Review

Explainable AI for Classification using Probabilistic Logic Inference: A Review

Aymen Bashir, Department of Artificial Intelligence, CureMD
Lahore, Pakistan

The research paper in question devised a method to explain classification decisions made by artificial intelligence. The goal of the paper is to make classification decisions systems more transparent that will increase their reliability. The decisions are made explainable using probabilistic knowledge interference approach while constructing Knowledge Bases (KBs) using linear programming. KBs consists of a set of clauses with each clause having a probability. Each clause is a combination of literals where each literal is a propositional variable. The paper proposes methods to construct a Knowledge Base with all possible clauses and their probabilities in an efficient manner and devise an algorithm for classification of new data. Two algorithms have been developed for the construction of Knowledge Bases (KBs).

1) KBs from decision trees
2) KBs directly from data

Authors of Reviewed Paper: Xiuyi Fan1 , Siyuan Liu1 , Thomas C. Henderson2
1Department of Computer Science, Swansea University, UK
2School of Computing, University of Utah, USA
{xiuyi.fan,siyuan.liu}@swansea.ac.uk,
tch@cs.utah.edu

Each path of decision tree act as a clause. The clauses also include path from root to each node to include all possible clauses and address the data with incomplete information or missing features. Ratio between positive and total samples is calculated as probability of a particular clause. The direct approach adds features values of all incoming data points as clauses and counts to find probability associated with each combination of feature values.

The querying approach uses an optimization technique to find the probability of a literal and not just the clause given a query. If the probability is greater than 0.5 then the literal is assigned to the query.

The performance of the classification using KBs from the two algorithms is tested on various datasets that include Titanic, Mushroom, Nursery and HIV-1 protease cleavage datasets, UK Parliament Bill and image dataset for vehicle classification. Results of both the algorithms promise good F1 scores that are comparable with other classification algorithms that include CART, multi-layer perceptron, Forest, and Support Vector Machine.

The explanation approach of the algorithm is compared with Shapley Value based approach for explanation for Titanic and Mushroom Datasets which also shows similar outcomes. Due to no ground truth to evaluate the accuracy of the explanations, a synthetic dataset was constructed that showed a better performance of Direct method than Shapley approach.

The paper provides a thorough insights into all the approaches being devised. A foundation is set before diving into more complex algorithms. All the approaches are represented in the form of algorithms and examples. The shortcoming of alternative approaches such as Nilson method to find complete conjunction set, are discussed. Computational and time complexity of the devised algorithm is also addressed and compared in the form of graphs. To assess the accuracy of explanations derived from the algorithms, a completed new dataset was created that increased the credibility of the method. Moreover, the approach considers the data with missing features which is a novelty since the Shapley method only works with data with all features available.

However, the method is non-parametric which means no tuning is involved. This means that the algorithm is not going to adjust to changing feature value relationships in the data. Moreover, for large datasets constructing knowledge bases will become computationally expensive. A thorough analysis of a Shapley approach is missing which would have helped to compare the two methods in terms of space, computational and time complexities.

Also, while testing the proposed algorithms on image dataset, feature extraction has been applied that reduces features to 12 per image. This approach brings the performance of the proposed algorithms in question since we do not use feature extraction in Deep Neural Network models. The two proposed algorithms must also be tested on images with more features.

Also, performance analysis that shows relationship of run time, number of variables, clauses and their lengths present in a KB is not explained in detail. We do not know why the clauses with length 30 per clause performs better than clauses with length 20 per clause for variables less than 9000.

The most important aspects of the paper include the performance analysis, comparison of the two proposed algorithms with baseline algorithms and the comparison of explanation results. The performance analysis is significant in determining if the algorithms are feasible in terms of time. The KBs with large number of variables and clauses can be solved using single CPU WITH Xeon 2660v2 processor and 32 GB RAM, however it still takes 10-20 seconds to process KBs with 10,000 clauses with the length of 20 and more. The success of the proposed algorithms in classifying various datasets with nearly same accuracy as other algorithms that validates the performance of the two proposed algorithms. Also, the synthetic dataset is an important aspect of the research since it designs the conditions to compare the performance of proposed method and Shapely method where proposed approach is more competitive. The most important aspect of the paper is the representation of explanation results in graphical form that indicates which features and how many features have remained same during a classification.

An important aspect that sets the proposed methodology apart from the previous work is that the proposed method can handle data with missing features or a smaller number of features. It does not require a query to have all the features. The explanation generation is convenient because the proposed method does not require reconstructing the trained model for sub-queries. The whole methodology can help find features that play a significant role in making a certain decision. The method has an advantage that data can be classified even with a smaller number of features which is especially useful in areas where the data is inconsistent and limited.

The proposed methodology is focused on the explanation aspect rather than the causability. The information of the features that play a role in the decision making of the AI is useful only to a limit since the information is present in technical form in many cases rather than human understandable form e.g., in the image classification, the knowledge of which pixels are playing important role in classification may not be useful if the features are not related to a physical understanding of the practical examples. However, the proposed methodology is useful where features can be explained in a way that it relates them to physical phenomenon. In both the cases, this paper is a significant contribution to making Artificial Intelligence explainable especially for datasets where features are limited and can be directly linked with practical phenomenon.

Overall, the paper is well organized and moves from basic to complex concepts and algorithms while addressing all the questions that may arise in the readers minds. The paper discusses all the techniques that could have been used but were not used due to some problems or limitations that were also discussed.

The literature review section briefly goes over the existing similar methods and their shortcomings. In this regard, ProbLog has been discussed that ProbLog assumes independence of variables that are not derived by Logic Programming. In contrast, the proposed algorithm uses Nilsson's probabilistic logic that makes no such assumption due to which uncertainty is reduced. Other KB approaches are not thoroughly explained due to a limited scope of the paper. Model agnostic explainers have also been discussed but it is unclear how the proposed approach can be compared with the techniques like LIME.

Summing up, the paper presents a novel approach in making classification understandable in terms of feature contribution in a decision. The results are competitive with the existing Shapely approach; however, the algorithm should be tested on more synthetic datasets to assess if the approach performs better than Shapely on other datasets as well. The comparison in term of computational cost should be kept in mind while further optimizing this technique. Moreover, the algorithms need to be tested on images that contains large number of features. Further developments should focus on explaining feature contribution in the form of physical phenomenon. The paper has increased the scope of feature contribution identification by allowing partial queries with missing features to be classified as well. This definitely is a step forward in the domain of Explainable AI.

References:

[1] Transfer Learning for Human & AI. 2021. *Explainable AI for Classification using Probabilistic Logic Inference - Transfer Learning for Human & AI*. [online] Available at:<https://transfer-learning.ai/paper/explainable-ai-for-classification-using-probabilistic-logic-inference/> [Accessed 28 March 2021].

[2] *Andreas Holzinger, Bernd Malle, Anna Saranti, Bastian Pfeifer, Towards multi-modal causability with Graph Neural Networks enabling information fusion for explainable AI,Information Fusion, Volume 71,2021,Pages 28-37, ISSN 1566-2535,https://doi.org/10.1016/j.inffus.2021.01.008.(https://www.sciencedirect.com/science/article/pii/S1566253521000142)*

[3] *Schmelzer, R., 2021. Understanding Explainable AI. [online] Forbes. Available at: <https://www.forbes.com/sites/cognitiveworld/2019/07/23/understanding-explainable-ai/> [Accessed 28 March 2021].*