Session 79PD, Using Predictive Analytics to Develop Assumptions

Moderator/Presenter:
Jonathan D. White, FSA, MAAA, CERA

Presenters:
Missy A. Gordon, FSA, MAAA
Brian M. Hartman, ASA
2018 SOA Health Meeting

DAVID WHITE, FSA, MBA, CERA – MODERATOR
BRIAN HARTMAN, PHD, ASA – PRESENTER
MISSY GORDON, FSA, MAAA - PRESENTER

Session 79, Using Predictive Analytics to Develop Assumptions
June 26, 2018
SOCIETY OF ACTUARIES
Antitrust Compliance Guidelines

Active participation in the Society of Actuaries is an important aspect of membership. While the positive contributions of professional societies and associations are well-recognized and encouraged, association activities are vulnerable to close antitrust scrutiny. By their very nature, associations bring together industry competitors and other market participants.

The United States antitrust laws aim to protect consumers by preserving the free economy and prohibiting anti-competitive business practices; they promote competition. There are both state and federal antitrust laws, although state antitrust laws closely follow federal law. The Sherman Act, is the primary U.S. antitrust law pertaining to association activities. The Sherman Act prohibits every contract, combination or conspiracy that places an unreasonable restraint on trade. There are, however, some activities that are illegal under all circumstances, such as price fixing, market allocation and collusive bidding.

There is no safe harbor under the antitrust law for professional association activities. Therefore, association meeting participants should refrain from discussing any activity that could potentially be construed as having an anti-competitive effect. Discussions relating to product or service pricing, market allocations, membership restrictions, product standardization or other conditions on trade could arguably be perceived as a restraint on trade and may expose the SOA and its members to antitrust enforcement procedures.

While participating in all SOA in person meetings, webinars, teleconferences or side discussions, you should avoid discussing competitively sensitive information with competitors and follow these guidelines:

- **Do not** discuss prices for services or products or anything else that might affect prices
- **Do not** discuss what you or other entities plan to do in a particular geographic or product markets or with particular customers.
- **Do not** speak on behalf of the SOA or any of its committees unless specifically authorized to do so.
- **Do** leave a meeting where any anticompetitive pricing or market allocation discussion occurs.
- **Do** alert SOA staff and/or legal counsel to any concerning discussions
- **Do** consult with legal counsel before raising any matter or making a statement that may involve competitively sensitive information.

Adherence to these guidelines involves not only avoidance of antitrust violations, but avoidance of behavior which might be so construed. These guidelines only provide an overview of prohibited activities. SOA legal counsel reviews meeting agenda and materials as deemed appropriate and any discussion that departs from the formal agenda should be scrutinized carefully. Antitrust compliance is everyone’s responsibility; however, please seek legal counsel if you have any questions or concerns.
Presentation Disclaimer

Presentations are intended for educational purposes only and do not replace independent professional judgment. Statements of fact and opinions expressed are those of the participants individually and, unless expressly stated to the contrary, are not the opinion or position of the Society of Actuaries, its cosponsors or its committees. The Society of Actuaries does not endorse or approve, and assumes no responsibility for, the content, accuracy or completeness of the information presented. Attendees should note that the sessions are audio-recorded and may be published in various media, including print, audio and video formats without further notice.
To Participate, look for Polls in the SOA Event App or visit health.cnf.io in your browser.

Find The Polls Feature Under More In The Event App.

Type health.cnf.io In Your Browser.

Choose your session.

Respond to Polls when they appear.
Poll: How familiar are you with Predictive Analytics?
How familiar are you with Predictive Analytics?

- I'm an expert: 44%
- Strong understanding: 7%
- Familiar: 45%
- Noobie: 4%
Identifying insureds for preventative care

Probability of claim

Incurred claims

Claim termination

Flagging miscoded or fraudulent claims

Evaluating new preventative care or treatments

Mortality & Lapse

Experience studies

Policyholder behavior due to rate increases

Risk scores & underwriting

Utilization & Improvement
Agenda

Using predictive analytics to develop assumptions
• Bias/variance tradeoff – Brian
• Developing assumptions – Missy
• Some other fun applications – Brian
• Question and answer
Bias/Variance Tradeoff

BRIAN HARTMAN, PHD, ASA
Assistant Professor of Statistics and Actuarial Program Director, Brigham Young University
June 26, 2018
Motivating Example
Motivating Example
Motivating Example
Motivating Example
Motivating Example
Motivating Example
Motivating Example
Motivating Example
Motivating Example
Motivating Example

Model Degree: 8
Training MSE: 14.896
Test MSE: 23.772
Motivating Example
Motivating Example
Motivating Example
Motivating Example
Motivating Example
Bias vs. Variance *(Hastie et al. 2009)*

The expected squared prediction error is:

\[
E \left[ (Y - \hat{f}(x))^2 \right] = \left( E[\hat{f}(x)] - f(x) \right)^2 + E \left[ \hat{f}(x) - E[\hat{f}(x)] \right]^2 + \sigma_e^2
\]

