Neural Network Based Bilingual Language Model Growing for Statistical Machine Translation

Rui Wang[†,*], Hai Zhao[†], Bao-Liang Lu[†], Masao Utiyama[‡], Eiichro Sumita[‡]

[†] Shanghai Jiao Tong University

[‡] National Institute of Information and Communications Technology

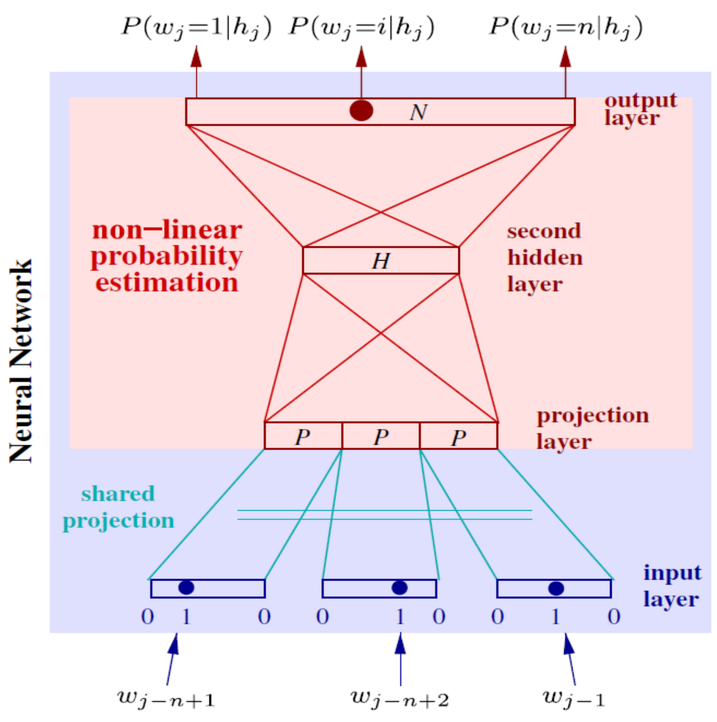
* Part of this work is finished when Rui Wang visited NICT

Introduction

Background: How to construct efficient large LM is an important topic in SMT. Most of the existing LM growing methods need an extra monolingual corpus, where additional LM adaption technology is necessary.

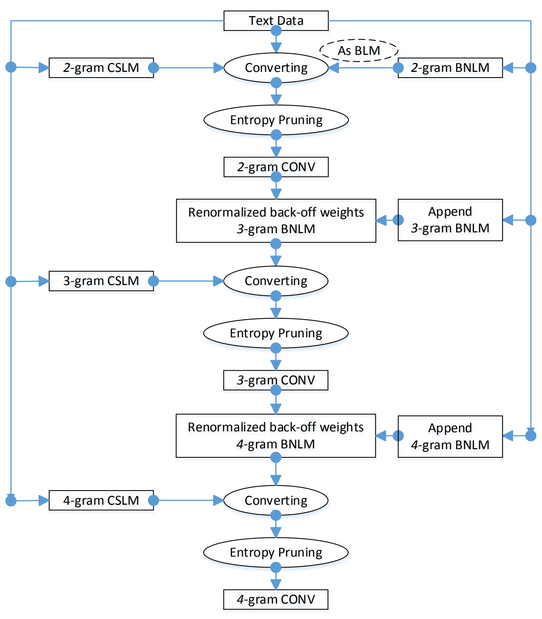
Main Contributions: We propose a novel neural network based bilingual LM growing method, only using the bilingual parallel corpus in SMT. The results show that our method can improve both the perplexity score for LM evaluation and BLEU score for SMT, and significantly outperforms the existing LM growing methods without extra corpus.

