RESEARCH ARTICLE   **OPEN ACCESS**

Text Summarization And Evaluation Methods:–An Overview

Najma Abdi Omar*, Nevcihan Duru **
(Department of Computer Engineering, Kocaeli University, Turkey)
** (Department of Computer Engineering, Kocaeli University, Turkey)*

ABSTRACT
Automatic text summarization introduced in the late 50's has evolved in its methods and usage, with the advent of technology it has been found to be useful especially in helping mankind absorb the mountains of data that is continuously streaming in. This paper serves to be an introductory paper for students or researchers new in the field of automatic text summarization, it introduces the concept of text summarization and where it currently stands also briefly touching on evaluation method hence giving the reader an all-round feel of text summarization and its techniques.
Keyword: Abstractive, Evaluation, Extractive, Hybrid Text Summarization, Text Summarization

I. INTRODUCTION

With the advent of social media, the world of big data and the advancement in technology, the human reader is bombarded with thousands of different information outlet, while this is a good thing when looked through the lenses of more research and knowledge source availability it is often more than the average reader can handle hence the need for text summarization technology where by reducing the contents of particular subjects into half or less than the original content the user is able to absorb the knowledge at a better rate, which brings us to the question what is Automatic Text Summarization(TS)? Automatic Text summarization is a technological art where a given content is summarized to a much smaller text size depending on the type of summary needed, it can be a two sentence summary to give idea on what the full content of the article is or a paragraph that can be able to answer the main questions that the user required from that particular source or sources. There are various applications of text summarization examples include and are not limited to: production of abstracts or text outlines, reviews and in advertisement. Text summarization began in the late 50's with the grandfather of the technology Luhn [1] who had actualized an idea that text could be summarized much like the human summaries except with no one sided bias and less than the effort required of human summarizer, this technology has come a long way from the time it started except it is still mostly based on the original idea through different applications. At the time text summarization started it was a way to automatically produce abstracts much like the normal abstracts in technical articles and in recent times it has gained popularity in its usage whereby it is used in other applications like sentiment analysis and advertisement etc. the remainder of this paper is outlined as follows: section 2 introduces text summarization the types and steps involved, section 3 speaks of the evolution of text summarization and where it currently is at, section 4 touches briefly on evaluation of text summary with conclusion on section 5.

II. TEXT SUMMARIZATION TECHNOLOGICAL ART

The term 'technological art' in reference to text summarization is because in a way it's a creative and helpful skill that is improving as the years go by with new angles as it grows. The goal of TS is basically to produce a shorter version of a text that has important information for a user. TS has come a long way from the time it was first introduced by Luhn [1] who wanted a way to create abstracts without the need for much human effort and the emotional bias that is normal when any human is dealing with anything for example if a human is given an article to summarize depending on the mood of the particular day will give a summary that maybe totally different from another summary that the same person would produce for the same article. Text summarization has been defined by various authors throughout its history, [2] stated of "A good summary contains a small number of sentences but captures most of the main ideas of the document" while [3] stated that summarizing entailed reduction in both length and complexity while keeping the essential information of document, [4] defined summarization as "process of recognizing and indicating the most important component of a document or a set of document" in all the given definition of TS they all

lead to the same thing which entails as [5] stated "identifying the most relevant portions of given text" and rearranging it to make a sensible summary.

Some of the main types [6] of summarization as seen over this years are grouped according to:

i. The input type: i.e. single document summarization, where the produced summary is either an abstract, article outline or just the headline vs the multiple document summarization where the produced summary gives a summary of several documents circling on the same event or update summaries from a continuous flow of information about a particular happening.

ii. The purpose of the summary: this is usually either a generic summary with no specific objective rather than to give the general content of a particular document vs a query-focused summary which gives a summary in respect to a user query to give the info needed or a summary that is part of a complex question answering system.

iii. The output type: this is where you find definition of text summarization, here there is either an extractive summary where the summary is an output containing the exact phrases pulled from the source text, this is where majority of TS work has been done on vs the abstractive summary whereby the summary is expressing ideas of the documents paraphrasing it in as much as the human summarizer will do, this type is usually the harder and more challenging type of the two summarization methods.

