



How to Effectively Use New & External Data

CAS Spring Meeting 2017

Presenter:
Maxime Lafleur-Forcier

May 23rd 2017



Agenda



- Why Do We Need Data?
- Traditional Data Usage & Collection
- Using External Data
- Geographical Information Science
- Unstructured Data
- Behavioral Data
- Client Relationship Management
- Conclusion

2

Why do we need data?



- Better Understand the Customer
 - Make better decisions:
 - What is the adequate premium?
 - Do we have the appetite for this risk?
 - Should we act on renewal for this policy?

Many of these decisions need to happen quickly in real-time

- Improve customer experience
 - Provide better advice to customers
 - Meet the insured's needs and expectations
 - Serve the customer in a timely yet efficient manner

Customer experience has recently become a growing area of focus

3

Traditional Data Usage & Collection



- Collected when Underwritten and used to describe the different entities



– Age, Gender, marital status, driving experience, Credit score, convictions, claims experience, ...



– Vehicle make\model, model year, modifications, ...



– Year built, amount of insurance, postal code, construction type, basement, roof type\age, ...

4

Interaction with the Customer



5

Using External Data



- Traditionally, links with external data for information unknown to customers, or little reliance on the customer's word:
 - Examples:
 - Credit Score
 - Claim History
 - Driving Convictions
- Greater availability of external data providers allows for the possibility to reduce handling time with pre-population
 - Property details from address
 - Customer information from driver's license
 - Vehicle characteristics and history from VIN



6

Considerations when selecting data providers 

- What are the potential benefits of the data?
 - Provide new uncorrelated information to other data points already available
 - Improve the predictive power of the claims model
 - Identifies adequately profiles with worst risk characteristics
 - Reduce the handling time with customers
 - Help me better serve or understand the customer
 - Improve the customer experience
- What are the costs?
 - Cost is not solely acquiring or purchasing data!
 - Is IT involvement required? What resources are required?
 - Does storage or new extractions to consume the data require system changes?
 - Are monitoring resources required to ensure data adequacy?
 - Maintenance fees?

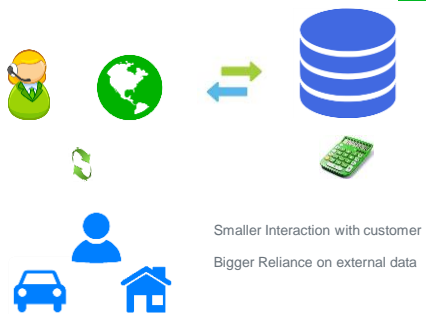
7

Considerations when selecting data providers 

- What are the risks?
 - Are there any other provider I don't know about I should consider?
 - Is the provider financially sound?
 - Is my data protected?
 - Are there any regulatory constraints that would prevent me from using the data?
 - Is the data accurate?
 - Are up-time and response time adequate?
 - What is the back-up plan if service not available or no-hit?

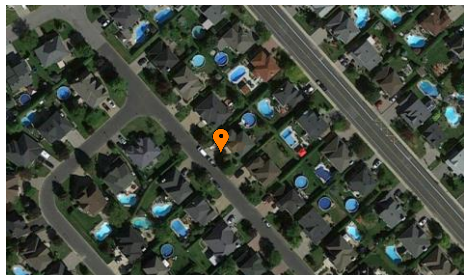
8

Interaction with the Customer 



9

Geographical Information Science



10

Territorial Analysis



- One element of the profile of house / customer or vehicle is the location, ie where the risk is located
- We typically describe this dimension at Postal Code/ Zip Code, FSA/County, Province/State ...
- To better describe the very high granularity of postal codes and low credibility, we try to find ways to regroup similar levels into homogenous groups
- Groups can be formed from:
 - Geographic proximity
 - Demographic data from census or other
- Demographics data can change over time, which might require major review after new census

11

Limitations of Postal Code



- New product availabilities might increase the need of territorial granularity
 - Flood:
 - Risk might be very different houses located close from each other
 - Postal code territory definition might not be able to properly differentiate those risks
 - Error factor even greater in rural areas, where postal code cover many houses that can be far apart
 - Other level of granularity which might be possible to capture
 - Distance from forest from forest maps
 - Elevation, slope, shape of the land
 - Infrastructure databases
- In Canada, Postal code definitions changes continuously which make maintenance of database difficult



12

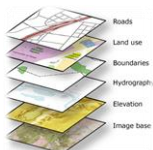
GIS – Geographical Information Systems



- A geographic information system (GIS) is a system designed to capture, store, manipulate, analyze, manage, and present spatial or geographical data

The Geographic Approach

- We use GIS to gather place-based information and organize it on a digital map. We then use the software analytical tools and capabilities to evaluate a decision.
- It is then possible to increase granularity of information about particular locations
- Once we understand the possible geographical constraints and possible consequences of our decision, we can then act in an informed and responsible manner.



13

Geolocalisation



■ There are multiple challenges to get an address geolocalised

- Need to have a clean, and recognized format of an address, which is often on a free form format, which can make analysis of active or historical policies more difficult
- Extensive database and expertise required to develop and maintaining a proper geolocalisation service, so choice of right provider is essential.
- Even if we can get the latitude and longitude, most systems do not have capacity to assign a territorial factor or set of rules from those features
- If want to be fully able to leverage the value of data, need to be able to have the process applied real time
- Need a contingency plan when address is not recognized and not able to return a proper geocoding

14

Interaction with customer



15

Unstructured Data



80% of the world's data is **unstructured**

Image credit: IBM 16

Unstructured Data



- Unstructured data is data that does not have a predefined structure, or data models
- It is sometimes defined as fat data, as it can be organized in millions of attributes.
- The most common type of unstructured data that is mined is text

One advantage
of being
disorderly
is that
one is
constantly
making
exciting
discoveries.

