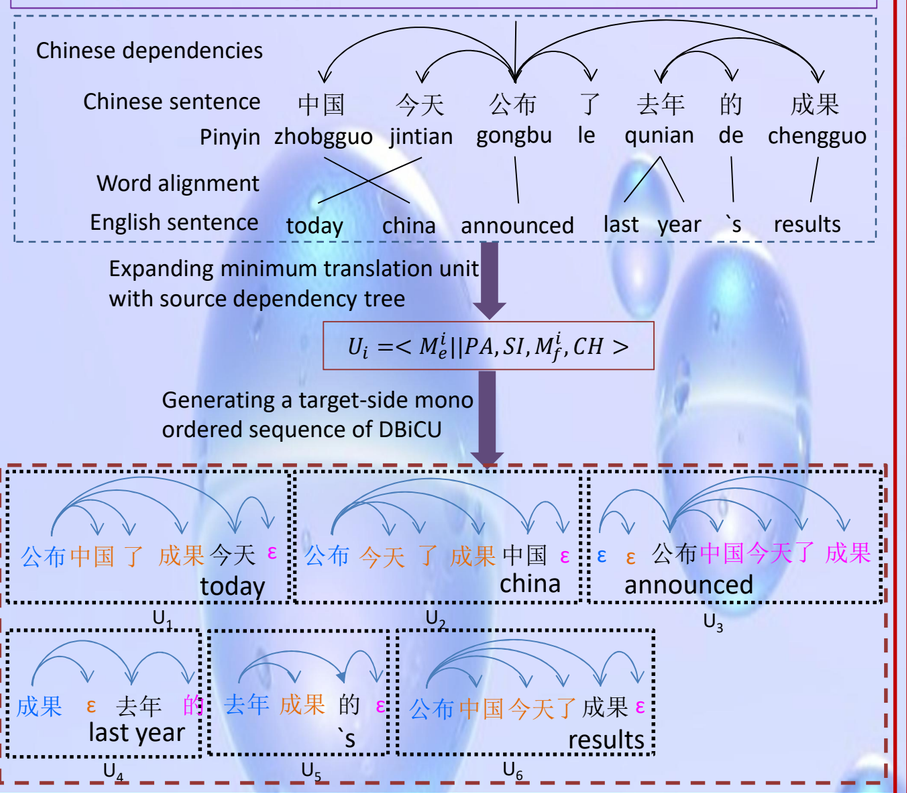
Translation Prediction with Source Dependency-Based Context Representation

Kehai Chen[1], Tiejun Zhao[1], Muyun Yang[1] and Lemao Liu[2]
[1]Machine Intelligence and Translation Laboratory, Harbin Institute of Technology, Harbin, China
[2]ASTREC, National Institute of Information and Communications Technology, Kyoto, Japan

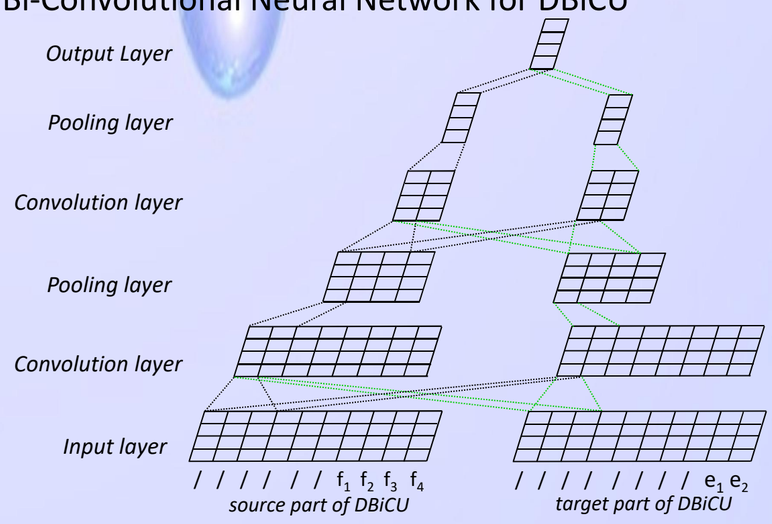
1. Dependency-Based Bilingual Compositional Sequence

- *Learning context representations is very promising to improve translation results, particularly through neural networks.*
- *Previous efforts process the context words sequentially and neglect their internal long-distance dependencies.*

Therefore, a novel neural network based on bi-convolutional neural network is proposed to represent the source-side dependency-based context for translation prediction.



