A Machine Learning Analysis of Twitter Sentiment to the Sandy Hook Shootings

Nan Wang, Blesson Varghese   *Queen's University Belfast*

Peter Donnelly   *University of Toronto*

Outline

- Motivation

- Methodology

- Machine Learning Approaches

- Case Study: Sandy Hook Elementary School Shooting

- Limitations

- Conclusion

- Q & A

Outline

- Motivation

- Methodology

- Machine Learning Approaches

- Case Study: Sandy Hook Elementary School Shooting

- Limitations

- Conclusion

- Q & A

A Machine Learning Analysis of Twitter Sentiment to the Sandy Hook Shootings

- Motivation
  - Apply and evaluate machine learning approaches for sentiment analysis on social network
  - Provide insights gathered from social networks to decision makers
  - Engage non-CS audiences with research outputs through interactive visualisation

IEEE
Advancing Technology
for Humanity

Outline

- Motivation & Contribution

- Methodology

- Machine Learning Approaches

- Case Study: Sandy Hook Elementary School Shooting

- Limitations

- Conclusion

- Q & A

Methodology
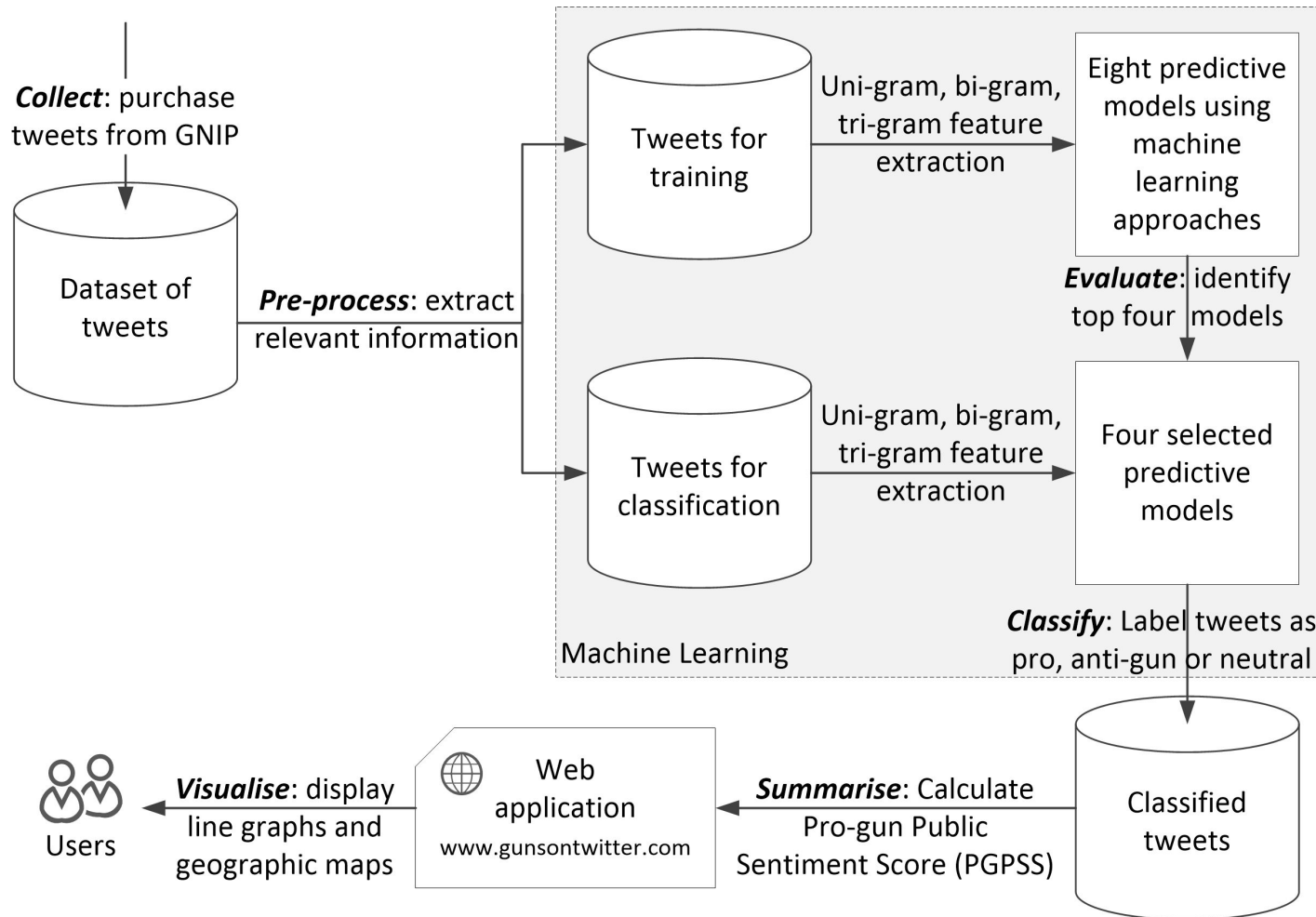
Methodology

- Pro-Gun Public Sentiment Score
  - Baseline $$PGPSS_1 \;\; = \;\; \frac{count_{(g,t)} \; (positive \; tweets)}{count_{(g,t)} \; (negative \; tweets)}$$

  - Correction for Volume of Tweets

$$PGPSS_2 \;\; = \;\; \frac{count_{(g,t)} \; (positive \; tweets)}{count_{(g,t)} \; (negative \; tweets)} \; * \; \frac{count_{(g,t)} \; (tweets)}{count_{(t)} \; (tweets)}$$

  - Correction for Volume of Tweets & Population

$$PGPSS_3 \;\; = \;\; \frac{count_{(g,t)} \; (positive \; tweets)}{count_{(g,t)} \; (negative \; tweets)} \; * \; \frac{count_{(g,t)} \; (tweets)}{count_{(t)} \; (tweets)} * \frac{population_g}{population}$$

g: geographic region   t: time frame

Outline

- Motivation & Contribution

- Methodology

- Machine Learning Approaches

- Case Study: Sandy Hook Elementary School Shooting

- Limitations

- Conclusion

- Q & A

Machine Learning Approaches

- Feature Extraction
  - N-gram

Uni-gram	Bi-gram	Tri-gram
Not	Not sure	Not sure if
sure	sure if	sure if gun
if	if gun	if gun shot
gun	gun shot	gun shot or
shot	shot or	shot or fire
or	or firework	or firework
firework		

