



Can Social Media measure Air Quality?

Prithviraj Pramanik

**Visiting Doctoral Fulbright Fellow,
Discovery Partners Institute/Argonne National Laboratory,**

**PhD Candidate
Department of Computer Science & Engineering
National Institute of Technology Durgapur, India**

Online @ Citizen Science with Application to Nuclear, Seismic
and Air Quality Monitoring (smr3596), ICTP, Trieste, Italy

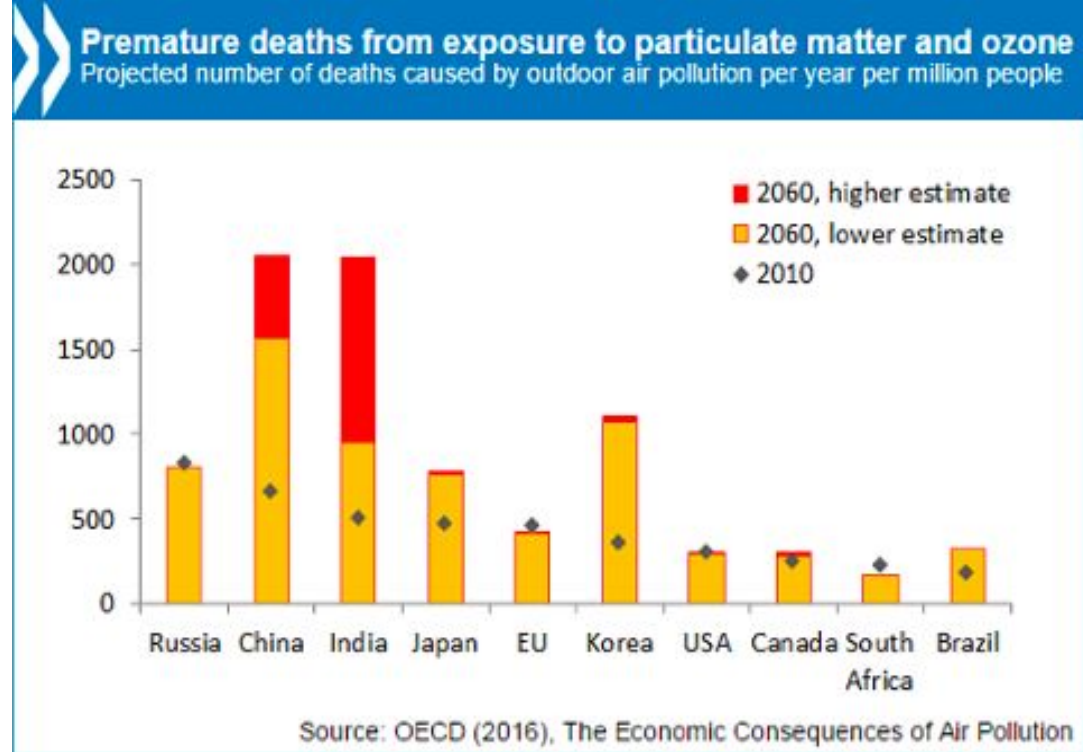
18.03.2021

Twenty-two of the world's 30 most polluted cities are in India: World Air Quality Report

It said that despite the improvement, Delhi ranked as the 10th most polluted city and the top polluted capital city in the world.



Published: 16th March 2021 05:05 PM | Last Updated: 16th March 2021 05:06 PM



- We breathe on an average 20,000 times a day. 10,000 liters of air is pass through our lungs with majority of them being pollutants.¹
- Air pollution is mainly associated with asthma and cardiovascular diseases.²
- Air pollution causes 9 times as many premature deaths than traffic crashes.³

¹<https://seetheair.wordpress.com/2017/01/31/we-breathe-11000-liters-of-air-daily/>

²https://www.env-health.org/IMG/pdf/8_patients_with_cardiovascular_disease_-_info_leaflet_-_en.pdf

³<http://www.vancouver.sun.com/health/pollution+nine+times+deadlier+than+crashes+researchers/9061897/story.html>