There are usually three general steps involved in text summarization: the first step usually is in the content selection i.e. what is going to be used for the summary what will be discarded this is where the pre-processing of text come to play for example, tokenization, stemming, stop word removal etc. and the choosing of the sentences to be in summary in terms of sentence score or relation to topic among other text selection criteria in this stage the sentences or the paragraphs that are going to the summary are selected based on the algorithm in placewhich is either supervised by use of training; machine learning approach by use of certain features like sentence length or cue phrases sentences are categorized as going to summary or not or unsupervised; like defining salient words by use of weight of words in a document TF-IDF or use of topic signature algorithms like mutual information smaller set salient words or the log-likelihood ratio, the next step then becomes the ordering of the chosen sentences/paragraphs, this is where the summarizer will either keep the order of the sentences as they appear in source documents or maybe change the order bringing the most relevant first etc. depending on the algorithm employed then the final step is putting together the final summary i.e. sentence realization this stage is also where paraphrasing could be done, concept declaration done or the original sentences could be used in the most basic applications.

III. TEXT SUMMARIZATION HISTORY AND JOURNEY

Text summarization started in the 50's with the works of people like Luhn, in the beginning the objective was to do away with the need for human summarizers and in recent times from extraction of sentences like in [1]to the extraction of paragraphs like in [7] the usefulness of TS cannot be stressed enough, with the huge growth in data and with things like social media where everyone has an opinion on any topic, where expert word and irrelevant information come in equal measure text summarization will always be a technology that will be required in various aspects of life. To understand the future of TS there is need to look at where it all began. Among the first work of TS was demonstrated in [1] where Luhn worked on creating abstracts for single document technical papers and magazine articles, he used what is now the basis for most TS methods and that is statistical information i.e. word frequency to first compute the significance of words and then used the significant words to compute the significant sentences. The sentences with the highest significant were then used in the automatic production of a summary, this was the birth of extractive text summarization.

Then moving forward there were other revelations andin 1995 many breakthroughs in terms of TS were realized in such areas as the first to come with a trainable method using Naïve Bayes, a machine learning classifier where a set of features are used to categorize inclusive sentences was used by Kupiec[3], there was other uses of machine learning classifiers in TS like [8] use of Maximum Entropy in extracting sentences. In this same time also there was also the use of TS for not only technical articles but also for commercial newsand the first multi-document summarizer was also introduced by McKeown and Radev [9] who presented a natural language processing system that could summarize a series of news article circling the same event, the system named (SUMMONS) was based on traditional language generation architecture entailing two parts a content planner and a linguistic component,(for more details the paper can be referred to).

Other ways of text summarization were introduced examples including use of topic signature [5] i.e. sets of related words with associated weights that relate to the head topic were used in pointing out to the sections with complex concepts. There is also

use of relevance measure; by use of information retrieval methods the sentences are ordered and selected according to their relevance and latent semantic analysis; LSA used in getting the semantically important sentences as introduced by (10). Then there was the text summarization by aiding the user in limiting queries to leading texts resulting to better summary precession[11]which was found to be acceptable in most general news.

In this period of research in this field various summarizers were introduced and used in summarization from the SUMMARISTin [5] a system that was designed to generate multilingual summaries to other examples like the Mead [12]summarizer a public available multi-document analyser that uses a combination of different algorithms and has a feature extractor, a sentence scorer and a sentence re-ranker as its three basic components. TextRank, a graph based ranking model was also introduced by [13] an improvement in the TS work. LexRank [14] another text summarizing method which groups sentences by processing the relationship between sentences.

In recent research areas with the improvement in technology, many improvement have been made in the existing TS systems, like in [15]relying on word distribution and lexical chain critical sentences in the articles are extracted. In 2015 due to the domain dependence nature of the traditional TS system [16]came up with a one size fits them all design which introduced the concept of domain independent framework which could be used for both extractive and abstractive TS. There was also introduction of event-based text summarization as in [17] where there was use of atomic events to describe the relationship between the important actors in the document, and also extending to the use of subevents as in [18] where the highlights of the document were grouped and sentences were ranked according to their direct relation to the subevent the sentences with the highest score were selected. In the recent past there was the additional step in TS where a list of hybrids came up in terms of hybrid of extractive and abstractive models for TS as in [19] where single long document abstractive summaries were constructed from the extractive summary by feeding the extracts into an RNN based encoder-decoder model, and in [20] where they integrated a hybrid of extractive and abstractive model for multi-document summarization using WordNet. Other hybrids in terms of techniques like mix of LexRank and TextRank two independent summarizing techniques which resulted to perform better than individual performance when combined in [21], the mix of statistical and linguistic feature in [22]were also found to be more effective when used together in one text summarizer. There is also borrowing other fields in text summarization, apart from information retrieval who's techniques have been a major influence in text summarization, recent works like [23] have borrowed sentiment analysis in text summarization by using it to figure out the key sentences of a doc, these are some of the major works that have and currently are influencing the text summarization field.

