Explainable AI for Classification using Probabilistic Logic Inference on Medical Dataset

Aymen Bashir
Department of Artificial Intelligence, CureMD
Lahore, Pakistan

Abstract— The goal of this paper is to implement the methodologies proposed in [1] on a dataset that has medical application. The explanation of classification decisions is most important for medical domain since one can can not apply AI based systems without the supervision of an expert. This is because we do not know why a certain decision is made and in medical domain, the risk can not be taken. Lack of understanding of why a certain decision has been made makes AI based systems unusable for medical purposes. Therefore, even if a system performs well, we can not deploy it without understanding what it is doing and how an error can be identified and corrected in a timely way. Explainable AI is the area of AI that addresses this issue by making the process of decision making in AI systems transparent. This paper aims to implement one of the methods proposed in [1] on a Show/NoShow Dataset [2].

Index Terms—Knowledge Bases, Probabilistic Logic Inference

1 Introduction

With the increase in AI systems, it has become important that we understand why a certain decision has been made. The goal of the paper is testing the methodologies proposed in [1] on a medical dataset and verify if the method is usable for a medical related dataset. The methodologies proposed work well with the limited number of features, therefore a medical image dataset could not be used due to lack of computational power. The Show/NoShow Dataset consists of features that help know why a patient misses his/her appointments. The features of the dataset include Gender, Age, Neighbourhood, Scholarship, Hypertension, Diabetes, Alcoholism, Handicap, SMS Received and others that will be discussed in detail further.

The considered paper proposes to construct a Knowledge Base (KB) that consists of probabibilties of all possible combinations of propositional variables variables. The paper has used the datasets having features that have been converted into Boolean. However, many real-life problems include features that are continuous. In continuous feature dataset, the proposed methodology explodes, and the methods become computationally expensive even for a classification problem.

2.1 Review Stage

The proposed methodology of classification and feature importance has been tested on datasets like Titanic, Mushroom, Nursery, HIV-1, UK parliament Bill and image dataset for vehicle classification. All these datasets are beginner level and the algorithms have been applied on image dataset after feature extraction that leaves 11 features to be analysed which greatly reduces the complexity of the tests. However, this is not the case in practical situations where input features can be very large in size especially in images and video datasets. The dataset of Show/NoSow consists of age and neighbourhood that consists of continuous values and had the most contribution in the output decision. However, making a Knowledge Base proved extremely computationally expensive and therefore, a complete knowledge base could not be formed due to limitations of resources.

2.2 Dataset Analysis

The Dataset of Show/NoShow is based on features including patient id. Appointment id, gender, scheduled day, appointment day, age, neighbourhood, scholarship and diseases and conditions including hypertension, diabetes, alcoholism, handicap, and the question that if the patient received the message or not. The dataset consists of various inconsistencies that had to be corrected before feeding it into a probabilistic model. Inconsistencies like incorrect spellings, unreasobale data like age of 1000 years, incorrect datatype were corrected.

After the corrections, data analysis was performed to see which features are highly correlated with the output and which ones have no relationship and can be ignored. For this purpose, a futhur analysis was done where features like date were converted into day, month and year and correlation was found to be inadequate to be considered for model training. Also, data like patient id and appointment id were assumed to be irrelevant since they were a random combination of unique numbers for each patient and did not corelate with the outcome in any case.
Also, datatypes had to be corrected.

For further analysis, number of unique values that a feature can take were found and their frequency in the dataset was analyzed. Binary values were found in gender, scholarship, hypertension, alcoholism, and if the patient has received the SMS or not. Age took continuous values and features like handicap and neighbourhood took a wide range of nominal values.

Label Encoders were used to to encode possible values of the feature from 0 to n. e.g., male and female were encoded as 0 and 1. The dataset was split into training and test set and decision trees were used to do the classification. The decision tree classifier from sklearn.tree was used which also provided insight about the most important features that contribute in the decision making. It turned out that gender and age played the most important role with 37 % contribution from each feature. Neighbourhood played 6% role.

