Improving Algorithms of Criminal Risk Assessment

Rachel Choi

Introduction

Artificial intelligence is a wide-ranging branch of computer science in which the focus is on building smart machines that are capable of performing tasks that typically require human intelligence. As advancements in machine learning and deep learning are creating a paradigm shift in almost every sector of the tech industry, artificial intelligence is creating a huge influence on the way we live, work, travel, and do business in the 21st century.

The influence of AI technology can also be seen in the criminal justice industry. Artificial Intelligence is widely used throughout the criminal justice system in the United States. The most commonly used are pretrial risk assessment algorithms, also called as risk assessment tools, which are designed to predict a defendant's future risk for reoffending. They influence judgments about guilt or innocence, bail, and sentencing. However, the algorithms are largely hidden from public view, and many critics have raised concerns that the results may be a source of bias (Angwin et al., 2016).

For my capstone research, I will explore the fragility of risk assessment tool in the US justice system by exploring the assessment algorithm and show how it is biased to raise awareness of possible unfair sentencing.

Background

Risk assessment instruments (RAIs) are designed to predict a defendant’s future risk for misconduct. These predictions inform high-stakes judicial decisions, such as whether to incarcerate an individual before their trial. For example, an RAI called the Public Safety

Assessment (PSA) considers an individual's age and history of misconduct, along with other factors, to produce three different risk scores: the risk that they will be convicted for any new crime, the risk that they will be convicted for a new violent crime, and the risk that they will fail to appear in court. A decision-making algorithm translates these risk scores into release-condition recommendations, with higher risk scores corresponding to stricter release conditions. RAIs influence a wide variety of judicial decisions, including sentencing decisions and probation and parole requirements (Chohlas-Wood, 2020).

Risk assessment tools are used in almost every state in the U.S. They are usually used in pre-trial, although they exist at sentencing, in prison management, and for parole determinations. There are also specific risk assessment tools for different functions in the criminal justice system, such as domestic violence risk or juvenile justice risks, with the understanding that different factors are used in those contexts than in a general criminal risk or violent criminal risk of rearrest or re-offense (Chohlas-Wood, 2020). As criminal justice algorithms have come into greater use at the federal and state levels, they have also come under greater scrutiny. Many criminal justice experts have denounced risk assessment tools as opaque, unreliable, and unconstitutional.

Literature Review

Another scholarly research that has been done in regards to the risk assessment tool is an analysis of a tool made by Northpointe, Inc., COMPAS (Correctional Offender Management Profiling for Alternative Sanctions). Northpointe created risk scales for general and violent recidivism and for pretrial misconduct. According to the COMPAS Practitioner's Guide, the scales were designed using behavioral and psychological constructs of very high relevance to recidivism and criminal careers. COMPAS has been used by the U.S. states of New York, Wisconsin, California, Florida's Broward County, and other jurisdictions (Larson et al., 2016). The research group conducted an analysis on COMPAS and found that black defendants were far more likely than white defendants to be incorrectly judged to be at a higher risk of recidivism, while white defendants were more likely than black defendants to be incorrectly flagged as low risk. The study was done on more than 10,000 criminal defendants in Broward County, Florida. The research group compared the recidivism risk categories predicted by the COMPAS tool to the actual recidivism rates of defendants in the two years after they were scored, and found that the score correctly predicted an offender's recidivism 61 percent of the time, but was only correct in its predictions of violent recidivism 20 percent of the time. In forecasting who would re-offend, the algorithm correctly predicted recidivism for black and white defendants at roughly the same rate (59 percent for white defendants and 63 percent for black defendants) but made mistakes in very different ways. It misclassified the white and black defendants differently when examined over a two-year follow-up period. It was found that black defendants were often predicted to be at a higher risk of recidivism than they actually were. Black defendants who did not recidivate over a two-year period were nearly twice as likely to be misclassified as higher risk compared to their white counterparts (45 percent vs. 23 percent). Black defendants were also twice as likely as white defendants to be misclassified as being a higher risk of violent recidivism; White violent recidivists were 63 percent more likely to have been misclassified as a low risk of violent recidivism, compared with black violent recidivists.

