Interpretable Machine Learning to Understand Participant Evolution in Longitudinal Cohort Study Data

Ashwini Gonibedu Dathatri
M_Nr: 223687
Chethan Mysuru Radhakrishna
M_Nr: 225919
Manish Bhandari Vipinraj Bhandari
M_Nr: 226011
Sinchana Ramakanth Bhat
M_Nr: 223738

1 Motivation

A general goal of medical research on epidemiological data is to identify risk factors for diseases and to contribute thus to prevention and diagnosis. Epidemiological scholars work mostly in a hypothesis driven way where there is a greater emphasis on the quality of previous knowledge which may sometimes be misleading. The results of these epidemiological studies can be enhanced by applying mining methods to identify factors that could modulate the outcome. In this project, we have access to a dataset from a large German population health study that contains information on medication, habits and consumption, anthropometric measurements, socio-demographics, laboratory values, genetic markers and much more. One special property of this study is that it is longitudinal, so for each cohort participant, we have data from the baseline examinations and two follow-ups that are five years apart. There are few applications of data mining on these kinds of data. However, there are very few studies on exploiting inherent temporal information. In previous work, a framework was built for modeling and extracting "the evolution" of cohort participants in longitudinal data. It has been shown that the evolution features can improve classification accuracy. In this project, our primary aim is to extend this framework by designing and implementing a visual exploration tool for the evolution features.

2 Introduction

The goal of science is to extract knowledge from information. And at the same time to explain the decisions made by the model. "If you can't explain it simply, you don't understand it well enough". Hence we are proceeding with Interpretable Machine Learning methods which explains the decisions on why the model behaves the way it behaves! Model-Specific and Model-Agnostic methods are the two types of methodologies which can be followed to interpret the models and give explanations.

3 Literature Survey

- Paper: Can we classify the participants of a longitudinal epidemiological study from their previous evolution?[3]

  This paper stands as a base to understand how the model is built and how evolutionary features are exploited using the data mining approach for which we have to apply the interpretable methods to justify the decisions.

- Book: Molnar, Christoph's "Interpretable machine learning. A Guide for Making Black Box Models Explainable"[1]

  The book clearly gave an idea to approach the problem with interesting interpretable ideas. The book clearly explains the importance of interpretability, the models and methods for the approach, evaluation and human explanations for the same.

- Paper: Exploring factors associated with pressure ulcers: A data mining approach[4]

  The paper gives a major justification as to why random forests are very handy in predictive modelling which is a base of the approach we have proposed.

- Paper: Cohort-Profile-The-Study-of-Health-in-Pomerania[5]

  We are dealing with the SHIP data, paper clearly gives the insights on how data is collected, who are in the study, data quality, time frames etc.

4 Research Questions

- What is the merit of an evolution feature towards classification accuracy improvement?
- What is the minimal set of evolution features that would result in better predictive performance?
- What is the effect of evolution feature on predicted class?
- What is the minimal change in the participant such that the predicted class label changes?

5 Subtasks

Design and implementation of a visual exploration tool which would include features to :

- Evaluate the merit of an evolution feature towards classification accuracy.
- Analyse the minimal change in the participant such that the predicted class label changes.
- Measure the influence of each evolution feature on the predicted class.
- Identify the minimal set of evolution features that would result in better predictive performance.

6 Expected results

- Visualisations that offer well defined interpretation of the influence of evolution features on classification accuracy.
- A set of evolution features that can be used as prime factors which play a major role in the prevention and diagnosis of diseases.

7 Methods

7.1 Task1: What is the merit of an evolution feature towards classification accuracy improvement?

Proposed Method: Rule Fit
The goal of our task is to measure the importance of each evolution feature towards classification accuracy. In other words, to measure the amount of variance in the target outcome that can be explained by the variance in the particular feature values across the dataset. One possible way to achieve this would be to fit a Random Forest to the dataset. Then go through all the splits produced by the feature and measure the amount of variance reduced compared to the parent node. Random forests outperform many advanced machine learning algorithms including Neural Networks (NNs) and Support Vector Machines (SVMs), it still produces a "black box" model which is difficult to interpret. Also, it does not efficiently explain the correlations among different features in the dataset.

