Topic-Controlled Text Generation

Cansen Çağlayan
Department of Computer Engineering
Atılım University
Ankara, Turkey
cansen.caglayan@atilim.edu.tr

Murat Karakaya
Department of Computer Engineering
Atılım University
Ankara, Turkey
murat.karakaya@atilim.edu.tr

2021 6th International Conference on Computer Science and Engineering (UBMK) | 978-1-6654-2908-5/21/$31.00 ©2021 IEEE | DOI: 10.1109/UBMK52708.2021.9558910

***Abstract*— Today, the text generation subject in the field of Natural Language Processing (NLP) has gained a lot of importance. In particular, the quality of the text generated with the emergence of new transformer-based models has reached high levels. In this way, controllable text generation has become an important research area. There are various methods applied for controllable text generation, but since these methods are mostly applied on Recurrent Neural Network (RNN) based encoder decoder models, which were used frequently, studies using transformer-based models are few. Transformer-based models are very successful in long sequences thanks to their parallel working ability. This study aimed to generate Turkish reviews on the desired topics by using a transformer-based language model. We used the method of adding the topic information to the sequential input. We concatenated input token embedding and topic embedding (control) at each time step during the training. As a result, we were able to create Turkish reviews on the specified topics.**

Keywords—text generation, review generation, controllable text generation, topic-controlled text generation.

I. Introduction

Text generation is one of the challenging tasks that has been worked on in the Natural Language Processing (NLP) area. The goal of text generation is for machines can use human language so that can generate meaningful texts. Text generation takes text (i.e. a sequence of words or characters) as input, processes the input text into semantic representations, and generates desired output text [1]. There are many application areas where text generation is employed. For example, machine translation, summarization, question answering, dialogue systems, or generating review, story, poetry, etc. Especially after the neural network technology, much more successful results began to be obtained. Various architectures have been used for text generation. Most popular ones are Recurrent Neural Network (RNN) based encoder-decoder models [2],[3], Convolutional Neural Network (CNN) based encoderdecoder models [4], Generative Adversarial Networks (GAN) [5], Reinforcement Learning [6] and Transformer [7].

In recent years, transformer have become very important and have started to achieve very successful results in text-based operations. Transformer is a model architecture eschewing recurrence and instead relying entirely on an attention mechanism to draw global dependencies between input and output [7]. Transformer model was first used in the language translation task. Successful results were obtained in English - German and English - French language translation. Recurrent layers, which are usually used in encoder-decoder models, have been replaced by a more parallel model that is completely attention-based. In this way, it has been observed that transformer can be trained much faster than the recurrent and convolutional based models for the translation task. Later, transformer started to be used in other NLP tasks. One strong example is Transformer-based Generative Pre-Training (GPT) model [8], which is one of the most successful models known in the task of text generation.

GPT models (GPT2, GPT3) developed by the OpenAI company were able to generate text with very high success rates. They have used huge text data to train these language models. GPT models predict the next word using all the words in a given text file using transformer [9]. However, there is often no control over the text generated if you use the GPT models directly. The process of having the generated text in the desired topic, style, sentiment, or personality is called controllable text generation.

Controllable text generation is the task of generating natural sentences whose attributes can be controlled [10]. Our study aims to generate Turkish reviews on the specified topic and different methods can be used for this task. Controllable text generation can be achieved by making changes in various parts of the process. One of them is to modify the sequential input (x_t). x_t is used to denote the word embedding of the token at time step t. As in our study, concatenation operations were generally performed for this task. That is, the model is trained by concatenating an additional control vector s to the sequential input x_t at each time step. In this way, the given input is not given to the model alone, but together with the determined control which is the topic in our study, and it is aimed to generate text in the desired control in the inference part. In general, the process is as follows. At time step t, the generator takes the word embedding x_t as input of the word that was predicted at step $t-1$ and predicts the word to be generated y_t at the current time step. Note that $x_t = y_{t-1}$. The input x_t can be concatenated with s at each time step to control the generation process. Hence, $\widetilde{x_t} = [x_t; s]$ [10].

II. Related Works

We can consider the related studies in two parts. In the first part, some studies that modify the sequential input to generate controlled text are mentioned. In these studies, style, personality, etc., may have been used as a control mechanism other than the topic. Also, the language models and architectures used to generate text in studies may be different. However, the method of adding control mechanism to the generated text is the same as in our study. In the second part, some studies that generate topic-controlled texts using different models and methods are mentioned.

