Learning Hypotheses
Decoding in an Image Text Recognition Pipeline

Jindřich Libovický

Week of Doctoral Studies,
Prague, May 19, 2014
Outline

Image Text Recognition

Learning the Decoding

Evaluation

Results

Future Work
Outline

Image Text Recognition

Learning the Decoding

Evaluation

Results

Future Work
Image Text Recognition

FREEDOM

Jindřich Libovický, Charles University in Prague, September 16, 2014
Learning Hypotheses Decoding in an Image Text Recognition Pipeline, 4/21
CMP: TextSpotter

- a tool developed at Centre for Machine Perception at the Czech Technical University
- input: an image, output: rectangles with words and their transcriptions
- scores well in the ICDAR competition
- only text localization and rectangle transcription in the competition
CMP: TextSpotter

Image Text Recognition

Jindřich Libovický, Charles University in Prague, September 16, 2014
Learning Hypotheses Decoding in an Image Text Recognition Pipeline, 6/21
Outline

Image Text Recognition

Learning the Decoding

Evaluation

Results

Future Work
Training Data Preparation

- generate graphs from images → match with annotation
Training Data Preparation

- generate graphs from images $\rightarrow$ match with annotation
- ICDAR train set – 229 images
Training Data Preparation

- generate graphs from images → match with annotation
- ICDAR train set – 229 images
- 1607 graphs generated, 812 matched with annotation
Training Data Preparation

- generate graphs from images → match with annotation
- ICDAR train set – 229 images
- 1607 graphs generated, 812 matched with annotation
- 568 used for training, 244 for intrinsic evaluation
Features

- originally 4 features: detected area similarity, OCR confidence, fitting the detected direction of text, simple language model
Features

- originally 4 features: detected area similarity, OCR confidence, fitting the detected direction of text, simple language model

**bigram features**

- width, height, area ratio
- top line and bottom line deviations
- patterns: Xx, xx, XX, numbers
- bigram character language model

**in total 20 features**
Features

- originally 4 features: detected area similarity, OCR confidence, fitting the detected direction of text, simple language model

**bigram features**
- width, height, area ratio
- top line and bottom line deviations
- patterns: Xx, xx, XX, numbers
- bigram character language model

**trigram features**
- spaces ratio
- top line, bottom line, and central line angles
- character patterns
- trigram character language model

*in total 20 features*  
*another 9 features*
Independent learning

- all edges from all training graphs
Independent learning

- all edges from all training graphs
- on the correct path \(\Rightarrow\) positive examples
  others \(\Rightarrow\) negative examples
Independent learning

- all edges from all training graphs
- on the correct path ⇒ positive examples
  others ⇒ negative examples
- path maximizes sum of the scores from the classifiers
Structured Learning

- problem needs to be defined in the following form:

\[
\hat{y} = \arg\max_{y \in Y_X} w^T \psi(x, y)
\]
Structured Learning

- problem needs to be defined in the following form:

\[
\hat{y} = \arg\max_{y \in \mathcal{Y}_x} w^T \psi(x, y)
\]

- $\mathcal{X}$ ... all possible graphs
- $\mathcal{Y}_x$ ... all possible paths graph $x \in \mathcal{X}$
- $\psi(x, y)$ ... feature vector for path $y$ in graph $x$

\[
\psi(x, y) = \sum_{e \in y} \phi(e)
\]

- $w$ ... weight vector
Structured Learning

- problem needs to be defined in the following form:

\[
\hat{y} = \arg\max_{y \in Y_x} w^T \Psi(x, y)
\]

- \(X\) ... all possible graphs
- \(Y_x\) ... all possible paths graph \(x \in X\)
- \(\Psi(x, y)\) ... feature vector for path \(y\) in graph \(x\)

\[
\Psi(x, y) = \sum_{e \in y} \phi(e) \quad \leftarrow \text{we want to guess this}
\]

- \(w\) ... weight vector \(\leftarrow \text{we want to learn this}\)
Structured Predicition

- Structured Perceptron
  - simple modification of the standard Perceptron algorithm

- Structured SVM
  - weights optimized by quadratic programming
  - not constant margin, but a loss function
  - exponential number of path in a graph \(\Rightarrow\) exponentially many inequalities for quadratic programming
  - approximative algorithm
Outline

Image Text Recognition

Learning the Decoding

Evaluation

Results

Future Work
Intrinsic Evaluation

- 30% of extracted graphs from the ICDAR train set
Intrinsic Evaluation

- 30% of extracted graphs from the ICDAR train set
- counting edges incomparable between bigram graphs and trigram graphs
Intrinsic Evaluation