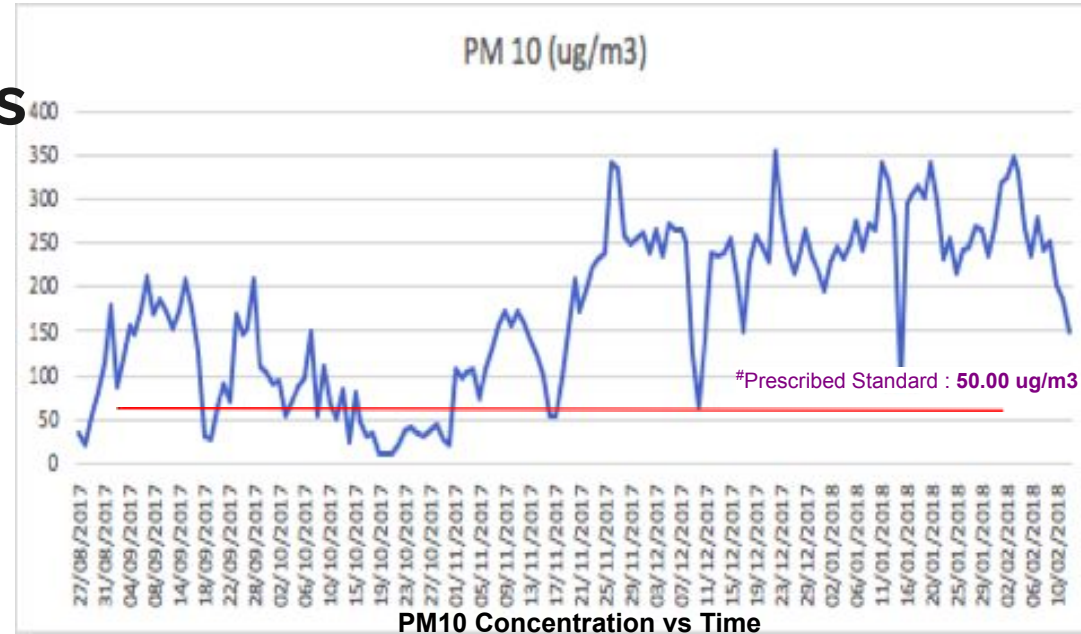
<https://www.newindianexpress.com/nation/2021/mar/16/twenty-two-of-the-worlds-30-most-polluted-cities-are-in-india-world-air-quality-report-2020-2277377.html>

<https://www.theguardian.com/environment/2020/apr/11/positively-alpine-disbelief-air-pollution-falls-lockdown-coronavirus>

