



The MIT-LL/AFRL System for IWSLT-2010

Wade Shen
HLT Group, MIT/Lincoln Laboratory
swade@ll.mit.edu

Tim Anderson and Ray Slyh
Air Force Research Laboratory
{timothy.anderson, raymond.slyh}@wpafb.af.mil

A. Ryan Aminzadeh
Department of Defense
ryan.aminzadeh@ugov.gov



Introduction

The MIT-LL/AFRL System is a Phrase-based MT system with a number of modifications for IWSLT.

- Multiple word alignments (CLA, Berkeley, GIZA)
- Rescoring models (IBM Model1, class-based LMs)
- Multiple optimization strategies (MERT, Simplex, MIRA)

Highlights for 2010

- CoMMA: MTT
- Improved Domain Adaptation
- MIRA Optimization
- System Combination

MIRA Optimization

- Implementation draws from [6, 5, 7]
- Online, iterative update of weights w_i :

$$w_i = w_{i-1} + \alpha * (\mathbf{h}(f, \hat{e}) - \mathbf{h}(f, e))$$

where \hat{e} denotes the oracle translation for a source sentence f , $\mathbf{h}(f, e)$ is a vector of model scores corresponding to the translation of f into e , and α is an update scaling parameter defined as follows:

$$\alpha = \max(0, \min(C, \frac{\mathcal{L}(\hat{e}, e) - (s^{i-1}(f, \hat{e}) - s^{i-1}(f, e))}{\|\mathbf{h}(f, \hat{e}) - \mathbf{h}(f, e)\|})$$

$$s^{i-1}(f, e) = w_{i-1} \cdot \mathbf{h}(f, e)$$

$\mathcal{L}(\hat{e}, e)$ defines a loss function (in our case, the BLEU score difference between the oracle translation, \hat{e} , and the current best translation, e . C is a limiter on the update scaling.

- Results:

System	Opt. Method	dev2
TALK PT + TALK LM	MERT	24.90
TALK PT + TALK LM	MIRA	25.27
TALK PT + TALK LM + Giga LM	MERT	25.91
TALK PT + TALK LM + Giga LM	MIRA	25.76
TALK PT + Cont. TALK LM + Giga LM	MERT	26.15
TALK PT + Cont. TALK LM + Giga LM	MIRA	25.87
(TALK + WMT) PT + TALK LM + Giga LM	MERT	23.91
(TALK + WMT) PT + TALK LM + Giga LM	MIRA	24.43

Improved Domain Adaptation

- Last year: MAP for phrase probability adaption

$$\hat{p}(s|t) = \lambda p_{iwslt}(s|t) + (1 - \lambda) p_{gp}(s|t)$$

$$\lambda = \frac{N_{iwslt}(s, t)}{N_{iwslt}(s, t) + \tau}$$

where p_{gp} and p_{iwslt} are phrase probability estimates from the general purpose and IWSLT-domain models respectively.

$$\lambda = \frac{N_{iwslt}(s, t)}{N_{iwslt}(s, t) + \tau}$$

- Modified MAP, uses *relative* counts of gp and iwslt models

$$\lambda = \frac{N_{iwslt}(s, t)}{N_{iwslt}(s, t) + N_{gp}(s, t) + \tau}$$

- Results

System	Arabic (dev7)	French (dev3)
IWSLT Model Only (baseline)	55.31	65.51
IWSLT MAP-adapted ([1])	58.85	67.39
IWSLT MAP-adapted (modified)	59.75	68.27

CoMMA-MTT

- Count Mediated Morphological Analysis + MultiThreshold Training
- CoMMA: Extends AP5 processing [2]
- CoMMA: Algorithm:

```

foreach w in |W|
  Apply AP5 diacritic normalization procedure
  if count(w) < t
    Apply AP5 segmentation of clitics, etc.
  else don't segment

```

- MTT: Apply CoMMA at diff. thresholds, combine data, train
- Experiments (Arabic dev6)

CoMMA Threshold	Mean BLEU	
	CoMMA	CoMMA-MTT
0	50.40	51.55
20	53.67	54.44
200	53.88	54.51
2,000	52.44	54.20
10,000	53.06	54.54

System Combination

- Same Algorithm as last year [4]
- Increased diversity of systems
 - MIRA optimized
 - variants of CoMMA-MTT systems (by threshold)
 - MBR Rescored systems
- Results:

Arabic-to-English Systems		
System	Features	BLEU
AE-primary 2009	2009 baseline	57.17
AE-primary	2010 combined system	58.69
AE-contrast2	2010 best individual system (baseline)	56.58
Turkish-to-English Systems		
System	Features	BLEU
TE-primary 2009	2009 baseline	60.01
TE-primary	2010 combined system (without CoMMA)	60.21
TE-contrast1	2010 combined system	60.78
TE-contrast4	2010 best individual system (MIRA)	58.85
French-to-English Systems		
System	Features	BLEU
FE-primary	2010 combined system	63.62
FE-contrast2	2010 best individual system (MBR)	63.22
TALK Task Systems		
System	Features	BLEU
TALK-primary	2010 combined system	26.50
TALK-contrast3	2010 best individual system	26.12

References

- [1] Shen, W., Delaney, B., Aminzadeh, A.R., Anderson, T., and Slyh, R. "The MIT-LL/AFRL IWSLT-2009 MT System," In Proc. Of the International Workshop on Spoken Language Translation, Tokyo, Japan, 2009.
- [2] Shen, W., Delaney, B., Anderson, T., and Slyh, R. "The MIT-LL/AFRL IWSLT-2007 MT System," In Proc. Of the International Workshop on Spoken Language Translation, Trento, Italy, 2007.
- [3] Shen, W., Delaney, B., and Anderson, T. "The MIT-LL/AFRL IWSLT-2006 MT System," In Proc. Of the International Workshop on Spoken Language Translation, Kyoto, Japan, 2006.
- [4] Rosti, A.V.I. and Matsoukas, S. and Schwartz, R., "Improved Word-Level System Combination for Machine Translation," In Proc. of ACL, 2006.
- [5] T. Watanabe, J. Suzuki, H. Tsukada, and H. Isozaki "Online large-margin training for statistical machine translation," In Proc. of EMNLP-CoNLL, 2007.
- [6] D. Chiang Y. Marton, and P. Resnik, "Online large-margin training of syntactic and structural translation features," In Proc of EMNLP, 2008.
- [7] D. Chiang, K. Knight, W. Wang, "11,001 new features for statistical machine translation," In Proc. NAACL/HLT, 2009.

