TiZen: Neural Title Generation for Scientific Papers

Harshil Jain*, Anubhav Jain*, Chandan Maji*, Abhisht Tiwari*, Naman Jain
University of California San Diego
hjain@ucsd.edu

ABSTRACT

Generating a catchy title for research papers plays a pivotal role in influencing the mindset of the reviewer. However, it can become an arduous task for an author to come up with a very compelling title, which captures the essence of the entire paper. It is the title that intrigues the reader's interest to read the whole document. Every year a lot of papers having high-quality research involved fail to catch the attention of the community because of the lack of an appealing title. Generating a title requires thought and time. In this fast-paced world, we propose a neural title generator TiZen. TiZen will help the authors to make their work stand out in the community and it generates the title with only the abstract of the paper. We have used abstractive and extractive text summarization along with some subtle ground rules to make the generated title catchy. A catchy title not only grabs the attention of the reader but also motivates him to read the entire content. This is precisely what TiZen has been built for. Results show that titles generated by TiZen are significantly closer to the title a person may predict to be.

KEYWORDS

Catchy Title Generation; Scientific Papers; Text Summarization

1 INTRODUCTION

Many scientists and researchers do not invest time in coming up with a title for their manuscripts, yet many others consider the title to be the most crucial element in a written piece [10]. Libraries organize and retrieve manuscripts using the titles as they play an important role in introducing the written content, attracting the relevant audience, identification of domain, and uniquely defining the manuscript. As research volume increases approximately by 8-9% every year [9], a title can help in highlighting a manuscript to the potential reader and contribute towards greater scientific impact (measured in terms of citations). Since it is the entry point of a manuscript and is the most read sentence out of all the manuscript's content, a title can influence the reader's judgement about the importance of the paper. [10] In some cases, it can act as an influential factor in the review process of a publication.

Thus, a title of a scientific manuscript should accurately summarise the work, must be easy to read, and should make the work stand apart from the clutter of the scientific publishing ecosystem through catchy words or phrases. Due to the subjective definition of catchiness, constraints on the length of the sentence, weak one-line summary generator algorithms, and difficulty in generating out-of-vocabulary words, the creation of an apt and successful title is a challenging task for researchers.

In the past, neural-based methods have been successful in summarisation problems because of their ability to generalize well owing to the large number of training features extracted from the data. We introduce TiZen, an automated tool to help researchers generate an appropriate, creative, and catchy title for research articles. We introduce catchiness to the title generated by the neural model through rule-based NLP techniques on the abstract of the paper.

*Equal Contribution

2 RELATED WORK

Earlier, researchers have utilized automatic text summarization techniques to produce a short and concise summary of a document while preserving the key content and overall meaning of the text [2]. Vasilyev et al. [19] proposed a method for generating titles for unstructured text documents by reframing the problem as a sequential question-answering task. Shvets et al. [15] tried to automate the process of title generation with various levels of informativeness to benefit from different categories of users. Various approaches have been developed for automatic text summarization and applied widely in several domains. For example, search engines are used for generating snippets for previewing documents [18]. There are several examples wherein news websites use text summarization to generate a short description of the news, especially news headlines, by using knowledge extractive approaches ([1][13][17]). Luhn [7] introduced a novel technique to extract salient sentences from the text using attributes such as word and phrase frequency. They proposed to weight the sentences of a document as a function of high-frequency words. Putra and Khodra [10] have tried to generate titles using template based approach and K-Nearest Neighbours models. However, there are no significant works in the literature that either utilize text summarization techniques or neural models for generating title for a scientific paper. Interestingly, there have been no significant previous works on introducing catchiness into a text document too. TiZen, however, is a combination of both neural and heuristic approaches, which makes it much more efficient and accurate.

