AJCAI2021 (author)

Help / Log out

My Submissions | AJCAI2021 | Premium | Conference | News | EasyChair

AJCAI2021 Submission 8

Proceedings version

Submission information updates are disabled. If you need to update your paper, contact AJCAI2021 chairs.

For all questions related to processing your submission you should contact the conference organizers. Click here to see information about this conference.

This submission is included in proceedings. Click on "Proceedings version" in the vertical menu to access the proceedings.

All **reviews sent to you** can be found at the bottom of this page.

Submission 8	
Title:	An Explanation Module for Deep Neural Networks Facing Multivariate Time Series Classification
Paper:	(Aug 27, 02:57 GMT)
Track:	Main Research Track
Author keywords:	Multivariate time series classification Explanation Module Deep Learning Data Mining
EasyChair keyphrases:	time series (288), neural network (246), deep neural network (206), multivariate time series classification (200), multivariate time series (147), feature map (130), recurrent neural network (95), convolutional neural network (79), time series classification (79), variable id importance (79), time series data (63), deep learning model (63), data mining (60), wafer dataset (60), classification accuracy (60), computer vision (60), multivariate time (60), explanation module (60), temporal feature (60), variable accuracy (50), deep learning (50), accuracy change (50), excessive length (50), total grouped convolution (47), time series forecasting (47), time series sequence (47), multivariate time series dataset (40), dataset accuracy (40), popular deep neural network (40), deep convolutional neural network (40)
Topics:	11. AI applications and innovations; e.g. healthcare; education; entertainment; smart city; IoT; blockchain.,, 16. Data-driven AI: big data; representation; analytics and visualisation,
Abstract:	Deep neural networks currently achieve state-of-the-art performance in many multivariate time series classification (MTSC) tasks, which are crucial for various real-world applications. However, the black-box characteristic of deep learning models impedes humans from obtaining insights into the internal regulation and decisions made by classifiers. Existing explainability research generally requires constructing separate explanation models to work with deep learning models or process their results, thus calling for additional development efforts. We propose a novel explanation module pluggable into existing

	deep neural networks to explore variable importance for explaining MTSC. We evaluate our module with popular deep neural networks on both real-world and synthetic datasets to demonstrate its effectiveness in generating explanations for MTSC. Our experiments also show the module improves the classification accuracy of existing models due to the comprehensive incorporation of temporal features.
Submitted:	Jun 28, 06:29 GMT
Last update:	Aug 22, 01:27 GMT
Author conflicts:	none

Authors

first name	last name	email	country	affiliation	Web page	corresponding?
Chao	Yang	chao.yang@student.uts.edu.au	Australia	University of Technology, Sydney		✓
Xianzhi	Wang	Xianzhi.Wang@uts.edu.au	Australia	University of Technology, Sydney		
Lina	Yao	lina.yao@unsw.edu.au	Australia	University of New South Wales		
Jing	Jiang	Jing.jiang@uts.edu.au	Australia	University of Technology, Sydney		
Guandong	Xu	Guandong.xu@uts.edu.au	Australia	University of Technology, Sydney		

Reviews

Review 1

Overall evaluation:

1: (A good submission - weak accept)
This paper proposes a flexible module to insert into a DNN to help learn more explainable classifiers. The module is generic and was intentionally designed to help explain multivariate time series classication (MTSC). The experiment results show a positive correlation between the predicted importance score and the real impact of the variable to test performance.

Strong points:
1. An interesting explanation module, carefully designed and well tested.
2. The importance score is empirically verified with the random noise interference effect.

Weak points:
1. Missing baselines. Why cannot just use the gradient of the loss to the input variable as the importance score, will it be better than the proposed method?
2. Relation to LIME?
3. Why do we have to use a learnable module? It seems to go against the idea of

	inference-time explanation since at this time the model should be fixed. 4. Fig. 1, the F_backbone should be FM_backbone? 5. The attention and self-attention part above Eqn. (1) is strange. Why it is a sel-attention? It is quite different from the sel-attention concept in NLP. 6. What is "text process"? (in the second para. above 4.4) 7. Table 2, the Accuracy change for variable 4 should be 4.9, not 0.49.

Review 2	
Overall evaluation:	**1**: (A good submission - weak accept) This paper proposes a module to augment a deep learning model to investigate variable importance for explaining multivariate time series classification tasks. The module is added to variants of deep neural network architectures specifically Recurrent Neural Networks (RNN) and Convolutional Neural Networks (CNN). Experiments are conducted on four real-world data sets. In addition, noise is added to the input variables to generate synthetic versions of the real-world data sets to demonstrate the efficacy of the approach. The authors claim that not only can that this module explain variable importance but it can also improve model performance. There is good motivation for the research, the research gap identified, and the contributions of the work clearly articulated. In terms of the architecture of the module, it is described well. However, more justification for the data sets used for the experiments should have been included. In addition there was no argument made for the train-test ratios used for each data set. With regard to the results there was only one model evaluation metric used and the results of which variables were important compared with the output of a competing approach for model interpretability such as Shapley values. Also as the average accuracy was reported then the standard deviations should have been included. Finally, there was no discussion on the limitations or weaknesses of the approach nor scope for future work.

Review 3	
Overall evaluation:	**3**: (Top 10% of accepted AJCAI papers - strong accept) The paper proposes a generic module for neural networks, which enables the interpretation of the prediction results for multivariate time-series data. Experiments validate the effectiveness of the proposed method by using different neural networks. Strengths - An straightforward interpretation method is proposed. - The module is general and can be incorporated into different neural networks. - Experimental results are good Weaknesses - While this method is generic, it is better to see how it automatically adjust itself to fit different neural networks.

Metareview

Metareview for paper 8	
Title:	An Explanation Module for Deep Neural Networks Facing Multivariate Time Series Classification
Authors:	Chao Yang, Xianzhi Wang, Lina Yao, Jing Jiang and Guandong Xu
Text:	Please address the reviewer comments - especially weak points highlighted by reviewer 2