Research

For my capstone research, I decided to investigate on current machine learning algorithms of the risk tools that are used in the American Justice system and identify features where bias can be introduced. Thus, the goal of this research is to discover the underlying accuracy of the recidivism algorithms. For the research, I used data that is two years worth of COMPAS scores from the Broward County Sheriff's Office in Florida. COMPAS algorithm is one of the most popular scores used nationwide and is increasingly being used in pretrial and sentencing.

Random Forest

To investigate on how the COMPAS scores were calculated and find important attributes of the scores, I used Random Forest on the dataset. Random forest is a supervised learning algorithm that builds multiple decision trees and merges them together to get a more accurate and stable prediction. To use Random forest, I first preprocessed the dataset by removing attributes that are irrelevant. Then I split the dataset into training set and test set, 70% training and 30% test. Then I changed the categorical features to numerical values by using OneHotEncoder and Pipeline. With these newly processed test and training sets, I tried to run the pipeline, but I was running into dimension issues.

Here is my work:
https://colab.research.google.com/drive/1vIRhqbWviCfgjsqJTP5mcx8JBAwsM0tU#scrollTo=MjcVSJSP6xha

Association Mining

The next algorithm that I used to explore risk assessment tools was Association Mining. Association rule mining is a procedure which aims to observe frequently occurring patterns, correlations, or associations from datasets found in various kinds of databases such as relational databases, transactional databases, and other forms of repositories. I chose to use this algorithm to see patterns in the dataset and see if there is any correlation between the attributes. To use Association Mining, I first preprocessed the dataset by removing attributes that are irrelevant.

Then I split the dataset into categorial and numerical data and converted the categorial attributes to numerical attributes as the apriori library requires the dataset to numerical. However, I was running into issues with the dataset conversion. Having done this, we realized association mining requires vectors of 0s and 1s, where 0 means that the data is not there and 1 means that the data is there. We realized that our data is not organized the same way, so we decided to move on.

Here is my work:

https://colab.research.google.com/drive/162X_7L5dp2VBpr-kiIXtZdZgsjX8pqFd#scrollTo=VopjAZbGRbZr

Weka

I also used Weka to explore the datasets. The first algorithm I tried was Random Forest. Here are the results.

```
=== Run information ===

Scheme:       weka.classifiers.trees.RandomForest -P 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1
Relation:     preprocessed-data-small
Instances:    4761
Attributes:   24
              sex
              age
              race
              juv_fel_count
              1_decile_score
              juv_misd_count
              juv_other_count
              2_priors_count
              days_b_screening_arrest
              c_days_from_compas
              c_charge_degree
              c_charge_desc
              is_recid
              r_days_from_arrest
              r_charge_desc
              is_violent_recid
              vr_charge_desc
              decile_score
              v_decile_score
              priors_count
              start
              end
              event
              two_year_recid
Test mode:    10-fold cross-validation

=== Classifier model (full training set) ===
```


```
=== Classifier model (full training set) ===
```


```
RandomForest

Bagging with 100 iterations and base learner

weka.classifiers.trees.RandomTree -K 0 -M 1.0 -V 0.001 -S 1 -do-not-check-capabilities

Time taken to build model: 2.4 seconds

=== Cross-validation ===
=== Summary ===

Correlation coefficient                  0.9751
Mean absolute error                      0.0626
Root mean squared error                  0.1173
Relative absolute error                 12.6633 %
Root relative squared error             23.5962 %
Total Number of Instances             4761
```

