FINE TUNING T5 MODEL FOR TEST EXAMINATION AND ASSESSMENT IN MUSIC EDUCATION FOR VISUALLY IMPAIRED STUDENTS

Simeon Monov, Nikolay Pavlov, Andrey Nikolov

Abstract. *Efficient teaching and examination of students with special educational needs require additional interaction accessibility, beyond what is provided by standard user interface instruments. Nowadays text-to-speech tools are abundant, but speech-to-text interaction and free-text answer understanding, and validation are still a new area to be researched and developed. In this paper we present the fine-tuning and application of T5 model for test examination and assessment in Music education for visually impaired students with focus on note durations.*

Key words: T5 model, AI, Music education, Visually impaired students.

Introduction

Pre-trained language models have been used for different tasks before [1]. T5, or Text-To-Text Transformer was developed by Google and was presented in *Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer* paper [2]. The abstract of the paper begins with "Transfer learning, where a model is first pre-trained on a data-rich task before being fine-tuned on a downstream task, has emerged as a powerful technique in natural language processing (NLP)". In the current work we utilize a pre-trained T5 model on Bulgarian language [3] and we further fine-tune it to be able to validate and assess answers on questions in music education and especially note durations for visually impaired students.

Chukanska and Koleva [4] has created an adaptive instrument for teaching note durations to visually impaired students. To improve the accessibility of their instrument, Chukanska and Koleva have stepped upon Azure Cognitive Services and related automation technologies [5, 6] for speech recognition and answer validations. Authors reported later that common complaints were failure to recognize voiced answers. In this work

we aim to improve the recognition of verbal answers and answer validation by employing a pre-trained T5 model, fine-tuned for validating and assessing answers about musical note durations. Note durations are based on strict mathematical rules, which facilitates the training of the model.

The T5 model

T5 is an encoder-decoder transformer model which converts all NLP problems into a text-to-text format. It is trained using teacher forcing. T5 transformers produce impressive results even by just a few epochs of training.

T5 models are capable of performing different NLP tasks, which include:

- **Language Translation**: Translating text from one language to another.
- **Text Summarization**: Condensing a long piece of text into a concise summary.
- **Question Answering**: Providing answers to questions based on the context given.
- **Text Classification**: Categorizing text into predefined classes.

Figure 1 shows a diagram of different NLP tasks that can be performed by the T5 model.

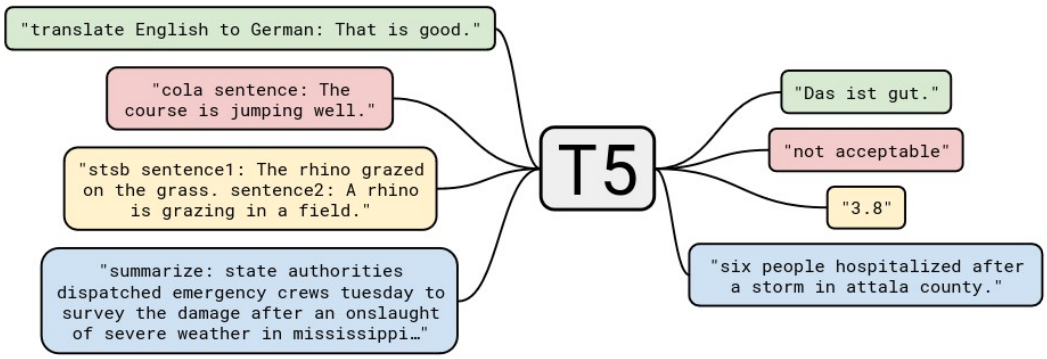

Figure 1. A diagram showing different tasks T5 model can perform [2]

One of the best abilities of T5 is that it is possible to virtually adapt to any text task it is given.

For the purpose of the current work, we use T5 for Text Classification tasks.

