International Journal of Scientific Research in Computer Science, Engineering and Information Technology
ISSN : 2456-3307 (www.ijsrcseit.com)
doi : https://doi.org/10.32628/CSEIT2390218

Utilizing Deep Learning Techniques for Text and Image Capturing Summarization in Information Retrievals

Dr. S. Selvakani*[1], Mrs K. Vasumathi[2], S.Divya[3]

[*1] Assistant Professor and Head, PG Department of Computer Science, Government Arts and Science College, Arakkonam, Tamil Nadu, India

[2] Assistant Professor, PG Department of Computer Science, Government Arts and Science College, Arakkonam, Tamil Nadu, India

[3] PG Scholar, PG Department of Computer Science, Government Arts and Science College, Arakkonam, Tamil Nadu, India

ARTICLEINFO

Article History:

Accepted: 13 March 2023
Published: 29 March 2023

Publication Issue
Volume 10, Issue 2
March-April-2023
Page Number
202-207

ABSTRACT

In this paper, a novel information retrieval and text summarization model based on deep learning (DL) is introduced. The model comprises three primary stages, including information retrieval, template generation, and text summarization. The initial step involves utilizing a bidirectional long short term memory (BiLSTM) technique to retrieve textual data. This approach considers each word in a sentence, extracts relevant information, and converts it into a semantic vector.

Keywords: Semantics, Information retrieval, Feature extraction, Data mining, Deep learning, Task analysis.

I. INTRODUCTION

Due to the rapid growth of content such as blogs, articles, and reports, retrieving data from vast amounts of text has become an arduous task. Automatic text summarization methods offer a solution by extracting meaningful information from extensive texts, preserving the original meaning and significant portions. Summarization is a crucial aspect of natural language understanding, aiming to produce a concise representation of the input text that captures its essence. Extractive approaches, which involve selecting and combining text fragments, are commonly used in successful summarization systems. Conversely, abstractive summarization strives to generate a summary from scratch, incorporating new elements not present in the original text.

Text-Image summarization involves summarizing a document containing both text and images into a

summary that includes both elements. This approach differs from pure text summarization and image summarization, which involves summarizing sets of images. Natural language processing (NLP) is a set of computational techniques aimed at automatically analyzing and representing human language. NLP research has advanced significantly, with the analysis of a sentence taking mere seconds instead of minutes. When summarizing text with images, the visual content of the image, such as object shape arrangement, and color, is considered alongside associated image data. These systems are faster and more efficient than conventional image retrieval methods. This paper proposes a new system that uses Gabor filtering to extract features, which are then optimized using lion optimization. Classification is performed using SVM for cuckoo search optimization and the decision tree method for lion optimization. The proposed method is tested based on various parameters, demonstrating that Lion optimization produces superior results compared to cuckoo search optimization.

II. RELATED WORK

Natural language processing (NLP) refers to a set of computational techniques used for automatically analyzing and representing human language, motivated by theory. NLP research has progressed from the era of punch cards and batch processing, where sentence analysis could take up to seven minutes, to the current era of Google and similar systems, where millions of web pages can be processed within a second. NLP empowers computers to perform a diverse range of natural language-related tasks at various levels, including parsing, part-of-speech (POS) tagging, machine translation, and dialogue systems.

Statistical natural language processing (NLP) has become the go-to approach for modeling intricate language tasks. In the early days of statistical NLP, it often faced the problem of the curse of dimensionality when attempting to learn joint probability functions of language models. To overcome this issue, there was a push to learn distributed representations of words that existed in a low-dimensional space [1].

Word embeddings, also known as distributional vectors, are based on the distributional hypothesis, which states that words with similar meanings tend to appear in similar contexts. Therefore, these vectors attempt to capture the characteristics of a word's neighbors. The primary benefit of distributional vectors is that they can capture the similarity between words, which can be measured using methods such as cosine similarity. As a result, word embeddings are frequently employed as the initial data processing layer in a deep learning model [2].