- 30% of extracted graphs from the ICDAR train set
- counting edges incomparable between bigram graphs and trigram graphs
- string measures comparing decoded and correct string:
Intrinsic Evaluation

- 30% of extracted graphs from the ICDAR train set
- counting edges incomparable between bigram graphs and trigram graphs
- string measures comparing decoded and correct string:
  - average Levenshtein distance
Intrinsic Evaluation

- 30% of extracted graphs from the ICDAR train set
- counting edges incomparable between bigram graphs and trigram graphs
- string measures comparing decoded and correct string:
  - average Levensthein distance
  - average relative Levensthein distance
Intrinsic Evaluation

- 30% of extracted graphs from the ICDAR train set
- counting edges incomparable between bigram graphs and trigram graphs
- string measures comparing decoded and correct string:
  - average Levenshtein distance
  - average relative Levenshtein distance
  - average length difference
Intrinsic Evaluation

- 30% of extracted graphs from the ICDAR train set
- counting edges incomparable between bigram graphs and trigram graphs
- string measures comparing decoded and correct string:
  - average Levenshtein distance
  - average relative Levenshtein distance
  - average length difference
  - full string accuracy
Extrinsic Evaluation

- ICDAR test data
Extrinsic Evaluation

- ICDAR test data
- text localization task
Extrinsic Evaluation

- ICDAR test data
- Text localization task
  - Shared task from ICDAR
Extrinsic Evaluation

- ICDAR test data
- text localization task
  - shared task from ICDAR
  - dismissed the textual content
Extrinsic Evaluation

- ICDAR test data
- Text localization task
  - Shared task from ICDAR
  - Dismissed the textual content
- Letter localization + correctness
Extrinsic Evaluation

- ICDAR test data
- text localization task
  - shared task from ICDAR
  - dismissed the textual content
- letter localization + correctness
- precision, recall, F1-measure
Extrinsic Evaluation

- ICDAR test data
- text localization task
  - shared task from ICDAR
  - dismissed the textual content
- letter localization + correctness

- precision, recall, F1-measure
- for rectangles 90% area overlap required
Outline

Image Text Recognition

Learning the Decoding

Evaluation

Results

Future Work
## Intrinsic Measures

<table>
<thead>
<tr>
<th></th>
<th>avg. edit dist.</th>
<th>avg. rel. edit dist.</th>
<th>avg. length diff.</th>
<th>full string acc.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>bigram edges</strong></td>
<td>baseline</td>
<td>.6471</td>
<td>.1317</td>
<td>.0336</td>
</tr>
<tr>
<td></td>
<td>indep. class.</td>
<td>.3320</td>
<td>.0682</td>
<td>.0615</td>
</tr>
<tr>
<td></td>
<td>S. Perceptron</td>
<td>.4631</td>
<td>.0917</td>
<td>.1352</td>
</tr>
<tr>
<td></td>
<td>S. SVM</td>
<td>.4385</td>
<td>.0817</td>
<td>-.0041</td>
</tr>
<tr>
<td></td>
<td>S. SVM + indep. cl.</td>
<td>.3770</td>
<td>.0798</td>
<td>.0574</td>
</tr>
<tr>
<td><strong>trigram edges</strong></td>
<td>indep. class.</td>
<td>.3975</td>
<td>.0749</td>
<td>.0451</td>
</tr>
<tr>
<td></td>
<td>S. Perceptron</td>
<td>.4877</td>
<td>.1035</td>
<td>.0902</td>
</tr>
<tr>
<td></td>
<td>S. SVM</td>
<td>.4016</td>
<td>.0768</td>
<td>.1148</td>
</tr>
<tr>
<td></td>
<td>S. SVM + indep. cl.</td>
<td>.3975</td>
<td>.0765</td>
<td>.0779</td>
</tr>
</tbody>
</table>
Extrinsic Measures

???
Outline

Image Text Recognition
Learning the Decoding
Evaluation
Results

Future Work
Future Work

- finish the extrinsic evaluation (in progress)
Future Work

- Finish the extrinsic evaluation (in progress)
- Employ structured prediction method with non-linear decision boundary
Future Work

- finish the extrinsic evaluation (in progress)
- employ structured prediction method with non-linear decision boundary
- automatically get more training data (in progress)
Future Work

- finish the extrinsic evaluation (in progress)
- employ structured prediction method with non-linear decision boundary
- automatically get more training data (in progress)
- publish the work
Thank you for your attention.