CS410 Technology Review
TensorFlow NLP and Its Applications in University Researches

Introduction

According to TensorFlow's official website [1], TensorFlow is an open source end-to-end machine learning platform that provides an comprehensive ecosystem of tools for developers and researchers. TensorFlow is actively maintained by Google and has been widely used in both the industry and research contexts since its release in 2015. NLP (Natural Language Processing) is one of the hot topics in machine learning and TensorFlow provides a powerful support for NLP projects. TensorFlow even has a dedicated tutorial about movie review text classification for beginners to get familiar with the tool [2]. This framework is very popular in the industry, while in the research area it's not as popular as its rival PyTorch, there are still many researchers who chose this framework for its unique qualities. This review will focus on TensorFlow's key features and workflows for NLP projects and discuss typical applications of this framework in university research.

Working Principles

TensorFlow ships with a variety of useful tools for NLP, and in this section I will focus on the working principles and workflow for starting a text classification project. The workflow is based on an official guide on the TensorFlow website [3]. Notice that TensorFlow supports more than one language (C++, Go, Java, JavaScript and Swift) and in this case I will choose python as the demonstration language.

Starting with gathering data, because data can come from different sources, there isn't a dedicated API for data gathering, but TensorFlow has very good compatibility with popular libraries for retrieving data, such as the python DB-API standards. Then, we want to capture the features of the dataset and turn them into objects that TensorFlow can understand. This is the part where vanilla python and frameworks such as sklearn are best at. TensorFlow also has great support for these frameworks and data structures. Users can choose to use functions such as TfidfVectorizer() in sklearn freely. The next part, building models, is where TensorFlow comes into play. In order to deliver good text classification results, the design of the model is very important. Many designers can have complicated layered structures, and TensorFlow's models defined in its keras library can implement the design easily and efficiently. A lot of predefined functions also come in handy such as Dropout() for dropping out. The addition of such functions and layers can be integrated into the model with as little effort as calling .add() with the model object. The syntax for building a model is very close to human-readable English language and thus the efficiency for development and code clarity are very high. Next up we want to actually train the model. TensorFlow defines three key parameters: Metric, Loss function and Optimizer; each is loaded with abundant predefined setups which can satisfy the needs for most

applications. The actual training process is triggered with the fit method, which is also highly integrated with handy arguments. After we train our models, we need to tune the hyperparameters to achieve best model performance. This step is made easy because all the parameters mentioned earlier are highly integrated, therefore, tweaking them requires minimum effort as we only need to change the values in the function directly. The hyperparameters can be anything from Learning Rate to Kernel size, and most of the standard parameters have been predefined by this all-encompassing framework. Finally, we want to serve the model for actual productivity. The fact that this framework is maintained by Google ensures that it's fully compatible with Google Cloud, and the deployment from TensorFlow to Google Cloud is extremely simple. Google also has detailed guides for the deployment [4], making it even more user-friendly.

Interesting University Research Applications

Although in the research area, PyTorch is more popular than TensorFlow, according to the statistics from the gradient [5], there are still usages of this framework in research where scalability and efficiency are also the targets. For example, in this student research project [6] at Somaiya College of Engineering, India, having considered the scalability and efficiency requirements of the product when entering the formal production stage, they chose TensorFlow to achieve the goal for robustness and simplicity. The project used standard machine learning models to design the NLP unit and TensorFlow built the design in a way that is "Quite efficient" according to the result section of the paper.

Another interesting research [7] at Zhejiang University, China about deep transfer learning in NLP which has a collaboration with the tech giant Alibaba also used TensorFlow to train their models. This research has actual industrial relations in Alibaba's online products which lead to high expectations of scalability and efficiency of the model. As a result, they made use of multiple efficient frameworks and chose TensorFlow for the standard layers, which proves to meet their expectations.

Conclusion

TensorFlow is a popular and powerful framework for machine learning. It supports NLP tasks really well with simple syntaxes and efficient model implementations. It also supports major machine learning toolkits well and has a streamlined deployment solution. It has a relatively steep learning curve compared to its rivals such as PyTorch but it offers great scalability and efficiency for the project and many university researchers choose TensorFlow for these benefits.

REFERENCES

[1] *Why tensorflow*. TensorFlow. (n.d.). Retrieved November 8, 2021, from https://www.TensorFlow.org/about.

[2] *Text classification with Movie Reviews : tensorflow hub*. TensorFlow. (n.d.). Retrieved November 8, 2021, from https://www.tensorflow.org/hub/tutorials/tf2_text_classification.

[3] Google. (n.d.). *Step 1: Gather data | text classification guide | google developers*. Google. Retrieved November 8, 2021, from https://developers.google.com/machine-learning/guides/text-classification/step-1.

[4] Google. (n.d.). *AI platform documentation | google cloud*. Google. Retrieved November 8, 2021, from https://cloud.google.com/ai-platform/docs?utm_source=DevSite&utm_campaign=Text-Class-Guide&utm_medium=referral&utm_content=cloud&utm_term=guide.

[5] He, H. (2020, October 30). *The state of Machine Learning Frameworks in 2019*. The Gradient. Retrieved November 8, 2021, from https://thegradient.pub/state-of-ml-frameworks-2019-pytorch-dominates-research-tensorflow-dominates-industry/.

[6] R. Singh, M. Paste, N. Shinde, H. Patel and N. Mishra, "Chatbot using TensorFlow for small Businesses," 2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT), 2018, pp. 1614-1619, doi: 10.1109/ICICCT.2018.8472998.

[7] Minghui Qiu, Peng Li, Chengyu Wang, Haojie Pan, Ang Wang, Cen Chen, Xianyan Jia, Yaliang Li, Jun Huang, Deng Cai, and Wei Lin. 2021. EasyTransfer: A Simple and Scalable Deep Transfer Learning Platform for NLP Applications. Proceedings of the 30th ACM International Conference on Information & Knowledge Management. Association for Computing Machinery, New York, NY, USA, 4075–4084. DOI:https://doi.org/10.1145/3459637.3481911