Centre for Next Generation Localisation

MaTrEx: the DCU MT System for NTCIR-8

Tsuyoshi Okita, Jie Jiang, Rejwanul Haque, Hala Al-Maghout, Jinhua Du, Sudip Kumar Naskar, Andy Way
Dublin City University, CNGL/School of Computing

National Development Plan 2007 - 2013
science foundation ireland
Dublin City University
University College Dublin
University of Limerick
Trinity College Dublin

Table Of Contents

MaTrEx:

- MaTrEx
  - Supertagged PB-SMT
  - Context-informed PB-SMT
  - Noise reduction
  - System combination
- We participated in Machine Translation subtasks:
  - Intrinsic EN-JP: the second among six participants (BLEU).
  - Intrinsic JP-EN: the fourth among seven participants (BLEU).
  - Extrinsic JP-EN: the first (Mean Average Precision) and the third (Recall@N).

NTCIR-8 Corpora

- Patent Corpus

	train set	dev set	test set
JP-EN	3,186,284	1,200/2,000	1,251
EN-JP	3,186,284	1,200/2,000	1,119

Table: Parallel corpus size of NTCIR-8

NTCIR-8 Corpora

- Unstructured complex sentences

Japanese: この第２のスライドブロック５のスライド移動によって、弾性糸ＳＹが、第１のスライドブロック４の下流側面と前記第２のスライドブロック５の上流側面との間で、確実に把持されるとともに、前記弾性糸ＳＹは、第２のスライドブロック５の下流側面と下側の固定ブロック６の上流側面とのスライドによって前記カッター刃１０の作用により切断される。

English: due to this slidable movement of the second slide block 5 , the elastic yarn sy is reliably held between the downstream side of the first slide block 4 and the upstream side of the second slide block 5 , and the elastic yarn sy is cut by the operation of the cutter blade 10 due to sliding between the downstream side of the second slide block 5 and the upstream side of the fixed block 6 at the lower side .

NTCIR-8 Corpora

- Translational Omission

  Japanese: 従来、スパッタリング用ターゲット（以下、単にターゲットと略称する）としてはプレーナ型（円板状もしくは角板状）のターゲットが広く使用されている。

  English: conventional sputtering targets extensively in use are of a planer type having a circular or square plate-like shape .

NTCIR-8 Corpora

- Equations in a sentence

  Japanese: 処理７３では、上記の取り込んだ信号Ｖ，θｆ，ｄ／ｄｔ（θｆ） から、目標ヨー角加速度ｄ／ｄｔ（ωＴ） を決定する。

  English: at a process 73 , target yawing angular acceleration d/dt (.omega. .sub.t) is determined based on the fetched signals v , .theta.f and d/dt (.theta.f) .

NTCIR-8 Corpora

Reference number

Japanese: この軸受けユニット（44 ）は、本体支柱（４）に固着した上・下部ブラケット（45 ）（46 ）と、この上・下部ブラケット（45 ）（46 ）に挿通した軸セット（47 ）と、軸セット（47 ）と製品容器（２）間を連結するアーム（48 ）とで構成する。
Japanese: 図１に示すガイド５とガイドローラ３はこのような案内を行うものであり、以下にその実施例を説明する。

English: the bearing unit 44 comprises upper and lower brackets 45 and 46 fixed on the body pillar 4 , a shaft set 47 inserted in said upper and lower brackets 45 and 46 , and an arm 48 interconnecting the shaft set 47 and the product container 2.
English: the guide 5 and the guide rollers 3 shown in fig. 1 are designed to provide such guidance , and embodiments thereof will be described hereinunder .

NTCIR-8 Corpora

Many parentheses

Japanese: 次に、直径５ｍｍの多数の空孔（９６ポイント）を有する開閉式の扉１４を開けて、大気中に浮遊している有機ガス１５を基板１０に２４時間吸着させる（ステップＳＡ２）。

English: then , in step 2 of fig. 4 , the cover 14 is opened for 24 hours so that gaseous organic substances 15 floating in an atmosphere are adsorbed to the silicon substrate 10 .

