Outcomes Models with Machine Learning

Sarah Gulliford PhD
Division of Radiotherapy & Imaging
sarahg@icr.ac.uk

AAPM 31st July 2017

Making the discoveries that defeat cancer

Radiotherapy Outcomes

Therapeutic Ratio

- Probability of Tumor Control
- Probability of Complication

Dose (Gy)

0 20 40 60 80 100 120

0.2 0.4 0.6 0.8 1

Rectum Prostate Bladder
Reviews of Outcome Modelling in Radiotherapy


What is Machine Learning?
Original concept based on the way that a human brain learns

- Algorithms designed to learn from the data
- No a priori knowledge of the relationship between the data
- Training using example cases
- Ability to generalise to unseen cases

Unsupervised learning
Data grouped together using common features
No reference made to corresponding output
‘Unlabelled data’

- Self organising maps (kohonen)
- Principal Component Analysis
Can be used for feature selection prior to a supervised learning approach
Supervised learning

Algorithms trained to relate input features to output (outcomes)
‘Labelled’ data
Iterative training using cost function to find best model

- Support Vector Machines
- Random Forest
- Neural Networks

Used for classification & regression

Common considerations (1)

Data splitting:

- Cross validation
- Bootstrapping (sampling with replacement)
  - Independent test set

TRIPOD guidelines


Common considerations (2)

Assessment of results:

- Receiver Operator Curve (ROC analysis)
- Calibration Curves
- Learning curves (bias/variance)
Common considerations (3)
Curse of Dimensionality:
High order data becomes sparse in a multidimensional space

http://www.visiondummy.com/2014/04/curse-dimensionality-affect-classification/

Common considerations (4)
Garbage in Garbage out:
Models are entirely dependent on the quality of the data

- Tumour/organ contouring consistency
- Intra/inter fraction motion
- Adaptive planning
- Reporting of events using standardised scales
- Quality Assurance

The curse of dimensions

Data stored as a jpeg
3 dimensional array
2816x2112x3
The curse of dimensions

Data stored as
2D matrix 2816 by 2112
Radiotherapy Outcomes

Dosimetry Features

Challenges of modelling dose-volume effects

- Dose-volume relationship to toxicity is complex and not well understood
- Highly correlated data
- Toxicity related to a number of factors including dose-volume effects
Challenges of modelling dose-volume effects

- Dose-volume relationship to toxicity is complex and not well understood
- No a priori knowledge of model required
- Highly correlated data
- Methods to deal with correlated data
- Toxicity related to a number of factors including dose-volume effects
  - Can include all types of data without knowing how the variables are related

Radiotherapy planning

- Parotid gland
- Spinal cord
- Brain stem
- Oral cavity
- Parotid gland
- Primary planning target volume
- Nodal planning target volume
### Patients

<table>
<thead>
<tr>
<th>Trial</th>
<th>Number available</th>
<th>Primary disease site</th>
<th>Radiotherapy technique</th>
<th>Concurrent chemotherapy</th>
</tr>
</thead>
<tbody>
<tr>
<td>PARSPORT</td>
<td>71</td>
<td>Oropharynx, hypopharynx</td>
<td>Bilateral; Conventional, IMRT</td>
<td>No</td>
</tr>
<tr>
<td>COSTAR</td>
<td>78</td>
<td>Parotid gland</td>
<td>Unilateral; Conventional, IMRT</td>
<td>No</td>
</tr>
<tr>
<td>Dose Escalation</td>
<td>30</td>
<td>Hypopharynx, larynx</td>
<td>Bilateral; IMRT</td>
<td>Yes</td>
</tr>
<tr>
<td>Midline</td>
<td>117</td>
<td>Oropharynx</td>
<td>Bilateral; IMRT</td>
<td>Yes</td>
</tr>
<tr>
<td>Nasopharynx</td>
<td>36</td>
<td>Nasopharynx</td>
<td>Bilateral; IMRT</td>
<td>Yes</td>
</tr>
<tr>
<td>Unknown Primary</td>
<td>19</td>
<td>Unknown primary</td>
<td>Bilateral; IMRT</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Dean et al. Rad Onc 120 (2016) 21–27

---

### Toxicity scoring

<table>
<thead>
<tr>
<th>Grade 1</th>
<th>Grade 2</th>
<th>Grade 3</th>
<th>Grade 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clinical oral mucositis</td>
<td>Clinical oral mucositis</td>
<td>Clinical oral mucositis</td>
<td>Clinical oral mucositis</td>
</tr>
</tbody>
</table>

- Erythema of the mucosa
- Patchy ulcerations
- Confluent ulcerations
- Tissue necrosis; significant spontaneous bleeding

Dose limiting toxicity

Treatment interruptions

Dean et al. Rad Onc 120 (2016) 21–27

---

### Toxicity scoring

- • Prospectively measured at baseline, weekly during and 1, 2, 3, 4 and 8 weeks post-radiotherapy
- • 351 patients with data available
- • Patients with baseline toxicity excluded
- • Peak grade < 3 vs >= 3
- • Patients with missing data excluded
- • Final Dataset 183 patients