Bias\(^2\) + Variance + Irreducible Error

A perfect model and infinite data would reduce the first two terms to zero, but with finite data and imperfect models, we will need to choose between minimizing bias and minimizing variance.
Bias vs. Variance in our Example
Increasing the Sample Size

Sample Size = 10000

MSE

Model Degree

Training Set
Test Set
Live Content Slide
When playing as a slideshow, this slide will display live content

Social Q&A
Using predictive analytics to develop assumptions

MISSY GORDON, FSA, MAAA
Principal and Consulting Actuary, Milliman - Minneapolis
June 26, 2018
Traversing Bias-Variance Tradeoff (BVT)

Variable interactions

Judgement/Manual

Robust/Automated

Variable selection / data credibility

A:E with credibility weighting

Classical Generalized Linear Model (GLM) with offset

Penalized GLM with offset

Domain knowledge

Gradient Boosting Machine (GBM)
## Traversing BVT

<table>
<thead>
<tr>
<th>Traversing BVT</th>
<th>A:E with credibility weighting</th>
<th>Classical GLM with Offset</th>
<th>Penalized GLM with Offset</th>
<th>GBM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data credibility</td>
<td>Judgement</td>
<td>Judgement</td>
<td>Cross validation</td>
<td>Cross validation to tune hyper-parameters to control for overfitting</td>
</tr>
<tr>
<td>Variable selection</td>
<td>Judgement</td>
<td>In-sample tests of fit</td>
<td>Cross validation</td>
<td>Automated process to minimize prediction error</td>
</tr>
<tr>
<td>Interactions</td>
<td>Judgement</td>
<td>Judgement</td>
<td>Judgement</td>
<td>Automated process to minimize prediction error</td>
</tr>
</tbody>
</table>

**Domain knowledge**
Classical GLM: challenges traversing BVT

1. Gives full credibility to data, unless using judgement

2. Violating underlying GLM assumptions may produce misguided conclusions relative to variable selection

3. Judgement to determine interactions and doesn’t handle multicollinearity well
Traversing BVT

Variable interactions

Judgement/Manual

Robust/Automated

A:E with credibility weighting

Classical GLM with offset

Penalized GLM with offset

GBM

Variable selection / data credibility
Penalized GLM: how it works

• Develops coefficients using GLM with offset
  • Similar to simultaneous A:E adjustments

• Penalizes (shrinks) coefficients
  • Similar to credibility weighting in A:E study
  • Controls for overfitting
  • No penalty (full data weight) = Classical GLM
Penalized GLM: coefficients after shrinking

Illustrative shrinking of coefficients for GLM with offset

Coefficient (B)

Variable $x_i$
Penalized GLM: how it traverses BVT

• Data credibility and variable selection by shrinking coefficients
  • Automates decision by minimizing the cross validation prediction error

• Judgement to determine interactions
  • Better handling of multicollinearity
  • Challenge remains of navigating complex interactions
Cross validation: automates traversing BVT

K-fold cross-validation

- Use subset of data to develop coefficients
- Calculate error of predicted values on holdout data
- Average error across the k tests

<table>
<thead>
<tr>
<th>3-Fold</th>
<th>Test 1</th>
<th>Test 2</th>
<th>Test 3</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>1  33%</td>
<td>Holdout</td>
<td>Use</td>
<td>Use</td>
<td>Test 1</td>
</tr>
<tr>
<td>2  33%</td>
<td>Use</td>
<td>Holdout</td>
<td>Use</td>
<td>Test 2</td>
</tr>
<tr>
<td>3  33%</td>
<td>Use</td>
<td>Use</td>
<td>Holdout</td>
<td>Test 3</td>
</tr>
<tr>
<td>Calibration data</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>MSE on holdout data</td>
</tr>
</tbody>
</table>

Automated process!
Penalized GLM: how it traverses BVT

- Test range of penalties (data credibility)
- Chose penalty that minimizes prediction error
- Automated process tests thousands of models with a few lines of code!

<table>
<thead>
<tr>
<th>Overfitting</th>
<th>Balanced</th>
<th>Underfitting</th>
</tr>
</thead>
<tbody>
<tr>
<td>No penalty</td>
<td>Minimize error</td>
<td>Full penalty</td>
</tr>
<tr>
<td>Fully trust data</td>
<td>Credibility of data</td>
<td>Don’t trust data</td>
</tr>
<tr>
<td>(Classical GLM)</td>
<td>(Credibility of data)</td>
<td>(Benchmark)</td>
</tr>
</tbody>
</table>
Minimizing prediction error

• Objective to minimize MSE or SSE

• Classical GLM: \[ \text{SSE} = \sum (Y - x\beta)^2 \]

• Ridge: \[ \text{SSE} + \lambda \sum \beta^2 \]
  - Shrinks coefficients, but remains > 0
  - Helps with multicollinearity
Minimizing prediction error