Centre for Next Generation Localisation

Table Of Contents

Supertagging

1. Lexicalized grammatical formalisms
   - Lexicalized Tree Adjoining Grammar (Schabes et al., 88)
   - Combinatory Categorial Grammar (Steedman, 00)
   - Head-Driven Phrase-Structure Grammar (Pollard and Sag, 94)
2. Supertagging: To separate lexical category assignment from the combinatory processes that make use of such categories in lexicalized grammatical formalisms.
   - Lexical category assignment: the assignment of informative syntactic categories to linguistic objects such as words or lexical predicates.
   - Combinatory processes: parsing and surface realization.

LTAG supertag

LTAG supertag sequence for the sentence
’ The purchase price includes taxes’

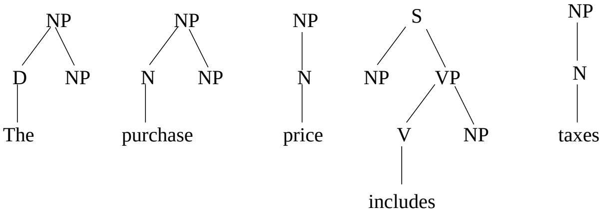

Figure: Supertags (Lexical syntax)

Supertagged PB-SMT [Hassan et al., ACL07]

- To incorporate the target-side supertag info.
- Better local reorderings.
- Only Japanese to English direction (supertaggers available only for English)
- This technique is challenging:
  - Long distance dependencies in Japanese may not be captured well by incorporating the supertag information in English.
  - Quality of parsing outputs of patent data are problematic. (reference numbers, many parentheses, long sentences, technical terms, and symbols).
- We use the HPSG supertagger ENJU (Miyao, Matsuzaki07) (instead of the CCG supertagger (Clark,02))

Supertagged PB-SMT

Incorrect parsing by supertagger

1) CCG supertagger [Clark et al., 03]

$The|DT|NP[nb]/N \quad guide|NN|N \quad 5|CD|N\backslash N \quad and|CC|conj$
$the|DT|NP[nb]/N \quad guide|NN|N \quad rollers|NNS|N \quad 3|CD|N\backslash N$
$shown|VBN|S[pss]\backslash NP \quad in|IN|((S\backslash NP)\backslash(S\backslash NP))\backslash NP \quad FIG|NN|N$
$.|.|. \quad 1|CD|N \quad are|VBP|(S[dcl]\backslash NP)\backslash(S[pss]\backslash NP)$
$designed|VBN|(S[pss]\backslash NP)\backslash(S[to]\backslash NP) \quad \ldots$

2) HPSG supertagger (ENJU [Miyao et al., 03;Matsuzaki et al., 07])

$The|[\langle D\rangle]N \quad guide|[D\langle N.3sg\rangle] \quad 5|N[\langle ADJP\rangle] \quad and|[N\langle CONJP\rangle N]$
$the|[\langle D\rangle]N \quad guide|[D\langle N.3sg\rangle] \quad rollers|[D\langle N.3sg\rangle] \quad 3|N[\langle ADJP\rangle]$
$shown|[NP.nom\langle V.bse\rangle NP.acc] \quad in|V[\langle P\rangle NP.acc] \quad FIG.|[D\langle N.3sg\rangle]$
$1|N[\langle ADJP\rangle] \quad are|[NP\langle V.be.bse\rangle VP.pas]$
$designed|[NP.nom\langle V.bse\rangle NP.accVP.inf] \quad \ldots$

Supertagged PB-SMT

- $t =< \phi_t, O_t >$ and $s =< \phi_s, O_s >$: target and source sentences (separate orderings from content)
  - ϕ_x: the bag of phrases that constitute x,
  - O_x: the order of the phrases

  - $P_w(t)$: target-language model,
  - $P(O_s|O_t)$: conditional (order) linear distortion probability,
  - $P(\phi_s|\phi_t)$: translation model from target-language bags of phrases to source-language bags of phrases.
  - ST: a supertag sequence of the same length as a target sentence t.
  - $P_{ST}(t, ST)$: language model for sequences of word-supertag pairs.

