TMU19 at NTCIR-15 Retrieval Task

Duy-Duc Le Nguyen[1,#], Yu-Chi Lang[1], Yung-Chun Chang[1,2]
[1]Graduate Institute of Data Science, Taipei Medical University
[2]Clinical Big Data Research Center, Taipei Medical University Hospital

RELATED WORKS

Boris et al. [1] proposed a method to process the original data through feature processing and then use machine learning models to identify human activities.

Ignatove [2] suggested to use CNNs to solve the human activity task.

METHODOLOGY

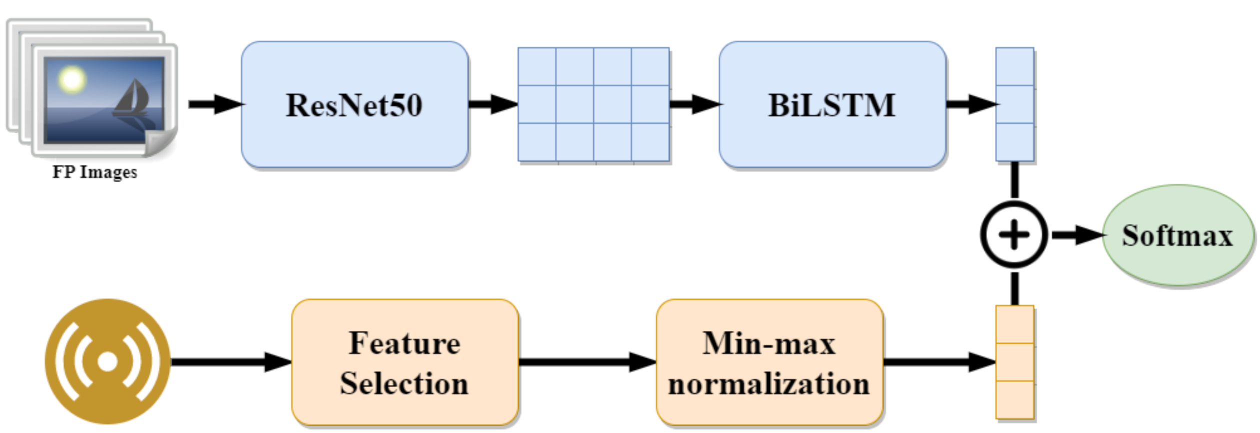

Image Feature Extraction: We extract *m* features from each image. We chose *ResNet50* [3] to be equivalent to provided ResNet probability outputs from the organizer. These features are stacked into a $m \times n$ matrix, As the model has to deal with limited labeled data, all ResNet’s parameters are locked and cannot be learnable.

Bidirectional Long Short-Term Memory (BiLSTM): Two independent *LSTM* [4] layers putting together to form a BiLSTM. We set the hidden size as m to match with the previous block. After processing through the BiLSTM block, its output with $2m$ scalar-value will be concatenated with 3,000-length rich multi-modal data.

Rich Multi-modal Data: We choose *Min-Max Normalization* to convert the original column data. This normalization significantly impacted on the model's performance.

EXPERIMENT & RESULT

Experiemental Settings:

- ResNet's output m is 2048 features
- hidden size of BiLSTM $L = m = 2048$.
- we initialize $n = 16$, the largest number of images in the dataset. If an entity has less than 16 images, it will be padded with a set of zero-value images.
- all images are resized to 256×256, then center cropped and normalized with ▯ = (0.485, 0.456, 0.406) , ▯ =(0.229, 0.224, 0.225).
- We obtain Adam optimizer with the weight decay factor. After trying several learning rates, the optimal learning rate with high slope and lowest training loss is $2e-5$.

We tried to use three common methods in linear regression models to select independent variables for the raw data: **forward selection, backward selection, and stepwise selection**.

After comparing these three methods, **only fifteen independent variables** are remaining.

We observed **an unpredictable scenario** that the correlation of sensor’s duration (89 or 90 seconds) while the sensor's min, max, or average value does not correlate.

According to the model's performance, we found that **deleting the variable hurt the performance.**

Result

Methods	Leave-One-Out	10-folds	Scoreboard
Random Forest	0.181	0.568	-
SVM	0.297	0.377	-
XGBoost	0.395	0.625	0.399
Our model	**0.540**	**0.638**	**0.465**

Validating: In both leave-one-out and 10-fold cross validation, our model performance grants the first place. XGBoost easily beats three traditional machine learning methods as it is the most evolutionary in decision-tree-based ensemble algorithms.

Testing in the organizer scoreboard: Our proposed model and XGBoost are used to submit into NTCIR-15 Retrieval Task scoreboard system. Our model's submission returns 0.465 score while XGBoost's performance is almost 0.4

REFERENCES

[1] Boris Ginsburg, Patrice Castonguay, Oleksii Hrinchuk, Oleksii Kuchaiev, Vitaly Lavrukhin, Ryan Leary, Jason Li, Huyen Nguyen, Yang Zhang, and Jonathan M. Cohen. 2019. Stochastic gradient methods with layer-wise adaptive moments for training of deep networks. *arXiv preprint arXiv:1905.11286 (2019).*

[2] Andrey Ignatov. 2018. Real-time human activity recognition from accelerometer data using Convolutional Neural Networks. *Applied Soft Computing 62, (2018)*, 915–922.

[3] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2015. Deep Residual Learning for Image Recognition. *arXiv:1512.03385 [cs]* (December 2015). Retrieved

November 1, 2020 from http://arxiv.org/abs/1512.03385
[4] Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long Short-Term Memory. *Neural Computation 9, 8 (November 1997)*, 1735–1780. DOI:https://doi.org/10.1162/neco.1997.9.8.1735