From Credit Risk to Explainable AI Research

FABRIZIO RUSSO

UCL - 20.05.22
Content

• Overview of modelling data from Credit Risk Agencies (CRAs)

• Credit Risk Modelling for Retail Application: from regression to machine learning

• Ongoing Research: Causal and Explainable Neural Networks

• Final Remarks
CRA data*

WHAT INFORMATION DO BANKS USE TO ASSESS YOUR CREDIT WORTHINESS?

*MATERIAL FROM RUSSO, 2019
# Application Credit Checks Data

## Credit Scores

### Closed User Group Information (CUG)
- Number of accounts
- Outstanding balance
- Repayment behaviour
- Types of accounts
- Credit card utilisation

### Credit Searches
- Recent credit activity

### Public Data
- Electoral roll
- Court judgements
- Bankruptcies
- Financial associates

### Associates
- Postcode Level

## Transactional Data

FROM CREDIT RISK TO EXPLAINABLE AI RESEARCH - RUSSO FABRIZIO

20/05/2022
Tasks that Use Extensive CRA Data

- Origination Strategy
- Reject Inference
- Customer Management Strategy
- Regulatory Impacts
- New Products
Data Analysis – Quantity & Types

Full Initial Dataset

<table>
<thead>
<tr>
<th>CRA</th>
<th>Applicant</th>
</tr>
</thead>
<tbody>
<tr>
<td>Applicant Stability</td>
<td>Financial Activity</td>
</tr>
<tr>
<td>Repayment Performance</td>
<td>Applicant Profile</td>
</tr>
<tr>
<td></td>
<td>Affordability</td>
</tr>
</tbody>
</table>

Good data quality variables

Predictive variables

Model variables
Data Analysis – Cleansing

Take-up can take up to 3 months
Application spikes in September and December

Unusual behaviour from December '13 to May '15
~200 cases per month with default value of ~100
Data Analysis – Feature Engineering

### Variable

<table>
<thead>
<tr>
<th>Variable</th>
<th>Overall IV</th>
<th>Data Group</th>
<th>Potential for Modelling</th>
</tr>
</thead>
<tbody>
<tr>
<td>Worst Status Last 6 Months</td>
<td>1.22</td>
<td>CUG</td>
<td>y</td>
</tr>
<tr>
<td>Number of Delinquent Accounts</td>
<td>1.22</td>
<td>CUG</td>
<td>y</td>
</tr>
<tr>
<td>Value of Delinquent Accounts</td>
<td>1.22</td>
<td>CUG</td>
<td>maybe</td>
</tr>
<tr>
<td>Months Since Delinquency</td>
<td>1.19</td>
<td>CUG</td>
<td>y</td>
</tr>
<tr>
<td>Value of Unsecured Delinquent Debt</td>
<td>1.18</td>
<td>CUG</td>
<td>no</td>
</tr>
<tr>
<td>Number of Unsecured Delinquencies</td>
<td>1.18</td>
<td>CUG</td>
<td>Y</td>
</tr>
<tr>
<td>Time Since Most Recent Default</td>
<td>1.05</td>
<td>CUG</td>
<td>y</td>
</tr>
<tr>
<td>Value of Defaults</td>
<td>1.03</td>
<td>CUG</td>
<td>no</td>
</tr>
<tr>
<td>Number of Defaults</td>
<td>1.03</td>
<td>CUG</td>
<td>Y</td>
</tr>
<tr>
<td>Months Since Mortgage Default</td>
<td>1.00</td>
<td>CUG</td>
<td>y</td>
</tr>
<tr>
<td>Value of Mortgage Default</td>
<td>0.99</td>
<td>CUG</td>
<td>maybe</td>
</tr>
<tr>
<td>Number of Mortgage Defaults</td>
<td>0.99</td>
<td>CUG</td>
<td>y</td>
</tr>
<tr>
<td>Confirmed at Address</td>
<td>0.31</td>
<td>ER</td>
<td>y</td>
</tr>
<tr>
<td>Number of Judgements</td>
<td>0.28</td>
<td>Public</td>
<td>y</td>
</tr>
<tr>
<td>Time Since Judgement</td>
<td>0.28</td>
<td>Public</td>
<td>Y</td>
</tr>
<tr>
<td>Time on ER at Current Address</td>
<td>0.27</td>
<td>ER</td>
<td>y</td>
</tr>
<tr>
<td>Number of All Public Judgement Records</td>
<td>0.26</td>
<td>Public</td>
<td>y</td>
</tr>
<tr>
<td>Time Since Bankruptcy</td>
<td>0.26</td>
<td>Public</td>
<td>y</td>
</tr>
<tr>
<td>Value of Bankruptcy</td>
<td>0.26</td>
<td>Public</td>
<td>y</td>
</tr>
<tr>
<td>Applicant Age</td>
<td>0.25</td>
<td>Internal</td>
<td>y</td>
</tr>
<tr>
<td>Confirmed at Current Address</td>
<td>0.18</td>
<td>ER</td>
<td>y</td>
</tr>
<tr>
<td>Worst Status of Active Accounts Last 12 Months</td>
<td>0.92</td>
<td>CUG</td>
<td>y</td>
</tr>
<tr>
<td>Credit Limit Utilisation</td>
<td>0.92</td>
<td>CUG</td>
<td>y</td>
</tr>
<tr>
<td>Worst Current Status</td>
<td>0.89</td>
<td>CUG</td>
<td>y</td>
</tr>
<tr>
<td>Worst Status Last 3 Months</td>
<td>0.83</td>
<td>CUG</td>
<td>y</td>
</tr>
<tr>
<td>Months Since Most Recent Delinquency</td>
<td>0.78</td>
<td>CUG</td>
<td>y</td>
</tr>
</tbody>
</table>

