Predictive Analytics and Applications to Insurance

A presentation to the Middle Atlantic Actuarial Club By Ben Williams

Turf Valley



© 2017 Willis Towers Watson. All rights reserved.

Willis Towers Watson III'I'II

Disclaimer

This document has been prepared to support a presentation to the Middle Atlantic Actuarial Club, delivered on May 11, 2017, and should not be used for any other purpose

Predictive Analytics and Applications to Insurance

Contents

- 1. What is Predictive Analytics?
- 2. Why is it useful?
- 3. What is required to do it?
- 4. How is it done?
- 5. A case study



What is Predictive Analytics?









What is Predictive Analytics?

- Using a model, based on historical data, to make predictions about the outcomes of future events
- The process of fitting a model involves using various techniques (predictive modeling, machine learning, data mining) to identify what characteristics are driving the process of interest, quantify their impact, and separate signal from noise
- The model allows predictions as a function of characteristics of an individual (i.e. it allows predictions at a granular level)
 - A model might predict the number of claims for an individual auto risk (vehicle and driver(s)) as a function of age, sex, address, vehicle, prior claim behaviour, etc.)
- These granular predictions can be used to predict outcomes across a large number of individuals
 - The same model can be used to forecast the total number of claims across a portfolio of personal lines auto risks

Why is it useful (in insurance)?

- By its nature, insurance deals with the outcomes of uncertain future events
- Having a better understanding of expected outcomes can lead to better results for insurance companies (and their clients!)



Why is it useful (in insurance)?

Applications that I personally am most familiar with

Property and Casualty

- Loss Cost
 - Frequency and severity of claims
- Demand
 - Probability of a quote being accepted
 - Probability of a policy renewing
 - Probability of buying an additional product
- Claims Triage
 - Probability of a claim being "complicated"

Life and related

- Mortality
 - Making more granular assumptions
- Lapse
 - Making more granular assumptions
- Underwriting
 - Predicting UW Class on the basis of limited information
- LTC Morbidity
 - Incidence, Utilization and Termination

What is required to do predictive analytics?

A problem

- "What is going wrong?" or "What would I like to understand better?"
- What is the process of interest?

Data

- Big and not so big
- Internal vs external sources
- Structured vs unstructured
- A modeler
 - Someone who knows how to "do the math"
- Hardware
 - What is required depends on the volume of data (# rows and columns)
- Software
 - That fits the appropriate type of model (R, SAS, Emblem)
- A domain expert
 - Someone who knows how to interpret the results
- A way to implement the solution
 - A model that is not implemented in some way is not useful

How is it done?

An overview of some methodologies

- Generalized Linear Models
- Decision Trees (Classification and Regression Trees)
- Ensemble Methods (Gradient Boosting Machines, Random Forests)
- Topic Modeling
- Others

An overview of some methodologies: GLM

GLMs (Generalized Linear Models) take the following form:

Link function (g=h⁻¹) determines how variables are related

y = h(Combination of variables)

 Include variables that are predictive; exclude those that are not

- Simplify if necessary using groups and curves
- Include combinations of variables if necessary
- Best combination is determined by experiment, using a combination of statistics and judgment

Error

Reflects the variability of the underlying process and can be any distribution within a broad family (exponential distributions)

An overview of some methodologies: GLM

- Model output is of tables and vectors
- These are multiplied together to give expected incidence for a given profile
- We will see an example of a GLM in the case study

Base		0.0018							
Gender * IncurredAge				DurYear		Marital_Status		Prem_Class	
		Gender		DurYear		Marital Status		Prem Class	
		Female	Male			Married	1.0000	Preferred	0.7800
IncurredAge	35-	0.2306	0.2560	1	0.2391	Single	1.3799	Standard	1.0000
	36	0.0022	0.2740	2	0.7034			Substandard	1.2050
	37	0.4799	0.0047	3	0.8754				
	38	0.4019	0.0040	4	1.0000				
	39	0.3335	0.3261	5	1.0973				
	40	0.0902	0.1340	6	1.1993				
	41	0.2187	0.1097	7	1.2748				
	42	0.1170	0.1801	8	1.3023				
	43	0.2816	0.0029	9	1.3835	Incidence:	=D2*E24*H20*K8*N9		
	44	0.1519	0.3636	10	1.3775				
	45	0.2467	0.2510	11	1.3779				
	46	0.1252	0.1255	12	1.4146				
	47	0.2470	0.0704	13	1.4276				
	48	0.1695	0.1180	14	1.4555				
	49	0.2775	0.2222	15	1.4211				
	50	0.2487	0.1840	16	1.4073				
	51	0.2750	0.1164	17	1.3820				
	52	0.2644	0.2192	18	1.3269				
	53	0.2700	0.1705	19	1.2720				
	54	0.3547	0.1437	20+	1.2688				
	55	0.3385	0.2185						
	56	0.3094	0.2321						
	57	0.3579	0.2915						