2.1 Our Contribution

In this paper, we address the task of automatically generating catchy titles for scientific papers. We have approached the task using the combination of both the abstractive and extractive summarization techniques. In the first stage, we extract the text filtered from the abstract using extractive text summarization. Next, we predict the title using abstractive text summarization. Then, we propose heuristics to make the title catchy. Although the topic of catchiness is subjective, we have introduced a "Catchiness Score" metric to measure how catchy is the generated title. We achieved interesting results over standard text summarization algorithms in terms of BLEU score and Catchiness. A survey conducted at IIT Gandhinagar confirms the authentication of the metric "Catchiness" and practically evaluates our model TiZen.

Outline: Section 3 describes the dataset used for conducting the experiments. Section 4 explains the components from the pipeline and architecture of TiZen. Section 5 describes the experimental setup and results.

3 DATASETS

The dataset is prepared from ACL Anthology and arXiv. We have used the OCR++ tool [16] to extract details of each paper. Out of the eight available features in the dataset, we used the paper title as the actual title, the abstract as training input to generate the titles and also the keywords if they were available. We collected 58,700 scientific papers for training and evaluation purposes.

4 METHODOLOGY

Data Preprocessing: We removed the incomplete entries of either the title or the abstract from the dataset. Then, we use extractive summarization and the top five relevant sentences from every abstract are extracted using the TextRank algorithm [8]. The standard pre-processing steps such as lower casing, stopwords and punctuation removal were conducted on the whole dataset. Then, we found out that 94% of the dataset had titles with less than 15 words, and 96% of the abstracts contained less than 200 words. We removed the samples which exceeded these limits and obtained a final dataset with abstract and title pairs. The dataset was split into a 90:10 ratio with 52,802 papers for training and 5,868 papers for testing. The cleaned abstract and the title was then fed to the abstract encoder and title encoder modules, respectively and finally heuristics were used for adding catchiness. A diagrammatic representation of the entire pipeline is depicted in Figure 1.

Pipeline Overview: TiZen, our entire pipeline can be divided into two parts
1. Generation of the title using abstractive and extractive summarization.
2. Adding catchiness to the generated title using heuristics.

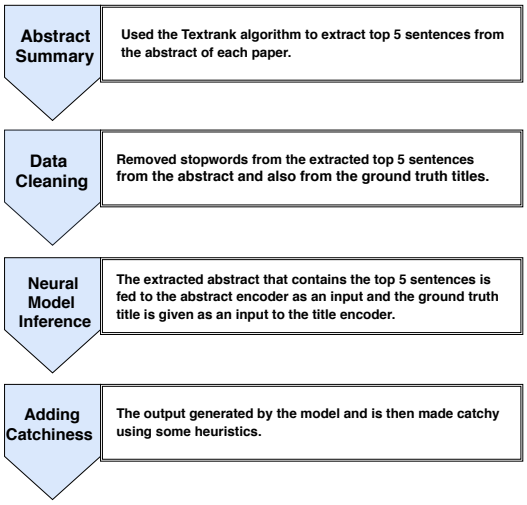

Figure 1: Methodology

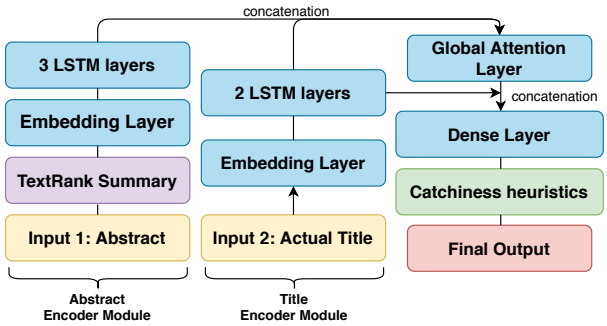

Figure 2: TiZen's Architecture

4.1 Architecture of Neural-Model

The entire architecture used for training the pipeline is as shown in Figure 2. The main motivation for choosing this architecture is that it is currently the most widely used model for abstractive summarization tasks and has achieved very good results on various NLP problems. The input to the system is the extracted abstract [after applying the TextRank algorithm and extracting the top 5 sentences] and the actual title; the former goes to the abstract encoder as an input and the latter to the title encoder.

