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
[3] Consolidated Edison Company of New York, 4 Irving Place, New York, NY, 10003, U.S.A.
Email: axinia@ccls.columbia.edu, rudin@mit.edu, becky@ccls.columbia.edu, isaacd@coned.com

Abstract

We present a manhole profiling tool, developed as part of the Columbia/Con Edison machine learning project on manhole event prediction, and discuss its role in evaluating our machine learning model in three important ways: elimination of outliers, elimination of falsely predictive features, and assessment of the quality of the model. The model produces a ranked list of tens of thousands of manholes in Manhattan, where the ranking criterion is vulnerability to serious events such as fires, explosions and smoking manholes. Con Edison set two goals for the model, namely accuracy and intuitiveness, and this tool made it possible for us to address both of these goals. The tool automatically assembles a "report card or "profile" highlighting data associated with a given manhole. Prior to the processing work that underlies the profiling tool, case studies of a single manhole took several days and resulted in an incomplete study; locating manholes such as those we present in this work would have been extremely difficult. The model is currently assisting Con Edison in determining repair priorities for the secondary electrical grid.

1. Introduction

In March of 2007, we began a collaboration with NYC's power utility, Con Edison, with the goal of prioritizing manholes (or more generally, manholes and service boxes, or "structures") in the secondary network for future upgrade and repair. Con Edison started a comprehensive inspection program in 2004; before this point, repairs were made mainly in response to an event (for instance, a power outage). The inspection program has generated a large number of possible future repairs that Con Edison would like to make pre-emptively, however, this list is long (and growing), with the consequence that repairs need to be prioritized carefully. Con Edison did not have an existing statistical method to determine repair priority, so our task was to assist with creating one. The goal was to rank structures with respect to the likelihood of experiencing a *serious* manhole event, meaning an explosion, fire, or smoke in a manhole. This paper describes one of the main mechanisms for reviewing the ranked list: a *profiling* tool for manholes.

Profiling, that is, summarizing information regarding specific manholes, is used in three ways we discuss here: elimination of outliers, removal of falsely predictive features by finding correlations that are not causal, and demonstration that data mining is valuable for determining repair priority. Con Edison hypothesized that the kind of vulnerable structures we present here did exist; however without a model predicting which structures are vulnerable, and the ability to review a structure's characteristics as provided by our profiling tool, it would have been impossible to identify vulnerable structures or verify their vulnerability.

At the start of the project, we tried to construct case studies of manholes that had been involved in past serious events. For example, we chose a large manhole explosion in a popular area of midtown New York City and tried to find all information about the "trouble hole" for this event, meaning the primary structure implicated in the event. We faced many hurdles working with historical data from several different sources within Con Edison. Our data consisted mainly of ten years of Emergency Control System (ECS) trouble tickets, which are entered mostly by call center personnel, plus several raw cable tables with records dating as far back as 1889. From the raw datasets, we attempted to determine whether the manhole had been involved in prior events indicating a problem, how many prior events had occurred nearby, and what cables were inside the manhole. Even after several days searching, our information was highly incomplete. We found records of several cables that were connected to the manhole, but had no basis to judge the state of those cables, e.g., on age or material. We were unable to compare the number of cables in the structure to the relative number of cables in other structures. In fact, we were unable to determine what percentage of the cables entering the structure we had located. Regarding nearby events, we could find no other clearly relevant tickets within that block. While it is likely there were several, we were using street addresses from the ECS tickets which were often inconsistent or incomplete. Since it was not possible to evaluate the vulnerability of an individual structure, it was not possible to create an effective success metric for data mining.

The project has been underway for over two years, and now we can construct a complete structure profile, or "report card," of any structure within Manhattan through a single command, using the Structure Profiling Tool that is the focus of this work. The profiling tool executes a set of database queries and organizes the results into a summary of everything used in determining the vulnerability of a structure: the history of events, the number, type and material of cables connected to the structure, the structure's inspection history, etc. It presents the results of all stages of the modeling process at a glance, including data cleaning and processing, interpretation of trouble tickets and incorporation of prior knowledge. The tool, along with the statistical results, allows a domain expert to easily assess the learning model's accuracy in pinpointing vulnerable structures.