 GLMs are the most widely used type of predictive model in the insurance industry, because they are easy to fit, easy to interpret and explain, and they allow a blend of statistics and business insight

An overview of some methodologies: Decision Trees

What do you think of when I mention **Titanic?**





I think of a well-known example of a decision tree!

An overview of some methodologies: Decision Trees

Survival of passengers of the Titanic



- Decision Trees (Classification and Regression Trees) determine a set of rules that segment observations
- The predicted value is average value within the terminal node
- It stops when no further splitting will improve results, or some stopping condition is satisfied

An overview of some methodologies: Decision Trees

Survival of passengers of the Titanic



Pros

- Outcome is a set of rules applying to different segments (easy to interpret, explain, and order)
- Good for identifying highdimensional features of data, as in the Titanic example (and these features can be used as variables in other modeling techniques)

Cons

 Highly dependent on the data, and therefore doesn't give predictions as good as some other techniques

An overview of some methodologies: Ensemble Methods

- To overcome this issue, generalizations of decision trees exist, such as:
 - Random Forests
 - Gradient Boosted Machines
- These are examples of <u>ensemble methods</u>, which use combinations of different underlying models



© 2017 Willis Towers Watson. All rights reserved. Proprietary and Confidential. For Willis Towers Watson and Willis Towers Watson client use only.

An overview of some methodologies: Random Forests

Roughly speaking, fitting a random forest involves

- Taking a large number of random samples of the data (with replacement)
- Fitting a simple tree on each sample
- The model is an average of these trees



The idea is that the combination of simple trees fit on different samples avoids overfitting to the data, and is more predictive than any single tree



- Output is the average of a set of tree-based rules
- Can give very good predictions, but interpreting the model can be complicated (some people would argue with this statement)

An overview of some methodologies: GBMs

Roughly speaking, fitting a Gradient Boosted Machine involves these steps:

- Fit a simple tree on a random sample of the data
- On another random sample
 - Calculate predictions according to previous tree
 - Calculate model residuals
 - Fit another simple tree on the residuals
- Repeat the process



