IAES International Journal of Artificial Intelligence (IJ-AI)
Vol. 14, No. 2, April 2025, pp. 1587~1596
ISSN: 2252-8938, DOI: 10.11591/ijai.v14.i2.pp1587-1596

Abstractive summarization using multilingual text-to-text transfer transformer for the Turkish text

Neda Alipour[1], Serdar Aydın[2]
[1]Department of Management Information Systems, Faculty of Economics and Administrative Science, Atatürk University, Erzurum, Türkiye
[2]Department of Software Engineering, Faculty of Engineering, Atatürk University, Erzurum, Türkiye

Article Info

Article history:

Received Oct 22, 2023
Revised Nov 17, 2024
Accepted Nov 24, 2024

Keywords:

Abstractive summarization
Dataset
Deep learning
Pre-trained
Turkish text

ABSTRACT

Today, with the increase in text data, the application of automatic techniques such as automatic text summarization, which is one of the most critical natural language processing (NLP) tasks, has attracted even more attention and led to more research in this area. Nowadays, with the developments in deep learning, pre-trained sequence-to-sequence (text-to-text transfer converter (T5) and bidirectional encoder representations from transformers (BERT) algorithm) encoder-decoder models are used to obtain the most advanced results. However, most of the studies were done in the English language. With the help of the recently emerging monolingual BERT model and multilingual pre-trained sequence-to-sequence models, it has led to the use of state-of-the-art models in languages with fewer resources and studies, such as Turkish. This article used two datasets for Turkish text summarization. First, Google multilingual text-to-text transfer transformer (mT5)-small model was applied on multilingual summarization (MLSUM), which is a large-scale Turkish news dataset, and success was examined. Then, success was evaluated by first applying BERT extractive summarization and then abstractive summarization on 1010 articles collected on the Dergipark site. Rouge measures were used for performance evaluation. This study is one of the first examples in the Turkish language and it is considered to provide a basis for future studies with good results.

Corresponding Author:

Neda Alipour
Department of Management Information Systems, Faculty of Economics and Administrative Science
Atatürk University
Erzurum 25240, Türkiye
Email: neda.alipour14@ogr.atauni.edu.tr

1. INTRODUCTION

With the advent of the internet in the digital age, there has been a massive increase in access to textual information. Automatic text summarization, which is one of the different natural language processing (NLP) tasks, helps to obtain more compact and efficient versions of text content in a shorter time by obtaining the most important information [1], [2]. Thus, it was tried to overcome the difficulties that emerged with the increase in data. With the increase in data and due to repetitive and irrelevant content, it is necessary to spend more time and effort to obtain important information by humans. For this reason, automatic text summarization has been one of the issues that should be studied and unavoidable lately. For automatic text summarization, text retrieval systems are used to display a summarized version of search results in search engines [3].

According to the Moreno [4] list, text summarization can be viewed from different angles, including single-document [5] and multi-document [6], [7] in terms of number of input documents, monolingual and multilingual [8] in terms of number of input languages, and extractive, abstractive and hybrid summarization in terms of output generation approach: extractive summaries define the candidate sentences according to features such as the length of the sentence, the position of the sentences relative to each other, and the ratio of nouns, by creating a sentence scoring mechanism for the most important sentences from the source and combines them to form the summary. Candidate sentences are ranked according to the specified features and scoring, and the candidate sentences at the top of the requirement are selected [9]. In abstractive summarization, new expressions are produced with sentences that are not found in the original text and are tried to obtain a summary by linguistic methods that are used for understanding and examining the text [10], [11]. Abstractive text summaries are more attractive, by using of complex natural language comprehension and rendering capabilities to produce human-like summaries. Therefore, in recent years, abstractive techniques in different languages have attracted more attention with advances in deep learning. Since text summarization can be seen as a sequence-to-sequence (Seq2Seq) task, there are different approaches to abstract text summarization, especially for the English language. Text summarization is a Seq2Seq task. Encoder-decoder architecture-based Seq2Seq models have gained significant attention in recent years. As a result, there has been a shift from long short-term memory (LSTM)-based models [12] to transformer-based models in encoder-decoder networks [13].

In recent years, studies have shown very good performance using pre-training Seq2Seq models on very large datasets to improve text summarization [14]–[17] and achieving state-of-the-art results in neural abstractive summarization [18]. Unfortunately, most of the research has been done in English only, and models that need pre-trained require large amounts of data and computational power. However, recently multilingual versions of bidirectional encoder representations from transformers (BERT) [19] the widely used two pre-trained multilingual Seq2Seq models multilingual text-to-text transfer transformer (mT5) [20] and multilingual bidirectional and auto-regressive transformers (mBART) [19], [21] have led to studies in several research areas for low-resource languages. The mT5 [20] model, which covers 101 different languages and is trained on a common language, is a multilingual version of the text-to-text transfer converter (T5) model. The mT5 model is a suitable option for most languages due to its multilingual feature.

It seems that for Turkish language extractive automatic text summarization studies have been done more. However, there are very few studies focused on abstractive text summarization for Turkish [22], [23]. In these studies, pre-trained Seq2Seq models were used less frequently. Previous studies in NLP have demonstrated that techniques developed for languages like English often perform poorly on morphologically rich languages, such as Turkish. This highlights the need for additional methods that account for the unique morphological structures in these languages [24]. For example, Turkish is an agglutinative language in which root words can acquire numerous derivatives and inflections. This characteristic results in a wide variety of unique word surface forms, leading to challenges with data sparsity [25].

Section 2 of this article summarizes related work and their achievements. Sections 3 and 4 provide an overview of the datasets and the mT5 method. Finally, sections 5 and 6 conclude the article by presenting the conclusions of the study and suggestions for future studies.

