RECURRENT NEURAL NETWORK LANGUAGE MODELING FOR CODE-SWITCHING CONVERSATIONAL SPEECH

Cognitive Systems Lab

Heike Adel, Ngoc Thang Vu
Franziska Kraus, Tim Schlippe, Haizhou Li, Tanja Schultz
heike.adel@student.kit.edu, thang.vu@kit.edu

Karlsruhe Institute of Technology

1. Overview

Goal

- Prediction of Code Switches based on textual features (words and POS tags)
- Extended structure of recurrent neural networks for Code-Switching

=> 10.8 % (2 %) relative improvement in terms of perplexity (WER) on the SEAME development set and 16.9 % (2.7 %) relative on the evaluation set

What is Code-Switching (CS)?

- Code-Switching speech is defined as speech that contains **more than one language**. It is a common phenomenon in multilingual communities.

2.1 The SEAME Corpus [D.C. Lyu et al., 2011]

SEAME = South East Asia Mandarin-English

- Conversational Mandarin-English Code-Switch speech corpus
- Temporarily provided as part of a joint research project by NTU and KIT
- About 63 hours of audio data and their transcriptions
- Four language categories: English, Mandarin, particles (Singapourean and Malaysian discourse particles) and others (other languages)
- Average number of CS per utterance: 2.6; very short monolingual segments **=> challenging bilingual task**

2.2 Code-Switching-Analyses of the Corpus

Prediction of Code-Switches

- Trigger words:

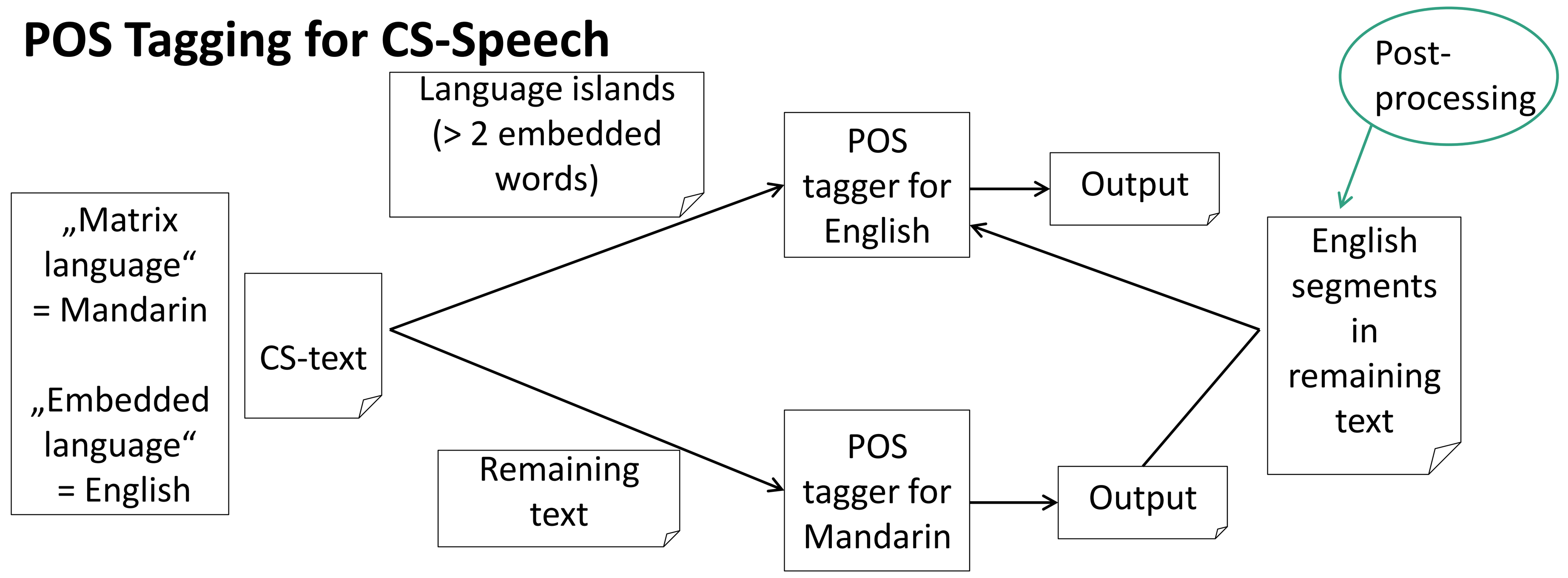
word	frequency	CS-rate
那个(that)	5261	53.43 %
我的(my)	1236	52.35 %
那些(those)	1329	49.44 %
一个(a)	2524	49.05 %
他的(his)	1024	47.75 %

Mandarin trigger words

word	frequency	CS-rate
then	6183	56.25 %
think	1103	37.62 %
but	2211	36.23 %
so	2218	35.80 %
okay	1044	34.87 %

English trigger words

POS Tagging for CS-Speech



- Trigger POS:

Tag	meaning	frequency	CS-rate
DT	determiner	11276	40.44 %
DEG	associative 的	4395	36.91 %
VC	是	6183	25.85 %
DEC	的 in a relative-clause	5763	23.86 %
M	measure word	2612	23.35 %

