

NTT SMT System for IWSLT 2008

Katsuhito Sudoh, Taro Watanabe, Jun Suzuki,
Hajime Tsukada, and Hideki Isozaki
NTT Communication Science Labs., Japan

Overview

- 2-stage translation system
 - k-best translation candidates are generated by **hierarchical phrase-based SMT**
 - The top-best candidate is chosen by a reranker based on **Ranking SVMs with large-scale sparse features**
- Evaluation on **Chinese-to-English challenge task**

Stage 1: Translation

- **Hiero** (Chiang, CL 2007) : in-house implementation
 - **Hierarchical phrase-based SMT**
 - CKY-based decoder
 - **Minimum Error Rate Training**
 - Decoder features are same as our IWSLT '06 system
 - Hierarchical and lexical translation probabilities
 - Insertion, deletion, and reordering penalties
 - Length penalties (words / hierarchical phrases)
 - Word 5-gram language model scores

Stage 2: Reranking

- Reorder k-best translation candidates after decoding
 - Ranking SVMs with large scale sparse features
 - Incorporate *context features*
 - Difficult to use in decoding (e.g. MIRA-based method)

Ranking SVMs (Joachims, 2002)

- Ranking samples (not classification)

- Trained using ordered k-best candidates e_1^*, \dots, e_k^*

- Metric: **Approximated BLEU**

- Converted to top-best vs. non-best pairwise difference pairs D

- $D = \{d_{ij} = e_i^* - e_j^* \mid e_i^* \in \langle \text{top - best} \rangle, e_j^* \in \langle \text{non - best} \rangle\}$

