Predicting appointment non-attendance using AI

Amy Nelson, Daniel Herron, Geraint Rees & Parashkev Nachev

@KFDigital18
Non-attendance

- Cost £1bn per annum
- Outpatient appointment non-attendance 2016-17: 8 million 8.5%
- Health opportunity loss
- Reminders:
  - Telephone 13-40%
  - Text 11-40%
  - Email 25%

Data from NHS Digital hospital outpatient activity reports 2010-17
Why do we need Artificial Intelligence?

• Human behaviour <-> environment <-> hospital
• This will always be a complicated problem
• Complex means required to solve it

• Need big data
• Need machine learning
Our objective

• Build model of sufficient complexity
• Use adequate number of variables

Build a classification model to predict non-attendance behaviour from operations data
Data

- 2 years of Radiology Information Services data at UCLH
- 22,318 MRI appointments and 17,295 patients
- 123 predictor variables:
  - Location
  - Previous scans
  - Previous DNAs
  - Total cost of patient
  - Specialty count
  - Grade of clinician
  - Referral/appointment month and day
Data

- 10% test set removed
- Data pre-processing
- 5 fold cross-validation
Model selection

- Logistic regression
- Support Vector Machine
- Random Forest
- AdaBoost

Model performance

Increasing model complexity
Performance on test data

- Top performing model: AdaBoost
- AUC 0.83 (+/- 0.01)

<table>
<thead>
<tr>
<th></th>
<th>PPV</th>
<th>Number to call</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>0.09</td>
<td>11</td>
</tr>
<tr>
<td>95% sensitivity</td>
<td>0.16</td>
<td>6.3</td>
</tr>
<tr>
<td>90% sensitivity</td>
<td>0.18</td>
<td>5.5</td>
</tr>
</tbody>
</table>
Top predictors

1. Number of previous DNAs
2. Referral to appointment time lag
3. Appointment day of the week
4. Appointment month of the year
5. Total cost of patient
6. Referral month of the year
7. Home address latitude
8. Travelling distance to hospital
9. Home address longitude
10. Referral day of the week
Dimensionality

- Current models tend to have under 10 variables
- Optimal performance not seen until 84 features
- We need a complex model
Implementation

- Open source
- Cost aside from labour is zero
- Improves with more data

- Although model is complex, running at test time is easy
- Predictions available at time of booking
Applications

1. Improved targeting of reminding
   • Calibrate volume and intensity to risk

2. Match appointment to patient
   • “Collaborative filtering”

3. Inferring service features that reduce attendance
   • If fewer people than predicted attend, could something be putting patients off attendance eg. accessibility?
Patient perspective

• Increases access to health care
• Reduced waiting time
• System anticipates what the patient needs
• Not a substitute for patients having a choice
Future directions

- More data – demographics, weather
- Other scheduling data
- Generalise across hospital
- Prediction, mapping, service insight
Thank you! Questions?
Supplementary

Class Imbalance

- Randomly undersample
- Synthetically oversample
- Class weights

Precision-Recall curve