PBMT vs NMT: Which helps translators the most?

Alessandro Cattelan - alessandro@matecat.com
Who am I?

VP Operations at Translated

Product Manager at MateCat

Project Coordinator at ModernMT
Overview

• Phrase-Based Machine Translation vs Neural Machine Translation

• Some key differences (training data, phrases vs sentences, generalisation vs specialisation, learning vs memorisation)

• Analysis (error annotation, quality rating, post-editing effort)

• Conclusions (how to use the data)
PBMT vs NMT
Phrase-Based Machine Translation
Phrase-Based Machine Translation

In a PBMT system, the building blocks (phrases) of a sentence are deconstructed and their translations are recombined to form a new sentence in the target language.
Phrase extraction

The Bologna Process has transformed the face of European higher education
Il Processo di Bologna ha cambiato il volto dell’ istruzione superiore in Europa

Credits: Marcello Federico, Head of HLT-MT Unit at FBK

#LocWorld35
Translation and Language Model

Phrases from parallel texts are stored in a translation model and retrieved as if they were units from a translation memory.

Phrases are recombined to form the target sentence and reordered based on the samples in the language model.
Translation and reordering

Credits: Marcello Federico, Head of HLT-MT Unit at FBK

#LocWorld35
Neural Machine Translation
Neural Machine Translation

PBMT memorises translation fragments and uses them as building blocks to compose new translations.

NMT learns and performs translation through an encoding-decoding process which converts source words into a numeric representation, from which it then generates the corresponding target words.
Phrases to Gradients

Tomorrow I will fly to the conference in Italy

Morgen fliege ich nach Italien zur Konferenz

Credits: Marcello Federico, Head of HLT-MT Unit at FBK

#LocWorld35
Encoding - Decoding

- Input: One word at a time
- "Machine", "Learning", "is", "Fun"

- Stateful Model
- First Recurrent Neural Network (Encoder)

- Encoded sentence
- 0.636, 0.122, 0.981

- Stateful Model
- Second Recurrent Neural Network (Decoder)

- Output: One word at a time
- "Aprendizaje", "automático", "es", "divertido"

Credits: Adam Geitgey -
Feed-Forward Neural Networks

Credits: Marcello Federico, Head of HLT-MT Unit at FBK

#LocWorld35
Key Differences
Training Data

PBMT:
- Monolingual data: 2-10B words
- Parallel data: 1B

NMT:
- Monolingual data: no need
- Parallel data: 100M (higher quality)
- Impossible to retrieve content from models (privacy)
Phrases vs Sentences

**PBMT** breaks the original text into phrases and retrieves their translations from the translation model, looking first for longer phrases and then shorter.

**NMT** takes in entire original sentences and encodes them into a numeric representation and decodes this representation into a target sentence.
Generalisation vs Specialisation

**NMT** outperforms PBMT in most cases, especially on texts that differ from the training data.

**PBMT** still is more effective when the content to translate is similar to the training data (e.g. custom engines for a specific product or customer).
Learning vs Memorization

PBMT and NMT learn differently

NMT tends to be better than PBMT on **new** inputs.

The opposite happens on inputs very similar to training examples.

Credits: [Marcello Federico](https://example.com), Head of HLT-MT Unit at FBK

#LocWorld35
Analysis
Data Collection

Data from projects translated by Translated.net in MateCat.

MT system used is Google Translate (API) both for PBMT and NMT.

