Integrating Multiple Knowledge Sources in Statistical Machine Translation

Sergio Penkale Andy Way

CNGL
DCU

Centre for Next Generation Localisation

April 22nd, 2009

Phrase-Based Statistical Machine Translation

Phrase-Based SMT systems are trained using these steps:

1. From a sentence-aligned parallel corpus, obtain an initial word alignment using statistical methods
2. Extend this word alignment using heuristics
3. All phrases *consistent* with the word alignment are extracted
4. These phrases are scored with several estimators
5. A phrase table is built

Phrase-Based Statistical Machine Translation (2)

Decoding

Given a source sentence f:

- Segment f into source phrases
- Search for target sentence e that maximises log-probability:

$$p(e|f) = exp(\sum_{i=1}^{n} \lambda_i h_i(e, f))$$

- Search is performed using phrases in phrase table

Marker-Based EBMT Phrase Extraction

Chunking

Given a set of *closed-class* words:

- Monolingually chunk each side of the corpus
- Create a new chunk for each marker word in the sentence
- Avoid marker-only chunks

Marker-Based EBMT Phrase Extraction (2)

Aligning

- Chunks are aligned using an edit-distance-style algorithm
- Translation probabilities
- Cognates information

Merging

- We have phrases obtained by two different methods
- We wish to merge them to create a new system
- How to best combine the knowledge sources?

We are interested in improving both:

- Translation quality
- Amount of additional phrases that get used

Merging (2)

Most common method:

1. Word-align
2. Extend word-alignment
3. Extract phrases
4. Add EBMT phrases to extracted phrases
5. Score phrases
6. Build phrase table

We propose some different merging methods

Merging (2)

Most common method:

1. Word-align
2. Extend word-alignment
3. Extract phrases
4. Add EBMT phrases to extracted phrases
5. Score phrases
6. Build phrase table

We propose some different merging methods

Merging (2)

Most common method:

1. Word-align
2. Extend word-alignment
3. Extract phrases
4. Add EBMT phrases to extracted phrases
5. Score phrases
6. Build phrase table

We propose some different merging methods

n-count

We merge n times.

1. Word-align
2. Extend word-alignment
3. Extract phrases
4. Add EBMT phrases to extracted phrases n times
5. Score phrases
6. Build phrase table

- Increase translation probability

n-count

We merge n times.

1. Word-align
2. Extend word-alignment
3. Extract phrases
4. Add EBMT phrases to extracted phrases n times
5. Score phrases
6. Build phrase table

- Increase translation probability

n-count

We merge n times.

1. Word-align
2. Extend word-alignment
3. Extract phrases
4. Add EBMT phrases to extracted phrases n times
5. Score phrases
6. Build phrase table

- Increase translation probability

n-count Results

Development test set:

System	BLEU	NIST	METEOR	EBMT%
no-EBMT	30.04	7.3780	59.61	23.75
1-count	30.19	7.4083	59.68	23.47
2-count	29.91	7.3541	59.38	23.53
4-count	29.96	7.3723	59.52	24.32
8-count	30.21	7.4101	59.41	26.41
16-count	29.71	7.3520	59.03	29.74

Results on testset follow the same pattern.

n-count Results

Development test set:

System	BLEU	NIST	METEOR	EBMT%
no-EBMT	30.04	7.3780	59.61	23.75
1-count	30.19	7.4083	59.68	23.47
2-count	29.91	7.3541	59.38	23.53
4-count	29.96	7.3723	59.52	24.32
8-count	30.21	7.4101	59.41	26.41
16-count	29.71	7.3520	59.03	29.74

Results on testset follow the same pattern.

n-count Results

Development test set:

System	BLEU	NIST	METEOR	EBMT%
no-EBMT	30.04	7.3780	59.61	23.75
1-count	30.19	7.4083	59.68	23.47
2-count	29.91	7.3541	59.38	23.53
4-count	29.96	7.3723	59.52	24.32
8-count	30.21	7.4101	59.41	26.41
16-count	29.71	7.3520	59.03	29.74

Results on testset follow the same pattern.

n-count Results

Development test set:

System	BLEU	NIST	METEOR	EBMT%
no-EBMT	30.04	7.3780	59.61	23.75
1-count	30.19	7.4083	59.68	23.47
2-count	29.91	7.3541	59.38	23.53
4-count	29.96	7.3723	59.52	24.32
8-count	30.21	7.4101	59.41	26.41
16-count	29.71	7.3520	59.03	29.74

Results on testset follow the same pattern.

n-count Results

Development test set:

System	BLEU	NIST	METEOR	EBMT%
no-EBMT	30.04	7.3780	59.61	23.75
1-count	30.19	7.4083	59.68	23.47
2-count	29.91	7.3541	59.38	23.53
4-count	29.96	7.3723	59.52	24.32
8-count	30.21	7.4101	59.41	26.41
16-count	29.71	7.3520	59.03	29.74

