Statistical Post-Editing

Miguel Domingo

Pattern Recognition and Human Language Technologies

March 31, 2015

Table of Contents

Aim

The idea behind Statistical Post-Editing (SPE) is being able to automatically post-edit a text translated by a Machine Translation (MT) system, by means of Statistical Machine Translation (SMT), in order to improve the translation quality.

This can be achieve by training an SMT system in which the output of the previous MT system becomes the new source language, and the target language remains the same.

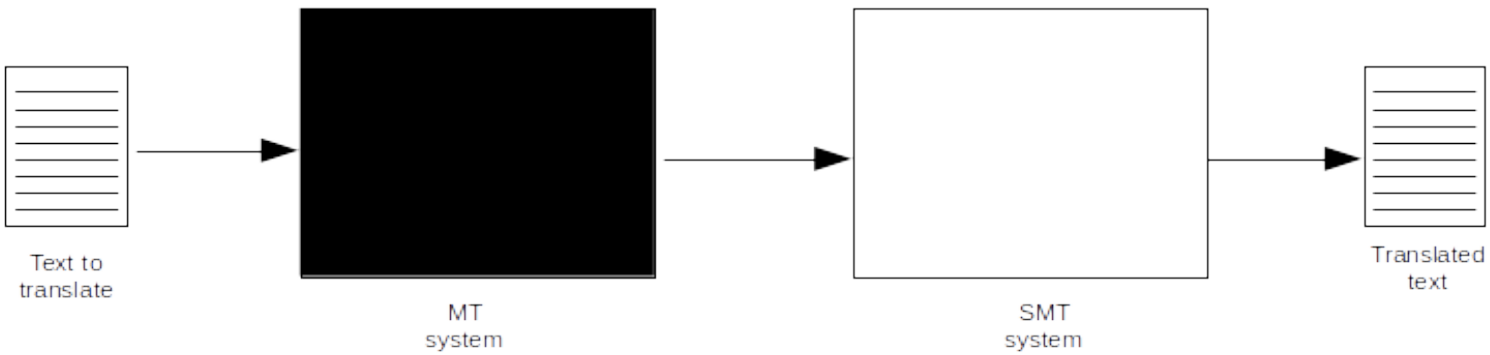

Table of Contents

Rule-Based Machine Translation

Aim

Improve the translation quality of a Rule Based Machine Translation (RBMT) system by means of SPE. This is achieve by training an SMT system which takes the RBMT system's output as input.

Rule-Based Machine Translation

Canadian Job Bank (Simard, Goutte, & Isabelle, 2007)

- *Content*: Job ads.
- *Language pairs*: English-French, French-English.
- *RBMT system*: Unknown.
- *SMT system*: Portage.
- *Goal*: Reduce human post-editing effort.

Rule-Based Machine Translation

Table: Experimental Results: For TER, lower (error) is better, while for BLEU, higher (score) is better. Best results are in bold.

Language	TER	BLEU
English-to-French		
Baseline	53.5	32.9
Portage *translation*	53.7	36.0
Baseline + Portage APE	**47.3**	**41.6**
French-to-English		
Baseline	59.3	31.2
Portage *translation*	43.9	41.0
Baseline + Portage APE	**41.0**	**44.9**

Rule-Based Machine Translation

Canadian Job Bank II (Isabelle et al., 2007)

- *Content*: Job ads.
- *Language pairs*: English-French, French-English.
- *RBMT system*: Unknown.
- *SMT system*: Portage.
- *Goal*: Reduce human post-editing effort.

Rule-Based Machine Translation

Table: Experimental Results: For TER, lower (error) is better, while for BLEU, higher (score) is better. Results for Automatic Post-Editing are in bold.

	English-to-French		French-to-English	
	TER	BLEU	TER	BLEU
T_1 (vanilla RBMT)	62.2	23.3	68.8	24.4
T_2 (customized RBMT)	53.5	32.9	59.3	31.2
Portage SMT	53.7	36.0	43.9	41.0
T_1 + APE	**48.6**	**39.8**	**41.5**	**44.2**
T_2 + APE	**47.3**	**41.6**	**41.0**	**44.9**

Rule-Based Machine Translation

Second Workshop on Statistical Machine Translation (Simard, Ueffing, & Isabelle, 2007)

- *Corpora*: Europarl v6 and News Commentary.
- *Language pairs*: English-French, French-English.
- *RBMT system*: Systran.
- *SMT system*: Portage.

Rule-Based Machine Translation

System features:

- The use of two distinct phrase tables, containing phrase pairs extracted from the Europarl and the News Commentary training corpora respectively.
- Multiple phrase-probability feature functions in the log-linear models, including a joint probability estimate, a standard frequency-based conditional probability estimate, and variants thereof based on different smoothing methods.
- A 4-gram language model trained on the combined Europarl and News Commentary target language quality corpora.
- A 3-gram adapted language model: this is trained on a mini-corpus of test-relevant target language sentences, extracted from the training material using standard information retrieval techniques.
- A 5-gram true casing model, trained on the combined Europarl and News Commentary target-language corpora.

