

Computer Assisted Translation with Neural Quality Estimation and Automatic Post-Editing

Ke Wang*, Jiayi Wang*, Niyu Ge, Yangbin Shi, Yu Zhao, Kai Fan†

Alibaba Group Inc.

{moyu.wk, joanne.wjy, niyu.ge, taiwu.syb}@alibaba-inc.com,
kongyu@taobao.com, k.fan@alibaba-inc.com

A Appendix

A.1 Pseudo code of QE tag computation

The computation of QE tags is quite similar to the famous Minimum Edit Distance problem and can be solved with dynamic programming in algorithm 1.

Algorithm 1 QE tag computation

Require: machine translation $\mathbf{m} = \{m_i\}_{i=1}^M$, post-editing $\mathbf{e} = \{e_i\}_{i=1}^N$.

```
1: Initialize the edit distance matrix  $d_{i,0} = i, d_{0,j} = j$  and QE tag  $q_i = 1$ .
2: for  $i = 1, \dots, M$  do
3:   for  $j = 1, \dots, N$  do
4:      $d_{i,j} = \min\{d_{i-1,j-1} + \mathbb{I}_{m_i \neq e_j}, d_{i,j-1} + 1, d_{i-1,j} + 1\}$ 
5:   end for
6: end for
7: while  $i > 0$  or  $j > 0$  do
8:   if  $i > 0$  and  $j > 0$  and  $d_{i-1,j-1} + 1 = d_{i,j}$  then
9:      $q_i = -1, i--, j--$ 
10:  else if  $j > 0$  and  $d_{i,j-1} + 1 = d_{i,j}$  then
11:     $q_i = +, j--$ 
12:  else if  $i > 0$  and  $d_{i-1,j} + 1 = d_{i,j}$  then
13:     $q_i = 0, i--$ 
14:  else
15:     $i--, j--$ 
16:  end if
17: end while
18: return  $\mathbf{q} = \{q_i\}_{i=1}^M$ 
```

A.2 Details of the Training Corpus

WMT APE shared-task provided both real APE triplets and a large a large-scale artificial synthetic corpus containing around 500K high quality and 4 million low quality synthetic triples. Table 1 shows the difference between them.

A.3 Case Study and Runtime Efficiency

As mentioned in the paper, the AOM is more suitable for translations that only require a few edit

Table 1: Details of the WMT 2017 APE Shared-Task Dataset. The BLEU and TER metrics are directly evaluated on machine translation and post-editions as references.

Source	# Sentence	Avg. Length	BLEU	TER
Real Triples	23,000	17.88	61.87	25.35
Artificial 500K	526,368	20.90	60.01	25.55
Artificial 4M	4,391,180	16.68	46.59	35.37
500K+20*Real	986,368	19.49	60.80	25.46
4M+500K+20*Real (Full Training data)	5,377,548	17.20	49.65	33.31

operations while GM is more preferable for low quality translations. To demonstrate this conclusion and prove the effectiveness of our QE-based automatic selector, some cases of translations with different qualities are shown in Table 2.

In case 1 and case 2, the translation is quite close to *pe*. Therefore, the AOM only need to predict tokens for a small number of [PLH]s. When there are relatively complete contexts provided, the AOM can achieve a higher performance than the GM. Moreover, after reading the source and the final output, the human translators did not even take any additional action to improve the translation quality.

In the opposite way, as shown in case 3 and case 4, there is a huge gap between *mt* and *pe*, and the input for AOM contains a considerable number of placeholders, which lacks enough contextual information. In these cases, our GM can auto-regressively regenerate the translation based on the given *mt* to guarantee the higher quality of the final output. Based on the QE selector, the translators only need to make very few efforts to correct the errors in the final generated APE of our model.

A practical point of the computer assisted translation via APE is its expense and computational cost. Compared with the traditional computer as-

* indicates equal contribution.

† indicates corresponding author.

Table 2: Examples of Crowdsourcing after APE. Tokens in “⟨⟩” indicates GM’s over corrections or AOM’s inaccurate translations due to too many consecutive [PLH] predictions, which leads inadequate contextual information. Tokens in “{}” highlights correct automatic editings.