Supertagged PB-SMT

- Noisy-channel Model can be rewritten as

$$
\begin{aligned}
&\arg\max_{t} P(s|t)P(t)\\
&= \arg\max_{t} \sum_{ST} P(s|t, ST)P_{ST}(t, ST)\\
&= \arg \max_{<\phi_t, O_t>} P(\phi_s|\phi_t)P(O_s|O_t)P_w(t)\\
&= \arg \max_{<t, ST>} P(\phi_s|\phi_t, ST)P(O_s|O_t)^{\lambda_o} P_{ST}(t, ST) \exp^{|t|\lambda_w}
\end{aligned}
$$

Context-informed PB-SMT [Haque et al., EAMT09]

- To incorporate the source-side supertag info.
- To capture source-side context to solve lexical ambiguity
- Only English to Japanese direction (supertaggers available only for English)
- This technique is challenging for Japanese:
  - Structural ambiguity in syntactic constituency rather than lexical ambiguity
  - Scrambling phenomenon (Harada,77) rearrange the order among the constituents of a sentence where the case particles can serve to identify the functions of the accompanying NPs within the sentence.

Context-informed PB-SMT (Scrambling phenomenon)

- English: rigid word order among constituents in a sentence.
  - John gave Bill Mary. (Different meaning).
  - John gave Mary Bill.
  - Bill gave John Mary.
  - Bill gave Mary John.
  - Mary gave Bill John.
  - Mary gave John Bill.
- Japanese: relatively free word order.
  - Kinoo Taroo-ga Ginza-de susi-o tabeta. (Same meaning).
  - Taroo-ga Ginza-de kinoo susi-o tabeta.
  - Kinoo susi-o Taroo-ga Ginza-de tabeta.
  - Susi-o kinoo Taroo-ga Ginza-de tabeta.
  - Ginza-de Taroo-ga kinoo susi-o tabeta.
  - Kinoo Ginza-de susi-o Taroo-ga tabeta.
  - ...

Context-informed PB-SMT

- Log-Linear PB-SMT can be rewritten as

$$\begin{aligned}\arg\max p(e|f) &= \arg\max \Sigma\lambda_i h_i(e,f) \\ &= \arg\max \Sigma\lambda_i \hat{h}_i(\hat{f}_k, CI(\hat{f}_k), \hat{e}_k, s_k)\end{aligned}$$

  - s_k: segmentation of source and target sentences.
- Context-informed feature ($\hat{h}_m$): CI may include any feature (lexical, syntactic, etc.)

$$\hat{h}_m(\hat{f}_k, CI(\hat{f}_k), \hat{e}_k, s_k)(= logP(\hat{e}_k|\hat{f}_k, CI(\hat{f}_k)))$$

Context-informed PB-SMT

For a given focus phrase $\hat{f}_k = f_{i_k} \ldots f_{j_k}$ of fixed window size $2l$ (experiments we use window size of $\pm$ 1 and $\pm$ 2),

- Lexical Features (CI_{lex})

  $$CI_{lex}(\hat{f}_k) = \{f_{i_k-l}, \ldots, f_{i_k-1}, f_{j_k+1}, \ldots, f_{j_k+l}\}$$

- Syntactic Features (Part-of-Speech tag)

  $$CI_{pos}(\hat{f}_k) = \{pos(f_{i_k-1}), \ldots, pos(f_{i_k-l}), pos(\hat{f}_k), pos(f_{j_k+1}), \ldots, pos(f_{j_k+l})\}$$

- Syntactic Features (Supertags)

  $$CI_{st}(\hat{f}_k) = \{st(f_{i_k-1}), \ldots, st(f_{i_k-l}), st(\hat{f}_k), st(f_{j_k+1}), \ldots, st(f_{j_k+l})\}$$

Noise Reduction [Okita, ACL09SRW]