RNNs are designed to process sequential information, and they are called "recurrent" because they perform the same task for each instance of the sequence, with the output depending on previous computations and results. Typically, a fixed-size vector is created to represent a sequence by feeding tokens one by one to a recurrent unit. RNNs have "memory" over previous computations and use this information in current processing, making them well-suited for NLP tasks such as language modeling, machine translation, speech recognition, and image captioning. CNNs and RNNs have been essential in sequence transduction applications using the encoder-decoder architecture[3].

LSTM-based models have additional "gates" that allow them to memorize longer sequences of input data. The question is whether the gates in the LSTM architecture provide good predictions and whether additional training data is needed to further improve prediction. The objective is to explore the extent to which additional layers of training data would be beneficial in tuning the parameters. The results show

that BiLSTM-based modeling with additional training data offers better predictions than regular LSTM-based models. Specifically, BiLSTM models provide better predictions compared to ARIMA and LSTM models[4].

Forecasting is a crucial but challenging part of time series data analysis, and the performance and accuracy of time series data analysis and forecasting techniques depend on the type of time series data and underlying context. Section II reviews related works, and Section III provides essential background and mathematical formulations. The procedure for experimental setup, data collection, and preparation is presented in Section IV, while Section V presents the pseudo-code of the developed algorithms[5].

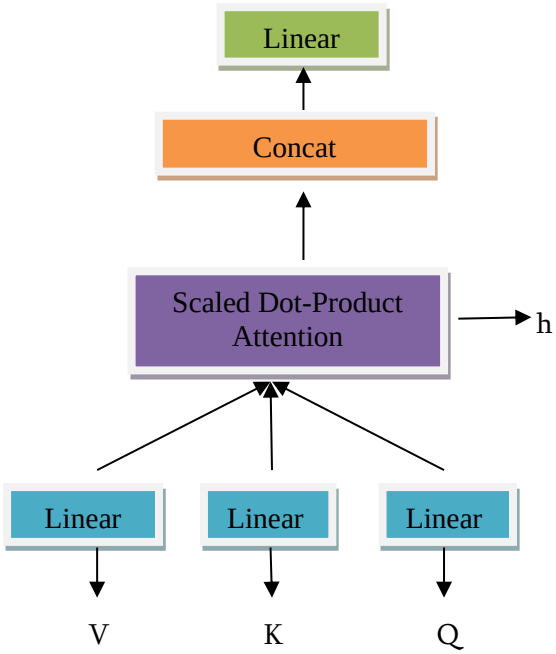

Figure 1.Multi-head Attention

Arabic is among the six official languages recognized by the United Nations (UN) and is classified as a Semitic language spoken by Muslims and Arabs worldwide. The global population of native Arabic speakers is approximately 400 million, with 30 different dialects, and it is estimated that around 1 billion Muslims speak Arabic. The Arabic alphabet includes 28 letters and the Hamza, and it is written from right to left. Capitalization is not used in Arabic, and the shape of Arabic letters varies based on their position within the word.

The term Natural Language Processing (NLP) refers to the research and practical application of using computers to understand and manipulate natural language in written or spoken forms for useful purposes. NLP has a wide range of applications, including sentiment analysis, machine translation, information retrieval, speech recognition, and expert systems. To improve the accuracy of NLP applications, various machine learning algorithms have been employed since accuracy is critical for many NLP tasks [6].

The size of image databases is constantly increasing due to advancements in technology, high-speed internet, and increased storage capacity in various devices. As a result, there is a growing need for the development of image retrieval systems. In the past, images were annotated manually with text, keywords, and tags, which were used as metadata to describe them. However, manual annotation becomes challenging and time-consuming with large datasets, leading to increased costs and the need for significant manual labor. [7]

The field of pattern recognition heavily relies on a type of neural networks known as Convolution Neural Networks (CNN) and conventional auto-encoder networks, which are typical Artificial Neural Network (ANN) models. In contrast, Recurrent-based Neural Networks (RNNs) use a feedback methodology to remember parts of past data. This is achieved through a loop in the network, which not only allows training to occur from input to output in a feed-forward manner but also enables the network to function like a memory, preserving important information.

