Joint Prediction of Word Alignment with Alignment Types

Anahita Mansouri Bigvand, Te Bu, Anoop Sarkar
Simon Fraser University, 8888 University Dr., Burnaby, BC, Canada
{amansour,tbu,anoop}@sfu.ca

Contributions

- We provide a new probabilistic model for word alignment where word alignments are associated with linguistically motivated alignment types.
- We propose a novel task of joint prediction of word alignment and alignment types and propose novel semi-supervised learning algorithms for this task.
- Our generative models with alignment types significantly outperform the models without alignment types.

Proposed Methods

- Baselines:
  - IBM Model 1 (Brown et al., 1993)
  - HMM-based word alignment model (Vogel et al., 1996)
- Generative HMM with alignment types (HMM+Type+Gen):

$$Pr(\mathbf{f}, \mathbf{a}, \mathbf{h}|\mathbf{e}) = \prod_{j=1}^{J} p(a_j|a_{j-1}, I)p(f_j|e_{a_j})p(h_j|f_j, e_{a_j})$$

  - EM training:

$$Pr(a_j = i, h_j = h|\mathbf{f}, \theta) = Pr(a_j = i|\mathbf{f}, \theta) \times Pr(h_j = h|a_j = i, \mathbf{f}, \theta)$$

$$\gamma_i(j, h) = \underbrace{\gamma_i(j)}_{\text{HMM posterior}} \times \underbrace{p(h|f_j, e_i)}_{\text{alignment type parameter}}$$

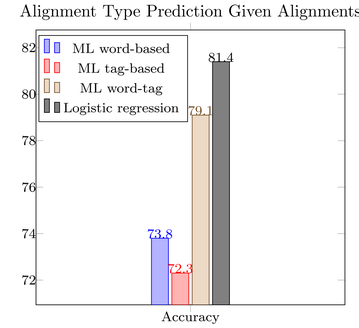
  - Decoding: Find the best word alignment and alignment types

$$V_i(j, h) = \max_{i', h'} \{V_{i'}(j-1, h')p(i|i', I)p(f_j|e_i)p(h|f_j, e_i)\}$$

- Discriminative HMM with alignment types (HMM+Type+Disc):
  - $p(h|f, e)$ are computed using a logistic regression classifier with 22 different features including lexical, POS tags, etc.
  - Decoding is similar to the decoding of HMM+Type+Gen.

Data Set

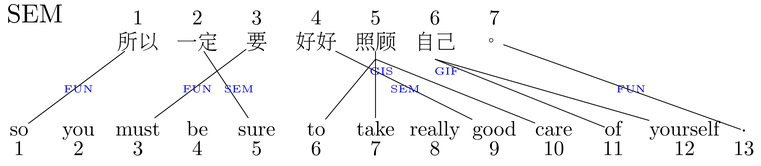
- GALE Chinese-English Word Alignment and Tagging Corpus released by LDC
  - Catalog numbers: LDC2012T16, LDC2012T20, LDC2012T24, LDC2013T05, LDC2013T23 and LDC2014T25
  - 22K sentences annotated with gold alignment and alignment types (20K sentences for training and 2K sentences for test)



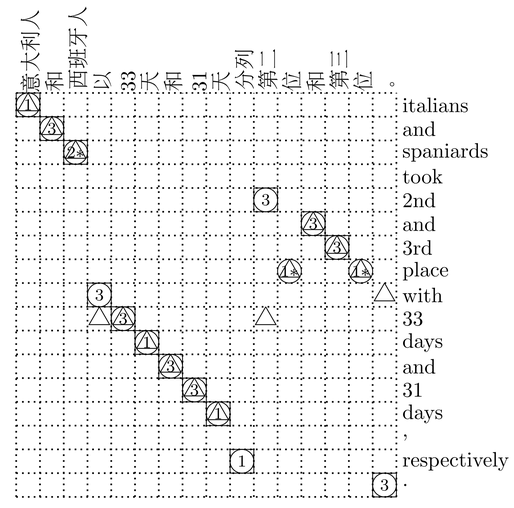
ID	Alignment Type	Meaning	Count
1	SEM	Semantic	159,277
2	GIS	Grammatically Inferred Semantic	81,235
3	FUN	Function	97,727
4	GIF	Grammatically Inferred Function	12,314
5	PDE	DE-Possessive	1,421
6	COI	Contextually Inferred	3,256
7	CDE	DE-Clause	1,608
8	TIN	Translated Incorrectly	1,116
9	MDE	DE-Modifier	4,615
10	NTR	Not Translated	34,090
11	MTA	Meta word	84

Word Alignment with Alignment types

- Alignment function $a : j \rightarrow i$
- Tagging function $h : j \rightarrow k$
- Example:
  $a_2 = 5$ and $h_2 = \text{SEM}$



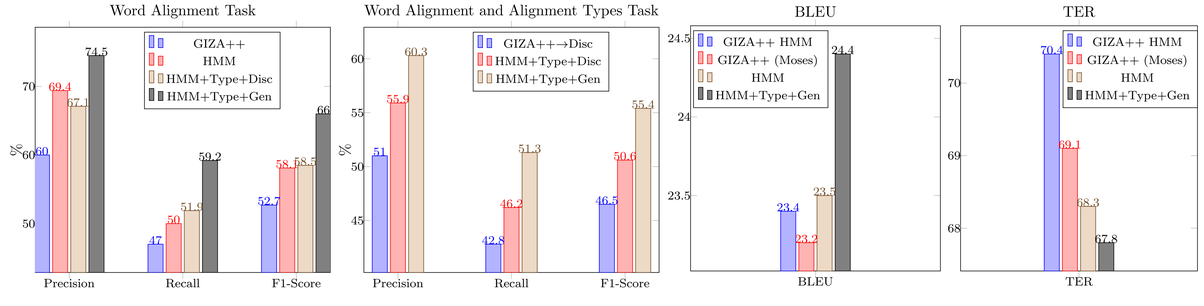
Example



○: HMM+Type+Gen, △: HMM □: Gold alignment

Experiments

- Training data: 20K sentences from LDC data (annotated with gold alignment and alignment types) + 1M sentences from Hong Kong (HK) Parliament proceedings
- Tasks: (1) Word alignment (2) Joint prediction of word alignment and alignment types
- Test data for these two tasks: 2K held-out sentences from LDC data
- Test data for machine translation experiments: 919 sentences of MTC part 4 (LDC2006T04)
- We built a baseline HMM similar the one proposed by Och and Ney (2003).
- Alignment type parameters of the HMM+Type+Gen model are initialized based on the maximum likelihood estimate of the 20K LDC data.



References

[1] Peter F Brown, Vincent J Della Pietra, Stephen A Della Pietra, and Robert L Mercer. 1993. The mathematics of statistical machine translation: Parameter estimation. Computational linguistics, 19(2):263–311.

[2] Stephan Vogel, Hermann Ney, and Christoph Tillmann. 1996. Hmm-based word alignment in statistical translation. In Proceedings of the 16th conference on Computational linguistics-Volume 2, pages 836–841. Association for Computational Linguistics.

[3] Franz Josef Och and Hermann Ney. 2003. A systematic comparison of various statistical alignment models. Computational linguistics, 29(1):19–51.