Sentiment Analysis of Twitter Response to Major World Events

Team 3

Group Members:
Juan Pablo Arrivillaga (jpa8c@virginia.edu)
Kyle Ferris (kcf5bw@virginia.edu)
Dong Hyuk Kim (dk2qd@virginia.edu)
Kaustav Saha (ks3hw@virginia.edu)

Problem Description

What is the problem?

How do different populations react to major world events such as a terrorist attack? Is there a difference in the topics they discuss in relation to the event? In this project we focus on a study of the November 2015 Paris attacks by probing tweets posted on the microblogging platform Twitter. In terrorism informatics, tracking, location and time of activities vary significantly and thus become extremely hard to predict. Intelligent information sharing techniques, applied to the unstructured content of texts, can lead to the discovery of hidden patterns for many real world disaster and crisis management situations. Twitter is known to be one of the channels where civilians break news of and respond to terrorist activities. It is also used as a method to notify the public of any latest updates, to express strong emotions, and as an information source for the authorities. Twitter is an avenue to break news faster than traditional news outlets. We have built a novel application applying opinion-mining and sentiment analysis to Twitter data. Our framework would enable users to visualize the civilian response to a terror attack, and determine any differences in response between the citizens of different regions.

Why is it important?

Understanding how the populations of different regions react differently to events such as a terrorist attack will help to increase understanding of cultural and political differences, and potentially aid in policy decisions and political campaigning.

Who cares about it?

Users could include politicians, international leaders, and NGOs. Polling is expensive and takes time. Generating a model for understanding reactions in real time would be faster and would complement the slower, more expensive polling approach. We propose a structured framework to harvest civilian sentiment and response on Twitter during terrorism scenarios. Coupled with intelligent data mining, visualization, and filtering methods, this data can be collated into a knowledge base that would be of great utility to decision-makers and the authorities for rapid response and monitoring during such scenarios.

Why does it remain unsolved?

While sentiment analysis has been extensively studied for some domains, such as product reviews, the performance on social media data is still unsatisfactory due to the distinct data characteristics. First, social media posts are always short and unstructured. For example, Twitter allows no more than 140 characters and uses many informal words such as “cooool” and “OMG” . The brevity of tweets can lead to difficulty in providing enough information for interpretation. Second, it is labor intensive and time consuming to obtain ground truth for training data, which is needed to build an effective supervised learning model.

Objectives

What are we proposing to do about the problem?

Topics and hashtags discussed on Twitter are analyzed by querying the Twitter Trending Topics list, a list of very frequently discussed topics updated on a regular basis. By monitoring the most talked about messages at any given time, we can use Twitter to chronicle the civilian response to major events from the moment news first breaks out.

We used Twitter’s Search API and Streaming API as our data sources. Twitter Search is used to query the topic and its past discussion right up to event; on the other hand the Streaming API is used to monitoring the real-time chatter on Twitter using their live HTTP stream to capture discussion about the event as it happens. We used Latent Dirichlet Allocation and Pachinko Allocation Model to generate topic models and compare generated topics from populations in distinct regions of the world. Using Tableau, we created a visualization of the manually annotated emotions obtained on a world map to understand the sentiments from different parts of the world.

How did we measure the success of our work?

To asses our topic models, human-interpreted coherence will be the ultimate goal. Using the MALLET software package, we also were able to get log-likelihood measures of the performance of the topic models. We were able to discern differences in sentiments among different geographic regions. Finally, we compared different topic modelling approaches given the above metrics.

Related Work

What have others (e.g., researchers, companies, etc.) done to address this problem?

There is limited research on the usage of Twitter in terrorism informatics. On a topic closely related to terrorism informatics, Hughes and Palen (2009) have surveyed the adoption and use of Twitter during mass convergence and emergency events, specifically those involving national security, in the perspective of “crisis informatics”. They allude that Twitter messages exhibit “features of information dissemination [supporting] information broadcasting and brokerage”, and that Twitter may be used as a tool for emergency response and communication by the authorities in order to provide aid and counter disinformation. Jungherr (2009) detailed the role of Twitter in social activism and looked into case studies whereby Twitter was an instrumental tool in disseminating information on terrorist attacks, political dissent, and acts of oppression. In the past, academic research has been done to explore reactions to terrorism or natural disasters on Twitter (Cheong and Lee, 2011). Cheong and Lee focused on assigning responses into seven distinct categories: Fear/Anxiety, Shock, Response, Need for Information and Updates, Threat Assessment, Casualties Assessment, and Response and Law Enforcement.

What are you doing that is similar to past work?

Similar to the work of Cheong and Lee, we have also created a geographical visualization. Cheong and Lee had grouped responses into several categories including emotional responses, practical concerns, and official communication.

