Rule Learning Vs. Tree Learning

Mohammad Doosti Lakhani – 98722278

Rule Learning

Rule based learning approach consists of extracting rules from given training data instances that can match the description of the desired instance with one of the possible classes. Training happens by splitting instances to two groups of positive samples which has the label of desired class and a group of negative samples which has a different label.

As the Rule Based Learning approaches cannot generalize very well w.r.t. available renown approaches such as Deep Learning or Evolutionary algorithms, they have been used as an auxiliary method to alleviate some of the challenges in the aforementioned state of the art approaches but still cannot catch the generalization and robustness of the approaches such as deep learning.

From *A. Fernandez et al.* [2] we can see they have adopted Genetic Algorithm to evolve the rules obtained via Rule Based Learning approaches. Fig 5 of this paper depicts a %7 percent increase in accuracy of Genetic combined with Rule based learning w.r.t. pure Rule Learning.

On the other side, Rule Learning approaches also have been adopted in state-of-the-art deep learning tasks such as Sentiment Analysis.[3]
In this paper, author have demonstrated that basic deep learning models almost all time catch some aspects in aspect sentiment analysis which are useless and not related. So, to overcome this problem, they have used a CNN to tag each aspect of an opinion sentence then combining rules from Rule based learning with feature extracted by CNN model to improve the performance of aspect extraction method and sentiment scoring method. They have achieved approximately %7-%12 increase in accuracy in their model.

But finally, Rule based learning approaches are not the favorite because of the following problems:

1. Not robust to noises, outlier data
2. Cannot be adopted from task to task
3. Updating a rule takes a lot of time and energy every time
4. Rules are to explicit and they increase in size overtime which cause redundancy
5. Needs human expert for tracing back their origin

Tree Learning

In Tree Learning, attributes from training data are the label of each inner node of tree that enable tree to do branching based on a threshold which is obtained from the possible values of that particular attribute. The splitting and selecting proper attribute happen through some heuristic learning approaches such as Gini index. Finally, each leaf is one of the desired classes.

Currently, trees are used mostly in Ensemble Learning approaches for imbalanced data which can achieve high scores regarding their simplicity in term of implementation and theoretically.

One of the challenges in Tree Learning is the size of tree which can be grown to a huge size for tasks with many attributes such text mining, etc. In this situation, it has been proven that the optimal trees can be achieved if the tree constructed in one parse not incrementally. This is not practical in large datasets so to handle this obstacle, hardware-software design need be considered. For instance, A huge speedup and approximately 1.2-1.3% increase in accuracy can be achieved by adopting mixed-integer optimization in finding optimal tree. [4]

We said that the real challenge in trees are the selection of appropriate attribute for a node as a splitter. But the challenge of choosing this attribute is hard as it related to information theory which focus on the task that which attribute (data) gives us most of the information regarding other possible attributes. So as we said previously in definition of Tree Learning, a heuristic decides about the selection of attribute so finding a optimal heuristic that can be used in different tasks with different data types and ranges can lead us to more generalization and more stability and robustness while achieving high score w.r.t. previously available heuristics such as Gini or Entropy.

Even though Tree Learning approaches are in use nowadays, but because of a few problems they are not scalable to modern real world problems such as optimal control problems. Some of these problems can be listed as below:

1. Small changes cause instability (advantage in ensemble learning)
2. Much higher time than other algorithms in large problems
3. Training is expensive and sometimes impossible in some cases

References

1. J. Furnkranz et al., Foundations of Rule Learning, Cognitive Technologies, DOI 10.1007/978-3-540-75197-7 2, © Springer-Verlag Berlin Heidelberg 2012
2. A. Fernandez et al., Genetics-Based Machine Learning for Rule Induction: State of the Art, Taxonomy, and Comparative Study, IEEE Transactions on Evolutionary Computation, 2010
3. P. Ray et al., A Mixed approach of Deep Learning method and Rule-Based method to improve Aspect Level Sentiment Analysis, Applied Computing and Informatics, Elsevier, 2019
4. D Bertsimas et al., Optimal classification trees, Springer, 2017