Primljen: 20.05.2021.
Prihvaćen: 12.11.2021.

Prethodno priopćenje
UDK: 004.85:004.93

Online service for accessible machine learning of prediction models

Online servis za pristupačno strojno učenje prediktivnih modela

[1]Ivan Zatezalo, [2]Ivan Dunđer
[1,2] Sveučilište u Zagrebu, Filozofski fakultet, Odsjek za informacijske i komunikacijske znanosti, Ivana Lučića 3, Zagreb, Croatia
e-mail:[1]ivanzatezalo@gmail.com, [2]ivandunder@gmail.com

Abstract: *The field of machine learning is getting more and more advanced every day. Many use this technology in order to start their projects or better their current applications. Due to such a great interest, the use of mathematical algorithms required to build artificial intelligence has been simplified through various frameworks that allow seamless implementation. One framework by the name Teachable Machine is available as an online service. Its ease of use and simplicity allows anyone to develop machine learning models that can then be applied for personal use or projects. Despite this, such a framework still requires specialized knowledge in order to make an effective and accurate prediction model. This paper aims to test the limits of this service by developing models of different complexity. One model will be tested using the upload feature present on the service, whereas the other one will be implemented within a mobile application and tested on real-world data. The results gathered from this testing will then give an insight into the performance and quality of the aforementioned online service.*

Keywords: *machine learning, image classification, Teachable Machine, model accuracy, TensorFlow, information and communication sciences*

Sažetak: *Polje strojnog učenja postaje sve naprednije iz dana u dan. Mnogi koriste ovu tehnologiju kako bi započeli nove projekte ili poboljšali postojeće aplikacije. Zbog ovako velikog interesa, upotreba matematičkih algoritama potrebnih za izgradnju umjetne inteligencije pojednostavljena je kroz korištenje raznih razvojnih okvira koji omogućuju* *jednostavnu implementaciju. Jedan razvojni okvir pod nazivom Teachable Machine dostupan je kao online servis. Zbog lakoće korištenja i jednostavnosti ovaj servis omogućuje svakome razvijanje modela za strojno učenje koji će moći biti upotrijebljeni za osobne svrhe ili projekte. Unatoč ovome, takav razvojan okvir i dalje zahtijeva specijalizirano znanje kako bi se izradio efektivan i precizan model za predviđanje. Cilj ovog rada je ispitati ograničenja ovog servisa razvijanjem modela različitih razina složenosti. Jedan model bit će testiran putem servisa uz pomoć funkcije prijenosa, dok će drugi model biti implementiran unutar mobilne aplikacije te testiran nad podacima iz stvarnog svijeta. Rezultati prikupljenim ovim testiranjem dat će uvid u sposobnost i kvalitetu ranije navedenog online servisa.*

Ključne riječi: *strojno učenje, klasifikacija slika, Teachable Machine, TensorFlow, točnost modela, informacijske i komunikacijske znanosti*

1. Introduction

The field of machine learning is developing rapidly and becoming increasingly more desired by many corporations and companies that use this technology to advance their applications. However, its use is not only present in professional or scientific use. Many are developing their own artificial intelligence in order to suit their own needs and wants. To create such intelligence, one would have to collect their own data and develop a model using complex machine learning algorithms that require vast knowledge and expertise of the field.

The term model in machine learning refers to the output of a machine learning algorithm ran on data. Essentially, a model represents what was learned by a machine learning algorithm. This knowledge that was gathered on a dataset throughout training and testing can then be used in future to make assessments or predictions on unseen data (Brownlee, 2020a). The process of creating artificial intelligence has been simplified through several machine learning frameworks such as TensorFlow[1] and PyTorch[2], which ease the use of mathematical algorithms. Still, there is a necessity in knowing various intricacies that happen under the hood of a machine learning system. These intricacies include types of neural networks used to train a model, activation functions, optimizers and so on. All of these are necessary in order to create a successful and precise machine learning model.

Because of the huge time investment necessary for building of proper and precise models, a web service by the name Teachable Machine was developed by Google. With its first version available in 2017, it truly allowed anyone to develop their own artificial intelligence

[1] TensorFlow. https://www.tensorflow.org/
[2] Pytorch. https://pytorch.org/

without expert knowledge of machine learning. This means that the training process is done automatically without the need to write complex code (Google, n.d.), which makes the technology of machine learning accessible to a vast range of potential users.

In order to properly assess the functionality and quality of this service the following research questions were raised:

1. How accurate is the model developed by the online service when using a small and simple dataset?
2. While building a more complex model, what are some of the problems indicated during the training process and how accurate is the model when testing it on real-world data?
3. Depending on the results, when should the online service be used and in what capacity?

