Dialect Translation:
Integrating Bayesian Co-segmentation Models with Pivot-based SMT

M. Paul, A. Finch, P. R. Dixon, and E. Sumita

Presented by Wael Salloum
Course: Bayesian Analysis for NLP
Instructor: Shay Cohen, Ph.D.
Columbia University
May 6th 2013

Problem Definition

- Task: Translate Dialects into Foreign (Target) Languages.
  - Source Japanese dialects: (Kumamoto, Kyoto, Okinawa, Osaka)
  - Target Languages:
    - Indo-European languages (English, German, Russian, Hindi)
    - Asian languages (Chinese, Korean)
- Problems:
  - Dialects are resource-poor languages:
    - Limited parallel data to train Statistical Machine Translation (SMT)
    - Limited NLP tools (e.g., word segmentation)

Approaches

- Direct Translation
- Pivot-based Translation:
  - SMT-based Pivot Translation: Dialect-to-Standard SMT followed by Standard-to-Target SMT
  - BCS-based Pivot Translation: Dialect-to-Standard **Transduction** followed by Standard-to-Target SMT

Statistical Machine Translation

- A maximization problem:

$$argmax_{trg}\ p(src|trg) * p(trg)$$

Dialect-to-Standard Transduction

- Bayesian co-segmentation (BCS) model
- Joint-source channel model (*n*-gram transliteration model)

Dialect-to-Standard Transduction

- **Transliteration**: character-to-character mapping to transfer *Dialect* sentences to *Standard* word segments.
- The paper uses a **Generative Bayesian Model**:
  - Avoids *over-fitting.*
  - Constructs *compact* models that have only a small number of *well-chosen* parameters.
  - Is based on *joint source channel model.*
  - Is *symmetric* w.r.t. source and target languages.

Joint-Source Channel Model

- A Dialect sentence: $\sigma = l_1, l_2, \ldots, l_L$ (l is a character)
- A Standard sentence: $\omega = s_1, s_2, \ldots, s_S$ (s is a word token)
- There exists an **alignment**

$$\gamma = <l_1 \ldots l_q, s_1>, \ldots, <l_r \ldots l_L, s_S>$$

of K transliteration units.

➔ The n-gram model: the transliteration probability of a transliteration pair $<l, s>_k$ depending on its immediate n preceding transliteration pairs:

$$P(\sigma, \omega, \gamma) = \prod_{k=1}^{K} P(<l, s>_k | <l, s>_{k-n+1}^{k-1})$$

Bayesian co-segmentation (BCS)

- Two Models:
  - A model for *generating* an outcome that has already been generated at least once before
  - A model for *assigning* a probability to an outcome that has not yet been produced
- The co-segmentation process is driven by a Dirichlet process. The underlying stochastic process for the generation of a corpus of bilingual phrase pairs (s_k,t_k):

$$
\begin{aligned}
G|_{\alpha,G_0} &\sim DP(\alpha, G_0) \\
(\mathbf{s}_k, \mathbf{t}_k)|G &\sim G
\end{aligned}
$$

Bayesian co-segmentation (BCS)

- The base measure G_0 controls the generation of novel sequence pairs: A joint spelling model to assign probabilities to them:

$$\begin{aligned} G_0((\mathbf{s},\mathbf{t})) &= p(|\mathbf{s}|)p(\mathbf{s}||\mathbf{s}|) \times p(|\mathbf{t}|)p(\mathbf{t}||\mathbf{t}|) \\ &= \frac{\lambda_s^{|\mathbf{s}|}}{|\mathbf{s}|!}e^{-\lambda_s}v_s^{-|\mathbf{s}|} \times \frac{\lambda_t^{|\mathbf{t}|}}{|\mathbf{t}|!}e^{-\lambda_t}v_t^{-|\mathbf{t}|} \end{aligned}$$

- The generative model:

$$\begin{aligned} & p((\mathbf{s}_k,\mathbf{t}_k))|(\mathbf{s}_{-k},\mathbf{t}_{-k})) \\ & \qquad = \frac{N((\mathbf{s}_k,\mathbf{t}_k)) + \alpha G_0((\mathbf{s}_k,\mathbf{t}_k))}{N+\alpha} \end{aligned}$$

Bayesian co-segmentation (BCS)

- **Sampling**: Blocked Gibbs sampler
- Extended the forward filtering / backward sampling DP algorithm to deal with bilingual segmentation.