The abstract encoder architecture consists of an embedding layer followed by three LSTM layers. The title encoder architecture consists of an embedding layer followed by two LSTM layers. The output of the abstract encoder and title encoder are concatenated and given to the Global Attention Layer. The output of the title encoder and the output of the Global Attention layer are further concatenated and provided as input to the time distributed dense layer, which gives the final output. We used Keras with Tensorflow backend as the deep learning framework. For training the network as described Figure 2, we used stochastic gradient descent with rmsprop with a learning rate of 0.001, a discounting factor (ρ) of 0.9 and categorical cross-entropy loss.

The output from global attention is fed to the time distributed dense layer. The output generated from the attention layer will make sense syntactically but may not be catchy. For inference, the weights are saved and the title encoder module is removed, the abstract of the paper is given as input on which again TextRank is applied and it is fed to the abstract encoder module.

4.2 Introducing Catchiness and Catchiness Score

The output from the architecture is made catchy in two ways:

- We look for words having more than three capitalized letters and if there are more than two such words, we choose the one which occurs the most number of times in the abstract. We add this word at the start of the title, followed by a colon and then we add the title generated by the neural model. For example, the research paper "You Only Look Once: Unified, Real-Time Object Detection" [11] will have the word YOLO frequently present in the abstract. The intuition behind the idea is that the authors

nowadays give a allocate unique name to the model, pipeline, or dataset proposed in their work for recognition and identification.
- If no word is found in Step 1, then we look for keywords mentioned just below the abstract in the scientific paper. We then randomly pick any of the keywords, and the title that is given as output is constructed as follows: a "#" concatenated with a random keyword, followed by a colon and then the title predicted by the neural model. The main thought behind this is that the keywords convey the entire topic that the paper is written on and hence appending it in the title makes the reader know what mainly the paper is about.: For example, "Domain Transfer" is a keyword mentioned below the abstract, we append "#Domain Transfer:" at the beginning of the title. If no keywords are found then this step is skipped, and the title predicted by our neural model is returned.

The notion of whether a title is catchy is very subjective we have proposed one possible metric which can quantify catchiness. The basic intuition behind the definition of catchiness is that less frequent or rare content words make a title catchy. The definition of **Title Catchiness** is as follows:

$$\mathrm{TC_G} = -\frac{\sum_{i=1}^{m} \textbf{doc\_count}[actual[i]]}{m}$$

$$\mathrm{TC_P} = -\frac{\sum_{i=1}^{n} \textbf{doc\_count}[predicted[i]]}{n}$$

The definition of **Catchiness Score** is as follows:

$$\mathrm{CS} = \mathrm{TC_G} - \mathrm{TC_P}$$

where **CS** is the Catchiness of predicted title over the actual title, TC_G is the Catchiness Score of the actual title, TC_P is the Catchiness Score of predicted title, **doc_count** contains the counts of words in the given document/scientific paper, **m** is the number of words in the actual title and **n** is the number of words in the title given as an output by TiZen. We are parsing one word at a time in the actual title and the predicted title. Thus, **actual[i]** represents the ith word in the actual title whereas the **predicted[i]** represents the ith word in the predicted title. The count for words that are not present in the hashmap is set to -10. The catchiness score for a title is document dependant as a title might be catchy for one document and not for another. This idea is clearly imposed in the definition of our metric.

If Catchiness Score, CS of predicted title over the actual title > 0, then the title predicted is catchy else not. The basic intuition behind the definition of Catchiness Score is that less frequent or rare words make a title catchy.

5 EXPERIMENTS AND RESULTS

TiZen has outperformed the baseline based on LexRank [4] and TextRank in terms of the BLEU score, as shown in Table **??**. The average BLEU score obtained on the test set is 0.54. It conveys of the title generated by TiZen is more closer to the actual title or closer to titles preferred by authors than that of TextRank and LexRank.

