelo Proposto

REFERÊNCIAS

- [1] V. B. ANDRADE, “Aplicação de um sistema composto MDRNN-SVM para reconhecimento de escrita cursiva”, Dissertação de Mestrado, Universidade de Pernambuco, Recife, 2002.
- [2] CONCI et al., “Computação Gráfica: Teoria e Prática” 2.ed – Rio de Janeiro, Elsevier, 2008.
- [3] A. Graves e J. Schmidhuber, “Offline Handwriting Recognition with Multidimensional Recurrent Neural Networks,” 22nd Annual Conference on Neural Information Processing Systems, vol. NIPS '08, pp. 545–552, 2008.
- [4] D. Verstraeten, “Reservoir Computing: Computation With Dynamical Systems”, Tese de doutorado, Ghent University, 2009.
- [5] J.-C. Simon. Off-line cursive word recognition. Proc. of the IEEE, 80(7):1150–1161, July 1992.
- [6] H. BUNKE, “Recognition of cursive roman handwriting - past, present and future”, Document Analysis and Recognition, International Conference on IEEE Computer Society, Los Alamitos, CA, USA, v. 1, p. 448, 2003.
- [7] Byron et al., “A Hybrid RNN Model for Cursive Offline” Brazilian Symposium on Neural Networks, 2012.
- [8] F. Camastra, M. Spinetti, e A. Vinciarelli, “Offline Cursive Character Challenge: a New Benchmark for Machine Learning and Pattern Recognition Algorithms”, 18th International Conference on Pattern Recognition, vol. 2, pp. 913–916, 2006.
- [9] R. M. O. Cruz, G. D. C. Cavalcanti, e T. I. Ren, “An Ensemble Classifier For Offline Cursive Character Recognition Using Multiple Feature Extraction Techniques,” The 2010 International Joint Conference on Neural Networks (IJCNN), pp. 1–8, 2010.
- [10] S. Hochreiter e J. Schmidhuber, “Long Short-Term Memory”, Neural Computation, pp. 1735–1780, 1997.
- [11] A. Graves, S. Fernández e J. Schmidhuber, “Multi-Dimensional Recurrent Neural Networks”, Istituto Dalle Molle di Studi sull'Intelligenza Artificiale, Manno, Suiça. 2007.
- [12] A. Graves, S. Fernández e J. Schmidhuber, “Multidimensional Recurrent Neural Networks”, In: Proc. of Int. Con. on Artificial Neural Networks, pp. 549–558, 2007.
- [13] Ferreira, Aida Araújo, “Um Método para Design e Treinamento de Reservoir Computing Aplicado à Previsão de Séries Temporais”, Tese de Doutorado, Cin-UFPE, 2011.