Saint Petersburg State University

Department of mathematical game theory and statistical decisions

Xu Feiran

Master's Research Report

High-Dimensional Explainable AI for Cancer Detections

Specialization 01.04.02

Applied Mathematics and Informatics

Master's Program Game Theory and Operations Research

Research advisor,

Petrosian Ovanes

Saint Petersburg

2022

1. Introduction.

Anomaly detection is a technique used to identify abnormal patterns that do not conform to expected behavior (Patcha A et al., 2007). In Hawkins' view, outliers are quite different from other observations in the sample, and therefore cannot be trusted (Hawkins D M, 1980). In the case of medical physiological data analysis, it means that if the patient's physiological index data deviates from other observations, then his probability of suffering from cancer will significantly increase. In this paper, we use the Isolation Forest anomaly detection algorithm to analyze the cancer data set, and explain the results of anomaly detection from the local and global perspectives.

At present, the practicability of artificial intelligence in recognition and prediction has reached a very high level. Several researchers are already proposed some AI approaches for cancer detection and treatment (Kuswanto H et al., 2019). However, the problems of ethics (Cath C et al., 2018) , trust (Lui A et al., 2018), and prejudice (Challen R et al., 2019) in its application in law, medical care, and ethnic minority-related activities can not be ignored. Therefore, for cancer diagnosis results based on Anomaly Detection, Explainable AI is necessary for credible and correct diagnosis results. Previous surveys have fully summarized the classification of XAI methods [6]. In a word, these methods could be divided into two types: global explanation and local explanation.

In the field of Explainable AI, the interpretation of high-dimensional data sets is more complicated. On the one hand, high-dimensional data means that the amount of computation increases sharply. On the other hand, the correlation between various features in the interpretation process will hinder Explainable AI.

2. Project.

I have completed two papers and they have both been published.1): Explainable AI: Using Shapley Value to Explain the Anomaly Detection System Based on Machine Learning Approaches [2]. 2):High-Dimensional Explainable AI for Cancer Detection [3]. In these thesis, although we got not bad results, there are still some problems that need to be improved and I will try my best to fix the problems that exist in my master's degree. The key idea of our work is to find a way to explain machine learning models by means of Shapley values [1].

2.1. Anomaly Detection using Isolation Forest [5].

I have completed this part of the work, which is described in detail in our published paper.

2.2. Anomaly Detection using Artificial Neural Network (ANN).

1. This part of my work is still a work in progress. During the performance testing phase of the model we found that when working on anomaly detection for the same dataset, ANN resulted in much higher prediction accuracy than the isolated forest algorithm($\approx$ 12%). This means that for some scenarios where anomaly detection is required, the ANN model will be in higher demand and it is therefore necessary to devise a feasible interpretation method for this model. In this part of the work, we try to implement a new approach to the interpretation of ANN models with high-dimensional inputs.

2. In the traditional process of calculating shapley values, we need to obtain the values of all coalitions and all coalition characteristic functions, and use these to calculate the shapley value for each participant. In our chosen example, however, the number of data features is defined as the number of participants, and the need to calculate the shapley values for 31 participants is essentially impossible on an ordinary personal computer. We therefore

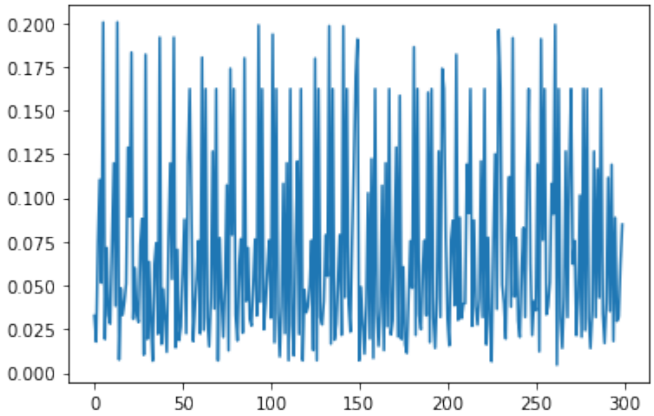
attempted to reduce the computational effort involved in calculating the shapley values by clustering the coalitions of participants and identifying those coalitions that were very close in their impact on the model output during the model exercise.