2. RELATED WORK

2.1. Pre-trained sequence-to-sequence models

In recent years, state-of-the-art results of transfer learning in NLP, which has been very effective, have emerged in different tasks. It has been determined that the concept of pre-training of a language model that can learn task-agnostic knowledge in natural language comprehension and then transfer it to subsequent tasks is successful [19], [26], [27]. However, new research is turning to pre-trained Seq2Seq models because pre-trained encoder models do not work well for tasks that require natural language generation and natural language understanding, such as text summarization and machine translation. Song *et al.* [15] proposed masked sequence to sequence pre-training (MASS) with help from BERT in reconstructing the rest of the sentence to create an encoder-decoder based language. A sentence containing a randomly masked part in the encoder part is used as input and the decoder part tries to guess this masked part. Thus, the MASS model can train the encoder and decoder together. Dong *et al*. [14] introduced a new pre-trained unified language model (UniLM) that incorporates bidirectional, unidirectional, and Seq2Seq predictive language modeling tasks. This model can be fine-tuned for tasks involving both understanding and generation of natural language. UniLM emerged using a shared transformer network and using certain self-attention masks to control context in which the prediction conditions are. Lewis *et al*. [17] trained with bidirectional and auto-regressive transformers (BART), one of the autoencoder Seq2Seq models to generate new text by first distorting the text and then learning a model. For this purpose, they used a standard transformer-based neural machine translation architecture. Fine-tuning BART has shown to be effective for text creation and comprehension tasks. Raffel *et al*. [28] provided an overview of transfer learning techniques for NLP, also they compared pre-training goals, transfer approaches, architectures, unlabeled datasets and other factors for language understanding. Xue *et al*. [20] introduced mT5, a pre-trained multilingual variant of T5 available in 101 languages, on a new common scan-based dataset. It demonstrated state-of-the-art performance in multilingual benchmarks by detailing the mT5's modified training and design. Liu *et al*. [21] found that multilingual denoising pre-training for a wide range of machine translation yields huge performance gains. Devlin *et al*. [19] offered mBART inspired by BART to pre-train the Seq2Seq model.

2.2. Abstractive text summarization

With the development of deep learning, encoder decoder Seq2Seq networks have started to gain more importance for abstractive text summarization. Rush *et al*. [29] proposed a neural network language model (NNLM), a neural local attention-based model that can be easily trained and scalable to training data for abstractive sentence summarization. Chopra *et al*. [10] proposed a convolutional attention-based encoder model as a simplified version of the encoder-decoder framework using a recurrent neural network (RNN) for abstractive sentence summarization. It was used two-layer LSTMs for the encoder-decoder containing 500 hidden units in each layer. Nallapati *et al*. [11] proposed a new dataset of multi-sentence summaries and several new models for abstractive text summarization using bidirectional LSTM-based encoder-decoder, such as feature rich encoder, modeling keywords, and a hierarchical encoder-decoder that is capable of capturing the document structure. Celikyilmaz *et al*. [30] extended the CommNet model of [31] on CNN/DailyMail and New York Times datasets for abstractive summarization with deep communication agents in an encoder-decoder architecture. Paulus *et al*. [32] introduced a new training method for abstractive summarization that combines standard supervised word prediction and a neural network model with reinforcement learning (RL) on the CNN/DailyMail dataset. Narayan *et al*. [33] proposed extreme summarization based on convolutional neural networks on a large-scale dataset by collecting online articles from the British Broadcasting Corporation (BBC) for single-document abstractive summarization and for creating a one-sentence news summary. Liu and Lapata [34] presented a general framework on the CNN/DailyMail news highlights dataset [35] and the New York Times Annotated Corpus [36] for both extractive and abstractive models, and on Xsum [33] for the BERT-based coder. in this model, a new fine-tuning program is proposed for abstractive summarization, which adopts different optimizers for encoder and decoder. Devlin *et al*. [19] introduced BART, a pre-training auto-encoder approach. According to the authors, BART works well for text generation and text comprehension tasks when fine-tuned. Zhang *et al*. [18] proposed pre-training with extracted gap-sentences for abstractive summarization (PEGASUS), which pre-trained the large transformer-based encoder-decoder for abstractive text summarization. PEGASUS selects and masks important sentences in the document and creates gap sentences as a pre-training target. They evaluated the best PEGASUS models for 12 downstream summaries covering science, news, stories, instructions, patents, emails, and bills. Qi *et al*. [37] introduced a new self-supervised objective, future n-gram prediction, which was tested on the CNN/DailyMail, Gigaword, and SQuAD 1.1 benchmarks for tasks like question generation and abstractive summarization. They also developed a Seq2Seq pre-trained model called ProphetNet, featuring an n-stream self-attention mechanism. In contrast to conventional Seq2Seq models, ProphetNet is optimized for n-step forward prediction, predicting the next n tokens based on previous context tokens at each time step. ProphetNet was pre-trained on both a base-scale dataset (16 GB) and a large-scale dataset (160 GB).

2.3. Turkish text summarization

In study by Altan [38], the system was developed by single Turkish document as input and scoring was carried out using features such as sentence location information and term frequency information, and summaries were obtained using a number of statistical methods. Kutlu *et al*. [39] proposed a general text summarization method based on sentence ordering. The system calculated sentence scores using surface-level features and produced summaries by selecting the highest-scoring sentences from the original documents. Features such as sentence position, title similarity, key phrase centrality and and term frequency were applied. The study emphasized the effectiveness of centrality as a feature and was one of the first to showcase the use of key phrases in summarizing Turkish texts [39], [40]. The authors argued that the cross method developed in the study outperformed other latent semantic analysis (LSA) methods. Ozsoy *et al*. [41] introduced two new LSA-based hashing algorithms and presented a general extractive text summarization system for Turkish, based on LSA. Pembe [42] proposed a rule-based approach for automatic document summarization based on information requests and text structure for search engines. After scoring the sentences using the position, title (the frequency of occurrence of the terms in the title in the sentence), query sentence and term frequency methods (the value obtained from the frequency of occurrence of the terms in the sentence in the whole document), scores were given according to the importance of the sentences and sentence selection was carried out. Güran [43] proposed a new weight value for extractive text summarization that can be used in text summarization methods based on LSA. In this study, a hybrid system was proposed with two different approaches that combine semantic and structural features for important sentence extraction. Abstractive text summarization studies using Seq2Seq models are very few and limited for Turkish texts. Scialom *et al.* [22] presented multilingual summarization (MLSUM), the first large-scale MLSUM dataset, in five different languages (Turkish, Russian, Spanish, German, and French), including over 1.5 million article/summary pairs from online newspapers, to evaluate Seq2Seq models. This study reports cross-language comparative analysis based on state-of-the-art systems. In the study of [44], an encoder-decoder model was developed for the prediction of abstractive Turkish news headlines and the system was trained with RNN. FastText model was used for word placement in news texts. Baykara and Güngör [23] evaluated several morphological tokenization methods using the pointer-generator model, presenting two large-scale datasets (HU-News and TR-News) to generate abstractive summarization in Turkish and Hungarian. They also compared the results obtained from the TR-News dataset with BERT-based models.

