RWTH Aachen Machine Translation System: {Arabic, Chinese, German}-English MT Track

Stephan Peitz, Markus Freitag, Saab Mansour,
Minwei Feng, Joern Wuebker
surname@cs.rwth-aachen.de

IWSLT 2012, Hongkong
December 6, 2012

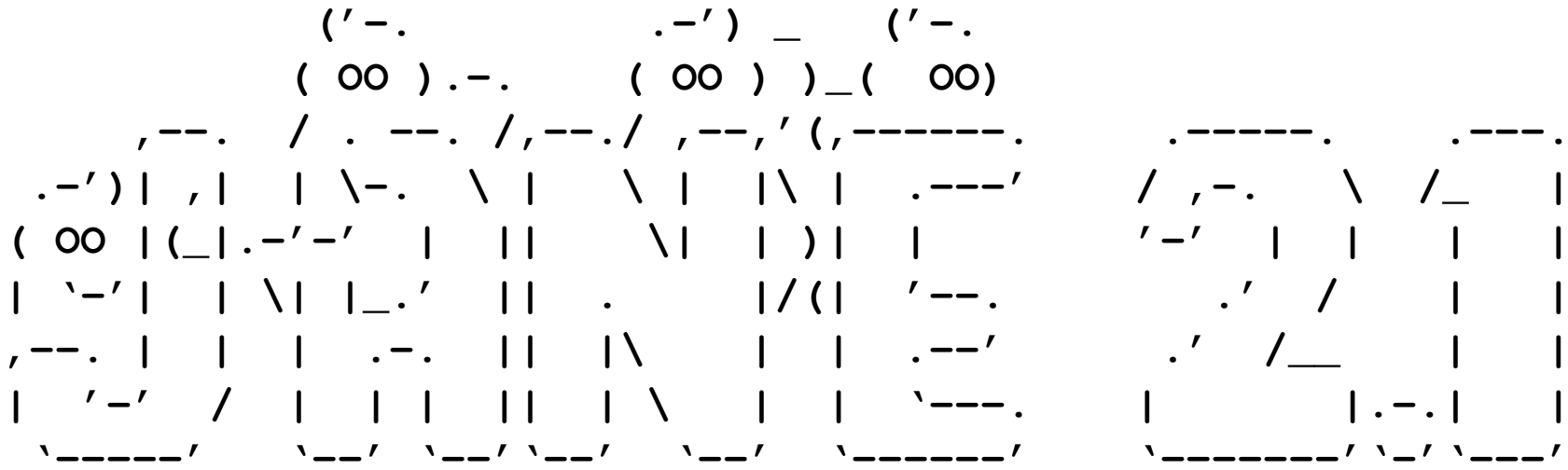
Human Language Technology and Pattern Recognition
Lehrstuhl für Informatik 6
Computer Science Department
RWTH Aachen University, Germany

Overview

- RWTH participated in 6 tracks this year:
  - English ASR
  - Arabic-English MT
  - English-French MT
  - Chinese-English MT
  - German-English MT
  - English-French SLT
- full results will be presented later today at the poster session:

The RWTH Aachen Speech Recognition and Machine Translation System for IWSLT 2012

Stephan Peitz, Saab Mansour, Markus Freitag, Minwei Feng, Matthias Huck, Joern Wuebker, Malte Nuhn, Markus Nußbaum-Thom and Hermann Ney



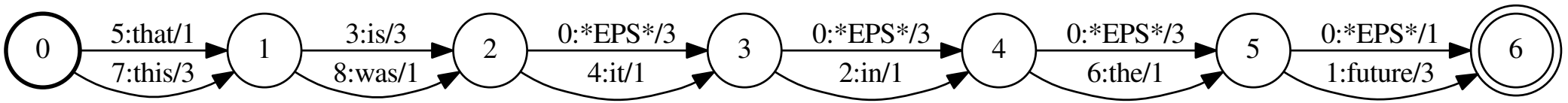
- RWTH's open-source translation toolkit
- new version Jane 2.1
- hierarchical phrase-based decoder [Huck & Peter+ 12]
- phrase-based decoder [Wuebker & Huck+ 12]
- applied in all MT and SLT tasks
- `http://www.hltpr.rwth-aachen.de/jane`

System Combination

- applied in following MT tasks:
  - Arabic-English
  - Chinese-English
  - English-French
- goal: produce consensus translation from multiple systems
- based on [Matusov & Leusch$^+$ 08]
- in this work:
  - create word alignment with METEOR [Banerjee & Lavie 05]
  - feature weights optimization with MERT [Och 03]
  - implementation based on OpenFst [Allauzen & Riley$^+$ 07]