Mandarin trigger POS

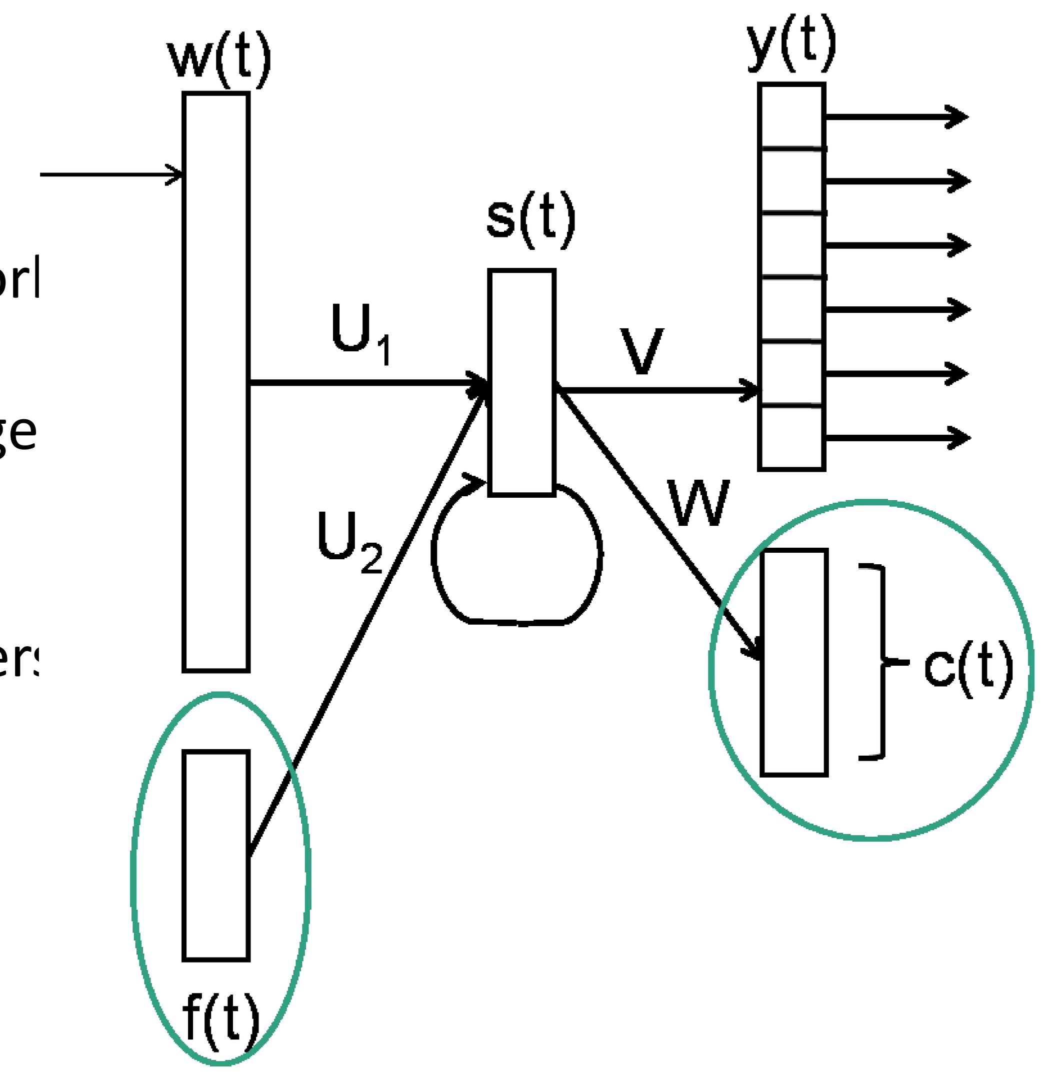
Tag	meaning	frequency	CS-rate
NN	noun	49060	49.07 %
NNS	noun (plural)	4613	40.82 %
RB	adverb	21096	31.84 %
JJ	adjective	10856	26.48 %
CC	coordinating conjunction	4400	24.05 %

English trigger POS

3. Recurrent Neural Network Language Model (RNNLM) for Code-Switching

- Input:
  - Word vector w(t)
  - Feature vector f(t) containing POS tags
- Hidden Layer: Vector s(t) containing the state of the network
- Output:
  - Vector c(t) with the probabilities for each language
  - Vector y(t) with probabilities for each word given its language
- U_1, U_2, V, W: weights for the connections between the layers

- Training with back-propagation through time (BPTT)
- Computation of the probabilities:
  $P(w_i \mid s(t)) = P(c_i \mid s(t)) \cdot P(w_i \mid c_i, s(t))$
- Reference to CS task: use words and **features** to not only determine the next word but also the **next language**



4. Experiments and Results

Perplexity Evaluation and Rescoring Experiments

- Rescoring of 100-best lists of our CS-ASR system [Vu, 2012] with different settings for language model weights (lz) and word insertion penalties (lp):
  $score = lz \cdot (\lambda \cdot score_{RNNLM} + (1-\lambda) \cdot score_{NGRAM}) + score_{AM} + lp \cdot |w|$
- RNNLM and the 3-gram LM of the ASR system are weighted equally (λ = 0.5)
- Performance Measure: **Mixed Error Rate** (MER): word error rates for English segments and character error rates for Mandarin segments

Model	PPL dev	PPL eval	MER dev	MER eval
3-gram	285.87	285.25	35.5 %	30.0 %
RNNLM	246.60	287.88	35.6 %	29.3 %
RNNLM + OF	239.64	269.71	34.9 %	29.4 %
RNNLM + FI	233.50	268.05	34.8 %	29.3 %
RNNLM + FI + OF	**219.85**	**239.21**	**34.7 %**	**29.2 %**

(OF: output factorization, FI: feature integration)