3. DATASET AND RESEARCH METHODOLOGY

In the text summarization area, most datasets are available in English, and datasets in other languages such as Turkish are limited. In this study, firstly, the MLSUM [21] news dataset and then the article dataset created by the author were used. MLSUM covers 5 languages as French, German, Spanish, Turkish, and Russian and is known as text summarization dataset. The MLSUM dataset was created from the popular CNN/DailyMail dataset. The article dataset was collected from the Dergipark and all subjects were included. There are 1010 articles in this dataset.

This section provides an overview of the mT5 [20] architecture that is the multilingual variant of the T5 model [28]. T5 is pre-trained text-to-text encoder-decoder transformer model which closely follows the originally proposed transformer architecture [13] and can be used for all text-based NLP problems [20] and covers the following goals: predicting the next word with language modeling, redefining the original text with de-shuffling, and predicting masked words with corrupting spans [5].

This approach is an NLP framework for generative tasks such as text summarization, question answering and text classification where the task format allows the model to generate text based on some input [20], [23]. As a result, the same hyperparameters and loss function are applied across each task [5]. Figure 1 illustrates the T5 model as a unified framework for downstream NLP tasks. Each downstream task in text-to-text format is represented by a different color: translation (green), linguistic acceptability (red), sentence similarity (yellow), and text summarization (blue) [28]. Although T5 model was trained only for English language, mT5 model was trained on 101 different languages (including Turkish) and inherits all capabilities of the T5 model. mT5 looks more powerful compared to other models such as BERT, cross-lingual language model with RoBERTa (XLM-R), and multilingual BERT [5].

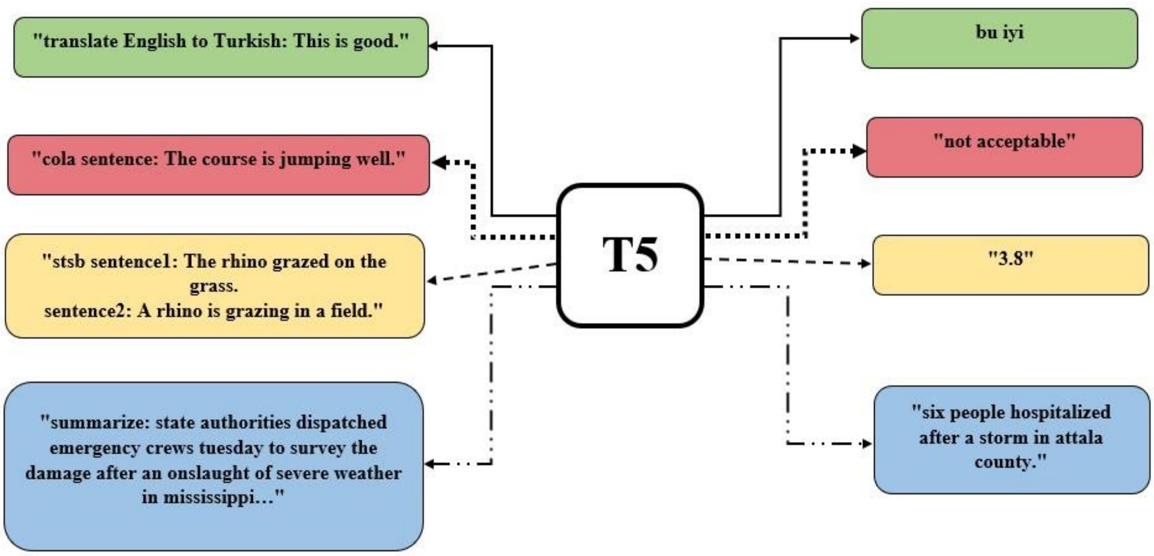

Figure 1. mT5 framework

4. RESULTS AND DISCUSSION

In this paper, fine-tune of mT5 was used for Turkish news and Turkish papers summarization. We used Adam optimizer, 8 and 16 batch size and 15 training epochs as fine-tune parameters. First of all, training was carried out on Turkish news with 8 and 16 batch sizes and 15 epochs. To evaluate the model, the results were evaluated with rouge metrics. Rouge metrics most commonly used to evaluate text summarization and translation values. In this study, the success of the model was examined with obtaining of Rouge 1, Rouge 2, and Rouge L. Rouge-N is a method that scores the sensitivity value between the reference abstract and the candidate abstract according to n-gram overlap. It tries to find the repetition rate of the parts divided by the number n in an N-gram word string. Similarly, Rouge-L value uses the longest common word subsequence between two different abstracts [45]. Train and validation loss for Turkish news with 8 batch size were shown in Figure 2. Train and validation loss for Turkish news with 16 batch size were shown in Figure 3.