System Combination

- select each hypothesis h in a set of hypotheses as primary system
  1. align all other hypotheses to h using METEOR
  2. construct confusion network
- unify all confusion networks
- add features to the arcs of the confusion networks
- find path with the best score ($=$ consensus translation)



System Combination

- used features in system combination
  - word counts of the single systems
  - language model
  - word penalty
  - binary feature to mark primary system
- features are combined in a log-linear model
- feature weights are optimized with MERT

- in this work:
  - improvements of up to 0.9 points in BLEU over best single systems

Arabic-English

- phrase-based decoder
- preprocessing: different Arabic segmentations
- applied techniques:
  - data selection for LM and TM training [Moore & Lewis 10]
  - phrase table interpolation of in-domain (*in*) and out-of-domain (*ood*)
  - system combination

Phrase Table Interpolation

- **linear interpolation**
  - $p(\tilde{f}|\tilde{e}) = \lambda p_{in}(\tilde{f}|\tilde{e}) + (1-\lambda) p_{ood}(\tilde{f}|\tilde{e})$
  - **interpolation weight λ was adjust on the development set**
- **log-linear interpolation**
  - **fits directly into the SMT log-linear framework**
  - **weights optimized using MERT**
  - **no improvement**
- **ifelse method [Haddow & Koehn 12]**

  if $(\tilde{f}, \tilde{e})$ exists in in-domain phrase table
  assign $p_{in}(\tilde{f}|\tilde{e})$
  else
  assign $p_{ood}(\tilde{f}|\tilde{e})$

Phrase Table Interpolation Results

system	dev2010		tst2010	
	BLEU	TER	BLEU	TER
TED	27.9	51.8	26.1	54.9
TED+UN	28.2	52.8	25.7	57.0
TED-linear-UN	29.0	51.0	26.8	54.6
TED-ifelse-UN	29.5	50.8	26.7	55.0

- TED: in-domain, UN: out-of-domain
- TED+UN: concatenation of in-domain and out-of-domain data

Arabic-English Results

system	tst2010	
	BLEU	TER
FST	26.5 +1.4	55.8 -1.2
SVM	26.6 +1.2	54.4 -3.0
HMM	26.9 +1.2	55.1 -1.8
CRF	26.9 +1.2	54.5 -2.2
MADA-D1	26.3 +1.6	55.4 -2.4
MADA-D2	26.9 +1.7	54.7 -2.4
MADA-D3	27.0 +1.6	54.0 -3.1
MADA-TB ALL	27.1 +1.0	54.4 -2.2
system combination	28.0 +1.0	53.4 -1.3

- a comparison between 2011 and 2012 systems, over tst2010
- for all segmentation methods: linear interpolation and same LM
- improvements of $>$ 1% BLEU on all setups, including final system

Chinese-English

- **decoders:**
  - **in-house phrase-based decoder (PBT)**
  - **hierarchical decoder (HPBT)**
- **applied techniques:**
  - **reverse translation**
  - **system combination**

Reverse Translation

- reverse direction decoding (right-to-left) [Finch & Sumita 09]
- same data as the standard direction system
- reverse the word order of the corpora and test sets
  - retrain the word alignment
  - recompute the language model
- employ on PBT and HPBT
- obtain four different translations
- apply system combination to gain benefits from two-direction decoding

Chinese-English Results

system	dev2010		tst2010	
	BLEU	TER	BLEU	TER
PBT	12.2	80.0	14.2	73.7
PBT-reverse	11.9	79.6	13.7	74.3
HPBT	12.7	80.0	14.7	74.5
HPBT-reverse	12.8	81.0	14.5	76.2
HPBT-withUN-a	12.1	81.4	14.1	76.0
HPBT-withUN-b	12.5	80.4	14.0	75.5
system combination	13.7	78.9	15.4	74.1

- HPBT-withUN-*
  - additional 800K bilingual sentences from UN data
  - differently optimized feature weights

German-English

- phrase-based decoder
- preprocessing:
  - compound splitting [Koehn & Knight 03]
  - POS-based long-range verb reordering [Popović & Ney 06]
- applied techniques:
  - forced alignment [Wuebker & Mauser$^+$ 10]
  - word class language model
  - two phrase tables (in-domain and out-of-domian)