- $p(\bar{e}|\bar{f})$: Obtain this indirectly (Word alignment $p(e|f)+$ phrase extraction heuristics)
  - word alignment: For a given a pair of sentence aligned bilingual texts, to find a lexical translation probability $p_{f_i} : e_i \rightarrow p_{f_j}(e_i)$ such that $\Sigma p_{f_j}(e_i) = 1$ and $\forall e_i : 0 \leq p_{f_j}(e_i) \leq 1$.
  - Phrase extraction: For a given word alignment, to extracts all consistent phrase pairs from a word aligned sentence pair [Och and Ney, 03].
- $p(\bar{e}|\bar{f})$: Obtain this directly (Phrase alignment [Marcu and Wong, 2002]) where computational complexity is $O(n^4)$.
  - $p(e,f) = \prod_{i=1}^{n} p(\bar{e}_i, \bar{f}_i | c_i) d(pos(e_i)|pos(e_{i-1}))$

Noise Reduction

- ‘Word alignment + phrase extraction’ approach is a compromise to solve phrase alignment.
- This makes new problem (N-to-m mapping object problem): For word alignment, empirical evidence has shown that n-to-m mapping objects, such as paraphrases, non-literal translations, and multiword expressions, appear as both noise (or outlier) and as valid training data [Fraser, 07; Okita, 09].
- (Noise aspects): If we collect ‘good points’, we may be able to avoid such noise [Okita,09].

Noise Reduction

Algorithm 1 Good Points Algorithm

Step 1: Train WB-SMT, and translate all the sentences to get n-best lists.
Step 2: Obtain the sentence-based cumulative X-gram ($X \in \{1, \cdots, 4\}$) score $S_{WB,X}$.
Step 3: Train PB-SMT, and translate all training sentences to get n-best lists.
Step 4: Obtain the sentence-based cumulative X-gram ($X \in \{1, \cdots, 4\}$) score $S_{PB,X}$.
Step 5: Remove sentence pairs where $S_{WB,2} = 0$ and $S_{PB,2} = 0$.
Step 6: The remaining sentence pairs after removal in Step 5 are used to train the final PB-SMT systems.

Noise Reduction

upon addition of diethyl ether , the solution became turbid .
ここ_に_ジエチルエーテル_を_加える_と_白濁_し_た_。

the rest of the operation is the same as the order (a) .
後_は_（_a_）_と_同じ_。

moreover , it is possible for the device to be automatically driven .
また_装置_を_無人_運転_化_する_こと_も_可能_で_ある_。

the smaller the index , the better the stopping ability is .
小さい_ほど_良好_で_ある_こと_を_示す_。

a method coping with this will be described later .
これ_に対する_対応_案_を_後_に_述べる_。

this means that no slip in the roll of the already wound long film occurs .
すなわち_、_フィルムロール_の_内部_で_は_フィルム_間_に_滑り_が_生じ_ない_。

when this process is repeated , the whole image is recorded .
これ_を_くり返す_こと_により_画像_全体_を_記録_する_という_もの_で_ある_。

a high feed rate will further improve the machining efficiency .
送り_量_が_大きい_こと_により_、_切削_加工_の_能率_が_一層_向上_する_。

System Combination

- Minimum Bayes-Risk-Confusion Network (MBR-CN) framework [Kumar and Byrne, 2004][Du et al., WMT2009] (Work very well in our recent MT evaluation campaigns).

$$\hat{e}_i = \arg\min_{e_i} \sum_{j=1}^{N} \{1 - BLEU(e_j, e_i)\}$$

- Confusion Network:
  - (backbone) output of MBR decoder, (other elements) other hypotheses are aligned by TER (NULL words are allowed).
  - (Each node in CN) votes (or some form of confidence measures), (Each arc in CN) an alternative word at that position in the sentence.
  - Features: 1) word posterior probability, 2) trigram and 4-gram target language model, 3) word length penalty, and 4) NULL word length penalty.

System Combination

System translations (3 translation outputs)

it does not go home
he does not to the home
he does not go house

Confusion networks

0.33	0.67	1.00	0.67	0.67	0.67	0.67	
it	does	not	go	(empty)	(empty)	home	backbone
he	(empty)		goes	to	the	house	
0.67	0.33		0.33	0.33	0.33	0.33	

Table Of Contents

Experimental Setup

- Baseline System: Standard log-linear PB-SMT system
  - word alignment by Giza++,
  - phrase extraction heuristics,
  - MERT (optimised by BLEU),
  - 5-gram language model with Kneser-Ney smoothing by SRILM, and
  - Moses [Koehn et al., 07].
- System Combination
  - Joshua (Hierarchical Phrase-Based system) [Li et al., 09],
  - Chart-based Moses decoder [Hoang et al., 09].