After training the model with batch sizes of 8, 16, and 15 epochs, the most commonly used Rouge metrics were used and the success of the model was examined by obtaining Rouge 1, Rouge 2, and Rouge L. Rouge-1, Rouge-2, and Rouge-L values were shown in Table 1. The success achieved was compared by the success of other studies. Ahuir *et al.* [46] worked on abstractive summarization with mT5 in Spanish and Catalan. They obtained the following values as Rouge-1, Rouge-2, and Rouge-L (Table 2). Pant and Chopra [47] worked on summative summarization with mT5 in Spanish and Greek documents. They just evaluated the Rouge-2 metric and got 13.1 for Spanish and 13.8 for Greek. When the values obtained are compared with the values of these two studies, it is obvious that the results are close to the values of the first study and better than the second study. To illustrate the performance of this model in more detail, two examples from the dataset are presented in the Table 3.

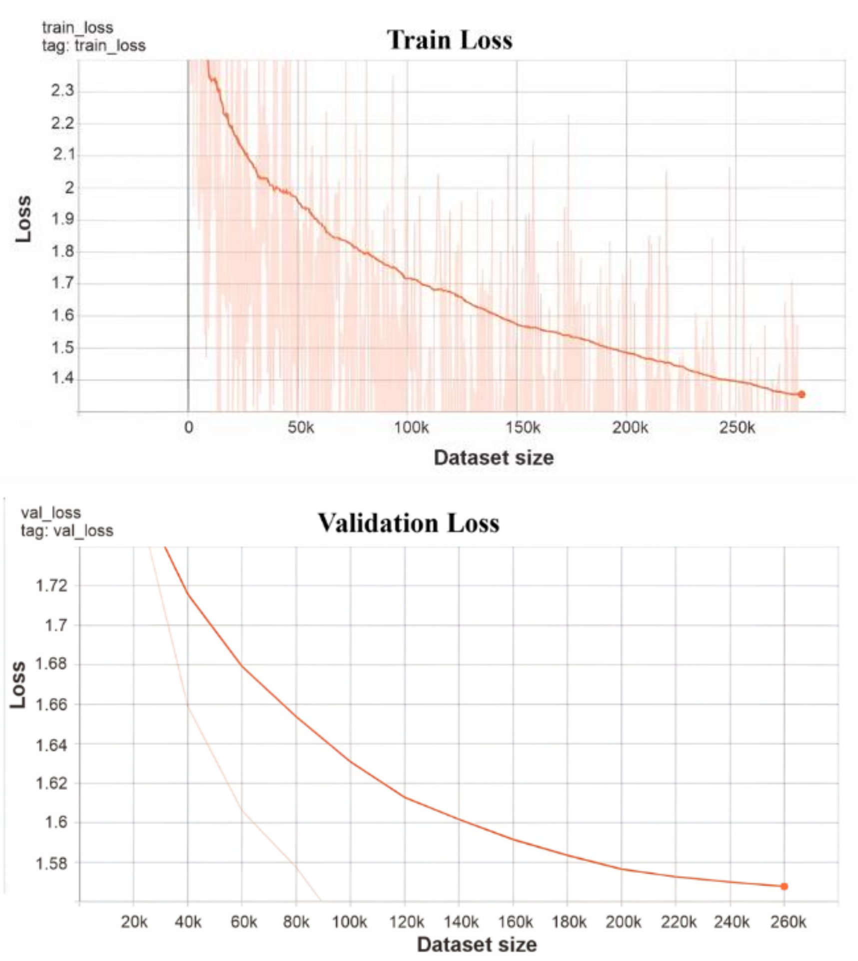

Figure 2. Train and validation loss for news dataset with 8 batch-size

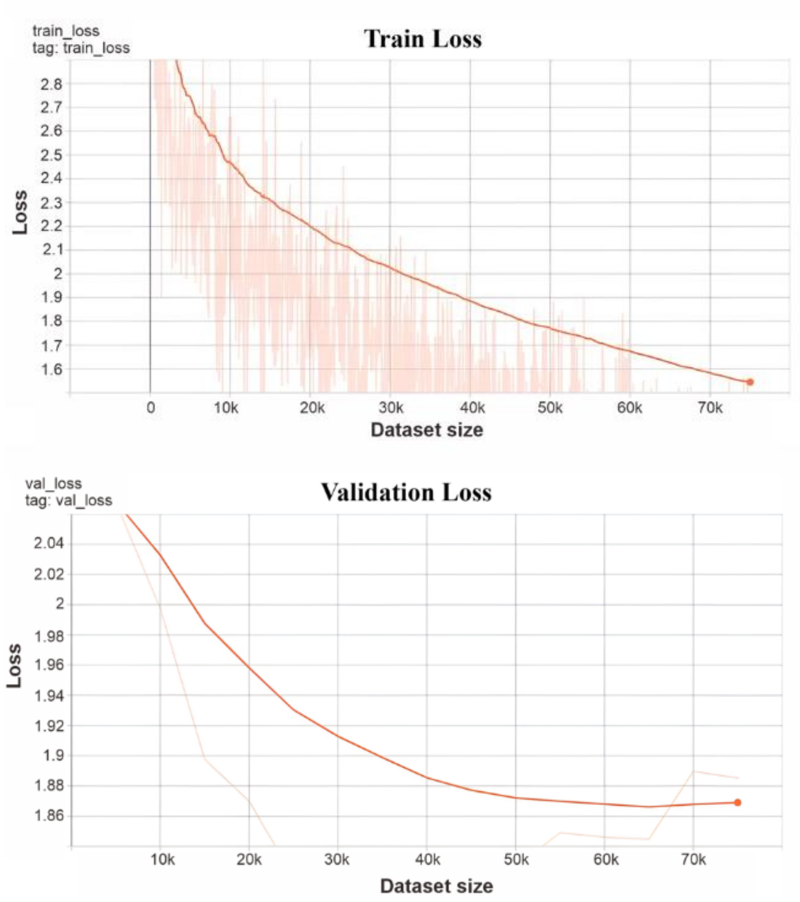

Figure 3. Train and validation loss for news dataset with 16 batch-size

Table 1. Rouge metrices for Turkish news dataset

Batch size	Rouge-1	Rouge-2	Rouge-L
8	31.98	21.11	30.93
16	28.43	17.61	27.40

Table 2. Rouge metrices in [46]

Language	Rouge-1	Rouge-2	Rouge-L
Spanish	30.61	12.36	23.53
Catalan	27.00	11.28	21.27

In the continuation of the study, it was aimed to increase the success by changing the hyper parameters. For this purpose, 0.00004 was selected for the learning rate and the system was retrained on the news dataset and the result was evaluated with Rouge metrics and shown in Table 4. Farahani *et al.* [5] have achieved 42.25, 24.36, and 35.94 successes in their study for Persian, respectively. In addition, Baykara and Güngör's [23] highest values obtained in their study for Turkish were 42.26, 27.81, and 37.96, respectively. When the values of this study are compared with above studies, it can be said that this study achieved a good result.

