The RWTH Large Vocabulary Arabic Handwriting Recognition System

OpenHaRT 2013 Workshop, Washington DC

Mahdi Hamdani[1], Patrick Doetsch[1], Michal Kozielski[1],
Hendrick Pesch[1], Amr El-Desoky Mousa[1]
Hermann Ney[1,2]

[1]Lehrstuhl für Informatik 6
Human Language Technology and Pattern Recognition
Computer Science Department, RWTH Aachen University
D-52056 Aachen, Germany

[2]Spoken Language Processing Group
LIMSI CNRS, Paris, France

Aug. 23, 2013

Outline

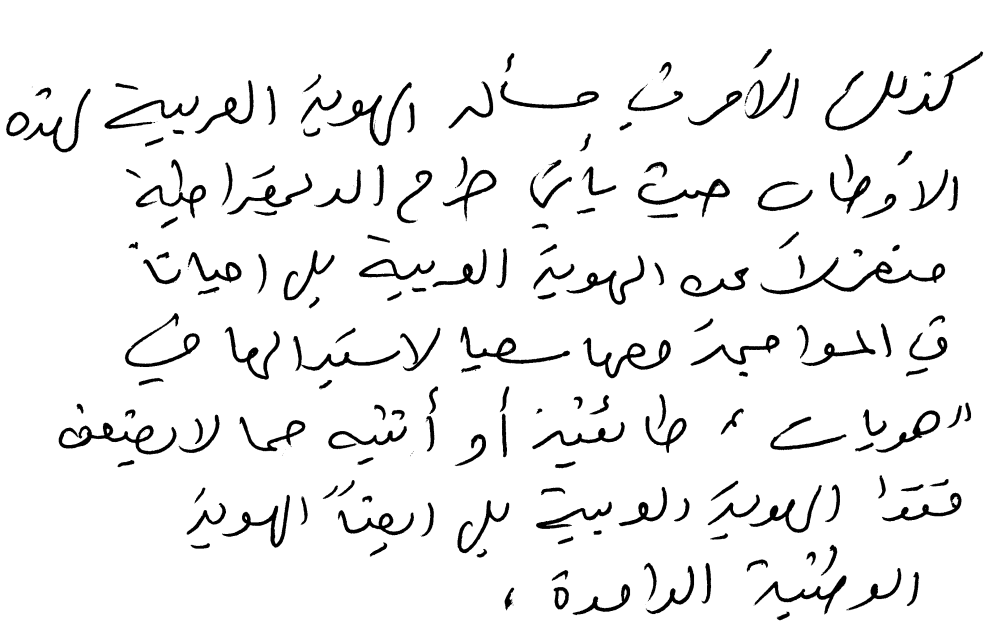
- Introduction
  - State of the Art
  - Feature Extraction
  - Visual Model
  - Language Model
  - Results
- Conclusions and Future Work

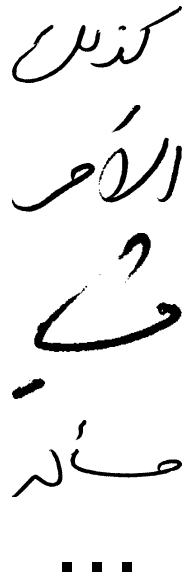
Introduction

What are the used databases?

OpenHaRT database

- Large Arabic Handwriting database
- Pages of handwritten text
- Pages are segmented into words and lines
- Paragraphs are typically multiple lines