I also tried Association Mining. Here are the results.

```
=== Run information ===

Scheme:       weka.classifiers.trees.RandomTree -K 0 -M 1.0 -V 0.001 -S 1
Relation:     preprocessed-data-small
Instances:    4761
Attributes:   24
              sex
              age
              race
              juv_fel_count
              1_decile_score
              juv_misd_count
              juv_other_count
              2_priors_count
              days_b_screening_arrest
              c_days_from_compas
              c_charge_degree
              c_charge_desc
              is_recid
              r_days_from_arrest
              r_charge_desc
              is_violent_recid
              vr_charge_desc
              decile_score
              v_decile_score
              priors_count
              start
              end
              event
              two_year_recid
Test mode:    10-fold cross-validation

=== Classifier model (full training set) ===
```

```
RandomTree
==========

is_recid < 0.5 : 0 (2497/0)
is_recid >= 0.5
|   event < 0.5
|   |   c_charge_desc = Aggravated Assault w/Firearm : 0.37 (1.73/0.01)
|   |   c_charge_desc = Felony Battery w/Prior Convict : 0.95 (5.18/-0)
|   |   c_charge_desc = Possession of Cocaine
|   |   |   decile_score < 1.5 : 0.08 (1.09/0.08)
|   |   |   decile_score >= 1.5
|   |   |   |   juv_fel_count < 0.5
|   |   |   |   |   2_priors_count < 3.5 : 1 (14.27/0)
|   |   |   |   |   2_priors_count >= 3.5
|   |   |   |   |   |   r_charge_desc = Felony Battery (Dom Strang) : 0 (0/0)
|   |   |   |   |   |   r_charge_desc = Driving Under The Influence : 1 (1/0)
|   |   |   |   |   |   r_charge_desc = Poss of Firearm by Convic Felo : 0 (0/0)
|   |   |   |   |   |   r_charge_desc = Battery : 1 (2/0)
|   |   |   |   |   |   r_charge_desc = Driving License Suspended : 1 (1/0)
|   |   |   |   |   |   r_charge_desc = Grand Theft (Motor Vehicle) : 0 (0/0)
|   |   |   |   |   |   r_charge_desc = Criminal Mischief>$200<$1000 : 0 (0/0)
|   |   |   |   |   |   r_charge_desc = Grand Theft in the 3rd Degree : 1 (1/0)
|   |   |   |   |   |   r_charge_desc = Possession of Cocaine
|   |   |   |   |   |   |   age < 43.5 : 0 (1/0)
|   |   |   |   |   |   |   age >= 43.5
|   |   |   |   |   |   |   |   1_decile_score < 7 : 0 (1/0)
|   |   |   |   |   |   |   |   1_decile_score >= 7 : 1 (1/0)
|   |   |   |   |   |   r_charge_desc = Poss Cocaine/Intent To Del/Sel : 0 (0/0)
|   |   |   |   |   |   r_charge_desc = Prowling/Loitering : 0 (0/0)
|   |   |   |   |   |   r_charge_desc = Operating W/O Valid License : 1 (2/0)
|   |   |   |   |   |   r_charge_desc = Possess Cannabis/20 Grams Or Less : 0 (1/0)
|   |   |   |   |   |   r_charge_desc = False Imprisonment : 0 (0/0)
```

```
|   |   |   |   |   |   r_charge_desc = Resist/Obstruct W/O Violence : 0 (0/0)
|   |   |   |   |   |   r_charge_desc = Possession of Cannabis : 0 (0/0)
|   |   |   |   |   |   r_charge_desc = Unlaw Use False Name/Identity : 1 (1/0)
|   |   |   |   |   |   r_charge_desc = Viol Pretrial Release Dom Viol : 0 (0/0)
|   |   |   |   |   |   r_charge_desc = Viol Injunct Domestic Violence : 1 (1/0)
|   |   |   |   |   |   r_charge_desc = Agg Battery Grt/Bod/Harm : 0 (0/0)
|   |   |   |   |   |   r_charge_desc = DOC/Fighting/Threatening Words : 0 (0/0)
|   |   |   |   |   |   r_charge_desc = DWLS Canceled Disqul 1st Off : 0 (0/0)
|   |   |   |   |   |   r_charge_desc = Trespass Struct/Conveyance : 0 (0/0)
|   |   |   |   |   |   r_charge_desc = Petit Theft : 0 (0/0)
|   |   |   |   |   |   r_charge_desc = Susp Drivers Lic 1st Offense : 0 (0/0)
|   |   |   |   |   |   r_charge_desc = Aggrav Stalking After Injunctn : 0 (0/0)
|   |   |   |   |   |   r_charge_desc = Pos Cannabis W/Intent Sel/Del : 0 (0/0)
|   |   |   |   |   |   r_charge_desc = Felony Battery w/Prior Convict : 0 (0/0)
|   |   |   |   |   |   r_charge_desc = Robbery / No Weapon : 0 (0/0)
|   |   |   |   |   |   r_charge_desc = Leave Acc/Attend Veh/More $50 : 0 (0/0)
|   |   |   |   |   |   r_charge_desc = Possess Drug Paraphernalia : 0 (0/0)
|   |   |   |   |   |   r_charge_desc = Possession of Hydromorphone : 0 (0/0)
|   |   |   |   |   |   r_charge_desc = Viol Injunction Protect Dom Violence : 0 (0/0)
|   |   |   |   |   |   r_charge_desc = Poss Wep Conv Felon : 0 (0/0)
|   |   |   |   |   |   r_charge_desc = Aggravated Assault W/Dead Weap : 0 (0/0)
|   |   |   |   |   |   r_charge_desc = Uttering a Forged Instrument : 0 (0/0)
|   |   |   |   |   |   r_charge_desc = Burglary Unoccupied Dwelling : 0 (0/0)
|   |   |   |   |   |   r_charge_desc = Felony Driving While Lic Suspd : 0 (0/0)
|   |   |   |   |   |   r_charge_desc = Theft/To Deprive : 0 (0/0)
|   |   |   |   |   |   r_charge_desc = Expired DL More Than 6 Months : 0 (0/0)
|   |   |   |   |   |   r_charge_desc = Poss Pyrrolidinovalerophenone : 1 (1/0)
|   |   |   |   |   |   r_charge_desc = Unlaw LicTag/Sticker Attach : 0 (0/0)
|   |   |   |   |   |   r_charge_desc = Prostitution/Lewd Act Assignation : 0 (0/0)
|   |   |   |   |   |   r_charge_desc = Fraudulent Use of Credit Card : 0 (0/0)
|   |   |   |   |   |   r_charge_desc = Fail Register Vehicle : 0 (0/0)
|   |   |   |   |   |   r_charge_desc = Leaving the Scene of Accident : 0 (0/0)
|   |   |   |   |   |   r_charge_desc = Extradition/Defendants : 0 (0/0)
```