2. Related work

Machine learning today is intensively applied in a wide range of domains and for various tasks. It can be used for automatic recognition of information behavior (Lugović and Dunđer, 2017), for detecting emotions in speech (Lugović et al., 2016) or text (Dunđer and Pavlovski, 2019a), for natural language processing tasks (Dunđer and Pavlovski, 2019b), document classification (Dunđer et al., 2015) or machine translation (Dunđer, 2020; Dunđer et al., 2020).

When it comes to image and object identification and classification, research on k-NN classifier for class discrimination has been done by Lodh and Parekh (2016). One work proposed a system that would locate and segment an object from the surrounding area in a photograph (Hart et al., 2004). Another research used a Convolutional Neural Network (CNN) approach to reach an average precision of roughly 90% in identification of national flags (Said and Barr, 2021). Classification of images of ancient coins while dealing with problems such as erosion, uneven illumination, and moderate background clutter has been explored by Anwar et al. (2019). Space object classification with CNN and Recurrent Neural Networks (RNN) showed promising results with accuracy of more than 97% (Jia et al., 2018; Linares, 2016). Deep Neural Networks (DNN) have been employed for animal recognition and identification and demonstrated to be robust and stable (Nguyen et al., 2017). DNN has also been used for classifying and detecting plant diseases from leaf images with an overall accuracy of ca. 96% (Sladojevic et al., 2016).

3. Research

The following subsections deal with dataset entry in the online service Teachable Machine, the development of machine learning models, and the training and validation process within the service.

3.1. Dataset entry

Development of a successful machine learning model entails collecting data. The data necessary for machine learning can be found in numerous forms such as images, sounds, texts etc. The online service Teachable Machine allows training of the model with data such as images, audio or poses (Google, n.d.). The collected data should be separated into different classes so that the service is able to correctly make an assessment about the information present inside the data. To demonstrate the training done by the service, 20 images were selected. Ten of those images contained images of dogs, and the other images of cats. For this reason, two classes were prepared using the service and images were fed into the website using the upload feature.

3.2. Developing of the machine learning model

Before the training of the model, users can opt to tinker with the settings found in the advanced menu. This menu allows users to change some of the machine learning variables such as epochs, a batch size and learning rate (Google, n.d.). Teachable Machine's development team defined epoch as: "a time period during which every sample in the training dataset has been fed through the training model at least once" (Google, n.d.). The development team also noted that this number should be tweaked up until a model gives good predictive results (Google, n.d.). The term batch size refers to a particular number of samples used in one iteration of training. For example, if one class contained 40 images and batch size was set to 8, the data will be split into 40/8=5 batches. Once all 5 batches have been fed through the model, one epoch will be complete (Google, n.d.). The last variable available inside the advanced tab is the learning rate. In short, this important parameter controls how quickly the model is adapted to the task. A learning rate that is too large can impact the model to converge too quickly to a suboptimal solution, while a learning rate that is too small can cause a model to not gain the ability to adapt itself to the current problem (Brownlee, 2020b).

3.3. Training and validation of the model

The training process splits the loaded datasets into two sets. The first contains 85% of the data and it is used to train and teach the model how to correctly classify the given samples into the classes that were given by the user. The rest of the data (15% of it) is then used for validation. The validation data is used for understanding and monitoring the model performance on the unseen data, and to track potential issues with underfitting and overfitting. These two are crucial to machine learning as the model cannot be properly trained and assessed without analyzing their effects on the model. Furthermore, correct interpretation of underfitting and overfitting are necessary in order to further tweak and tune the model in the right direction.

The generated models can then be exported in order to be used in any website or application. Users can also test the accuracy on the new data via the service itself as it allows users to upload new data for testing.

4. Results and discussion

The model developed with dog and cat images after 10 epochs reached almost an accuracy of 100% on the testing data. The model's accuracy was then tested against new data via the upload feature in the online service. It showed promising results as only 1 out of 10 images was incorrectly classified. The problematic image contained a picture of a wolfdog. A possible reason for this incorrect classification can be attributed to the dog's pointy ears, as they are almost always present within the cat family, while some dog species do not have pointy ears. This example paints a clearer picture into how exactly machine learning works. Here, based on prior entries, i.e. the provided dataset, the model predicted that the image is more akin to the class named cat.

The following information is important and crucial as it allows users to further tweak and change the model in order to better fit the need. After retraining the model with 15 images per class and a more diverse section of dog breeds, the wolfdog was now correctly classified alongside other images. The correct classification of the image is shown in figure 1.

