Using Social Sensors for Detecting Power Outages in the Electrical Utility Industry

Konstantin Bauman, Alexander Tuzhilin, Ryan Zaczynski
Stern School of Business, New York University

WITS
December 12, 2015
Power outages is a big problem

• 3,634 power outages in 2014, affecting 14.2 million people
• 2008-2013 the US has 2,987 outages on average affecting 21.6 million people annually
• estimated losses in excess of $150 billion annually

Solution:
Detect PO as fast as possible!

Question:
HOW?
How to detect Power Failure

Solution 1: Call center
  • increasing cost per call
  • take a lot of time
  • difficult to reach utilities during extensive power outages

Solution 2: “Smart” Grid
  • total cost being estimated at $338 to $476 billion
  • fully implemented by only 2030

Solution 3: Social media
  We focus on Social Media approach
The power of tweets @ outages

- People do tweet in case of power outages

ICE STORM | TORONTO 2013
2,641 TWEETS IN 24 HOURS
RESEARCH QUESTION

How can we use social media (e.g. Twitter) for real-time power outage event detection?
OUR APPROACH

Use *automated* data mining algorithms and burst detection methods for *real-time* power outage detection based on Twitter.

Konstantin Bauman, Stern School of Business NYU
OUR SOLUTION

1(a) Core set of key concepts

1(b) Concept extension

2 Collecting tweets containing key concepts

3(b) Identification of Class 1 tweets

4 Power outage tweet burst detection

5 Aspect extraction of power outages

Predictive model

Database D

Report on power outage including:
- Start time
- Detection time
- Set of corresponding tweets
- Possible reason of power outage
- Weather condition in the region

Konstantin Bauman, Stern School of Business NYU
MANY WAYS TO SAY THAT YOUR POWER IS OUT

Step 1: Building a set of Key Concepts

- Identify the set of core concepts $K$
- Compute closure $C$ of set $K$ by finding all “similar” concepts based on synonyms, variations, slang terms, misspellings, etc.
MANY WAYS TO SAY THAT YOUR POWER IS OUT

Core concepts:
“power outage,” “no power,” “electric failure”

The Closure of core concepts:
• 110 key concepts:
  - power outage
  - power outages
  - power out
  - power’s out
  - powers out
  - power blackout
  - power brownout
  - power disruption
  - power off
  - power flicker
  - power flickering
  - power flickered
  - energy outages
  - energy out
  - energy’s out
  - energys out
  - energy blackout
  - energy brownout
  - energy disruption
  - energy off
  - energy flicker
  - energy flickering
  - energy flickered
  - energy failure
Step 2: Collecting Tweets containing key concepts

• Use *Twitter API* to collect tweets having at least one key concept in real time

• Outage detection within regions served by different power utility companies
  - 281 region in US
NOT ALL TWEETS WITH KEY CONCEPTS REPORT POWER OUTAGE

Relevant
• “Wow #%! we have a power outage rn?”
• “#ferguson power outages due to lightning”
• “8/16 8:38PM - Power outage.”

Irrelevant
• “can there be an earthquake or power outage so that i can go home?”
• “#KONE Widespread power outage hits Barstow - Victorville Daily Press http://t.co/4pmcVqJoRb inlandempire”
• “Flashlight cap. Perfect for power outage http://t.co/gAaXTI”
NOT ALL TWEETS WITH KEY CONCEPTS REPORT POWER OUTAGE

Step 3: Predictive Model

**Class 1**: Tweets posted by individuals witnessing power outages and immediately tweeting about them.

**Class 0**: all other tweets.

**Features for learning**

- length of the tweet in symbols/words/sentences,
- presence of a URL link (True/False),
- if the tweet is a re-tweet (True/False),
- if the user name contains certain special words, such as ``news'', ``police'', ``power'', etc.,
- sentiment, single words, and others.
Performance of the Predictive Models

We use Logistic Regression because:

- it shows good classification performance
- the model is simple and fast in predicting new labels
Step 4: Identification of Power Outages

• We cannot rely on a single tweet

TWO PEOPLE, SAME BUILDING, THE UTILITY NEVER SAW IT
Step 4: Identification of Power Outages

We use burst detection algorithm introduced in (Kleinberg 2002)

- efficient, dealing with underlying noises
- does not require human intervention
Step 5: Aspect Extraction of Power Outages

Can we extract any additional information from tweets?

We try to identify two aspects of power outages:

• its reason, such as equipment failure or public accident
• the weather condition at the time of the outage.

This is accomplished using a set of predefined keywords.

Example: for **Vegetation** category we use keywords: “tree,” “limb,” “branch,” “vines,” and “trunk.”
RESULTS

Utility Power Outage Data

• Utility company in a large municipal region in US
• Power Outages for the period 10/25/2014 - 1/25/2015

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Identified PO bursts</td>
<td>16</td>
</tr>
<tr>
<td>Detected PO confirmed by utilities</td>
<td>15</td>
</tr>
<tr>
<td>Precision</td>
<td>93.7%</td>
</tr>
<tr>
<td>PO in validation data</td>
<td>298</td>
</tr>
<tr>
<td>PO identified by system</td>
<td>109</td>
</tr>
<tr>
<td>PO discussed on Twitter</td>
<td>147</td>
</tr>
<tr>
<td>Total Recall</td>
<td>36.5%</td>
</tr>
<tr>
<td>Twitter based Recall</td>
<td>74.1%</td>
</tr>
</tbody>
</table>

Konstantin Bauman, Stern School of Business NYU
RESULTS (cont.)

Outage Data Based on Reliable Twitter Accounts

**Reliable sources**: news organizations, police departments and other “official” Twitter accounts.

Power Outages for the period 8/25/2014 - 1/25/2015

We identify *reliable* Twitter user names based on a set of key words, such as:

“news”, “police”, “power”, “electricity”, “weather”, “alert,” etc.
RESULTS (cont.)

Outage Data Based on Reliable Twitter Accounts

<table>
<thead>
<tr>
<th>Identified PO bursts</th>
<th>3750</th>
</tr>
</thead>
<tbody>
<tr>
<td>Detected PO confirmed by utilities</td>
<td>152 (from 300)</td>
</tr>
<tr>
<td>Detected PO confirmed by inspection</td>
<td>296 (from 300)</td>
</tr>
<tr>
<td>Precision</td>
<td>50.6%</td>
</tr>
<tr>
<td>Precision Manual</td>
<td>97.6%</td>
</tr>
<tr>
<td>PO in validation data</td>
<td>4205</td>
</tr>
<tr>
<td>PO identified by system</td>
<td>169 (from 300)</td>
</tr>
<tr>
<td>PO discussed on Twitter</td>
<td>242 (from 300)</td>
</tr>
<tr>
<td>Total Recall</td>
<td>56.3%</td>
</tr>
<tr>
<td>Twitter based Recall</td>
<td>69.8%</td>
</tr>
</tbody>
</table>
CONCLUSION

We presented a novel power outage detection method that
• filters the tweets containing key concepts
• identifies the tweets referring to real power outages
• detects bursts among these identified tweets
• identifies possible reasons of the outage and the weather
  conditions in the region at that time.

We validated our method on two datasets and showed that it has
• high precision measure - 93.7% and 97.6%
• good recall measure - 36.5% and 56.3%

The system identified
• possible reason of power outage in 5.36% of the outages
• weather conditions in the region in 10.32% of the outages
Thank you!

Konstantin Bauman
Stern School of Business NYU

kbauman@stern.nyu.edu