Deep Learning for Text Summarization: A Brief Survey

B. Tauro, T. Salah, T. Itani and M. Serry, Heriot – Watt University

**Abstract** - Deep learning has been widely adopted in recent years as it provides more promising results to traditional text summarization techniques, in this paper we evaluate sample papers that utilize deep learning for text summarization weither extractive or abstractive In order to evaluate overall challenges and trends in the field.

**Index Terms** - Text Summarization, Abstractive Text Summarization, Extractive Text Summarization, Deep Learning.

—————————— ◆ ——————————

# Introduction

*Automatic text summarization* is the task of producing a concise summary while preserving key information and overall meaning [1]. Previously, text summarization was dominated by traditional machine learning techniques [2]. In recent years, Deep learning has demonstrated more promising results to pre-existing approaches [3], Namely in both the extractive and abstractive paradigms. In this paper we briefly review the usage of deep learning models from sample papers to discuss motivations, challenges, and overall trends in the field.

# Models Used

A close up of a logo

Description automatically generatedIn this section, we review the deep learning models that were used in our sample papers.

## Restricted Boltzmann machines

Restricted Boltzmann machines (RBMs) are probabilistic graphical models that can be interpreted as stochastic neural networks. They are a variation of Boltzmann machines that learn more efficiently [4]. They have various applications outside of text summarization, including dimensionality reduction, topic modeling and feature learning [5].

A close up of a map

Description automatically generated

Figure 2 - The basic structure of a deep auto encoder.

Figure 1 - The basic structure of restricted Boltzmann machine (RBM).

In an RBM, we have layers where no two units within the same group are connected, in contrast to a BM where each node is connected to all other nodes symmetrically [5].

## Deep Auto-Encoders

A Deep Auto-Encoder is feed forward neural network with the fuction of reconstructing an input x [7]. They are a deep generative model that learns a binary encoding capable of fast and efficient retrieval times in comparison to latent analysis and and when used as a filtering mechanism for more traditional approaches such as TF-IDF [7].

They are a feed forward network with the exact number of nodes in both the input and output layers, each layer is composed of RBMs with a compressed bottle neck layer being used as a gating mechanism between the encoding and decoding process of the model. This allows the network to recreate the input from sparse features while also recognizing the most vital features in the input space [6].

## Pointer Generator Networks

Initially introduced by Nallapati et al. and See et al. they help solve the challenge of OOV words and factual errors. They perform better than other models for multi-sentence summaries [10].

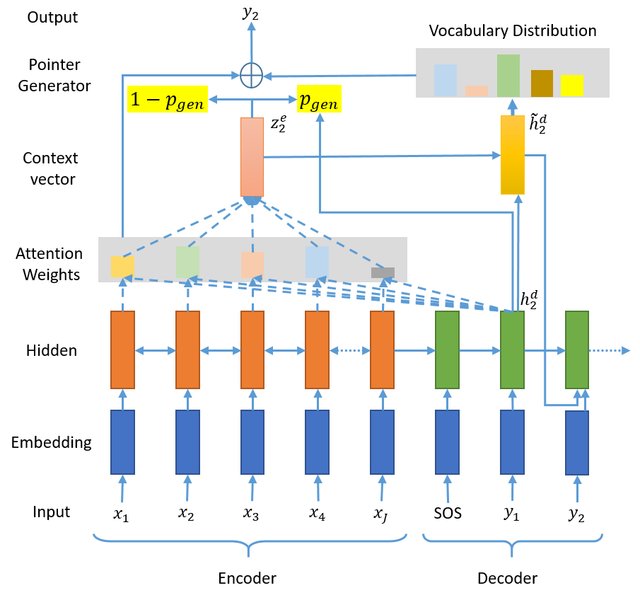


Figure 4 – Pointer Generator Network.

The basic idea is to choose between generating a word from the fixed vocabulary or copying one from the source document at each step of the generation. It brings in the power of extractive methods by pointing [10].

# Experiments

In this section, we give a brief overview for the experiments carried out in our selected papers and present their motivations, context, and methodology.

## Experiment: A

Verma, S. and Nidhi propose an extractive summarization approach that involves a mixture of feature engineering and using established methods in wider literature. The research was primarily motived by a proposition that perhaps better performance can be achieved by using simpiler RBMs mixed with carefully engineered features, it is applied to summarize factual reports from several domains [8].