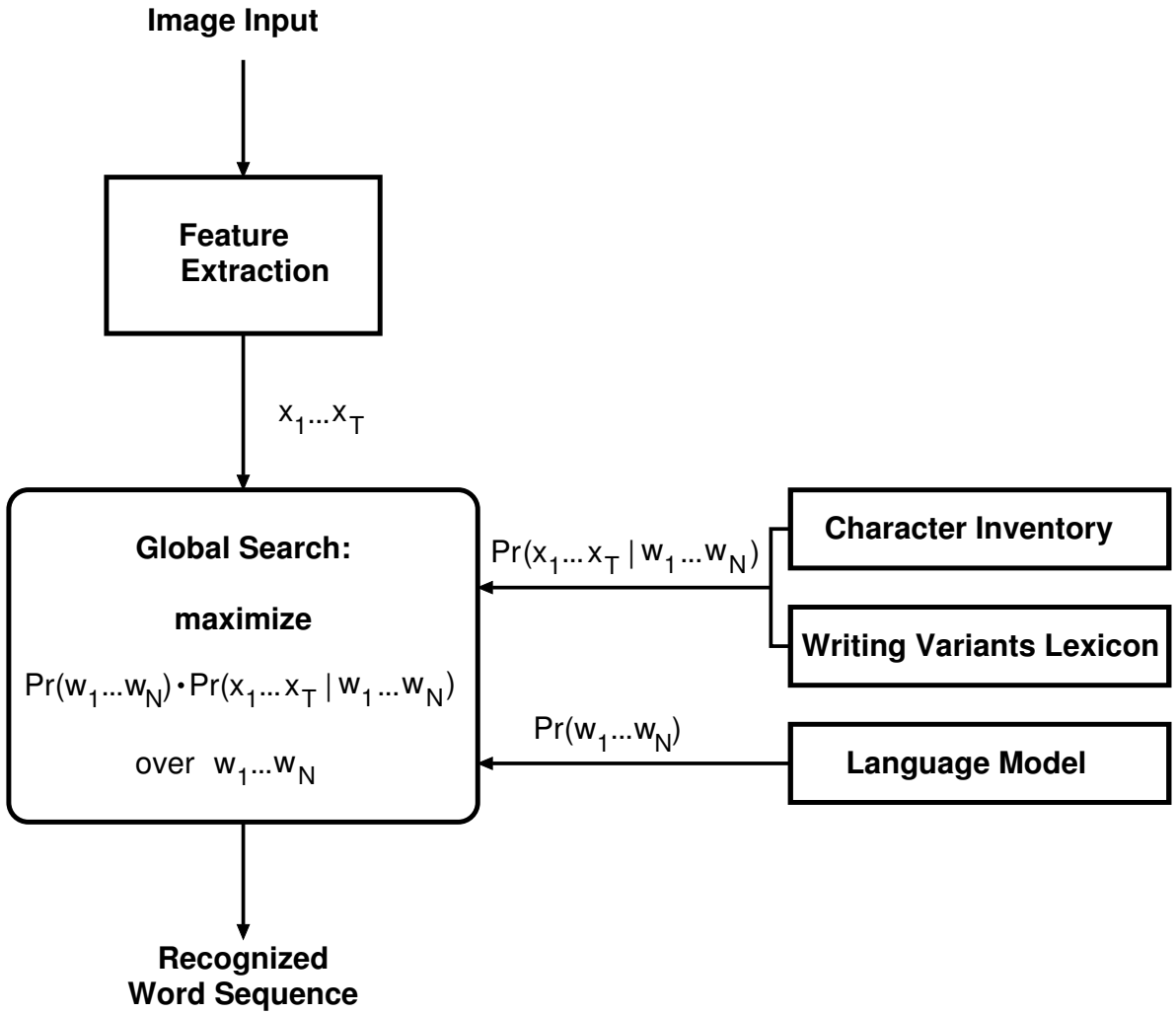




Introduction...