Several groups have found ways to combine domain expertise with statistical models to assist with power grid operations (including our colleagues [4]); for instance, Lambert-Torres et al. [5] were concerned with the manual checking of electrical load data from experiments. Their system assists technicians by automatically identifying patterns of interest. Dola and Chaudhury [3] used decision trees to classify faults and delays at substations. Evaluative profiling tools are useful in many application areas, e.g., mining emergency department data at a medical center [2] or designing customer profiles for marketing applications [1].

In Section 2 we describe the raw Con Edison data, and in Section 3 an outline of the processing and modeling for constructing the ranked list. Section 4 illustrates the ways in which evaluation can be performed via profiling.

2 Raw Con Edison Data

We have heterogeneous data regarding *events, cables* and *inspections*. The most valuable information we have is the ECS table, which contains data from past events (e.g., "no lights," "flickering lights," "smoking manhole"). ECS is Con Edison's trouble ticket database, which was started in the 1970's and became electronic after Hurricane Gloria in the late 1980's.[1] Each ECS ticket includes a collection of notes about an event (called the "remarks"), partially recorded in free-text by Con Edison dispatchers and partially automated log entries. The free-text may include the name of the customer reporting the call, the type of equipment missing an electrical phase, whether Con Edison crews were able to park their vehicle in the area, whether someone reported smoke, etc. The ticket system is mainly used for recording logistics and record keeping, so a ticket may not contain a description of the event it represents. Tickets can exceed several hundred lines and may contain misspellings, irregular shorthand notations, and formatting problems. For instance, there are over 40 different ways that "service box" is written, e.g., "`S/B`", "`BOX`", "`SERV BX`." Part of an example ticket is provided in Figure 1.[2] The ticket contains a collection of key-value pairs, containing important information such as the location of the event as a (noisy and imprecise) street address, date of event, and trouble type as a 2-3 letter code (such as "ACB" for AC burnout, "SMH" for smoking manhole, or "MHF" for a manhole fire).

ticket	ME031019735
house_no	
cp	W
street	24
arty	ST
trouble type substring	ACB
recvd_datetime	2003-12-27
cross_st	7 AV

remarks:

```
FRANK CURTIS REPORTS IN M544466 S/W/C W.24ST &
7TH AVE CREW
FOUND AC B/O'S...........................FH
12/27/03 00:33 MDEHINOJOS DISPATCHED BY 11511
12/27/03 01:15 MDEHINOJOS ARRIVED BY 55988
12/27/03 02:50 HINOJOSA/#9 REPORT ALL BURNOUTS
CLEARED .
12/27/03 02:53 MDEHINOJOS COMPLETE BY 55988
```

Figure 1. Partial ECS trouble ticket. The trouble type is ACB (alternating current burnout).

We were also given cable data and inspections data. The inspections data required only some interpretation and cleaning. The cable data, on the other hand, were very noisy. We went through several iterations of requesting cable data from different departments within Con Edison; important fields were often missing, and cables often could not be matched to the corresponding structures.

[1] Several of the Con Edison engineers still refer to ECS tickets as "B-Tickets" in reference to the time when tickets were written directly onto carbon copy and the "B" copy was the one that was filed.

[2] Ticket identifiers and structure identifiers have been made anonymous, and street addresses have been altered.

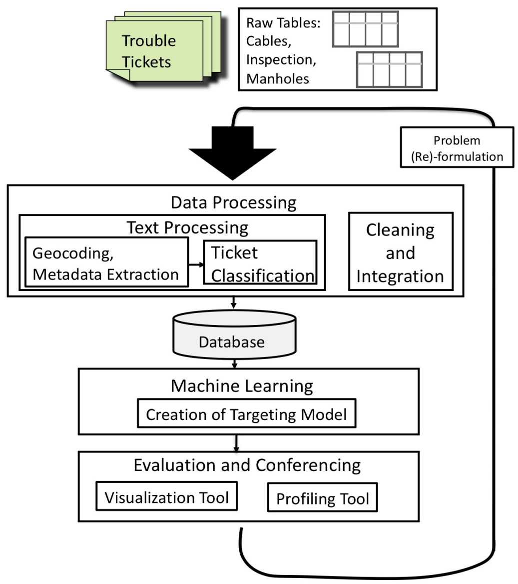

Figure 2. Process Diagram

3 Manhole Event Prediction

The process for manhole event prediction is shown in Figure 2. The focus of this paper is the evaluation of the model; details about the model itself can be found in [8]. The bulk of the initial processing, leading to the creation of a relational database, concentrates mostly on the trouble tickets. Since we aimed to predict which structures will incur serious events, we needed to define for each event: 1) whether the event was serious (ticket classification), 2) where the event occurred, that is, which structures were the trouble hole and which structures were otherwise involved, 3) what characterized those structures, in terms of cables, inspections, and other factors.