### Data Analysis

**WoE = LnOdds(attribute) – LnOdds(population)**

**IV = Avg_{Good}(WoE) – Avg_{Bad}(WoE)**

---

**Monotonic Trend – Worst Status Last 6 Months WOE**

Monotonic trend (decreasing)
From Regression to Machine Learning*

CAN WE GET MORE OUT OF THE SAME DATA?

*MATERIAL FROM RUSSO ET AL, 2019. SEE FROM RISK SCORECARDS WITH MACHINE LEARNING (BRIZIO1RUSSO.GITHUB.IO)
Algorithm Comparison

GINI Comparison on Test Sample

- Internal Scorecard
- Internal & CRA Scorecard
- Elasticnet
- Classification Tree
- Random Forest
- XGBoost
- 1HL Neural Network

ROC Curves

GINI Comparison on Test Sample

11
FROM CREDIT RISK TO EXPLAINABLE AI RESEARCH - RUSSO FABRIZIO
Performance Comparison

Score Distribution by Outcome

Bad Rate by Quintile

Performance Comparison

FROM CREDIT RISK TO EXPLAINABLE AI RESEARCH - RUSSO FABRIZIO
Regulatory Considerations

Governance

Transparency

Consistent Decisions & Treat Customers Fairly
From Transparent Machine Learning to Causal XAI*

HOW DO WE GO FROM ASSESSING MODELS EX-POST TO MAKING SURE THEY LOOK AT THE RIGHT RELATIONSHIPS?

*MATERIAL FROM RUSSO & TONI, 2022.
Causal Discovery and Injection for Feed-Forward Neural Networks

- In finance many hard problems are tackled with models (e.g. fraud, pricing, credit scoring, trading, planning etc.)

- Practitioners often have a lot of domain (causal) knowledge

- Regulation is quite strict in requiring model stakeholders to understand and “own” their models

- Machine Learning models (e.g. Neural Networks) do not easily allow knowledge integration nor interpretation

Causal Injection into Neural Networks

- Introducing causality into neural networks not only makes them more robust and reliable, but it is also a step towards their interpretability
Formal Set-up

**Supervised Learning setting**

- $\mathbf{X} = [X_1, \ldots, X_d] \in \mathcal{X} \subseteq \mathbb{R}^d$ (input features)
- $Y \in \mathcal{Y} \subseteq \mathbb{R}$ (target)
- $\mathcal{P}_{\mathbf{X},Y}$ joint distribution of input and target (DGP)
- $\mathcal{D} = \{(X_i, Y_i), i \in \{1, \ldots, N\}\}$
  - $N$ i.i.d samples from $\mathcal{P}_{\mathbf{X},Y}$
- $f_Y: \mathcal{X} \rightarrow \mathcal{Y}$
- Goal: find $\hat{f}_Y$ in $\mathcal{H}$ (hypothesis space)
- $\mathcal{H}$ too complex $\rightarrow$ Regularize

