Improving alignments for better confusion networks for combining machine translation systems

Author：Ayan, Necip Fazil and Zheng, Jing and Wang, Wen

Professor：陳嘉平

Reporter：陳逸昌

Outline

- Introduction
- Related work
- System combination with confusion networks
- Confusion networks with word synonyms and two-pass alignment
- Evaluation and results
- conclusions

Introduction(1/2)

- Various techniques in hypothesis combination from different systems using
  - Sentence level scores
  - Re-decoding source sentences using phrases
  - Word-based combination techniques using confusion networks

- In this paper, they used the confusion network decoding.

Introduction(2/2)

- One of the crucial steps in confusion network decoding is the alignment of hypotheses.

- Extends the alignment approaches by
  - Handle synonyms
  - Two-pass alignment strategy

Related work

- Sentence-level combination
- Phrase-level combination
- Word-level combination

Sentence-level combination

- Choosing one hypothesis among multiple MT system outputs.

- The selection criterion can be a combination of translation model and language model scores with multiple comparison test.

Phrase-level combination

- The input systems provide some internal information such as phrases used by the system.
- Extracting a new phrase translation table from each system’s target-to source phrase alignments and re-decoding the source sentence using this new translation table and a language model.

Word-level combination

- Word-level combination chooses the best translation units from different translations and combine them.
- After reordering hypotheses and aligning to each other, the combination system builds a confusion network and chooses the path with the highest score.

System combination with confusion networks

- Step
  1. Pick a skeleton hypothesis
  2. Reorder all the other hypotheses by aligning them to the skeleton translation
  3. Build a confusion network
  4. Decode the confusion network
  5. Optimize feature weights on development set and re-decode
- Two important choices:
  - Selection of the skeleton hypothesis
  - Alignment of other hypotheses to the skeleton

Selection of the skeleton hypothesis

- Choosing the hypothesis with Minimum Bayes Risk when it is used as the reference against the other hypotheses.

- Choosing the best hypotheses from each system and using each of those as a skeleton in multiple confusion networks.

Alignment of other hypotheses to the skeleton

- Multiple string-matching algorithm based on Levenshtein edit distance. (Bangalore et al., 2001)
- A heuristic-based matching algorithm. (Jayaraman and Lavie, 2005)
- Using GIZA++ with possibly additional training data. (Matusov et al.,2006)
- Using TER between the skeleton and a given hypothesis. (Sim et al., 2007;Rosti et al., 2007)

Confusion networks with word synonyms (1/2)

- When building a confusion network, the goal is to put the same words on the same arcs as much as possible.
- Their goal is to create equivalence classes and modify the alignment algorithm to give priority to the matching of words that are in the same equivalence classes.

Confusion networks with word synonyms (2/2)

- Use WordNet to extract synonyms of the words that appear in all hypotheses
- Augment each skeleton word with all synonymous words that appear in all the hypotheses.
- Modify TER script to handle words with alternatives using an additional synonym matching operation.

Extracting synonyms from WordNet

- Allow matching words that have the same stem or variations of the same word with different part-of-the-speech (POS) tags.

- For a example, it is clear that the verbs “wait” and “expect” have the same meaning but TER is unable to align these two words to each other if there are different word positions.

Extracting synonyms from WordNet

- Expect:
  - Sense 1:expect, anticipate => evaluate, pass judgment, judge
  - Sense 2:ask, require, expect => demand
  - Sense 3:expect, look, await, wait
  - Sense 4:expect => see, consider, reckon, view, regard
  - Sense 5:expect => expect, look, await, wait
  - Sense 6:have a bun in the oven, bear, carry, gestate, expect => give birth, deliver, bear, birth, have
- Wait:
  - Sense 1:wait
  - Sense 2:wait, hold off, hold back => act, move
  - Sense 3:expect, look, await, wait
  - Sense 4:wait, waitress => work

ycchen@tiger: /home/ycchen [125x45]

Connection Edit View Window Option Help

```
4 senses of wait

Sense 1
wait

Sense 2
wait, hold off, hold back
       => act, move

Sense 3
expect, look, await, wait

Sense 4
wait, waitress
       => work
ycchen@tiger:~$ wn expect -synsv

Synonyms/Hypernyms (Ordered by Estimated Frequency) of verb expect

6 senses of expect

Sense 1
expect, anticipate
       => evaluate, pass judgment, judge

Sense 2
ask, require, expect
       => demand

Sense 3
expect, look, await, wait

Sense 4
expect
       => see, consider, reckon, view, regard

Sense 5
expect
       => expect, look, await, wait

Sense 6
have a bun in the oven, bear, carry, gestate, expect
       => give birth, deliver, bear, birth, have
ycchen@tiger:~$
```

Augmenting references

- Avoid over-generation of synonyms, they make the assumption that words w_i and w_j are synonyms of each other only if w_i appears in the synonym list of w_j ,and w_i appears in the synonym list of w_j.

