The UPV Handwriting Recognition and Translation System for OpenHaRT 2013

DSIC/ITI
Universitat Politècnica de València
E-46022 València (Spain)

UNIVERSITAT POLITÈCNICA DE VALÈNCIA

I. Khoury, A. Giménez, J. Andrés, A. Juan, J.A. Sánchez
`[ialkhoury,agimenez,jandres,ajuan,jandreu]@dsic.upv.es`

August 23, 2013

Site Introduction

- Pattern Recognition and Human Language Technology research group (PRHLT)
- From the Universitat Politècnica de València (UPV)
  - DSIC and DISCA of the Universitat Politècnica de València (UPV)
  - Instituto Tecnológico de Informática (ITI) from UPV
- Interests:
  - Multimodal Interaction
  - Machine Translation
  - Handwritten Text Recognition (HTR) and Document Analysis
  - Automatic Speech Recognition and Understanding
  - Image Analysis and Computer Vision
  - Transcription and Translation of Video lectures (transLectures) [1]

Site Introduction (Cont.)

- Related work and current research in HTR:
  - HTR using Bernoulli and Gaussian HMMs applied to:
    - Arabic IFN/ENIT database [9]
    - Arabic APTI database for Printed Arabic [10]
    - NIST OpenHaRT 2010 and 2013 (LDC) corpus
    - IAM database [7]
  - BHMMs using discriminative training

Transcription System

- Image Processing
  - Scaling to a given height (30 pixels)
  - Image Binarization using Otsu method
- Text Processing
  - Adding shape information to Arabic transcripts
- Feature extraction
  - Window extraction to a given width (9 pixels)
  - Window repositioning to its center of mass
    - Vertical, Horizontal, and Both directions (Vertical)
- HMM system using Bernoulli mixtures (BHMM)
  - Fixed number of states (6 states per character)
  - Mixture components per state (128)
  - Tri-character approach
  - EM algorithm for training and recognition
  - 5-grams Language Model (LM) for recognition
  - Grammar Scale Factor (GSF) on LM (30)

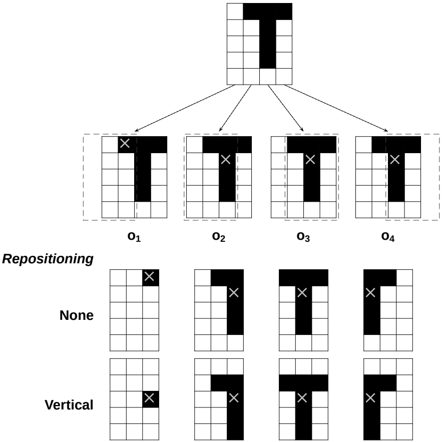

Figure: Example of transformation of a 4×5 binary image (top) into a sequence of 4 15-dimensional binary feature vectors $O = (\mathbf{o}_1, \mathbf{o}_2, \mathbf{o}_3, \mathbf{o}_4)$ using a window of width 3. After window extraction, the standard method is compared with the vertical repositioning. Mass centers of extracted windows are also indicated.

Translation System

- Our system is based on a state-of-the-art log-linear translation system (Moses toolkit)
- Standard moses features
  - Phrased-based model
    - Phrase translation probabilities (both directions)
    - Lexical weights (both directions)
  - Language Model (5-grams trained with SRILM)
  - Distance-based reordering model
  - Word penalty
  - Lexicalized reordering model

Translation System (Cont.)

- Text processing
  - tokenization:
    - English was tokenized with Moses tokenization tools
    - Arabic was tokenized with *MADA+TOKAN* tools
  - Removing long sentences (longer than 150 words)
- Standard Moses training
  - Alignment extraction
  - Phrase extraction
  - MERT

Submissions

- Document Image Recognition (DIR)
  - Two systems followed the constrained training condition
  - Trained with our BHMMs approach
  - The contrastive system was trained using the complete data
  - The primary system was trained with less data
- Document Text Translation (DTT)
  - Two systems: Different training conditions
  - Trained with Moses toolkit
  - For the constrained training condition:
    - We used only the LDC resources for the OpenHaRT'13
  - For the unconstrained training condition:
    - We used the MultiUN and TED corpus (IWSLT 2011)
    - Aligned on sentence level using the Champollion tool
    - Sentences were selected according to the infrequent *n*-grams score

Submissions (Cont.)

- Document Image Translation (DIT) Given a handwritten image f, it can be expressed as follows

$$y^{\star} = \underset{y \in Y}{\operatorname{argmax}}\ p(y|f) = \underset{y \in Y}{\operatorname{argmax}} \sum_{x} p(x|f)\, p(y|x) \qquad (1)$$

  where,
  f: input image
  x: candidate recognized source (Arabic) text
  y: candidate translated sentence (in English) corresponding to f.

  - Three systems followed the constrained training condition
  - The probability $p(x \mid f)$ in Eq. (1) was approximated by the primary DIR transcription system
  - The key difference among systems lay in the translation subsystems

Submissions (Cont.)

Translation subsystems for the DIT task (Three Systems)

- The first DIT system (DIT1),
  Eq. (1) was approximated as follows,

$$\begin{aligned} y^* &\approx \operatorname*{argmax}_{y \in Y} [\max_{x}\{p(x|f)\, p(y|x)\}] \\ &\approx \operatorname*{argmax}_{y \in Y} [p(y|\max_{x}\{p(x|f)\})] \end{aligned} \tag{2}$$

  *The $p(y|x^{\star})$ was approximated by the primary DTT translation system
- The input image was recognized by the primary DIR transcription system, and the recognized text was fed into the primary DTT translation system.

Submissions (Cont.)