Experiments

Approaches

- Direct Translation
- Pivot-based Translation:
  - SMT-based Pivot Translation: Dialect-to-Standard SMT followed by Standard-to-Target SMT
  - BCS-based Pivot Translation: Dialect-to-Standard **Transduction** followed by Standard-to-Target SMT

Direct Translation

Table 2: SMT-based Direct Translation Quality

BLEU (%)

SRC / TRG	ja (160k)	ja (20k)	ja_{ku} (20k)	ja_{ky} (20k)	ja_{ok} (20k)	ja_{os} (20k)
en	56.51	32.84	32.27	31.81	30.99	31.97
de	51.73	26.24	25.06	25.71	24.37	25.18
ru	50.34	23.67	23.12	23.19	22.30	22.07
hi	49.99	21.10	20.46	20.40	19.72	20.96
zh	48.59	33.80	32.72	33.15	32.66	32.96
ko	64.52	53.31	52.93	51.24	49.40	51.57

SMT-based Pivot Translation

Table 3: SMT-based Pivot Translation Quality

BLEU (%)

SRC / TRG	ja_{ku}	ja_{ky}	ja_{ok}	ja_{os}
	($\mathrm{SMT}_{SRC \rightarrow ja}$+$\mathrm{SMT}_{ja \rightarrow TRG}$)			
en	52.10	50.66	45.54	49.50
de	47.51	46.33	39.42	44.82
ru	44.59	43.83	38.25	42.87
hi	45.89	44.01	36.87	42.95
zh	45.14	44.26	40.96	44.20
ko	60.76	59.67	55.59	58.62

Dialect-to-Standard Transduction

Table 4: Dialect to Standard Language Transduction

BLEU (%)

Engine	SRC (decoding)	ja_{ku}	ja_{ky}	ja_{ok}	ja_{os}
			($SRC \rightarrow ja$)		
BCS	(monotone)	91.55	86.74	80.36	85.04
SMT	(monotone)	88.39	84.87	74.27	82.86
	(reordering)	88.39	84.73	74.26	82.66

BCS-based Pivot Translation

Table 5: BCS-based Pivot Translation Quality

BLEU (%)

SRC / TRG	ja_{ku}	ja_{ky}	ja_{ok}	ja_{os}
	($BCS_{SRC \rightarrow ja}$+$SMT_{ja \rightarrow TRG}$)			
en	52.42	50.68	45.58	50.22
de	47.52	46.74	39.93	45.60
ru	45.29	44.08	38.39	43.53
hi	45.72	44.71	37.60	43.56
zh	45.15	43.92	40.15	44.06
ko	60.26	59.14	55.33	58.13

Comparison of Approaches

Table 6: Gains of BCS-based Pivot Translation
BLEU (%)

SRC / TRG	ja_{ku}	ja_{ky}	ja_{ok}	ja_{os}
	on SMT-based Pivot (Direct) Translation			
en	+0.32 (+20.15)	+0.02 (+18.87)	+0.04 (+14.59)	+0.72 (+18.25)
de	+0.01 (+22.46)	+0.41 (+21.03)	+0.51 (+15.56)	+0.78 (+20.50)
ru	+0.70 (+22.17)	+0.25 (+20.89)	+0.14 (+16.09)	+0.66 (+21.46)
hi	-0.17 (+25.26)	+0.70 (+24.31)	+0.73 (+17.88)	+0.61 (+22.60)
zh	+0.01 (+12.43)	-0.34 (+10.77)	-0.81 (+7.49)	-0.14 (+11.10)
ko	-0.50 (+7.33)	-0.53 (+7.90)	-0.26 (+5.93)	-0.49 (+6.56)

Thank You

Questions?