		Same and the second sec	Share - Store - Color	State - State		Strate - State - State - State				Same and the second second
and the second second			and the second sec			and the second	Strady Strady			
Contraction of the second	and the second second second	and the second s	and the second s	Star of the Star of the Star	and the second		State - State - State - State - State	and the second sec	Same and the second	The second second
and the second second	and the second second	and the second sec	and the second second	and the second second second			and the second second		State - State - State - State - State	State of the second
State State	State State State	Star Street Street and Street	Star Street Branches	Same Street Street Street	and the second second second	State - State	State - State of the State of State	Star Street Street Street	and the second	Star Street Street
Contraction of the second	and the second second	The second s	The second se		and the second sec		Same and the second sec	Starting and the second second	the second s	Same and the second
	and the second			State of the State		Same State State State	Same of the second second		The second s	and the second second
Contraction of the second	and the second sec		Strate Strate Strate Strate					and the second	Strate Contraction of the second s	State of the state
	The second second	Solar - Sector - Sect	$\frac{\mu_{1}}{2\pi m^{2}h^{2}} = \frac{1}{2\pi m^{2}} \frac{1}{2\pi m^{2}} = \frac{1}{2\pi m^{2}} \frac{1}{2\pi m^{2}} = \frac{1}{2\pi m^{2}} \frac{1}{2\pi m^{2}} \frac{1}{2\pi m^{2}} \frac{1}{2\pi m^{2}} = \frac{1}{2\pi m^{2}} $		A CONTRACT OF A CONTRACT OF		Same - State -		Solution - Constant -	Same Street Contraction
A STATE OF A STATE OF	Martin - Martin (Martin - Martin Martin)	Whether States (1977) - States (1975)		Martin " and a first of the second se	Salah - 201 - 10 - 201 -	Second - Second Street St	Whether - Walter and State - S		Share and the second	Same Street Contractor
	The second s	States - States - States - States - States		and the second sec			With the second s	where the second se		Same and the second
and the second of the	San Street Barry Street Street	when the second s	Same Street Barrier	Same Street and a second second	Same and the second sec	Salar - Salar - Salar - Salar - Salar				
			Whether with the state of the state of the				Manage - Man		Strate - State - State - State - State	The state of the s
	Same and the second					wither with the second s	and the second	and the second second second	And a state of the second state of the	and the second second
and the second sec	Same and the second		Sparse - Sparse (State - Sparse - Spars		State - State - State - State - State	and a second	Same -	And States and States	Same - Same - Same - Same - Same - Same	The state of the state
	The state of the state of the	Same and the second	Shark - Share	State of the State	and the second s	Same and the second	and the second	and the second second		Same and the second
a state and a state of the	and the second second		The second	Same and the second	Same and the second		and the second sec	and the second second second	Sandara - S	
	Same and the second second			and the second sec	where the second s	and the second sec	Strate State State State		State of the State	State of the state
	The second s			and a state of the			Strate Strategy and Strategy and Strategy			and the second second
Contraction of the second	Same of the second s		$\sum_{\substack{\alpha \in [1,\infty] \\ \alpha \in [1,\infty] }} \sum_{\substack{\alpha \in [1,\infty] \\ \alpha \in [1,\infty] }} \sum_{\substack{\alpha \in [1,\infty] \\ \alpha \in [1,\infty] }} \sum_{\substack{\alpha \in [1,\infty] \\ \alpha \in [1,\infty] \\ \alpha \in [1,\infty] }} \sum_{\alpha \in [1,\infty] \\ \alpha \in [1,\infty] \\$	and the second s			Street Street Street Street Street	and the second second	and the second sec	and the second second
and the second second	and the second second	The second s	There was a state of the second state of the	and the second second second	with the state of	and the second sec		State of the second	Sources and states	and the factor
Contraction of the second	and the second		$=_{\substack{\alpha_1, \alpha_2, \alpha_3}} \cdots \geq_{\substack{\alpha_{n-1}, \alpha_{n-1}}}^{n-1} (\beta_1)^{\alpha_1 \beta_1 \alpha_2 \alpha_3 \alpha_3 \alpha_3 \alpha_3 \alpha_3 \alpha_3 \alpha_3 \alpha_3 \alpha_3 \alpha_3$	State - State			Same of the second second	the second s	and the second s	The second se
Same Street	and the second second						State - State State - State - State	Star Star Star	State of the second	and the second second
	And the second s	Strate Strategy		The second secon			Silve Street Street Street		Same and the second	State - State - State - State
and the second second	and the second s		State State State State				and the second second	and the second		
and the second	States - States - States - States - States	And the second s	$\frac{1}{ \mathcal{U} _{\mathcal{U}}\mathcal{U} _{\mathcal{U}}} = \frac{1}{ \mathcal{U} _{\mathcal{U}}} \frac{1}{ \mathcal{U} _{\mathcal{U}}}} \frac{1}{ \mathcal{U} _{\mathcal{U}}} \frac{1}{ \mathcal{U} _{\mathcal{U}}} \frac{1}{ U$	and the second s	man - 2 - Contraction		States of the second	ware and the second	State of the Contraction of the	and the second
a state and the state of the	and the second second	Salar Street Street Street	Tables - State	State of the second second	Salar Salar Salar	Strate - State - State - State - State	State of the second second	Sales and	The second se	Section of the sectio
and the second	THE OF THE PARTY OF	Solar - Solar	Same of the Real Property in the State	and the second second	Salar-Salar States	Same and the second	The state of the s		The second s	Same and the second
	San and the second s	Star - Star - Star - Star - Star	She of the state of	State of the second	and the second second	State - State - State - State	and the second sec	The second s	the second s	The second s
Contraction of	Same and the second second	State State State State State	The second s	and the second second	State of the second	Same and the second	State of the second second	and the second s	State of the second sec	and the second
	and the second second	Same State State State	Same State State State	San Street Street	and the second second	Same and the second second	State - State - State - State	State of the second	The state of the s	
and the second sec	and the second second	The state of the state of the state of the	The second secon	Same of the other states of the	Same and the second	and the second	State of the second second		States - States - States - States	States - States - Land
	and the second second	Salar - Salar - Salar - Salar - Salar	The second s	and the second se	man the R. C. Street		Same Street and Street and Street	AND STREET, DOOR ST.	Allowing the second sec	States of the States of the States
A CONTRACTOR OF A CONTRACTOR A	Same and the second second	and the second s	Same - State -	Salar - Sa	and the second second	San States and States	State of the second	and the second s		Same and the second second
and the second second	and the second second						and the second			Same of the second s
and the second	The second s	and the second sec					Same and the second		Martin - Constanting of the Constant of the South	The second s
a de la competencia d	and the second second second	and the second	Same State States				and a start of the start of the	When the second second second		and the second second
Contraction of the second	Same of the second s	Salar - Salar (Charles - Salar Salar	and the state of the second	and the second second	Star - Star - Star - Star - Star	and the second second	and the second second	Star Street Street	State of the second sec	and the second second
a state of the state of the	Star Starting and start	Same of the second of the		The second second second	State of the second second		States - Sign - States - Sign	State of the second state		Same and the second
a de la contraction	Mar - Sarah Congrad State	State of the second state	State State State State State	and the second second	and the second second second	and the second	The second secon	The second s	The second s	State of the State of the State
	William William Contraction of the	State of the second second		Same of the second s	and the second	Contraction of the second seco	State - State - State - State - State	and a start and a start of the	Same of the Contraction of the	Same and the second second
State States	Same and the second	State - State	State - State - State	and the second second	Same State State State	Same and the second sec	Strate State State State	and the second sec		
a - 2 - Contraction of the second	The second second	Shall - State	And the State of Stat	The second s	and the second second	Same State State State State	and the second second	and the second second	and the second	and the second
	Winter William Provide State	Sille Contraction	The second secon	and the second second	Same and the second second	Same and the second	and the second sec	State of the second	Same - State -	The state of the second