In addition, another study, which is in English NLP studies but has not been done in Turkish until now, was processed by the author and its effect on text summarization was examined. For this purpose, all the plural suffixes in the text and the suffixes following the plural suffix were deleted from the word and thus the number of words was reduced and the effect on the text summarization was examined. After removing the plural suffix and training the system with this dataset, the result was evaluated with Rouge metrics and shown in Table 5. Sometimes meaningless summaries were produced after removing plural suffixes, and the result was also very low compared to other results.

Table 3. Examples of abstractive summarization for Turkish news dataset

Example	Main	Original abstract	mT5 abstract
Example 1	"Fenerbahçe kulübü, İtalya birinci futbol ligi ekiplerinden Empoli'nin orta saha oyuncusu Miha Zajc'ı kadrosuna kattığını resmen açıkladı. Sarı-lacivertli kulübün internet sitesinde yer alan açıklamada, 24 yaşındaki Zajc ile 4,5 yıllık anlaşmaya varıldığı belirtildi. Zajc transferi için Empoli'de kiralık olarak forma giyen Salih Uçanın haklarından vazgeçildiği de duyuruldu. Açıklamada, "Kulübümüz, İtalya Seri A ekiplerinden Empoli takımında forma giyen merkez orta saha ve ofansif orta saha oyuncusu Miha Zajc bonservisiyle birlikte kadromuza katmak üzere kulübüyle ve futbolcuyla anlaşmaya varmıştır. 24 yaşındaki Sloven oyuncu Miha Zajc, 4,5 sezon boyunca sarı-lacivertli forma ile mücadele edecek. Oyuncumuz Miha Zajc'a Fenerbahçeye hoş geldin diyor, çubuklu ile nice başarıla diliyoruz. Ayrıca, bu transfer kapsamında kulübümüz Salih Uçan üzerindeki haklarından vazgeçerek, Empoli ile Salih Uçan'ın anlaşmasına müsaade etmiştir" ifadeleri yer aldı."	"Fenerbahçe Sloven orta saha oyuncusu Miha Zajc'ı 4,5 yıllığına transfer ettiğini resmen açıkladı. Bu sezon İtalya birinci futbol ligi ekibi Empoli'de 21 resmi maçta görev alan Zajc, 3 gol kaydetti."	"Fenerbahçe kulübü, Empoli'nin orta saha oyuncusu Miha Zajc'ı kadrosuna kattığını resme açıkladı."
Example 2	"Muğla'nın Bodrum ilçesinde, içerisinde askeri personelin bulunduğu minibüs su kanalına düştü. Kazada minibüste bulunan 3 asker hafif şekilde yaralandı. Kaza, bugün akşam saatlerinde Bodrum-Milas karayolu üzerinde meydana geldi. Askeri personel taşıyan 48 TN 173 plakalı minibüs, Güvercinlik istikametine giderken, sağanak yağış sonrası kayganlaşan yolda sürücü direksiyon hakimiyetini kaybetti. Kontrolden çıkan minibüs, önce yol kenarında bulunan su kanalına düştü, daha sonra da kayalıklara çarparak durabildi. Kazanın ardından olay yerine gelen Muğla 911 Arama Kurtarma ekipleri hafif yaralı askerleri araştan çıkararak sağlık ekiplerine teslim etti. Yaralan 3 askerden 2'si Bodrum Devlet Hastanesi'ne, 1 asker ise özel bir hastaneye kaldırıldı. Tedaviye alınan 3 askerin de sağlık durumu iyi olduğu öğrenildi."	"Muğla'nın Bodrum ilçesinde askeri personel taşıyan askeri minibüs kaza yaptı. Yol kenarında bulunan su kanalına düşen minibüste bulunan 3 asker hafif şekilde yaralandı."	"Muğla'nın Bodrum ilçesinde, içerisinde askeri personelin bulunduğu minibüs su kanalına düştü. Kazada 3 asker yaralandı."

Table 4. Rouge metrices for Turkish news dataset with changing learning rate

Batch size	Learning rate	Rouge-1	Rouge-2	Rouge-L
8	0.00004	58.76	52.98	58.45

Table 5. Rouge metrices for Turkish news dataset by removing plural suffixes

Batch size	Rouge-1	Rouge-2	Rouge-L
8	10.55	3.89	10.21

After the news dataset, the model was tested on the paper dataset. However, since the paper dataset was large, the papers were first reduced to 26 lines with the BERT extractive summarization method. Because the smallest paper had 26 lines. Thus, the size of the paper dataset was reduced and given to the system. The paper data were trained with 8 batch size and 15 epochs. because, better result was obtained with this batch size on Turkish news. Train and validation loss for papers with 8 batch size were shown in Figure 4. After training the model with batch sizes of 8 and 15 epochs, the most commonly used Rouge metrics were used and the success of the model was examined by obtaining Rouge 1, Rouge 2, and Rouge L. Rouge-1, Rouge-2, and Rouge-L values were shown in Table 6.

Table 6. Rouge metrices for papers dataset

Batch size	Rouge-1	Rouge-2	Rouge-L
8	18.34	4.62	17.63

Until now, no abstractive summarization process with mT5 has been performed on the paper dataset. In addition, it is not possible to compare the results of this study with other studies, because Turkish article dataset was created by the author. Also, the reason for the low rouge metrics is that the text is meaningful but long. So, the produced texts may differ from the actual abstracts. Therefore, the results of this study can form a basis for future summarization studies. An example from the dataset is presented in the table to illustrate the performance of this model in more detail. Since the text is long, only the summary and the summary produced with the model are presented in Table 7.

