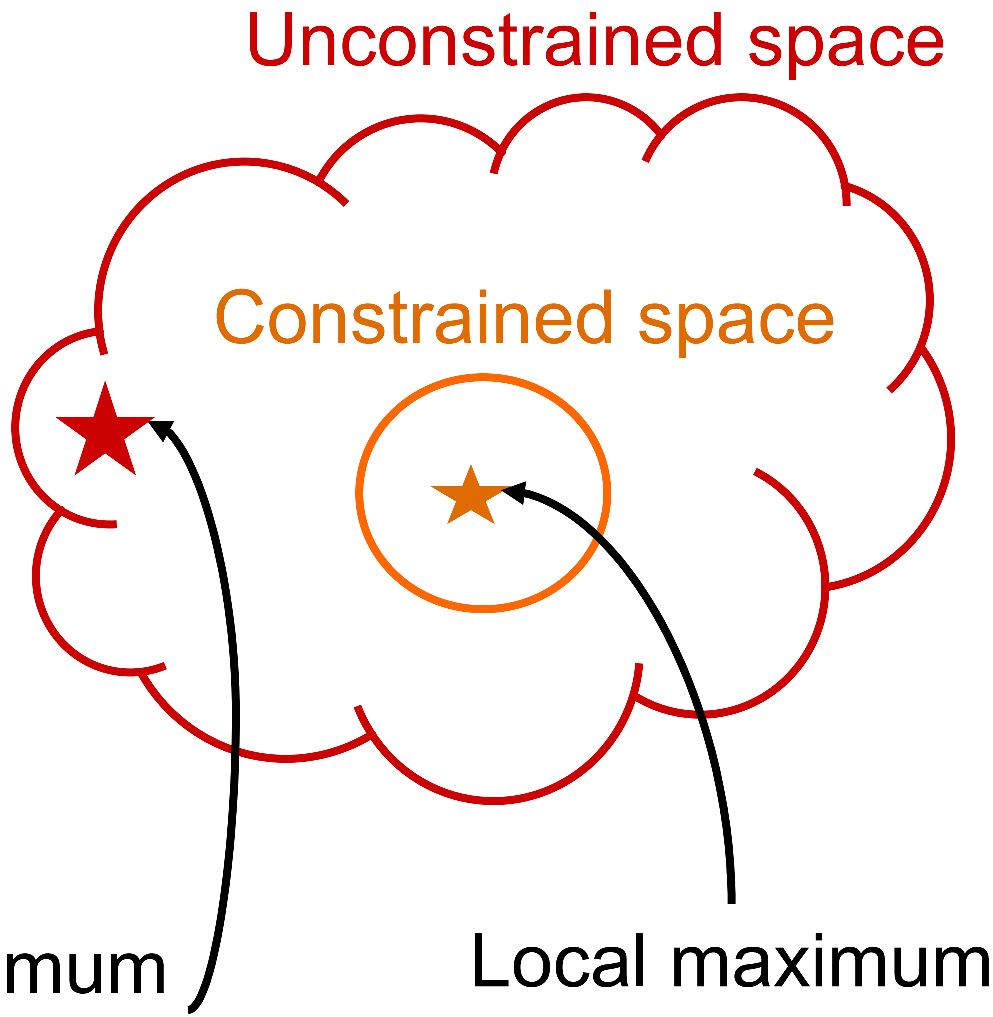
Reordering Constraints for Phrase-based MT

Richard Zens, Hermann Ney, Taro Watanabe, and Eiichiro Sumita

Karim Filali Oct 27, 2004

Problem

- Search problem in Statistical MT is NP-hard…
- unless we constrain possible re-orderings



Review: Statistical MT

- Source channel approach
  - Source: e_1^I Target: f_1^J

$$\hat{e}_1^I = argmax_{e_1^I}\{Pr(e_1^I)Pr(f_1^J|e_1^I)\}$$

- Alternative: model Pr(E|F) directly

$$\hat{e}_1^I = argmax_{e_1^I}\{\sum_{m=1}^{M} \lambda_m h_m(e^I, f^J)\}$$

Phrase-based MT /Translation Templates

- Train translation model in the two directions f -> e and e -> f
- For each direction, calculate Viterbi alignment
- Create intersection alignment matrix from the above alignments
- Determine word classes
- Estimate word translation parameters and translation template probabilities
- For search, first find best segmentation of source sentence into alignment templates

Log-linear Model: Feature Functions Used

- Phrase translation model
- Word translation model
- Tri-gram LM
- Class-based five-gram LM
- Word penalty
- Alignment template penalty