An overview of some methodologies: Topic Modeling

- Most insurers have text data such as loss adjuster and underwriter notes
- This data is unstructured, meaning that its (potentially rich and useful) content cannot easily or accurately be restated as structured columns of data
- Topic modeling is a data mining technique that determines, from a set of unstructured data, a list of topics (sets of co-occurring words), which can describe specific events or ideas
- It is then possible to determine, for each observation, to what extent each topic is present, and use this information to improve predictive models



Topic 1 includes the words "injure lower back spinal column vertebrae" Topic 2 includes the words "heavy box moving company"

An overview of some methodologies: Others

Penalized Regression

- Variations on the basic GLM set-up, including Lasso, Ridge Regression, and Elastic Net, each of which have their own properties
- For example, Lasso is good for understanding which, among a large number of variables, are most important in explaining the signal
- Neural Nets
 - Inspired by biological neural networks (i.e. the brain)
- Genetic Algorithms
 - Use concepts from the theory of evolution, such as random mutations and survival of the fittest, to create child models containing combinations of features of their parents, some of which will be better
- The choice of model will be determined by considerations such as the requirements of its end-users and the tools available
 - For example, claim models in regulated P&C lines will have to be explained to a regulator; most regulators are familiar with GLMs, but may not (yet) be familiar with GBMs

- The SOA carried out an inter-company LTC experience study in 2015. Over 20 companies contributed data, corresponding to
 - Over 15 million exposures
 - Over 170 thousand claims
- For this study, we developed the GLMs for the basic components of LTC morbidity:
 - Incidence (probability of going on claim in any given year)
 - Benefit utilization (expected proportion of benefits utilized per month on claim
 - Termination rates (probability of going off claim in any given month)
- The full study is available from the SOA
 - https://www.soa.org/experience-studies/2015/2000-2011-ltc-experience-basic-table-dev/
- Here, I'll present some results from a simplified incidence model

Variables included in the simplified model



Observed results (Substandard has lower incidence than Standard) are unintuitive



Rescaled Predicted Values - Prem_Class

Standardizing by other variables in the model leads to a more intuitive result



Rescaled Predicted Values - Prem_Class

 Standardizing by other variables in the model shows that effect of duration is not as strong as indicated by observed results



Rescaled Predicted Values - DurYear

This model also included some interactions



Interaction between Attained Age and Marital Status



Rescaled Predicted Values - IncurredAge

Predictive Analytics and Applications to Insurance

Contents

- 1. What is Predictive Analytics?
- 2. Why is it useful?
- 3. What is required to do it?
- 4. How is it done?
- 5. A case study