The encoder-decoder architecture involving CNNs and RNNs has been widely used for sequence transduction applications. However, these architectures face a bottleneck due to sequential processing during the encoding step. To overcome this, Vaswani et al. introduced the Transformer which eliminated the need for recurrence and convolutions in the encoding step, relying solely on attention mechanisms to capture the global relationships between input and output. This led to a more parallelizable architecture that required less training time and achieved positive results on tasks such as translation and parsing.

III. METHODOLOGY

In contrast to traditional summarization models, the previous approach relied on manual feature engineering that required domain expertise and labeled data. Furthermore, this method was based on human-generated summaries, which were used to mitigate the impact of manual feature engineering and data labeling. Instead, in this approach, the summary is generated using a DBN model. The sentences are embedded to determine a concise model that is then fed into the DBN, which consists of input, hidden, and output layers with a single component. Similar to word embeddings, distributed representation for sentences can also be learned in an unsupervised fashion. The result of such unsupervised learning is “sentence encoders”, which map arbitrary sentences to fixed-size vectors that can capture their semantic and syntactic properties. Usually an auxiliary task has to be defined for the learning process.

These types of neural networks are useful for modeling the (linear or non-linear) relationship between the input and output variables and thus functionally perform like a regression-based modeling. In other words, through these networks a functional mapping is performed through which the input data are mapped to output data.

The object detection process of Faster R-CNN involves two stages. The first stage, referred to as the Region Proposal Network (RPN), generates object proposals. An intermediate-level CNN feature map is convolved with a small network. The network predicts an objectness score that is agnostic to class and a bounding box refinement for anchor boxes of various scales and aspect ratios at each location of the feature map.

The presented algorithm demonstrates that the mixed signal dimension can be smaller than the independent component separation dimension while still achieving good precision. This means that the true nature of an object can be obtained using a smaller number of signal acquisition devices, resulting in reduced processing costs. The method is based on high order statistical analysis of signals and utilizes a blind source separation technology.

IV. EXPERIMENTS

The problem of summarization is framed as a binary classification task, where the classifier distinguishes between important and unimportant sentences. This is accomplished using datasets generated with specific techniques. Our approach differs from previous methods of summarization, which rely on either manual feature engineering based on domain knowledge or large labeled data sets. Instead, our method does not rely on either of these and can be easily applied to areas where human-generated summaries are available, without the need for manual feature crafting or data labeling.

We expand the existing DailyMail corpora by extracting the captions and images from the HTML-formatted documents. This augmented dataset is called E-DailyMail. The DailyMail and CNN datasets

are commonly used datasets for neural document summarization, initially compiled by gathering human-generated highlights and news stories from news websites in a study by Hermann et al. in 2015.

TABLE I

ANALYSIS OF GENERATED IMAGE CAPTIONS WITH THE EXISTING

MODEL	BLUE-1	BLUE-2	BLUE-3	BLUE-4
Nearest neighbor	0.48	0.281	.166	0.1
Google NIC	0.66	0.461	0.329	0.246
LRCN	0.62	0.442	0.304	-
AICRL-RestNet50	0.731	0.562	0.41	0.326
BiLSTM +Attention	0.758	0.734	0.746	0.73

The presented algorithm demonstrates that the mixed signal dimension can be smaller than the separation independent component dimension, resulting in good precision. This means that the true nature of the object can be captured using a limited number of signal acquisition devices, leading to a reduction in processing costs. The success achieved through independent component analysis (ICA) has led to its increased use in various applications. The proposed approach involves three main stages: firstly, information retrieval and template generation using a Bi-LSTM model; secondly, text summarization using a DBN model; and lastly, caption generation for images using both CNN and RNN techniques. These stages will be discussed in further detail in the following sections.

The process involves applying attention to the output of one or more layers of a CNN in each case, achieved by predicting weights for each spatial location in the CNN output. However, finding the optimal number of image regions inevitably involves a trade-off between coarse and fine levels of detail that cannot be resolved. Moreover, the random placement of regions with respect to image content can make it harder to detect objects that are not well aligned with regions and to associate visual concepts related to the same object.

Figure 2. Final Generate Image Capturing

V. CONCLUSION

The presented model consists of three main stages: information retrieval, template generation, and text summarization. Initially, the BiLSTM approach is utilized to retrieve textual data, where each word in a sentence is assumed to extract information and embed it into a semantic vector. Then, the DL model is used for template generation. The DBN model is used as a text summarization tool to summarize textual content, and image captions are generated. The model's performance is evaluated using the Gigaword corpus and DUC corpus.