**Causal framework (Pearl, 2009)**

- Causal Structure is a DAG $G = \langle V, E \rangle$
  - $V = \{Y, X_1, \ldots, X_{d+1}\}$ the set of vertices
  - $E \subseteq V$ the set of edges
- $v_i = f_i (\text{pa}_i, u_i)$
  - $v_i$ is a value for $V_i \in V$ with parents $\text{Pa}_i$ having values $\text{pa}_i$
  - $f_i$ any function
  - $u_i$ representing the errors due to omitted factors
(a) Example DAG from Kyono, Zhang and Schaar 2020.
(b) Adjacency Matrix produced by CASTLE (Kyono, Zhang and Schaar 2020) when fitted to the synthetic data produced following the DAG to the left.
Causal Injection – The Intuition

**Objective:**

have the network use only the relationships contained in the DAG i.e. predict each of the features using only its parents.
Joint Network

Predict the target while reconstructing all other input features
Joint Network

Predict the target while reconstructing all other input features
Joint Network

Predict the target while reconstructing all other input features

Is this input-output relationship contemplated in my causal DAG?
Joint Network

Predict the target while reconstructing all other input features

NO?

➢ “Semantic” Regularization
Limitations of Proposed Algorithm

- It requires a complete DAG (covering all variables considered in the problem and the data)
- Full causal DAG is rare and often impractical to build
- We propose a second algorithm that involves Subject Matter Experts (SMEs) providing their input
Algorithm 2 – Human-AI Collaboration

Figure 2: Input graph $G_p$, as partial causal knowledge for the Adult dataset, in the form of an adjacency matrix $W$. Blue represents edges; missing edges in white (hard constraints).

Figure 3: Example of computed DAG for Adult dataset (see Section 5.3.3). Cyan nodes at the top are computed causes for the target ("Income>50K"), edges coming out of the target are in blue while in purple are the edges into nodes that cannot be caused (as per basic assumptions in Section 5.3.2).
HCI Causal Injection - Results

Table 1: Experiments with real data in the financial/economics sector. We report MSE (AUC) for regression (classification) across different sample sizes of the training data (best results in bold). We also detail, for each dataset, the number of features/nodes $|V|$ and the number of edges $|E|$ in the injected DAG (for our method) and in the (graph drawn from the) underlying adjacency matrix (for CASTLE). NA indicates a data size $(N)$ bigger than the full dataset. CASTLE and Injected columns refer to Section 5.3.1, for Partial and Refined columns see Sections 5.3.2 and 5.3.3, respectively.

<table>
<thead>
<tr>
<th>Data size $(N)$</th>
<th>REGRESSION (Metric: MSE)</th>
<th>CLASSIFICATION (Metric: AUC)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>California $</td>
<td>V</td>
</tr>
<tr>
<td>100</td>
<td>7.05 (12.81)</td>
<td>2.94 (2.63)</td>
</tr>
<tr>
<td>500</td>
<td>2.33 (1.39)</td>
<td>2.25 (1.07)</td>
</tr>
<tr>
<td>1000</td>
<td>2.96 (4.12)</td>
<td>1.68 (1.14)</td>
</tr>
<tr>
<td>2000</td>
<td>3.86 (3.68)</td>
<td>1.71 (0.57)</td>
</tr>
<tr>
<td>5000</td>
<td>4.91 (7.41)</td>
<td>1.51 (0.62)</td>
</tr>
<tr>
<td>10000</td>
<td>1.74 (1.70)</td>
<td>1.16 (0.31)</td>
</tr>
<tr>
<td>20000</td>
<td>0.66 (0.08)</td>
<td>1.02 (0.35)</td>
</tr>
</tbody>
</table>
Conclusion

- CRA Data is what enables (more) accurate credit worthiness assessment in UK
- Logistic Regression is to this day the most used technique for its interpretability
- Other ML algorithms can achieve similar levels of transparency
- Statistical relationship is not the same as Causal Relationship
- High-stakes decision models should look at both statistical and causal relationships
Questions?

GET IN TOUCH

fabrizio@imperial.ac.uk
briziorusso.github.io
References


Acknowledgments
Credit risk material was adapted from presentations given while the author was working at 4most Europe, thanks to the team that helped with its preparation.