Figure 1. *Correct classification of an image containing a wolfdog using the service Teachable Machine*

Source: Authors

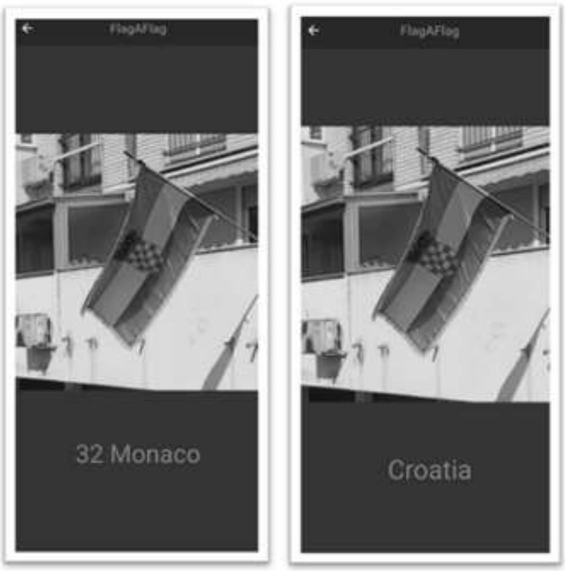
This is one of the ways a model developed though the service Teachable Machine can be enhanced and retuned. For this particular model to be ready for experimental use in the real world, further testing and experimentation has to be conducted within the dataset and model to assess potential problems and obstacles that may occur during the real-world classification.

While creating a simple binary classifier was rather successful, the problems started to arise when creating a more complex image classifier. The complex classifier contained more than 4,000 images, and in total 48 different classes. This also changed the classification challenge, as the classification was no longer binary but had multiple classes. In order to compare the model developed with the Teachable Machine service, a custom model was developed using the framework TensorFlow. The classification challenge dealt with national flag identification. The custom model yielded validation accuracy of 94% after 100 epochs. Similar results were achieved by the service as its validation accuracy hovered around 94% after 100 epochs.

The problems surfaced during the testing of the accuracy of both models against real-world data. Both models were implemented inside an Android mobile application designed to classify flags. In order to properly assess the accuracy of both models, over 100 images were selected from the user’s gallery or were captured by a phone camera. After real-world testing, the accuracy of the custom model was very similar to the accuracy achieved during the training. Out of 118 images 113 (95,6%) were correctly classified, whereas the model developed with Teachable Machine service showed accuracy of only 30% on the real-world data. This means that the drop of accuracy is very significant.

Such a big difference between the results can be attributed to the amount of tinkering and tuning that can be done within the custom model, opposingly, very few variables can be adjusted and changed within the Teachable Machine service. One potential problem of the low accuracy of the real-world data could be attributed to the lack of data shuffling. The lack of shuffling can contribute to overfitting, as the model over time learned the order and particular features of the dataset too well. With data shuffling, a model will receive a different combination of the dataset in each epoch and this way the order of data should not influence the output of the model (Bouda, 2017). Another reason for the poor performance could be assigned to the lack of data augmentation. Data augmentation allows for image rotation, mirroring and normalization in order to simulate the unpredictability of real-world data. This technique also further expands the dataset to compensate for the lack of data. The lack of these features makes it hard to make a robust model based on a complex dataset ready for real-world experimentation and use (Bhandari, 2020). Figure 2 shows the difference in classification between a custom-built model and the model generated through the online service.

Figure 2. *Left: incorrect classification of the Croatian flag using the Teachable Machine model; right: correct classification done via the custom model developed with the framework TensorFlow*



Source: Authors

So, when exactly can one rely on this online service? It remains a good introduction to machine learning as it allows anyone to quickly train their own artificial intelligence without facing many challenges during the development process. In case of developing a smaller model or for educational purposes, the mentioned Google service seems to be a good option, but if the goal is to develop a model with a large dataset or many different classes that would be used for commercial purposes, it is necessary to develop a custom model trained with a specific purpose in mind.

5. Conclusion

This paper delved into an online service by the name Teachable Machine. In order to properly assess the functionality of this service, two models were developed. The first model developed using a smaller and simpler dataset yielded great results after additional retuning, as it managed to classify 100% of the testing data correctly using the upload feature. The second or the more complex model was developed using over four thousand images assigned to almost 50 different classes and was compared to a custom model developed using TensorFlow. The main goal of the models was to predict the correct class of images unseen during the process of training. Both models were implemented within a mobile application and tested on over a 100 images. The custom-built model showed much better performance than the complex model trained using the service as it reached an accuracy of only 30%. The poor performance was attributed to the lack of data shuffling and data augmentation. Lack of these features is usually assigned to overfitting as the model showed good performance during the training process, but the accuracy was found to be significantly lower during the real-world testing.

