Notes on Neural Machine Translation

compiled by D.Gueorguiev 7/19/2025

# Introduction

# Sequence Transduction (ST)

## Continuous Translation Models (CTM)

//TODO: Finish this section (mainly using [1], [2], [3])

## Recurrent Continuous Translation Models (RCTM)

RCTM are a class of probabilistic

//TODO: Finish this section (mainly using [4])

# References

[1] [Sequence Transduction with Recurrent Neural Networks, Alex Graves, 2012](https://github.com/dimitarpg13/nlp_concepts/blob/main/literature/articles/neural_machine_translation/Sequence_Transduction_with_Recurrent_Neural_Networks_Graves_2012.pdf)

[2] [Continuous Space Language Models for Statistical Machine Translation, Holger Schwenk et al, 2006](https://github.com/dimitarpg13/nlp_concepts/blob/main/literature/articles/neural_machine_translation/Continuous_Space_Language_Models_for_Statistical_Machine_Translation_Schwenk_2006.pdf)

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

[4] [Continuous Space Translation Models with Neural Networks, Le Hai Son et al, 2012](https://github.com/dimitarpg13/nlp_concepts/blob/main/literature/articles/neural_machine_translation/Continuous_Space_Translation_Models_with_Neural_Networks_Le-Hai-Son_2012.pdf)

[5] [Recurrent Continuous Translation Models, Nal Kalchbrenner, Phil Blunsom, Oxford U., 2013](https://github.com/dimitarpg13/transformers_intro/blob/main/articles_and_books/Recurrent_continuous_translation_models_Kalchbrenner_Blunsson_OxfordU_2013.pdf)

[6] [Neural Machine Translation by Jointly Learning To Align and Translate, Dzmitry Bahdanau, K. Cho, Yoshua Bengio, 2016](https://github.com/dimitarpg13/transformers_intro/blob/main/articles_and_books/NeuralMachineTranslationByJointlyLearningToAlignAndTranslateBahdanau2015.pdf)

[7] [On The Properties of Neural Machine Translation: Encoder-Decoder Approaches, K. Cho, B. van Merrienboer, Dzmitry Bahdanau, 2014](https://github.com/dimitarpg13/transformers_intro/blob/main/articles_and_books/OnthePropertiesOfNeuralMachineTranslationEncoderDecoderApproaches.pdf)

[8] [Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation, K. Cho, B. van Merrienboer, C. Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, Yoshua Bengio, 2014](https://github.com/dimitarpg13/transformers_intro/blob/main/articles_and_books/LearningPhraseRepresentationsUsingRNNEncoderDecoderForStatisticalMachineTranslation.pdf)

[9] [Generating Sequences With Recurrent Neural Networks, Alex Graves, U of Toronto, 2014](https://github.com/dimitarpg13/deep_learning_and_neural_networks/blob/main/literature/articles/Generating_Sequences_With_Recurrent_Neural_Networks_Graves_2014.pdf)

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

[11] [Fast and Robust Neural Network Joint Models for Statistical Machine Translation, J. Devlin et al, Raytheon, 2014](https://github.com/dimitarpg13/nlp_concepts/blob/main/literature/articles/neural_machine_translation/Fast_and_Robust_Neural_Network_Joint_Models_for_Statistical_Machine_Translation_Devlin_2014.pdf)

[12] [Neural Machine Translation of Rare Words with Subword Units, Rico Sennrich, Barry Haddow, Alexandra Birch, 2015](https://github.com/dimitarpg13/nlp_concepts/blob/main/literature/articles/neural_machine_translation/Neural_machine_translation_of_rare_words_with_subword_units_UEdinsburgh_2016.pdf)

[13] [Bidirectional Recurrent Neural Networks, Mike Schuster, Kuldip Paliwal, IEEE Transactions on Signal Processing, 1997](https://github.com/dimitarpg13/transformers_intro/blob/main/articles_and_books/Bidirectional_Recurrent_Neural_Networks_Schuster_Paliwal_1997.pdf)

[14] [Neural Networks for Pattern Recognition, Christopher M. Bishop, 1995](https://github.com/dimitarpg13/transformers_intro/blob/main/articles_and_books/Neural_Networks_for_Pattern_Recognition-Bishop_1995.pdf)

[15] [Introduction to Finite-State Devices in Natural Language Processing, TR-96-13, Emmanuel Roche and Yves Schabes. 1996](https://github.com/dimitarpg13/nlp_concepts/blob/main/literature/articles/neural_machine_translation/Introduction_to_Finite-State_Devices_in_Natural_Language_Processing_roche_schabes_1996.pdf)