IEEE
Advancing Technology
for Humanity

Machine Learning Approaches

- Feature Extraction

  - Hashtags

  #PrayForNewtown, #NRA, #guncontrol

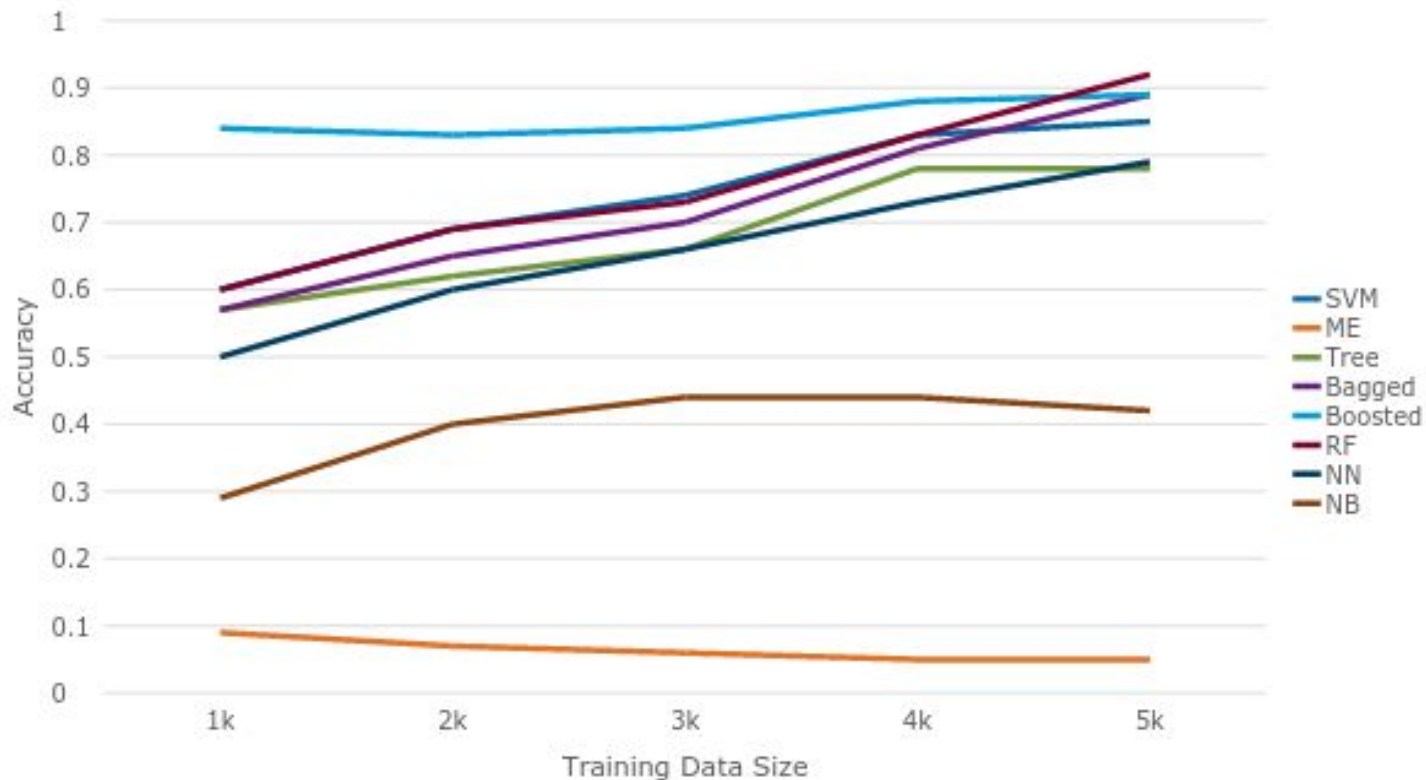 - Reply/Mention Tags

  @BarackObama, @Death, @cnnbrk