3. Current problems encountered.

1. In our previous paper, we clustered participants, but there are some problems with this process. In our example, we have 31 features, and clustering of data with a total of 31 is obviously not very appropriate. Although we get good results due to some features of the breast cancer dataset, for a more general case, the results of clustering of dozens of features may not be available. Therefore, we attempt to cluster coalitions to increase the accuracy of clustering results.

2. When we tried to cluster the coalition of participants, we found that this was not a convergent process. When we attempted to cluster for coalitions formed by participants, we found that this was not a convergent process. As shown in the figure below, at each iteration we add 100 new coalitions for clustering and calculate the average of the distance from all points in each category to the cluster centroid combined, but this value does not tend to plateau as the number of iterations increases.

In the clustering task described above, we have clustered each coalition closely according to its characteristic function value, which would lead to a significant loss of information in the clustering process, so we would like to add another attribute to the coalition to change the one-dimensional clustering into two-dimensional clustering. We do not yet have a clear goal for the selection of the second dimension.



4. A possible thing for our research.

In the course of our research, we found a paper that perhaps the concepts mentioned in the paper would be helpful in refining our algorithm.

Multivariate Shapley Interactions: Original concept. This research is from the paper "Interpreting Multivariate Shapley Interactions in DNNs" [4].

$$\begin{aligned} B\left(S_{ij}\right) &\stackrel{\text{def}}{=} \phi\left(S_{ij} \mid N'\right) - \left[\phi\left(i \mid N_i\right) + \phi\left(j \mid N_j\right)\right] \\ &= \sum_{S \subseteq N \setminus \{i,j\}} \frac{(n-|S|-2)!|S|!}{(n-1)!} [\Delta f(S,i,j)] \end{aligned}$$

In this research the authors defined $B(S_{ij})$ to measure the interaction between player i and j in coalition $S = i, j$. Furthermore, they also extend it to multi-players case:

$$B([A]) \stackrel{\text{def}}{=} \phi\left([A] \mid N_{[A]}\right) - \sum_{i \in A} \phi\left(i \mid N_i\right) \tag{1}$$

By definition of $B([A])$, we can know:

1. if $B([A]) > 0$, means that the players in coalition A has positive interaction. The $\phi(A) > \sum \phi(a_i), a_i \subset A$
2. if $B([A]) < 0$, means that the players in coalition A has negative interaction. The $\phi(A) < \sum \phi(a_i), a_i \subset A$

This approach from theory side help us well explain the interaction among coalitions. The author use this for NLP explaining, which they already well eplain the NN by considering each input words as player and the interaction based on words. The orders is more important compare with our case.

5. Future work.

In the work that follows, we will build on the concepts mentioned above to find a feasible and good solution that will refine our proposed interpretation algorithm. Refinement of the interpretation method for machine learning models with high-dimensional inputs.

References

1. Contributions to the Theory of Games[M]. Princeton University Press, 1953.

2. Zou J, Xu F, Petrosian O. Explainable AI: Using Shapley Value to Explain the Anomaly Detection System Based on Machine Learning Approaches[J]. , 2020, 7(1): 355-360.

3. Zou J, Xu F, Zhang Y, et al. High-Dimensional Explainable AI for Cancer Detection[J]. International Journal of Artificial Intelligence, 2021, 19(2): 195-217.

4. Zhang H, Xie Y, Zheng L, et al. Interpreting multivariate shapley interactions in dnns[C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2021, 35(12): 10877-10886.

5. Liu F T, Ting K M, Zhou Z H. Isolation forest[C]//2008 Eighth IEEE International Conference on Data Mining. IEEE, 2008: 413-422.

6. Adadi A, Berrada M. Peeking inside the black-box: A survey on Explainable Artificial Intelligence (XAI)[J]. IEEE Access, 2018, 6: 52138-52160.