Initial observations of the dataset and the features showed that the following types of relationships exist between the features and the target outcomes:

- Linear Relations: Some of the evolution features exhibit linear relationship with the target outcome like A_Diff_f_t_t [3] which measures the change in the original feature values over each time step.
- Feature Interactions: Evolution feature interactions such as Cluster_t [3] and dist_to_centroid_t [3] is highly correlated since the former represents the membership of a cohort participant data point in the cluster and the latter measures the distance of the data point from the centroid.

Rule Fit approach provides a good explanation of the relationships mentioned above. Here we extract the minimal set of frequent decision rules from a random forest to interpret the feature correlations. At the same time, we use Least Absolute Shrinkage and Selection Operator (LASSO) to fit a sparse linear model to extract a minimal set of evolution features that are linearly affecting the target outcome.

This approach is easily interpretable and also, explains the variance in the target outcome by effectively explaining the linear variations in the feature values and the effect of feature interactions on the model outcome.

This approach is a model specific approach which lets us take advantage of the behaviour of the model while interpreting it. However, this can be easily combined with model – agnostic methods to interpret the machine learning model locally (at the level of individual participant variations).

7.2 Task 2: What is the minimal set of evolution features that would result in better predictive performance?

Proposed Method: Rule Fit
The rule fit approach extracts frequent decision rules and the features that affects the target output linearly. These extractions accumulates to a minimal set of such evolution features that efficiently explain the variance in the target outcome. In other words, these are the most important features and feature interactions, that would have an impact on improving the predictive performance of the model.

7.2.1 Extracting minimal set of decision rules:

To achieve this, we first train a random forest to fit the dataset which includes both original features and the evolutionary features. As shown in figure 1, we extract the if-then rules like $(x2 < 3) \quad and \quad (x5 < 7)$, generated by each path from root-to-leaf node of the resulting decision tree. We then add them as decision rules to the set of mined association rules. Here we discard the predictions available at the leaf nodes since we are only interested in the rules to be extracted.

A coverage metric including support and confidence levels are used to measure the frequency of the extracted rules. A fine-tuned threshold can be set on the support levels of these rules and a minimal set can be extracted.

7.2.2 Extract minimal set of linear features:

To achieve this we train a sparse linear model that uses aforementioned Least Absolute Shrinkage and Selection Operator (LASSO). On each iteration the weights learned for each feature in the dataset are optimized by penalizing large weights as shown in below equation.

The term $||\beta||_1$, is used to add the penalty to the weights. As the learning progresses, the features that are not linearly influencing the predictive perfor-

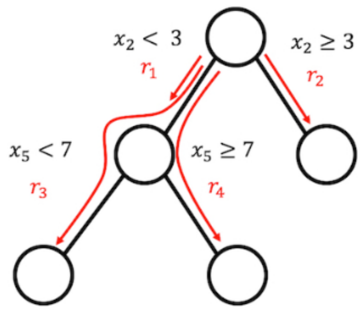

Figure 1: Rule Extraction[1]

$$min_\beta \left(\frac{1}{n} \sum_{i=1}^{n} (y^{(i)} - x_i^T \beta)^2 + \lambda ||\beta||_1 \right)$$

Figure 2: Minimize weights[1]

mance of the model get 0 weights. The term λ, is used as threshold to select the minimal set of linear features. This threshold can be fine-tuned by performing cross validations.

7.3 Task 3: What is the effect of evolution feature on predicted class?

Proposed Method: Individual Conditional Expectation (ICE) Plot
There exists an alternative approach to interpretability in machine learning which can provide explanations to these kind of counter factual statements as defined in our task 3. Those alternative approaches are termed as model-agnostic techniques where the model is treated as a black box, providing model flexibility.