Are there commercial products that accomplish what you are trying to do? What are their characteristics? Where are their gaps?

As far as we know, there are no commercial products that accomplish exactly what we are trying to do. There are a number of free tools and guides available for performing sentiment analysis and text mining on Twitter data.

What about your work is novel? What gaps does it fill?

We utilized the capability of conjoint structured data and unstructured content mining in extracting deep knowledge from noisy twitter messages, through our proposed structured framework. The novelty of our work lies in differentiating the emotions into actual categories like fear, anxiety, threat, etc., rather than the generic use of positive/negative/neutral response in sentiment analysis. Our goal is to explore the underlying trends in sentiment with respect to important global events. We also plot segregated emotions to different regions of the world to understand how different parts of the world react to a world event.

Approach

Overall Approach Summary

1. Extract tweet Data (geotagged) using popular trending hashtags on major events such as elections, Paris Attacks, etc. (http://hashtagify.me/).
2. Filter Nose using stop word removal, whitespace removal, punctuation removal.
3. Perform Topic Modeling using basic LDA approach and its extensions of Pachinko Allocation.
4. Formulate 8 emotions (Refugee Migration,Terrorist_Attack, Fear, Religious_Extremism, Political_Conflict,Support_Refugee,Threat,Response) from the top-10 words of the topics.
5. Segregate tweet based on these 8 emotions of their point of origin.
6. Do a demographic analysis by using Tableau and Heat-Maps based on the segregated tweets.

What data did you use? How did you preprocess it?

Our data consisted of tweets over the period of November 17, 2015 to November 26, 2015, filtered to include only tweets that included certain relevant hashtags. The Paris terrorist attacks occurred on November 13 so most of the tweets truly reflected the emotions that we intend to study. Using http://hashtagify.me/ , we identified the top trending hashtags (30 hastags) based on the Paris attack from as the following:

#Terrorist,	***#ISIS,***	***#Syria,***	***#Iraq,***	***#ISIL,***
#AK47,	***#ParisAttacks,***	***#Raqqa,***	***#France,***	***#Taliban,***
#IslamicState,	***#Islam,***	***#Muslim,***	***#Jihadists,***	***#Obama,***
#Turkey,	***#Iran,***	***#USA,***	***#JeSuisParis,***	***#NotAfraid,***
#CharlieHebdo,	***#ParisShooting,***	***#ViveLaFrance,***		***#refugees,***
#Suicidebombing,	***#Terrorism,***	***#AlQaeda,***	***#Pakistan,***	
#Hollande	***#Putin***			

Thus our final dataset was a total of 60,998 tweets from 11/17/15-11/26/15 with the above hashtags. We further divided our data into two subsets, consisting of tweets from major cities and tweets from smaller cities. 4,673 were classified as originating from major cities (39 major cities), and 56,325 from smaller cities (Around 1000 cities around the world) .

We next cleaned the data before performing the topic modeling. All letters were converted to lowercase and URLs were removed from the tweets. Furthermore, a set of common English stopwords were filtered out of the tweets. The MALLET Java library for natural language processing was used to perform the topic modeling.

What analyses did you perform?

We performed Latent Dirichlet Allocation (LDA) topic modeling on the corpus with a total topic count of 400. We then manually went through all 400 generated topics and selected those that were coherent and related to one of our topics of interest, as defined below :-

Fear	Anxiety, Anxiousness, Catastrophic, Concern, Disaster, Fear, Insecure, Panic, Scared, Terror, Trouble, Warning, Worry
Political Conflict	Assad,Syria,Russia,Turkey,Policy,EU,Emergency,Embargo, Sanction,Xenophobia,France,USA
Refugee Migration	Open,Border,Ship,Boat,Refugees,Racist,Child,Women,Sleep,Food, Transport
Religious Extremism	Bigotry,Shame,Muslim,Execute,Trust,Intolerance,Hindu,Christian, Islam,Hate,Foul Play
Response	Breaking News, Call, Incident, Phone, Report, Situation, Unconfirmed,Act, Asap, Escape, Evacuate, Flee, Help, Hide, Run
Support Refugee	Welcome,Refugee,Commitment,Homeless,Support,Obama, Republican,Asylum,Trump
Terrorist Attack	Attack, Bomb, Bullet, Collapse, Crash, Explosion, Fire, Gun, Hijack, Hit, Hostage, Plane, Responsibility,Rifle, Shoot, Struck, Suicide, Terrorist
Threat	Threats, Accident, Shock, Aback, Floor, God, Bless, Omg, Shock, Stun, Sudden

For the dataset containing the minor cities we were left with 77 Topics after manual annotation whereas for the Major Cities we worked with 124 Topics out of 400. Most of the rest of the topics left out were coherent but unrelated to our domain of interest. For example, there were dozens of topics that could have easily been interpreted as relating to travel and sightseeing, or the

Thanksgiving holiday. The remaining topics were then manually classified into one of our eight topic groups.