The second DIT system (DIT2):

- Followed a similar approach to the first DIT system
- The source part of each bilingual training pair was substituted by the transcription obtained by the primary DIR system
- The new training data set produced in this way was used to train the translation system
- It was expected to better handle the noisy output of the DIR system
- Better performance than the primary DTT in development set but worse performance in the test set

Submissions (Cont.)

The third DIT system (DIT3):

- Different approximation of Eq. (1) was used

$$y^{\star} = \underset{x \in \text{NBest}(f)}{\text{argmax}} \left\{ \underset{y \in \text{NBest}(f|x)}{\text{argmax}} \{p(x|f)\,[p(y|x)]^{\theta}\} \right\} \quad (3)$$

- Introducing a scaling factor θ
- The search space was approximated by N-best lists
- Each input image was first recognized using the primary DIR system into 100-Best transcriptions, and then each transcription was translated using the primary DTT system into 100-Best translations

Data statistics

Table: Data (lines) used for training each system and its training conditions.

System/Condition	Constrained	Unconstrained
DIR1	779,100	-
DIR2	789,874	-
DIT (recognition part)	779,100	-

Table: Data (segments) used for training each system and its training conditions.

System/Condition	Constrained	Unconstrained	
Corpus	LDC	MultiUN	TED
DTT	40,580	19,956	2,205
DIT (translation part)	40,580	-	-

Tools and Means

For Text Processing:

- Moses tokenization tools [5]
- MADA+TOKAN [8] toolkit
- Champollion Toolkit (CTK) [6]

For Handwritten Text Recognition:

- TLK toolkit [2]

For Machine Translation:

- MGIZA++ [3] to establish the word alignments .
- Moses toolkit [5]

Results (Line Condition)

Table: Submitted systems for DIR and line segmentation condition together with their Word Error Rate (WER%)

System	Reference	WER [%]	
		Eval'10	Eval'13
DIR1	p-1_1_20130425	29.08	29.32
DIR2	c-1_2_20130425	-	**29.20**
UPV PRHLT	OpenHaRT'10	47.45	-

- The DIR2 system slightly outperforms the DIR1 system
  - Expected improvement: DIR2 was trained with more data
- Both DIR1 and DIR2 systems outperform the DIR system of the 2010 evaluation (UPV PRHLT)
  - Trained with more mixture components (128) per state
  - We used a bigger language model for recognition.

Results (Line Condition)

Table: Submitted systems for (DTT and DIT) and line segmentation condition together with their BLEU score

System	Reference	BLEU [%] Eval'10	BLEU [%] Eval'13
DTT Constrained	p-1_1_20130425	22.53	21.93
DTT Unconstrained	p-1_1_20130425	25.18	24.10
DIT1	p-1_1_20130425	16.51	16.95
DIT2	c-1_2_20130425	16.58	16.52
DIT3	c-1_3_20130425	18.13	**17.49**

- The Unconstrained DTT system significantly outperforms the Constrained DTT system.
  - The usage of an additional data (around $20K$) significantly improved the translation accuracy in the DTT system.
  - Sentence selection according to the infrequent *n*-grams score [4]
- The DIT3 shows better performance over DIT1 and DIT2

Conclusion

- The UPV Recognition and Translation System for the NIST OpenHaRT'13 evaluation.
- Submissions:
  - Two systems for the DIR task (constrained training condition)
  - One system for the DTT task (both training conditions)
  - Three systems for the DIT task (constrained training condition)
- Results for the DIR task outperform previous results in OpenHaRT 2010 evaluation
- Results for DTT and DIT tasks are very promising

The translectures project. http://translectures.eu/tlk. http://www.translectures.eu/, 2013.

The translectures-upv team. the translectures-upv toolkit (tlk). http://translectures.eu/tlk. http://www.translectures.eu/tlk/citing-tlk/, 2013.

Qin Gao and Stephan Vogel.
Parallel implementations of word alignment tool.
In *In Proc. of the ACL 2008 Software Engineering, Testing, and Quality Assurance Workshop*, 2008.

Guillem Gascó, Martha-Alicia Rocha, Germán Sanchis-Trilles, Jesús Andrés-Ferrer, and Francisco Casacuberta.
Does more data always yield better translations?
In *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics*, pages 152–161, Avignon, France, April 2012. Association for Computational Linguistics.

P. Koehn, H. Hoang, A. Birch, C. Callison-Burch, M. Federico, N. Bertoldi, B. Cowan, W. Shen, C. Moran, and R. Zens.
Moses: Open source toolkit for statistical machine translation.
In *Annual meeting-association for computational linguistics*, volume 45, page 2, 2007.

X. Ma.
Champollion: A robust parallel text sentence aligner.
In *LREC 2006: Fifth International Conference on Language Resources and Evaluation*, page 489–492, 2006.

U. V. Marti and H. Bunke.
The IAM-database: an English sentence database for offline handwriting recognition.
IJDAR, 5(1):39–46, 2002.

Owen Rambow Nizar Habash and Ryan Roth.

Mada+tokan: A toolkit for arabic tokenization, diacritization, morphological disambiguation, pos tagging, stemming and lemmatization.
In Khalid Choukri and Bente Maegaard, editors, *Proc. of the 2nd Int. Conf. on Arabic Language Resources and Tools*, Cairo, Egypt, April 2009. The MEDAR Consortium.

M. Pechwitz et al.
IFN/ENIT - database of handwritten Arabic words.
In *CIFED '02*, pages 21–23, Hammamet (Tunis), oct 2002.

Fouad Slimane, Rolf Ingold, Slim Kanoun, Adel M. Alimi, and Jean Hennebert.
A new arabic printed text image database and evaluation protocols.
pages 946–950. IEEE, 2009.