# Neural Machine Translation Introduction

D. Gueorguiev 2/19/24

## Statistical Machine Translation (SMT)

The goal of statistical machine translation (SMT) is to produce target string from source string where models are estimated from examples. Among the possible target strings the one with max probability is chosen. The Bayes relation is used to introduce a target language model:

(1)

where is the *translation model* and is the *language model*. That is, the *translation model* is the probability that the source string is the translation of the target string while *language model* is the probability of seeing the target language string .

## Sequence to Sequence Network (Seq2Seq)

Drawback of Deep Neural Networks – can only be applied to problems whose inputs and targets can be sensibly encoded with vectors of fixed dimensionality. This is a significant limitation since many important problems are best expressed with sequences whose lengths are not known a-priori. For example, speech recognition and machine translation are sequential problems. Likewise, question answering can be seen as mapping a sequence of words representing the question to a sequence of words representing the answer.

## Continuous Space Language Models (CSLM) for Statistical Machine Translation

In most statistical approaches to machine translation the basic units of translation are phrases that are composed of one or more words. At the heart of the translation systems are models that estimate the translation probabilities for pairs of phrases, one phrase being from the source language and the other from the target language. Such models count phrase pairs and their occurrences as distinct if the surface forms of the phrases are distinct. Although distinct phrase pairs often share significant similarities, linguistic or otherwise, they do not share statistical weight in the models’ estimation of their translation probabilities. Besides ignoring the similarity of phrase pairs, this leads to general sparsity issues. The estimation is sparse or skewed for the large number of rare or unseen phrase pairs, which grows exponentially in the length of the phrases, and the generalization to other domains is often limited.

Continuous representations have shown promise at tackling these issues. Continuous representations for words are able to capture their morphological syntactic and semantic similarity.

//TODO: finish the section on Continuous Space Language Models

## References

[1] [Language Models are Unsupervised Multitask Learners, Alec Radford et al, 2018](https://github.com/dimitarpg13/large_language_models/blob/main/articles/neural_machine_translation/Language_Models_are_Unsupervised_Multitask_Learners_Radford_2018.pdf)

[2] [Neural Machine Translation of Rare Words with Subword Units, Rico Sennrich, Barry Haddow, Alexandra Birch, 2015](https://github.com/dimitarpg13/large_language_models/blob/main/articles/neural_machine_translation/Neural_Machine_Translation_of_Rare_Words_with_Subword_Units_Rico_Sennrich_2016.pdf)

[3] [Sequence to Sequence Learning with Neural Networks, Ilya Sutskever et al, Google, 2014](https://github.com/dimitarpg13/large_language_models/blob/main/articles/neural_machine_translation/Sequence_to_Sequence_Learning_with_Neural_Networks_Ilya_Sutskever_Google_2014.pdf)

[4] [Neural Machine Translation by Jointly Learning to Align and Translate, D. Bahdanau, K. Cho, Yoshua Bengio, 2016](https://github.com/dimitarpg13/large_language_models/blob/main/articles/neural_machine_translation/Neural_Machine_Translation_by_Jointly_Learning_to_Align_and_Translate_Bahdanu_2013.pdf)

[5] [Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation, K. Cho et al, 2014](https://github.com/dimitarpg13/large_language_models/blob/main/articles/neural_machine_translation/Learning_Phrase_Representations_using_RNN_Encoder%E2%80%93Decoder_for_Statistical_Machine_Translation_Cho_2014.pdf)

[6] [Recurrent Continuous Translation Models, N. Kalchbrenner, P. Blunsom, U of Oxford, 2013](https://github.com/dimitarpg13/large_language_models/blob/main/articles/neural_machine_translation/Recurrent_Continuous_Translation_Models_Kalchbrenner_2013.pdf)

[7] [Continuous Space Translation Models for Phrase-Based Statistical Machine Translation, Holger Schwenk, 2012](https://github.com/dimitarpg13/large_language_models/blob/main/articles/neural_machine_translation/Continuous_Space_Translation_Models_for_Phrase-Based_Statistical_Machine_Translation_Schwenk_2012.pdf)

[8] [Effective Approaches to Attention-based Neural Machine Translation, MH Luong et al, 2015](https://github.com/dimitarpg13/large_language_models/blob/main/articles/neural_machine_translation/Effective_Approaches_to_Attention-based_Neural_Machine_Translation_Luong_2015.pdf)

[9] [Neural Machine Translation, Quinn Lanners, 2019, Medium](https://medium.com/data-science/neural-machine-translation-15ecf6b0b)

[10] [A Gentle Introduction to Neural Machine Translation, J. Brownlee, Machine Learning Mastery, 2019](https://machinelearningmastery.com/introduction-neural-machine-translation/)

[11] [NLP From Scratch: Translation with a Sequence to Sequence Network and Attention, Sean Robertson, PyTorch Tutorial](https://pytorch.org/tutorials/intermediate/seq2seq_translation_tutorial.html)

[12] [A Neural Probabilistic Model, J. Bengio et al, 2003](https://github.com/dimitarpg13/large_language_models/blob/main/articles/A_Neural_Probabilistic_Language_Model_bengio_2003.pdf)

[13] [Recurrent neural network based language model, T. Mikolov et al, 2010](https://github.com/dimitarpg13/large_language_models/blob/main/articles/Recurrent_neural_network_based_language_model_mikolov_interspeech_2010.pdf)

[14] [On the Properties of Neural Machine Translation: Encoder-Decoder Approaches, Cho et al., U de Montreal, 2014](https://github.com/dimitarpg13/transformers_intro/blob/main/articles_and_books/OnthePropertiesOfNeuralMachineTranslationEncoderDecoderApproaches.pdf)