$$D' = \{d_{ij} = e_i^* - e_j^* \mid 1 \leq i < k, 1 < j \leq k, i < j\}$$

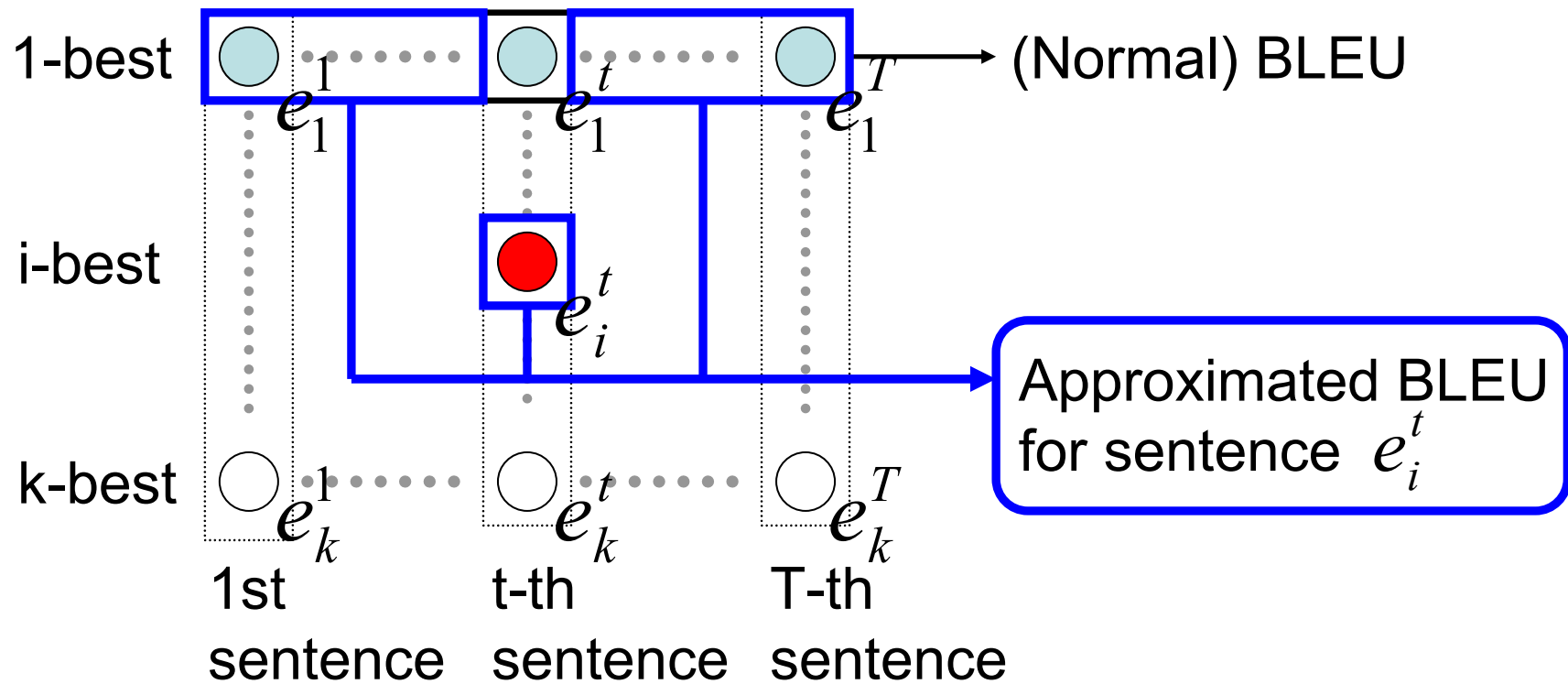
- Optimizing *classification* SVMs on D

- Test: choose highest-scored candidate

Approximated BLEU

- BLEU : document-wise score
 - **Requires re-computation in every iteration**
 - Not suitable for independently assigning scores to k-best candidates
- Approximated BLEU (Watanabe, IWSLT 2006)
 - Sentence-wise approximation of document-wise BLEU (*not sentence-wise BLEU*)
 - **Independently calculated for each candidate**
 - Constant throughout optimization

Approximated BLEU (cont'd)



Reranker Features

- Intra-sentence features
 - Word alignments
 - Source-target word pairs aligned by IBM Model 1
 - Target-source direction was also considered
 - Alignment bigram : $a(i)*a(i+1)$
 - Word pairs
 - Arbitrary source-target unigram/bigram pairs within each sentence
 - Target-side skip bigrams

Reranker Features (cont'd)

- Inter-sentence feature
 - Context-dependent word pairs
 - Arbitrary pair of [target word unigram] and [source/target word unigram **in the previous sentence**]