Based on the research results, the service Teachable Machine can be widely used for creation of simpler classifiers intended for educational purposes, however, its efficiency and performance are subpar in the issues dealing with complex classification. For this reason, it is important to develop custom models that can properly process and assess the data in more challenging ways.

References

1. Anwar, H. et al. (2019). CoinNet: Deep Ancient Roman Republican Coin Classification via Feature Fusion and Attention.

https://www.researchgate.net/publication/335420202_CoinNet_Deep_Ancient_Roman_Republican_Coin_Classification_via_Feature_Fusion_and_Attention (7.5.2021.)

2. Bhandari, A. (2020). Image Augmentation on the fly using Keras ImageDataGenerator! https://www.analyticsvidhya.com/blog/2020/08/image-augmentation-on-the-fly-using-keras-imagedatagenerator/ (19.5.2021.)
3. Bouda, T. (2017). Day 43: Shuffle - 100 days of algorithms. https://medium.com/100-days-of-algorithms/day-43-shuffle-b5abe4644c23 (19.5.2021.)
4. Brownlee, J. (2020a). Difference Between Algorithm and Model in Machine. https://machinelearningmastery.com/difference-between-algorithm-and-model-in-machine-learning/ (19.5.2021.)
5. Brownlee, J. (2020b). Understand the Impact of Learning Rate on Neural Network Performance. https://machinelearningmastery.com/understand-the-dynamics-of-learning-rate-on-deep-learning-neural-networks/ (19.5.2021.)
6. Dunđer, I. (2020). "Machine Translation System for the Industry Domain and Croatian Language", *Journal of Information and Organizational Sciences*, vol. 44(1), 33-50.
7. Dunđer, I. and Pavlovski, M. (2019a). "Behind the Dystopian Sentiment: a Sentiment Analysis of George Orwell's 1984". U: *42nd International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, 577-582.
8. Dunđer, I. and Pavlovski, M. (2019b). "Through the Limits of Newspeak: an Analysis of the Vector Representation of Words in George Orwell's 1984". U: *42nd International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, 583-588.
9. Dunđer, I., Seljan, S. and Pavlovski, M. (2020). "Automatic Machine Translation of Poetry and a Low-Resource Language Pair". U*: 43rd International Convention on Information, Communication and Electronic Technology (MIPRO)*, 1034-1039.
10. Dunđer, I., Seljan, S. and Stančić, H. (2015). "Koncept automatske klasifikacije registraturnoga i arhivskoga gradiva". U: *48. savjetovanje hrvatskih arhivista (HAD)*, 195-211.
11. Google. (n.d.). Teachable Machine. https://teachablemachine.withgoogle.com/ (19.5.2021.)
12. Hart, E., Cha, S.-H. and Tappert, C. C. (2004). Interactive Flag Identification Using Image Retrieval Techniques.

https://www.researchgate.net/publication/221296911_Interactive_Flag_Identification_Using_Image_Retrieval_Techniques (28.4.2021.)

13. Jia, B. et al. (2018). “Space object classification using deep neural networks”. U: *2018 IEEE Aerospace Conference*, 1-8. https://doi.org/10.1109/AERO.2018.8396567
14. Linares, R. (2016). Space Object Classification Using Deep Convolutional Neural Networks. https://www.researchgate.net/publication/308904236_Space_Object_Classification_Using_Deep_Convolutional_Neural_Networks (7.5.2021.)
15. Lodh, A. and Parekh, R. (2016). “Computer Aided Identification of Flags using Color Features”, *International Journal of Computer Applications*, vol. 149(11), 1-7.
16. Lugović, S. and Dunđer I. (2017). “Automatic Information Behaviour Recognition”. U: *40th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO 2017)*, 1429-1432.
17. Lugović, S., Dunđer, I. and Horvat, M. (2016). “Techniques and applications of emotion recognition in speech”. U: *39th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, 1278-1283.
18. Nguyen, H. et al. (2017). “Animal Recognition and Identification with Deep Convolutional Neural Networks for Automated Wildlife Monitoring”. U: *2017 IEEE International Conference on Data Science and Advanced Analytics (DSAA)*, 40-49. https://doi.org/10.1109/DSAA.2017.31
19. Said, Y. and Barr, M. (2021). “Countries flags detection based on local context network and color features”, *Multimedia Tools and Applications*, vol. 80, 14753-14765.
20. Sladojevic, S. et al. (2016). Deep Neural Networks Based Recognition of Plant Diseases by Leaf Image Classification. Computational Intelligence and Neuroscience, vol. 2016. https://doi.org/10.1155/2016/3289801