A. Modifying Sequential Input

Various tasks and models may have been used in the studies to generate controlled text. The main point here is that all of them have made changes in the sequential input

978-1-6654-2908-5/21/$31.00 ©2021 IEEE

(x_t) . Noraset [11] used this technique for definition modeling: the task of estimating the probability of a textual definition, given a word being defined and its embedding. They concatenate the word embedding vector s of the word to be defined at each time step. Zhou [12] created a dialogue system with the same way but they used an external source as a control mechanism. Prabhumoye [13] also concatenate the hidden representation of the external source for Wikipedia update generation process. Harrison [14] concatenate a side constraint s which represents style and also personality attribute into the generation process. Chandu [15] also concatenate the personality representation as a control mechanism at each time step of the story generation process [10].

B. Topic-Controlled Text Generation

The purpose of topic-controlled text generation is that the generated text is on the desired topic. There are studies on this task using different models, techniques, and datasets. Tang [16] created a conversational system to chat naturally with humans and proactively guide the conversation to a designated target topic with reinforcement learning. The goal is to guide the conversation naturally to the target topic. Prabhumoye [17] also investigated agents that can make chit-chat and goal-oriented (topic) dialogue in a game environment using reinforcement learning. Li [18] combined reinforcement learning, generative adversarial networks, and recurrent neural networks as a new model to generate category sentences. They concatenated the controllable information and the sentence distribution to be the prior information.

The number of studies on this subject using transformers is very low. However, it is increasing and very successful results are obtained. Keskar [19] proposed a model called CTRL, which is a conditional transformer language model, trained to condition control codes that govern style, topic, and taskspecific behavior. Chang [20] used a transformer-based language model to generate text in chosen topics but their data was unlabeled.

III. EXPERIMENT

This section explains the dataset, preprocessing the data, the language model we used, and the changes we made to the model to generate controlled text as a conditioned language model.

A. Dataset

Multi-Class Classification Dataset for Turkish [1] is a dataset that contains the reviews of the products and review categories in Turkish. It has 430k lines, 32 categories like electronic, education, food, sports, tourism. Each category roughly has 13k reviews from different Turkish websites. We chose this dataset because it is very useful to generate Turkish text in different categories. We handled 32 different categories as topics.

[1] https://www.kaggle.com/savasy/multiclass-classification-data-for-turkish-tc32?select=ticaret-yorum.csv

Input 1 (review sequence) :	petshop ürün iade sorunu 385 lira tutarında kedi lazer oyuncağı satın aldım.
Input 2 (topic) :	alışveriş (0)
Output :	ürün iade sorunu 385 lira tutarında kedi lazer oyuncağı satın aldım. ürünün

Fig.1. Sample data.

B. Preprocessing Data

The dataset consists of two columns, text (reviews) and category of review (topic). Topics are kept as text (shopping, electronic, education, etc.). First, an integer value is assigned to each topic. Shopping-0, electronic-1, education-2, food-3,…,transportation-31. Then, each review is vectorized using the text vectorization layer.

For the model to learn to predict the next word in text generation, the last word of each sequence was prepared as the target word. The words in the sequence are shifted in one position to create our output sequence. Thus, for a word at position i, the target is the word at $i + 1$. The model uses all words up to position i to predict the next word. In addition, there is the topic information for each sequence. In the final dataset, we have the following: words up to position i of the vectorized sequence as input 1, the topic of each sequence as input 2, and the target word at position $i + 1$. To give these two inputs (review sequence and topic) to the model, we combined them as a single input and finalized the new dataset formed with the target word using the data pipeline functions. Fig. 1. shows the final state of the data after the preprocessing operations, for this example the maximum sequence size is 12 and the integer value of the topic is 0 which is shopping in the text format.

C. Language Model

We used an autoregressive language model that called a miniature version of the GPT. The model consists of a single Transformer block with causal masking in its attention layer [21]. Causal masking prevents flow of information from future tokens to current token. Transformer is an attention-based encoder-decoder type architecture. The encoder maps an input sequence into an abstract continuous representation that holds all the learned information of that input. The decoder takes continuous representation and step by step generates a output while also being fed the previous output [22]. GPT models consist of only the decoder structure of the transformer to generate text.

