TMU19 at NTCIR-15 MART Retrieval Task

Duy-Duc Le Nguyen[1,#], Yu-Chi Liang [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

ABSTRACT

This paper proposes a new method for predicting user activities at the NTCIR-15 Micro-activity Retrieval Task. Additional concepts from ResNet generated features following with a Bidirectional Long-Short Term Memory block helps our neural network paying more attention in the corresponding class. Our model received an upright result on the scoreboard.

KEYWORDS

Rich Multi-modal Data, Activity Recognition

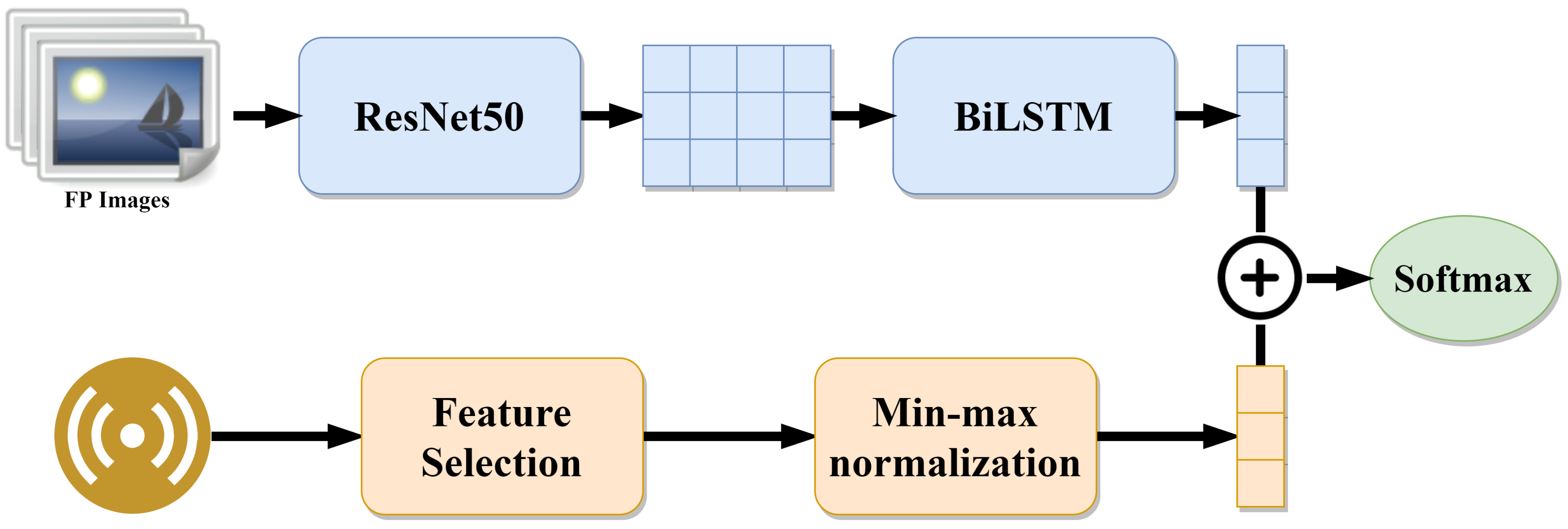
INTRODUCTION

We are living in a multi-modal data world — we hear sounds, smell scents, touch surfaces, see things and taste flavors. A query will be defined as a multi-modal problem if it covers multiple modalities. As this research area is quite fresh, we have decided to step in NTCIR15-MART task to find a novel method. This project can bring lots of advantageous application to the real world. In a clinic environment, AI applications will diagnose patient's state and activity based on their biological data, breathing rate, voice and images. Hence, doctors and nurses can promptly meddle in case of emergencies.

METHODLOGY

Image Feature Extraction: We extract features from each image. There are various methods that generating crucial insight from images such as DenseNet, AlexNet and ResNet. We chose ResNet50 to be equivalent with provided ResNet probability outputs from the organizer. These features are stacked into a matrix. Since the model has to deal with limited labeled data, all of the parameters in Image Feature Extraction block are locked and cannot be learnable.

Bidirectional Long Short-Term Memory (BiLSTM): Two independent LSTM layers putting together to from a BiLSTM. This structure allows the model to have both backward and forward information about time-series data and tackle the vanishing gradient problem from traditional recurrent neural network architectures. Using bidirectional will run input in two ways, one from past to future and one from future to past. The differs this approach from unidirectional is that BiLSTM can learn itself how and when to forget and when not to using gates. As the result, the model can understand the context better. We set the hidden size as to match with the previous block.


FP Images
ResNet50
BiLSTM
Softmax
Feature Selection
Min-max normalization

EXPERIMENT & RESULT

280 activities are released with labels as a training set and 140 activities are a testing set. All scores are ranked in mAP (mean average precision). We obtain Random Forest, Support Vector Machine and XGBoost as the baseline models. The performance are described as the table below:

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**

In both leave-one-out and 10-fold cross validation, our model performance grants the first place. XGBoost easily beats ther traditional machine learning methods as it is the most evolutionary in decision-tree-based ensemble algorithms. Our proposed model and XGBoost are used to submit into NTCIR15-MART scoreboard system. Our model’s submission returns 0.465 score in the organizer evaluation system while result is almost 0.4.

CONCLUSION

In summary, our approach in separately splitting difference modality data into two neural networks can achieve a fine mean average precision score in NTCIR15-MART retrieval task. There is a lot of room to optimize in our approach. Hence, we will keep configuring this model in the future.

Acknowledgments

This research was supported by the Ministry of Science and Technology of Taiwan under grant MOST 107-2410-H-038-017-MY3, MOST 109-2410-H-038-012-MY2, and MOST 107-2634-F-001-005.