

Machine Learning for Predicting Constraints

An Initial Study using an Automotive Industry Dataset: 99.998%

05-August-2022

Technical Report: Kedge-22-001

Contacts: www.kedge360.com

Kedge Systems Private Limited

Summary

- Constraint Systems/Solvers¹ have been widely used in various intelligent systems
- This study evaluated the use of ML in predicting “unknown” constraint systems, by learning from a **small** set of labelled sample data
- This has many applications like error diagnostics, failure prediction etc.,
- This study focused on an industrial *automotive* data and *quickly* got up to **99.998%** prediction accuracy

¹https://en.wikipedia.org/wiki/Constraint_programming

Background

- In a typical Constraint System:
 - First, rules over attributes will be modeled
 - Then, constraint solvers are typically used to find valid/matching solution(s), under given conditions
- In the scenarios where the system cannot be modeled for various reasons:

Can we use ML to learn the model from sample solutions, and then predict validity of future solutions?

- This study took an industrial data from automotive domain² and performed an analysis

²<https://www.sciencedirect.com/science/article/pii/S000437020100162X>

Experimental Setup

- **Reference constraint system:** Automotive data² having:
 - ~1.4 Trillion valid solutions
 - >100 Trillion invalid solutions
- **Synthetic dataset:** For evaluation purpose, a custom solution generator tool was built to randomly create solutions for the system with valid/invalid labels
- **Test:** Using a (20%, 10%, 4%, 2%, 1%, 0.2%) part of the synthetic-dataset, train a ML model and check its accuracy on remaining part (80%, 90%, 96%, 98%, 99%, 99.8%) of the synthetic-dataset

²<https://www.sciencedirect.com/science/article/pii/S000437020100162X>

Dataset: Train/Test/IAT

- **Train/Test:**

- By varying few parameters in the synthetic solution generator tool, a variety of a well-balanced (50%/50%) set of five million solutions was created for Train/Test purpose

- **Independent Accuracy Test (IAT) data:**

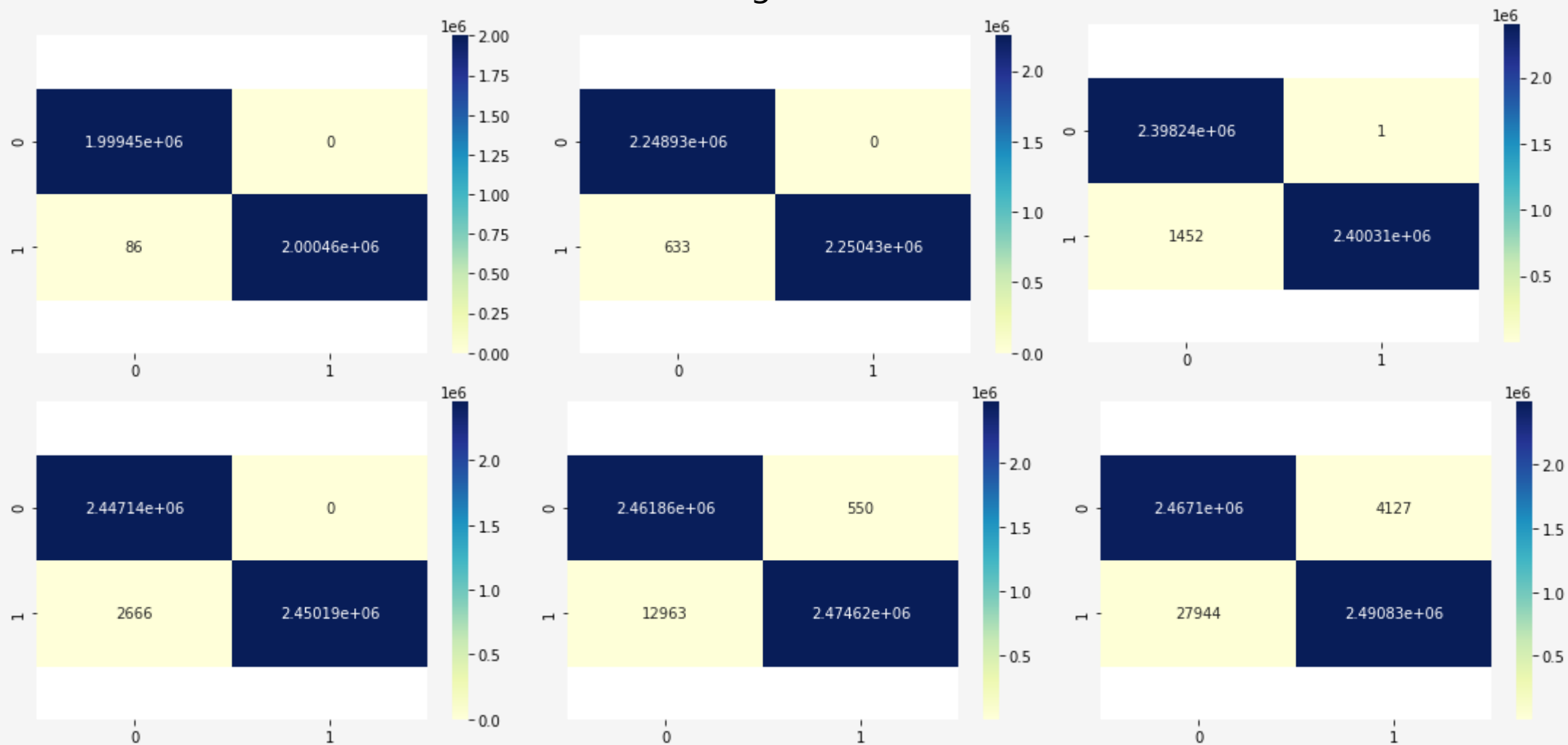
- With some changes to the solution generator tool, another variety of five million solutions were created for doing an independent test of the model accuracy.
- Some changes were deliberately introduced to vary the type of random valid/invalid solutions being generated.
- This set had a rough 47.7% / 52.3% balance between valid/invalid labels.