DP Algorithm using IBM Constraints

$$Q(1, \emptyset, \$) = 1$$

$$Q(j, S, e) = \max\{$$

$$\max_{e', \tilde{e}}\{$$

$$\max_{j-M \leq j' < j} Q(j', S, e')p(f_{j'}^{j-1}|\tilde{e})p(\tilde{e}|e'),$$

$$\max_{(j',l) \in S', S = S' \setminus (j',l)} Q(j, S', e')p(f_{j'}^{j'+l-1}|\tilde{e})p(\tilde{e}|e')\},$$

$$\max_{j-M \leq j' < j, S' : S = S \cup (j', j-j') \wedge |S'| < k} Q(j', S', e)\}$$

$$Q(J+2, \emptyset, \$) = \max_e Q(J+1, \emptyset, e)p(\$|e)$$

ITG constraints for Phrase-based MT

- Inversion Transduction Grammars
  - A -> [AA] | <AA> | f/e | f/ε | ε/e
  - [ ] : concatenation
  - < > : inverted concatenation
- Example: [Leusch’02]
  - “We will meet at noon in the lobby”
  - “We will meet in the lobby at twelve o’clock’
  - [we/we will/will meet/meet < [at/at noon/twelve ε/o’clock] [in/in the/the lobby/lobby] >]

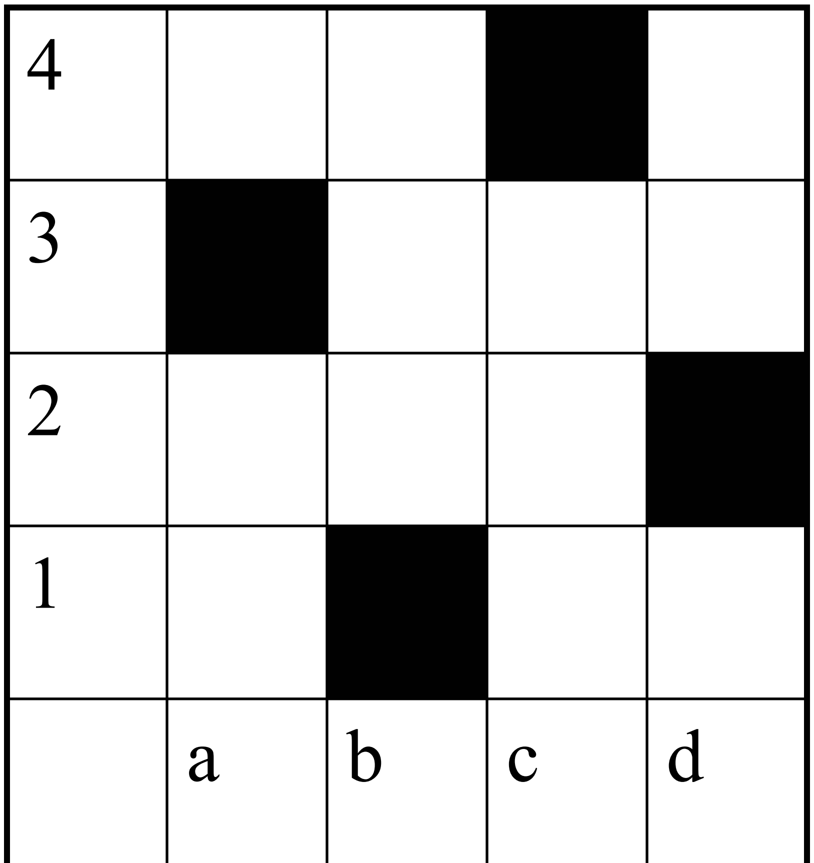
DP algorithm using ITG Constraints

$$Q_{j_l,j_r,e_b,e_t} = \max_{j_l \leq k < j_r, e', e''} \{ Q^0_{j_l,j_r,e_b,e_t},$$

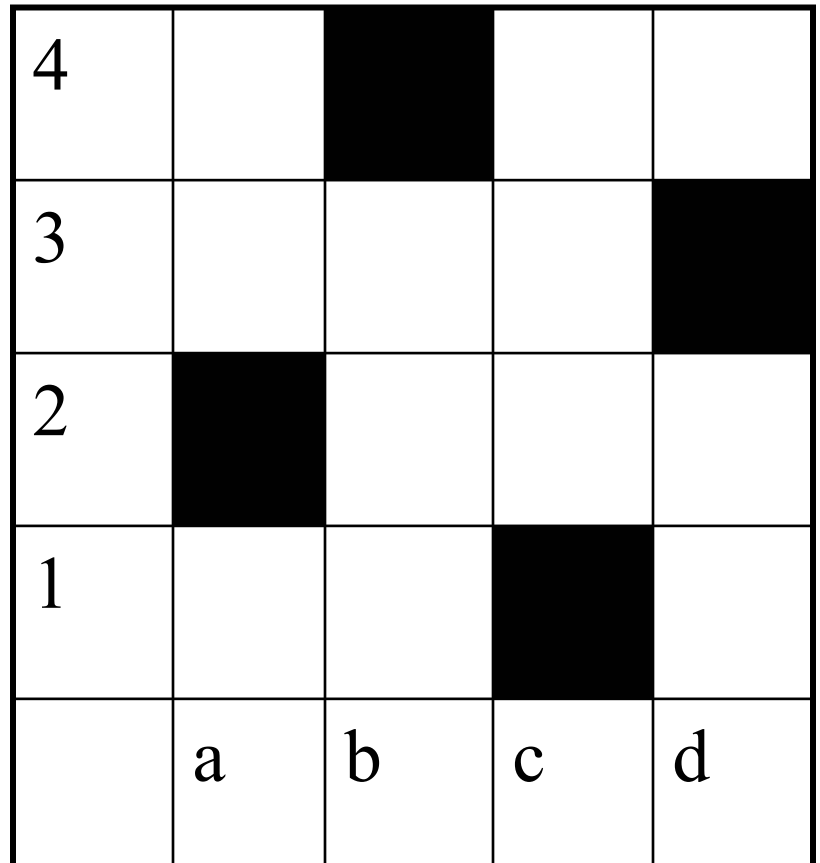
$$Q_{j_l,k,e_b,e'} Q_{k+1,j_r,e'',e_t} p(e''|e') p_m,$$