T5 for Bulgarian NLP tasks

There are multiple variants of T5 models, but it lacks good models pre-trained on Bulgarian data. Multilingual models [7, 8, 16] which include Bulgarian language exist, but recent works demonstrate that monolingual models perform better on tasks on the language that they are trained on, than models pre-trained on multilingual corpora [3, 9, 10, 11]. For the purpose of the current work, we have chosen to use BGT5 model [3], which is a T5 model pre-trained on massive data of Bulgarian language. This model shows better performance on different Bulgarian language tasks than other multilingual models.

We choose BGT5 because it is capable of understanding Bulgarian language and matching different answers with same meaning.

Dataset

Every large language model (LLM) needs to be "taught" or fine-tuned to perform a downstream task such as text classification, which is achieved by Supervised Learning most of the time.

For the purpose of fine-tuning on a downstream task, the data is the most important thing. Sometimes it is very difficult to find a good dataset for a particular task especially for tasks that require language other than English. In our work we use data synthesis and generate synthetic dataset with questions in music education. In recent years with the rapid development of different LLMs synthetic dataset generation is used a lot to further train LLMs and has proven to work well on wide range of tasks [12, 13, 14, 15].

Music theory questions and their answers follow strict mathematical rules and can be easily generated using a generator. We have developed such a generator for generating different variants of questions about music notes durations with their respective answers. The dataset was balanced between positive and negative answers. Further, additional noise was added into the dataset to avoid over-fitting of the model.

The dataset contained 35785 data points. Figure 2 shows some examples of dataset data points with positive and negative answers.

```
{
  "question": "Колко са седем четвъртини?",
  "answer": "седем удара",
  "truth": "true"
},
{
  "question": "Колко удара са три четвъртини?",
  "answer": "два",
  "truth": "false"
},
```


Figure 2. Dataset data points examples

Model training and evaluation

In our work we are using a BGT5-base variant of the model. The model was trained for 3 epochs with learning rate of 0,0003 and batch size of 16. The dataset was split into 80% training and 20% validation data. We used 1 NVIDIA V100 GPU with 32GB. Training time took 36:27 minutes for training and 08:45 minutes for validation per epoch. The training and validation losses observed are listed in Table 1.

Table 1. Observed training and validation losses during training

Epoch	Training loss	Validation loss
1	0.262363	0.145156
2	0.220144	0.139303
3	0.198087	0.135779

The model was evaluated using a validation dataset and the results show an exact match of 77.89 and a scope of 77.89 F1.

Manual evaluation was performed too, showing the good performance of the model. Table 2 shows some results of the manual evaluation.

Table 2. Observed results from manual model evaluation

Question	Answer	Result
В една половина нота, колко осмини има?	четири ноти	'Reference truth': 'true', 'Predicted truth': 'true'
Колко удара са две четвъртини?	два удара	'Reference truth': 'true', 'Predicted truth': 'true'
Колко осмини има в цяла нота?	осем	'Reference truth': 'true', 'Predicted truth': 'false'

We observed that fine-tuning a T5 model using synthetic data on the task of question answers classification on music note duration questions in Bulgarian language shows very good results and can be effectively used in systems for music education for visually impaired students.

Acknowledgments

This work is supported by the project MUPD23-FMI-009 of the Scientific Fund of the Paisii Hilendarski University of Plovdiv, Bulgaria.

References

[1] V. Naneva, K. Stefanova, The Application of Machine Learning in Business Intelligence, *Proc. of International conference on Applied Internet and Information Technologies*, 16 October, 2020, Zrenjanin, Serbia, p. 169–174, ISSN: 978-86-7672-342-3, https://aiitconference.org/archive/Proceedings_AIIT2020.pdf.

[2] C. Raffel, N. Shazeer, A. Roberts, K. Lee, S. Narang, M. Matena, Y. Zhou, W. Li, P. Liu, Exploring the limits of transfer learning with a unified text-to-text transformer, *The Journal of Machine Learning Research*, 2020, Vol. 21, Issue 1, pp. 5485–551.