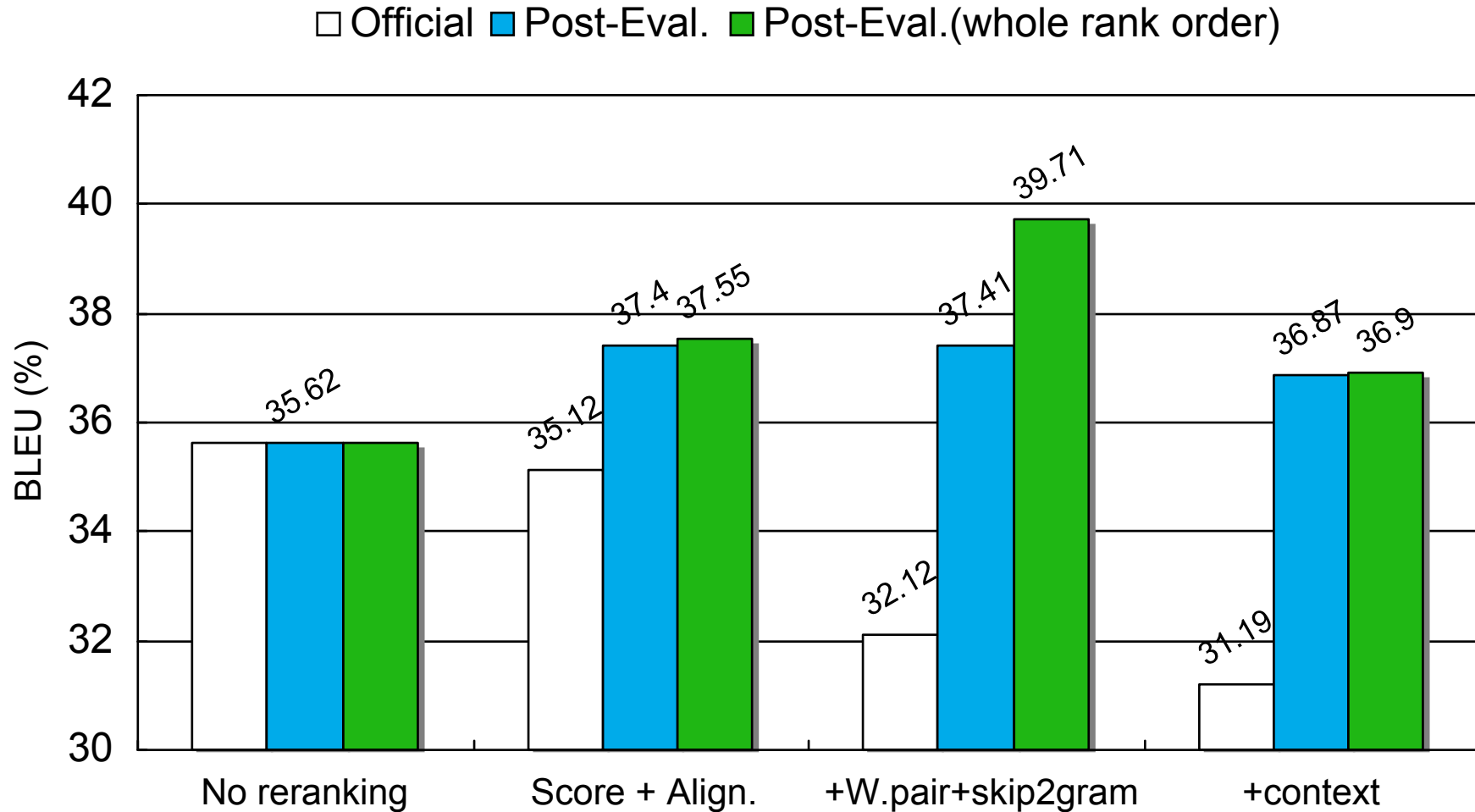
Pegasos

- Fast optimization algorithm for linear-kernel SVMs (Shalev-Shwartz et al., ICML 2007)
 - Use sub-gradients calculated based only on k samples in each iteration
 - Learning time does *not* depend on data size