In order to measure the effect of each evolution feature from the minimal set over the predicted class, we are using one such model-agnostic approach called Individual Conditional Expectation (ICE) plot.

An example of the plot is given below:

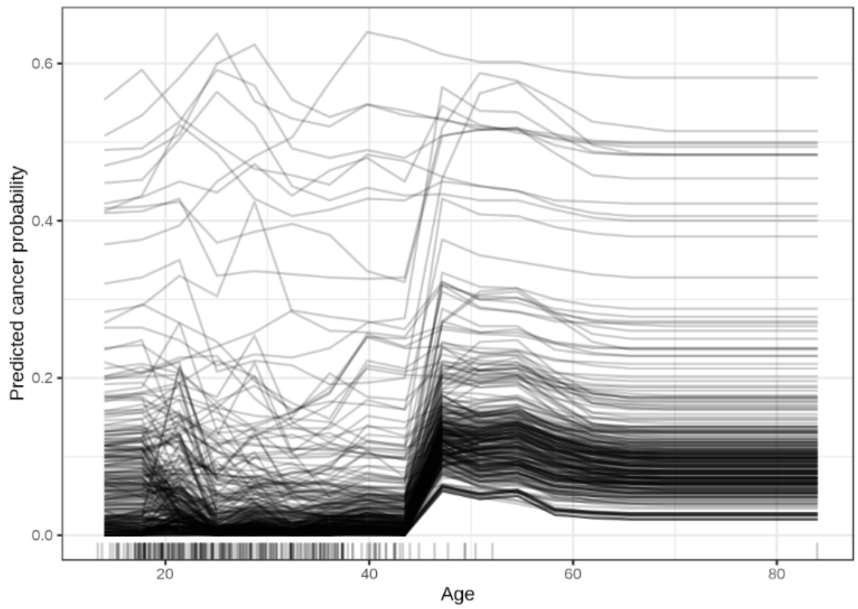

Figure 3: Example of ICE plot[1]

In this plot, each line indicates one instance which depicts the instance's prediction changes(cancer probability) over variations in a single feature(Age) belonging to the minimal set of features extracted during task2. In this case, the values of all other features remain constant. As mentioned above, we would use Rule fit as our black box model to predict the value for the variants of an instance for a feature.

Motive to choose ICE plots are:

- Helps us uncover heterogeneous effects of the features on the prediction – provide insights regarding the correlations between the feature under interest and the remaining features of an instance.
- More intuitive to understand
- Considering ICE plots also help us to accomplish Task 4 as described below.

7.4 Task 4: What is the minimal change in the participant such that the predicted class label changes?

Proposed Method: ICE Plot and Robustness Computation
In this task, we try to identify the minimal changes in the participant such that

the predicted class label changes. Identifying these changes are significant for participants who have been diagnosed positive so that they can understand the changes which they have to undergo so that the possibility that they are further diagnosed as negative increases. We try to identify these changes with respect to the minimal features identified from the previous task since they have the maximum impact on the class label prediction.

An approach for this is to combine the Individual Conditional Expectation (ICE) plot with the method of robustness computation[2]. In this method, we try to identify the minimal change in the participant by finding out the point of least robustness with respect to predicted class label. To do this we first set a threshold at which the change in class label takes place and then we perform small variations($\pm\Delta x$) at each point along the x-axis to identify the point where these small variation results in large change in the predicted outcome. The point at which the maximum change occurs for this small change is the point of least robustness and at this point along with the small change (Δx) gives us the required minimal change of an individual with respect to the feature value.

8 Timeplan — Roles — Responsibilities

Roles and responsibilities regarding:

- Implementation of ICE and determining minimal change using robustness are handled by Ashwini.
- Pre-processing and evaluation are handled by Chethan.
- Implementation of Rule Fit and documentation are handled by Manish.
- Exploratory data analysis and visualization are handled by Sinchana.

