Potholes and Bad Road Conditions - Mining Twitter to Extract Information on Killer Roads

Swati Agarwal1 Nitish Mittal2 Ashish Sureka3


Research Motivation and Aim

- Increasing trend of adaptation of social media by Indian Government and public agencies to reach out to the people, public citizens use Twitter to post their complaints and report the incidents to the concerned authorities.
- Complaints on killer roads contain the information about road irregularities and other issues causing high risks and discomfort to the citizens.
- To develop a system that can automatically identify complaint reports and overcome the challenge of manual inspection.
- To extract important information from tweets such as the exact geographical location which can be used to locate the fault

Bad Road Conditions Complaints on Twitter

- Inhendreds
  - @NHNINDIA @inhindia big pothole on service lane near hyatt gurgaon on NH8. Cons r getting stuck daily. Pls help in fixing it.
  - @Vijay_Ind @VijayIndio15 Potholes and Bad Road Conditions- Mining Twitter to Extract Information on Killer Roads

Identifying the Geographical Location

- Using feature extraction methods like Rule-Based, Core NLP, POS Tagging, Semantic Similarity, and Lexical Knowledgebase, the complaint reports in our Experimental Dataset can be used to discover insights from road related concepts.
- We propose a content-disambiguation model to enrich our linguistic features for identifying useful and less informative reports.

Experimental Dataset

- First study on mining citizen's complaints and reports on killer roads posted on official Twitter handle of Government.
- We investigate the efficacy of spatial, contextual and linguistic features for identifying useful information from complaint posts. We build a text-analysis based model to enrich spatial information from complaint posts. We build a text-analysis based model to enrich spatial and linguistic features (geographical location metadata) in a tweet that can be used to discover insights from noisy data.
- We propose an approach for identification of complaints reported in road irregularities and bad road conditions.
- We address the challenge of noisy and user-generated data by performing syntactic enrichment on raw tweets.
- We propose to use the application of ConceptNet lexical Knowledgebase, named entity recognizer and geocoding location APIs to extract important components of a killer road complaint.
- In order to conduct our experiments for this research, we create the very first database of citizens' complaints on killer roads and highways reported to official public agencies on Twitter.
- We make our dataset publicly available for the research community so that our results can be used for benchmark, comparison and further extension.

High-Level Diagram of the Proposed Framework

- Example of a Complaint on a Killer Road: "A farmer died y'day while jumping pothole, road condition is horrible, cant even work, pls help."
- Travelling on NH3, not a single street light between gadgetsab and manut on such a busy route @WCM/Indirapuram @NHNINDIA
- Figura: Concrete Examples of Complaints on Killer Roads and Citizens' Discomfort Reported to Official Twitter handle of Ministry of Transport, Road and Highways, Government of India- Addressing Various Road and Transport Related Issues such as Pothole, Dysfunctional Streetlights, Unplanned Construction on Highways and Breakers near Intersection and Sharp Turns.

Distribution of Distinct Geographical Locations

- We propose an approach for identification of complaints reported in road irregularities and bad road conditions.
- We address the challenge of noisy and user-generated data by performing syntactic enrichment on raw tweets.
- We propose to use the application of ConceptNet lexical Knowledgebase, named entity recognizer and geocoding location APIs to extract important components of a killer road complaint.
- In order to conduct our experiments for this research, we create the very first database of citizens' complaints on killer roads and highways reported to official public agencies on Twitter.
- We make our dataset publicly available for the research community so that our results can be used for benchmark, comparison and further extension.

Summary and Conclusion

- We propose an approach for identification of complaints reported in road irregularities and bad road conditions.
- We investigate the efficacy of spatial, contextual and linguistic features for identifying useful information from complaint posts. We build a text-analysis based model to enrich spatial and linguistic features (geographical location metadata) in a tweet that can be used to discover insights from noisy data.
- We propose an approach for identification of complaints reported in road irregularities and bad road conditions.
- We address the challenge of noisy and user-generated data by performing syntactic enrichment on raw tweets.
- We propose to use the application of ConceptNet lexical Knowledgebase, named entity recognizer and geocoding location APIs to extract important components of a killer road complaint.
- In order to conduct our experiments for this research, we create the very first database of citizens' complaints on killer roads and highways reported to official public agencies on Twitter.
- We make our dataset publicly available for the research community so that our results can be used for benchmark, comparison and further extension.

Suggestions and Collaboration Possibilities

- Please email: Swati Agarwal swatiga08@bits-pilani.ac.in, Nitish Mittal nitishmittal94@gmail.com, Ashish Sureka ashish.sureka@ashoka.edu.in
- Author Homepage: http://www.swati-agarwal.in and http://www.ashish-sureka.in


Figura: Performance Results of Proposed Approach

Figura: Distribution of Distinct Geographical Locations (Places and States) Identified in the Complaint Reports in our Experimental Dataset.

Presentation at CoDS 2018

Saturday - January 13, 2018, 16:00 - 17:30 Session (Research Track - 5)