Ticket Classification: We located terms (or variations of terms) appearing in certain contexts to support metadata tagging, for instance indicating whether work was performed (“`SHUNTS`,” “`CLEARED`,” “`CUT FOR REPLACEMENT`”), and whether the ticket represents a serious event (“`FIRE`,” “`SMOKE`,” “`BLOWN`,” “`WIPEOUT`”). We computed the number of free-text lines in the remarks as a proxy for the amount of work performed. These features were used to classify tickets as representing a serious event or potential precursor event (see [6]).

Finding the Trouble Hole: From the ticket remarks, we extracted names of structures (taking into account the possible variations). This was combined with trouble hole information found in several of Con Ed’s specialized databases. The list of (ticket, structure) pairs was pruned by checking that a geocoded version of the ticket address is close to the geographic coordinate of the structure. A structure is a trouble hole for the event if it is the most often mentioned structure in the ticket or if it appears in one of the specialized databases. Trouble hole information was used for both features and labels for learning. If a structure was mentioned in a ticket but was not the trouble hole, it indicates that the structure was tangentially involved in the event; this information was used to construct features for learning.

Cleaning and Integration: In several tables, structures are indexed only by a noisy structure name and grid coordinate, e.g., (S/BOX 2415, 45-G). Specialized data cleaning and pattern matching was used to assign Con Edison structure identifiers, in order to construct a true relational database summarizing events, cables, inspections, etc.

Machine Learning Model: From the database, we formed features and labels for the *rare event supervised ranking* task. The model was trained to predict events in 2005 and tested on events in 2006. For training and testing, a positive label was assigned to the trouble holes for serious events within the specified year, and other structures were given negative labels. Features were derived from the properties of the structure and its events and inspections prior to the given timeframe. To rank the structures, we implemented supervised ranking methods that are designed to concentrate on the top of the list [8, 7]. These models find a linear combination of features to produce a real-valued score for each structure, and this score is used to rank the structures. The coefficients of the linear combination are chosen to maximize a weighted version of the Area Under the ROC Curve that favors the left side of the curve (the top of the ranked list).

Most of the machine learning features we developed are good individual predictors, although many (particularly the inspection features) are not. Important features include: the number of past precursor and serious tickets in which the structure was mentioned, the number of past precursor and serious tickets for which the structure was the trouble hole, and the number of main phase cables entering the structure.

The final ranked list (called the “targeting model”) incorporates a correction to the machine learning results for domain knowledge that is not present in the database. This correction stems from domain experts’ hypothesis regarding correlations that are not causal, as discussed in Section 4.2.

4 Structure Evaluation

The Structure Profiling Tool produces a report for a single structure based on the stages of the modeling process, containing: a summary of all ECS tickets within 60 meters of the structure and all additional tickets that the structure was mentioned in, a roster of all cables entering the structure, a list of the structure’s past inspection results, and the cover type. This profile allows a domain expert to approximately judge the relative vulnerability of a structure at a

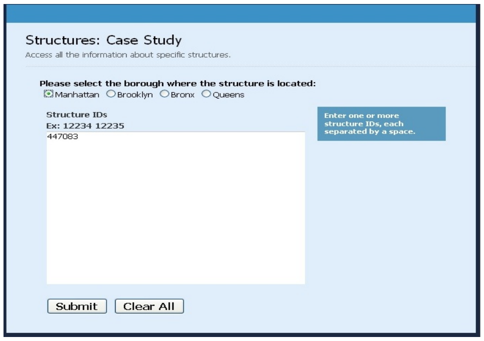

Figure 3. Profiling Tool Interface

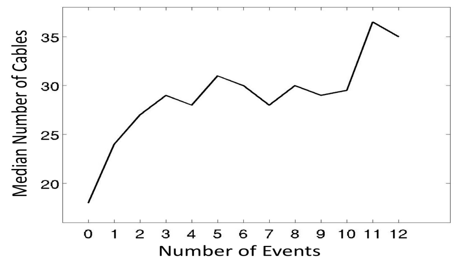

Figure 4. Median Number of Cables vs. Number of Events

glance. In order to receive the profile for a given structure, the structure number is entered into a web-based interface (see Figure 3) to our POSTGRESQL database.