(…)

```
|   |   c_charge_desc = Cruelty to Animals : 0 (0/0)
|   |   c_charge_desc = Trespass Private Property : 0.98 (1.9/0)
|   |   c_charge_desc = Unauth C/P/S Sounds>1000/Audio : 0 (0/0)
|   |   c_charge_desc = Obstruct Officer W/Violence : 0 (0/0)
|   |   c_charge_desc = Cause Anoth Phone Ring Repeat : 0 (0/0)
|   |   c_charge_desc = Poss Unlaw Issue Id : 0.98 (1.9/0)
|   |   c_charge_desc = PL/Unlaw Use Credit Card : 0.98 (1.9/0)
|   |   c_charge_desc = Possession of LSD : 0 (0/0)
|   |   c_charge_desc = Tamper With Witness : 0 (0/0)
|   |   c_charge_desc = Possession Of Cocaine : 0.98 (1.9/0)
|   |   c_charge_desc = Harm Public Servant Or Family : 0.98 (1.9/0)
|   |   c_charge_desc = Possess Cannabis 1000FTSch : 0 (0/0)
|   |   c_charge_desc = Consp Traff Oxycodone  4g><14g : 0 (0/0)
|   |   c_charge_desc = Consume Alcoholic Bev Pub : 0 (0/0)
|   |   c_charge_desc = Shoot Into Vehicle : 0.45 (1.9/0)
|   |   c_charge_desc = Battery Spouse Or Girlfriend : 0.98 (1.9/0)
|   |   c_charge_desc = Delivery Of Drug Paraphernalia : 0 (0/0)
|   |   c_charge_desc = Theft : 0 (0/0)

Size of the tree : 15561

Time taken to build model: 0.02 seconds

=== Cross-validation ===
=== Summary ===

Correlation coefficient                  0.9314
Mean absolute error                      0.0433
Root mean squared error                  0.1825
Relative absolute error                  8.7487 %
Root relative squared error             36.6999 %
Total Number of Instances             4761
```

These results are helpful by getting a grasp of how the data will look when the algorithms are run.