The main point of the transformer is to compute the self-attention between the different words in a sequence and then re-represent the words in a sensible manner [23]. Self-attention is an attention mechanism relating different positions of a single sequence in order to compute a representation of the sequence [7]. In this way, high successes are achieved in text generation training to preserve the meaning especially if they contain long sequences. The generation of each word conditions on all previous words in the sequence, not only on the last generated word [24].

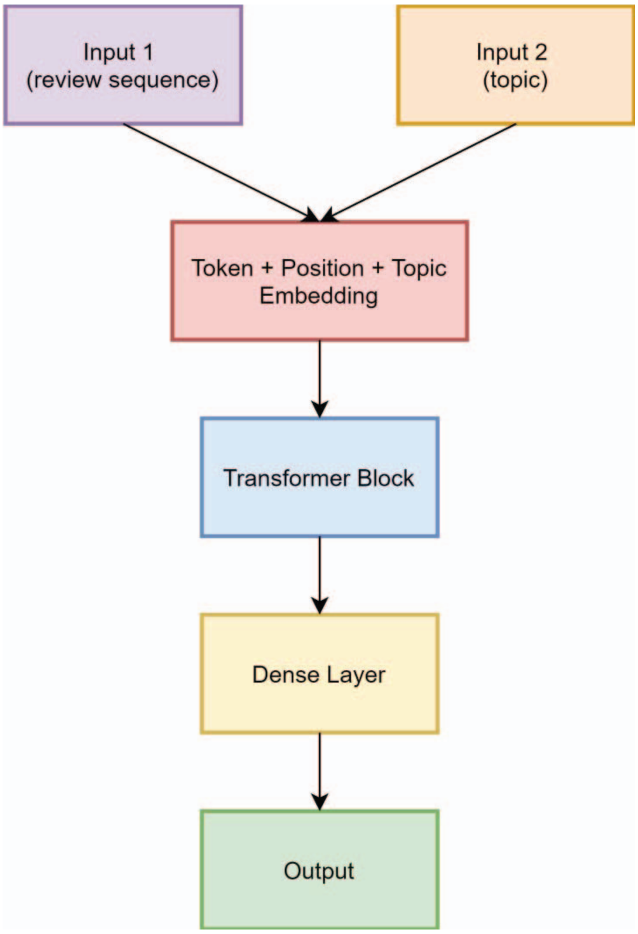

Fig. 2. Model summary.

D. *Conditioned Language Model*

Transformer uses token embedding to represent the words of the sequence and position embedding to inject positional information of words. The transformer has no recurrence like recurrent neural networks, some information about the positions of words must be added into the token embedding to keep the meaning of words according to their order in the sequence [22]. In addition to these two embeddings, we have created a topic embedding in our model. Topic embedding is a representation of the topic information for each review. Then we put it all together. We concatenated the token and position representation at each time step and the embedding of the topic information to which the input sequence belongs. Since we have already prepared our dataset as two inputs (review sequence and topic number), we have obtained the embedding representation of each element of the training data, including token, position, and topic information. Then, we passed this topic information added display through the GPT kind transformer decoder block. We trained our model with the outputs given by passing through a dense layer. We used *sparse categorical cross-entropy* as the loss function.

The embedding representation of the topic vector, which is the control mechanism s, has been added to the embedding representation of the review sequence, which is the sequential input x_t at each time step. In this way, the input to be used for the text generation training was given to model together with a control mechanism as shown in the Fig. 2.

Generated Review	Topic Number	Topic
ürün değişimi yapılmıyor , 22 mayıs 2020 tarihinde aldığım ve numaralı siparişimdeki bir türlü bahanelerle iptal etmek için müşteri hizmetlerine ulaşmak üzere bir hafta içi boş yere götürdüm . Ürün için bir an önce beni arayıp soran yok	0	alışveriş
hastanesi 'nde yaşanan mağduriyet , ben de bulunan bir haftadır beni çok yüksek ateş ölçer ile ilgili hiçbir şekilde bana ait olmayan bir tane ilaç 1 aydır her gün	26	sağlık
otel 'den para İadesi yapmıyor , yaklaşık 1 hafta sonu önce yapmış olduğum rezervasyon için numarası ile birlikte verilen sözlerin tutulmasını talep edilen sürede. Üstelik sadece iki kişilik her şey dahil olmak üzere bu kadar kötü bir hafta sonu için bu konuda	30	turizm

Fig. 3. Some best generated reviews and topics.