Machine Learning Approaches

- Modelling
  - Support Vector Machine (SVM)
  - Naïve Bayes (NB)
  - Maximum Entropy (ME)
  - Decision Tree (Single, Bagged, Boosted)
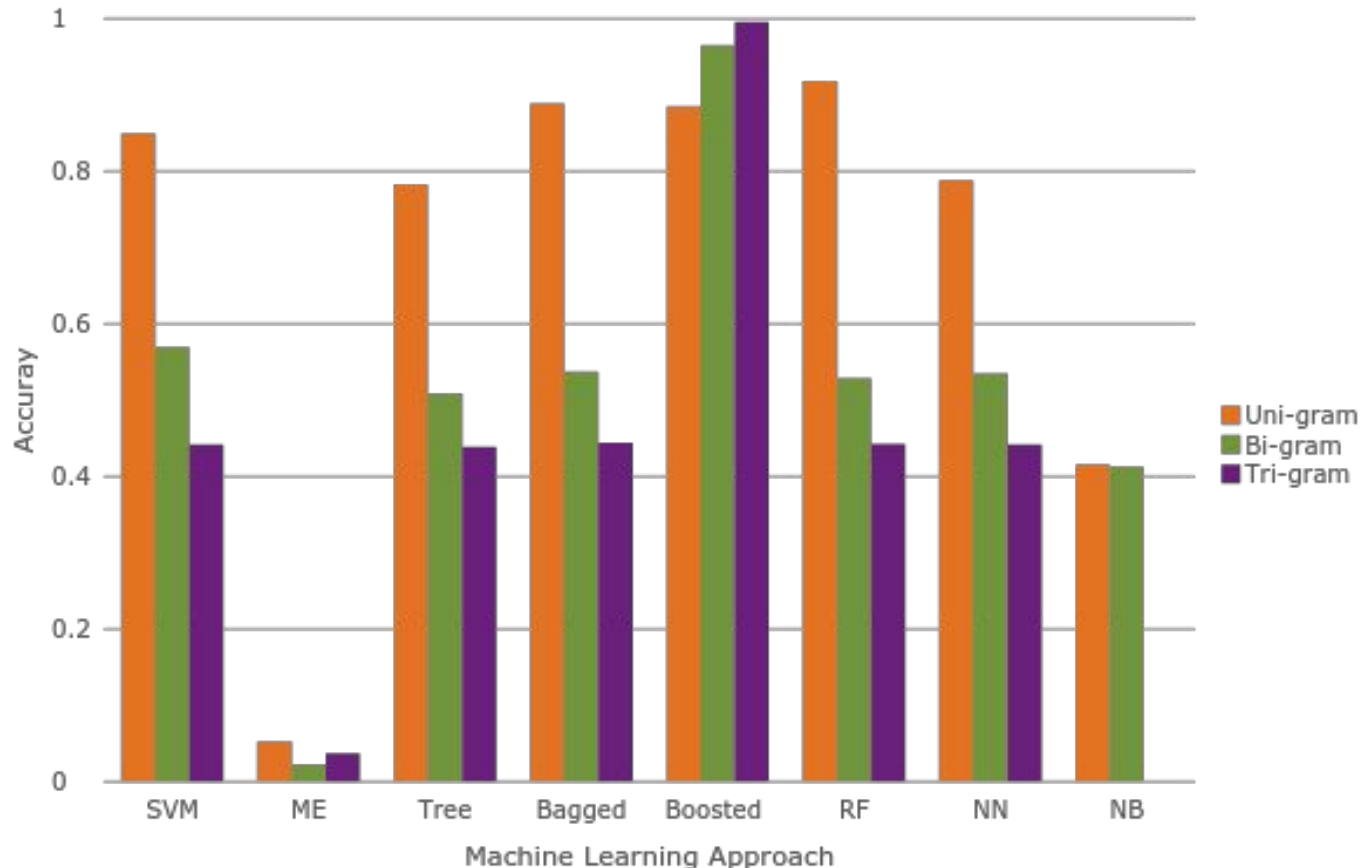 - Random Forest (RF)
  - Neural Network (NN)