27-08-2017 to 14-2-2018

*Source: Central Pollution Control Board

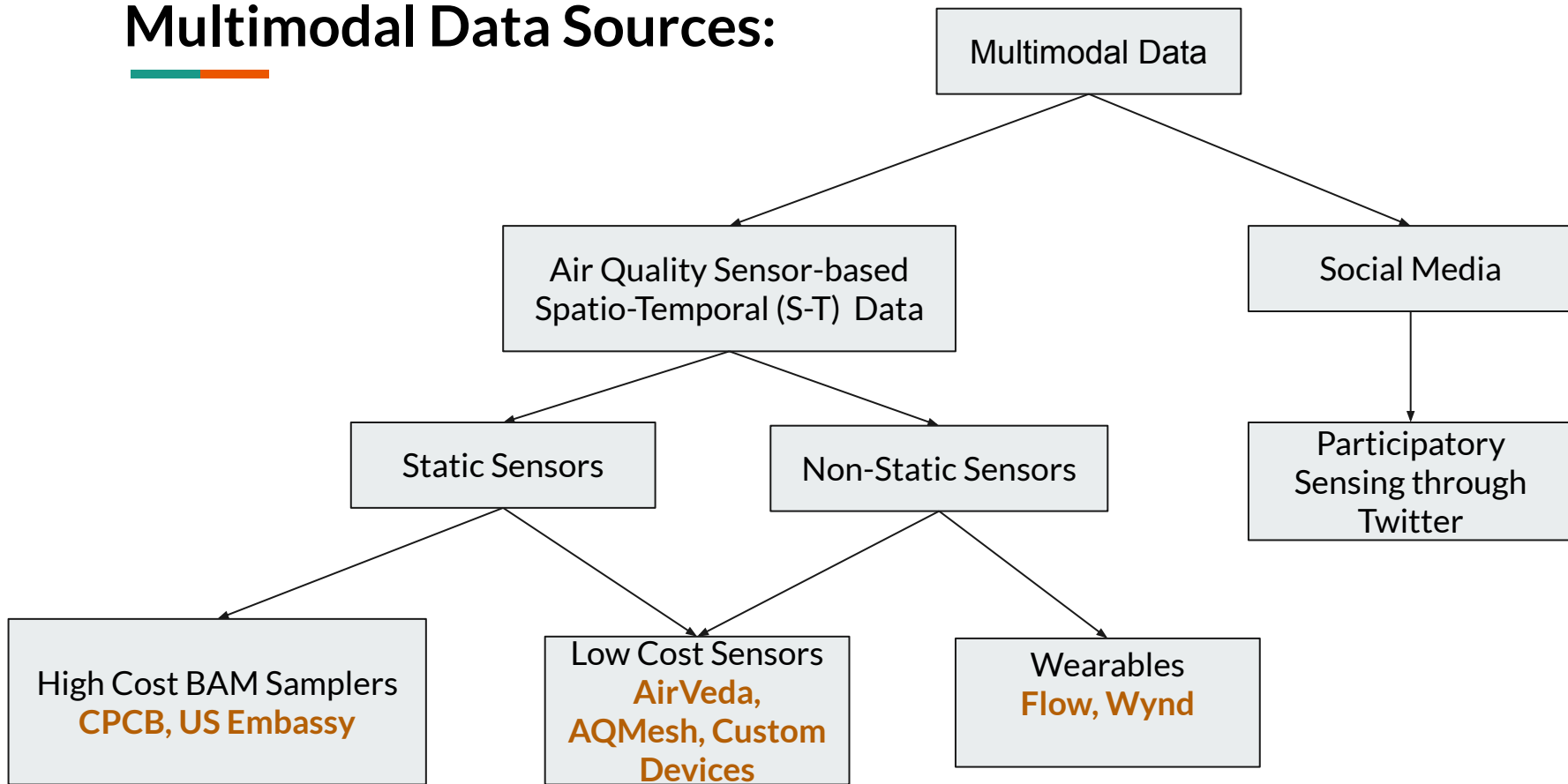
Durgapur: State of Affairs



Primary Pollutants: PM_{2.5}, PM₁₀, SO₂, NO₂, CO, O₃
Absence of Infrastructure leads to multimodal data source

