An Overview of Machine Translation and its Applications

Gao Zhao-Ming
National Taiwan University
Overview

• Early years in MT
• Rule-based MT
• Statistical MT
• Neural MT
• Speech-to-speech translation
• Applications of MT
Early years in MT

• The first proposal of MT was proposed by Warren Weaver based on information theory in 1949.
• The first public demonstration of an MT system was in 1954, which was in fact of a toy system of 250 words.
• The publication of the ALPAC report in 1966 gave a serious blow to MT, because it concluded that ‘fully automatic high-quality machine translation’ was not possible.
Rule-based MT

Three typical procedures of rule-based MT are analysis, transfer, and synthesis.
The mainstream approach between 1970s and mid 1990s.
Relies heavily on bilingual lexicon and bilingual transfer rules encoded by linguists.
Makes use of parsers, morphological analyzers as well as other software for processing linguistic information.
Statistical MT

- First proposed by Peter Brown and his colleagues at I.B.M. in 1988.
- Requires large sentence-aligned parallel corpora. SMT is the result of machine learning. The training data are sentence-aligned parallel corpora.
- Use statistical models to estimate the maximum likelihood of a sentence in the target language given an input sentence in the source language.
- Requires a large target language corpus to derive a target language model for fluency and a large parallel corpus to derive a translation model for bilingual correspondences.
- Does not require a lexicon or transfer rules encoded by linguists.
- Algorithms for statistical MT are based on mathematics and thus largely language-independent.
Model 3

The fertility problem is addressed in IBM Model 3. The fertility is modeled using probability distribution defined as:

\[ n(\phi \vee f) \]

For each foreign word \( j \), such distribution indicates how many output words \( \phi \) it usually translates. This model deals with dropping input words because it allows \( \phi = 0 \). But there is still an issue when adding words. For example, the English word *do* is often inserted when negating. This issue generates a special *NULL* token that can also have its fertility modeled using a conditional distribution defined as:

\[ n(\varnothing \vee NULL) \]

The number of inserted words depends on sentence length. This is why the NULL token insertion is modeled as an additional step: the fertility step. It increases the IBM Model 3 translation process to four steps:

1. *ja nie pójdę do domu*
2. *ja nie pójdę do domu*
3. *I do not go to the house*
4. *I do not go to the house*

The last step is called distortion instead of alignment because it is possible to produce the same translation with the same alignment in different ways.\(^{[5]}\)

IBM Model 3 can be mathematically expressed as:

\[
P(S \mid E, A) = \prod_{i=1}^{I} \Phi_i \ln(\Phi_i \mid \epsilon_{a_i}) \cdot \prod_{j=1}^{J} t(f_j \mid \epsilon_{e_j}) \cdot \prod_{j=1}^{J} \prod_{j \neq 0} d(j \mid a_j, I, J) \cdot (1 - \Phi_0)^{p_0} p_0^{I+J}\]

where \( \Phi_i \) represents the fertility of \( \epsilon_{a_i} \), each source word \( a \) is assigned a fertility distribution \( n \), and \( I \) and \( J \) refer to the absolute lengths of the target and source sentences, respectively.\(^{[6]}\)
Moses is the most popular open source toolkit for statistical machine translation (SMT).
Neural MT

1. Like SMT, neural MT is the result of machine learning and the training data are sentence-aligned parallel corpora.
2. Neural MT differs from statistical MT in that it is based on deep neural network (i.e. deep learning).
3. Unlike SMT, neural MT does not create a phrase table (which in fact involves correspondences of bilingual ngrams rather than parsed bilingual phrases)
4. GPUs are required to run Neural MT systems.
5. Neural MT has the potential of tackling structures that SMT is difficult to deal with.
Google Translate can translate sentences and phrases although it does require parsers and linguistic knowledge.
Using deep learning, Google Translate can now tackle many constructions involving relative clauses.
OpenNMT is an open source toolkit for neural MT.
Applications of Neural MT

• The mainstream of Neural MT is based on sequence-to-sequence model using recurrent neural net (RNN), which can also be used in developing a ChatBot.

• The training of a ChatBot is similar to developing a neural MT. Both require an input string and output string. Both approaches require large training data.
Neural MT and Automatic captioning of an image also share many features in common.

When the model is presented with scenes similar to what it’s seen before, it will often re-use human generated captions.

So does it really understand the objects and their interactions in each image? Or does it always...
Comparisons of Statistical MT and Neuro MT

• Both require large sentence-aligned parallel corpora to work.
• In building statistical MT, we can access word alignment probabilities and phrase table, which are absent in neural MT.
• Neural MT can tackle constructions that are difficult in statistical MT, such as relative construction.
Many forms of violence and hardship people as the country’s civil war has. But the Syrian government’s attack pointed to one of the war’s irrefutable deliberate targeting of civilians by President Assad’s military, in this case with a weapon impossible to use precisely.

Syrians on both sides in this fight have bloodshed and sectarian fury given war. The victims of the cluster bomb tactic as collective punishment, a war crime that is against the rebels.

The munitions in question — Soviet-era bombs that drop indiscriminately.