Data collected before and after April 2017 (when GNMT was introduced).
Language Pairs

- German to English
- English to French
- English to Italian
- English to German
- English to Portuguese
- English to Spanish
Methodology

- Qualitative Analysis
  - Error Annotation
  - Quality Rating

- Quantitative Analysis
  - Post-Editing Effort
Error Annotation
Error Annotation

- Task: Annotate errors found in the raw MT output from PBMT and NMT
- 100 segments per language pair
- Randomised data so that the translators didn’t know which system generated the translation
- Four translators per language pair, each annotated all segments
- Seven error categories
## Error Annotation - Categories

<table>
<thead>
<tr>
<th>Category</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grammar</td>
<td>Any issues which affect language quality (e.g. morphology, word order, concordance etc.).</td>
</tr>
<tr>
<td>Mistranslation</td>
<td>The translation does not carry the same meaning as the source sentence.</td>
</tr>
<tr>
<td>Omission</td>
<td>Information from the original sentence is missing in the translation.</td>
</tr>
<tr>
<td>Spelling</td>
<td>Spelling or typographical errors.</td>
</tr>
<tr>
<td>Style</td>
<td>Linguistic issues which make the sentence sound awkward in the target language.</td>
</tr>
<tr>
<td>Terminology</td>
<td>Translation is correct, but the terminology is not adequate for the context.</td>
</tr>
<tr>
<td>None</td>
<td>No errors detected.</td>
</tr>
</tbody>
</table>
Error Annotation
Results by language pair
Error Annotation
Overall
Quality Rating
Quality Rating

- Task: Select which raw MT output is easier to post-edit to get to a high quality translation.
- 100 segments per language pair
- Randomised data so that the translators didn’t know which system generated the translation
- Four translators per language pair, each evaluated all segments
Which is easier to post-edit?

- PMBT
- NMT
- Same quality

Occurrences

EN-FR  EN-DE  EN-ES  EN-IT  EN-PTBR  DE-EN

#LocWorld35
Post-Editing Effort
Post-Editing Effort - A definition

Post-Editing Effort: the percentage of edits required to modify the suggestions provided by the MT system in order to get to a good quality translation.
Post-Editing Effort

Post-Editing Effort is calculated by comparing the suggestion from the MT system with the final translation.

The function used to calculate the post-editing effort compares the average number of steps required to edit the suggestion in order to produce a professional translation. These steps could either be changing synonyms, correcting numbers or casing, adjusting punctuation, changing the tag positions, etc.
Post-Editing Effort

Changing synonyms, correcting numbers or casing, adjusting punctuation, changing the tag positions, etc. all have different weights which have been calculated to reflect the effort required to correct each of these issues.
# Post-Editing Effort - Examples

<table>
<thead>
<tr>
<th>Suggestion</th>
<th>Translation</th>
<th>Post-Editing Effort</th>
</tr>
</thead>
<tbody>
<tr>
<td>hi!</td>
<td>Hi!</td>
<td>2%</td>
</tr>
<tr>
<td>hi all!</td>
<td>Hi all!</td>
<td>2%</td>
</tr>
<tr>
<td>tests</td>
<td>experiments</td>
<td>100%</td>
</tr>
<tr>
<td>Long tests</td>
<td>Long experiments</td>
<td>50%</td>
</tr>
</tbody>
</table>
Post-Editing Effort & Productivity

Post-Editing Effort shows a good inverse correlation with translators’ productivity. The better the MT quality, the lower the post-editing effort and the higher the throughput of translators.
Post-Editing Effort - Analysis

We collected data on the Post-Editing Effort for the six language pairs over a period of 18 months and noted the impact due to the introduction of GNMT.
What do I need this for?
Negotiating Rates

Transparency on the actual benefits of MT help set a level playing field with clients and vendors.

Our experience: Clients, account/project managers and translators work on the same platform and have access to the same data on productivity. This makes it easier to negotiate rates.
Post-Editing Effort correlates with the daily throughput of translators.

Our experience: We use post-editing effort together with time to edit to estimate the daily productivity of translators and use that to dynamically assess the required TAT.
Increasing Quality

Machine translation can boost productivity but may lead translators to output sub-optimal translations. Post-editing effort helps to evaluate the final quality.

Our experience: Post-editing effort is one of the quality metrics that we use to identify good translators. Translators with higher post-editing effort rates are preferred.
Thank you

Alessandro Cattelan - alessandro@matecat.com