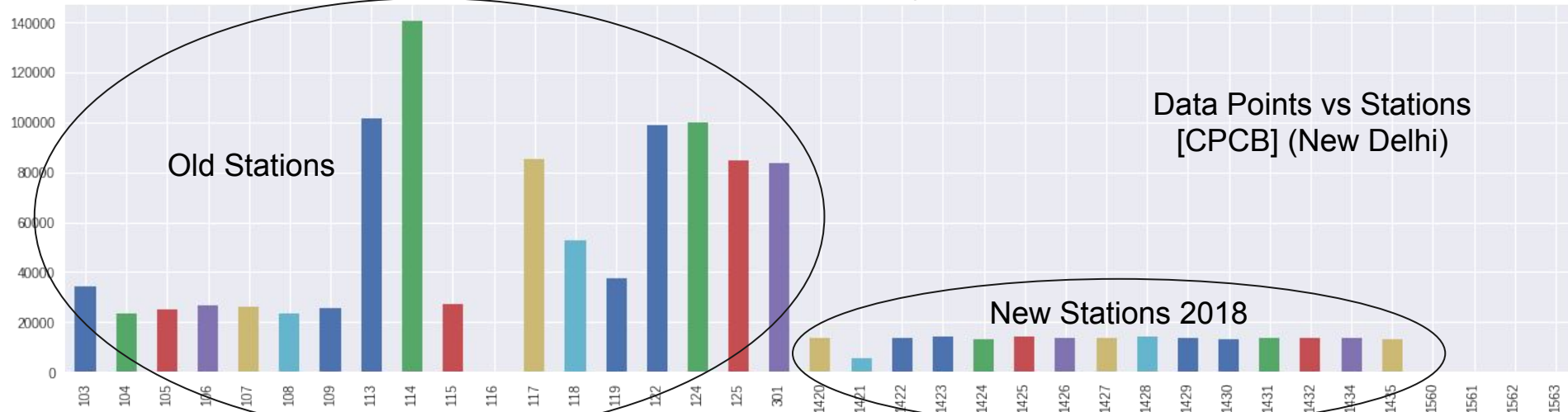
#WHO Air quality guidelines for particulate matter, ozone, nitrogen dioxide and sulfur dioxide Global update 2005
Summary of risk assessment

Multimodal Data Sources:



Data Source: CPCB, India

- Has only 3% (131) of the required 4000 Environment Monitoring Stations.
- New Delhi has been considered the test-bed
- We have collected data from Jan. 2015(5 Stations) - Aug. 2018(31 Stations) [Number of data points for each station is shown below]
- Suffers heavily the issue of missing values
- Primary Pollutants considered are PM_{2.5}, PM₁₀, CO, O₃, SO₂, NO₂



Data Source: Low Cost Commercial & Custom Devices

- AirVeda (PM2.5, PM10, CO₂)
- Flow (VOC, PM2.5, NO₂)
- Custom Device (CO, NO₂, CO₂, SO₂, PM2.5, PM10)



<https://flow.plumelabs.com/>
<http://www.airveda.com/airveda-pm-co2-air-quality-monitor>

Arindam Ghosh et al., "Air Quality Monitoring using Low-cost Sensing Devices", MobiSys Poster 2016.

Praveen Kumar Sharma et al., "On Detecting Acceptable Air Contamination in Classrooms using Low Cost Sensors ", COMSNETS WACI 2017.

Data Source: Social Media Feed from Twitter

- Based on hashtags (**Air Pollution Delhi**)
- Collected from **January 2015 - August 2018** (160K)
- Sample Tweets:

@PetroleumMin @AgriGol @narendramodi @pcraindia Finger crossed .We will see the result how effectively it has been implemented in Northern state of India. As the crop residue play a important role of causing air pollution in Capital Delhi during winter session.

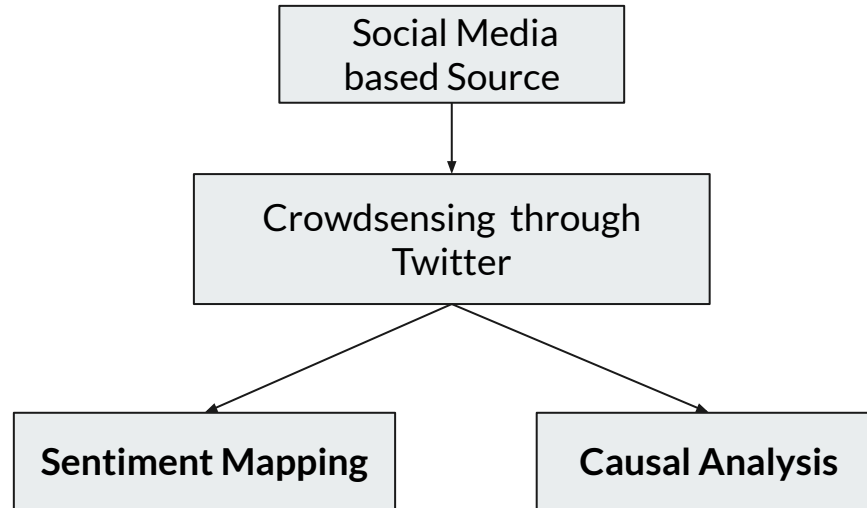
From Delhi air to Bengaluru lakes, will environment finally be an election issue in 2019? #airpollution #pollution #nopollution #plasticpollution <http://citizenmatters.in/delhi-air-pollution-bengaluru-lakes-environment-lok-sabha-2019-election-7722>

