Machine learning and fairness in commercial insurance

Insurance Data Science Conference
London, UK

July 2018
Cytora uses artificial intelligence to improve risk targeting, selection, and pricing for commercial insurance.
Decisions that personally affect us are increasingly data-driven

- Credit scoring
- Insurance
- Airport security
- Crowd monitoring
- Reoffending rates
- Advertising
Defining fairness
Fairness means treating individuals from different groups equally.
Fairness through unawareness is not sufficient to guarantee equal treatment for individuals in protected groups.
**Careful consideration is necessary when designing decision systems**

<table>
<thead>
<tr>
<th>Data</th>
<th>Modelling</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Inherent data biases</td>
<td>- Quantify feature contributions</td>
</tr>
<tr>
<td>- Reasoned vetting of variables</td>
<td>- Tune for fairness</td>
</tr>
<tr>
<td>- True measures of latent risk</td>
<td>- Bias in, bias out</td>
</tr>
<tr>
<td>- Measure the protected attribute</td>
<td></td>
</tr>
</tbody>
</table>
Protected attributes encoded in “harmless” rating factors

Sex

Protected attr
Rating factor

Car colour

Driving aggression

Latent
Target

Accident

“Counterfactual Fairness”
Strategies for fairer pricing in insurance
1. Observe relevant rating factors

- Sex
- Telemetrics
- Driving aggression
- Accident

Protected attr
Rating factor

Latent
Target
2. Adjust premiums to optimise metrics of fairness

- Profit (accuracy)
- False positive rate, equal opportunity (FP/N)
- False negative rate (FN/P)
- Equalised odds (FPR & FNR)
- Equality of opportunity (FPR)
- Calibration (true probabilities)
- Demographic parity

Confusion matrix per protected group

<table>
<thead>
<tr>
<th>Actual</th>
<th>Predicted</th>
</tr>
</thead>
<tbody>
<tr>
<td>No loss</td>
<td>TN</td>
</tr>
<tr>
<td>Loss</td>
<td>FP</td>
</tr>
<tr>
<td>No loss</td>
<td>FN</td>
</tr>
<tr>
<td></td>
<td>TP</td>
</tr>
</tbody>
</table>

Berk, Heidari, Jabbari, Kearns, Roth (2017) "Fairness in Criminal Justice Risk Assessments: The State of the Art"

Kleinberg, Mullainathan, Raghavan (2016) "Inherent Trade-Offs in the Fair Determination of Risk Scores"
3. Design and train algorithms with fairness baked-in

- **Structural models**
  - Kilbertes, et. al. (2018)  "Avoiding discrimination through causal reasoning"
  - Kusner, Loftus, Russell, Silva (2018)  "Counterfactual Fairness"

- **Penalised / constrained loss functions**
  - Zafar, et. al. (2017)  "Fairness Beyond Disparate Treatment & Disparate Impact"
  - Zhao, et. al. (2017)  "Men Also Like Shopping: Reducing Gender Bias Amplification using Corpus-level Constraints"

- **Model inspection**
  - Tan, Caruana, Hooker, Lou (2018)  "Detecting Bias in Black-Box Models Using Transparent Model Distillation"
Example: Restaurant shutdown

**Rating factor:** Cuisine type = Krusty Burgers

**Protected attribute:** Shelbyville or Springfield resident

**Solutions:**
1. Observe management quality, menu, online reviews...
2. Geographic analysis of offered premium (adjust?)
3. Use fairness-calibrated algos
Summary

Data-driven modelling and machine learning can improve fairness by

1) Find better approximations of latent risk
2) Quantify effects on domain-specific fairness metrics
3) Calibrate decision making process to optimise fairness
Questions?

oliver@cytora.com

Learn more at cytora.com