RWTH OCR: System Overview

Recognition

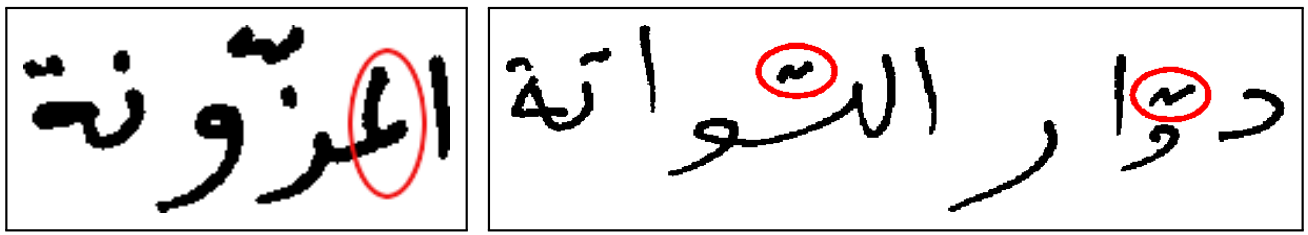


Introduction...

Arabic Language Specificities

Arabic

- Right to left cursive writing: 28 base characters
- Ligatures, diacritics optional in handwriting!
- Letter can have many shapes (position dependent)



(a) Ligatures (b) Diacritics

State of the art

Arabic Handwriting Recognition Competition

- ICDAR (2005, 2007, 2009 and 2011) and ICFHR 2010
  - Best systems are based on Hidden Markov Models (HMMs) or Long Short Term Memory (LSTM) Neural Networks
  - Graves and Schmidhuber, Offline Handwriting Recognition with Multidimensional Recurrent Neural Networks, NIPS 2008 (TU Munich)
  - Doetsch et al. Comparison of Bernoulli and Gaussian HMMs using a vertical repositioning technique for off-line handwriting recognition, ICFHR 2012 (RWTH Aachen)

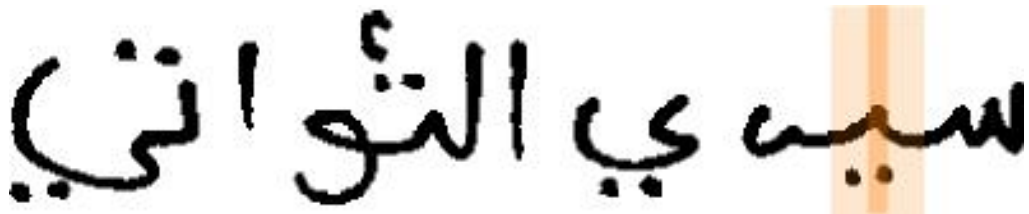
NIST 2010 OpenHaRT evaluation

- Best system based on HMMs
- Bianne-Bernard et al., "Dynamic and Contextual Information in HMM Modeling for Handwritten Word Recognition," IEEE TPAMI, vol. 33, no. 10, pp. 2066-2080, October, 2011 (a2ia)

Feature Extraction

Appearance-Based

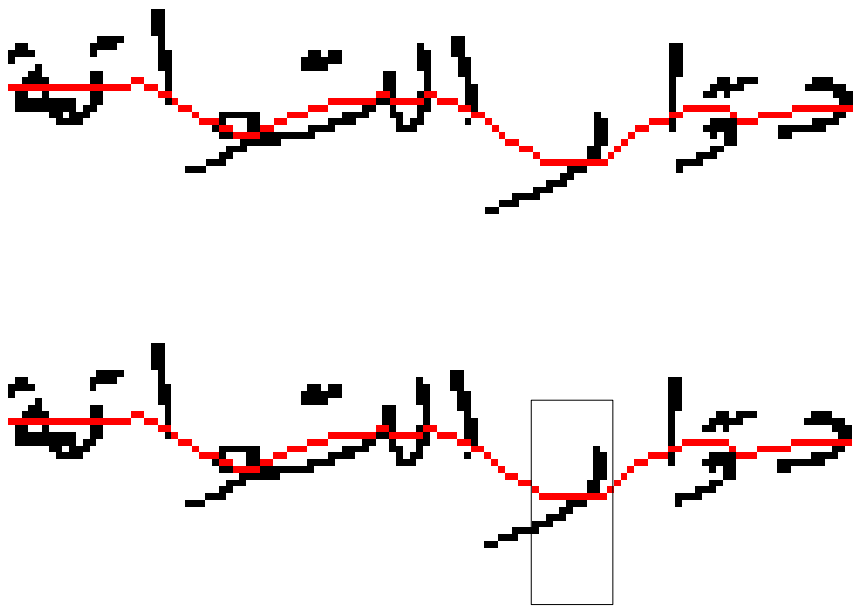
- Images are scaled to the same height
- Recognition of characters within a context
- Sliding window, PCA reduction
- Typically: large context-window with maximum overlap



Feature Extraction...