#Delhi pollution: Parl panel asks govt to revisit tree cutting policies. Emphasising strict and time bound steps to deal with growing menace of air pollution in Delhi, Parliamentry penal asked Delhi Govt to further stepup its efforts

Over 17 lakh cases of acute respiratory infection were reported in Delhi from 2013 to 2017. It's disappointing that the government is still not understanding the urgency and looking for solutions to curb India's air pollution crisis. #MyRightToBreathe

Air pollution in NCR Delhi is health emergency but no political parties paying attention on this.@PMOIndia @drharshvardhan @ArvindKejriwal
Delhi pollution: 17 lakh respiratory infections, 981 deaths in 5 years in Capital | Delhi News - Times of India

Social Media Based



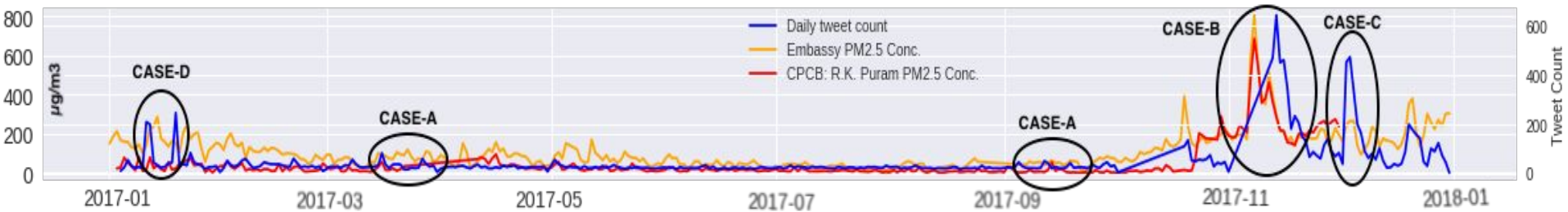
- Sentiment Mapping: Can the latent sentiment of the social media users posting air pollution related social messages help in pollution estimation?
- Causal Analysis: Analysis of which is causing what in relating S-T pollution measurements with Air Pollution based messages.



Evaluation

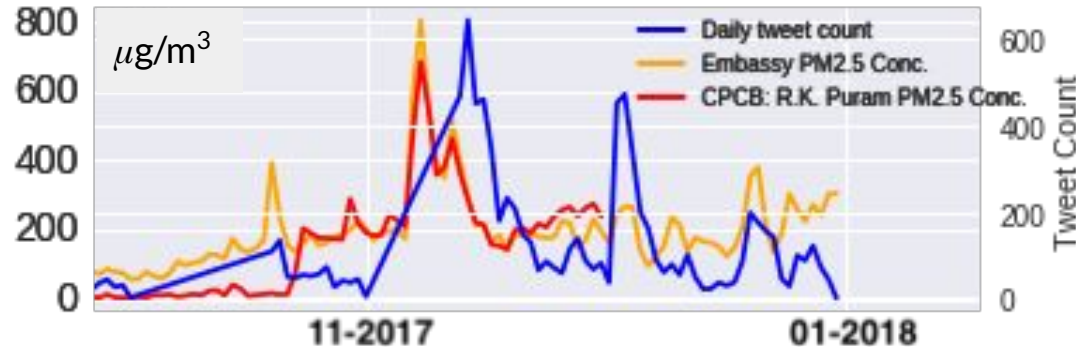
- New Delhi taken as test-bed
- PM2.5 used as Major Pollutant
- Preliminary Results show it can roughly detect all the major events related to Delhi Air Pollution