Generated Review	Topic Number	Topic
[UNK] .com .tr 'den Şikayetçiyim , merhaba , [UNK] numaralı kargom ile beraber iki adet ürün satın aldım . [UNK]) . bir daha doğrusu [UNK] ve bu süreçte [UNK] .	0	alışveriş
2 adet küçük yatırımcıyı mağdur etti ! , [UNK] [UNK] ve en son zamanlarda 2 adet küçük bir sürü para verip aldığımız yere gittim ve daha önce de bu konuda yardımcı olamayacaklarını söyleyip başka bir de	26	sağlık
[UNK] numaralı siparişimin gönderilmemesi , bu firmadan almış olduğum halde 1 yıl boyunca bir şey sormak için bir ay içi [UNK] dair bir türlü bahanelerle oyalanıyoruz . [UNK] . Üstelik sadece iki kişilik bir şey	30	turizm

Fig. 4. Some worst generated reviews and topics.

We used the *top-k sampling* approach. With the *top-k sampling* method at each time step, the model generates the probability of each word in the vocabulary being the likely next word. We choose k is 10. Therefore randomly sample from the 10 most likely candidates from the distribution. Then, subsequent time steps generated words based on the previously selected words [25].

We generated approximately 100 reviews for each topic using our conditioned language model, which we modified by adding topic embedding to the sequential input representation.

IV. RESULT

At the end of the study, the reviews generated for each topic were examined with human evaluation. According to the visible results, the model is successful in generating Turkish reviews on the specified topic. In some examples generated sentences are not completely meaningful because the language model can be improved with more transformer layers and epochs. However, according to analyzes made by the human, the frequently generated text is about the topic that we wanted to generate. In addition, the model can complete the next word by learning the rules of the Turkish language. Some of the best and worst reviews generated on different topics are shown in Fig. 3. and Fig. 4. We observe that in the best generated reviews, they contain words that belong to the desired subject. In addition, the features of the Turkish language, complementing each other and forming a meaningful whole, have been learned to some extent by the model. But in the worst reviews, we sometimes see that is irrelevant to the expected topic, or that it produces too much unknown (UNK).

V. CONCLUSION & FUTURE WORK

In the generation of controlled text, which is created by adding the control to the sequential input, the number of studies using transformers is quite low. Due to the fact that transformer models are very efficient, they will be seen more in this field in the future. In this study, we aimed to generate Turkish reviews on different topics by using modifying the sequential input method. However, the language model used for generation can be improved with more layers and more epochs. Although we have achieved successful results according to human evaluations, we plan to make quantitative evaluations in the future. For this, we will train a transformer-based topic classification model. Thus, using the classification model, we will be able to measure whether the reviews generated by the conditional language model are on the right topic. In addition, we plan to compare different methods to which the control mechanism can be applied and to obtain quantitative results between methods. Due to the lack of time, we explained this study through the generation process and qualitative results.

Controlled text generation is a constantly evolving research area that can be done using different methods in a wide variety of tasks. We plan to continue our work by developing the language model and implement different methods to generate topic-controlled text.

REFERENCES

[1] W. Yu, C. Zhu, Z. Li, Z. Hu, Q. Wang, H. Ji, and M. Jiang, "A Survey of Knowledge-Enhanced Text Generation," vol. 1, no. 1, pp. 1–44, 2020.

[2] K. Cho, B. Van Merrienboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using RNN encoder-decoder for statistical machine translation," EMNLP 2014 - 2014 Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference, pp. 1724– 1734, 2014.

[3] I. Sutskever, O. Vinyals, and Q. V. Le, "Sequence to sequence learning with neural networks," Advances in Neural Information Processing Systems, vol. 4, no. January, pp. 3104–3112, 2014.

[4] J. Gehring, M. Auli, D. Grangier, D. Yarats, and Y. N. Dauphin, "Convolutional sequence to sequence learning," 34th International Conference on Machine Learning, ICML 2017, vol. 3, pp. 2029–2042, 2017.