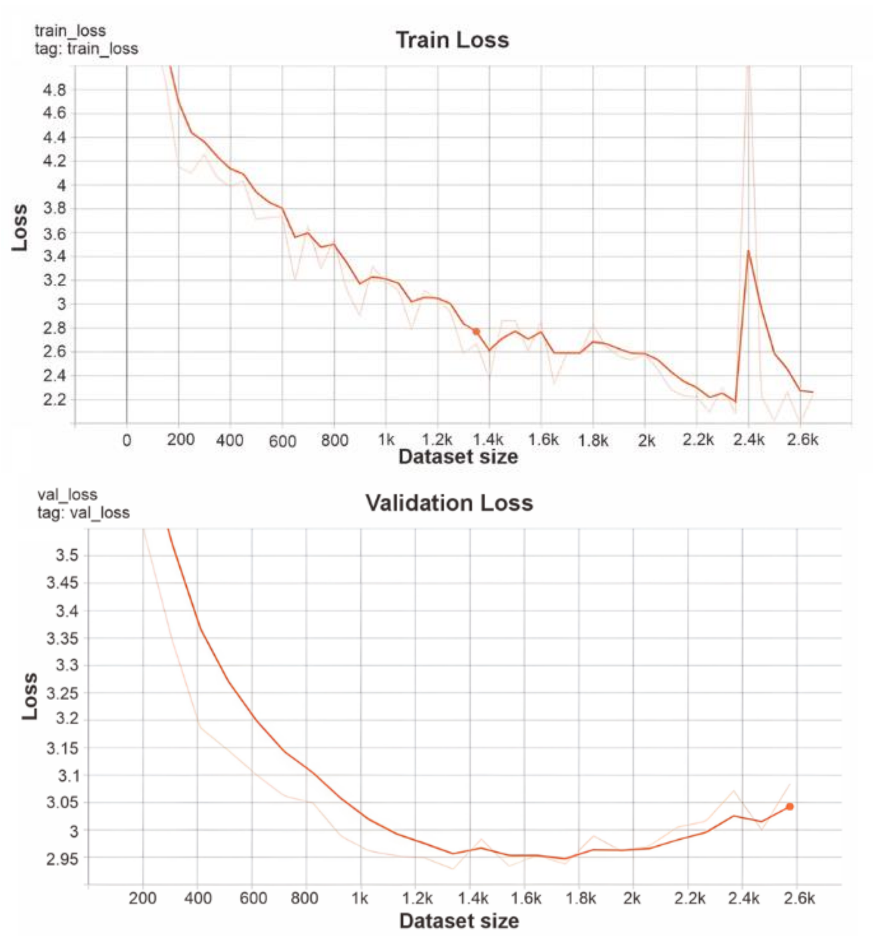

Figure 4. Train and validation loss for papers dataset

Table 7. Examples of abstractive summarization for papers dataset

Original Abstract	mT5 Abstract
"ikinci dünya savaşı sonrası ülkelerarası gelişmişlik farklarının belirginleşmesiyle azgelişmiş veya geri kalmış ülkelerin ekonomik olarak kalkınması son derece ciddi bir sorun olarak ortaya çıkmıştır. bu süreçte gelişmişlik farlılıkları bakımından benzer olmayan ülkelerin benzer büyüme modellerini uygulamalarının mümkün olmadığı görülmüş ve kalkınma çabalarında yeni arayışlara yönelmiştir. bu dönemde oluşmaya başlayan büyüme teorilerinin özünü savaş sonrası savaştan etkilenen ekonomilerin kalkındırılması oluşturmuştur. bu doğrultuda gelişen büyüme teorileri ülkelerin gelişme çabalarında önemli rol oynamıştır. büyümenin temel belirleyicileri üzerinde yapılan değerlendirmeler ile gelişme yolunda ivme kazanılmıştır. ancak ülkelerin kalkınmalarında temel belirleyicilerden olan eğitim faktörü beşeri sermaye oluşumuna katkı sağlayarak iktisadi büyümede önemli olmaktadır. eğitim beşeri sermaye teorisinin kilit unsurlarından biridir çünkü bilgi ve beceriyi geliştirmenin birincil yolu olarak görülmektedir. buna göre eğitim düzeyi emek kalitesini ölçmenin bir yolu olarak ele alınmaktadır. nitelikli eğitim ise beşeri sermaye oluşumunun temelini oluşturmaktadır. büyüme teorilerinin gelişimi."	"bu çalışmada türkiye de mesleki ve teknik eğitim kurumları itibariyle nüfus beşeri sermayenin gelişmesiyle birlikte yürütülmüştür. bu kapsamda türkiye de mesleki ve teknik eğitim kurumları itibariyle ekonomik büyüme modellerinin ortaya çıktığı bir dönemdir. türkiye de mesleki ve teknik eğitim kurumları itibariyle insan sermayesi ile sağlanmıştır. bu dönemde ekonominin sürdürülebilirliğinin artmasına yönelik sonuçlar ortaya çıkan nitelikli işgücü ihtiyacı karşılamaktadır. bu durumun sonunda beşeri sermayenin gelişmesine katkı sağladığı düşünülmektedir."

5. CONCLUSION

There are a limited number of studies on abstractive summarization with pre-trained models in Turkish texts. In this study, a pre-trained mT5 model was used to summarize Turkish texts. This model was first tested on MLSUM dataset which is news dataset, and then on the article dataset created by the author. Since the articles were long, their sentences were reduced by the BERT extractive summarization. Due to a lack of studies and dataset in article dataset, preparing the dataset was one of the most important limitations and difficulties of the study and our paper could not be compared to any earlier study and can now serve as a baseline for any future studies in this field. Another most important limitation is the system inadequacy. Text datasets require high hardware, and although this work ran on Google Colab Pro Plus, it was met with hardware failure errors in most cases. In addition, Turkish language is an agglutinative language and NLP studies are very difficult in Turkish. In future studies, it can be examined whether success has changed by enlarging the article dataset. In addition, the success of the model can be evaluated by enriching the system hardware, changing the hyperparameters and doubling the dataset. On the other hand, in addition to text summarization, question generator can be important research.