German-English Results

system	dev2010				tst2010			
	BLEU		TER		BLEU		TER	
allData	29.0		49.5		27.5		51.6	
TED	29.9	+0.9	48.4	-0.9	28.4	+0.9	50.3	-1.3
+ForcedAlignment	30.3	+0.4	47.7	-0.7	28.5	+0.1	49.9	-0.4
+ShuffledNews	31.1	+0.8	47.9	+0.2	29.2	+0.7	50.2	+0.3
+WordClassLM	31.2	+0.1	47.8	-0.1	29.8	+0.6	49.7	-0.5
+oodDataTM	31.9	+0.7	47.4	-0.4	30.3	+0.5	49.3	-0.4
+Gigaword	32.6	+0.7	46.4	-1.0	30.8	+0.5	48.6	-0.7

- allData: all available bilingual data vs. TED: in-domain data
- oodDataTM: additional out-of-domain translation model
- incremental improvement of translation quality

Thank you for your attention

Stephan Peitz

peitz@cs.rwth-aachen.de

http://www-i6.informatik.rwth-aachen.de/

References

[Allauzen & Riley+ 07] C. Allauzen, M. Riley, J. Schalkwyk, W. Skut, M. Mohri: OpenFst: A General and Efficient Weighted Finite-State Transducer Library. In *Proceedings of the Ninth International Conference on Implementation and Application of Automata, (CIAA 2007)*, Vol. 4783 of *Lecture Notes in Computer Science*, pp. 11–23. Springer, 2007. 4

[Banerjee & Lavie 05] S. Banerjee, A. Lavie: METEOR: An Automatic Metric for MT Evaluation with Improved Correlation with Human Judgments. In *43rd Annual Meeting of the Assoc. for Computational Linguistics: Proc. Workshop on Intrinsic and Extrinsic Evaluation Measures for MT and/or Summarization*, pp. 65–72, Ann Arbor, MI, June 2005. 4

[Finch & Sumita 09] A. Finch, E. Sumita: Bidirectional phrase-based statistical machine translation. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 3 - Volume 3*, EMNLP '09, pp. 1124–1132, Stroudsburg, PA, USA, 2009. Association for Computational Linguistics. 12

[Haddow & Koehn 12] B. Haddow, P. Koehn: Analysing the Effect of Out-of-Domain Data on SMT Systems. In *Proceedings of the Seventh Workshop on*

Statistical Machine Translation, Montreal, Canada, June 2012. Association for Computational Linguistics. 8

[Huck & Peter+ 12] M. Huck, J.T. Peter, M. Freitag, S. Peitz, H. Ney: Hierarchical Phrase-Based Translation with Jane 2. *The Prague Bulletin of Mathematical Linguistics*, Vol. 98, pp. 37–50, Oct. 2012. 3

[Koehn & Knight 03] P. Koehn, K. Knight: Empirical Methods for Compound Splitting. In *Proceedings of European Chapter of the ACL (EACL 2009)*, pp. 187–194, 2003. 14

[Matusov & Leusch+ 08] E. Matusov, G. Leusch, R. Banchs, N. Bertoldi, D. Dechelotte, M. Federico, M. Kolss, Y.S. Lee, J. Marino, M. Paulik, S. Roukos, H. Schwenk, H. Ney: System Combination for Machine Translation of Spoken and Written Language. *IEEE Transactions on Audio, Speech and Language Processing*, Vol. 16, No. 7, pp. 1222–1237, 2008. 4

[Moore & Lewis 10] R. Moore, W. Lewis: Intelligent Selection of Language Model Training Data. In *ACL (Short Papers)*, pp. 220–224, Uppsala, Sweden, July 2010. 7

[Och 03] F.J. Och: Minimum Error Rate Training in Statistical Machine Translation. In *Proc. of the 41th Annual Meeting of the Association for Computational Linguistics (ACL)*, pp. 160–167, Sapporo, Japan, July 2003. 4

[Popović & Ney 06] M. Popović, H. Ney: POS-based Word Reorderings for Statistical Machine Translation. In *International Conference on Language Resources and Evaluation*, pp. 1278–1283, 2006. 14

[Wuebker & Huck+ 12] J. Wuebker, M. Huck, S. Peitz, M. Nuhn, M. Freitag, J.T. Peter, S. Mansour, H. Ney: Jane 2: Open Source Phrase-based and Hierarchical Statistical Machine Translation. In *International Conference on Computational Linguistics*, Mumbai, India, Dec. 2012. To appear. 3

[Wuebker & Mauser+ 10] J. Wuebker, A. Mauser, H. Ney: Training Phrase Translation Models with Leaving-One-Out. In *Proceedings of the 48th Annual Meeting of the Assoc. for Computational Linguistics*, pp. 475–484, Uppsala, Sweden, July 2010. 14