Exploratory Analysis of Twitter feed with Ground Truth



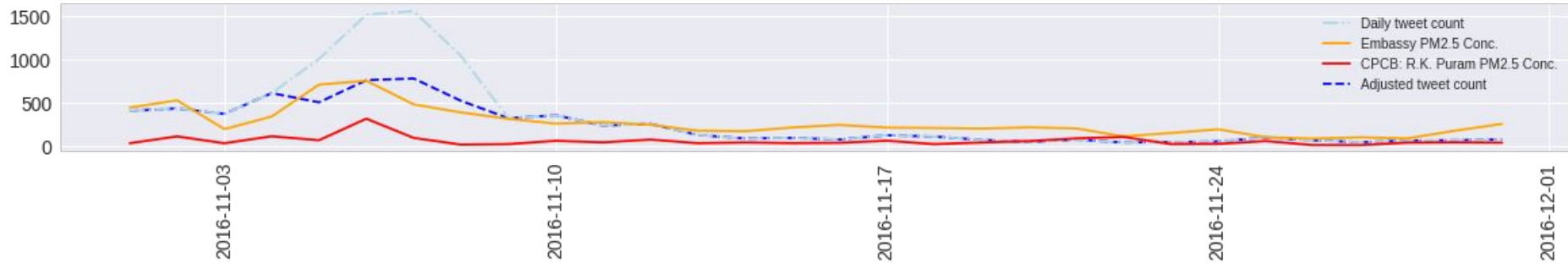
- Case A: the inherent changes Social Perception with the rise & fall of PM2.5 (a primary pollutant).
- Case B: the time drift between social perception and actual ground truth.
- Case C: the sudden spurt of Tweets without air pollution increment.
- Case D: sudden peaks in Tweets after highly polluted days.

Drift: Twitter feed vs Ground Truth



The drift from the air sensor ground truth with Tweets in the initial level when heavy pollution occurs: As seen the figure, Case B occurs here. The major cause behind this is the social perception takes more time in forming than the sensors.

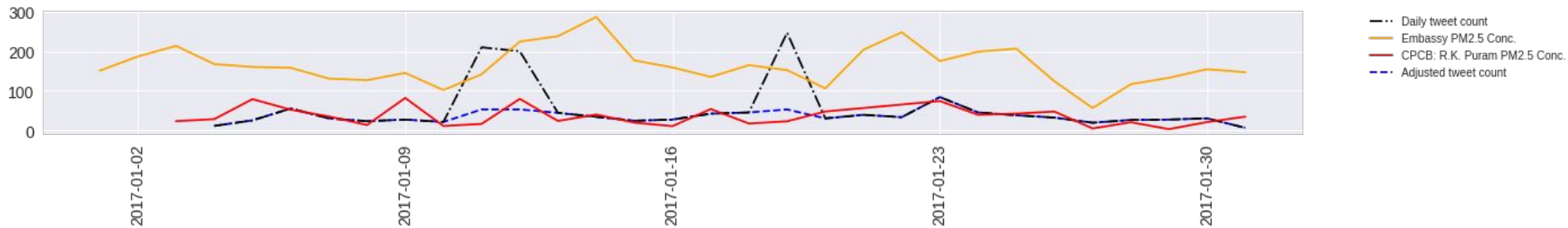
Trending Topic in Social Media



The effect of being a trending topic on highly polluted days: When social media starts trending due to air pollution instead of the raw social perception other irrelevant topics emerge and they try to piggyback on these events to draw attention as seen in Case B in the figure. Identifying the signature of these tweets & users from Case A can help us normalize the latent social sensing for better accuracy by removing these noises