Machine Learning Approaches

- Evaluation

Machine Learning Approaches

- Evaluation

Outline

- Motivation & Contribution

- Methodology

- Machine Learning Approaches

- Case Study: Sandy Hook Elementary School Shooting

- Limitations

- Conclusion

- Q & A

Case Study: Sandy Hook Elementary School Shooting

- ## Data Description
  - Timeframe

  Friday, 12/07/2012 00:00:01 GMT ~ Tuesday, 01/15/2013 23:59:59 GMT
  - Data Size   *7 million tweets*

- ## Triple-class Sentiment
  - Positive

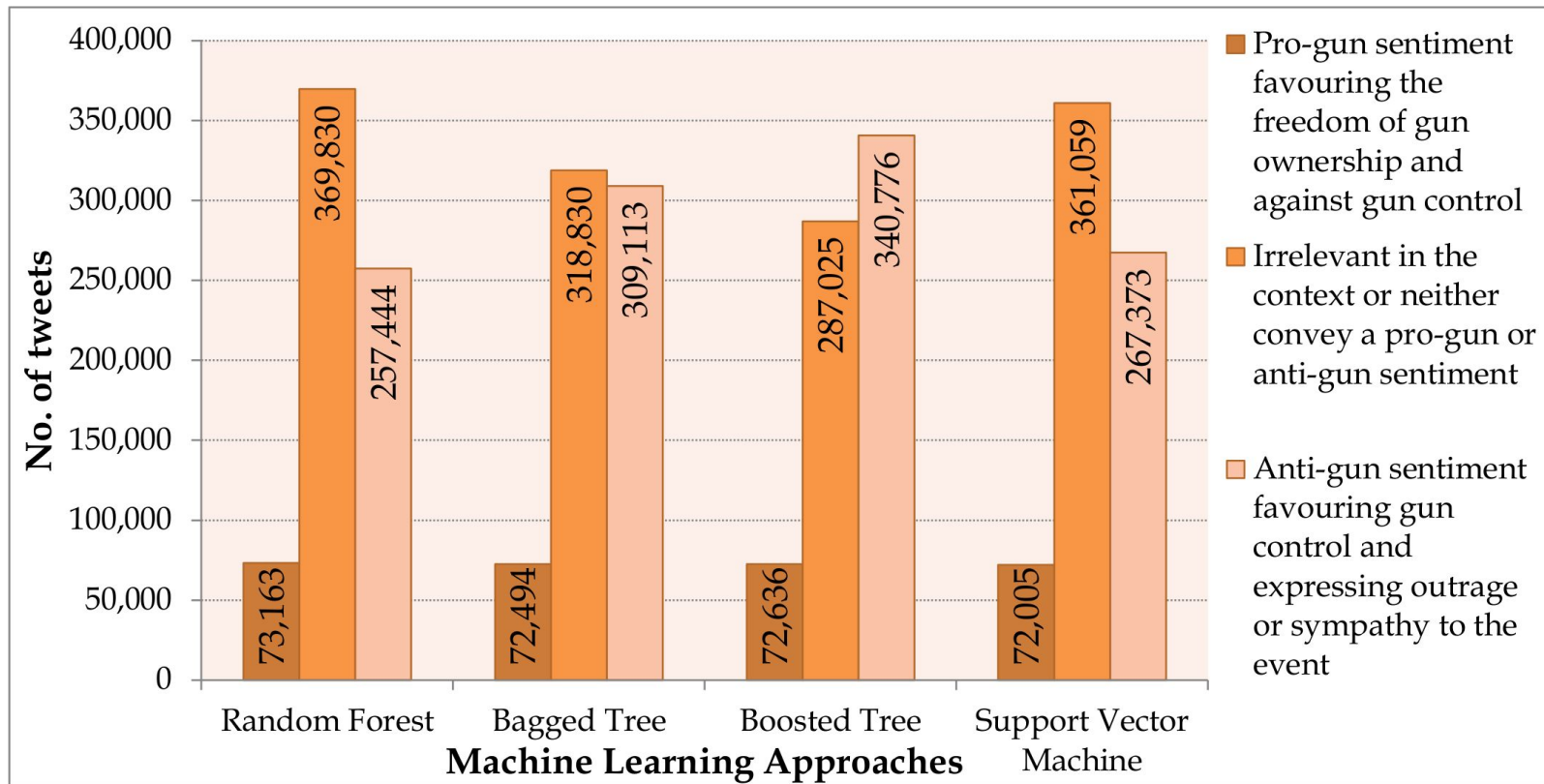 *"The only thing that stops a bad guy with a gun, is a good guy with a gun"*
  - Negative

