The A2iA Arabic Handwritten Text Recognition System at the OpenHaRT2013 Evaluation

Christopher Kermorvant, A2iA, Paris
A2iA ATR research group : J. Louradour, T. Bluche, B. Moysset, F. Benzeghiba, Maxime Knibbe
Arabic Handwriting Evaluations

Small vocabulary isolated word recognition has reached a plateau in 2009 (IFN-ENIT evaluations)

OPENHART IS THE REFERENCE EVALUATION FOR ARABIC HTR
Overview of A2iA's system

- Image pre-process
- Line segmentation
- Feature extraction character decoding
- Optical models RNN coeff.
- RNN
- Forms hypothesis graph
- Lexical decoding LM rescoring
- Textual corpora
- Vocabulary
- Language model
- KALDI
- SRILM
- Paragraph transcription
Image pre-processing

- NoRisk pre-processing
  - upscale in 600 DPI
  - Gaussian denoising
  - Deskew
  - Deslant

- Suppl. pre-processing
  - Line re-detection
  - Rule line removal
  - CC cleaning

THE PRE-PROCESSING USEFULL IN 2010 WITH FEATURE-BASED GMM ARE NOT USEFULL WITH RNN
Overview of A2iA’s system
Optical Model : **2D-LSTM RNN**
Optical Model: 2D-LSTM RNN

Note: peephole connections (in blue) not used
Optical Model: 2D-LSTM RNN

DEEP RECURRENT NEURAL NETWORKS
Optical Model: 2D-LSTM RNN
Optical Model: training several RNN

Phase 1-2-3 Train

Split 0
Split 1
Split 2
...
Split 10

WORD LEVEL TRAINING

RNN 0
RNN 1
RNN 2
...
RNN 10

Phase 1-2-3 Train

LINE LEVEL TRAINING

RNN 0
RNN 1
RNN 2
...
RNN 10
Overview of A2iA’s system
Language Model: vocabulary selection

OPENHART

Phase 1-Train
Phase 2-Train
Phase 3-Train

C(w_i) = \sum_{k=1}^{K} \lambda_k C(w_{ik})

E.M. OPTIMIZATION

Phase 1-Dev
Phase 2-Dev

TOP 60K

OOV: 9%

Phase 1-Eval
Language Model: N-gram interpolation

\[ P(w|h) = \sum_{k=1}^{K} \lambda_k P(w|h) \]

- **Phase 1-Train**
- **Phase 2-Train**
- **Phase 3-Train**

EM OPTIMIZATION

PRUNING

3-Gram
KN SMOOTHING

PPX: 975
Overview of A2iA's system

- **Image pre-process**
- **Line segmentation**
- **Feature extraction**
  - Optical models
  - RNN coeff.
- **Character decoding**
- **RNN**

**Training image corpora**

**Test image corpora**

**Textual corpora**

**Vocabulary**

**Language model**

**Optical models**

**RNN coeff.**

**forms hypothesis graph**

**Lexical decoding**

**LM rescoring**

**Paragraph transcription**

**KALDI**

**SRILM**
Decoding: composition of WFST

WFST composition and decoding with KALDI
Rover Combination

- Usually, combining several RNN helps a lot
- We had many variants (RNN initialization, hyperparameters, etc)
- Best Single recognizer : 25% WER (Phase 1-Eval)
- Best ROVER combination : 23.3% WER

SMALL IMPROVEMENT WITH ROVER
OpenHaRT’13 Evaluation

Constrained conditions

WER divided by 2 since 2010
Error analysis: writing conditions

PEN
- Ink
- Pencil

PAPER
- Unlined
- Lined

SPEED
- Careful
- Normal
- Fast
Error analysis: scribes

- **SEX**
  - Female: 18.6
  - Male: 21.1

- **COUNTRY**
  - TUNIS
  - IRAQ
  - SAUDI ARABIA
  - ALGERIA
  - MOROCCO
  - EGYPT
  - SYRIA

- **AGE**

* not statistically significant
OpenHaRT Evaluation

Evaluation conditions closer to historical document processing (US Census) than to digital mailroom (mail sorting in companies)

☛ See the Maurdor Evaluation
Thank you!