Visualization of Manhole and Precursor-Type Events for the Manhattan Electrical Distribution System

Haimonti Dutta[1], Cynthia Rudin[1], Becky Passonneau[1], Fred Seibel[1], Nandini Bhardwaj[1], Axinia Radeva[1], Zhi An Liu[1], Steve Ierome[2], Delfina Isaac[2]
[1]Columbia University, Center for Computational Learning Systems, New York, NY 10115
[2]Consolidated Edison Company of New York, New York, NY 10003
{haimonti | rudin | becky}@ccls.columbia.edu, fseibel@sprintmail.com, nb2338@columbia.edu, axinia@hotmail.com, zl2153@columbia.edu, {ieromes | isaacd} @coned.com

ABSTRACT

We present a visualization framework for analyzing the Consolidated Edison Company of New York (Con Edison) trouble tickets for the Manhattan electrical distribution system. The Con Edison Emergency Control System (ECS) is a work management tool that documents all events that occur in the electrical distribution system. The trouble ticket generated from ECS is a record of an event affecting the secondary (low-voltage) electrical distribution system, such as a manhole fire, manhole explosion, smoking manhole, no-light event, flickering light event, side-off partial outage, or burnout. The visualization tool outlined here is used alongside our preliminary statistical and machine learning work for predicting future manhole events. ECS tickets stored in our PostgreSQL database are displayed using Google Earth's satellite images of Manhattan as a backdrop. The ability of this tool to display events relative to the surrounding buildings has already yielded some highly promising directions for our ongoing analysis.

INTRODUCTION

The Con Edison Emergency Control Systems (ECS) "tickets" database is a rich resource for data mining containing approximately 1 million tickets. Each ECS ticket (or "trouble ticket") is a report of an event affecting the New York City electrical distribution system as recorded by a Con Edison dispatcher. The category of "manhole events" includes manhole fires, manhole explosions and smoking manholes. It is not an easy task to predict manhole events; however, if areas of vulnerability to these events could be identified, it would assist Con Edison in prioritizing inspections, repairs and follow-up work. It is precisely our goal to pinpoint areas of vulnerability based on historical ECS data.

The visualization tool presented in this work provides the first graphical view of historical ECS tickets. Such a method for graphically viewing data and results is an essential tool in many ways, for instance, for gaining physical intuition for underlying geospatial trends, for identifying and analyzing historical "hot spots" (i.e., localized concentrations of events) and even for visually determining useful density-of-events estimates. Also, we have found it extremely useful for viewing the relationship of events to the buildings, since it is not possible to understand this relationship any other way. An important example of this is the corner shown in Figure 3 (top), which is located in Manhattan's Upper East Side. It is very clear that the "hot spot" surrounds only a particular set of buildings, and is not characteristic of the nearby area. This type of important geospatial intuition would be lost without the tool that we describe here.

In its present form, the visualization tool displays all trouble tickets in a selected area. There are two parts to the application – (1) The Apache Tomcat (version 6.0.13) server and (2) Google Earth that needs to be installed on the client end. The server, hosted at Columbia University, deploys a java servlet application. The computations, database access and retrieval operations take place on the server side. Google Earth enables the actual visualization on the client side. These two components communicate with each other using the KML document format, an XML-based language schema for expressing geographic annotation and visualization on Web-based online maps and earth browsers. We will describe the ECS database, the architecture of the system, and a description of its use.

ECS DATA

An ECS ticket is a record of an event, recorded mainly in free-text by a Con Edison dispatcher. The relevant events for the secondary system are labeled with various "trouble types" describing the category and seriousness of the event. A sample portion of an ECS ticket for a manhole fire is shown in Figure 1.

ticket	ME00000503	recvd_datetime	2000-01-20 14:46:00-05
house_no	S/E/C	cross_st	BROADWAY
cp	W	comp_datetime	2000-01-20 21:50:00-05
street	189	system_job	ME00000503
arty	ST	lines	15
orig	EDSMHF	remarks (abbreviated)	FIRE DEPT
act	EDSMHF	REPORTS CONDITION ORANGE----------MC	

Figure 1: Sample ECS Ticket

For this ticket (as for many other tickets) the exact geographic location of the event is not recorded in the ticket; only a free text address is given, in this case "S/E/C W 189 ST" (here, S/E/C indicates "southeast corner") with cross street "BROADWAY." Thus, geographic coordinates must be assigned to each ticket before these data are analyzed geo-spatially. This geocoding was performed using "Columbia Geostan," a geocoding system we developed specifically for the ECS database [Rudin, 2007a]. After each ECS ticket was assigned a geographic coordinate, we performed a preliminary analysis using statistics and machine learning coupled with visualization techniques [Rudin, 2007b].