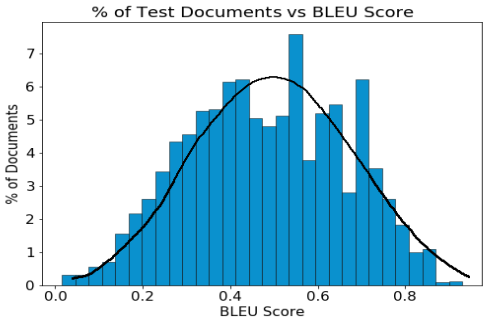

Figure 3: % of Test Documents vs BLEU Scores

On evaluating with our Catchiness Score metric, TiZen model obtains 2.66 on the test set. TiZen performs extremely well compared to the baseline models based on Word-Frequency, TextRank, and LexRank according to the Catchiness Score metric on the test set, as shown in Table 1.

	Without heuristics		With heuristics	
Model	BLEU	CS	BLEU	CS
Word Frequency	0.04	−3.80	0.08	-0.54
TextRank	0.10	0.11	0.18	0.24
TiZen	0.49	1.73	**0.54**	**2.66**

Table 1: Evaluation using BLEU and Catchiness Scores

Figure 3 shows the percentage of test documents against their corresponding BLEU scores. We can clearly see that a greater percentage of the documents have high BLEU scores. It is analogous to the Gaussian curve, which indicates that TiZen generates a good quality title for most % of the documents, and very few documents have poor titles generated. The reason that a Gaussian curve fit is good is that it shows that mots papers achieve an average BLEU title.

A good proportion of papers have a Catchiness Score around 3. This shows that according to the defined metric for catchiness above, most of the papers predicted by TiZen are catchy and thus validates the model.

Table 2 shows the results obtained by running TiZen on a few sample papers. The table mentions the Catchiness Score of predicted title over the actual title. These examples clearly show the robustness and richness of TiZen in generating titles for scientific papers in terms of uniqueness, explainability and identification. For example, the title "YOLO9000: Better, Faster, Stronger" does not convey anything about object detection, however the title generated by TiZen, "YOLOv2: Object Detection Using Single Stage CNN" clearly shows that it is describing object detection thus making it catchy.

We conducted a survey at IIT Gandhinagar to validate the result of TiZen titles through a randomized and double-blind trial in which subjects were unaware of which titles were human or machine-generated. We conducted the survey with 40 subjects in the groups of four and showed 20 different title pairs of title generated by TiZen and actual title to each group. We considered the majority

Titles	Catchiness Score (*CS*)	Abstract of Paper
Actual Title - Multiresolution Recurrent Neural Networks: An Application to Dialogue Response Generation **Predicted Title** - Learning to Generate Spoken Language Models with Recurrent Neural Networks	-0.5	We introduce the multiresolution recurrent neural network, which extends the sequence-to-sequence framework to model natural language generation as two parallel discrete stochastic processes: a sequence of high-level coarse tokens, and a sequence of natural language tokens. ... [14]
Actual Title - FeRoSA: A Faceted Recommendation System for Scientific Articles **Predicted Title** - FeRoSA: A Recommendation System for Scientific Articles	0.3	The overwhelming number of scientific articles over the years calls for smart automatic tools to facilitate the process of literature review. Here, we propose for the first time a framework of faceted recommendation for scientific articles (abbreviated as FeRoSA) which apart from ensuring quality retrieval of scientific articles for a query paper ... [3]
Actual Title - Weakly-Supervised Deep Learning for Domain Invariant Sentiment Classification **Predicted Title** - #DomainTransfer: Domain Adaptation for Sentiment Classification Using Supervised Learning	1.7	The task of learning a sentiment classification model that adapts well to any target domain, different from the source domain, is a challenging problem. Majority of the existing approaches focus on learning a common representation by leveraging both source and target data during training ... Keywords: Sentiment Analysis, Domain Transfer, Weakly labeled datasets [5]
Actual Title - YOLO9000: Better, Faster, Stronger **Predicted Title** - YOLOv2: Object Detection Using Single Stage CNN	2.9	We introduce YOLO9000, a state-of-the-art, real-time object detection system that can detect over 9000 object categories. First we propose various improvements to the YOLO detection method, both novel and drawn from prior work. The improved model, YOLOv2, is ... [12]
Actual Title - SEAGLE: Sparsity-Driven Image Reconstruction under Multiple Scattering **Predicted Title** - SEAGLE: Multi Scale Gradient Descent For Image Enhancement	3.2	Multiple scattering of an electromagnetic wave as it passes through an object is a fundamental problem that limits the performance of current imaging systems. In this paper, we describe a new technique—called Series Expansion with Accelerated Gradient Descent on Lippmann-Schwinger Equation (SEAGLE)—for robust imaging under multiple scattering ... [6]

