News, Media and Situational Awareness

Using headlines to provide real time updates for responders

Brendan C, Kevin P, John H. aka Team Pottery Shack

The Problem

There is currently no forum for rounding up and archiving relevant news during a live disaster event.

The Goal

Leverage news feeds relevant to specific disasters, gathered from multiple sources, to create a webpage that presents these live feeds under one umbrella (on one page). This is similar to the Google News feature.

To do this we utilized natural language processing and machine learning.

The Process

Step one: Finding the data

The Process

Step one: Finding the data

Two sources for data:

1. New York Times API

2. newsapi.org

Search worldwide news with code

The Process

Step one: Finding the data

Using The New York Times API, we collected articles related to specific disaster types:

 Hurricane

 Earthquake

 Fire

 Tornado

 Blizzard

 Flood

The Process

Step two: cleaning the data

Having relevant data was a key factor for success in this project.

We created criteria to help determine data relevancy:

The Process

Step two: cleaning the data

Having relevant data was a key factor for success in this project.

We judged data relevancy based on two main criteria:

1. **Is this article relevant to an ongoing event (at the time written)?**

2. **Is the article relevant to first-responders / stakeholders?**

The Process

Step two: cleaning the data

In order to implement the criteria, we manually read each headline and marked the entries as either relevant or irrelevant.

The Process

Step two: cleaning the data

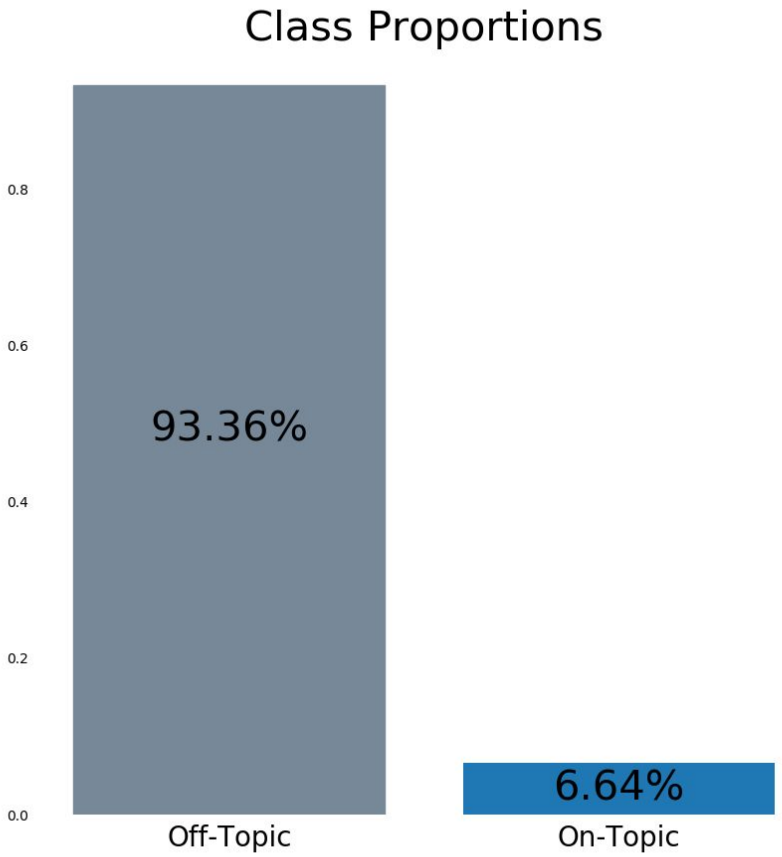
In order to implement the criteria, we manually read each headline and marked the entries as either relevant or irrelevant.

This wasn’t as bad as it seems.

The Process

Step three: modeling

Severely Imbalanced Classes!



The Process

Step three: modeling

$$\text{Recall} = \frac{\text{Relevant Articles Identified}}{\text{Relevant Articles Identified} + \text{Relevant Articles Missed}}$$

Our model had a recall score of **77% on unseen articles.**

The Process

Step three: modeling

Once we had a trained model, we connected it to the news API to provide newer, up-to-date news articles to the user.

The Process

Step four: the website

We then created a website that allows a user to search for keywords and specific dates.

Our goal for the website was a simple and easy **user experience.**

The search returns a list of the most relevant articles for the time period selected.

Closing remarks...

With proper scope and resources, this project would be improved by

- Diversifying data by acquiring paid access to the news API
- Dedicated UX team to improve usability, website look and feel
- A clever name

Website Demo