A COMPARISON OF RECOGNITION STRATEGIES FOR PRINTED/HANDWRITTEN COMPOSITE DOCUMENTS

Bastien Moysset, Ronaldo Messina and Christopher Kermorvant

Motivation

Evaluate a RNN-based recognition system trained to recognize both handwritten and printed texts at the same time.

Summary of the different strategies used in the process

- Page image
- Text zone detection
- Writing type detection
  - a. Ground-truth
  - b. LITIS system
  - c. LITIS detector
- Paragraph image
- Text line segmentation
  - a. Ground-truth
  - b. IRISA detector
  - c. LITIS detector
  - d. Handwritten text line segmentation
- Text line image
- Text line recognition
  - a. Printed text recognizer
  - b. Handwritten text recognizer
  - c. Mixed text recognizer 1 label
  - d. Mixed text recognizer 2 labels
- Text line transcription

Database (Maurdor)

- Six classes, subset C1 comprises mainly forms. (English-only docs.)

Motivation

- Paragraph extraction algorithms on real-world images usually output blocks containing both types.
- Mixed RNN to deal with zones with single or both types.

Optical models

- Usual MDLSTM RNN; CTC training.
- Single-type: RNN-PRN and RNN-HWR.
- Mixed: 1 or 2 labels per character; in the latter labels are “tagged”: i.e. a_PRN and a_HWR.
- Number of units in all layers is the same, except the last layer for Mixed2.

Experimental Results

- (Mis-)Matched conditions: models for printed (RNN-PRN) and handwritten (RNN-HWR) text.
- Word-error-rate as figure of merit.

Located paragraphs with different optical models.

- Under mismatched conditions, single-type RNN’s performance is very low.
- Mixed RNN are nearly as good as the matched conditions.

Located paragraphs with GT/automatically detected type.

- On properly detected zones, matched-type RNNs are slightly better.
- Handwritten zones remain challenging for both detection and recognition.

Test on complete dataset and restricted to C1 documents

- GT and automatic text/type detection methods.

- On more difficult text zones, mixed-type is slightly better.
- C1 documents: the task is simplified, but text detection remains challenging.
- “Cascade” of errors affects the whole system.

Perspectives and future work

- Improve line detection for mixed-type zones.
- Train the RNN on snippets containing both handwritten and printed text.
- Test the generalization of RNNs to learn languages with different scripts, i.e. Roman and Arabic.
- Assess the impact of unbalanced data in our training dataset; there are much more printed material in the database.
- Tune the architecture of the RNN, allowing for more complex internal representations.
- Use the type information in the language models.

{bm,rm,ck}@a2ia.com