  "We NEED strict gun control. #Newtwon"
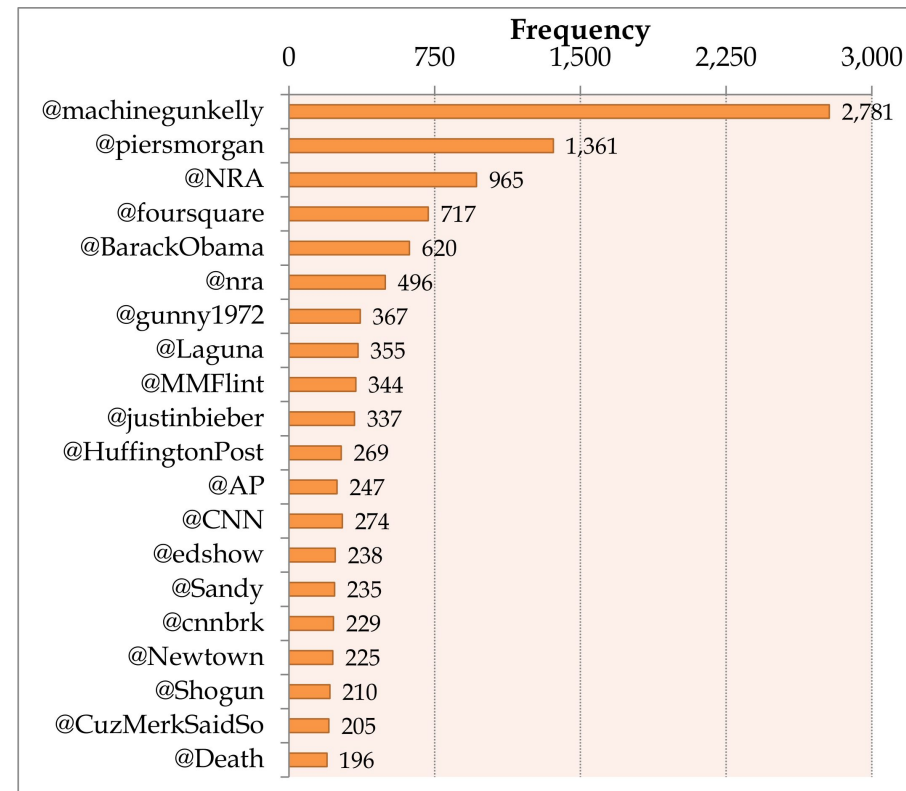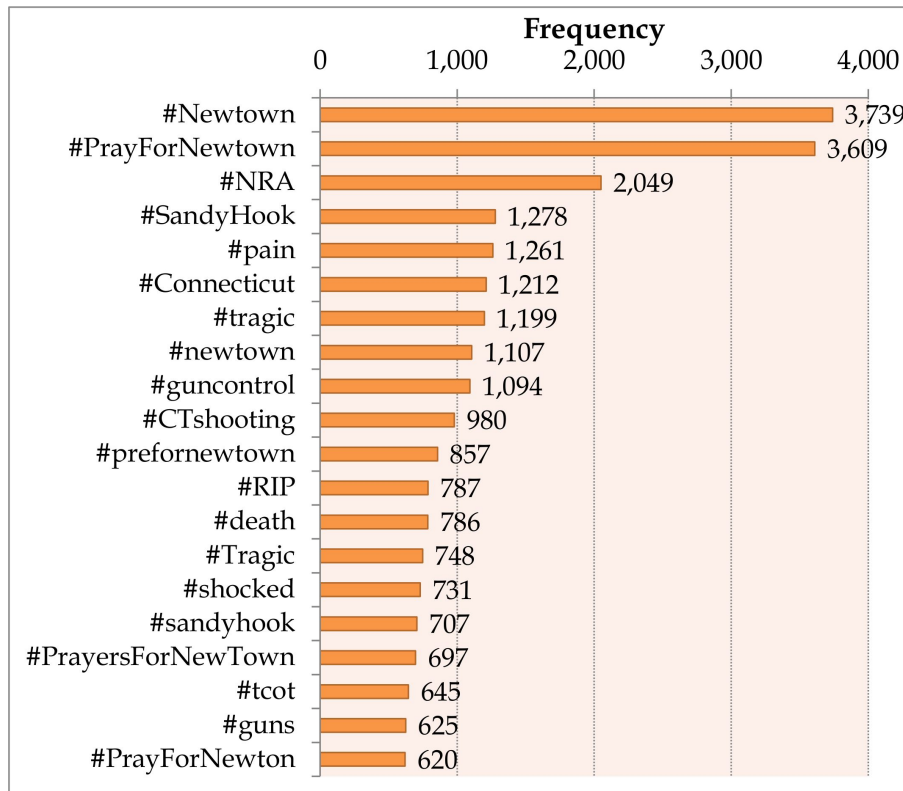 - Neutral