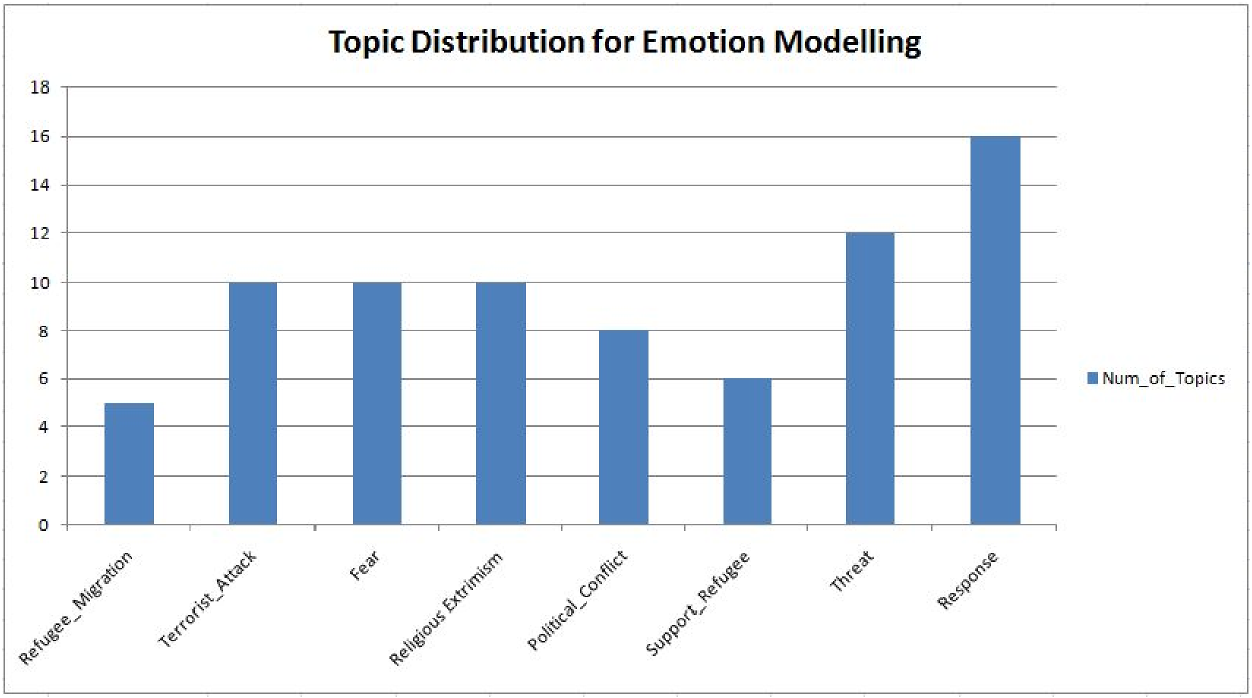


Fig:Topic Distribution for Emotion Modelling for All cities Dataset containing 77 topics.

