NICT-2 Translation System for WAT2016: Applying Domain Adaptation to Phrase-based Statistical Machine Translation

Kenji Imamura and Eiichiro Sumita
National Institute of Information and Communications Technology

Summary

- Domain adaptation method of Imamura+ (2016) was applied to WAT2016 data.
  - Japan Patent Office Corpus (JPC) was regarded as a mixture of four domain corpora.
    - Domain adaptation was effective on the patent data even if the domains are different.
  - We added ASPEC as the fifth domain, but there were no effects.
    - The patent data was not effective to the scientific paper domain.
- Google n-gram language models are added as external knowledge.
  - Our domain adaptation can easily incorporate such knowledge.

Domain Adaptation (Imamura+ 2016)

Adaptation of Weight Vector

- Feature weights are optimized using feature augmentation (Daumé 2007).
  - A feature space is expanded to common and domain-specific spaces.
  - All domains are simultaneously optimized/adapted.

Feature Space

	Common	Domain 1	Domain 2	⋯	Domain D
Domain 1 Data	$\Phi_c(f_1, e_1)$	$\Phi_1(f_1, e_1)$	∅	⋯	∅
Domain 2 Data	$\Phi_c(f_2, e_2)$	∅	$\Phi_2(f_2, e_2)$	⋯	∅
	⋮	⋮	⋮		⋮
Domain D Data	$\Phi_c(f_D, e_D)$	∅	∅	⋯	$\Phi_D(f_D, e_D)$

Optimization

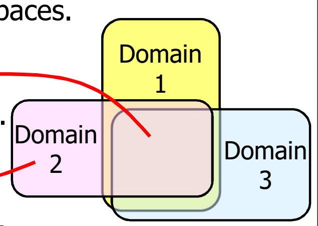
Feature Weights	w_c	w_1	w_2	⋯	w_D

Subspaces used for the translation in the domain 1

Subspaces used for the translation in the domain 2

Adaptation of Feature Vector

- Models are changed according to the feature spaces.
  - For the common space, we use a corpus-concatenated model, which is trained from corpora of all domains.
  - For the domain specific spaces, we use single-domain models, which are trained from specific domain data.



- This distinction matches meanings of the spaces.

Decoding Procedure

1. Phrase pairs are retrieved from both the corpus-concatenated and single-domain phrase tables.
2. Features of the corpus-concatenated model are located to the common space, and those of the single-domain model are located to the domain-specific space.
3. During search of the best hypothesis, the likelihoods are computed using only the common space and domain-specific space of the input sentence.

Domain/Corpora

- Japan Patent Office Corpus (JPC) was regarded as a mixture of four domains.
- Asian Scientific Paper Excerpt Corpus (ASPEC) was used as the fifth domain corpus.
  - The language pairs: Japanese-English (Ja-En) and Japanese-Chinese (Ja-Zh).

Corpus	Domain	#training sents. Ja-En pair	#training sents. Ja-Zh pair
JPC	Chemistry	250k	250k
	Electricity	250k	250k
	Machine	250k	250k
	Physics	250k	250k
ASPEC	ASPEC	1,000k	672k

Implementation Notices

Empty Value

- A value of feature functions when phrases appear only one of the corpus-concatenated or single domain models (unknown probability).

Common	Domain i
Φ_c(translation, 翻訳)	Φ_i(translation, 翻訳)
(-1.7, -6.3, -2.2, -7.6)	?? if the pair does not exist in the phrase table.

- We experimentally set to maximize the BLEU score of the development set.
  - This time, empty= -7 (i.e., exp(-7) = 0.0009).

Large Monolingual Corpora

- External knowledge such as language models constructed from large monolingual corpora is located to the common space while increasing the dimension.
- Language models are constructed from Google n-gram, and added as the external knowledge.
  - The back-off models are estimated using maximum likelihood.
  - English Data : Web 1T 5-gram Version 1 (LDC2006T13)
    Japanese Data: Web Japanese N-gram Version 1 (http://www.gsk.or.jp/catalog/gsk2007-c/)

Optimization

- Independent optimization of Imamura+ (2016) was used.
  - Each domain is optimized one-by-one.
  - Optimization algorithm: K-best Batch MIRA.

Translation System

- Phrase-based SMT with preordering.
  - Two preorderers:
    (1) Top-Down BTG (w/o external knowledge), and
    (2) In-house preorderer tuned to patents
    (w/ external knowledge, using Berkeley Parser).
  - Moses clone decoder.

Experimental Results

Settings

- Domain Adaptation vs. {Single-Domain / Corpus Concatenation}
  - Evaluation Metric: BLEU
    Statistical Testing: MultEval ($p<0.05$).
  - The scores are different from the official scores.

JPO Corpus (w/o External Knowledge)

- Corpus Concatenation: JPC was regarded as one domain corpus.
- Single Domain Model: If we divided JPC into 4 domains, the translation quality decreased because the number of the training sentences in each domain is reduced.
- Domain Adaptation: The BLEU scores were the highest.

Method	JPC Ja-En	JPC En-Ja	JPC Ja-Zh	JPC Zh-Ja
Corpus Concatenation	36.22	38.03 (-)	32.92 (-)	39.68 (-)
Single-Domain Model	35.12 (-)	37.40 (-)	31.96 (-)	38.15 (-)
Domain Adaptation	**36.29**	**38.48**	**33.36**	**39.85**

JPO and ASPEC Corpus (w/ External Knowledge)

- On JPC, Google n-gram the language models and domain adaptation were both effective. They can be combined.
- On ASPEC, domain adaptation was not effective. This might be because the corpus size of ASPEC is large.

	Method	JPC Ja-En	JPC En-Ja	JPC Ja-Zh	JPC Zh-Ja
w/o GN	Corpus Concatenation	35.81 (-)	38.62 (-)	32.76 (-)	39.96 (-)
	Single-Domain Model	33.90 (-)	38.19 (-)	31.78 (-)	38.74 (-)
	Domain Adaptation	36.25	39.58	**33.53**	40.76
w/ GN	Corpus Concatenation	36.03 (-)	39.48 (-)		40.14 (-)
	Single-Domain Model	34.35 (-)	39.04 (-)		38.90 (-)
	Domain Adaptation	**36.40**	**40.32**		**40.77**

	Method	ASPEC Ja-En	ASPEC En-Ja	ASPEC Ja-Zh	ASPEC Zh-Ja
w/o GN	Corpus Concatenation	22.20 (-)	33.94 (-)	28.95 (-)	37.62 (-)
	Single-Domain Model	22.79	34.80	**29.47 (+)**	38.96 (-)
	Domain Adaptation	22.80	34.91	29.28	39.18
w/ GN	Corpus Concatenation	22.10 (-)	34.55 (-)		38.15 (-)
	Single-Domain Model	**22.87 (+)**	**35.42**		39.74 (-)
	Domain Adaptation	22.74	35.36		**39.87**

References

- Hal Daumé III. 2007. Frustratingly Easy Domain Adaptation. In Proc. of ACL-2007, pp. 256-263.
- Kenji Imamura and Eiichiro Sumita. 2016. Multi-domain Adaptation for Statistical Machine Translation Based on Feature Augmentation. In Proc. o AMTA-2016. pp. 79-92.