Test Results

- A simple Deep Neural Network model was quickly built for this evaluation
- Test
 - For the Train/Test dataset of 5 million:
 - By varying the Train data set size from 1 Million (20%) to 10k (0.2%), and using remaining solutions for test, test accuracy was in the range of **99.998% to 99.357%**
 - On IAT dataset:
 - The models built using previous set of data for 20%-train to 0.2%-train cases, were used to predict labels of 5 million IAT solutions, and the prediction accuracy was in the range of **99.996% to 99.03%**
 - **So, for the 0.2% Model, built using 10,000 solutions out of > 100 Trillion solutions in the Automotive data, accuracy of 99.03% was obtained even though the IAT dataset was created by using a different variety of data generation tool.**
 - **For the 20% Model, built using 1 Million solutions out of > 100 Trillion solutions in the Automotive data, accuracy of 99.996% was obtained for the IAT dataset**

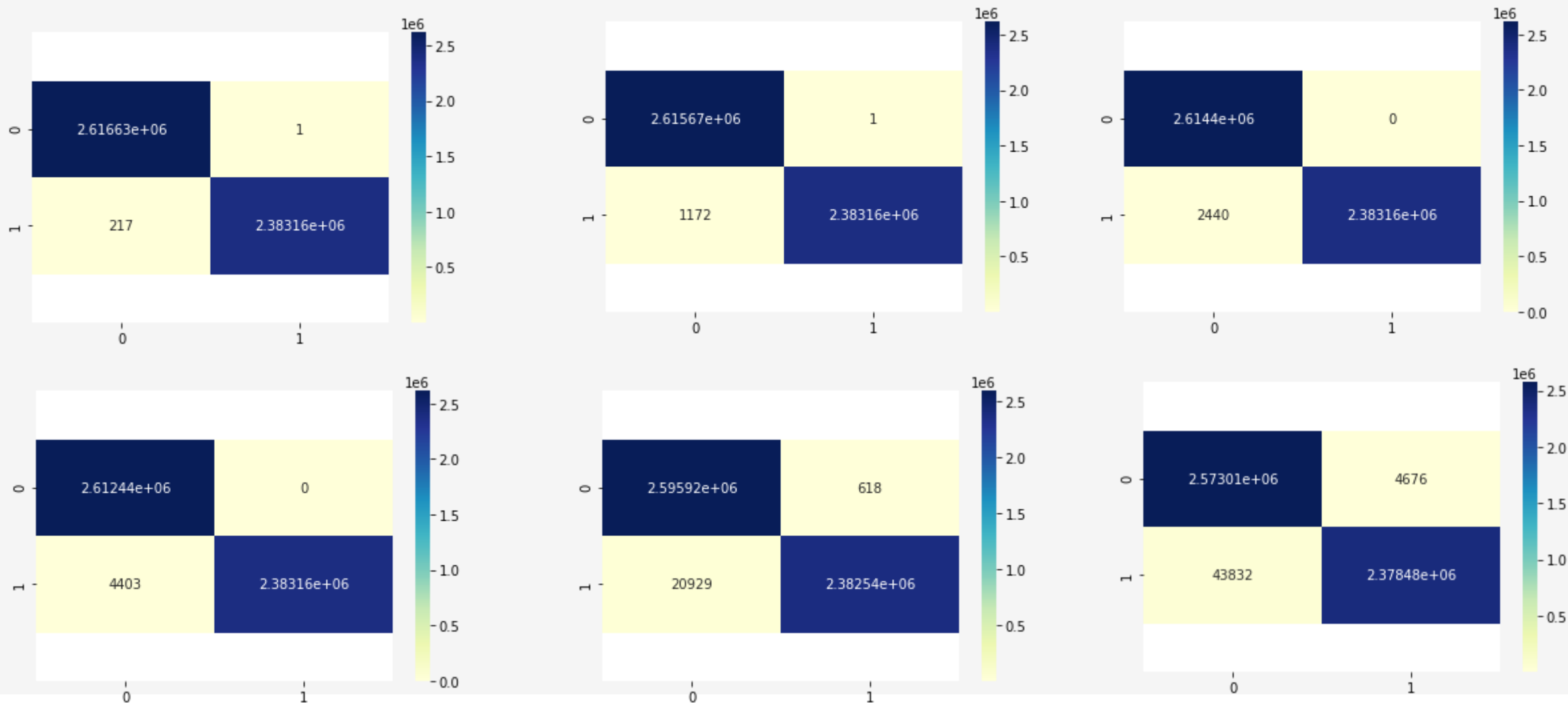
Confusion Matrices for Train/test dataset

Prediction of Test dataset with Models built using 20%, 10%, 4%, 2%, 1%, 0.2% of train/test dataset



Confusion Matrices for IAT dataset

Prediction of IAT dataset with Models built using 20%, 10%, 4%, 2%, 1%, 0.2% of train/test dataset



Conclusion

- This initial study done over a short period of time got a very good prediction accuracy for the used constraint system
- The quick initial study results are very exciting with accuracy up to 99.998%
- More such experiments with a variety of reference constraint systems needs to be done
- Future work includes application of this technique to fault diagnosis, failure prediction, etc., (where reference system might not always be available, or even when available might not be accurate)