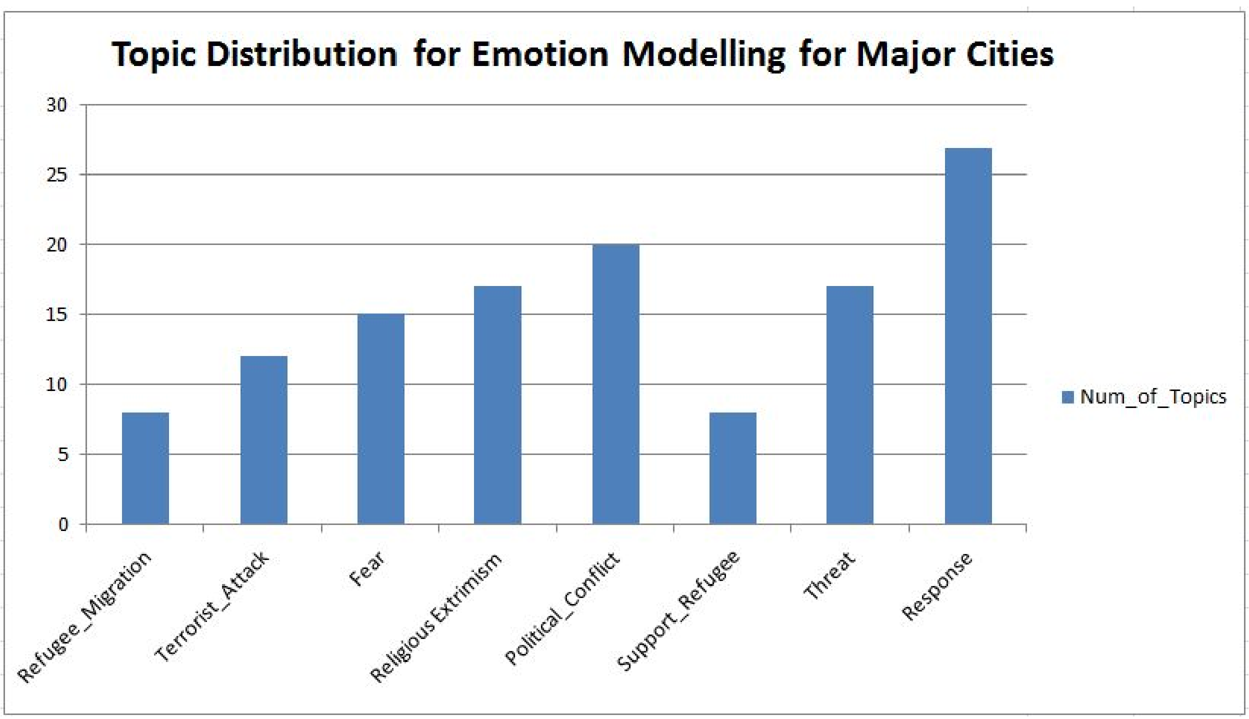


Fig:Topic Distribution for Emotion Modelling for Major cities Dataset containing 124 topics.

The next step was to generate a value of each topic group for each tweet. For each tweet, the relevant topic values in each group were added up. The end result was that each observation had 11 components: tweet text, longitude, latitude, and a value for each of our meta-topics indicating how relevant they are to the individual tweet.

For the Pachinko Allocation Model we used 20 super-topics subdivided into 200 sub-topics. We worked with the smaller dataset of the major cities as the Algorithm did-not scale well on the larger data-set.

What models did you build?

We created topic models using Latent Dirichlet Allocation (LDA). LDA is an unsupervised machine learning technique which identifies latent topic information in large document collections. LDA is a member of a family of models known as probabilistic topic models. The underlying theme of these models is that documents are a mixtures of topics and a topic is a probability distribution over words. The models are generative, that is, they describe a procedure for generating documents from some number of topics. One can imagine a set of topics, which are merely a set of probabilities for every word in a corpus. From there, a document is generated by choosing a topic at random according to some underlying probability distribution, and from this selected topic, a word is chosen at random weighted by the probabilities that define the topic.

In the LDA model, a Dirichlet prior is used for both the per document topic distributions and the per-topic word distributions. Finally, to generate a topic-model, statistical inference techniques are used to infer the set of topics, the latent variable, that would generate the words observed in the documents. The LDA analysis was implemented using the MALLET software package from UMASS Amherst.

The Pachinko Allocation Model (PAM). PAM aims to model correlations between topics. PAM captures correlations using a directed acyclic graph (DAG), where topic nodes occupy the interior levels and the leaves are words. PAM divides the topic-model space into super-topics which are further divided into sub-topics. These parameters can be manually tweaked in MALLET software package.

What evaluation setup will you use?

We compared the results of our topic modeling to the sentiment analysis topics proposed above. The topics generated from our LDA analysis are similar to the topics proposed for doing our sentiment analysis, so our hypothesis is validated. Additionally, we used evaluation setup with perplexity as a measure of model fit. Perplexity, which has traditionally been used for evaluating topic models, is a type of log likelihood measure. A low value is indicative of a better topic model. We varied the number of topics as our input parameter and studied it's effect on perplexity measures.

In addition, we created a geographic visualization in Tableau. This visualization showed us the location of all geo-tagged tweets in our sample. In addition, the tweets were color coded with the color corresponding to the emotion most relevant to each tweet. There is also the ability to hover over an individual tweet and see detailed information about that specific observation, such as the exact emotion proportions and the text of the tweet in addition to latitude and longitude. This visualization allows users to evaluate at a glance the difference between topics discussed in different regions, allowing us to evaluate if our result is successful.

Evaluation

How well does your approach perform according to the metrics you describe in the Objectives section?

Our visualizations show a distinct difference in countries topics. There are distinct differences in the reactions of different country populations to major events. For example, tweeters in the UK are primarily concerned with political conflict, while those in the Middle East tweet about religious extremism. A good example is shown in the below graph where we see someone in Saudi Arabia who had tweeted that Muslim Refugees from Syria cannot be trusted. The populations of Eastern European countries are more concerned with migration and refugees. This visualization tool could be easily used by policy makers or political analysts to test already existing hypotheses about population reactions, or to analyze events as they unfold by feeding in new tweet data. Therefore, we consider our approach to be a success.