REFERENCES

[1] A. Nenkova and K. McKeown, "A survey of text summarization techniques," in *Mining Text Data*, Boston, MA: Springer US, 2012, pp. 43–76, doi: 10.1007/978-1-4614-3223-4_3.
[2] H. P. Edmundson, "New methods in automatic extracting," *Journal of the ACM*, vol. 16, no. 2, pp. 264–285, 1969, doi: 10.1145/321510.321519.
[3] A. Turpin, Y. Tsegay, D. Hawking, and H. E. Williams, "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*, New York, USA: ACM, 2007, pp. 127–134, doi: 10.1145/1277741.1277766.
[4] J. M. T. Moreno, *Automatic text summarization*. Hoboken, New Jersey: John Wiley & Sons, 2014, doi: 10.1002/9781119004752.
[5] M. Farahani, M. Gharachorloo, and M. Manthouri, "Leveraging parsbert and pretrained mT5 for persian abstractive text summarization," in *26th International Computer Conference, Computer Society of Iran, CSICC 2021*, IEEE, 2021, pp. 1–6, doi: 10.1109/CSICC52343.2021.9420563.
[6] J. Christensen, Mausam, S. Soderland, and O. Etzioni, "Towards coherent multi-document summarization," in *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Association for Computational Linguistics, 2013, pp. 1163–1173.
[7] A. Nenkova, L. Vanderwende, and K. McKeown, "A compositional context sensitive multi-document summarizer," in *Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval*, New York, USA: ACM, 2006, pp. 573–580, doi: 10.1145/1148170.1148269.
[8] M. Gambhir and V. Gupta, "Recent automatic text summarization techniques: a survey," *Artificial Intelligence Review*, vol. 47, no. 1, pp. 1–66, 2017, doi: 10.1007/s10462-016-9475-9.
[9] V. Gupta and G. S. Lehal, "A survey of text summarization extractive techniques," *Journal of Emerging Technologies in Web Intelligence*, vol. 2, no. 3, pp. 258–268, 2010, doi: 10.4304/jetwi.2.3.258-268.
[10] S. Chopra, M. Auli, and A. M. Rush, "Abstractive sentence summarization with attentive recurrent neural networks," in *2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies,* 2016, pp. 93–98, doi: 10.18653/v1/n16-1012.
[11] R. Nallapati, B. Zhou, C. D. Santos, Ç. Gulçehre, and B. Xiang, "Abstractive text summarization using sequence-to-sequence RNNs and beyond," in *CoNLL 2016 - 20th SIGNLL Conference on Computational Natural Language Learning, Proceedings*, Feb. 2016, pp. 280–290, doi: 10.18653/v1/k16-1028.
[12] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997, doi: 10.1162/neco.1997.9.8.1735.
[13] A. Vaswani *et al.*, "Attention is all you need," in *31st Conference on Neural Information Processing Systems*, 2017, pp. 1–11.
[14] L. Dong *et al.*, "Unified language model pre-training for natural language understanding and generation," *Advances in Neural Information Processing Systems*, vol. 32, 2019, doi: 10.48550/arXiv.1905.03197.
[15] K. Song, X. Tan, T. Qin, J. Lu, and T. Y. Liu, "MASS: masked sequence to sequence pre-training for language generation," in *36th International Conference on Machine Learning, ICML 2019*, California: PMLR 97, 2019, pp. 10384–10394.
[16] S. Rothe, S. Narayan, and A. Severyn, "Leveraging pre-trained checkpoints for sequence generation tasks," *Transactions of the Association for Computational Linguistics*, vol. 8, pp. 264–280, 2020, doi: 10.1162/tacl_a_00313.
[17] M. Lewis *et al.*, "BART: denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension," in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, Stroudsburg, USA: Association for Computational Linguistics, 2020, pp. 7871–7880, doi: 10.18653/v1/2020.acl-main.703.
[18] J. Zhang, Y. Zhao, M. Saleh, and P. J. Liu, "PEGASUS: pre-training with extracted gap-sentences for abstractive summarization," in *37th International Conference on Machine Learning, ICML 2020*, PMLR 119, 2020, pp. 11265–11276.
[19] J. Devlin, M.-W. Chang, K. Lee, K. T. Google, and A. I. Language, "BERT: pre-training of deep bidirectional transformers for language understanding," in *Proceedings of NAACL-HLT 2019*, 2019, pp. 4171–4186.
[20] L. Xue *et al.*, "MT5: a massively multilingual pre-trained text-to-text transformer," *arXiv-Computer Science*, pp. 1-17, Oct. 2020.
[21] Y. Liu *et al.*, "Multilingual denoising pre-training for neural machine translation," *Transactions of the Association for Computational Linguistics*, vol. 8, pp. 726–742, 2020, doi: 10.1162/tacl_a_00343.
[22] T. Scialom, P. A. Dray, S. Lamprier, B. Piwowarski, and J. Staiano, "MLSUM: the multilingual summarization corpus," in *EMNLP 2020 - 2020 Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference*, 2020, pp. 8051–8067, doi: 10.18653/v1/2020.emnlp-main.647.
[23] B. Baykara and T. Güngör, "Turkish abstractive text summarization using pretrained sequence-to-sequence models," *Natural Language Engineering*, vol. 29, no. 5, pp. 1275–1304, 2023, doi: 10.1017/S1351324922000195.
[24] G. Eryiğit, J. Nivre, and K. Oflazer, "Dependency parsing of turkish," *Computational Linguistics*, vol. 34, no. 4, pp. 357–389, 2008, doi: 10.1162/coli.2008.34.4.627.
[25] D. Z. Hakkani-Tür, K. Oflazer, and G. Tür, "Statistical morphological disambiguation for agglutinative languages," *Computers and the Humanities*, vol. 36, no. 4, pp. 381–410, 2002, doi: 10.1023/A:1020271707826.
[26] A. Radford, K. Narasimhan, T. Salimans, and I. Sutskever, "Improving language understanding with unsupervised learning,"*Open AI*, pp. 1-12, 2018.
[27] Z. Yang, Z. Dai, Y. Yang, J. Carbonell, R. Salakhutdinov, and Q. V. Le, "XLNet: generalized autoregressive pretraining for language understanding," *Advances in Neural Information Processing Systems*, vol. 32, 2019.
[28] C. Raffel *et al.*, "Exploring the limits of transfer learning with a unified text-to-text transformer," *Journal of Machine Learning Research*, vol. 21, 2020.
[29] A. M. Rush, S. Chopra, and J. Weston, "A neural attention model for abstractive sentence summarization," *arXiv*, Sep. 2015.