Rule-Based Machine Translation

Table: System performance on WMT-06 test (*test2006* for the Europarl domain and *nc-devtest2007* for the News Commentary domain).

	en → fr	fr → en
Europarl (> 32M words/language)		
Systran	23.06	20.11
Portage *translation*	31.01	**30.90**
Systran + Portage	**31.11**	30.61
News Commentary (1M words/language)		
Systran	24.41	18.09
Portage *translation*	25.98	25.17
Systran + Portage	**28.80**	**26.79**

Rule-Based Machine Translation

Statistical Post-Editing of a Rule-Based Machine Translation System (Lagarda et al., 2009)

- *Corpora*: Parliament and Protocols.
- *Content*: Proceedings of parliamentary sessions; medical protocols.
- *Language pairs*: English-Spanish.
- *RBMT system*: Unknown.
- *SMT system*: Moses.

Rule-Based Machine Translation

Table: Automatic evaluation for *Parliament* and *Protocols* tests.

	Parliament		*Protocols*	
	BLEU	TER	BLEU	TER
RBMT	29.1	46.7	29.5	48.0
SMT	**49.9**	**34.9**	22.4	59.6
APE	48.4	35.9	**33.6**	**46.2**

Table of Contents

Domain Adaptation

Aim

Adapt an SMT system trained for another domain (out-of-domain) in order to improve the translation quality for the new task (in-domain). This is achieve by training a new SMT system which takes the previous SMT system's output as input.

Domain Adaptation

Canadian Job Bank (Simard, Goutte, & Isabelle, 2007)

- *Content*: Job ads.
- *Out-of-domain corpus*: Canadian Hansard.
- *Language pairs*: English-French, French-English.
- *SMT system*: Portage.
- *Goal*: Reduce human post-editing effort.

Domain Adaptation

Table: Portage translation - Portage APE system combination experimental results.

Language	TER	BLEU
English-to-French		
Portage *Job Bank*	**53.7**	**36.0**
Portage *Hansard*	76.9	13.0
+ Portage APE	64.6	26.2
French-to-English		
Portage *Job Bank*	**43.9**	**41.0**
Portage *Hansard*	80.1	14.0
+ Portage APE	57.7	28.6

Domain Adaptation

EMEA (Rubino et al., 2012)

- *Out-of-domain corpora*: Europarl v6, United Nations, News Commentary v6 (monolingual) and Shuffled News from 2007 to 2011 (monolingual).
- *Language pair*: French-English.
- *SMT systems*: Moses and Google Translate (*com*).

Domain Adaptation

Approach:

- Language model build on monolingual out-of-domain data (LM_g).
- Language model build on in-domain data (LM_m).
- Lineal interpolation of previous language models (LM_{g+m}).
- Translation model build from the out-of-domain data (TM_g).
- Translation model build from the in-domain data (TM_m).
- Translation model build from all the parallel corpora (TM_{g+m}).
- Standard SPE.
- Use of an oracle[1] to decide which lines to post-edit.

[1] A classifier build with Support Vector Machines (SVM) based on a linear kernel.

Domain Adaptation

Table: BLEU score of the different SMT systems when translating the test corpus from the medical domain.

SMT system	BLEU (*oracle*)
TM_g LM_g	29.9
+ SPE_m LM_m	43.4 (44.2)
+ SPE_m LM_{g+m}	**45.6 (47.0)**
TM_g LM_{g+m}	38.2
TM_g LM_m	39.2
+ SPE_m LM_m	42.7 (44.2)
+ SPE_m LM_{g+m}	42.5 (44.4)
com	44.9
+ SPE_m LM_m	46.8 (53.3)
+ SPE_m LM_{g+m}	**47.9 (53.5)**
TM_m LM_m	46.4
TM_{g+m} LM_m	47.2
TM_{g+m} LM_{g+m}	47.3

Table of Contents

Phrase-Based Machine Translation

Aim

Improve the translation quality of an SMT system by means of SPE. This is achieve by training a new SMT system which takes the previous SMT system's output as input.

Phrase-Based Machine Translation

Canadian Job Bank (Simard, Goutte, & Isabelle, 2007)

- *Content*: Job ads.
- *Language pairs*: English-French, French-English.
- *SMT system*: Portage.
- *Goal*: Reduce human post-editing effort.

Phrase-Based Machine Translation

Table: Portage translation - Portage APE system combination experimental results.

Language	TER	BLEU
English-to-French		
Portage *Job Bank*	**53.7**	36.0
+ Portage APE	**53.7**	**36.2**
French-to-English		
Portage *Job Bank*	**43.9**	41.0
+ Portage APE	**43.9**	**41.4**

Phrase-Based Machine Translation

Symantec (Béchara et al., 2011)

- *Content*: Technical software user help information.
- *Language pairs*: English-French, French-English.
- *SMT system*: Moses.

Phrase-Based Machine Translation

Approach:

- Standard SPE (PE).
- Context-aware PE (PE-C).
- Context-aware PE filtering out "#f tags" in the output (PE-CF).
- Context-aware PE filtering context information by thresholding word alignment strengths (PE-CF′).