Their proposed approach consists of three phases: *feature extraction*, *feature enhancement*, and *summary generation*, which work together to generate a coherent and understandable summary. In each phase various features are explored to improve the set of sentences selected for the summary. Lastly, an RBM is used to enhance and abstract those features to improve generated summaries in comparison to summaries generated by human experts [8].

## Experiment: B

Mahmood Yousefi-Azar, Len Hamey use an Auto Encoder (AE) to refine the features in the term frequencies of a document for summarization, using local and global vocabularies. They investigate the effect of adding noise to the term frequency before processing it with the encoder, creating a set of AEs’ called the Ensemble Noisy Auto Encoder (ENAE) [7].

A screenshot of a cell phone

Description automatically generated

Figure 3 – Ensemble Noisy Auto Encoder.

This ensemble adds random noise to the input term frequencies which changes the network from a feed forward model to a stochastic run model. They run their experiments on a corpus of emails [7].

## Experiment: C

Singhal, S. and Bhattacharya, A propose explore and review different techniques that can help overcome the issues of absurdity and repetitveness in generated summaries that existing deep learning approaches for abstractive text summarization produce [9].

They proposed solutions that address a wide range of possible improvements: usage of a large corprus, using more linguistically rich features for extraction such as Part of Speech Tagging (PoS), named entity recognition, TF-IDF weighting scheme, hierarchical attention models, bi-directional RNNs on both word and sentence level, Pointer Generator Networks (PGNs) which solves the out of vocabulary problem [9].

They also proposed a coverage mechanism that aims to solve the repetition problem by penalizing repetitive words. Similarly, they proposed Intra-Attention mechanism that avoids repeating words that have been output already by the decoder [9].

# Evaluation

# Conclusions

# References

1. Andrew Turpin, Yohannes Tsegay, David Hawking, and Hugh E Williams. 2007. Fast generation of result snippets in web search. In Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval. ACM, 127–134.
2. Dong, Y., 2018. A survey on neural network-based summarization methods. arXiv preprint arXiv:1804.04589.
3. K˚ageb¨ack, M., Mogren, O., Tahmasebi, N., and Dubhashi, D. (2014). Extractive summarization using continuous vector space models. In Proceedings of the 2nd Workshop on Continuous vector space space models and their Compositionality (CVSC)@ EACL, pages 31–39.
4. Fischer, Asja & Igel, Christian. (2012). An Introduction to Restricted Boltzmann Machines. 14-36. 10.1007/978-3-642-33275-3\_2.
5. A. Sharma, " Restricted Boltzmann Machines — Simplified", October 2nd, 2018. [Online]. Available: https://towardsdatascience.com/restricted-boltzmann-machines-simplified-eab1e5878976. [Accessed May 4, 2020].
6. Hinton, G. and Salakhutdinov, R. (2011), Discovering Binary Codes for Documents by Learning Deep Generative Models. Topics in Cognitive Science, 3: 74-91. doi:10.1111/j.1756-8765.2010. 01109.x
7. Mahmood Yousefi-Azar, Len Hamey, Text summarization using unsupervised deep learning, Expert Systems with Applications, Volume 68, 2017, Pages 93-105, ISSN 0957-4174, https://doi.org/10.1016/j.eswa.2016.10.017. (<http://www.sciencedirect.com/science/article/pii/S0957417416305486>)
8. Verma, S. and Nidhi, V., 2017. Extractive summarization using deep learning. arXiv preprint arXiv:1708.04439.
9. Singhal, S. and Bhattacharya, A., 2015. Abstractive Text Summarization.
10. Oriol Vinyals, Meire Fortunato, and Navdeep Jaitly. Pointer networks. In Advances in Neural Information Processing Systems, pages 2692–2700, 2015.
11. Abigail See, Peter J Liu, and Christopher D Manning. Get to the point: Summarization with pointergenerator networks. arXiv preprint arXiv:1704.04368, 2017.
12. J Dragomir R Radev, Eduard Hovy, and Kathleen McKeown. 2002. Introduction to the special issue on summarization. Computational linguistics 28, 4 (2002), 399–408.