Charles University in Prague Institute of Formal and Applied Linguistics

Possible Approaches to Decoding Strings in Scene Text Recognition

Jindřich Libovický

ÚFAL Annual Seminar, Sedec–Prčice, September 16, 2014

Motivation

Scene Text Recognition

Learning the Decoding

Local features optimization

Global features optimization

Motivation

Scene Text Recognition

Learning the Decoding

Local features optimization

Global features optimization

Motivation

Why am I telling you this?

Why am I telling you this?

?

Scene Text Recognition

Motivation

Scene Text Recognition

Learning the Decoding

Local features optimization

Global features optimization

Scene Text Recognition

Motivation

Scene Text Recognition

Learning the Decoding

Local features optimization

Global features optimization

What is STR?

Scene Text Recognition



Source: ICDAR Robust Reading Dataset

What is STR?

Scene Text Recognition



(200, 532) (358, 575) FREEDOM

Source: ICDAR Robust Reading Dataset

Hypotheses Decoding

CASTLE CAMPBELL

Jindřich Libovický, Charles University in Prague, September 16, 2014 Possible Approaches to Decoding Strings in Scene Text <u>Recognition, 8/ 18</u>

Scene Text Recognition

Hypotheses Decoding

CASTLE CAMPBELL

Jindřich Libovický, Charles University in Prague, September 16, 2014 Possible Approaches to Decoding Strings in Scene Text <u>Recognition, 8/ 18</u>

Scene Text Recognition

Scene Text Recognition

Hypotheses Decoding



Scene Text Recognition

Hypotheses Decoding



Learning the Decoding

Motivation

Scene Text Recognition

Learning the Decoding

Local features optimization

Global features optimization

many features for the edges

- many features for the edges
- learned by Structured Perceptorn and Structured SVM

- many features for the edges
- learned by Structured Perceptorn and Structured SVM



TAKPOROS → TANFORDS

character retreival F1: $.696 \rightarrow .775$

- many features for the edges
- learned by Structured Perceptorn and Structured SVM



TAKPOROS → TANFORDS

character retreival F1: $.696 \rightarrow .775$

word retreival F1: $.392 \rightarrow .393$

Moral of the story



 in fact stacking of various local classifiers + the features they were trained from

Source: ...

Jindřich Libovický, Charles University in Prague, September 16, 2014 Possible Approaches to Decoding Strings in Scene Text Recognition, 11/ 18

Learning the Decoding

Moral of the story



Learning the Decoding

- in fact stacking of various local classifiers + the features they were trained from
- in general hard to include non-linearity to the decions

Source: ...

Moral of the story



 in fact stacking of various local classifiers + the features they were trained from

Learning the Decoding

- in general hard to include non-linearity to the decions
- missing global features

Source: ...

Local features optimization

Motivation

Scene Text Recognition

Learning the Decoding

Local features optimization

Global features optimization

CRFs minimize energy of a variable assignment



- CRFs minimize energy of a variable assignment
- binary variables is / is not in word



- CRFs minimize energy of a variable assignment
- binary variables is / is not in word
- minimum energy + "true" assignments must form a path



- CRFs minimize energy of a variable assignment
- binary variables is / is not in word
- minimum energy + "true" assignments must form a path
- can be encoded as linear programming task



Local features optimization Unsupervised feature learning

tons of unlabeled data

Local features optimization Unsupervised feature learning

- tons of unlabeled data
- RBM, autoencoders ...

Local features optimization Back-prob in online learning



edges from the decoded path

edges from the ground truth path

Global features optimization

Motivation

Scene Text Recognition

Learning the Decoding

Local features optimization

Global features optimization

Greedy Agent

Global features optimization



choses edges greedily

Global features optimization

Greedy Agent



- choses edges greedily
- can use the previous decoding to make prediction

Global features optimization

Greedy Agent



- choses edges greedily
- can use the previous decoding to make prediction
- generalized SEARN for learning



Thank you for your attention.