The labels for the supervised ranking task depend heavily on interpretation of the trouble tickets, meaning that there is no gold standard evaluation measure. We focus on three main methods of evaluation: the presence of outliers, the use of profiles for eliminating falsely predictive features, and the search for potentially vulnerable structures. The ranking model has performed well on a "blind" prediction task described in [8], in which significant improvement over random guessing was shown. However, this statistical evaluation is not enough: the features of the model need to be meaningful, and it should prioritize structures that truly have signs of vulnerability.

4.1 Elimination of Outliers

"Outliers" are examples that receive an extreme score (either high or low), generally due to noise in the features. The top portion of the list is prioritized for repair and should not contain outliers. In our model, highly ranked structures generally possess characteristics that we have found to be statistically predictive, namely a large number of cables and involvement in several tickets. Of the top 1000 structures, all but one has above the average number of cables, and that structure was highly ranked for a good reason; it was involved in 9 events, has many aluminum cables (which indicate potential vulnerability), and has never been inspected. In fact, all but 2 of the top 100 structures are within the top 10th percentile for number of cables, and for the top 100 structures, the average number of events is almost 5 times higher than for the rest of the list (3.39 events vs. 0.7 events). A mark of the quality of data processing is that features processed independently are correlated; Figure 4 illustrates this by showing the median number of cables for structures involved in a given number of events. Note that 95% of structures are involved in ≤ 3 events.

Let us profile a prototypical structure ranked within the top 100, namely Manhole 544466 named in the ticket of Figure 1, located in the Chelsea neighborhood of Manhattan. We have found out also from our blind prediction test on 2007 data that this structure was the trouble hole for three SMH (smoking manhole) tickets in 2007; indeed it was correct for this structure to have been given a very high priority.

Table 1 gives an abbreviated (and slightly reformatted) version of the first piece of the report, which is a summary of the structure's involvement in past events. Each row of the table represents a ticket involving the structure.[3] The columns include: the ticket identifier; "trouble type," e.g., FLT (flickering lights), ACB (AC burnout), LV (low voltage), UDC (underground DC), SO (side off, which is a partial outage); date of the event; identifier for whether the structure was the trouble hole for the ticket; number of free-text (manually typed) lines in the remarks as an estimate of seriousness; and an identifier for the variations of "`SHUNTS`" or "`CLEARED`" in the remarks, indicating repair work. These terms appear in 36% of tickets.

MH 544466 was mentioned in nine events, and it was the trouble hole for five of them. This is unusual: only 0.004% (204/51219) of structures have participated in more than 9 events. Furthermore, MH 544466 has 107 cables; the mean number of cables in a manhole is 27 cables and only 593 structures out of 51219 (1.1%) have more than 100 cables. Furthermore, MH 544466 has not yet been inspected (recall that Con Edison's inspection program is new), which is an additional reason for it to be prioritized.

The current model is well-tuned to minimize the presence of outliers, but earlier versions were not. The initial version of the model relied only on geocoding and trouble type, and not on the ECS remarks. The initial model aimed to predict whether a structure would be within 30 meters of a serious event (rather than being the trouble hole for a serious event). This is an easier learning problem because it

[3] In the full report there are 66 other events within the 60 meter radius of the structure, but these events do not directly involve the structure and have been omitted.

Table 1. Events For MH 544466

Ticket	Type	Date	TH	Line	Sh
ME05115280	FLT	2005-08-28	*	37	*
ME05113163	LV	2005-07-22	*	32	*
ME05113027	ACB	2005-07-20	*	105	*
ME04108213	SO	2004-04-29		31	*
ME04102014	ACB	2004-01-29	*	4	*
ME03119735	ACB	2003-12-27	*	5	*
ME03112885	SO	2003-08-05		29	
ME00108482	ACB	2000-06-20		55	*
ME00101101	UDC	2000-01-26		30	

was easier to satisfy the criteria for a positive label; statistically, the model yielded very good results. However, the most highly ranked structures were those near "hotspots," or areas with many nearby events. Some of the highly prioritized structures would not be considered vulnerable (and would not be ranked highly in our current model). For instance, there is another structure at the same intersection as MH 544466 that has only 20 cables and has never been involved in an event. The initial model would have ranked both structures highly. The current model ranks the first structure highly, and the second structure at 34K/51K, or about three-quarters of the way down the list.