Dean et al. Rad Onc 120 (2016) 21–27
Clinical data

• Age
• Sex
• Primary disease site
• Definitive radiotherapy vs postoperative radiotherapy
• Concomitant treatments
  - Induction chemotherapy
  - Concurrent chemotherapy regime (cisplatin/carboplatin/both)
• No smoking, alcohol or genetic data

Dean et al Rad Onc 120 (2016) 21–27

Oral mucositis modelling

Grade 3 ‘Confluent ulceration’

Oral cavity

Dose-volume histogram
  • fractional dose

Spatial features
  • 3D moment invariants

Dean et al Rad Onc 120 (2016) 21–27

Spearman’s Correlation Matrix

Dean et al Rad Onc 120 (2016) 21–27
Penalised Logistic Regression

Logistic regression technique extended to mitigate for highly correlated data.

- Ridge Regression – some coefficients set to zero
- Least absolute shrinkage and selection operator LASSO regularisation – coefficients reduced

Random Forests

Ensembles of decision trees created and initialised using a randomly selected subset of the available data cases.

- The final result is aggregated from the contributions of each tree.
  - outcome classification this will be the most votes (i.e.) class chosen by the most trees
  - regression the outcome will be averaged across all the trees.

Support Vector Machines

Classify data by translating variables into a higher dimensional space where they are linearly separable.

Ideally a boundary can be found that completely separates the two possible classes and maximises the distance between them.

Mapping achieved using a Kernel function:
- Radial Basis function
- Polynomial function

Support Vector Machines

Computationally intensive to solve; however, it is possible to characterise the prediction function using only a subset of training data (support vectors).


PLR – no spatial features

Dean et al. Rad Onc. 120 (2016) 21–27
Random Forest – no spatial features

Dean et al Rad Onc 120 (2016) 21–27

Results

Spatial information did not improve predictive performance

Dean et al Rad Onc 120 (2016) 21–27

Representing Dose distribution using dose surface Maps

3D dose distribution Rectal Dose Surface Map
Using dose-surface maps to predict radiation-induced rectal bleeding: a neural network approach

Florian Buettner, Sarah J. Guillford, Steve Webb and Mike Partridge

Institute Department of Physics, Institute of Cancer Research and Royal Marsden NHS Foundation Trust, Sutton, Surrey, SM2 5PT, UK.

Artificial Neural Network

- Input layer
- Hidden layer
- Output layer

Weighted sum on each node $\sum w_{ij}$

Non linear activation function

Backpropagation of errors

Dose surface map ANN architecture

locally connected NN
2 hidden layers
a-c) individualised weights
d) shared weights
2 output nodes

Ensemble Learning

Ensemble learning incorporates groups of neural networks each with different starting conditions and selected subset of training data sets 250 NN initialised. Results aggregated

Expert ensemble

Results of each NN are assessed and if they improve the performance of the ensemble they are “voted in”.


Patients

Prostate cancer UK-MRC RT01 trial

Compared 64Gy vs 74Gy (circa 1998-2001) 388 patients with data
Used to predict rectal bleeding >= Grade 2 (RMH score) simple outpatient management/transfusion

Patients with baseline toxicity excluded

329 patients 53 patients with G2 Rectal Bleeding


Results

| Table 3: Performance of all locally connected nets. AUC$_{\text{cat}}$ denotes the AUC calculated from all nets in the ensemble and AUC$_{\text{cat}}$ the AUC derived from the inputs only. |
|---|---|---|---|
| AUC$_{\text{cat}}$ | AUC$_{\text{cat}}$ | AUC$_{\text{cat}}$ |
| 0.59 | 0.60 | 0.67 |
| 0.59 | 0.60 | 0.67 |
| 0.59 | 0.60 | 0.67 |
| 0.59 | 0.60 | 0.67 |
| 0.59 | 0.60 | 0.67 |

The number of unique weights and biases was calculated by taking off the unique connections between the nodes: 1484 = 205 + 1484 + 0 + 3 + 2390 + 522; 370 = 1065 + 370; 36 + 2 + 2 = 11; 36 + 36 + 2 + 3 + 3 + 2 + 3 + 3 + 3 + 2 = 80; 36 + 36 + 2 = 62; 80 + 80 = 82.

Compare results with fully connected NN using DSH data

AUC 0.59

Why such a low AUC?

• Incomplete characterisation of spatial information
• Model architecture
• Inter & Intra fraction rectal motion/filling
• Only dosimetry in the model

What’s missing?

☐ Clinical factors (age, diabetes etc)
☐ Other therapies (hormones)
☐ Genetic variants (SNPS)

Why don’t we use Machine Learning more?

Reputation mystical black box

Wide variety of techniques (which approach is appropriate?)

The road less trodden

Summary

• Evidence that Machine Learning approaches are complimentary to traditional statistical techniques and each other.

• Data hungry: more variables need more datasets

• Require rigorous methodology and independent validation
Acknowledgements

Prof David Dearnaley
Prof Christopher Nutting
Dr Florian Buettner
Dr Julia Murray
Dr Jamie Dean
Mr Matt Sydes
Prof Emma Hall

Patients and Colleagues
who report, collect and
share data