[30] A. Celikyilmaz, A. Bosselut, X. He, and Y. Choi, “Deep communicating agents for abstractive summarization,” in *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 2018, pp. 1662–1675, doi: 10.18653/v1/n18-1150.

[31] S. Sukhbaatar, A. Szlam, and R. Fergus, “Learning multiagent communication with backpropagation,” *Advances in Neural Information Processing Systems*, pp. 2252–2260, 2016.

[32] R. Paulus, C. Xiong, and R. Socher, “A deep reinforced model for abstractive summarization,” in *6th International Conference on Learning Representations, ICLR 2018 - Conference Track Proceedings*, 2018, pp. 1–13.

[33] S. Narayan, S. B. Cohen, and M. Lapata, “Don’t give me the details, just the summary! topic-aware convolutional neural networks for extreme summarization,” in *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, EMNLP 2018*, 2018, pp. 1797–1807, doi: 10.18653/v1/d18-1206.

[34] Y. Liu and M. Lapata, “Text summarization with pretrained encoders,” in *EMNLP-IJCNLP 2019 - 2019 Conference on Empirical Methods in Natural Language Processing and 9th International Joint Conference on Natural Language Processing, Proceedings of the Conference*, Aug. 2019, pp. 3730–3740, doi: 10.18653/v1/d19-1387.

[35] K. M. Hermann *et al.*, “Teaching machines to read and comprehend,” in *Advances in Neural Information Processing Systems*, Cambridge: MIT Press, 2015, pp. 1693–1701.

[36] E. Sandhaus, “The new york times annotated corpus,” *Abacus Data Network*, V1, 2008. [Online]. Available: https://hdl.handle.net/11272.1/AB2/GZC6PL

[37] W. Qi *et al.*, “ProphetNet: predicting future n-gram for sequence-to-sequence pre-training,” in *Findings of the Association for Computational Linguistics Findings of ACL: EMNLP 2020*, 2020, pp. 2401–2410, doi: 10.18653/v1/2020.findings-emnlp.217.

[38] Z. Altan, “A Turkish automatic text summarization system,” in *Proceedings of the IASTED International Conference: Applied Informatics*, 2004, pp. 311–316.

[39] M. Kutlu, C. Ciğir, and I. Cicekli, “Generic text summarization for turkish,” *Computer Journal*, vol. 53, no. 8, pp. 1315–1323, 2010, doi: 10.1093/comjnl/bxp124.

[40] Y. S. Kartal and M. Kutlu, “Machine learning based text summarization for turkish news,” in *2020 28th Signal Processing and Communications Applications Conference, SIU 2020 - Proceedings*, IEEE, 2020, pp. 1–4, doi: 10.1109/SIU49456.2020.9302096.

[41] M. Ozsoy, I. Cicekli, and F. Alpaslan, “Text summarization of turkish texts using latent semantic analysis,” in *Proceedings of the 23rd International Conference on Computational Linguistics (Coling 2010)*, 2010, pp. 869–876.

[42] F. C. Pembe, “Automated query-biased and structure-preserving document summarization for web search tasks,” *Ph.D Thesis,* Department of Computer Engineering, Boğaziçi University, Istanbul, Turkey, 2010.

[43] A. Güran, “Automatic tex summarization system,” *Ph.D Thesis,* Department of Computer, Yıldız Technical University, Istanbul, Turkey, 2013.

[44] E. Karakoc and B. Yilmaz, “Deep learning based abstractive turkish news summarization,” in *27th Signal Processing and Communications Applications Conference, SIU 2019*, IEEE, 2019, pp. 1–4, doi: 10.1109/SIU.2019.8806510.

[45] C.-Y. Lin, “ROUGE: a package for automatic evaluation of summaries,” in *Text Summarization Branches Out*, 2004, pp. 74–81.

[46] V. Ahuir, L. F. Hurtado, J. Á. González, and E. Segarra, “Nasca and nases: two monolingual pre-trained models for abstractive summarization in catalan and spanish,” *Applied Sciences*, vol. 11, no. 21, 2021, doi: 10.3390/app11219872.

[47] M. Pant and A. Chopra, “Multilingual financial documentation summarization by team_tredence for FNS2022,” in *Proceedings of the 4th Financial Narrative Processing Workshop*, 2022, pp. 112–115.

BIOGRAPHIES OF AUTHORS

Neda Alipour holds a doctor of management information systems degree from Atatürk University, Türkiye in 2022. He also received his B.Sc. (information technology engineering) from Tabriz University, Iran in 2011 and M.Sc. (MIS) from Atatürk University, Türkiye in 2017 and respectively. His research includes natural language rocessing, deep learning, e-commerce, and e-government. She can be contacted at email: nedaalipoor@yahoo.com or neda.alipour14@ogr.atauni.edu.tr.

Serdar Aydın is currently an associate professorship in Department of Software Engineering in Türkiye, Atatürk University. His research includes social sciences and humanities, science, technology, and society. He has published over 70 papers in international journals and conferences. He can be contacted at email: serdar@atauni.edu.tr.