Cluster Bombs in Syria
Syrian government jets have been targeting civilian areas with Soviet-era bombs that drop indiscriminately.
百度翻译

<table>
<thead>
<tr>
<th>自动检测</th>
<th>中文</th>
<th>英语</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABC</td>
<td>DEFG</td>
<td>HUKLMN</td>
</tr>
<tr>
<td>阿拉伯语</td>
<td>丹麦语</td>
<td>韩语</td>
</tr>
<tr>
<td>索比亚尼亚语</td>
<td>德语</td>
<td>荷兰语</td>
</tr>
<tr>
<td>保加利亚语</td>
<td>俄语</td>
<td>捷克语</td>
</tr>
<tr>
<td>波兰语</td>
<td>法语</td>
<td>罗马尼亚语</td>
</tr>
<tr>
<td>芬兰语</td>
<td>葡萄牙语</td>
<td>西班牙语</td>
</tr>
<tr>
<td>文言文</td>
<td>日语</td>
<td>希腊语</td>
</tr>
<tr>
<td>中文</td>
<td>西班牙语</td>
<td>意大利语</td>
</tr>
<tr>
<td>英语</td>
<td>希腊语</td>
<td>越南语</td>
</tr>
<tr>
<td>粤语</td>
<td>中文繁体</td>
<td></td>
</tr>
</tbody>
</table>
百度翻译（繁簡體中文轉換）

我等一下要先搭捷運再換公車。
百度翻譯（文言文翻百話文）

人非生而知之者，孰能無惑？惑而不彊師，於為惑也終不解矣！

您輸入的可能是：中文
Speech-to-speech translation

- Components of Speech-to-speech translation
  - 1. Automatic speech recognition (ASR): ASR takes speech as input and outputs its text. The mainstream approach of ASR is deep learning.
  - 2. MT: MT takes a text in one language as input and converts it into another language.
  - 3. Text-to-speech: Text-to-speech systems take text as input and output speech.
Skype Translator is a speech-to-speech translation system.
Applications of MT

- The tools used in statistical MT and neural MT can also be used in many other applications.
Sentence alignment

- Sentence alignment is the input data for SMT and neural MT.
- The parallel corpora developed for SMT and neural MT are valuable resources not only for MT but also for computer-aided translation and contrastive linguistics (cf. Opus Corpus project).
- Sentence alignment can turn a parallel corpus into a bilingual concordancer or translation memory, which are important tools for computer-aided translation and contrastive linguistics.
The output of Linguee bilingual search is based on sentence alignment.
Word alignment

- Word alignment is normally based on the input of sentence-aligned parallel corpora. The output of word alignment is the probability of a word or a sequence of words in one language corresponding to a word or a sequence of words.
- With word alignment, we can automatically derive a bilingual lexicon.
- Word alignment can also be used to compare the dialect differences at word level.
- In corpus-based translation study, the techniques of word alignment can be used to compare the preferences of lexical translations of one or more than one translators.
Automatic highlighting of translation equivalents using word alignment in the TotalRecall e-Learning Project at Tsing Hua U.

A careful look at statistics for trade between the two countries, however, shows that Taiwan's deficit is not quite what it seems.

Going back to what was said earlier, though the Japanese firms can do the job both well and cheaply, the longstanding trade imbalance between Taiwan and Japan surpassed US$12 billion last year, so it is really not possible to just sit and watch Japanese firms monopolizing the engineering projects in Taiwan. Further, the Japanese have always been very reserved about technology transfer.
The TAUS search is empowered by word alignment.
We can automatically identify the lexical divergences between Taiwan and China based on word alignment probabilities using TAUS data.

<table>
<thead>
<tr>
<th></th>
<th>台灣譯名</th>
<th>中國大陸譯名</th>
</tr>
</thead>
<tbody>
<tr>
<td>mouse</td>
<td>滑鼠: 0.9924</td>
<td>鼠标: 0.990767</td>
</tr>
<tr>
<td>cursor</td>
<td>游標: 0.998853</td>
<td>光标: 0.949564</td>
</tr>
<tr>
<td>network</td>
<td>網路: 0.616029</td>
<td>网络: 0.85752</td>
</tr>
<tr>
<td>video</td>
<td>視訊: 0.970871</td>
<td>视频: 0.888748</td>
</tr>
<tr>
<td>server</td>
<td>伺服器: 0.850816</td>
<td>服务器: 0.870501</td>
</tr>
<tr>
<td>image</td>
<td>影像: 0.999</td>
<td>图像: 0.766613</td>
</tr>
<tr>
<td>device</td>
<td>裝置: 0.783774</td>
<td>设备: 0.821117</td>
</tr>
</tbody>
</table>
We can also use word alignment to identify different lexical translations of two Chinese translations of “the old man and the sea”.
Using comparable corpora for identification of bilingual phraseology

• Comparable corpora are more easily available than parallel corpora.
• It is possible to align some phrases and collocations in comparable corpora.
Using comparable corpora for computer-aided translation

1. It is very time-consuming to develop translation memory (TM) (i.e. bilingual sentences aligned at the sentence level).
   
   Chen (2012) proposes a novel method to develop TM from scratch. In order to find the frequently-occurring English expressions in contracts, Chinese contracts are compiled and automatically translated into English using Google Translate.

   A comparable authentic corpus of English contracts is manually compiled using keywords and ngrams.

   Fuzzy match function in a computer-aided translation system is used to find authentic sentences that are similar to the sentence in the machine-translated text.

   The similar sentences that are found in the authentic comparable corpus might contain translation equivalents of frequent expressions in contracts.
Trends for MT

- Using deep learning to tackle constructions that are difficult to deal with in SMT, e.g. relative clauses.
- Using comparable corpora for MT.
- Improving the accuracy of translations involving lexical ambiguity and structural ambiguity.
- Speech-to-speech translation such as Skype Translator.