• LASSO: \( \text{SSE} + \lambda \sum |\beta| \)
  • Can shrink coefficients to 0
  • Provides automatic feature (variable) selection

• Elastic net: \( 
\text{SSE} + \lambda \left( \alpha \sum |\beta| + (1 - \alpha) \sum \beta^2 \right) \)
  • Blend of Ridge and LASSO
  • Helps with multicollinearity and provides feature selection
    • \( \alpha \) controls the blend
    • \( \alpha = 0 \) then Ridge, \( \alpha \in (0, 1) \) then Elastic net, \( \alpha = 1 \) then LASSO
In-sample vs. out-of-sample tests

• Model objective
  - Minimize prediction error on future data

• Training error
  - Optimistic and decreases by adding variables

• Two fixes
  - In-sample tests: theoretical formula increases training error based on number of variables
  - Out-of-sample tests: directly estimates prediction error
In-sample vs. out-of-sample tests

<table>
<thead>
<tr>
<th>In-sample tests</th>
<th>Out-of-sample tests</th>
</tr>
</thead>
<tbody>
<tr>
<td>AIC, BIC, Adjusted $R^2$</td>
<td>Separate train/test datasets</td>
</tr>
<tr>
<td>p-values to prune parameters</td>
<td>k-fold cross validation</td>
</tr>
</tbody>
</table>

**Pros**
- Model selection using all data
- Fast to calculate
- No theoretical formulas
- Compare across algorithms

**Cons**
- Relies on theoretical formulas
- May misguide if assumptions violated
- Harder (or not possible) to compare across algorithms
- Computationally expensive
- Potential to misuse if not setup properly (information leak)
Traversing BVT

Variable interactions

Judgement/Manual

Robust/Automated

A:E with credibility weighting

Classical GLM with offset

Penalized GLM with offset

GBM

Variable Selection / Data Credibility
GBM: how it works

• Develops layers of “A:E” adjustments

• Layers of decision trees to minimize error
  • Slices data to create variable buckets
  • At each point tests every variable and possible slice to minimize error
GBM: how it traverses BVT

• Variable selection and interactions
  • Non-parametric model
  • Automates decisions by minimizing the prediction error
  • Handles complex interactions and provides information on variable importance

• Data credibility incorporated using cross validation to tune hyperparameters that control for overfitting
Model interpretability vs. accuracy

- Low interpretability, low predictive accuracy: Classical GLM with offset
- Low interpretability, high predictive accuracy: Penalized GLM with offset
- High interpretability, low predictive accuracy: A:E with credibility weighting
- High interpretability, high predictive accuracy: GBM
Why might we want lower accuracy?

• Stepping stone
  • Isolate changes from one model to the next

• Assumption format
  • Higher inference: multiplicative factors
  • Lower inference: sets of tables or seriatim

• Purpose / materiality
How can we measure uncertainty?
Live Content Slide
When playing as a slideshow, this slide will display live content

Social Q&A
Predicting High-cost members

• Using HCCI data (47M members over 6 years)
• Predicted which members are likely to be high-cost next year (>100K, 250K, 500K, 1M).
• Compared different models and sampling methods
• Showed that the more flexible xgboost works really well in our application.

Using an Asymmetric Cost Matrix for Wellness Interventions

• Depending on the relative savings of the intervention for high-cost members, we show which members (and how many) should be given the intervention.

Health Claim Costs by Disease

- Performed model selection:
  - AIC
  - Random forest
  - Bayesian parallel model selection

- Bayesian parallel model selection was the most accurate, but the most computationally expensive.

- With four candidate distributions, the most commonly used (gamma) distribution was not selected as the best for any of the diseases.

<table>
<thead>
<tr>
<th>Distribution Selected by RF</th>
<th>lognormal</th>
<th>gamma</th>
<th>Lomax</th>
<th>log-skew-t</th>
<th>AIC Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>lognormal</td>
<td>23</td>
<td>0</td>
<td>2</td>
<td>106</td>
<td>131</td>
</tr>
<tr>
<td>gamma</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Lomax</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>25</td>
<td>27</td>
</tr>
<tr>
<td>log-skew-t</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>161</td>
<td>162</td>
</tr>
<tr>
<td><strong>AIC Total</strong></td>
<td><strong>24</strong></td>
<td><strong>0</strong></td>
<td><strong>4</strong></td>
<td><strong>292</strong></td>
<td><strong>320</strong></td>
</tr>
</tbody>
</table>

Predicting Group Health Costs

- Built a flexible Bayesian nonparametric model to predict future year costs by group.
- Bayesian nonparametric model outperformed the standard model for 84% of renewal groups and 88% of new groups.

Predicting LTC Lapse and Mortality Rates

- Compared many different models to predict lapse and mortality rates in long-term care insurance.
- Found some methods which significantly outperformed the methods currently used in practice.

Questions and answers
Live Content Slide
When playing as a slideshow, this slide will display live content

Social Q&A