Intrinsic Evaluation (JP-EN)

Systems	BLEU	#OOV
System combination	27.61*	321
HPB-SMT 1	26.86*	314
PB-SMT 1	26.51*	194
Noise reduction (PB-SMT)	24.01	443
PB-SMT 2+	23.91*	316
Preprocessing (PB-SMT)+	23.82	194
HPB-SMT 2	23.30	303
Supertag (ENJU) 1	20.68	430
Supertag (ENJU) 2	18.27	426
System combination (unofficial run)	28.43	331

Table: Intrinsic evaluation results (JP-EN). Noted that we trained over 3,200k training corpus for the systems marked with + and over 600k training corpus for other systems.

Intrinsic Evaluation (EN-JP)

Systems	BLEU
System combination	33.03
HPB-SMT 1	32.50
PB-SMT 1	30.53
PB-SMT 2^+	30.08
Noise reduction	29.53
Preprocessing (PB-SMT)$^+$	27.93
HPB-SMT 2	27.23
Context supertag (Base)	26.83
Context supertag (Superpair)	26.45
Context supertag (CCG)	26.38
Context supertag (LTAG)	26.38
Context supertag (CCG-LTAG)	26.22
Context supertag (POS)	26.21

Table: Intrinsic evaluation results (EN-JP).

Extrinsic Evaluation (JP-EN)

Systems	BLEU	MAP	r@100	r@200	r@500	r@1000
PB-SMT 1^+	<u>24.00</u>	<u>0.21</u>	<u>0.55</u>	0.63	<u>0.72</u>	<u>0.78</u>
HPB-SMT 2	23.71	0.18	0.53	0.59	0.68	0.73
HPB-SMT 1	23.48	0.18	0.53	0.59	0.68	0.74
PB-SMT 2	22.35	<u>0.21</u>	<u>0.55</u>	<u>0.64</u>	0.70	0.76

Table: Extrinsic evaluation results. The column shows the evaluated measure whether it is BLEU, MAP (Mean Average Precision) or Recall@N (which is abbreviated in a table as r@N). It is noted that we trained over 3,200k training corpus for the systems marked with $^+$ and over 600k training corpus for other systems.

Experiments

その結果、記録層３０の合成磁化は、ほぼ０となる。

(HPB-SMT2) the result synthesis , the magnetization of the recording layer 3 becomes almost 0 .

(PB-SMT1) as a result , the magnetization of the recording layer 3 and the synthesizing substantially .

(Noise) as a result , the recording layer 30 , a combination of magnetization becomes almost zero .

(Supertag) as a result , the recording layer 3 of the synthetic magnetization , the substantially zero .

(Syscombo4) as a result , the magnetization of the recording layer 3 and becomes almost 0 .

(Syscombo7) as a result , the magnetization of the recording layer 3 of magnetization becomes almost zero .

(Reference) As a result, the composed magnetization in the recording layer 30 becomes almost 0.

Conclusions and Further Works

- 2nd best system for EN-JP.
- System combination strategy is effective in both EN-JP (0.75 BLEU points) and JP-EN (0.53 BLEU points) even though we combine only four 1-best translation outputs.
- Supertagged PB-SMT and context-informed PB-SMT seem to have difficulties probably due to the typical characteristics of Japanese.
- Further Works:
  - Appropriate preprocessing method to deal with equations, parentheses, and symbols may improve the overall performance.
  - A word lattice-based decoding approach may reduce OOV words.

Acknowledgement

Thank you for your attention.

- This research is supported by the Science Foundation Ireland (Grant 07/CE/I1142) as part of the Centre for Next Generation Localisation at Dublin City University.
- Irish Centre for High-End Computing.
- Travel Support by NTCIR-8 organizers.

Dublin City University Trinity College Dublin University College Dublin University of Limerick

Centre for Next Generation Localisation

ICHEC Irish Centre for High-End Computing