ARCHITECTURE

The visualization tool interfaces directly with our PostgreSQL 8.2 database (allowing us to visualize the most recently cleaned and processed data), and also with Google Earth 4.2 (beta) locally. Figure 2 illustrates the architecture of our visualization tool. The Apache Tomcat 5.5 server is hosted on a machine at Columbia University. The java servlets residing on the server allow a user to input a

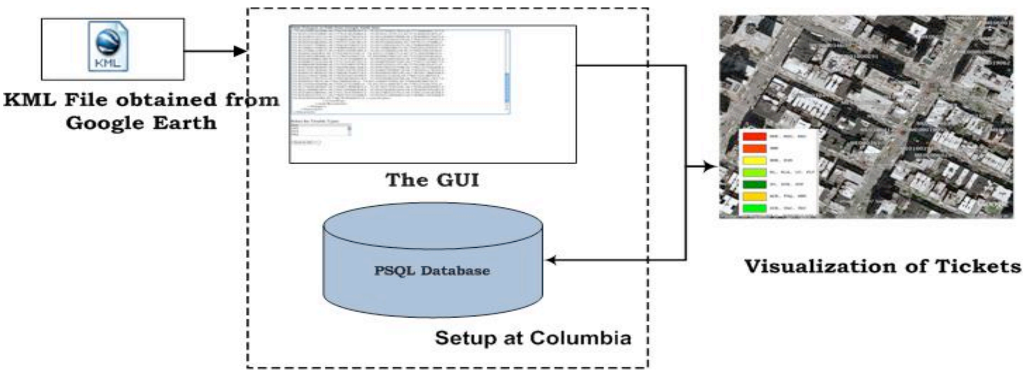

Figure 2: Architecture of the Visualization Framework

Keyhole Markup Language (KML) file. The file specifies the region of interest to a user using place marks, images, polygons, 3D models, textual descriptions and other features provided by the KML language. In addition, the user can select the trouble type(s) that are of interest. The servlets then connect to the PSQL database and retrieve the geographic coordinates, i.e., latitude and longitude of the ECS tickets computed by Columbia Geostan. Finally, these coordinates are mapped onto Google Earth. In order to display the results, a zip file is downloaded to the default directory set by the browser. The file contains the following: (a) 128 .PNG files - these are the images of dots needed to mark ticket locations (b) One or more query<ACT>.kml files - these contain the kml information relevant for each particular trouble type. For example, queryMHX.kml contains all the data relevant to the trouble type MHX (manhole explosion) (c) merge.kml file - if more than one trouble type is selected by a user, then we provide the facility to visualize the data from different trouble types at once. This is accomplished by taking advantage of the NetworkLink feature provided by Google Earth. The merged file is color-coded and the groupings are explained in the legend. If individual (rather than merged) kml files are visualized, the legend is unnecessary and will not appear.

SPATIO-TEMPORAL VISUALIZATION FOR MACHINE LEARNING SUPPORT

Our visualization tool for ECS tickets, used in conjunction with our preliminary statistical and machine learning techniques, has yielded some powerful results and observations. One major conclusion that was observed both visually and statistically is that a *long-term* history of events in the local geographic area can be useful for prediction of manhole events in the future. On the other hand, it is extremely difficult to predict manhole events using a short-term (on the order of 2 months) history of past events; for instance, we have computed that 76% of the manhole events had no precursor events within 60 meters and 60 days prior (5,166 out of 6,785). Thus, every short-term-only prediction algorithm would miss at least half of the manhole events. This observation is somewhat counterintuitive since one might ordinarily expect a series of precursor events to signal an upcoming manhole event. In fact, we have found that there are a series of localized long-term clusters of events (hot spots), which tend to have a historical concentration of ECS events spread over several years. By pinpointing these areas through both machine learning and visualization tools, we are able to identify such areas of future vulnerability to manhole events.

One localized region pinpointed by our analysis is the Upper East Side corner illustrated in Figure 3 (top). The southeast corner of this intersection is a historically active hot spot. Using the visualization tool, it is clear that the events in this area surround only one set of buildings. Upon our visit to this location, we noted that these buildings appeared to be significantly older than the other buildings in this area, an observation that would be either substantiated or rebuked by cable installation dates. It is a realistic assumption that the age of the low voltage electrical infrastructure could be a reliable predictor of manhole events; cable data is thus expected to be an important part of our future analysis. This is an instance where observations made using the visualization tool determine the future directions of our analysis.

We have found through our experience with the visualization tool that there are a significant number of intersections that have a high density of trouble tickets. For instance, we have located over 30 intersections in Manhattan with over 50 tickets apiece. One such intersection in Manhattan's East Village is shown in Figure 3 (bottom). The area surrounding this intersection has numerous tickets over many years.

ACKNOWLEDGEMENTS Thanks to Serena Lee, Won Choe, Stavros Livanos, Aaron Prazan and Maggie Chow from Con Ed, and David Waltz, Roger Anderson, Albert Boulanger, Hatim Diab, Jawwad Sultan and Tom Lippincott from Columbia.

Figure 3*[1]*: (Top) Image of an intersection in Manhattan's Upper East Side with ECS tickets superimposed by our visualization tool. The tickets seem mostly to surround one set of buildings. Each dot represents the location of an ECS ticket, and the color represents the trouble type. Each red dot represents a manhole event. (Bottom) Image of the area surrounding an intersection in Manhattan's Lower East Side.

[1] Enlarged images are provided as supplemental material.

BIBLIOGRAPHY

Cynthia Rudin, Becky Passonneau, Axinia Radeva, Zhi An Liu. Columbia/Con Edison Project on Secondary System Events, Phase 1 Final Report. Center for Computational Learning Systems, Columbia University, December 2007a.

Cynthia Rudin, Becky Passonneau, Axinia Radeva, Haimonti Dutta, Nandini Bhardwaj, Jawwad Sultan. Columbia/Con Edison Project on Secondary System Events, Phase 2 Final Report. Center for Computational Learning Systems, Columbia University, December 2007b.