  "Not sure if gun shot of firework."

IEEE
Advancing Technology
for Humanity

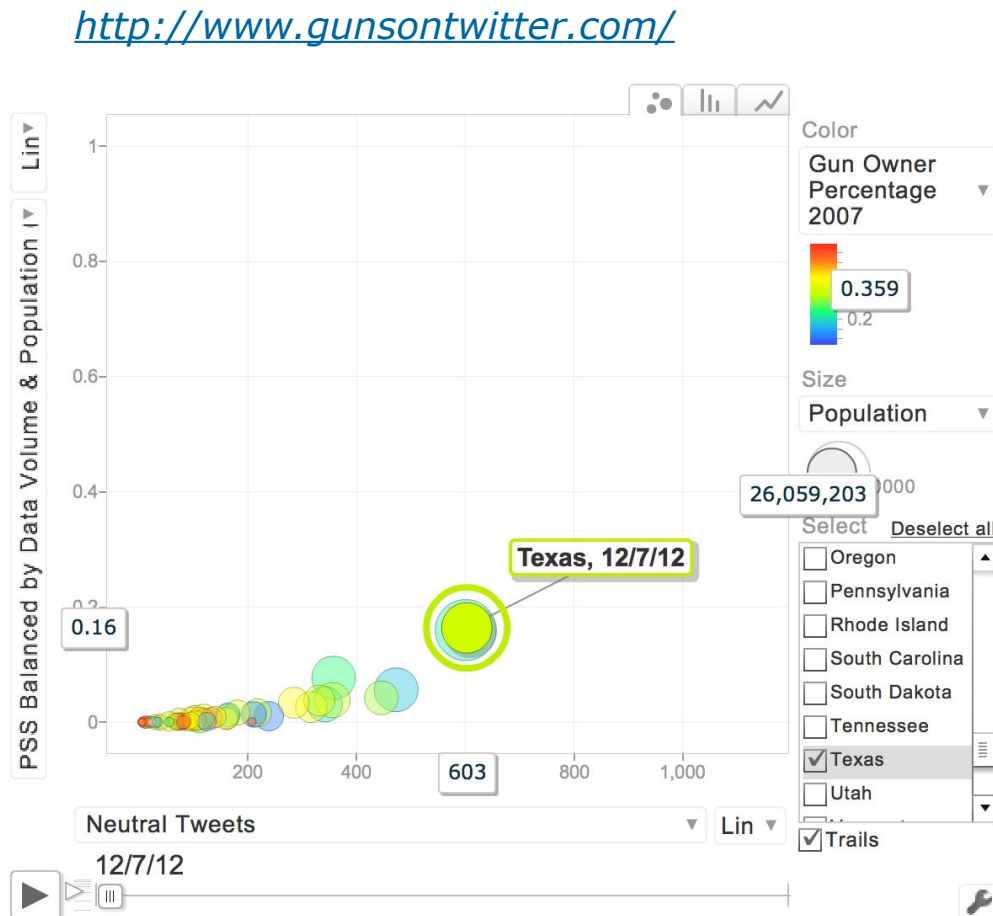# Case Study: Sandy Hook Elementary School Shooting

