

Example-based machine translation

Example-based machine translation (EBMT) is a method of machine translation often characterized by its use of a bilingual corpus with parallel texts as its main knowledge base at run-time. It is essentially a translation by analogy and can be viewed as an implementation of a case-based reasoning approach to machine learning.

Contents

Translation by analogy

History

Example

Phrasal verbs

See also

References

Further reading

External links

Translation by analogy

At the foundation of example-based machine translation is the idea of translation by analogy. When applied to the process of human translation, the idea that translation takes place by analogy is a rejection of the idea that people translate sentences by doing deep linguistic analysis. Instead, it is founded on the belief that people translate by first decomposing a sentence into certain phrases, then by translating these phrases, and finally by properly composing these fragments into one long sentence. Phrasal translations are translated by analogy to previous translations. The principle of translation by analogy is encoded to example-based machine translation through the example translations that are used to train such a system.

Other approaches to machine translation, including statistical machine translation, also use bilingual corpora to learn the process of translation.

History

Example-based machine translation was first suggested by Makoto Nagao in 1984.^[1] He pointed out that it is especially adapted to translation between two totally different languages, such as English and Japanese. In this case, one sentence can be translated into several well-structured sentences in another language, therefore, it is no use to do the deep linguistic analysis characteristic of rule-based machine translation.

Example

Example of bilingual corpus

English

Japanese

How much is that **red umbrella**? Ano **akai kasa** wa ikura desu ka.

How much is that **small camera**? Ano **chiisai kamera** wa ikura desu ka.

Example-based machine translation systems are trained from bilingual parallel corpora containing sentence pairs like the example shown in the table above. Sentence pairs contain sentences in one language with their translations into another. The particular example shows an example of a *minimal pair*, meaning that the sentences vary by just one element. These sentences make it simple to learn translations of portions of a sentence. For example, an example-based machine translation system would learn three units of translation from the above example:

1. *How much is that X* ? corresponds to *Ano X wa ikura desu ka*.
2. *red umbrella* corresponds to *akai kasa*
3. *small camera* corresponds to *chiisai kamera*

Composing these units can be used to produce novel translations in the future. For example, if we have been trained using some text containing the sentences:

President Kennedy was shot dead during the parade. and *The convict escaped on July 15th.* We could translate the sentence *The convict was shot dead during the parade.* by substituting the appropriate parts of the sentences.

Phrasal verbs

Example-based machine translation is best suited for sub-language phenomena like phrasal verbs. Phrasal verbs have highly context-dependent meanings. They are common in English, where they comprise a verb followed by an adverb and/or a preposition, which are called the particle to the verb. Phrasal verbs produce specialized context-specific meanings that may not be derived from the meaning of the constituents. There is almost always an ambiguity during word-to-word translation from source to the target language.

As an example, consider the phrasal verb "put on" and its Urdu/Hindi meaning. It may be used in any of the following ways:

- Ram put on the lights. (Switched on) (Urdu/Hindi translation: *Jalana*)
- Ram put on a cap. (Wear) (Urdu/Hindi translation: *Pahenna*)

See also

- Programming by example
- Translation memory
- Natural Language Processing

References

1. Makoto Nagao (1984). "A framework of a mechanical translation between Japanese and English by analogy principle". In A. Elithorn and R. Banerji. *Artificial and Human Intelligence* (<http://www.mt-archive.info/Nagao-1984.pdf>) (PDF). Elsevier Science Publishers.

Further reading

- Carl, Michael; Way, Andy (2003). *Recent Advances in Example-Based Machine Translation* (<https://www.springer.com/de/book/9781402014000#aboutBook>). Netherlands: Springer. doi:10.1007/978-94-010-0181-6 (<https://doi.org/10.1007%2F978-94-010-0181-6>). ISBN 978-1-4020-1400-0.

External links

- [Cunei \(http://sourceforge.net/projects/cunei/\)](http://sourceforge.net/projects/cunei/) - an open source platform for data-driven machine translation that grew out of research in EBMT, but also includes recent advances from the [SMT](#) field
-

Retrieved from "https://en.wikipedia.org/w/index.php?title=Example-based_machine_translation&oldid=850267990"

This page was last edited on 14 July 2018, at 20:13 (UTC).

Text is available under the [Creative Commons Attribution-ShareAlike License](#); additional terms may apply. By using this site, you agree to the [Terms of Use](#) and [Privacy Policy](#). Wikipedia® is a registered trademark of the [Wikimedia Foundation, Inc.](#), a non-profit organization.