Results on testset follow the same pattern.

two-tables and *three-tables*

two-tables:

- Build a phrase table without adding EBMT phrases
- Build another phrase table using only EBMT phrases
- Decode using phrases from either table
- Each phrase table obtains a separate set of log-linear weights
- Phrases of each kind are always available, regardless of search beams

three-tables:

Phrases common to both methods are placed on a separate table.

two-tables and *three-tables* results

Development test set:

System	BLEU	NIST	METEOR	EBMT%
no-EBMT	30.04	7.3780	59.61	23.75
two-tables	30.09	7.3807	59.64	29.53
three-tables	29.97	7.3669	59.49	25.24

Testset:

System	BLEU	NIST	METEOR	EBMT%
no-EBMT	30.79	7.5590	60.25	24.21
two-tables	30.78	7.5496	60.26	29.46

phrase-penalty

- One of the model features is set equal for all phrases
- We increase the value of this feature for EBMT phrases

phrase-penalty Results

Development test set:

System	BLEU	NIST	METEOR	EBMT%
no-EBMT	30.04	7.3780	59.61	23.75
pp-1.5	30.22	7.4037	59.63	27.16
pp-2	30.31	7.4040	59.83	29.48
pp-2.5	29.85	7.3704	59.52	24.97

Testset:

System	BLEU	NIST	METEOR	EBMT%
no-EBMT	30.79	7.5590	60.25	24.21
pp-1.5	30.84	7.5776	60.22	27.25
pp-2	30.94	7.5805	60.32	29.55
pp-2.5	30.66	7.5481	60.17	25.26

phrase-penalty Results

Development test set:

System	BLEU	NIST	METEOR	EBMT%
no-EBMT	30.04	7.3780	59.61	23.75
pp-1.5	30.22	7.4037	59.63	27.16
pp-2	30.31	7.4040	59.83	29.48
pp-2.5	29.85	7.3704	59.52	24.97

Testset:

System	BLEU	NIST	METEOR	EBMT%
no-EBMT	30.79	7.5590	60.25	24.21
pp-1.5	30.84	7.5776	60.22	27.25
pp-2	30.94	7.5805	60.32	29.55
pp-2.5	30.66	7.5481	60.17	25.26

feature

- Merge as in *1-count*
- When scoring, add additional feature to the model
- Set feature to 1 if it's an EBMT phrase, 0 otherwise
- MERT will determine the appropriate weight of the feature

Devtest:

System	BLEU	NIST	METEOR	EBMT%
no-EBMT	30.04	7.3780	59.61	23.75
feature	30.11	7.3832	59.69	35.29

Testset:

System	BLEU	NIST	METEOR	EBMT%
no-EBMT	30.79	7.5590	60.25	24.21
feature	31.11	7.6004	60.55	35.09

feature

- Merge as in *1-count*
- When scoring, add additional feature to the model
- Set feature to 1 if it's an EBMT phrase, 0 otherwise
- MERT will determine the appropriate weight of the feature

Devtest:

System	BLEU	NIST	METEOR	EBMT%
no-EBMT	30.04	7.3780	59.61	23.75
feature	30.11	7.3832	59.69	35.29

Testset:

System	BLEU	NIST	METEOR	EBMT%
no-EBMT	30.79	7.5590	60.25	24.21
feature	31.11	7.6004	60.55	35.09

Conclusions

- The most commonly used method 1-count doesn't allow us to fully exploit the additional resources
- The feature method is easy to implement and allows for better translation quality and increased usage of the additional phrases

Thank You

Results Devtest

Development test set

System	BLEU	NIST	METEOR	EBMT%
no-EBMT	30.04	7.3780	59.61	23.75
1-count	30.19	7.4083	59.68	23.47
2-count	29.91	7.3541	59.38	23.53
4-count	29.96	7.3723	59.52	24.32
8-count	30.21	7.4101	59.41	26.41
16-count	29.71	7.3520	59.03	29.74
two-tables	30.09	7.3807	59.64	29.53
three-tables	29.97	7.3669	59.49	25.24
pp-1.5	30.22	7.4037	59.63	27.16
pp-2	30.31	7.4040	59.83	29.48
pp-2.5	29.85	7.3704	59.52	24.97
feature	30.11	7.3832	59.69	35.29

Results Testset

Test set

System	BLEU	NIST	METEOR	EBMT%
no-EBMT	30.79	7.5590	60.25	24.21
1-count	30.78	7.5775	60.24	23.47
2-count	30.76	7.5582	60.20	23.64
4-count	30.71	7.5609	60.15	24.34
8-count	30.83	7.5969	60.18	26.64
16-count	30.42	7.5386	59.86	29.71
two-tables	30.78	7.5496	60.26	29.46
pp-1.5	30.84	7.5776	60.22	27.25
pp-2	30.94	7.5805	60.32	29.55
pp-2.5	30.66	7.5481	60.17	25.26
feature	31.11	7.6004	60.55	35.09