IV. EVALUATION OF TEXT SUMMARIES
Evaluation of the summaries produced has been one of the earliest challenges of text summarization, how good is the summary? How do you judge the goodness or badness of a summary? Throughout the history of summarization various ways of evaluation have been proposed ranging from the first of the types like evaluation by humans which beats the purpose of summarization which was mainly reduce human effort and bias, with the advancement of technology automatic ways of evaluating summaries have been proposed in the following parts the types of evaluations in existence will be discussed in brief: categorically we have two types of evaluation, the extrinsic type which basically checks how helpful a given summary is and the second type intrinsic which evaluates a summary by comparison to a basis of some sort either by another summary or a human summary. The two categories are further classified as follows the figure 1below[24] gives a clear overview of the methods.
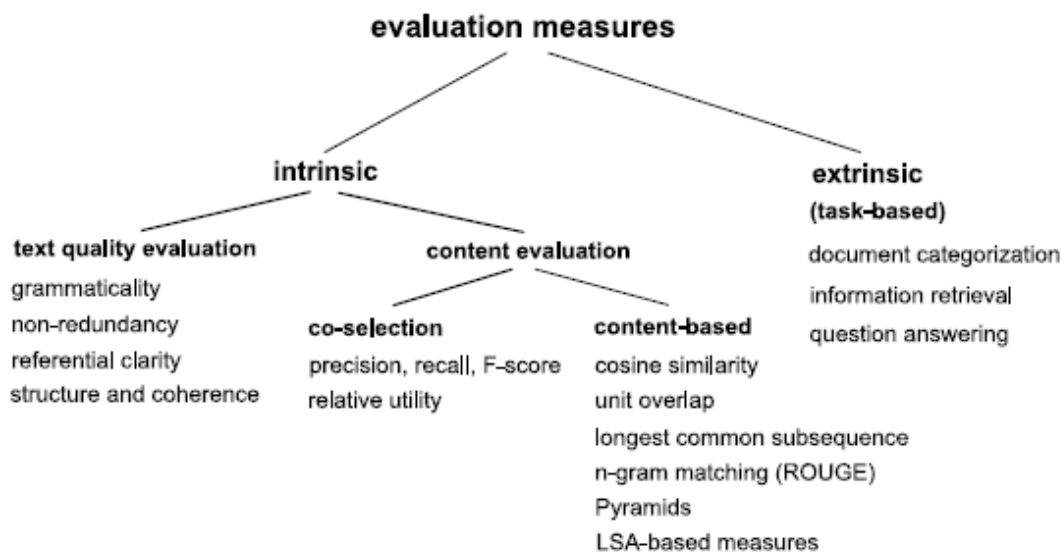
fig1. The evaluation measures [24]

The intrinsic evaluation measures are inclusive of text quality measure which looks at the basic structure of the summary produced in terms of grammar i.e. did the summary have a good grammatical construct, it also checks for redundant information the information given should have a natural flow of inclusion of no repetitive information and it also checks whether the references in terms of noun pronouns etc. are given due diligence, this type of evaluation also checks the text coherence giving an A if the sentences are well structured and coherent and an E if the quality is poor. The co-selection evaluation criteria checks the relevance of the given summary as per the document and a human given summary giving points according to the relevance score. The Content based evaluation on the other hand is as a result of complete unreliability on co-selection methods of evaluation some of the most used evaluation metrics is in this category with the example of ROUGE and its different variation based on similarities of n-grams. The extrinsic evaluation also termed as task based on the other hand, do not do sentence by sentence analysis but rather check the usefulness of the summary as a whole, was it able to achieve its objective? Is the user satisfied after using the summary, it measures according to the task at hand majorly can it categorize a document properly given sets of documents, can the summary be used in question answering in place of the source document etc. Even though evaluation has come a long way, work in improving evaluation measures is always ongoing and maybe in the future there will be an entire different sets of evaluation metrics or at the very least an improvement of the existing types.