17

Unstructured Data



- Text Data:
 - from claims or underwriting notes
 - More and more software or open source algorithms exist to analyze and classify this data
- Voice Files:
 - phone calls from contacts with customers
 - Much bigger data in terms of size
 - Not as many tools exist to parse this data, but more and more available
 - Voice to text, sentiment analysis, footprint
- Reports:
 - digitalization of historical claims files
 - extracting information from different ordered reports
- Picture Recognition:
 - Can extract home characteristics from satellite or drone imagery
 - Used in claims to assess damages of a vehicle, or see if picture is reused in multiples claims files

18

Unstructured Data



■ Challenges with unstructured data: Text Mining

- When most data warehouses were created long time ago or more, unstructured data was not in a need to be transferred with other data elements, so data harder to extract today
- There might be multiple free form fields in the systems, so searching where the valuable information is stored can by itself be very challenging
- Traditional tools and statistical methods used cannot adequately extract the information
- Until new open-source software packages become recently available, software packages were more expensive and not performing that well

19

Text Mining



■ Steps for text mining projects (or any data project):

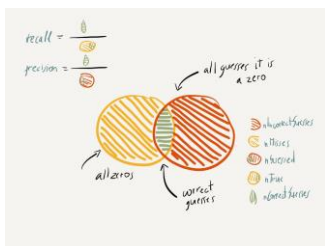
- Find the problematic you want to resolve
 - Identify fraudulent claims
 - Reason for Customers contacts (from SMS, Emails, ...)
 - Causes of loss
 - Customer satisfaction
- Evaluate the required data, which might be a mix of structured and unstructured
- Will my hardware be able to process the data
 - You might know too late if it will work or not
- What will the solution look like:
 - Modification of a question or process
 - Will there need to be a real-time decision to be made?
- What tools and techniques will I use
 - Open source vs Commercial software

20

Text Mining: Validation



- Precision is the percentage of items predicted relevant that are really relevant
- Recall is the percentage of really relevant items that are within the items predicted relevant

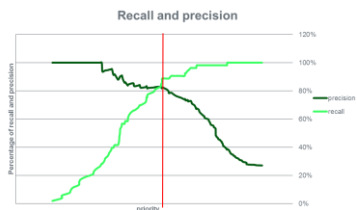


21

Text Mining: Validation



- Precision is the percentage of items predicted relevant that are really relevant
- Recall is the percentage of really relevant items that are within the items predicted relevant



22

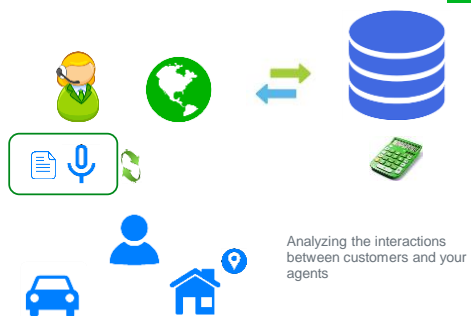
Voice Analytics



- Voice Analytics:
 - Fast growing field and more and more area of focus
 - One of main component is Voice recognition (voice to text).
 - Sentiments analysis
 - Voice/Speaker recognition
- With artificial intelligence, it is also possible to respond to interactions with actual insurance advices
- Challenges of Voice Analytics
 - Software might be expensive and hard to integrate with existing platforms
 - Many of existing software are a dead-end. Insights are harder to circle back with other data points

23

Interaction with customer



24

Behavioral Data



25

Behavioral Data



- Data produced as a result of multiple actions/decisions of individuals over a certain period of time
- With the arrival of mobile and smart devices, it has become possible to track more and more actions of the customers
- Behavioral data can capture huge amounts of data points, one every second or more
- Eg: of behavioral data:
 - Credit score
 - Usage Based Insurance
 - Digital Footprint
 - Social Media
- 2 Possible outcome of understanding behaviour:
 - New feature of a customer (score)
 - Incentive to modify the customer's behaviour to improve risk profile

26

Behavioral Data



- Usage Based-Insurance:
 - From a device connected to vehicle or from the cell phone data, while insured is driving, one record every second or so that is transferred with insurer
 - Score is usually computed real-time and results is shared back with customer so he can try to improve
 - Each customer will produce hundreds of gigabytes of data every year, so this is when we really need to have a platform for this data
- Balance between improving the models and insightful score to customers
 - Machine learning algorithm will be useful at identifying similar behaviors
 - But scores might become more difficult to share with customers or with regulators
- Also possible to improve the score by changing the behavior based on meta data:
 - Type of road
 - Relative traffic
 - Weather

27

Customer Relation Management



- To be able to effectively track previous actions undertaken with customers and optimize decisions, a 360 view of your customers is required
- CRM is leveraged to optimize the customer experience by integrating the proper technology and strategies; they enable the compilation of information about customers from different channels, and track the points of contact between them and the provider
- Decrease number of contacts with customer but optimize relevance and impact of each contact



31

Interaction with the Customer



32

Conclusion



- Currently, the quantity of available data is rapidly increasing and this trend will continue, and even accelerate
- More and more devices will get connected, data points collected, more information to analyze and insight to act upon
- Data is a very big asset for insurance companies, and for all companies in general
- Better understanding of the data can lead to better decisions and better customer satisfaction

33
