A Machine Learning Approach to Hypohtesis Decoding in STR

Jindřich Libovický[1], Lukáš Neumann[2],
Pavel Pecina[1], Jiří Matas[2]

[1] Charles University in Prague
[2] Czech Technical University in Prague

First International Workshop on Robust Reading, 2/11/2014

Obsah

Outline

Outline

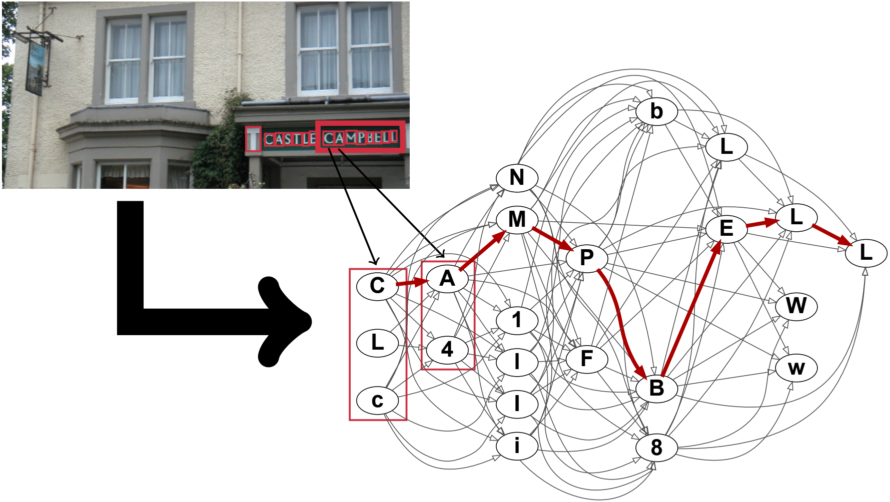
What is it?

Source: ICDAR Robust Reading Dataset

(200, 532) (358, 575)

FREEDOM

TextSpotter Pipeline



Outline

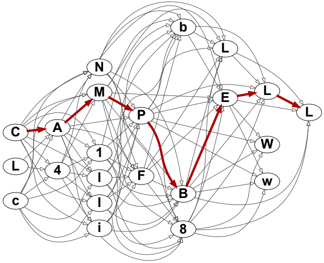
Problem Definition

- we define a hypothesis graph $G = (V, E, s, t, l, \phi)$
  - V, vertices $\sim$ recognition hypohteses with transrition lables $l : V \rightarrow$ `[a-zA-Z0-9]`
  - E, edges iff characters can follow each other
  - s, $t \in V$ – technical start and end
  - feature function $\phi : E \rightarrow \mathbb{R}^n$
- second order graph – nodes are character bigrams
- learned weight vector $\mathbf{w} \in \mathbb{R}^n$
- decoding string – maximum path from s to t w.r.t. $\mathbf{w}^T \cdot \phi(e)$

Preparing data

(511, 294) (621, 311) `CAMPBELL`

$\Longleftrightarrow$

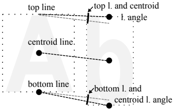


- ICDAR train set: 229 images, 849 words
- extracted 812 word graphs (not only whole words, but also prefixes of length $\geq$ 4)
- 70 % training, 30 % intrinsic evaluation

Features

Bigram features

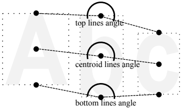
- angles of the top / bottom from the centroid line



- ratios of width, height, and area of the characters
- conditional character bigram probability
- patterns: two digits, `xX`, `xx|XX`, `Xx`
- OCR confidence

Trigram features

- adjacent top / bottom / centroid lines angles



- adjacent spaces ratio
- conditional character trigram probability
- patterns: digits only, `Xxx`, `xxx|XXX`, `xXX|xXx|xxX|XxX`

Learning Procedures

- independent classification
  - various classifiers from WEKA
  - train locally, decode globally
  - result of the independent classifier added a feature to other methods
- structured perceptron
- structured SVM

Outline

Introduction

Decoding Model

Evaluation

Conclusion

Intrinsic Evaluation

- the same type of data as in training (244 randomly picked graphs)
- measured quantities:
  - average edit distance from the ground truth string
  - average normalized edit distance from the ground truth string
  - accuracy – proportion of correctly decoded graphs

Intrinsic Evaluation Results

model	*order*	$\bar{d}$	$\bar{d}_r$	$\bar{a}$
TextSpotter	–	.647	.134	.685
Random Forest	1^{st}	**.332**	**.068**	.807
Structured Perceptron	1^{st}	.463	.092	.738
Structured SVM	1^{st}	.439	.082	.750
Str. Perc + Rand. for.	1^{st}	.377	.080	**.816**
Str. SVM + Rand. for.	1^{st}	.377	.080	**.816**
Random Forest	2^{nd}	.398	.075	.779
Structured Perceptron	2^{nd}	.488	.104	.701
Structured SVM	2^{nd}	.402	.077	.775
Str. Perc + Rand. for.	2^{nd}	.504	.101	.725
Str. SVM + Rand. for.	2^{nd}	.398	.077	.779

Extrinsic Evaluation

- tested on the ICDAR 2013 test set
- localization, character and word retreival

copy ceritre →
copy centre

CA lBOT PLACF →
CA BOT PLACE

TAKPOROS →
TANFORDS

Extrinsic Evaluation Results

model	*order*	loc.	char.	words
TextSpotter	–	**.715**	.696	.392
Random Forest	1^{st}	.673	.693	.376
Structured Perceptron	1^{st}	.710	.703	.389
Structured SVM	1^{st}	.673	.703	**.393**
Str. Perc + Rand. for.	1^{st}	.701	**.775**	.372
Str. SVM + Rand. for.	1^{st}	.701	**.775**	.372
Structured Perceptron	2^{nd}	.707	.704	.389
Structured SVM	2^{nd}	.680	.701	.385
Str. Perc. + Rand. for.	2^{nd}	.694	.705	.383
Str. SVM + Rand. for.	2^{nd}	.692	.709	.381

Conclusion

- better language modelling has a potential to improve STR performance
- only errors within words were fixed
  → joint word decoding and line splitting may help
- future goals: recognize longer segments of text for further processing

Any questions?

Thank you for your attention.