Charles University in Prague
Institute of Formal and Applied Linguistics

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

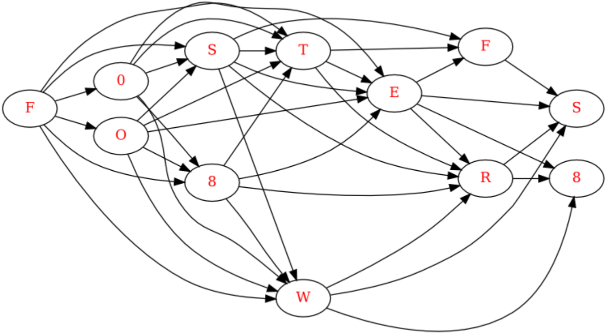
→ FREEDOM

CMP: TextSpotter

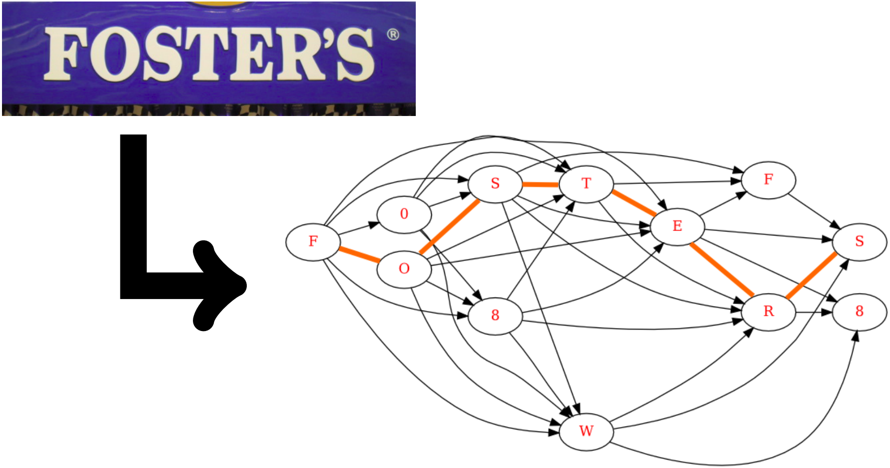
- a tool developed at Centre for Machine Perception at the Czech Technical University
- input: an image, output: rectangles with words and their transcriptions
- scores well in the ICDAR competition
- only text localization and rectangle transcription in the competition

CMP: TextSpotter

CMP: TextSpotter



CMP: TextSpotter



Outline

Training Data Preparation

- generate graphs from images → match with annotation

Training Data Preparation

- generate graphs from images → 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

in total 20 features

trigram features

- spaces ratio
- top line, bottom line, and cetral line angles
- character patterns
- trigram character language model

another 9 features

Independent learning

- all edges from all training graphs

Independent learning

- all edges from all training graphs
- on the correct path ⇒ positive examples
  others ⇒ negative examples

Independent learning

- all edges from all training graphs
- on the correct path ⇒ positive examples
  others ⇒ negative examples
- path maximizes sum of the socres from the classifies

Structured Learning

- problem needs to be defined in the following form:

$$\hat{\mathbf{y}} = \operatorname*{argmax}_{\mathbf{y} \in \mathcal{Y}_{\mathbf{x}}} \mathbf{w}^T \Psi(\mathbf{x}, \mathbf{y})$$

Structured Learning

- problem needs to be defined in the following form:

$$\hat{\mathbf{y}} = \underset{\mathbf{y} \in \mathcal{Y}_{\mathbf{x}}}{\operatorname{argmax}} \mathbf{w}^T \Psi(\mathbf{x}, \mathbf{y})$$

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

$$\Psi(\mathbf{x}, \mathbf{y}) = \sum_{e \in \mathbf{y}} \phi(e)$$

  - $\mathbf{w}$ ... weight vector

Structured Learning

- problem needs to be defined in the following form:

$$\hat{\mathbf{y}} = \operatorname*{argmax}_{\mathbf{y} \in \mathcal{Y}_{\mathbf{x}}} \mathbf{w}^T \Psi(\mathbf{x}, \mathbf{y})$$

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

$$\Psi(\mathbf{x}, \mathbf{y}) = \sum_{e \in \mathbf{y}} \phi(e)$$ ← we want to guess this

  - $\mathbf{w}$ ... weight vector ← we want to learn this

Structured Predicition

- Structured Percetpron
  - simple moficiation of the standard Perceptron algorithm
- Structured SVM
  - wieghts optimized by quadratic programming
  - not constant margin, but a loss function
  - exponential number of path in a graph ⇒ exponentially many inequalities for quadratic programming
  - approximative algorithm

Outline

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 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 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 Levensthein distance
  - average relative Levensthein distance
  - average length diffrence

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
  - average length diffrence
  - 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

		avg. edit dist.	avg. rel. edit dist.	avg. length diff.	full string acc.
bigram edges	baseline	.6471	.1317	.0336	.6933
	indep. class.	.3320	.0682	.0615	.8074
	S. Perceptron	.4631	.0917	.1352	.7377
	S. SVM	.4385	.0817	-.0041	.7500
	S. SVM + indep. cl.	.3770	.0798	.0574	**.8156**
trigram edges	indep. classs.	.3975	.0749	.0451	.7787
	S. Perceptron	.4877	.1035	.0902	.7008
	S. SVM	.4016	.0768	.1148	.7746
	S. SVM + indep. cl.	.3975	.0765	.0779	.7787

Extrinsic Measures

???

Outline

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.