Future Work

In the future, I can use another algorithm J48, which is used to examine data categorically and continuously. J48 develops a decision node utilizing the expected estimations of the class. It can deal with particular characteristics, lost or missing attribute estimations of the data and varying attribute costs. Thus, J48 would be a good fit for the data set. J48 is a specific algorithm for Weka, and this is what I received when I ran J48 algorithm on Weka.

```
=== Run information ===

Scheme:       weka.classifiers.trees.J48 -C 0.25 -M 2
Relation:     preprocessed-data-small-weka.filters.unsupervised.attribute.NumericToNominal-Rfirst-last
Instances:    4761
Attributes:   24
              sex
              age
              race
              juv_fel_count
              1_decile_score
              juv_misd_count
              juv_other_count
              2_priors_count
              days_b_screening_arrest
              c_days_from_compas
              c_charge_degree
              c_charge_desc
              is_recid
              r_days_from_arrest
              r_charge_desc
              is_violent_recid
              vr_charge_desc
              decile_score
              v_decile_score
              priors_count
              start
              end
              event
              two_year_recid
Test mode:    10-fold cross-validation

=== Classifier model (full training set) ===
```

```
J48 pruned tree
------------------

is_recid = 0: 0 (2497.0)
is_recid = 1: 1 (2264.0/135.0)

Number of Leaves  :     2

Size of the tree :      3


Time taken to build model: 0.1 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances        4626               97.1645 %
Incorrectly Classified Instances       135                2.8355 %
Kappa statistic                          0.943
Mean absolute error                      0.0534
Root mean squared error                  0.1634
Relative absolute error                 10.7905 %
Root relative squared error             32.8722 %
Total Number of Instances             4761

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.949    0.000    1.000      0.949    0.974      0.945    0.969     0.983     0
                 1.000    0.051    0.940      1.000    0.969      0.945    0.969     0.924     1
Weighted Avg.    0.972    0.023    0.973      0.972    0.972      0.945    0.969     0.957

=== Confusion Matrix ===
```

```
=== Confusion Matrix ===

    a    b   <-- classified as
 2497  135 |    a = 0
    0 2129 |    b = 1
```

With this result, I think I can explore more on the dataset and try running other algorithms such as JRip.

What I have learned

Throughout the research, I was able to get a deeper understanding on the topic of machine learning. I was able to get to know more about the different algorithms and use them to explore datasets. Using multiple algorithms, such as random forest and association mining, I was able to deepen my knowledge of the algorithms by understanding the logic behind them. I was also able to learn the importance of finding other algorithms to run on the dataset when an algorithm does not work. For example, my initial plan of using Random Forest did not go as planned due to some dimension issues in the code. Thus, I decided to move on to Association Rule Mining and was able to discover new relations of the dataset.

Moreover, I was able to learn more about the risk assessment tool itself, how widely it is used and how it is raising concerns regarding racial bias. It was such a fascinating experience to work on an interdisciplinary subject of computer science and criminal justice. This research made me think about how my software product could impact in diverse fields and the importance of creating reliable products as an computer science engineer.

References

Chohlas-Wood, Alex. “Understanding Risk Assessment Instruments in Criminal Justice.” *Brookings*, Brookings, 17 June 2020, www.brookings.edu/research/understanding-risk-assessment-instruments-in-criminal-justice.

Larson, J., Mattu, S., Kirchner, L. & Angwin, J. (2016). How We Analyzed the COMPAS Recidivism Algorithm. Pro Publica. Retrieved October 20, 2020, from https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm