## ICPR (Milan-2020-online)

# Learning to Sort Handwritten Text Lines in Reading Order through Estimated Binary Order Relations

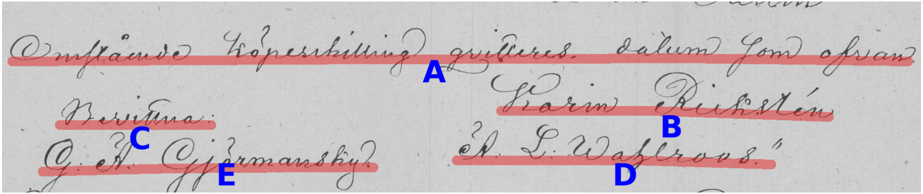
**Lorenzo Quirós, Enrique Vidal**

Pattern Recognition and Human Language Technology Research Center
Universitat Politècnica de València, Spain

{loquidia,evidal}@prhlt.upv.es

January 14th, 2020

# Motivation




- Advances in Handwritten Text Recognition and Document Layout Analysis make it possible to extract data from digitized documents.
- But, without the reading order that data cannot be transformed into information.

# Proposed Approach

- Traditional methods are mainly based on heuristics and deeply dependent on domain knowledge.
- The structure of handwritten text documents commonly does not follow a standard structure but the art of the writer.

- We propose to address the problem as a pair-wise probabilistic sorting problem, at baseline level.
- Order-relation operator is learned by a supervised machine learning algorithm.

# Proposed Approach [cont...]

Let $S$ be a set of $n$ baselines of a page image. Obtain a permutation $\mathbf{z}$ of $S$ that renders its elements (baselines) in reading order.

$$\mathbf{z} = (s_1, \nu_1), \ldots, (s_n, \nu_n), \nu_i \text{the index of } s_i$$

Proposed solution:

Addressed as a pair-wise probabilistic sorting problem;
Step 1: lets define the pair-wise relations

$$([s_i, s_j], y), \quad y = \begin{cases} 1 & \text{if } \nu_i < \nu_j \\ 0 & \text{otherwise} \end{cases} \quad 1 \leq i, j \leq n, \, i \neq j$$

Then we estimate the conditional distribution $P(y \mid s_i, s_j)$.

# Proposed Approach [cont...]

Step 2: sort baselines

$$H = [h_{s',s}] \in \mathrm{R}^{n\times n}, s', s \in S$$

$$h_{s',s} = \begin{cases} \arg\max_{y=\{0,1\}} P(y \mid s', s) & \text{if } s' \neq s \\ 0 & \text{otherwise} \end{cases}$$

Then sort the rows of $H$ by the number of zeros on it:

$$\nu_i^\star = \sum_{j=1}^{n} (1 - h_{s_i,s_j}), \quad 1 \leq i \leq n$$

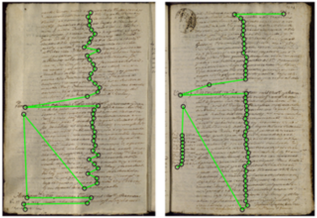
$$\mathbf{z}^\star = (s_1, \nu_1^\star), \ldots, (s_n, \nu_n^\star)$$

# Order operator estimator

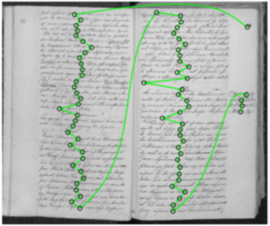
- *RDF:*
  - trained for each dataset using all the possible pairs of baselines per page
  - validation set was reserved to select the model hyperparameters
  - number of decision trees was set to 10
  - maximum depth of each tree was set to 10
- *MLP:*
  - composed of only one hidden layer
  - the number of hidden neurons was set to twice the number of input features
  - ReLU as activation function
  - ADAM optimizer and a learning rate of 0.001
  - training pairs were randomly generated on the fly.

# Datasets

OHG



FCR



ABP

# Quantitative Results

Table: Results for OHG, NAF and ABP datasets. Average over 10 experiments is reported for each metric, normalized Spearman Footrule ($\rho(\boldsymbol{t}, \boldsymbol{\nu})$) and Kendall's tau ($\overline{K}(\boldsymbol{t}, \boldsymbol{\nu})$) distance (in %), absolute Kendall's tau distance ($K(\boldsymbol{t}, \boldsymbol{\nu})$).In all the metrics the lower the better.

| dataset | model | $\rho(\boldsymbol{t}, \boldsymbol{\nu})$ | $\overline{K}(\boldsymbol{t}, \boldsymbol{\nu})$ | $K(\boldsymbol{t}, \boldsymbol{\nu})$ |
|---|---|---|---|---|
| OHG | TDLR | 2.91 | 1.49 | 12.97 |
| | RDF | 0.74 | 0.39 | 3.69 |
| | MLP | 0.67 | 0.34 | 3.31 |
| FCR | TDLR | 31.84 | 17.6 | 606.97 |
| | RDF | 0.95 | 0.54 | 21.64 |
| | MLP | 0.94 | 0.51 | 16.11 |
| ABP | TDLR | 10.93 | 7.67 | 3953.00 |
| | RDF | 6.36 | 3.94 | 2623.20 |
| | MLP | 7.83 | 4.93 | 2936.10 |

# Qualitative Results: OHG

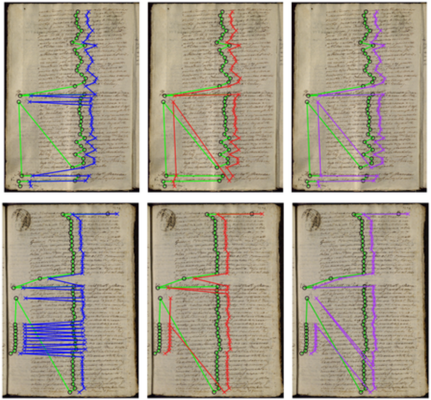


Figure: Two examples (left-right) of results obtained on OHG dataset where ground-truth is depicted in green and the hypothesis in blue for the TDLR approach, red for RDF and violet for MLP.

# Qualitative Results: FCR

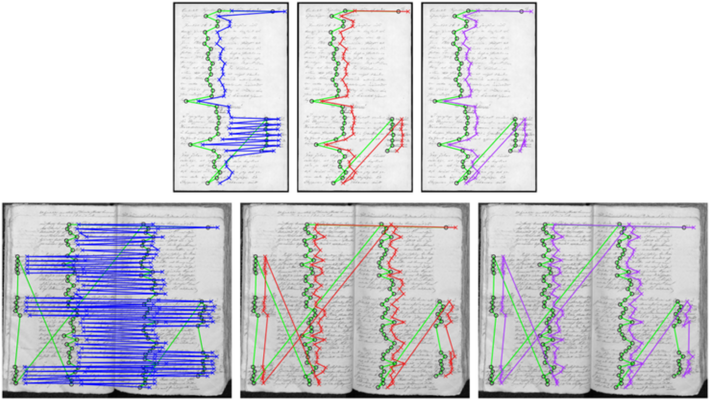


Figure: Two examples (left-right) of results obtained on FCR dataset where ground-truth is depicted in green and the hypothesis in blue for the TDLR approach, red for RDF and violet for MLP.

# Qualitative Results: ABP

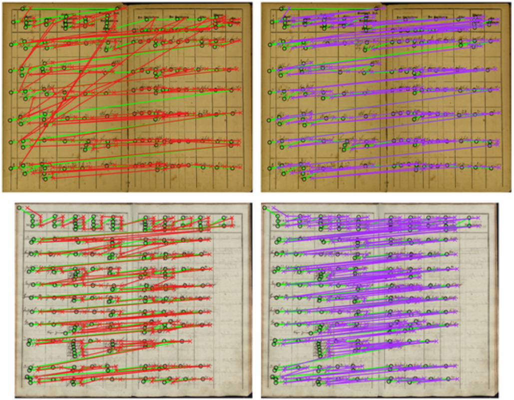


Figure: Two examples (left-right) of results obtained on ABP dataset where ground-truth is depicted in green and the hypothesis in blue for the TDLR approach, red for RDF and violet for MLP. (TDLR approach is dropped to improve readability)

# Conclusions

- We have introduced an automatically learned binary order-relation operator and demonstrated their effectiveness to solve the reading order problem in handwritten text documents.
- Experiments were carried out using three complex datasets and two different binary classifiers, with promising results in homogeneous datasets such as OHG and FCR.
- Although, further work has to be done in order to improve the binary classifier to handle heterogeneous datasets like ABP.

# Future Work

- Explore more powerful models for heterogeneous datasets as well as more complex features, such as the probabilistic index of each text line.
- Equally important is to extend the proposed method to take into account the full context of each page.

# Thanks for Your Attention !