Window Repositioning Technique

- Images Scaling
- Center of gravity calculation
- Window repositioning
- Features are gray scale pixel values



Visual Model

Context Dependent Model

Cart Decision tree

- Nodes are tagged with questions
- Leaves are tagged with class labels
- Questions concern the visual classes

Types	Examples of Characters	Images
Small Ascenders	س, ش, ز	
Descenders	و, ر, ز	
Occlusions	ه, ة, ف	

Discriminative training

Motivation

HMM Training Approaches

- Maximum Likelihood (ML) training
  - Separate model constructed for each class
  - Only in-class information is available!
- Discriminative training
  - Aim is to separate the classes
  - Classifier performance reflected by objective function
  - Training Criterion: e.g. Minimum Phone Error (MPE)

Framewise Training Approaches

- (Discriminative) training with fixed segmentation
- Segmentation has to be provided
- Here: Neural networks

Discriminative HMM training

Objective Function

Definitions

- $\boldsymbol{X}$: Sequence of observation vectors over time
- $\boldsymbol{W}$: Written word sequence
- $\boldsymbol{p_\Lambda(W|X)}$: Class posterior distribution with parameter set $\boldsymbol{\Lambda}$

Training

- Training examples $\boldsymbol{(X_r, W_r)}$
- Criterion

$$\hat{\boldsymbol{\Lambda}} = \boldsymbol{\arg\min_\Lambda \{\sum_{r=1}^{R} L[p_\Lambda(W_r|X_r)] + reg\_term(\Lambda)\}}$$

with a loss function $\boldsymbol{L[p_\Lambda(W_r|X_r)]}$

Discriminative training

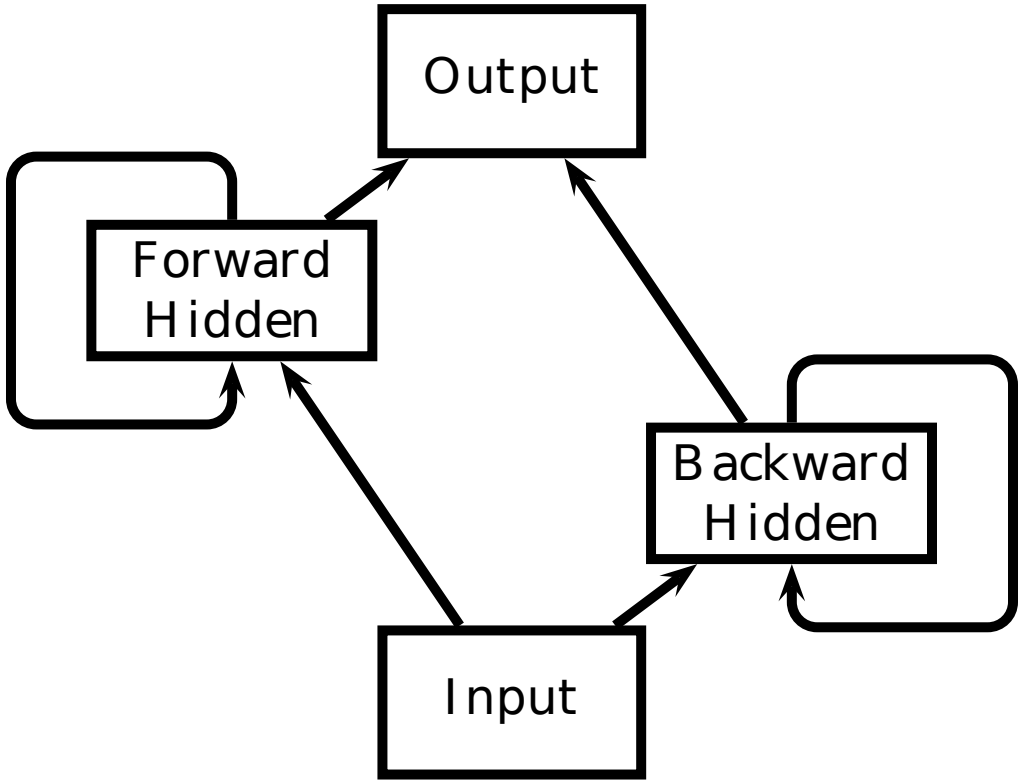
Training Criterion...