2. Bi-Convolutional Neural Network for DBiCU



Input Layer: The input layer includes two matrices for a DBiCU U:

Us= {w1,...,wi}
Ut= {w1,...,wj}

Convolutional Layer: A convolutional layer in the network contains two filters $W_m \in R^{d \times k}$, and m = {0,1}, Let the filter window size be t (e.g., t=3), the filter W_m generates the feature y_k^m as follows:

$$y_k^m = \sigma(W_m([w_i + w_{i+1} + w_{i+2}] + [w_j + w_{j+1} + w_{j+2}]) + b)$$

Pooling Layer: For the output feature map of the convolution layers, column-wise max over windows of t=2 consecutive columns is performed:

$$p_i^m = \max[y_{2i-1}^m, y_{2i}^m]$$

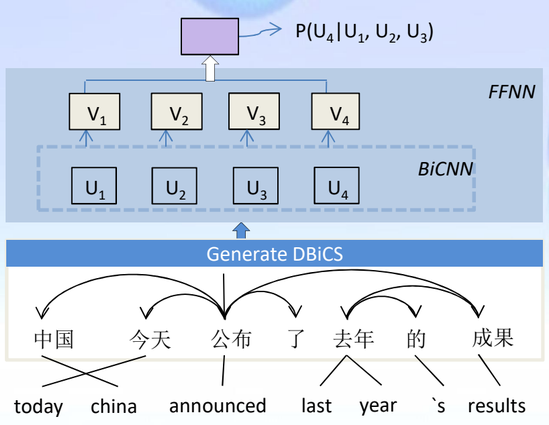
Output Layer: The output layer is typically a fully connected layer multiplied by a matrix, In the paper, row-wise averaging from pooling layers is performed:

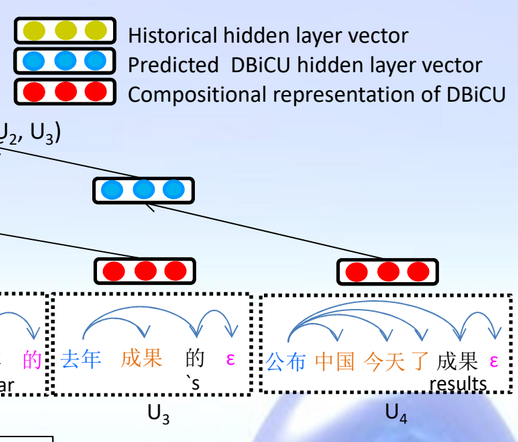
$$V = \text{average}(p_0, p_1)$$

Therefore, the above BiCNN plays a role of function φ parameterized by θ1, which maps a DBiCU U into V :

$$\boldsymbol{V = \varphi(U; \theta_1)}$$

3. Translation Prediction with DBiCS Representation





$$\prod_{i=4}^{l} P(U_i|U_{i-1}, U_{i-2}, U_{i-3}; \theta) = \prod_{i=4}^{l} \frac{\exp(\emptyset(V_{i-1}, V_{i-2}, V_{i-3}; \theta_2))}{Z(U_{i-1}, U_{i-2}, U_{i-3}; \theta)}$$

$$Z(U_{i-1}, U_{i-2}, U_{i-3}; \theta) = \sum_{\bar{U}} \exp(\emptyset(\bar{V}, V_{i-1}, V_{i-2}, V_{i-3}; \theta_2))$$