CSLM

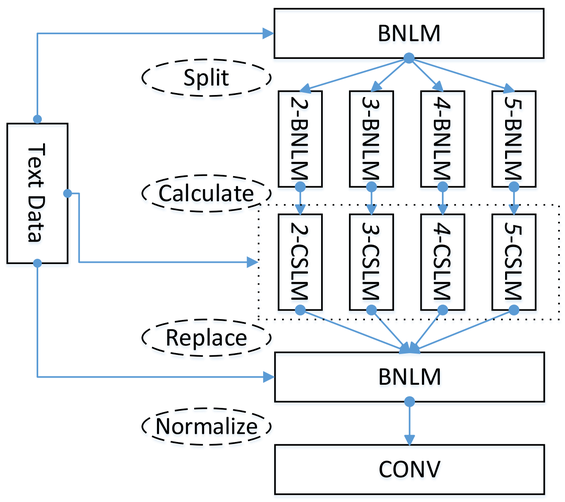


$$P(w_i|h_i) = \begin{cases} \frac{P_c(w_i|h_i)}{1-P_c(o|h_i)} P_s(h_i), & if\ w_i \in shortlist \\ P_b(w_i|h_i), & otherwise \end{cases}$$

Existing CSLM Converting Methods



Arsoy et. al.'s Method in ICASSP 2013



Wang et. al.'s Method in EMNLP 2013

Bilingual LM Growing

Connecting Phrases w_1^i if :

(1) w_1^k is the right (rear) part of one phrase βw_1^k in the phrase table, or

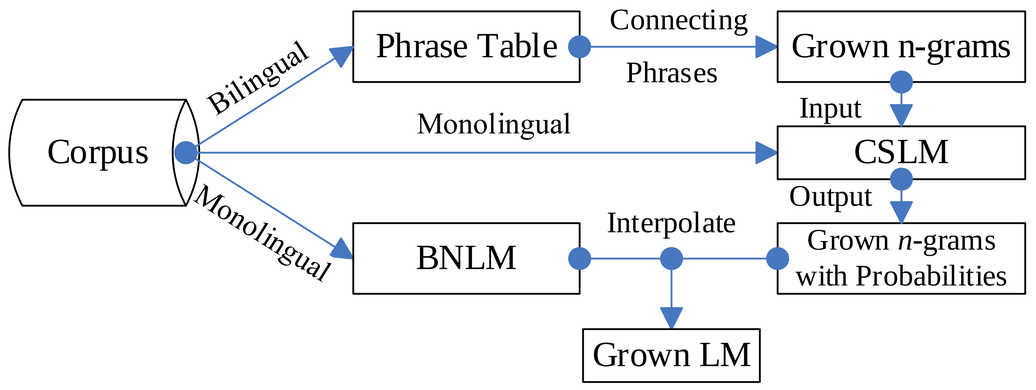
(2) w_{k+1}^i is the left (front) part of one phrase $w_{k+1}^i \gamma$ in the phrase table.

Ranking the Connecting Phrases (Grown *N*-grams):

$$P_{target}(e) = \sum_f P_{source}(f) \times P(e|f)$$

$$P_{connect}(w_1^k w_{k+1}^i) = \sum_{k=1}^{i-1} (\sum_\beta P_{target}(\beta w_1^k) \times \sum_\gamma P_{target}(w_{k+1}^i \gamma))$$

Calculating Probabilities of Grown *N*-grams Using CSLM:



Experiments and Results

Corpus:

(1) NTCIR-9: 1 million sentences from Chinese to English

(2) TED : 186K sentences from Chinese to English (additional monolingual corpus is hard to obtain)

SMT Results:

LMs	N-grams	PPL	BLEU
BNLM	73.9M	108.8	32.19
CSLM-RE	N/A	**97.5**	32.42
Wang2013	73.9M	104.4	32.60
Arsoy-1	217.6M	103.3	32.55
Arsoy-2	458.5M	103.0	32.39
Arsoy-3	712.2M	102.5	32.49
BI-1	223.5M	101.9	33.02+
BI-2	464.5M	100.6	**33.25++**
BI-3	705.5M	100.1	33.24++

NTCIR-9

LMs	N-grams	PPL	BLEU
BNLM	7.8M	87.1	12.41
Wang2013	7.8M	85.3	12.73
BI-1	23.1M	79.2	12.92
BI-2	49.7M	78.3	13.16
BI-3	73.4M	**77.6**	**13.24**

TED

Decoding Time on Test Data:

LMs	Decoding Time (sec.)
BNLM	15.3
CSLM	186.5
Bilingual Grown LM (BI-2)	16.5