- Tweets Statistics

Case Study: Sandy Hook Elementary School Shooting

- Tweets Statistics

**Frequency** (chart 1)

Term	Frequency
#Newtown	3,739
#PrayForNewtown	3,609
#NRA	2,049
#SandyHook	1,278
#pain	1,261
#Connecticut	1,212
#tragic	1,199
#newtown	1,107
#guncontrol	1,094
#CTshooting	980
#prefornewtown	857
#RIP	787
#death	786
#Tragic	748
#shocked	731
#sandyhook	707
#PrayersForNewTown	697
#tcot	645
#guns	625
#PrayForNewton	620

Frequency (chart 2)

Term	Frequency
@machinegunkelly	2,781
@piersmorgan	1,361
@NRA	965
@foursquare	717
@BarackObama	620
@nra	496
@gunny1972	367
@Laguna	355
@MMFlint	344
@justinbieber	337
@HuffingtonPost	269
@AP	247
@CNN	274
@edshow	238
@Sandy	235
@cnnbrk	229
@Newtown	225
@Shogun	210
@CuzMerkSaidSo	205
@Death	196

IEEE
Advancing Technology for Humanity

Case Study: Sandy Hook Elementary School Shooting

- Visualisation http://www.gunsontwitter.com/
  - Motion Chart

Case Study: Sandy Hook Elementary School Shooting

- Visualisation
  - Line Graph
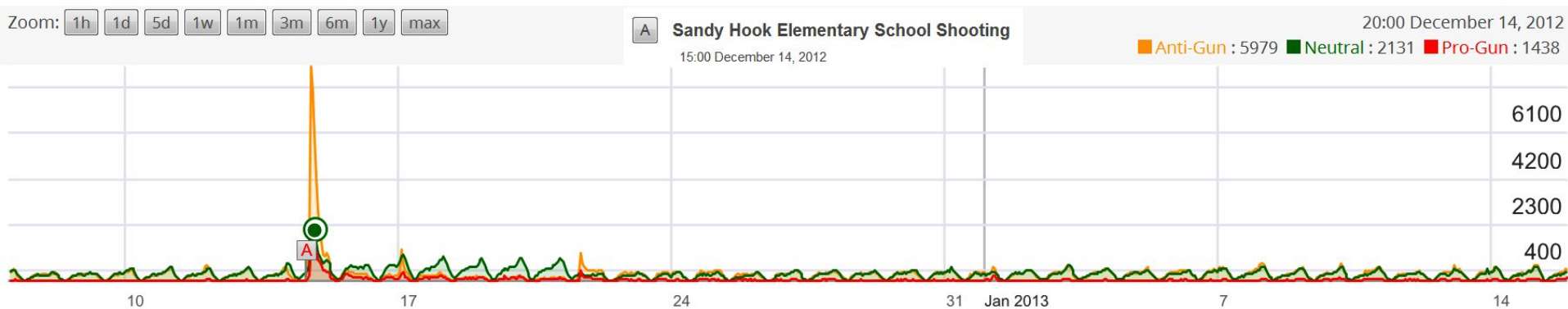