4.2 Elimination of Falsely Predictive Features

We discussed earlier a correction factor in the model for the presence of *falsely* predictive features. These are factors that are correlated with the presence of serious events, but are not causes of such events. These factors were discovered by troubleshooting earlier models using the profiling tool. One such example involves aluminum cables, which Con Edison considers to be an important factor in determining vulnerability. Since our cable data is from a snapshot in time, it is often impossible for us to determine whether a structure had aluminum cables at the time of the event: after the event, the cables are replaced and only the replacement cable appears in the database. This causes a misleading anti-correlation between aluminum cables and manhole events. A similar misleading correlation occurs between serious events and cable age, since many of the structures that have recently experienced serious events tend to contain newer cables, and another misleading correlation occurs with the number of neutral cables, since replacement bundles of cable tend to contain more neutrals. The profiling tool helps us locate these misleading features, which are not used for learning, and incorporated instead into the correction factor.

4.3 Pinpointing Vulnerable Structures

An important goal for us was to be able to pinpoint vulnerable structures that had not already been repaired. Con Edison performs necessary repairs for every event, so we were concerned that this approach might find only structures that were known to Con Edison. However this did not prove to be the case, and we were able to make recommendations useful for Con Edison's inspection and pre-emptive replacement programs; for instance, at the time of our inspections data, 35% of the top 1000 ranked structures had not been inspected, 32% had not yet been replaced with a vented cover through Con Edison's cover replacement program, and 20% have at least 5 aluminum cables.

The service box chosen for the next structure profile, SB 137521, is located on the northern tip of Manhattan, is ranked within the top 5% of the list, but is not a completely prototypical highly ranked structure: it has a sufficient number of cables and mentions in tickets, however, it was involved mainly in serious events (unlike MH 544466). It possesses several non-statistically predictive properties that are possible signs of vulnerability. Blind test results indicate that our prediction was accurate, in that this structure was the trouble hole for a smoking manhole event in 2008.

Table 2 (top) shows the structure's history of involvement in past events. It was the trouble hole for three events in 1999, including a smoking manhole and two AC burnouts. In all three tickets, burnouts were reported to have been cleared, meaning that problems with the structure's cables had accumulated between events. This demonstrates that structures that have been recently repaired are still potentially vulnerable to events, likely due to area-wide problems. The structure was again a trouble hole for a serious smoking manhole event in 2002 (and then again in 2008). The structure's only inspection indicates that "Tier 1" (necessary) repairs were made/recommended at the time of the inspection, which is also a sign of possible vulnerability (Table 2 middle).

The cable roster in Table 2 (bottom) indicates that SB 137521 has a large number of cables (48) considering that it is a service box. The mean number of cables in a service box is 23, and only 1571 service boxes out of 36285 (4.3%) have more than 48 cables. The type of cable is given in the rightmost column "M/S", as either "Main," connecting two structures, "Service" or "St. Light" connecting the network to the services or streetlights. The insulation material "BB" stands for BARE & BARE which indicates neutral cables (which do not carry current). What is most interesting is that the structure possesses several extremely old cables, namely, a set of service cables from 1919 and mains from 1929 and 1934. This is not uncommon, a full 4.6% of cables in Manhattan were installed before 1930; however, the presence of old cables is a reason for the structure to be highly ranked.

Another important purpose served by the Structure Profiling Tool is to illustrate the inherent difficulty in locating vulnerable structures. Consider the profile of another struc-

Table 2. Structure Profile: SB 137521

Information about past events:

Ticket	Type	Date	TH	Line	Sh
ME02107658	SMH	2002-06-30	*	13	
ME99105563	ACB	1999-05-07	*	4	*
ME99104859	ACB	1999-04-15	*	7	
ME99102789	SMH	1999-02-22	*	13	*

This structure has a Solid-Metallic cover.
Inspection information:

InspectionDate	T1	T2
2006-08-16	3	0

Cable information:

From	To	#Cable	Installed	Mtrl	M/S
SB137521	MH37518	1	1929	BB	Main
SB137521	MH37518	3	1929	RL	Main
SB137521	MH37518	1	1934	BB	Main
SB137521	MH37518	6	1934	RL	Main
SB37522	SB137521	6	1934	RL	Main
SB37522	SB137521	1	1929	BB	Main
SB37522	SB137521	3	1929	RL	Main
SB37522	SB137521	1	1934	BB	Main
SB137521	SB37523	3	1999	RN	Main
SB137521	SB37523	2	1949	BB	Main
SB137521	SB37523	6	1949	RN	Main
SB137521	SB37523	2	1999	BB	Main
	SB137521	7	1929	RL	Service
	SB137521	4	1919	RL	StLight
	SB137521	2	1919	RL	StLight

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

ture, SB 3412, that was the trouble hole for a serious manhole fire in 2006. This manhole fire would have been difficult to predict, since the structure had no prior involvement in events whatsoever, no pending repairs, only 23 cables, none of which are particularly old, and no currently present aluminum cables. In fact there was only one (not-serious) ticket within 60 meters of the structure within the 12 year span of our data. Reading the ticket for the manhole fire, it is discussed that the cables replaced were not aluminum, and the suspected cause of the incident was insulation breakdown; it is exactly the type of event that our model was designed to predict, but without any obvious way to predict it. There are several structures each year that are trouble holes for serious events, but without any clear predictive factors. In 2007 for instance, $\sim$20% of the serious events (9/44) possessed fewer than 20 cables.

5 Final Remarks

Con Edison insisted on "root causes" over "black boxes," meaning a model that was interpretable as well as predictive. We have found that Con Edison managers seem to respond as much to profiles of highly ranked structures as to statistical results based on rank tests (ROC Curves and rank statistics - our usual means of evaluation). With each new version of the model, we submit not only the model results, but also profiles of structures from the top, middle and bottom of the ranked list. The profiling tool (deemed the "report card" tool by Con Edison) has assisted with eliciting expert feedback, providing an explanation and refinement of the factors within the model, ensuring correctness, justifying the approach to management, and assisting with actionability of the result. Many machine learning applications to energy related (and other) tasks are likely to be subjected to the same type of qualitative assessment, namely that results that can be given meaningful causal interpretations. A carefully designed profiling tool can be extremely useful for such tasks.

Acknowledgment

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, and Diego Fulgueira.

References

[1] G. Adomavicius and A. Tuzhilin. Using data mining methods to build customer profiles. *Computer*, 34(2):74–82, 2001.

[2] A. Ceglowski., L. Churilov, and J. Wassertheil. Knowledge discovery through mining emergency department data. In *Proceedings of the 38th Annual Hawaii International Conference on System Sciences (HICSS '05)*, pages 142c–142c, Jan. 2005.

[3] H. Dola and B. Chowdhury. Data mining for distribution system fault classification. In *Proc. 37th Annual North American Power Symposium*, pages 457–462, Oct. 2005.

[4] P. Gross, A. Boulanger, M. Arias, D. L. Waltz, P. M. Long, C. Lawson, R. Anderson, M. Koenig, M. Mastrocinque, W. Fairechio, J. A. Johnson, S. Lee, F. Doherty, and A. Kressner. Predicting electricity distribution feeder failures using machine learning susceptibility analysis. In *The Eighteenth Conference on Innovative Applications of Artificial Intelligence IAAI-06*, Boston, Massachusetts, 2006.

[5] G. Lambert-Torres, D. da Silva Filho, and C. de Moraes. Intelligent data mining in power distribution company for commercial load forecasting. In *IEEE Power Engineering Society General Meeting*, pages 1–6, June 2007.

[6] R. Passonneau, C. Rudin, A. Radeva, and Z. A. Liu. Reducing noise in labels and features for a real world dataset: Application of NLP corpus annotation methods. In *Proceedings of the 10th International Conference on Computational Linguistics and Intelligent Text Processing (CICLing)*, 2009.

[7] C. Rudin. The P-Norm Push: A simple convex ranking algorithm that concentrates at the top of the list. *Journal of Machine Learning Research*, 2009. In press.

[8] C. Rudin, R. Passonneau, A. Radeva, H. Dutta, S. Ierome, and D. Isaac. Predicting manhole events in Manhattan : A case study in extended knowledge discovery. Accepted for publication to Machine Learning, 2009.