Minimum Phone Error (MPE)

$$L^{(\mathrm{MPE})}[p_\Lambda(X_r, \cdot), W_r] = \sum_{W \in \cdot} E(W, W_r) \frac{p_\Lambda(X_r, W)^\gamma}{\sum_V p_\Lambda(X_r, V)^\gamma} \quad (1)$$

$E(W, W_r)$: Measure of correctly transcribed characters in W
X_r: observation vector sequence
W_r: transcription word sequence
γ: approximation level (controls the smoothness of the criterion)

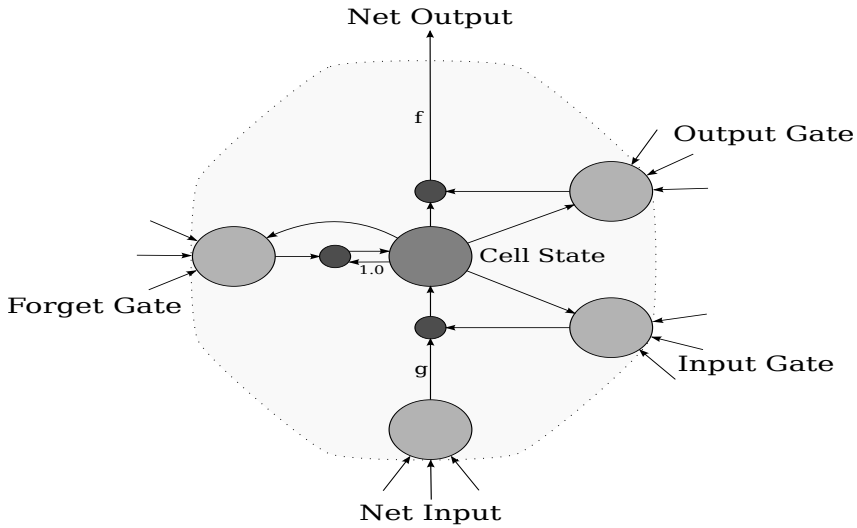
Neural Network Feature Extraction



- Framewise training with recurrent neural networks
- Non-linear feature transformation
- Full context modeling through bidirectional topology

Long Short Term Memory (LSTM)

- Constant error flow without loss of short time lag capability
- Replace units by memory cells



- Input Gate: Protects error flow inside cell from irrelevant inputs
- Output Gate: Protects error flow of other cells from irrelevant inputs
- Forget Gate: Provides a way to reset cell state

ANN/HMM Combination

- Idea: Train neural network on aligned feature vectors (x_1^T, s_1^T)
  - Alignment must be provided by previously trained HMM
- Tandem: Use posterior probabilities as features to train HMM
- Posteriors highly correlated: LDA/PCA to n dimensions
  $\Rightarrow$ Requires full retraining of HMM
- Hybrid: Use posterior probabilities as state emission probability of HMM

$$p_t(x_t|s) \stackrel{!}{=} \frac{p_t(s|x_t)}{p(s)^\alpha}$$

- α: Priori scaling factor

Language Model Training

Morphological Decomposition

MADA toolkit

- Morphological Analysis and Disambiguation for Arabic (Habash et al. 2009)
- Tool for morphological and contextual analysis of raw Arabic text
- Examines all possible analyses for each word
- Selects the analysis that matches the current context best
- Tokenize the disambiguated text generated by MADA
- A tokenization scheme is provided

Language Model Training...