Table 2: Comparison of actual and predicted titles by `TiZen`.

poll from each pair in a group. Thus, we received a survey of 200 different papers where `TiZen` obtained 122 votes out of 200. This confirms that the titles generated by `TiZen` are indeed more catchy than the original titles which the authors had decided.

6 CONCLUSION AND FUTURE WORK

We have proposed a novel technique for automatic generation of catchy titles for scientific papers. Currently, `TiZen` is trained on computer science scientific papers, and thus the model may not perform well on non-computer science papers. We have introduced a novel concept of catchiness and have to quantify the concept through the Catchiness Score metric for the first time. Experiments show that our metric gives a high score to appealing, catchy, and unique titles, whereas it allocates a low score to cliche and obsolete titles. In the future, we plan to explore the advanced neural models such as the transformer and self-attention layers in our model to improve contextual learning. We plan to incorporate more sections such as the Conclusion, Methodology and Introduction from the paper for title generation to make up for the lack of information in the abstract. Also, we plan to explore the possibility of adding more heuristics based on domain, subject, and publication type to make the title catchy.

REFERENCES

[1] Mehdi Allahyari, Seyed Amin Pouriyeh, Mehdi Assefi, Saied Safaei, Elizabeth D. Trippe, Juan B. Gutierrez, and Krys Kochut. 2017. A Brief Survey of Text Mining: Classification, Clustering and Extraction Techniques. *CoRR* abs/1707.02919 (2017). arXiv:1707.02919 http://arxiv.org/abs/1707.02919

[2] Mehdi Allahyari, Seyed Amin Pouriyeh, Mehdi Assefi, Saeid Safaei, Elizabeth D. Trippe, Juan B. Gutierrez, and Krys Kochut. 2017. Text Summarization Techniques: A Brief Survey. *CoRR* abs/1707.02268 (2017). arXiv:1707.02268 http://arxiv.org/abs/1707.02268

[3] Tanmoy Chakraborty, Amrith Krishna, Mayank Singh, Niloy Ganguly, Pawan Goyal, and Animesh Mukherjee. 2016. FeRoSA: A Faceted Recommendation System for Scientific Articles. 528–541. DOI:http://dx.doi.org/10.1007/978-3-319-31750-2_42

[4] Günes Erkan and Dragomir R. Radev. 2004. LexRank: Graph-based Lexical Centrality as Salience in Text Summarization. *J. Artif. Intell. Res. (JAIR)* 22 (2004),

457–479. http://dblp.uni-trier.de/db/journals/jair/jair22.html#ErkanR04

[5] Pratik Kayal, Mayank Singh, and Pawan Goyal. 2020. Weakly-Supervised Deep Learning for Domain Invariant Sentiment Classification. *Proceedings of the 7th ACM IKDD CoDS and 25th COMAD* (Jan 2020). DOI: http://dx.doi.org/10.1145/3371158.3371194

[6] Hsiou-Yuan Liu, Dehong Liu, Hassan Mansour, Petros Boufounos, Laura Waller, and Ulugbek Kamilov. 2017. SEAGLE: Sparsity-Driven Image Reconstruction under Multiple Scattering. *IEEE Transactions on Computational Imaging* PP (05 2017). DOI: http://dx.doi.org/10.1109/TCI.2017.2764461

[7] H. P. Luhn. 1958. The Automatic Creation of Literature Abstracts. *IBM J. Res. Dev.* 2, 2 (April 1958), 159–165. DOI: http://dx.doi.org/10.1147/rd.22.0159

