Cultural Machine Translation – Challenges and Solutions*

Extended Abstract†

Fernand Cohen
ECE Dept., Drexel University
Philadelphia, PA 19104
USA
fscohen@coe.drexel.edu

ABSTRACT

Translation is not the mere act of literal conversion of a word from one language to another, but rather an intricate process of unravelling the inherent meaning that exist in the original text and of finding ways to faithfully express it through the host language [1]. The biggest challenge for human translators and even more so for machine translation (MT) is the ability to correctly understand the cultural nuances of what is written or said. Polysemy stands at the core of the misreading, misconception, and misinterpretation of a word, phrase and sentence. Sources of polysemy range from the use of slang, dialect, hegemony, stereotyping, and biases – factors that reside in cultural semantics. Hence for a successful MT, a multi-layered foundation that pays very close attention to cultural semantics, with the intricate interdependency of a word on its cultural context within its immediate and removed neighboring words and sentences properly captured needs to be laid out. Whereas MT has been a success story when dealing with simple texts such as in the menu in the coffee shop, weather report, road sign, etc., it is less so when dealing with complex and nuanced texts. There are there major thrusts in MT research dealing with word disambiguation, paraphrasing and alignment between bitexts. In this talk we concentrate on word/phrase/sentence disambiguation. We propose a language model (MRF Gibbs model [2]) that embeds a multi-layered key words dependencies built on a semantic graph to help lexically disambiguate word, phrases, and sentences that lend themselves to different possible meaning and interpretations. The model looks at semantics cliques of words (key words) and assigns Gibbs potentials and conditional probabilities in proportion to the importance and degree of interactions between a given word and its neighbors within the clique. The joint probability of words in a sentence, paragraph, or whole text, is expressed in terms of the product of these Gibbs conditional probabilities of a word taking a particular meaning given the state (meaning) of its neighbors within the clique. The model allows for the determination of what the keys words are from bilingual corpora as well as from expert translators' feedback to unravel the cultural and nuances that exist in the original text and of finding ways to faithfully express it through the host language. The Gibbs model exhibits the Markov property and is analogous to a physical system describing particles interacting between themselves to form specific pattern with Gibbs potential dictating such patterns. Efficient (Maximum likelihood estimators (MLE)) estimates for the Gibbs potentials are obtained using bilingual parallel corpora and can be augmented using maximum a posterior probability (MAP) estimators that naturally factors in the beliefs of expert translator(s). We map the expert's belief after translating it into expert Gibbs parameters. This allows for the seamless combination of expert translator and corpora data together to improve the accuracy of MT pointing to the dominance of one kind (corpus versus expert translator) of learning over the other depending on the degree of belief (variance of the Gibbs potential parameters) derived from the corpus versus the cultural expert translator. This collective learning lays the foundation for a culturally meaningful and faithful MT framework that goes well beyond the current literal state of MT. We also obtain the most likely translation (highest probability) of a given text (one sentence, paragraph(s), a whole chapter, etc.) using a computationally efficient iterative relaxation algorithm that changes the possible meaning of ambiguous words as long as the joint likelihood (probability) of the text to be translated increases. This modeling approach goes well beyond deep and state of the art Neural Network (NN) learning as it delves deeper into the essential modeling elements and factor the various layers that exist within the original language both culturally and linguistically to faithfully capture the meaning (semantics) of the translated text. We test our proposed approach on testbeds in the medical domain as well as in the literary and fiction domain, and our results will be compared to the state-of-the-art LSTM Neural Network (NN) approach with Glove word-vector [3] to solve the disambiguation task run on the same dataset for comparison. It outperforms the deep learning NN approach when the learning data is limited. This stems from the fact that our approach explicitly allows for the determination of what the keys words are from bilingual corpora as well as from expert translators' feedback to unravel the cultural and nuances that exist in the original text and of finding ways to faithfully express it through the host language.

KEYWORDS

Cultural translation, machine translation, polysemy, MRF Gibbs model, MLE, MAP, Iterative relaxation.

Biography

With the advent of globalization, accurate machine translation becomes an essential and important tool in various domains ranging from politics, to commerce, to social media, etc., as human translation can no longer meet the increasing demands. In the early 1950s, research on Machine Translation (MT) was first conducted by Yehosha Bar-Hillel at MIT [4]. Later in 1954, the Georgetown-IBM experiment was carried out by an IBM 701 machine. (Nye, M., 2016 [5]) The experiment was a milestone in Machine Translation history. It proved the feasibility of Machine Translation and funding from both government and market started to finance Machine Translation projects. However, the pace of MT came slowly in 1960s. ALPAC, the Automatic Language Processing Advisory Committee, reported that lack of progress was made and that the 10-year research fell short of expectations (Ueno, 1986 [6]). In late 1980s and early 1990s, thanks to IBM models, Machine Translation was back to public view and showed both better effectivity and accuracy. To date, most of the Statistical Machine Translation (SMT) models used today have the IBM models as their basis (Collins, 2013 [7]). In the last 20 years, tremendous great work have been made in MT domain, mainly in SMT (Kumar and Byrne, 2004 [8]) and Neural Machine Translation (He et al., 2016 [9]). Present popular MT engines, such as Google Translate, Microsoft Translator, Baidu Translate are based on either based on SMT or NMT, or hybrid of these two approaches, with some language pairs applying SMT and some using NMT.

REFERENCES

[1] Bassnett, S. (2013), Translation studies, Routledge.

[2] Cohen, F. (1986). Markov Random Fields for Image Modelling & Analysis. In: U. Desai, ed., Modelling and application of stochastic processes, Springer Publisher, First Edition, pp. 243-278.

[3] Cho, K., Van Merriënboer, B., Bahdanau, D., & Bengio, Y. (2014). On the properties of neural machine translation: Encoder-decoder approaches. arXiv preprint arXiv:1409.1259.

[4] Bar-Hillel, Y., The present state of research on mechanical translation, Am. Document. 2, 229-237 (1951, appeared in 1953.
.

[5] Nye, M., 2016. *Speaking in Tongues: Science's centuries-long hunt for a common language*. Distillations. 2 (1): 40–43. Retrieved 1 February 2017.

[6] Ueno, T., 1986. パーソナルコンピュータによる機械翻訳プログラムの制作 (in Japanese). *Tokyo: (株)ラッセル社*. p. 16. ISBN 494762700X.

[7] Collins, M., 2013. Statistical Machine Translation: IBM Models 1and 2. *COMS W4705: Natural Language Processing lecture notes*, pp.1.

[8] Kumar, S and Byrne, W., 2004. Minimum Bayes-risk decoding for statistical machine translation[R]. *Johns Hopkins Univ Baltimore MD Centre for Language and Speech Proceedings (CLSP)*, 2004.

[9] He, D.et al, 2016. Dual learning for machine translation. *In Advances in Neural Information Processing Systems* pp. 820-828.