Word Beam Search: A Connectionist Temporal Classification Decoding Algorithm

Harald Scheidl, Stefan Fiel, Robert Sablatnig

FAKULTÄT FÜR !NFORMATIK
Faculty of Informatics
TU WIEN

Computer Vision Lab

Introduction

- Text encoded by paths in RNN output
  - Characters can be repeated: "ab" → "aaab"
  - And can be followed by blanks: "aaab" → "aaa-b"
- Word Beam Search (WBS) decodes RNN output

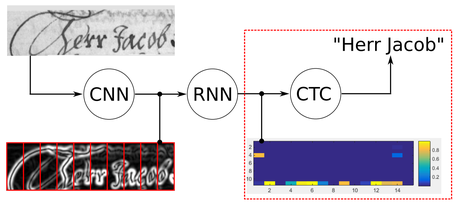

Fig.: CTC-trained neural network

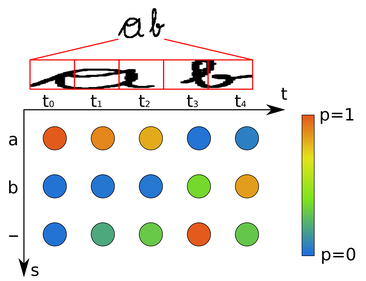

Fig.: RNN output containing character-probabilities

Related Algorithms

- Best path decoding
  - Collapse best-scoring path
- Token passing
  - Sequence of dictionary words, word-level LM
- Vanilla beam search (VBS)
  - Tree of beams, character-level LM

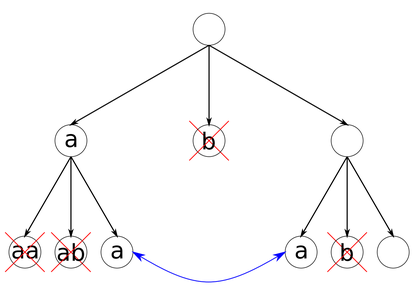

Fig.: Tree of beams

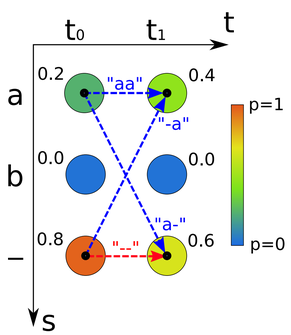

Fig.: RNN output with paths

Proposed Method

- Beam state
  - Inside word: constrain beam by prefix tree
  - Between words: allow arbitrary many non-word characters

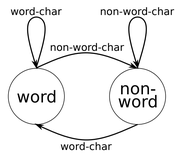

Fig.: Beam states

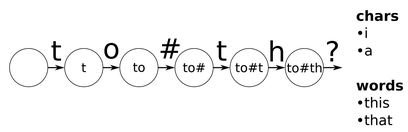

Fig.: Extending a beam in word-state

- Word-level LM with 4 possible scoring-modes
  - Only constrain words
  - N-gram score whenever beam finishes a word
  - N-gram lookahead: possible words given a prefix
  - N-gram sampled lookahead: subset of words

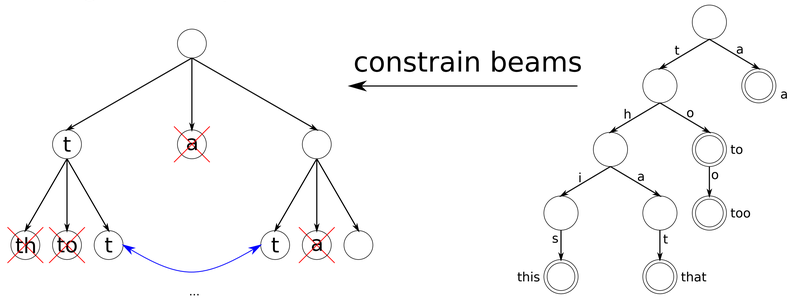

Fig.: Constrained tree of beams

Fig.: Prefix tree

Example

- Best path decoding: "ssla." → collapse → "sla." ✗
- Dictionary contains: "see", "sea", "as", "is", "else"
- Token passing: "sea" ✗
- VBS and WBS: beam width is 2, results see below

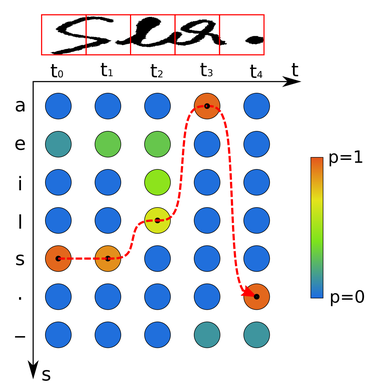

Fig.: RNN output with best path

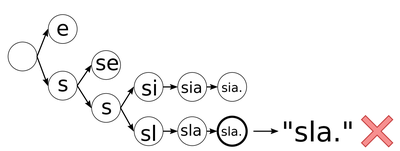

Fig.: VBS beams

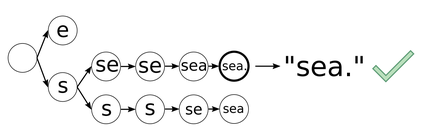

Fig.: WBS beams

Results

- Evaluation on IAM and Bentham HTR datasets
- Two different texts to train LM
  - Perfect: text of test-set (OOV-rate 0%)
  - Rudimentary: text of training-set + 370k word list

	Perfect LM	Rudimentary LM
Best Path	8.77 / 29.07	8.77 / 29.07
Token Passing	10.46 / 12.37	(not feasible)
VBS	8.27 / 27.34	**8.48** / 28.24
WBS W	5.62 / 11.01	8.95 / 24.19
WBS N	5.33 / **9.77**	10.00 / 23.88
WBS N+F	5.23 / 9.82	8.61 / **22.86**
WBS N+F+S	**5.21** / 9.78	8.62 / 22.91

Table: CER [%] / WER [%] on IAM dataset

	Perfect LM	Rudimentary LM
Best Path	5.60 / 17.06	5.60 / 17.06
Token Passing	8.16 / 9.24	(not feasible)
VBS	5.35 / 16.02	5.55 / 16.39
WBS W	4.22 / 7.90	**5.47** / 14.09
WBS N	4.07 / **7.08**	6.15 / **13.90**
WBS N+F	**4.05** / 7.36	6.76 / 18.00
WBS N+F+S	4.06 / 7.39	6.75 / 18.06

Table: CER [%] / WER [%] on Bentham dataset

- Dependence of running time on dictionary size
  - Token passing: only feasible for 2k words
  - Small dependence: VBS and WBS modes W and N
  - Large dependence: WBS modes N+F and N+F+S

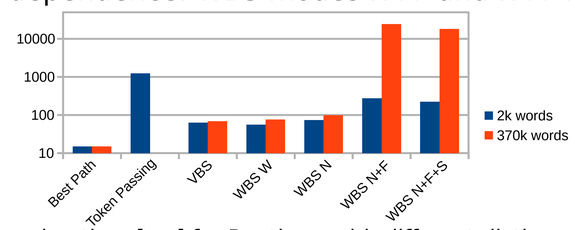

Fig.: Running time [ms] for Bentham with different dictionary sizes

Conclusion

- **Decodes** output of CTC-trained neural network
- **Words constrained** by dictionary
- Allows arbitrary number of **non-word characters between words**
- Optional **word-level LM**
- **Faster than token passing**

Code:
github.com/githubharald/CTCWordBeamSearch
Contact:
harald_scheidl@hotmail.com