[8] Rada Mihalcea and Paul Tarau. 2004. TextRank: Bringing Order into Text. In *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, Barcelona, Spain, 404–411. https://www.aclweb.org/anthology/W04-3252

[9] Monarch Parmar, Naman Jain, Pranjali Jain, P. Jayakrishna Sahit, Soham Pachpande, Shruti Singh, and Mayank Singh. 2020. NLPExplorer: Exploring the Universe of NLP Papers. *Advances in Information Retrieval* Advances in Information Retrieval (2020), 476–480. DOI: http://dx.doi.org/10.1007/978-3-030-45442-5_61

[10] Jan Wira Gotama Putra and Masayu Leylia Khodra. 2017. Automatic title generation in scientific articles for authorship assistance: a summarization approach. *Journal of ICT Research and Applications* 11, 3 (2017), 253–267.

[11] Joseph Redmon, Santosh Kumar Divvala, Ross B. Girshick, and Ali Farhadi. 2015. You Only Look Once: Unified, Real-Time Object Detection. *CoRR* abs/1506.02640 (2015). arXiv:1506.02640 http://arxiv.org/abs/1506.02640

[12] Joseph Redmon and Ali Farhadi. 2016. YOLO9000: Better, Faster, Stronger. *CoRR* abs/1612.08242 (2016). arXiv:1612.08242 http://arxiv.org/abs/1612.08242

[13] Guergana Savova, James Masanz, Philip Ogren, Jiaping Zheng, Sunghwan Sohn, Karin Kipper-Schuler, and Christopher Chute. 2010. Mayo Clinical Text Analysis and Knowledge Extraction System (cTAKES): Architecture, component evaluation and applications. *Journal of the American Medical Informatics Association : JAMIA* 17 (09 2010), 507–13. DOI: http://dx.doi.org/10.1136/jamia.2009.001560

[14] Iulian Vlad Serban, Tim Klinger, Gerald Tesauro, Kartik Talamadupula, Bowen Zhou, Yoshua Bengio, and Aaron C. Courville. 2016. Multiresolution Recurrent Neural Networks: An Application to Dialogue Response Generation. *CoRR* abs/1606.00776 (2016). arXiv:1606.00776 http://arxiv.org/abs/1606.00776

[15] Alexander Shvets et al. 2019. Improving Scientific Article Visibility by Neural Title Simplification. *arXiv e-prints* arxiv (Apr 2019). arXiv:cs.IR/1904.03172

[16] Mayank Singh, Barnopriyo Barua, Priyank Palod, Manvi Garg, Sidhartha Satapathy, Samuel Bushi, Kumar Ayush, Krishna Sai Rohith, Tulasi Gamidi, Pawan Goyal, and Animesh Mukherjee. 2016. OCR++: A Robust Framework For Information Extraction from Scholarly Articles. *CoRR* abs/1609.06423 (2016). arXiv:1609.06423 http://arxiv.org/abs/1609.06423

[17] Elizabeth D. Trippe, Jacob B. Aguilar, Yi H. Yan, Mustafa V. Nural, Jessica A. Brady, Mehdi Assefi, Saeid Safaei, Mehdi Allahyari, Seyedamin Pouriyeh, Mary R. Galinski, Jessica C. Kissinger, and Juan B. Gutierrez. 2017. A Vision for Health Informatics: Introducing the SKED Framework.An Extensible Architecture for Scientific Knowledge Extraction from Data. *arXiv e-prints* arxiv (Jun 2017). arXiv:q-bio.QM/1706.07992

[18] 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 (SIGIR '07)*. Association for Computing Machinery, New York, NY, USA, 127–134. DOI: http://dx.doi.org/10.1145/1277741.1277766

[19] Oleg Vasilyev, Tom Grek, and John Bohannon. 2019. Headline Generation: Learning from Decomposable Document Titles. *arXiv e-prints* (Apr 2019). arXiv:cs.CL/1904.08455