MADA Toolkit

- Different variables for the tokenization scheme

Type	Variable	Description
Prefix	QUES	The "question" proclitic (e.g. أَ)
	CONJ	"Conjunction" proclitic (و and ف)
	PART	"Article" proclitic
	FUT	Future marker (سـ)
	NART	Negative articles only (لَا and مَا)
	DART	Definite article (اَلْ)
Radical	REST	Remainder of the word
Suffix	PRON	Enclitics

For example, the word وسيكَاتبهَا will be decomposed to و (conjunction) + سـ (future marker clitic) + يكَاتب (rest) + هَا (suffix).

Language Model Training...

Vocabulary selection

Vocabulary Selection

- The M most frequent full-words are not decomposed
- The selected vocabulary contains new elements which are prefixes, suffixes and stems
- Prefixes and suffixes are tagged with a special marker ("+")
- Recognition of unknown words is possible by combining the vocabulary elements

Language Model

- Collected in domain text
- Decomposition using MADA toolkit
- Standard n-gram LM trained using the SRILM toolkit (Stolcke et al. 2012)

Results

Tasks

Arabic Handwriting Recognition Competition

- IfN/ENIT database
- Limited vocabulary size

OpenHaRT 2013 evaluation

- Constrained task
  - LM restricted to the training text
- Unconstrained task
  - Additional data used for the LM training
  - The used data is collected from publicly available newspapers and web-forums
  - 1 billion running words

Results...

Results on the IfN/ENIT task

System	WER [%]	CER [%]
GHMM, MLP Tandem (ICDAR'11)	**5.9**	**4.7**
GHMM, MLP Hybrid (ICDAR'11)	10.3	8.1
GHMM	13.1	10.6
+ Repo.	6.4	4.6
GHMM, LSTM Tandem	7.2	5.6
+ Repo., [Doetsch et al., 2012]	**4.8**	**3.7**
BHMM, UPV, [Doetsch et al., 2012]	6.2	-
MD-LSTM, TUM, [Graves et al., 2009]	6.6	-

Results...

OpenHaRT Data Description

Table: OpenHaRT data statistics

	Train set	Dev set
# of pages	42,148	470
# of paragraphs	182,879	1,832
# of words	4,361,056	48,832
# of characters	23,324,011	266,121
avg number words/paragraph	23.85	26.65
avg number characters/word	5.35	5.45

Results...

Results on OpenHaRT for the constrained task

Table: Results of the RWTH handwriting recognition system on the OpenHaRT constrained task

System	Vocabulary size	WER [%]	CER [%]
Baseline	99k	27.4	10.9
Sub-lexical approach	94k	**26.8**	**10.1**

OOV rates

- Baseline full-words: 8.29%
- Sub-lexical: 5.70%

Results...

Results on OpenHaRT for the unconstrained task

- Vocabulary size: 200k
- OOV: 3.5%

Table: Results of the RWTH handwriting recognition system on the OpenHaRT unconstrained task

System	WER [%]	CER [%]
GHMM CI	33.2	15.4
GHMM CD	25.9	10.1
+BLSTM	19.9	5.9
+MPE	17.0	4.5

Conclusions and Future Work

Conclusions

- Morphological Decomposition using MADA toolkit
- Improvement up to 1% in the constrained task
- Same results of the baseline system in the unconstrained task
- Lexicon flexibility with competitive results in very large vocabulary task

Future Work

- Comparison of the morphological decomposition with other types of decomposition (e.g. part of words)
- Combination with character based language models

Thank you for your attention

Mahdi Hamdani, Patrick Doetsch, Michal Kozielski,
Hendrick Pesch, Amr El-Desoky Mousa
Hermann Ney

hamdani@i6.informatik.rwth-aachen.de