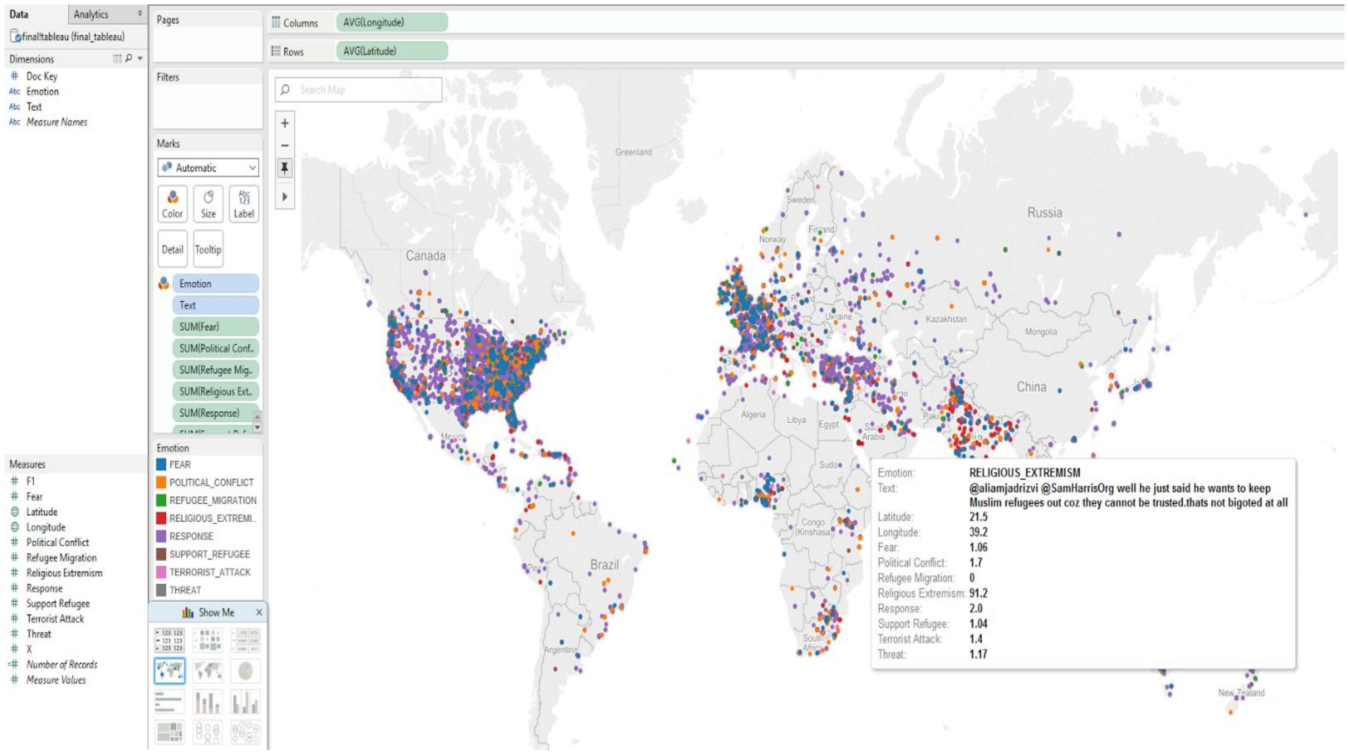


Fig : Distribution of Emotions around the World using LDA.

