TAURO, BRUCE H00228269 - SALAH, TAMER H00343334 – ITANI, TAREK H00292565 – SERRY, MOHAMED H00313456

DEEP LEARNING FOR TEXT SUMMARIZATION

A Brief Survey

MAY 9, 2020
HERIOT WATT UNIVERSITY
Dubai Campus

Table of Contents

Deep Learning for Text Summarization: A Brief Survey

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

Abstract - In this paper we evaluate sample papers that utilize deep learning for text summarization weither extractive or abstractive to evaluate overall trends in the field.

— — — — — — — — ◆ — — — — — — — — —

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 this paper we briefly review the usage, experimentation, and evaluation of deep learning models from sample papers to illustrate overall trends in the field.

MODELS USED

In 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].
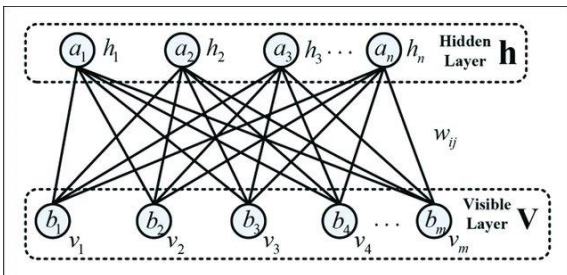

Figure 1 - The basic structure of restricted Boltzmann machine (RBM).

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 [7]. they have 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].
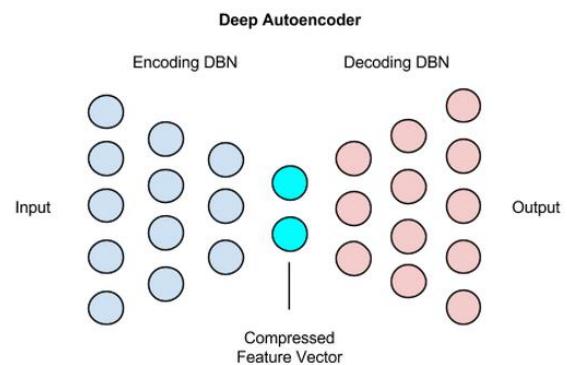

Figure 2 - The basic structure of a deep auto encoder.

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].
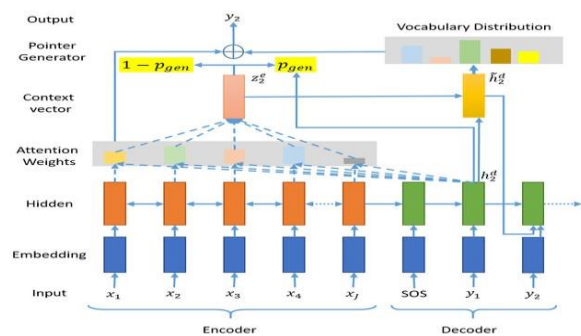

Figure 4 – Pointer Generator Network.

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 & APPROACHES

In this section, we give a brief overview of the methodology and overall approach used in our sample papers while mentioning experiments carried out where suitable.

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 [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 extracts a hierarchical representation of the data that initially did not have much variation, hence discovering the latent factors that can further enrich generated summaries. Generated summaries are then compared to expert-made summaries [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].
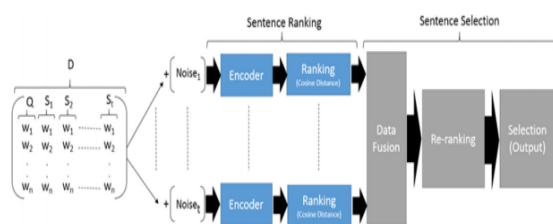

Figure 3 – Ensemble Noisy Auto Encoder.

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 explore different techniques that can help overcome the issues of absurdity and repetitveness in summaries gener-

ated by existing deep learning approaches [9].

They proposed solutions that address a wide range of possible improvements for overall better abstractive summaries such as: the usage of a large corpus, 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 and various other improvements [9].

In review, the authors carried no experiments on their own but merely surveyed results from related work.

EVALUATIONS

In this section, we briefly evaluate key results from experiments carried out in our sample papers.

