Critical Review

Explainable AI for Classification using Probabilistic Logic Inference

Explainable Artificial Intelligence in simpler words is AI that can be understood by humans. It is where the user is enriched with both prediction and reasoning of why a certain class was chosen for a query. The following paper by X. Fan et al proposes a method of explainable AI using probabilistic Knowledge bases that provides an explanation for the prediction made.Knowledge base is the representation of dataset in the form of clauses with respect to their probabilities.It can be formulated using two methods: (1) the tree method and (2) the direct method.The paper also gives algorithms that explain the working principle of formulating Knowledge bases from the respective methods.

Furthermore, the author explains that the method he proposed of constructing KB's is inconsistent as when a query and KB is introduced together, it leaves no solution for the probability distribution making it inconsistent. However, the author ensures that with taking the computation as just an optimization problem, the inconsistency becomes tolerable.The computation of literal probabilities is done irrespective of the probability distribution.He highlights this as the core of his inference method.He proposes a method of linear programming to get rid of inconsistency which also leads the proposed solution to be non-parametric.Moreover, the author also adds that he is using the local approach of finding relative features for a specific query instance. He introduces the method of subqueries which leads to incorporating knowledge similar to clauses and introduces acceptability for incomplete and imperfect knowledge.

To evaluate the performance analysis, the authors used 4 classifiers CART, Multi layer perceptrons, random forest and Support vector machines(SVM). The models were trained and tested on both synthetic and non synthetic datasets and F1 scores were calculated for comparison.The results were compared with the state of the art explainable AI named SHAP. The results for direct KB method on synthetic datasets gave higher accuracies than that of SHAP, ensuring the efficiency of the proposed solution.Lastly, the authors mention 4 research directions for future enhancements that they are going to explore which include (1) development of classification techniques, (2) semantics for inconsistencies in the Knowledge base, (3) richer explanation generation and (4) enhancements in knowledge incorporation.

The paper presents a practical nonparametric solution to Explainable AI with accuracies higher than that of SHAP. It is backed up with mathematical logic, proposing solutions to the shortcomings coming along each algorithm. The main core development of this paper is the ability to make inconsistencies tolerable and due to that being able to incorporate domain knowledge and process datasets that are incomplete or imperfect .Furthermore, the theory is supported by comparative tests with well known classifiers on well known datasets. Overall the paper presented results with visualizations and the issues were countered with future enhancements by the author. The main motivation behind this paper is to make AI a more trustable field by giving out reasonings. Incorporating AI in medicine has been challenging because patients need a diagnosis showing the symptoms that caused the disease. Therefore, Explainable AI using knowledge bases is a promising solution in the upcoming future that is going to make decision accuracy more accurate and trustable.

However, there are few points that were neglected by this paper that include the extra computational cost of formulating knowledge basis.This paper does not discuss the limitations of technology that come due to requirement of large memory storage and high speed processor to compute the algorithms. Furthermore, on increasing feature values the knowledge base increases exponentially keeping in note the creation of (nCk) subqueries. Thus, resulting in exponentially growing costs. Due to this margin, this algorithm has a limitation of using limited feature values, making it infeasible for large datasets. Moreover, for performance analysis, the algorithm is only compared with SHAP and that only proved fruitful on synthetic dataset. Whereas, the paper should have compared it with more than one similar domain to understand the performance irrespective of the limitations of the compared algorithm.

In conclusion, the respective paper represents details of different aspects with mathematical and probabilistic logics enlightening the reader about explainable AI in depth. The details show that the paper is well researched with 40 citations of included researched papers making it an authentic resource for reference. The study in it proves to be performing better than the widely known SHAP algorithm, making it an effective and efficient choice. However, more work should be done on defining the limitations of the model.