Phrase-Based Machine Translation

Table: Experimental results (BLEU).

Language pair	Baseline	PE	PE-C	PE-CF	PE-CF′
English-to-French	**60.30**	60.15	46.89	58.55	**60.30**[2]
French-to-English	61.60	62.25	57.33	61.36	**63.89**[3]

[2] Obtained with a word alignment level of 0.7.

[3] Obtained with a word alignment level of 0.8.

Human Evaluation Experiment

Symantec II (Béchara et al., 2012)

- *Content*: Technical software user help information.
- *Language pairs*: French-English.
- *RBMT system*: Systran.
- *SMT system*: Moses.

Human Evaluation Experiment

Table: BLEU and TER scores for the RBMT, SMT and SPE systems.

	RBMT	SMT	RBMT+SPE	SMT+SPE
BLEU	23.26	**65.43**	64.63	65.14
TER	61.07	**23.92**	24.62	24.12

Human Evaluation Experiment

Evaluation task:

- 10 different translators from different backgrounds.
- Source sentence + 2 possible translations (from different systems).
- 200 sentences per translator.
- For each sentence they have either to choose the best translation or mark them as equal.

Human Evaluation Experiment

Table: Number of sentences chosen as "best" by each of the evaluations.

	Human evaluation	S-BLEU	TER
SMT vs RBMT			
SMT	**97**	**162**	**161**
RBMT	52	16	9
Tie	51	26	30
SMT vs RBMT+SPE			
SMT	28	**125**	**124**
RBMT+SPE	40	50	46
Tie	**132**	25	30
RBMT vs RBMT+SPE			
RBMT	40	16	11
RBMT+SPE	**99**	**162**	**162**
Tie	61	22	26
SMT+SPE vs RBMT			
SMT+SPE	**107**	**167**	**161**
RBMT	46	49	41
Tie	47	25	30
SMT+SPE vs RBMT+SPE			
SMT+SPE	27	46	46
RBMT+SPE	47	49	41
Tie	**126**	**105**	**113**

Human Evaluation Experiment

Statistical Post-Editing of a Rule-Based Machine Translation System (Lagarda et al., 2009)

- *Corpora*: Parliament and Protocols.
- *Content*: Proceedings of parliamentary sessions; medical protocols.
- *Language pairs*: English-Spanish.
- *RBMT system*: Unknown.
- *SMT system*: Moses.

Human Evaluation Experiment

Table: Automatic evaluation for *Parliament* and *Protocols* tests.

	Parliament		*Protocols*	
	BLEU	TER	BLEU	TER
RBMT	29.1	46.7	29.5	48.0
SMT	**49.9**	**34.9**	22.4	59.6
APE	48.4	35.9	**33.6**	**46.2**

Human Evaluation Experiment

Evaluation task:

- 2 professional evaluators.
- Analysis of the suitability of the output of each system.

Human Evaluation Experiment

Table: Human evaluation for *Parliament* and *Protocols* corpora. Percentage of suitable translated sentences for each system.

	Parliament	*Protocols*
RBMT	58	60
SMT	60	-
APE	**94**	**67**

Table of Contents

Conclusions

RBMT:

- SPE improves RBMT's translation quality.
- Its quality is not always better than translating directly with an SMT system.

Domain Adaptation:

- SPE is good for adapting from one domain to another (improving translation quality).
- Can fail if domains are too specific and unrelated.

PBMT:

- SPE doesn't seem to be suitable for this task.

References

Béchara, H., Rubino, R., He, Y., Ma, Y., & Genabith, J. (2012). An evaluation of statistical post-editing systems applied to rbmt and smt systems. In *Proceedings of coling 2012* (pp. 215–230).

Béchara, H., Ma, Y., & Genabith, J. (2011). Statistical post-editing for a statistical mt system. In *Proceedings of the mt summit xiii* (pp. 308–315).

Isabelle, P., Goutte, C., & Simard, M. (2007). Domain adaptation of mt systems through automatic post-editing. In *Proceedings of the mt summit x* (pp. 10–14).

Lagarda, A. L., Alabau, V., Casacuberta, F., Silva, R., & Dìaz-de Liaño, D. (2009). Statistical post-editing of a rule-based machine translation system. In *Proceedings of the north american chapter of the association for computational linguistics - human language technologies (naacl hlt) 2009* (pp. 217–220).

Rubino, R., Huet, S., Lefèvre, F., & Lenarés, G. (2012). Statistical post-editing of machine translation for domain adaptation. In *Proceedings of the european association for machine translation* (pp. 221–228).

References

Simard, M., Goutte, C., & Isabelle, P. (2007). Statistical phrase-based post-editing. In *Proceedings of naacl* (pp. 508–515).

Simard, M., Ueffing, N., & Isabelle, P. (2007). Rule-based translation with statistical phrase-based post-editing. In *Proceedings of the acl-2007 workshop on statistical machine translation (wmt-07)* (pp. 203–206).