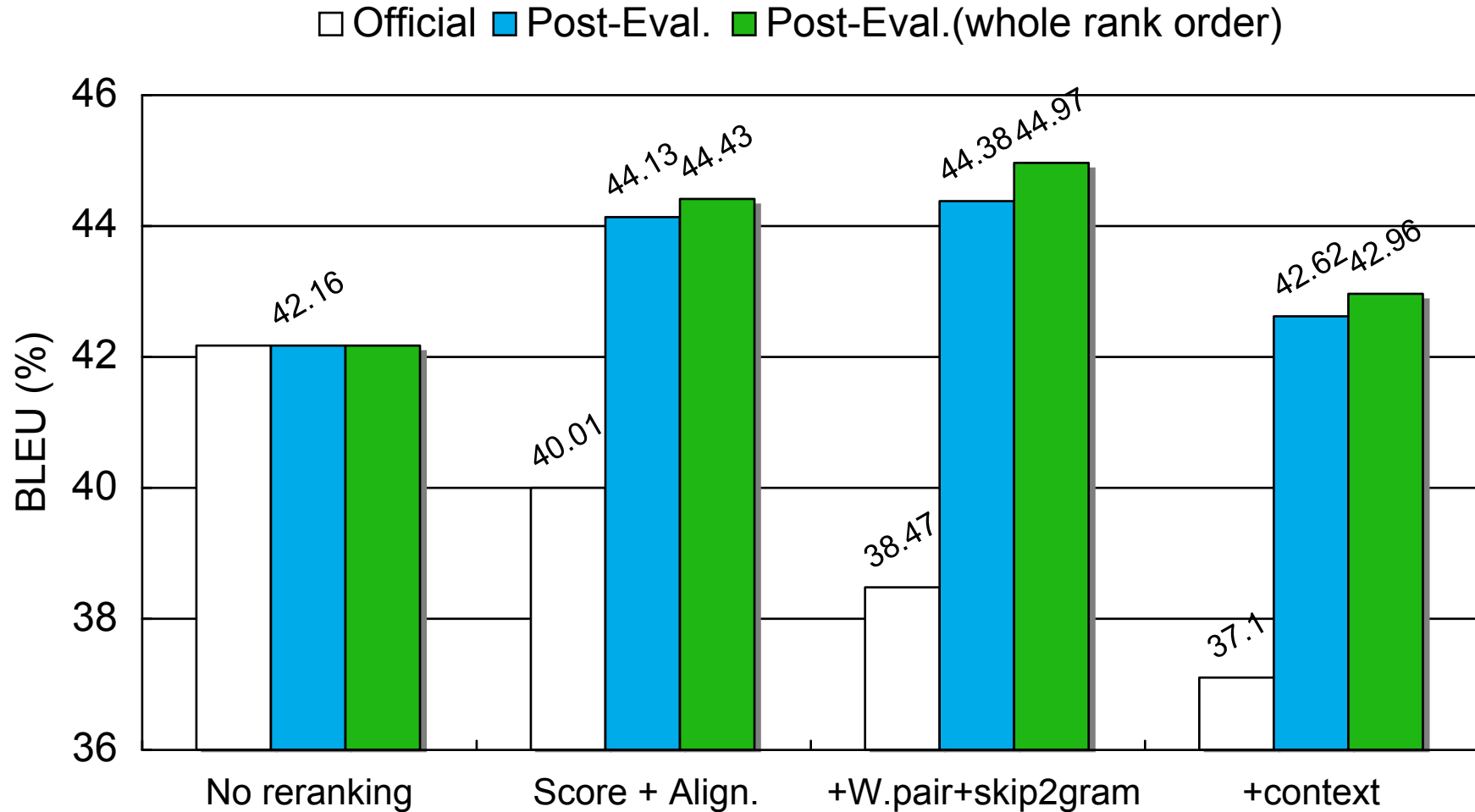
Post-evaluation

- Optimize SVM soft-margin parameter
 - 2-/3-fold cross validation on devset.CT_CE (246 sentences)
 - We didn't optimize it in the official evaluation!!
- Use the whole rank order in training R-SVMs
 - The whole rank order did not increase BLEU in our development phase

Results (ASR 1-best input)



Results (Clean input)



Results : Summary

- Reranking with *optimized soft-margin parameters* achieved good BLEU results
- Alignment-independent features were effective
- Context features were *not* effective

Discussion

- Reranker chose **adequate** candidates
 - Word alignment features captured *lexical correspondence*
- Reranker chose **fluent** candidates
 - (Skip-)Bigram features captured *target-side natural word order*
 - Bigram pair features captured *source-target co-occurrence* of bigrams
- Reranker failed to utilize context information
 - Context features turned out to capture many *general word co-occurrence*

Distinctive Word Alignment Features

ST: ?-<EOS> / 吗

TS: 吗-<EOS> / <\$.>

ST: 可以 / can

TS: ? / 吗

ST: tell-me / 请问

TS: 吗-<EOS> / ?

ST: i-would / 我-想

TS: 我-想 / i*like

ST: would-like / 想

TS: 在-哪里 / where

ST: you-have / 有

TS: 最近-的 / nearest^the

ST: <BOS>-i / 我

Distinctive Bigram Features

Bigram: ?-<EOS>

Bigram: .-<EOS>

Bigram: me-the

BigramPair: <BOS>-我 / <BOS>-i

BigramPair: <BOS>-我 / would-like

BigramPair: 吗-<EOS> / <BOS>-can

BigramPair :吗-<EOS> / ?-<EOS>

BigramPair: 多少-钱 / how-much

BigramPair: 多少-钱 / ?-<EOS>

BigramPair: <BOS>-能 / <BOS>-can

BigramPair: 给-我 / give-me

SkipBigram: would-*to

SkipBigram: <BOS>-*would

SkipBigram: <BOS>-*can

SkipBigram: do-*have

SkipBigram: tell-*the

Distinctive Context Features

TargetContext: for -> ?
TargetContext: is -> ?
TargetContext: a -> you
TargetContext: i -> you
TargetContext: . -> is
TargetContext: ? -> can
TargetContext: please -> ?
TargetContext: , -> can
SourceContext: 的 -> ?
SourceContext: 一 -> me
SourceContext: 吗 -> me
SourceContext: 我 -> .

Conclusion

- NTT's 2-stage SMT system
 - Hierarchical phrase-based SMT decoder
 - SVM-based reranker with sparse features
 - Achieved **39.71%(ASR), 44.97%(clean) BLEU** in Chinese-to-English challenge task
 - Reranker effectively utilized both monolingual and bilingual sparse features
 - Current context-dependent features are not effective