.

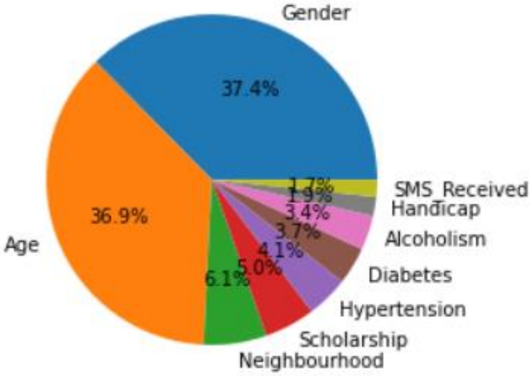

Figure 1. Importance of Features from Sklearn

The model achieved a test accuracy of 75.9%.

2.3 Knowledge Base Construction

The next step involved constructing a Knowledge Base. Algorithm 3 from [1] was implemented. A power set was constructed using chain and combinations from itertools. However, the combinations for all data entries exceeded the limit of the machine that was available. This caused us to only make a protype knowledge base with a very few data entries and few f eatures. But important features like age and gender were not skipped.

A knowledge base was constructed using Direct approach by applying the following algorithm:

Algorithm

Procedure Direct KB
Keys = []
Counts =[]
Labels = []
KB =[]

for each data entry in dataset:
feature values = {} (a:{v}| a feature has value v)
label = label of the data entry
S = powerset (feature values)
For each entry in S:
Check if entry is in keys:
Get index i of key.
Counts[i] = Counts[i]+1
Label[i] = Label[i]+label
Else:
Append new key.
Counts[i] = Counts[i]+1
Label[i] = Label[i]+label
For index in keys:
r = label[i]/counts[i]
KB.append.
KnowledgeBase = dataframe[keys, KB]
Return KnowledgeBase

The next step required to construct an optimization model which is able to tell probability of particular feature based on the Knowledge Base. For this purpose, optimization method was used in which an incoming query is compared with the knowledge bases by finding the probability with the most similarity with the query feature values and then a probability is assigned to the query. If the probability is greater than 0.5, the query is considered positive which in our means not missing an appointment.

Algorithm:

```
For keys in KnowledgeBase:
Diff = KB[key] -query[key]
    If Diff >threshold:
        Found_key = last_key
    Else:
        Found_key = key
P = probability[key]
If  P > 0.5
    Query['No_Show']=1
Else:
    Query['No_Show']=0
```

The algorithm needs further developments with respect to optimization and utilization of all the available data but due to contstraints of time and low computational power of available machine, it was not possible to completely implement the methodologies in [1].

Conclusion:

The implementation of the algorithm to generate-Knowledge bases directly revealed some shortcomings of the algorithm. First, it takes a large number of clauses for continuous data since combinations can exceed the working capacity of the machine which also costs more in terms of time. However, the algorithm might perform efficiently on

7.3 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,h*

ttps://doi.org/10.1016/j.inffus.2021.01.008.(https://www.sciencedirect.com/science/article/pii/S1566253521000142)

[3] Hoffman, R. R., Mueller, S. T., Klein, G., & Litman, J. (2018). Metrics for explainable AI: Challenges and prospects. arXiv:1812.04608.

[4] Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. Science, 349, 255–260.

[5] Montavon, G., Samek, W., & Müller, K.-R. (2017). Methods for interpreting and understanding deep neural networks. arXiv:1706.07979.

[6] Shortliffe, E. H., & Buchanan, B. G. (1975). A model of inexact reasoning in medicine. Mathematical Biosciences, 23, 351–379.

[7] Tenenbaum, J. B., Kemp, C., Griffiths, T. L., & Goodman, N. D. (2011). How to grow a mind: Statistics, structure, and abstraction. Science, 331, 1279–1285.