Are Recent Terrorism Trends Reflected in Social Media?
Ivana Terziyska, Xiao Luo
Indiana University-Purdue University Indianapolis

Introduction
Social media plays an important role in shaping an audience’s beliefs and sentiments regarding current events. A comparison between significant features of objectively accurate, holistic datasets and user features of social media data would give insight into the spread of misinformation and which aspects of events are reflected in user behavior. We use the Global Terrorism Database (GTD), a focus of previous comparisons to diverse datasets, to highlight trends in terrorism, a frequent and controversial focus of media coverage as a comparison to data obtained using the Twitter API to assess features of social media data pertinent to relevant queries. We present a connection between terrorism as recorded in the GTD and terrorism as spoken about in tweets to determine discrepancies in the trends and features of both data sets. We create a self-organizing map (SOM) of each dataset to identify clusters present both in the GTD and Twitter data to link our findings to broader conclusions about media bias, public perception on terrorism, and the in trends in both datasets.

Methods
I: Data Bases and Collection
We used data from Global Terrorist Database (GTD) containing 52234 instances of terrorist attacks from 2012-2015 with 45 different features. Additionally, 1,785 Tweets containing the keywords “terrorist” and “terrorism” collected using the Twitter REST streaming API.

II: Feature Selection
The input weights of each feature were analyzed to extract the most relevant features and analyze features that comprise clusters in the broader database. A four feature subset was chosen based off this analysis, containing the Global Terrorist Index (GTI), Region ID, crit1 (motive for the attack), and targ_type

III: Pre-Processing
Tweets were filtered by geolocation and search query; the tweet location and content was used to generate a value corresponding to the four features selected in the GTD through review according to the features mentioned in the GTD Codebook. Additional features including sentiments and favorites were also considered.

IV: Visualization:
The Self Organizing Map (SOM) was used as a proposed method to analyze trends and similarities in both Twitter and GTD data. SOM is an unsupervised machine learning technique that utilizes artificial neural networks (ANN) and calculates the best matching unit (BMU) to map data to a specific neuron (sample hit). Its ability to represent higher dimensional data makes it an effective tool for data visualization. Below are two SOMs of GTD data and Twitter data, each with four features. On the left is the neighborhood distances map, representing how closely neighboring neurons are related; on the right is the sample hit matrix, representing how many data points are assigned to each neuron.

Results
The clustering results for each dataset showed that social media data clusters together and weights input features differently than the objective GTD data. Additional features such as favorites and user sentiment affected the way the data clustered together, indicating a relationship between the input features used. The discrepancy in clustering between GTD and Twitter data may be explained by bias, disproportionate representation, or change in terrorist trends not reflected in previous GTD records. Examining and comparing diverse data sets using SOM was proven to be a viable method in extracting and measuring features of data to give insight on terrorism trends and user perception.

Future Work
This research can be further extended to model social media trends over a longer period of time, for example mapping an entire year of GTD data to a year of Twitter data. To better visualize global perception and information on terrorism, a more comprehensive and inclusive demographic sample can be obtained to eliminate possible bias in this study. Additionally, other features of the GTD can be explored aside from the ones tested in this experiment. Developing a repository of terms and classification of tweets may be a useful mechanism to better translate GTD features to social media.

This research was made possible with the support of the Indiana University-Purdue University Indianapolis Department of Computer Information and Information Science, as well as through funding from the National Science Foundation and the United States Department of Defense. The author(s) would like to thank their mentor Dr. Xiao Luo, as well as Dr. Feng Li, Dr. Eugenia Fernandez, and Sheila Walter for their support.