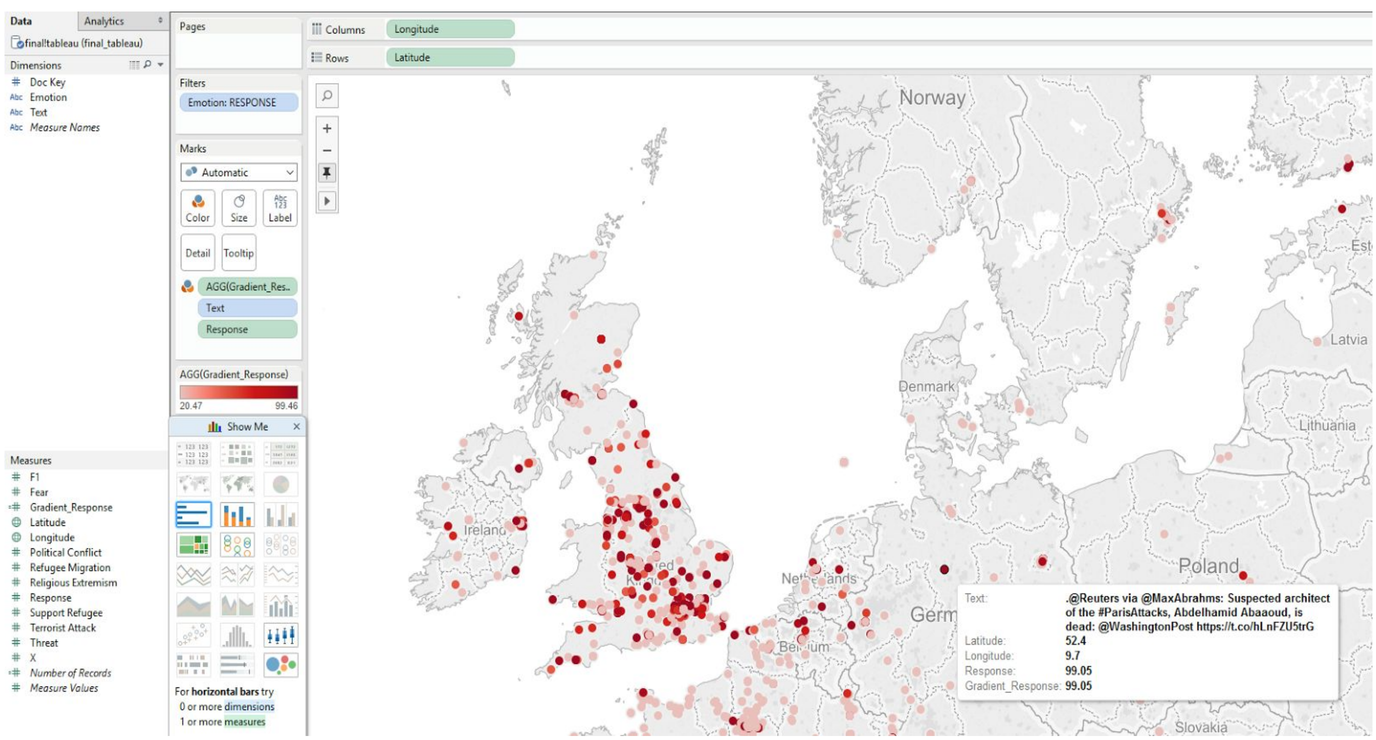


Fig : Gradient of Response in UK/France/Germany using LDA.

We also used gradient analysis of the different emotions. We plotted the points on the map relating to a particular emotion and put a color gradient describing the proportion of the emotion in the data. For example, as shown in the graph above we can see that when we form a gradient of the ‘Response’ Emotion we can observe darker red points in UK/France which depicts that people are concerned about Response in that part of the world. Response may comprise of breaking news,evacuation,government action or citizen work.We found a very good tweet from someone in Germany which stated that the suspected Architect of Paris Attack - Abdelhamid Abaaoud is dead.

