Association Rule Mining for Error Detection in Predictive Models

Alon Harari, Idan Lazar

Abstract

Machine learning models can make systematic errors that follow specific patterns within the data. Error analysis methods typically focus on individual features and simple statistics, potentially missing complex feature interactions that lead to prediction errors. We tried to use apriori to create rules based on errors in data to get a better explanation on why certain errors happen and secondly, try to use these rules to recognize "problematic" rows, and change their prediction. Experimental results show that our approach can identify error patterns but the performance is still lacking.

Problem Description

Classic error analysis sometime lack in intuitive explanation why some errors occurred, it mainly focuses on individual features and less on complex interactions, it can be hard to understand exactly what should be done to improve the model based on the classic analysis, and the analyst based on the error should create manual insights and manually improve the model based on the insights.

Solution Overview

Our solution takes the model errors and runs the apriori algorithm over them. Given the rules we show the top 10 most relevant rules(using confidence and lift) that could explain the error. These rules are represented. After that, we take these rules and try to implement them on the existing model to try and fix the wrong predictions.

The full solution using these steps:

1. First, we take an existing model(in our case xgBoost) and using it to predict the train data. then we find all the rows that the model made a mistake in the prediction.

2. Next, we format the data to be suitable for apriori algorithm, we are discretizing numerical features and formatting categorical features.

3. Now, the error data is ready for the rule mining. We run the apriori algorithm and we generate with it the rules, we only take rules with a minimum 0.1 support. Later we take the top 10 rules and show them to the analyst.

4. Next, we take the already trained model, but adding to it another layer, after the original model prediction we find rows that correspond to the rules, and we fix the prediction accordingly.

The solution gives us 2 things, it gives us new error analysis based on association mining rules, but also uses them in practice to make the model better.

Experimental Evaluation

To evaluate our approach, we conducted experiments on four different datasets, adult income dataset, heart disease dataset, breast cancer dataset, email spam, credit card fraud.

We used xgboost as a baseline model to run our experiment on. We checked the metrics before implementing the error rules and after. Metrics details:

Rule Quality Metrics:

Support: The proportion of error cases covered by the rule

Confidence: The probability that the rule correctly identifies an error

Lift: How much more likely the rule is to indicate an error compared to random chance

Error Coverage: The percentage of model errors that can be explained by the discovered rules

Model Performance Metrics:

Accuracy: The proportion of correct predictions

F1 Score: The harmonic mean of precision and recall

ROC AUC: Area under the ROC curve (for binary classification)

Results

Need to fill here

Related Work

Our inspiration came from the fact that there are not much studies on the topic of improving model with association rules. But we did find some research that tackle the topic in few different angles.

1. **Rule Mining for Correcting Classification Models** (Das et al., 2022): This work did something very similar to our approach. The difference is that they created their own rules generator by using smart pruning technique while we used directly apriori algorithm, their approach discovers rules of the form X → δ, where X represents a condition that identifies a subpopulation and δ is an optimized correction amount to be added to the model's prediction score for instances in that subpopulation. While we find rules of X -> error where X is a subpopulation and error is the classification of the model that he got wrong, we then find these subpopulations and change their classification. We got inspiration on how to implement the rules mechanism from here.
2. **BoostClean: Automated Error Detection and Repair for Machine Learning** (Krishnan et al., 2017): BoostClean is a system that automates error detection and repair in machine learning pipelines. BoostClean uses statistical boosting to select an ensemble of error detection and repair combinations that maximize model accuracy. BoostClean uses boosting to combine different cleaning strategies, while my solution uses association rule mining to identify specific patterns in errors**.** My solution focus on rules which are more interpretable, Boostclean is more complex. This paper gave us the idea to improve the model automatically based on rules.

In general my approach is more straightforward then the other papers, and give us simple rules to explain errors in the data.

Conclusion

In this research, we have developed and evaluated a framework for automatically discovering and explaining patterns in predictive model errors using association rule mining. Our experiments confirm that model errors can follow patterns that can be captured through association rules. These patterns frequently involve interactions between multiple features that might be missed by traditional error analysis methods. Association rules can be used for explainability in machine learning. We thought that the solution will be more straightforward but we discovered that using these rules to improve the model can be challenging, also we believe that with more research and work we can use the rules better by creating features based on rules.

References

1. Das, A., Du, X., Wang, B., Shi, K., Gu, J., Porter, T., & Cardie, C. (2022). Rule Mining for Correcting Classification Models. arXiv:2201.05017v3 [cs.CL].
2. Krishnan, S., Franklin, M. J., Goldberg, K., & Wu, E. (2017). BoostClean: Automated Error Detection and Repair for Machine Learning. arXiv:1711.01299v1 [cs.DB].
3. Ribeiro, M. T., Singh, S., & Guestrin, C. (2018). Anchors: High-Precision Model-Agnostic Explanations. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 32, No. 1).