

Supplementary Material for Neural Chinese Address Parsing

Hao Li and Wei Lu

StatNLP Research Group
Singapore University of Technology and Design

hao_li@mymail.sutd.edu.sg

luwei@sutd.edu.sg

Pengjun Xie and Linlin Li

DAMO Academy
Alibaba Group

chengchen.xpj@alibaba-inc.com

linyan.111@alibaba-inc.com

Transition Features
N-gram Features
5-gram; 4-gram; 3-gram; bigram; unigram
Character Features
digit, Chinese digit, punctuation characters; URL; direction words; English characters
Segment Features
segment string; segment length; segment suffix

Table 1: Features used in two baselines ℓ CRF and s CRF.

1 Experimental Setup

We describe the discrete features used in two baselines ℓ CRF and s CRF respectively.

- ℓ CRF is the standard first-order linear CRF model (Lafferty et al., 2001) with discrete features for sequence labeling tasks. Such discrete features include transition features, n-gram features and character features mentioned in Table 1.
- s CRF is based on the standard semi-Markov CRF (Sarawagi and Cohen, 2004) with discrete features. Such discrete features include transition features, n-gram features, character features and segment features mentioned in Table 1.

References

- John Lafferty, Andrew McCallum, and Fernando CN Pereira. 2001. [Conditional random fields: Probabilistic models for segmenting and labeling sequence data](#). In *Proc. of ICML*.
- Sunita Sarawagi and William W Cohen. 2004. [Semi-markov conditional random fields for information extraction](#). In *Proc. of NIPS*.