Modifications to TER script

- They modified the publicly available TER script, TERCOM, to match words in the same equivalence class at an additional synonym cost.
- Used a slightly higher cost for synonym matching, a cost of 0.1.
- All the other operations have a cost 1.0.

Extended skeleton

- Skeleton: at the same time expect israel to abide by the deadlines set by .
- Hypothesis: in the meantime , we are waiting for israel to abide by the established deadlines .
- Extended skeleton: at the same time expect_waiting israel to abide by the deadlines set_established by .

Alignment

```
ONE-PASS ALIGNMENT
khaddam                                 receives                      italian         house  speaker
khaddam                                 receives the head of the      italian chamber of     deputies
khaddam will meet with the president of the                           italian chamber of     deputies
```

- One of the major problems occurs when the hypotheses include additional words that do not appear in the skeleton translation.
- To overcome this issue, they employ a two-pass alignment strategy.

Two-pass alignment strategy

- First pass, build a confusion network, and an intermediate reference sentence is created.
  - When create the intermediate reference, the number of positions for a given word is bounded by the maximum number of occurrences of the same word in any hypothesis.
- Second pass uses this intermediate reference as the skeleton to generate the final confusion network.

WITHOUT SYNONYM MATCHING and ONE-PASS ALIGNMENT:

at	the		same	time					expect	israel	to	abide	by
at	the		same	time	we				expect	israel	to	abide	by
at	the		same	time	,	we	are	waiting	for	israel	to	abide	by
at	the		same	time	we				expect	israel	to	abide	by
at	the		same	time	,	we			expect	israel	to	abide	by
at	the		same	time	,	waiting			for	israel	to	comply	with
in	the	meantime	,	waiting					for	israel	to	abide	by

WITH SYNONYM MATCHING and TWO-PASS ALIGNMENT:

at	the	same	time				expect		israel	to	abide	by
at	the	same	time		we		expect		israel	to	abide	by
at	the	same	time	,	we	are	waiting	for	israel	to	abide	by
at	the	same	time		we		expect		israel	to	abide	by
at	the	same	time	,	we		expect		israel	to	abide	by
at	the	same	time	,			waiting	for	israel	to	comply	with
in	the		meantime	,			waiting	for	israel	to	abide	by

Features

- Word score
  - Uniform weighting:1/n
  - Rank-based weighting1/(1+r)
  - TM-based weighting
- NULL-word (or epsilon) insertion score
- Word penalty
- Language model score
  - 4-gram LM nearly 3.6 billion words using the SRILM toolkit.

Evaluation and results

- Combination on three system outputs.
  - the major difference between the three systems is they were trained on different subsets
- Arabic-English
- Chinese-English

Data

Data for Training/Tuning/Testing	Arabic-English		Chinese-English	
	# of segments	# of tokens	# of segments	# of tokens
Training Data (System1)	14.8M	170M	9.1M	207M
Training Data (System2)	618K	8.1M	13.4M	199M
Training Data (System3)	2.4M	27.5M	13.9M	208M
Tuning Set (Input Systems)	1800	51K	1800	51K
Tuning Set (System Combination)	1259	37K	1785	55K
Test Set - NIST MTEval'05	1056	32K	1082	32K
Test Set - NIST MTEval'06	1797	45K	1664	41K
Test Set - NIST MTEval'08	1360	43K	1357	34K

Results

System	MT'05	MT'06	MT'08
System 1	53.4	43.8	43.2
System 2	53.9	46.0	42.8
System 3	56.1	45.3	43.3
No Syns, 1-pass	56.7	47.5	44.9
w/Syns, 2-pass	57.9	48.4	46.2

Table 2: Lowercase BLEU scores (in percentages) on Arabic NIST MTEval test sets.

System	MT'05	MT'06	MT'08
System 1	35.8	34.3	27.6
System 2	35.9	34.2	27.8
System 3	36.0	34.3	27.8
No Syns, 1-pass	38.1	36.5	27.9
w/Syns, 2-pass	38.6	37.0	28.3
No Syns, 1-pass, tuning set w/webtext			28.4
w/Syns, 2-pass, tuning set w/webtext			29.3

Table 3: Lowercase BLEU scores (in percentages) on Chinese NIST MTEval test sets.

- Arabic-English
  - Improvements of 1.8 to 2.9BLEU point on three test sets over the best input system.
- Chinese-English
  - Improvements of 0.5 to 2.7BLEU point on three test sets over the best input system.

Results

Synon.	2-pass	MT'05	MT'06	MT'08
No	No	56.7	47.5	44.9
Yes	No	57.3	47.8	45.2
No	Yes	57.7	48.0	45.9
Yes	Yes	57.9	48.4	46.2

Table 4: Comparison of Synonym Matching and Two-pass Alignment on Arabic-English

- synonym matching on its own yields improvements of 0.3-0.6 BLEU points.
- Two-pass alignment turns out to be more useful than synonym matching.

Future work

- Includes a more effective use of existing linguistic resources to handle alignment of one word to multiple words.

- Planning to extend word lattices to include phrases from the individual systems for more grammatical outputs.