The probabilistic model we use is combined with a generation algorithm that produces precise abstractive summaries. Our next objective is to enhance the grammatical accuracy of the summaries through data-driven methods and expand the system's capabilities to generate summaries at the paragraph level.

In conclusion, we anticipate the emergence of additional deep learning models that integrate internal memory (acquired from the data) with

external memory (inherited from a knowledge base) to enhance performance. The combination of symbolic and sub-symbolic AI will play a crucial role in advancing from NLP to true natural language comprehension.

Advancements in technology have led to the development of image retrieval systems through the increase in storage devices, high-speed internet, and capacity. In the past, images were manually annotated with tags, keywords, and texts, known as metadata. Images contain visual information such as color, texture, shapes, and spatial information. CBIR systems are used to search for images in large databases based on the present visual information

VI. REFERENCES

[1]. D. Jain, M. D. Borah, and A. Biswas,"Fine-tuning textrank for legal document summarization: A Bayesian optimization based approach", Proc. Forum Inf. Retr. Eval., Hyderabad India, Dec. 2020, pp. 41_48.

[2]. H. Oufaida, O.Nouali, and P. Blache,Minimumredundancy and max- imum relevance for single and multi-document Arabic text summariza-tion, J. King Saud Uni v.-Comput. Inf. Sci., vol. 26, no. 4, pp. 450_461, Dec. 2014.

[3]. H. Yamada, S. Teufel, and T. Tokunaga, "Designing an annotation scheme for summarizing Japanese judgment documents," Proc. 9th Int. Conf. Knowl. Syst. Eng. (KSE), Oct. 2017, pp. 275_280.

[4]. J. Chen and H. Zhuge, "Abstractive text-image summarization using multi-modal attentional hierarchical RNN," Proc. Conf. Empirical Methods Natural Lang. Process., Brussels, Belgium, 2018, pp. 4046_4056

[5]. K. Agrawal, "Legal case summarization: An application for text sum- marization," Proc. Int. Conf. Comput. Commun. Informat. (ICCCI), Jan. 2020, pp. 1_6.

[6]. K. Merchant and Y. Pande,"NLP based latent semantic analysis for legal text summarization," Proc. Int. Conf. Adv. Comput., Commun. Informat. (ICACCI), Bengaluru, India, Sep. 2018, pp. 1803_1807, doi: 10.1109/ICACCI.2018.8554831.

[7]. M. Farsi, D. Hosahalli, B. Manjunatha, I. Gad, E. Atlam, A. Ahmed, G. Elmarhomy, M. Elmarhoumy, and O. Ghoneim, "Parallel genetic algo-rithms for optimizing the SARIMA model for better forecasting of the NCDC weather data," Alexandria Eng. J., vol. 60, no. 1, pp. 1299_1316, 2021.

[8]. R. K. Venkatesh, "Legal documents clustering and summarization using hierarchical latent Dirichlet allocation,"IJ-AI, vol. 2, no. 1, pp. 27_35, Mar. 2013.

[9]. S. Bhattasali, J. Cytryn, E. Feldman, and J. Park, "Automatic identi_cation of rhetorical questions," Proc. 53rd Annu. Meeting Assoc. Comput. Linguistics, vol. 2, Beijing, China, 2015, pp. 743_749.

[10]. Z. Malki, E. Atlam, G. Dagnew, A. Alzighaibi, E. Ghada, and I. Gad, "Bidirectional residual LSTM-based human activity recognition,"Com-put. Inf. Sci., vol. 13, no. 3, pp. 1_40, 2020.

Cite this article as :

Dr. S. Selvakani, Mrs K. Vasumathi, S. Divya, "Utilizing Deep Learning Techniques for Text and Image Capturing Summarization in Information Retrievals", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 9, Issue 2, pp.202-207, March-April-2023. Available at doi : https://doi.org/10.32628/CSEIT2390218

Journal URL : https://ijsrcseit.com/CSEIT2390218