[3] S. Monov, N. Pavlov, D. Trifonova, Pretraining and validating T5 model on Bulgarian data [Conference presentation], *International conference IMEA'2023*, 29 Nov – 01 Dec, 2023, Pamporovo, Bulgaria, https://imea2023.fmi-plovdiv.org/wp-content/uploads/2023/11/4_11_Abstract_Monov_Pavlov_Trifonova_51-1.pdf.

[4] Y. Chukanska, G. Koleva, Adaptive tools for teaching note durations to visually impaired students [Conference presentation], *International conference IMEA'2022*, 23 – 25 Nov, 2022, Pamporovo, Bulgaria, https://imea2022.fmi-plovdiv.org/wp-content/uploads/2022/11/IMEA-2022-Scientific-Program_final.pdf.

[5] Microsoft, Azure Cognitive Services, 2023, https://azure.microsoft.com/en-us/products/cognitive-services/.

[6] M. Mateev, Implementing Document Automation using Modern Data Lake and Automated Workloads for Big Environmental Projects, *Proc. of 3rd Annual International Conference on AI and Data Science*, Athens, 2022.

[7] L. Xue, N. Constant, A. Roberts, M. Kale, R. Al-Rfou, A. Siddhant, A. Barua, C. Raffel, mT5: A massively multilingual pre-trained text-to-

text transformer, arXiv preprint arXiv:2010.11934, 2020, DOI: https://doi.org/10.48550/arXiv.2010.11934.

[8] H. Chung, L. Hou, S. Longpre, B. Zoph B, Y. Tay, W. Fedus, E. Li, X. Wang, M. Dehghani, S. Brahma, A. Webson, Scaling instruction-finetuned language models, 2022, arXiv preprint arXiv:2210.11416, DOI: https://doi.org/10.48550/arxiv.2210.11416.

[9] F. Baly, H. Hajj, et al., Arabert: Transformer-based model for arabic language understanding, *Proc. of the 4th Workshop on Open-Source Arabic Corpora and Processing Tools, with a Shared Task on Offensive Language Detection*, 2020, pp. 9–15.

[10] J. Cañete, G. Chaperon, R. Fuentes, J. Pérez, Spanish pre-trained bert model and evaluation data, In to appear in PML4DC at ICLR 2020, 2020.

[11] D. Nguyen, A. Nguyen, Phobert: Pre-trained language models for Vietnamese, 2020, arXiv preprint arXiv:2003.00744.

[12] R. Tang, X. Han, X. Jiang, X. Hu, Does synthetic data generation of llms help clinical text mining?, arXiv preprint arXiv:2303.04360, 2023.

[13] H. Gupta, K. Scaria, U. Anantheswaran, S. Verma, M. Parmar, S. Sawant, S. Mishra, C. Baral, Targen: Targeted data generation with large language models, 2023, arXiv preprint arXiv:2310.17876.

[14] M. Josifoski, M. Sakota, M. Peyrard, R. West, Exploiting asymmetry for synthetic training data generation: Synthie and the case of information extraction, 2023, arXiv preprint arXiv:2303.04132.

[15] A. Askari, M. Aliannejad, E. Kanoulas, S. Verberne, A test collection of synthetic documents for training rankers: Chatgpt vs. human experts, *Proc. of the 32nd ACM International Conference on Information and Knowledge Management*, 2023, pp. 5311–5315.

[16] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, V. Stoyanov, Roberta: A robustly optimized bert pretraining approach, 2019, arXiv preprint arXiv:1907.11692.

Simeon Monov[1], Nikolay Pavlov[2], Andrey Nikolov[3]
[1,2,3] Paisii Hilendarski University of Plovdiv,
Faculty of Mathematics and Informatics,
236 Bulgaria Blvd., 4003 Plovdiv, Bulgaria
Corresponding author: smonov@uni-plovdiv.bg