Evaluation: A

Verma, S. and Nidhi's evaluation of the system-generated summaries is done based on three basic measures: Precision, Recall and F-Measure [8].
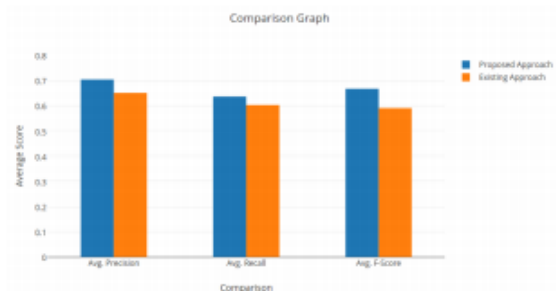
As observed from the above figure, with equal parameters for the RBM the proposed approach has more promising performance overall than previous approaches [8].

Evaluation: B

Mahmood Yousefi-Azar, Len Hamey Use the ROUGE and ROUGE2 metrics for evaluations they provide baselines for auto summarization results of the different models based on the number of sentences included in summary [7].
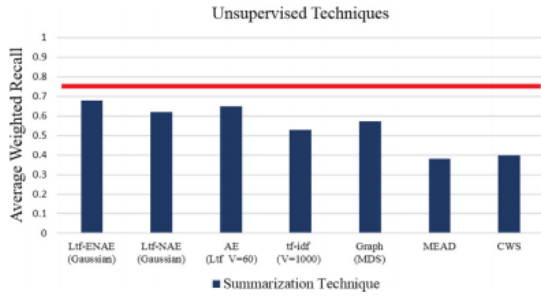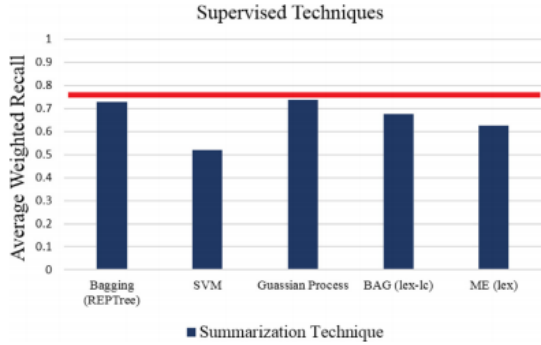
Fig. 11. Unsupervised techniques comparison, including the proposed model on, BC3.

From the results it is seen ENAE models have lower results than tf-idf when single sentences are used. However, when key value summarization is performed ENAE uniform has the highest scores across all models, and the ENAE models being on average equal to or better than tf-idf.

Evaluation: C

Singhal, S. and Bhattacharya suggested several metrics for evaluating abstraction-based summaries, each with accompanying preconditions to help the reader evaluate which metric would suit their goals. as in each data set, target summaries may or may not be provided [9].

1) If a target summary is missing: a similiarty measure between the summary and the source document in context of present topics, utilizing techniques such as LSA or LDA to identify latent topics is the proposed approach [9].

2) If a target summary is present, the metrics like ROUGE are more suited as they are essential string-matching metrics. A notable variation of ROUGE is ROUGE-N which measures the overlap of n grams between the generated and target summaries as a bigger N implies fluency in the generated summary [9].

Lastly, the DUC-2004 and Gigword data sets were recommened but the most used data set is the Daily Mail dataset [9].

CONCLUSIONS

In Conclusion, we found that Verma, S. and Nidhi provide a reasonable approach to summarization. However, many 'given conclusions' are to be observed throughout the paper without clear evidence or justification. Which makes it difficult to assess the quality of the results.

In Contrast, Mahmood Yousefi-Azar, Len Hame suggested a new method of auto summarizing documents using focused keywords and subjects to generate the summary. The evaluation of the model and number of experiments was remarkably in depth.

Lastly, Singhal, S. and Bhattacharya often did not provide detailed insights for their recommendations or the experiment carried out. Furthermore, we disagree with the Metrics used, we believe that the use of ROUGE and topic modeling is not suitable for real world application if there is error in the given text. The authors also acknowledged this.

Overall, we find that there is seems to be not enough focus on interprability or customization for the generated summaries with the end user in mind. none of the authors suggested any improvements based on the user preferences and without any customization to the user location, mother tongue or preferences. we believe text summarization should incorporate user preferences and the context and adjust the model accordingly.

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.