High Quality s Translation Case		
	SRC	In List view , click any column header to sort by that criteria .
	MT	Klicken Sie in der Listenansicht auf eine beliebige Spaltenüberschrift , um nach dieser Kriterien sortieren .
	PE	Klicken Sie in der Listenansicht auf eine beliebige Spaltenüberschrift , um nach diesen Kriterien zu sortieren .
	MT (sub-word)	..klicken ..Sie ..in ..der ..Listenansicht ..auf ..eine ..beliebige ..Spalten überschrift .., ..um ..nach ..dieser ..Kriterien ..sortieren ...
Case1	Predicted QE Tag	1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1
	TER vs Predicted TER	11.76 vs 11.11
	AOM Input	..klicken ..Sie ..in ..der ..Listenansicht ..auf ..eine ..beliebige ..Spalten überschrift .., ..um ..nach [PLH] ..Kriterien [PLH] ..sortieren ...
	AOM Output	..klicken ..Sie ..in ..der ..Listenansicht ..auf ..eine ..beliebige ..Spalten überschrift .., ..um ..nach {..diesen} ..Kriterien {..zu} ..sortieren ...
	GM Output	..klicken ..Sie ..in ..der ..Listenansicht ..auf ..eine ..beliebige ..Spalten überschrift .., ..um ..nach ..dieser ..Kriterien {..zu} ..sortieren ...
	Final Output	Klicken Sie in der Listenansicht auf eine beliebige Spaltenüberschrift , um nach diesen Kriterien zu sortieren .
	Translator Edit	no action
High Quality s Translation Case		
	SRC	You can justify all text in a paragraph either including or excluding the last line .
	MT	Sie können den gesamten Text eines Absatzes mit oder ohne die letzte Zeile .
	PE	Sie können den gesamten Text eines Absatzes mit oder ohne die letzte Zeile ausrichten .
	MT (sub-word)	..Sie ..können ..den ..gesamten ..Text ..eines ..Absatzes ..mit ..oder ..ohne ..die ..letzte ..Zeile ...
	Predicted QE Tag	1 1 1 1 1 1 1 1 1 1 1 1 2 1
Case2	TER vs Predicted TER	6.67 vs 6.67
	AOM Input	..Sie ..können ..den ..gesamten ..Text ..eines ..Absatzes ..mit ..oder ..ohne ..die ..letzte ..Zeile [PLH] ...
	AOM Output	..Sie ..können ..den ..gesamten ..Text ..eines ..Absatzes ..mit ..oder ..ohne ..die ..letzte ..Zeile {..ausrichten} ...
	GM Output	..Sie ..können ..den ..gesamten ..Text ..eines ..Absatzes ⟨..entweder ..einschließlich⟩ ..oder ..ohne ..die ..letzte ..Zeile ..löschen ...
	Final Output	Sie können den gesamten Text eines Absatzes mit oder ohne die letzte Zeile ausrichten .
	Translator Edit	no action
Low Quality Translation Case		
	SRC	In Start Number , enter the number to assign to the first PDF on the list .
	MT	Wählen Sie unter “ Number , ” geben Sie die Nummer für die erste PDF-Datei in der Liste aus .
	PE	Geben Sie unter “ Startnummer ” die Nummer für die erste PDF-Datei in der Liste ein .
	MT (sub-word)	..wählen ..Sie ..unter ..“ ..Number .., ..” ..geben ..Sie ..die ..Nummer ..für ..die ..erste ..PDF - Datei ..in ..der ..Liste ..aus ...
Case3	Predicted QE Tag	-1 1 1 2 -1 -1 -1 -1 -1 1 1 1 0 -1 -1 1 1 1 1 -1 1 1 1 -1
	TER vs Predicted TER	35.29 vs 54.55
	AOM Input	[PLH] ..Sie ..unter ..“ [PLH] [PLH] [PLH] [PLH] [PLH] [PLH] [PLH] ..die ..Nummer [PLH] [PLH] ..PDF - Datei [PLH] ..der ..Liste [PLH] ...
	AOM Output	{..geben} ..Sie ..unter ..“ ..Start ⟨..geben ..Sie ..zum ..Zuweisen⟩” ..die ..Nummer ..der ..ersten ..PDF - Datei ..über ..der ..Liste {..ein} ...
	GM Output	{..geben} ..Sie ..unter ..“ {..Start nummer} ..” ..die ..Nummer ..für ..die ..erste ..PDF - Datei ..in ..der ..Liste ..an ...
	Final Output	Geben Sie unter “ Startnummer ” die Nummer für die erste PDF-Datei in der Liste an .
	Translator Edit	an→ein
Low Quality Translation Case		
	SRC	The Illustrator text is converted to HTML text with basic formatting attributes in the resulting web page .
	MT	Die Illustrator Text HTML-Text mit grundlegenden Formatierungsattribute in der erstellten Webseite konvertiert wird .
	PE	Die Illustrator-Text wird in HTML-Text mit grundlegenden Formatierungsattributen in der erstellten Webseite konvertiert .
	MT (sub-word)	..die ..Illustrator ..Text ..HTML - Text ..mit ..grundlegenden ..Formatierung s attribute ..in ..der ..erstellten ..Webseite ..konvertiert ..wird ...
Case4	Predicted QE Tag	-1 3 3 1 1 1 1 1 1 1 1 -1 1 1 1 1 1 0 1
	TER vs Predicted TER	35.29 vs 33.33
	AOM Input	[PLH] ..Illustrator [PLH] [PLH] ..Text [PLH] [PLH] ..HTML - Text ..mit ..grundlegenden ..Formatierung s [PLH] ..in ..der ..erstellten ..Webseite ..konvertiert ...
	AOM Output	..in ..Illustrator - Der ..Text ..in ..in ..HTML - Text ..mit ..grundlegenden ..Formatierung s {..attributen} ..in ..der ..erstellten ..Webseite ..konvertiert ...
	GM Output	..der ..Illustrator {- Text ..wird ..in} ..HTML - Text ..mit ..grundlegenden ..Formatierung s {..attributen} ..in ..der ..erstellten ..Webseite ..konvertiert ...
	Final Output	Der Illustrator- Text wird in HTML-Text mit grundlegenden Formatierungsattributen in der erstellten Webseite konvertiert .
	Translator Edit	Der→Die

sisted translation crowdsourcing, machine translation + human post-editing, our additional automatic post-editing does increase the computational cost, which is roughly equivalent to another machine translation model. In general, the crowdsourcing is charged by hours. The numbers in our findings suggest a promising budget cut associated with CAT crowdsourcing. However, this extra APE module may lead to a latency increase by ~ 400 ms, which is still far below the average time cost by human post-editing. Even for an online crowdsourcing system, a well-designed concurrent mechanism should make the translators not feel any delay. From the perspective of architecture scale, the APE model can be deployed in the identical processing unit for the machine translation model and be called successively in a pipeline. The only concern is that the memory storage capacity should be large enough to store more parameters.