φ is a feedforward neural network parameterized by θ_2; and $\theta = \langle \theta_1, \theta_2 \rangle$ denotes all the model parameters including both of BiCNN and feedforward neural networks. Formally, we maximize the regularized log-likelihood on the training data:

$$\ell(\mathcal{U}; \theta) = \sum_{i=1}^{N} (logP(U_i|U_{i-1}, U_{i-2}, U_{i-3}; \theta)) - \alpha \log^2 Z(U_{i-1}, U_{i-2}, U_{i-3}; \theta)$$

$$Z(U_{i-1}, U_{i-2}, U_{i-3}; \theta) \approx \sum_{\bar{U} \in NB(U_i)} \exp(\emptyset(\bar{V}, V_{i-1}, V_{i-2}, V_{i-3}; \theta_2))$$

Where NB(U_i) denotes the neighborhood of a gold DBiCU, i.e. $U_i =< M_e^i || PA, SI, M_f^i, CH >$.

4.Experiments

Chinese-English NIST Results

System	MT03	MT04	MT05	AVG
basline	34.59	35.41	33.12	34.47
+BiLM	35.11+	35.79+	33.56+	34.8+
+OSM	35.24+	36.05+	33.83+	35.04+
+DBiLM	35.31+	35.75+	33.8+	34.95+
+DBiCSLM	35.53+*	36.71+*	34.14+*	35.28+*

Table 1 Effect of DBiCS

System	MT03	MT04	MT05	AVG
baseline	34.59	35.41	33.12	34.47
+NNJM	35.74+	36.79+	34.29+	35.60+
+DBiCSFFNN	35.56+	36.61+	33.92+	35.36+
+DBiCSNNLM	36.43+*	37.57+*	34.84+*	36.28+*
baseline+NNLM	35.13	36.22	33.58	34.97
+NNJM	35.82+	36.56+	34.50+	35.63+
+DBiCSFFNN	35.74+	36.86+	34.16+	35.59+
+DBiCSNNLM	36.76+*	37.97+*	35.21+*	36.64+*

Table 2 Effect of DBiCSNNLM

System	MT03	MT04	MT05	AVG
baseline(Dec)	34.59	35.41	33.12	34.47
SOUL	34.73+	35.96+	33.42+	34.70+
NNJM	35.02+	36.10+	33.72+	34.94+
JTRFFNN	34.81+	35.76+	33.26+	34.60+
MTURNN	35.10+	36.16+	33.89+	35.05+
DBiCSNNLM	35.21+*	36.40+*	33.77+	35.12+

Table 3 Effect on K-best Rescoring

Output sample sentences on the Chinese-English NIST Results

Example1: Translation Prediction on Ambiguous Words.

Ref: these dangerous people have seriously affected the normal immigration policy

Baseline: these dangerous elements seriously affected the normal immigration policy

NNJM : these dangerous elements have seriously affected the normal immigration policy

Src: 这些 危险 分子 严重 影响 了 正常 的 移民 政策
(pinyin) : *zhexie weixian fenzi yanzhong yingxiang le zhengchang de yimin zhengce*
This work: these dangerous people seriously affected the normal immigration policy

Example2: Translation Prediction on Word Forms

Ref: turkey is an important us ally in nato. it is now resisting pressures to join the us - led war against iraq

Baseline: turkey is a key nato ally of the united states , is now resisted pressure to join the us - led war plan against iraq

NNJM: turkey is a key nato ally of the united states , is to resist pressure to join the us - led war plan against iraq

Src:土耳其 是 美国 的 重要 北约 盟友 ，现 正 抗拒 压力 ，以 加入 美 领导 下 的 对 伊 作战
(pinyin): *tuerqi shi meiguo de zhongyao beiyue mengyou xian zheng kangju yali yi jiaru mei lingdao xia de dui yi zuozhang*
This work: turkey is a key nato ally of the united states , is resisting pressure to join the us - led war plan against iraq