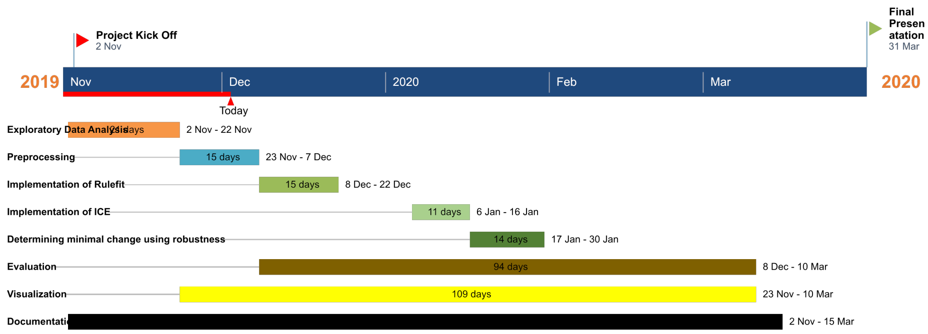

Figure 4: Timeline, Roles and Responsibilities

References

[1] Christoph Molnar. *Interpretable Machine Learning*. 2019. `https://christophm.github.io/interpretable-ml-book/`.

[2] S. Mostaghim. Chapter 6: Robust optimization and constraint handling. `http://www.is.ovgu.de/Teaching/SS+2019/Evolutionary+Multi_Objective+Optimization-p-674.html`, 2019.

[3] Uli Niemann, Tommy Hielscher, Myra Spiliopoulou, Henry Volzke, and Jens-Peter Kuhn. Can we classify the participants of a longitudinal epidemiological study from their previous evolution? In *Proceedings of the 2015 IEEE 28th International Symposium on Computer-Based Medical Systems*, CBMS '15, pages 121–126, Washington, DC, USA, 2015. IEEE Computer Society.

[4] Dheeraj Raju, Xiaogang Su, Patricia A. Patrician, Lori A. Loan, and Mary S. McCarthy. Exploring factors associated with pressure ulcers: A data mining approach. *International Journal of Nursing Studies*, 52(1):102 – 111, 2015.

[5] Henry Völzke, Dietrich Alte, Carsten Oliver Schmidt, Dörte Radke, Roberto Lorbeer, Nele Friedrich, Nicole Aumann, Katharina Lau, Michael Piontek, Gabriele Born, Christoph Havemann, Till Ittermann, Sabine Schipf, Robin Haring, Sebastian E Baumeister, Henri Wallaschofski, Matthias Nauck, Stephanie Frick, Andreas Arnold, Michael Jünger, Julia Mayerle, Matthias Kraft, Markus M Lerch, Marcus Dörr, Thorsten Reffelmann, Klaus Empen, Stephan B Felix, Anne Obst, Beate Koch, Sven Gläser, Ralf Ewert, Ingo Fietze, Thomas Penzel, Martina Dören, Wolfgang Rathmann, Johannes Haert-

ing, Mario Hannemann, Jürgen Röpcke, Ulf Schminke, Clemens Jürgens, Frank Tost, Rainer Rettig, Jan A Kors, Saskia Ungerer, Katrin Hegenscheid, Jens-Peter Kühn, Julia Kühn, Norbert Hosten, Ralf Puls, Jörg Henke, Oliver Gloger, Alexander Teumer, Georg Homuth, Uwe Völker, Christian Schwahn, Birte Holtfreter, Ines Polzer, Thomas Kohlmann, Hans J Grabe, Dieter Rosskopf, Heyo K Kroemer, Thomas Kocher, Reiner Biffar, Ulrich John, and Wolfgang Hoffmann. Cohort Profile: The Study of Health in Pomerania. *International Journal of Epidemiology*, 40(2):294–307, 02 2010.