'#BREAKING Delhi CM ArvindKejriwal calls for an emergency cabinet meeting at 12:30 PM on issue of air pollution in Delhi.',
 'Delhi Air Pollution: Ranji Players Wear Masks During Practice Session: <http://youtu.be/zr4r0hkcsIQ?a\xa0> via @YouTube',
 'Delhi air pollution closes schools for three days
http://www.theguardian.com/world/2016/nov/06/delhi-air-pollution-closes-schools-for-three-days?CMP=Share_AndroidApp_Tweet\xa0...,
 "#India #National Air pollution in Delhi: Cong pulls up 'habitual' AAP <http://ift.tt/2eL8UNU\xa0>",
 "@IndiaToday Thanks 2 pollution Rs 12.50 per bottle of fresh air Canada's Vitality Air seeks market in New Delhi
<http://m.indiatoday.in/story/canned-air/1/657956.html\xa0...> via @indiatoday",
 '@NewsX @newsx @TimesNow @CNNnews18 @IndiaToday 24x7 Delhi air pollution is over the top. India is not Delhi. Visit Korba & coal,power plant cities',
 'RT @thinkprogress: Delhi applies emergency measures over toxic air pollution
<https://thinkprogress.org/delhi-applies-emergency-measures-over-toxic-air-pollution-b0c686ac5d05#.dud846q45\xa0...> pic.twitter.com/SVTYH3otey',
 'Air pollution: Delhi shuts schools, bans construction work - The Times of India <http://toi.in/fApAoY65/a18cq\xa0> via @timesofindia',
 'Paramilitary fights #Delhi air pollution despite shortage of masks.'

Fact vs Opinion



The effect of high-profile air quality reports/ruling on regular days: We are predominantly focusing on the latent sentiment sensing, but when some Landmark reports like World Health Organization (WHO), Supreme Court rulings on Air Pollution & Environment, etc. occur, as seen in Case C in the figure, there is a massive outbreak of fact-based tweets on air pollution. **Challenge lies in weeding out these Tweets to rule out the false positive of air pollution.**



Learnings of using Twitter for Participatory Sensing

- i) Drift in the Tweets with respect to ground truth
- ii) Mitigation of Adversities of trending
- iii) Identifying Facts vs Opinion
- iv) Bot based social awareness
- v) Localisation Tweets for fine grained estimation

These problems require some novel approaches for using Twitter for Air Quality Estimation

Jiang et al. (PlosOne, 2015) has used social media for pollution sensing in Sina-Weibo.



Conclusion

The preliminary analysis of using Twitter for **participatory Sensing** has been performed.

Analysis for automatic classification needs to be performed. Other avenues like

- proactive measures for interactive Tweet Bot for Social Awareness
- the problem of localization (only 3-5% Tweets Geotagged by default)

have been kept in the future scope of work.



Thank You :)

Any Question?
prithvirajpramanik@yahoo.co.in



References

- Zheng et al., “U-air: When urban air quality inference meets big data,” in Proceedings of the 19th ACM SIGKDD. ACM, 2013, pp. 1436–1444.
- Y. Zheng et al., “Urban computing: concepts, methodologies, and applications,”ACM Transactions on Intelligent Systems and Technology(TIST), vol. 5, no. 3, p. 38, 2014
- J. Y. Zhu et al., “An extended spatio-temporal granger causality model for air quality estimation with heterogeneous urban big data,”IEEE Transactions on Big Data, vol. 3,no. 3, pp. 307–319, Sept 2017.
- W. Jiang et al., “Using social media to detect outdoor air pollution and monitor air quality index (aqi): A geo-targeted spatio temporal analysis framework with sina weibo(chinese twitter),”PLOS ONE, vol. 10, pp. 1–18, 10 2015.
- WHO Air quality guidelines for particulate matter, ozone, nitrogen dioxide and sulfur dioxide Global update 2005 Summary of risk assessment
- Arindam Ghosh et al, “Air Quality Monitoring using Low-cost Sensing Devices”, MobiSys Poster 2016.
- Praveen Kumar Sharma et al., “On Detecting Acceptable Air Contamination in Classrooms using Low Cost Sensors”,COMSNETS WACI 2017.
- C. W. J. Granger, Investigating Causal Relations by Econometric Models and Cross-spectral Methods, Econometrica, Vol. 37, No. 3 (Aug., 1969), pp. 424-438