Distributed Data Analytics for Real-Time Monitoring and Detection of Flash Floods in Smart City

PI: ***Nirmalya Roy & Aryya Gangopadhyay***
PhD Students: Bipendra Basnyat, P.E., Amrita Anam, & Neha Singh
Department of Information Systems, University of Maryland, Baltimore County

Motivation

- Social media & micro-blogging posting nowadays are spontaneous, instantaneous, & overwhelmed, particularly during any eventful or crisis situations
- Multimodal data analytics and knowledge extraction are vital
- Build a proof of concept information extraction model using flash floods in Ellicott city
- Similar models can be employed to more critical and severe natural disaster situations

Research Approach

- Investigated the narrative (texts) and visual (images) components of Twitter feeds to improve the results of queries by exploiting the deep contexts of each data modality
- Employed Latent Semantic Analysis (LSA)-based techniques to analyze the texts
- Implemented Discrete Cosine Transformation (DCT) to analyze the images
- Established cross-correlations between the textual and image dimensions of queries to reciprocate each other

Process Flow for Social Sensing Module

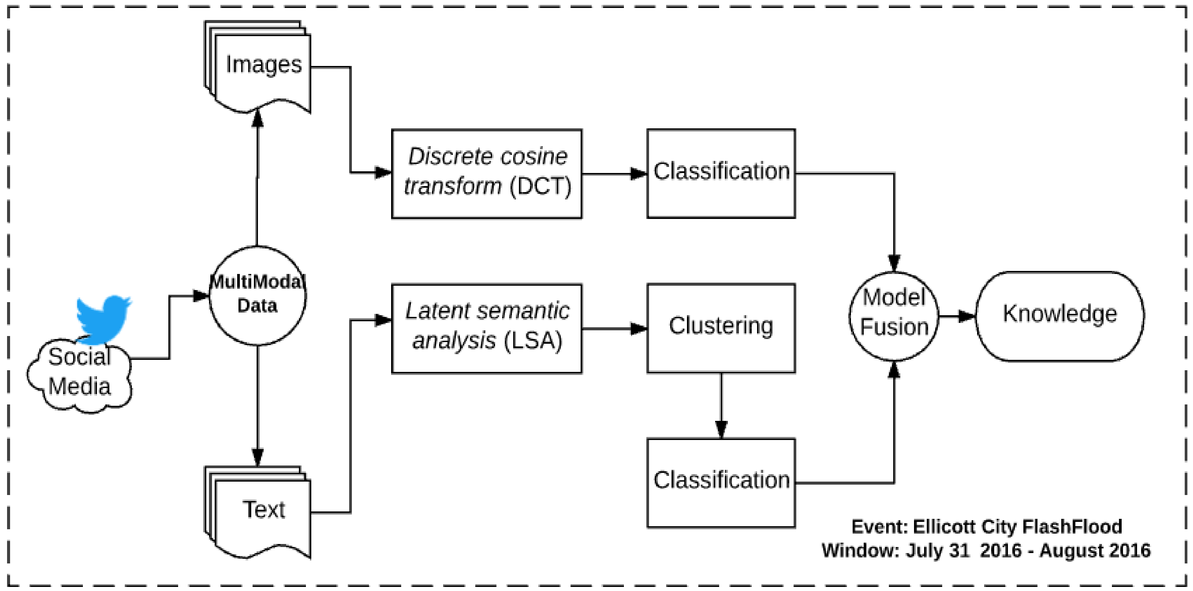
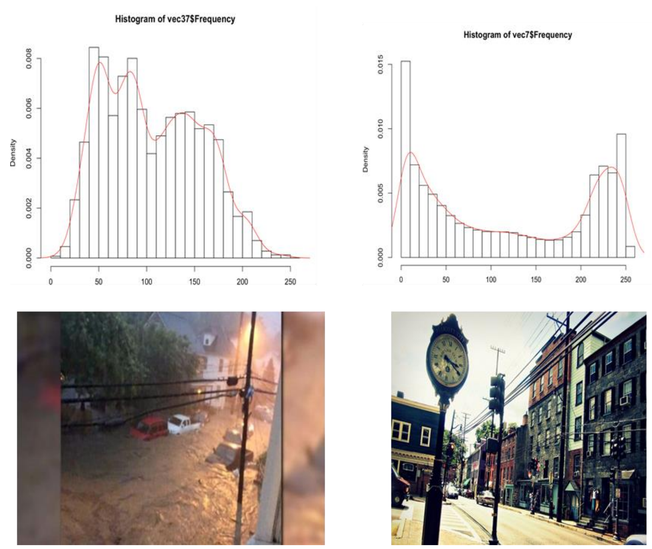


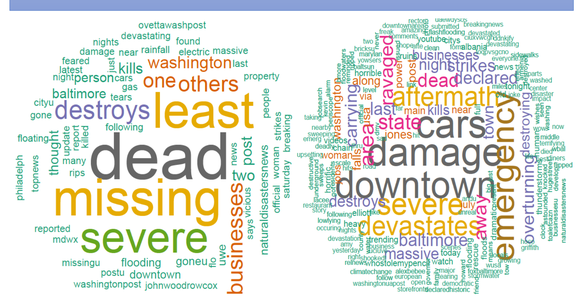
Image Analysis



Input: a corpus with p images;
Output: list of top k DCT coefficients for each image;
for $i \leftarrow 1$ **to** p **do**
  Convert image i to grayscale;
  Preform DCT on image i;
  Select the k largest coefficients such that 90% of energy is preserved;
end

Algorithm 1: Analyzing Images

Text Analysis



Input: a corpus with p Tweets(text);
Output: Tweets Cluster & Classification;
for $i \leftarrow 1$ **to** p **do**
  Perform LSA on Tweets i;
  Generate Labels on Tweets i;
  Preform Clustering og the Tweets i using KNN;
  Classify k into three categories as: "help/rescue", "damage/aftermath" and "casualty";
end

Algorithm 2: Analyzing Texts

- This work is partially supported by the NSF CNS grant 1640625
- B. Basnyat et. al., "Analyzing Social Media Texts and Images to Assess the Impact of Flash Floods in Cities", in Proceedings of the 2nd IEEE International Workshop on Smart Service Systems (SmartSys), co-located with IEEE SmartComp, May 2017

Datasets

- 648 Twitter feeds
- Timespan: July 31st 2016 to August 2016

Tweets with Images	152
Tweets without Images	477
Total	648

Image Analysis Results

Classifiers	With Outliers		Without Outliers	
	1D DCT	2D DCT	1D DCT	2D DCT
Decision Trees	27%	43%	65%	63%
Naive Bayes	52%	65%	67%	66%

Text Analysis Results

- 86.79% from tf-idf vectors
- 92.45% from SVD

Classifier Detection Accuracy

- help/rescue with 92.6%
- damage/aftermath with 90%
- casualty with 100%

Future Work

- Multi-model data fusion -- Image, text & sensor data streams
- Spatiotemporal modality analysis
- Bipartite graph based approach
- Elevation model

Mobile, Pervasive & Sensor Computing Lab
http://mpsc.umbc.edu/
Mobile, Pervasive & Sensor Computing Lab