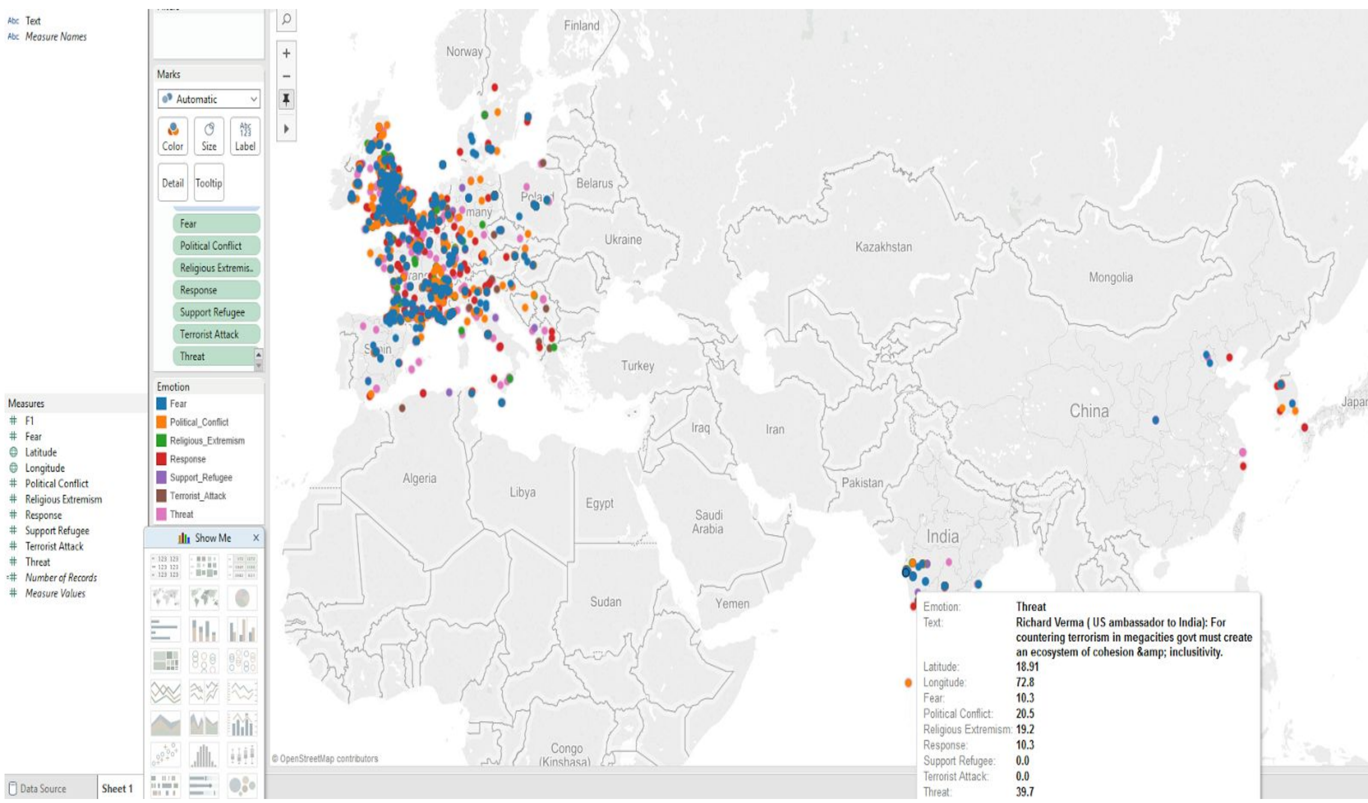


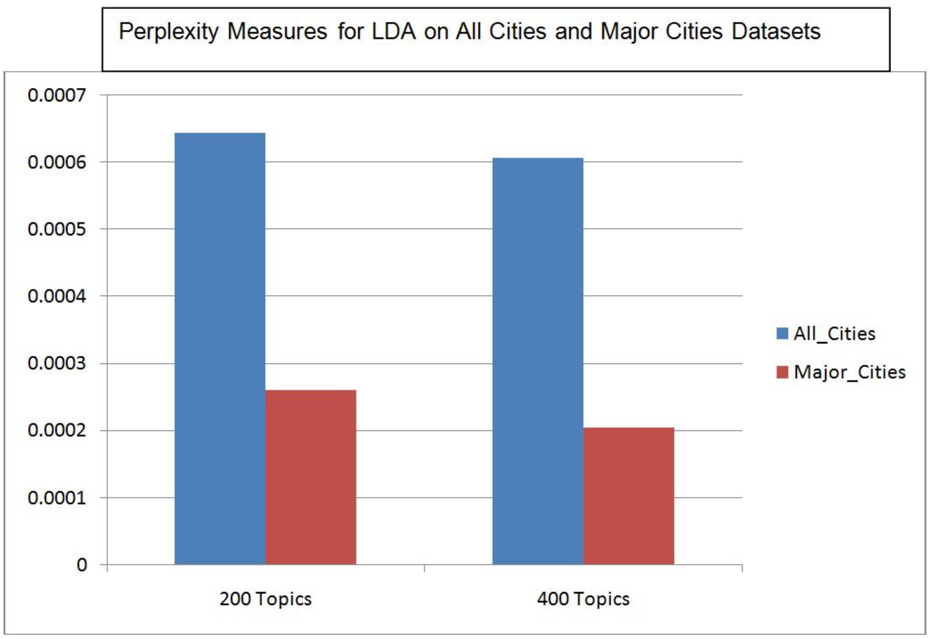
Fig : Distribution of Emotions around the World using Pachinko Allocation.

Using the Pachinko Allocation Model for Emotion Modelling, we can see that Fear and Political Conflict have an even distribution in both UK and France whereas in India there is a higher proportion of the Threat emotion. One of the important tweets was from Richard Verma,US Ambassador to India, regarding countering terrorism in mega-cities.

The measure traditionally used for topic models is the perplexity of held-out documents defined as :-

$$\text{perplexity(test set)} = \exp\{-L(w)/\text{count of tokens}\}$$

which is a decreasing function of the log-likelihood L(w) of the unseen documents. The lower the perplexity, the better the model. We varied the number of topics as our input parameter. We used 200 and 400 topics for LDA Topic Modelling on the all cities and major cities data-set and found that the major cities data-set performs slightly better.



In what situations does your approach perform well?

