PEARL: Prototype Learning via Rule Learning

Tianfan Fu*
tfu42@gatech.edu
Georgia Institute of Technology
Atlanta, Georgia, USA

Tian Gao*
tgao@us.ibm.com
IBM Research
Yorktown Heights, USA

Cao Xiao
cao.xiao@iqvia.com
Analytics Center of Excellence, IQVIA
Cambridge, USA

Tengfei Ma
tengfei.ma1@us.ibm.com
IBM Research
Yorktown Heights, USA

Jimeng Sun
jsun@cc.gatech.edu
Georgia Institute of Technology
Atlanta, Georgia, USA

ABSTRACT

Deep neural networks have demonstrated promising prediction performance on many health analytics tasks. However, the interpretability of the deep models is often lacking. In comparison, classical interpretable models such as decision rule learning do not lead to the same level of accuracy as deep neural networks (DNN) and can also be too complex to interpret (e.g., due to large tree depths). In this work, we propose Prototype LeArNing via Rule Learning (PEARL), which iteratively constructs a decision rule list to guide a neural network to learn representative prototypes that can be explained by the associated rules. The resulting prototype neural network inherits both the prediction power of DNNs and interpretability associated with rules, thus can provide accurate and interpretable predictions. Evaluated on real world health datasets, PEARL demonstrates state-of-the-art accuracy to various DNN baselines and interpretable results that are simpler than standard decision trees can provide.

CCS CONCEPTS

• **Computer systems organization** Embedded systems; *Redundancy*; Robotics; • **Networks** Network reliability.

KEYWORDS

Deep Learning; Interpretable Machine Learning; Healthcare

ACM Reference Format:

Tianfan Fu*, Tian Gao*, Cao Xiao, Tengfei Ma, and Jimeng Sun. 2019. PEARL: Prototype Learning via Rule Learning. In *10th ACM Int'l Conference on Bioinformatics, Computational Biology and Health Informatics (ACM-BCB '19)*, September 7–10, 2019, Niagara Falls, NY, USA. ACM, New York, NY, USA, 1 page. <https://doi.org/10.1145/3307339.3343250>

1 MOTIVATION

Although deep learning models can produce accurate predictions and classifications [1] in electronic health records (EHR) data, they are often treated as black-box models that lack interpretability and transparency of their inner working [3].

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

ACM-BCB '19, September 7–10, 2019, Niagara Falls, NY, USA

© 2019 Copyright held by the owner/author(s).

ACM ISBN 978-1-4503-6666-3/19/09.

<https://doi.org/10.1145/3307339.3343250>

Decision rule or rule list can be interpretable to domain experts as they are expressed in simple logical forms [4]. However, the accuracy of rule-based models is often lower than deep neural networks.

Prototype learning is another interpretable model inspired by case-based reasoning [2], where observations are classified based on their proximity to a prototype point in the dataset.

2 METHOD: PEARL

PEARL is an integrative prototype learning neural network that combines rule learning and prototype learning on deep neural networks to harness the benefits of these methods.

- **Interpretation Module (Rule Learning)** employs a rule list classifier (e.g., a decision tree) to provide interpretable prototype definitions.
- **Prediction Module (Neural Network)** contains a patient representation learning and a prototype learning network for prediction.
- **Iterative Data Reweighting.** To enable the iterative learning of prototypes and rule list, we use a data re-weighting procedure based on results from the prediction module.

ACKNOWLEDGEMENTS

This work was supported by the National Science Foundation award IIS-1418511, CCF-1533768 and IIS-1838042, the National Institute of Health award 1R01MD011682-01 and R56HL138415.

REFERENCES

- [1] Edward Choi, Mohammad Taha Bahadori, Andy Schuetz, Walter F. Stewart, and Jimeng Sun. 2016. Doctor AI: Predicting Clinical Events via Recurrent Neural Networks. In *Machine Learning for Healthcare Conference*. JMLR.org, Stanford, USA, 301–318.
- [2] Janet L. Kolodner. 1992. An Introduction to Case-based Reasoning. *Artificial intelligence review* 6, 1 (1992), 3–34.
- [3] Zachary Chase Lipton. 2016. The Mythos of Model Interpretability. *CoRR* abs/1606.03490 (2016), 03490.
- [4] Ronald L. Rivest. 1987. Learning Decision Lists. *Machine learning* 2, 3 (1987), 229–246.