Report Cards for Manholes : Eliciting Expert Feedback for a Learning Task

Axinia Radeva[1], Cynthia Rudin[2], Rebecca Passonneau[1], Delfina Isaac[3]

[1] Center for Computational Learning Systems
Columbia University 475 Riverside Drive MC 7717
New York, NY 10115 USA

[2] Sloan School of Management
Massachusetts Institute of Technology
Cambridge, MA 02139 USA

[3] Consolidated Edison Company of New York
4 Irving Place
New York, NY 10003 USA

Motivation

The Secondary electrical system of NYC:
- Provides power directly to residential/ commercial buildings
- A dense network of structures (manholes and service boxes) and cables

Collaboration between CCLS/ConEdison/MIT
- Started March of 2007
- To prioritize electrical service structures for future inspection and repair.
- Apply data mining and machine learning techniques to disparate sources of raw noisy data

Heavy data processing and integration evaluated using:
- Structure profiling tool
- Visualization tool

Visualization of ECS tickets, structures and cables

Manhole Event Prediction

May 2, 2008 manhole fire Harvard Square Cambridge, MA

Important features can be easily checked trough the structure profile credited from the profiling tool:
- Total Mentions
- Total Mentions Recently
- Number of Times the Trouble Hole
- Number of Times the Trouble Hole Recently
- Number of Main Phase Cables

Machine learning model

Training goal: predict trouble holes for *serious* events in 2005, using data prior to 2005

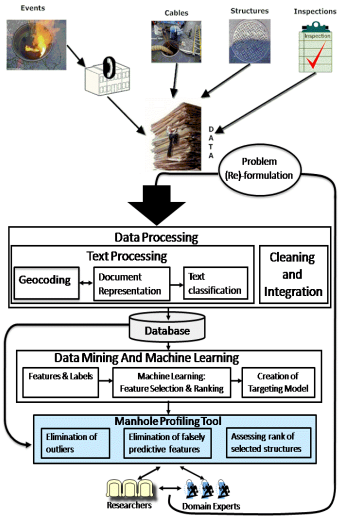
Testing goal: predict for 2006

For each event we define:

- whether the event was serious (*ticket classification*)
- which structures were the trouble hole and which structures were mentioned (*finding the trouble hole*)
- what characterized those structures, in terms of cables, inspections, and other factors

Main Sources of Data and Structure Ranking Process Diagram

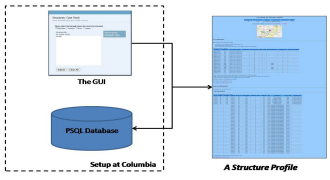
Data Source	Cols	Description
Emergency Control System (ECS) Tickets (Front)	26	Trouble tickets—reports of events, structured fields, hundreds of types (~20-30 for our models)
ECS Remarks (Back)	3	Trouble tickets, partially recorded in free text and partially automated log entries.
Electrical Incident (ELIN)	68	Serious events
Electric Shock Reports (ESR/ENE)	66	Electric shock events
Structures	32	Manholes, service boxes; locations
Covers	5	Structures with/without *vented* covers
Inspections	37	Inspection dates, repairs recommended, performed
Cables extract from Property Records	29	Type (main, service, street light), location, material, number of phase & neutral, size,



Secondary Stricture Ranking Process Diagram

Summary of Profiling Tool

1. From raw data we created a relational database integrating disparate noise sources of information
2. Profiling Tool interfaces directly with the database to create a structure profile (report card) for any of well known information for a given structure
3. Viewing profiles of structures on the top/bottom/middle of the list allows researchers and domain experts to evaluate the model



Architecture of the Profiling Tool Framework

Main Uses of Profiling Tool

Elimination of Outliers

Top portion of the list should not contain outliers

Initial model had *outliers* and relied on:
- Geocoding
- Trouble types

Current model No *outliers* and relied on:
- Geocoding
- Trouble class rules
- ECS remarks (*trouble hole*)

Elimination of Falsely Predictive Features

The profiling tool helps us locate *falsely* predictive features such as:

- Misleading anti-correlation between aluminum cables and manhole events
- Misleading anti-correlation between the serious events and age of the cables
- Misleading anti-correlation between the number of neutral cables and serious events

This misleading features were not used for the learning, instead they were incorporated into a corrections factor

Assessing Rank of Selected Structures

Profiling tool illustrates the inherent difficulties of locating vulnerable structures.

Characteristics of the highly ranked structures:

- Large number of cables
- Involvement in several events

Prototypical highly rankled structure:

Case Study for Structure MH54466
Its geographic coordinates are:
LAT: 40.744105 and LON: -73.995822

ECS Information
Here is the list of ECS tickets where either:
- the ticket occurred within 60 meters of the structure, or
- the structure was a trouble hole, or
- the structure was mentioned and the ticket occurred within 200 meters of the structure.

There are a total of 75 tickets.