Our approach performs well when the tweets are properly geo-tagged. Additionally, our approach performs well after a particular world-event when there is an large activity in twitter related to one topic, with clearly defined hashtags. This leads to a much larger corpus, increasing the efficacy of our analysis. The other thing that would really help us if people intending to comment in English actually completed their entire sentence in English rather than describing something in the local language using English Alphabets. If there are a larger number of words with synonymous meanings which fall into one of our 8 categories, then both LDA and Pachinko Algorithm would perform much better.

Where does your approach break down?

One of the bottlenecks would be when the tweets are not properly geo-tagged. Low volume tweeters often do not enable the geolocation feature in twitter. These precludes us from using their tweets for analysis. Another hurdle is that often tweets contain raw words without proper English expressions. This makes it very difficult for topic-modeling usage. One of the problems we faced in the tweets was that even though a portion of the tweet was in English, the rest was in native language but written in the Latin alphabet. This scenario would be pretty confusing for the topic-modeling platform.

How does your approach stack up against other known approaches?

The major difference between our work and the previous approach lies in the way we perform the sentiment analysis. In previous approaches and related research literature, sentiment analysis has mostly categorized content into three sub-divisions of positive/negative/neutral. We take a step forward from that and use a form of granular categorization into nuanced emotions. For our case, we have actually subdivided contents into 8 emotions (Fear, Political Conflict, Refugee Migration, Religious Extremism, Response, Support Refugee, Terrorist Attack and Threat) rather than the generic 3. Whereas in the work of Lee and Cheong, they simply mapped all tweets in their sample without any further visual information, we used Tableau to map color-coded tweets, with the color corresponding to the major emotion of each tweet. In addition, it is possible to hover over an individual tweet and see the total emotion breakdown as well as the actual text of the tweet and the latitude and longitude of the point.

Conclusions and Recommendations

What have you learned by doing this work?

When manually annotating topics and grouping into meta emotions, we got very good coherence and geographic separation. The biggest issue with our project remains that it depends on manual annotation of topics. The biggest benefit of our project was reducing the dimensions needed to be manually annotated. For the minor-city dataset using LDA we could find a distribution of emotions such as Fear, Political Conflict and Response in US whereas in UK there seemed to be a predominant effect of Political Conflict. In France there was a great proportion of Response. In other parts of the world there seemed to be Religious Extremism especially in Middle East and India. With Pachinko Allocation Algorithm we obtained somewhat different results with Fear and Political Conflict emotions dominating in both UK and France whereas in India there is a higher proportion of Fear. This gives as in-depth understanding of the sentiments from different parts of the world.

What are your final recommendations with regard to addressing the problem you have identified?

These technical findings should be expanded in the future, and combined with traditional sociological and political methods to help understand cultural differences.

What should be done to better address this problem in the future?

To address this problem in the future, it should be expanded to include languages other than English. Considering many of our countries of interest are in the Middle East, bringing in an Arabic speaker and expanding the model to include tweets in Arabic would allow for a better sample of data. Other languages to consider include French, German, Spanish, and Urdu.

In addition, accessing more geo-tagged tweets would prove useful. If Twitter introduced a change in policy leading to more users posting tweets with coordinates attached, the analysis could be richer.

Possible other avenues to look into would be Scalability of Pachinko Model, Spatial clustering and Amazon Web Services Mechanical Turk.

References

Cheong, Marc; Lee, Vincent C. S. "A microblogging-based approach to terrorism informatics: Exploration and chronicling civilian sentiment and response to terrorism events via Twitter." *Information Systems Frontiers* 13.1 (Mar2011): 45-59.

D. M. Blei, A. Y. Ng, M. I. Jordan, Latent Dirichlet allocation, J. Mach. Learn. Res. 3 (2003) 993–1022.

Brummette, John; Fussell Sisco, Hilary "Using Twitter as a means of coping with emotions and uncontrollable crises." *Public Relations Review* 41.1 (March 2015): 89-96.

Jonathan Chang, Jordan Boyd-Graber, Chong Wang, Sean Gerrish, and David M. Blei. Reading Tea Leaves: How Humans Interpret Topic Models. Neural Information Processing Systems, 2009.

Hughes, A. & Palen, L. (2009), Twitter adoption and use in mass convergence and emergency events. In Proc. 6th International ISCRAM Conference.

Jungherr, A. (2009). The DigiActive Guide to Twitter for Activism. Available from http://www.digiactive.org/wp-content/uploads/digiactive_twitter_guide_v1-0.pdf.

McCallum, Andrew Kachites. "MALLET: A Machine Learning for Language Toolkit." http://mallet.cs.umass.edu. 2002.

############################## END OF REPORT ##################################