V. CONCLUSION AND FUTURE WORK

This paper has given a brief overview of what text summarization is, the different type of text summarization, the journey that was text summarization and where text summarization is currently, there was also a brief touch on what the evaluation methods for text summarization are. From the studied trend it has been noted that text summarization has advanced in methods from the start of it and that the work on TS is yet to be completed, the ongoing research shows that the future is headed in hybrid of the different types mentioned. In the future a combination of the different algorithms, features and extractive abstractive methods integrating it with events could yield better results much more than expected in both single and multi-document summarization, where a positive answer can be given to the challenge/statement posed by [7] in his conclusion as to whether a machine will be able to extract the best abstracts given its nature when even human beings cannot agree on which paragraphs (sentences) represent an article best, maybe this will be the version of Turing test for text summarization.

REFERENCES

[1]. H.P Luhn (1958). The Automatic Creation of Literature Abstracts*
[2]. John Conroy, Dianne P. O'Leary (2001) Text Summarization via Hidden Markov Models and Pivoted QR Matrix Decomposition
[3]. Julian Kupiec, Jan Pedersen, Francine Chen (1995). A Trainable Document Summarizer
[4]. Mehdi Jafari, Amir Shahab Shahabi, Jing Wang, Yongrui Qin, Xiaohui Tao, Mehdi

Gheisari(n.a). Automatic Text Summarization Using Fuzzy Inference

[5]. Chin- Yew Lin, Eduard Hovy (n.a) The Automated Acquisition of Topic Signature for Text Summarization

[6]. Dan Jurafsky & Chris Manning (n.a) Introduction to Summarization-NLP-retrieved from https://www.youtube.com/watch?v=EZLCOrr l0Wc

[7]. Mandar Mitra, Amit Singhal, Chris Buckley (n.a). Automatic Text Summarization by Paragraph Extraction

[8]. Miles Osborne, (2002). Using Maximum Entropy for Sentence Extraction

[9]. Kathleen McKeown, Dragomir R. Radev (1995). Generating Summaries of Multiple News Articles

[10]. Yihong Gong, Xin Liu (2001) Generic Text Summarization Using Relevance Measure and Latent Semantic Analysis

[11]. Mark Wasson (n.a) Using Leading Text for News Summaries: Evaluation Results and Implications for Commercial Summarization Applications

[12]. Dragomir R. Radev et& all (2000). MEAD- A Platform for Multidocument Multilingual Text Summarization

[13]. Rada Mihalcea, Paul Tarau ().TextRank: Bringing Order into Texts

[14]. Erkan G, Radev D R. (2004) LexRank: Graph-based lexical centrality as salience in text summarization

[15]. Regina Barzilay, Michael Elhadad (1997). Using Lexical Chains for Text Summarization

[24]. Josef Steinberger, Karel Jezek (2009). Evaluation Measures for Text

[16]. Yogesh Kumar Meenaa, Dinesh Gopalani (2015). Domain Independent Framework for Automatic Text Summarization

[17]. Elena Filatova, Vasileios Hatzivassiloglou (n.a) Event-Based Extractive Summarization

[18]. Naomi Daniel, Dragomir Radev, Timothy Allison (n.a). Sub-event based multi-document summarization

[19]. Shuai Wang, Xiang Zhao, Bo Li, Bin Ge, Daquan Tang (2017). Integrating Extractive and Abstractive Models for Long Text Summarization. 2017 IEEE 6th International Congress on Big Data

[20]. Harsha Dave, and Shree jaswal (2015). Multiple Text Document Summarization System using Hybrid Summarization Technique. 2015 1st International Conference on Next Generation Computing Technologies (NGCT-2015)

[21]. Ailin Li, Tao jiang, Qingshuai Wang, Hongzhi Yu (2016). The Mixture of TextRank and LexRank Techniques of Single Document Automatic Summarization Research in Tibetan. 2016 8th International Conference on Intelligent Human-Machine Systems and Cybernetics

[22]. Joel Larocca Neto, Alex A. Freitas, Celso A. A. Kaestner (n.a). A Trainable Text Summarizer Employing Statistical and Linguistic Features

[23]. Nidhika Yadav, Niladri Chatterjee (2016). Text Summarization using Sentiment Analysis for DUC Data. 2016 International Conference on Information Technology