$$Q_{k+1,j_r,e_b,e'} Q_{j_l,k,e'',e_t} p(e''|e') (1 - p_m) \}$$

Forbidden Reordering Patterns



(1)



(2)

Beam Search Algorithm

- Start with beam search decoder for unconstrained algorithm
- Let
  - jc: current source sentence position
  - jn: candidate source sentence position
  - cov[j]: jth position has been translated
- Add constraints
  - If jn < jc for all jn < j < jc: cov[j] -> cov[j+1]
  - If jc < jn for all jc < j < jn: cov[j] -> cov[j-1] (cannot move from uncovered position to covered one)
- We can prove
  - above constraints equivalent to ITG constraints, and
  - we can generate all reorderings that don't violate ITG contraints

Experiments

Setup

- Japanese-English tasks
  - Basic Travel Expression Corpus (BTEC): travel phrasebook entries
  - Spoken Language Database (SLDB): hotel reservation transcriptions
- Evaluation Criteria:
  - WER
  - PER
  - BLEU
  - NIST

BTEC Corpus Statistics

		Japanese	English
train	sentences	152 K	
	words	1044 K	893K
	vocab	17047	12020
dev	sentences	500	
	words	3361	2858
test	sentences	510	
	words	3498	-

SLDB Corpus Statistics

		Japanese	English
train	sentences	15 K	
	words	201 K	190 K
	vocab	4757	3663
test	sentences	330	
	words	3940	-

BTEC WER Results: Short vs. Long Sentences

	% WER			
	sentence length			time (ms)
reorder	short	long	all	
mon	11.4	16.6	12.7	73
skip 1	10.8	13.5	11.4	134
2	10.8	13.4	11.4	169
free	10.8	13.8	11.5	194
ITG	**10.6**	**12.2**	**11.0**	164

SLDB WER Results: Short vs. Long Sentences

	% WER			
	sentence length			time (ms)
reorder	short	long	all	
mon	32.0	52.6	48.1	911
skip 1	31.9	51.1	46.9	3175
2	32.0	51.4	47.2	4549
free	32.0	51.4	47.2	4993
ITG	**31.8**	**50.9**	**46.7**	4472

BTEC Results (510 sents): All Metrics

	error rates (%)		accuracy measures	
reorder	WER	PER	BLEU (%)	NIST
mon	12.7	10.6	86.8	14.14
skip 1	11.4	10.1	88.0	14.19
2	11.4	10.1	88.1	14.20
free	11.5	10.0	88.0	14.19
ITG	**11.0**	**9.9**	**88.2**	**14.25**

SLDB Results (330 sents): All Metrics

	error rates (%)		accuracy measures	
reorder	WER	PER	BLEU (%)	NIST
mon	48.1	35.5	54.4	9.45
skip 1	46.9	35.0	56.8	9.71
2	47.2	35.1	**57.1**	9.74
free	47.2	34.9	**57.1**	9.75
ITG	**46.7**	**34.6**	**57.1**	**9.76**

Conclusions

- ITG constraints best compared to
  - Unconstrained search (*)
    - Best in theory but 1) search space too large and 2) re-ordering probabilities not reliably estimated
    - (*) Difference doesn't seem significant for BLEU/NIST scores
  - IBM constraints
    - Performance equivalent to unconstrained case for skips > 2
  - Monotone search
    - Works reasonably well for short sentences (<10 words)

Related Work

- Zens et al. “A comparative study on reordering constraints in SMT” ‘03
  - Word-based translation
  - Better BLEU % score using IBM constraints (42.5 vs 37.1)
  - Slightly better WER using ITG constraints (45.6 vs 46.2)
  - Extended ITG constraints: better alignment coverage
- Leusch et al. “A novel string-to-string distance measure with applications to MT evaluation” ‘02
- Doi and Sumita “Splitting Input Sentence for MT using LM with sentence similarity”