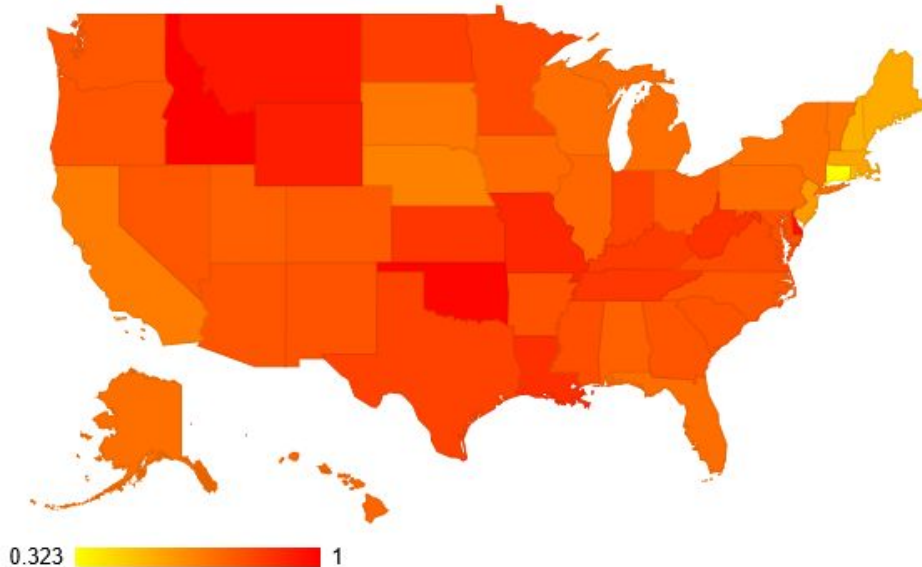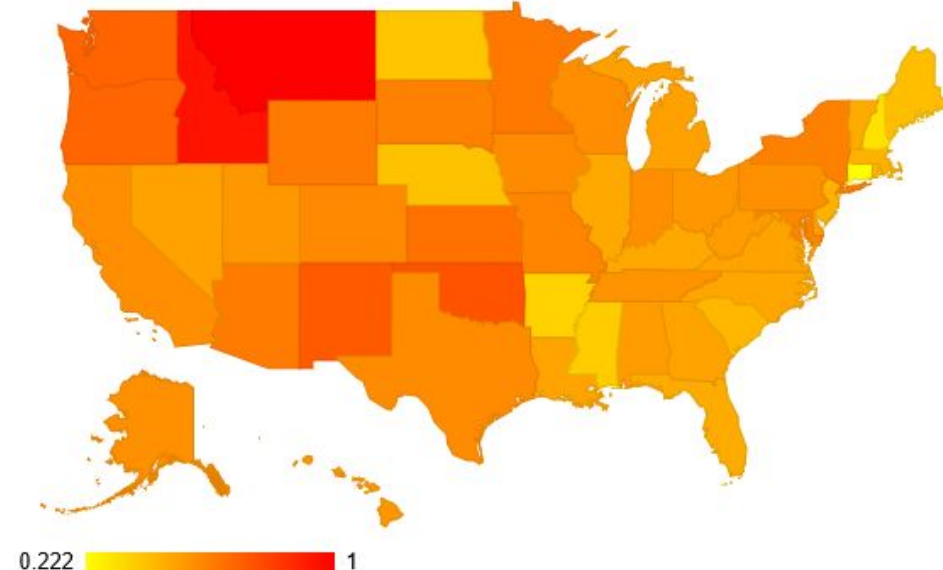# Case Study: Sandy Hook Elementary School Shooting

- Visualisation
  - Geo Map

Baseline PGPSS

0.323 ▮ 1

12/07/2012 ~ 01/15/2013

0.222 ▮ 1

12/13/2012 ~ 12/15/2012

◆IEEE
Advancing Technology for Humanity

Outline

- Motivation & Contribution

- Methodology

- Machine Learning Approaches

- Case Study: Sandy Hook Elementary School Shooting

- Limitations

- Conclusion

- Q & A

Limitations

- Size of Twitter Corpus

- Complexity of Feature Selection
    - emoticon
    - Part-Of-Speech tagging

- Trade-off between Performance & Computing Power

IEEE
Advancing Technology
for Humanity

Outline

- Motivation & Contribution

- Methodology

- Machine Learning Approaches

- Case Study: Sandy Hook Elementary School Shooting

- Limitations

- Conclusion

- Q & A

Conclusion

- This paper

  - Evaluates of machine learning approaches for twitter sentiment analysis

  - Investigates tweets' relevance to gun violence

  - Visualises public sentiment related data on multiple geographic/temporal level interactively

Q & A

IEEE
Advancing Technology
for Humanity