Ticket	Type	Date	TH	Line	CR	Sh	60m	Mentioned
...	...	...	...	...	...	...	...	...
ME05015280	FLT	2005-08-28	*	37	*	*		*
...	...	...	...	...	...	...	...	...
ME05013163	LV	2005-07-22	*	32		*		*
ME05013027	ACB	2005-07-20	*	105		*	*	*
ME04008213	SO	2004-04-29		31		*	*	*
ME04002014	ACB	2004-01-29	*	4		*	*	*
ME03019735	ACB	2003-12-27	*	5		*	*	*
ME03012885	SO	2003-08-05		29	*		*	*
ME00008482	ACB	2000-06-20		55		*	*	*
ME00001101	UDC	2000-01-26		30			*	*
...	...	...	...	...	...	...	...	...

This structure was a trouble hole 5 times.
This structure has Solid-Metallic cover.

Inspection Information

The structure was not inspected at all.

Cable Information

From	To	#Cable	Size	Installed	Mtrl	Type	Dist(ft)
MH4240	MH54466	6	4/0	1935	RL	Main	317.547
MH4240	MH54466	6	4/0	1951	RN	Main	317.547
MH4240	MH54466	2	4/0	1951	BB	Main	317.547
MH60922	MH54466	2	4/0	2005	BB	Main	196.959
...	...	...	...	...	...	...	...
MH44471	MH54466	2	4/0	1967	BB	Main	74.279
SB42422	MH54466	2	4/0	1983	BB	Main	144.389
MH54466	SB42422	3	500	1983	RN	Main	144.389
MH54466		2	40	1926	RL	Service	
MH54466		2	#6	2007	RN	StLight	

The total number of cables connected to the structure is: 107

Not a completely prototypical highly ranked structure

Case Study for Structure SB137521
Its geographic coordinates are:
LAT: 40.868966 and LON: -73.920203

ECS Information
Here is the list of ECS tickets where either:
- the ticket occurred within 60 meters of the structure, or
- the structure was a trouble hole, or
- the structure was mentioned and the ticket occurred within 200 meters of the structure.

There are a total of 14 tickets.

Ticket	Type	Date	TH	Line	CR	Sh	60m	Mentioned
...	...	...	...	...	...	...	...	...
ME02007658	SMH	2002-06-30	*	13		*	*	*
...	...	...	...	...	...	...	...	...
ME99005563	ACB	1999-05-07	*	4		*	*	*
ME99004859	ACB	1999-04-15	*	7	*		*	*
ME99002789	SMH	1999-02-22	*	13		*	*	*

This structure was a trouble hole 4 times.
This structure has Solid-Metallic cover.

Inspection Information

InspectionDate	T1A	T1B	T2C&R	T2M	T2S	ReasonForVisit
2006-08-16	1	2	0	0	0	Ad-Hoc Inspection

Cable Information

From	To	#Cable	Size	Installed	Mtrl	Type	Dist(ft)
MH37518	SB137521	1	4/0	1929	BB	Main	39.303
MH37518	SB137521	3	4/0	1992	RL	Main	39.303
MH37518	SB137521	1	4/0	1934	BB	Main	39.303
MH37518	SB137521	6	4/0	1934	RL	Main	39.303
SB137521	SB37522	6	4/0	1934	RL	Main	171.51
SB137521	SB37522	1	4/0	1929	BB	Main	171.51
SB137521	SB37522	3	4/0	1929	RL	Main	171.51
SB137521	SB37522	1	4/0	1934	BB	Main	171.51
SB137523	SB137521	3	500	1999	RN	Main	42.200
SB137523	SB137521	2	4/0	1994	BB	Main	42.200
SB137523	SB137521	6	4/0	1994	RN	Main	42.200
SB137523	SB137521	2	4/0	1999	BB	Main	42.200
SB137521		7	4/0	1929	RL	Service	
SB137521		4	40	1919	RL	StLight	
SB137521		2	#12	1919	RL	StLight	

The total number of cables connected to the structure is: 48

Conclusions

The big challenges of the project:
- massive quantities of data, tremendous amount of noise
- working with free-text containing telegraphic/fragmentary language and technical jargon
- no operational definition of what we are trying to predict ("serious events")
- unclear how to formulate as a machine learning task (choose a bipartite ranking framework, meaningful time periods)

The profiling tool assisted with:
- eliciting expert feedback
- providing an explanation and refinement of the factors within the model
- ensuring correctness
- justifying the approach to management
- assisting with actionability of the result.

We identified many vulnerable manholes that Con Ed has not previously identified. Currently, ConEdison is using Manhattan model for prioritizing their repair work.
In Manhattan and Brooklyn we have ranked a total of 115,083 structures.

Acknowledgments

This work was supported by a grant from Con Edison. Thanks to our collaborator Steve Ierome. Also thanks to Dave Waltz, Roger Anderson, Serena Lee, Robert Schimmenti, Artie Kressner, Maggie Chow, Ashish Tomar, Nandini Bhardwaj, Zhi An Liu, Charles Collins, Diego Fulgueira, Jing-Yeu Miaw, and Sara Stolbach Rosenthal.