[5] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," Communications of the ACM, vol. 63, no. 11, pp. 139–144, 2014.

[6] V. Srinivasan, S. Santhanam, and S. Shaikh, "Natural Language Generation Using Reinforcement Learning with External Rewards," 2019.

[7] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," vol. 2017-December, pp. 5999– 6009, Jun. 2017.

[8] A. Radford, K. Narasimhan, T. Salimans, and I. Sutskever, "(OpenAI Transformer): Improving Language Understanding by Generative Pre-Training," OpenAI, pp. 1–10, 2018.

[9] A. Radford, J. Wu, R. Child, D. Luan, D. Amodei, and I. Sutskever, "Language Models are Unsupervised Multitask Learners," 2019.

[10] S. Prabhumoye, A. W. Black, and R. Salakhutdinov, "Exploring Controllable Text Generation Techniques," pp. 1–14, 2020.

[11] T. Noraset, C. Liang, L. Birnbaum, and D. Downey, "Definition modeling: Learning to define word embeddings in natural language," 31st AAAI Conference on Artificial Intelligence, AAAI 2017, pp. 3259–3266, 2017.

[12] K. Zhou, S. Prabhumoye, and A. W. Black, "A dataset for document grounded conversations," Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, EMNLP 2018, pp. 708–713, 2020.

[13] S. Prabhumoye, C. Quirk, and M. Galley, "Towards content transfer through grounded text generation," NAACL HLT 2019 - 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference, vol. 1, pp. 2622–2632, 2019.

[14] V. Harrison, L. Reed, S. Oraby, and M. Walker, "Maximizing stylistic control and semantic accuracy in NLG: Personality variation and discourse contrast," DSNNLG 2019 - 1st Workshop on Discourse Structure in Neural NLG, Proceedings of the Workshop, pp. 1–12, 2019.

[15] K. Chandu, S. Prabhumoye, R. Salakhutdinov, and A. W. Black, ""My Way of Telling a Story": Persona based Grounded Story Generation," pp. 11–21, 2019.

[16] J. Tang, T. Zhao, C. Xiong, X. Liang, E. P. Xing, and Z. Hu, "Target-guided open-domain conversation," ACL 2019 - 57th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference, pp. 5624–5634, 2020.

[17] S. Prabhumoye, M. Li, J. Urbanek, E. Dinan, D. Kiela, J. Weston, and A. Szlam, "I love your chain mail! Making knights smile in a fantasy game world: Opendomain goal-oriented dialogue agents," 2020.

[18] Y. Li, Q. Pan, S. Wang, T. Yang, and E. Cambria, "A Generative Model for category text generation," Information Sciences, vol. 450, pp. 301–315, 2018.

[19] N. S. Keskar, B. McCann, L. R. Varshney, C. Xiong, and R. Socher, "CTRL: A Conditional Transformer Language Model for Controllable Generation," pp. 1–18, 2019.

[20] H.-S. Chang, J. Yuan, M. Iyyer, and A. McCallum, "Changing the Mind of Transformers for TopicallyControllable Language Generation," 2021.

[21] A. Nandan. (2020). "Text generation with a miniature GPT," [Online].Available:https://keras.io/examples/generative/text_generation_with_miniature_gpt/ (visited on 06/21/2021).

[22] M. Phi. (2020). "Illustrated Guide to Transformers - Step by Step Explanation," [Online]. Available: https: / / towardsdatascience . com / illustrated - guide - to - transformers - step - by - step - explanation - f74876522bc0 (visited on 06/22/2021).

[23] J. Gong, H. Ma, Z. Teng, Q. Teng, H. Zhang, L. Du, S. Chen, M. Z. A. Bhuiyan, J. Li, and M. Liu, "Hierarchical Graph Transformer-Based Deep Learning Model for Large-Scale Multi-Label Text Classification," IEEE Access, vol. 8, pp. 30 885–30 896, 2020.

[24] T. Mihaylova and A. F. Martins, "Scheduled sampling for transformers," in ACL 2019 - 57th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Student Research Workshop, Association for Computational Linguistics (ACL), Jun. 2019, pp. 351–356.

[25] A. Fan, M. Lewis, and Y. Dauphin, "Hierarchical neural story generation," vol. 1, pp. 889–898, 2018.