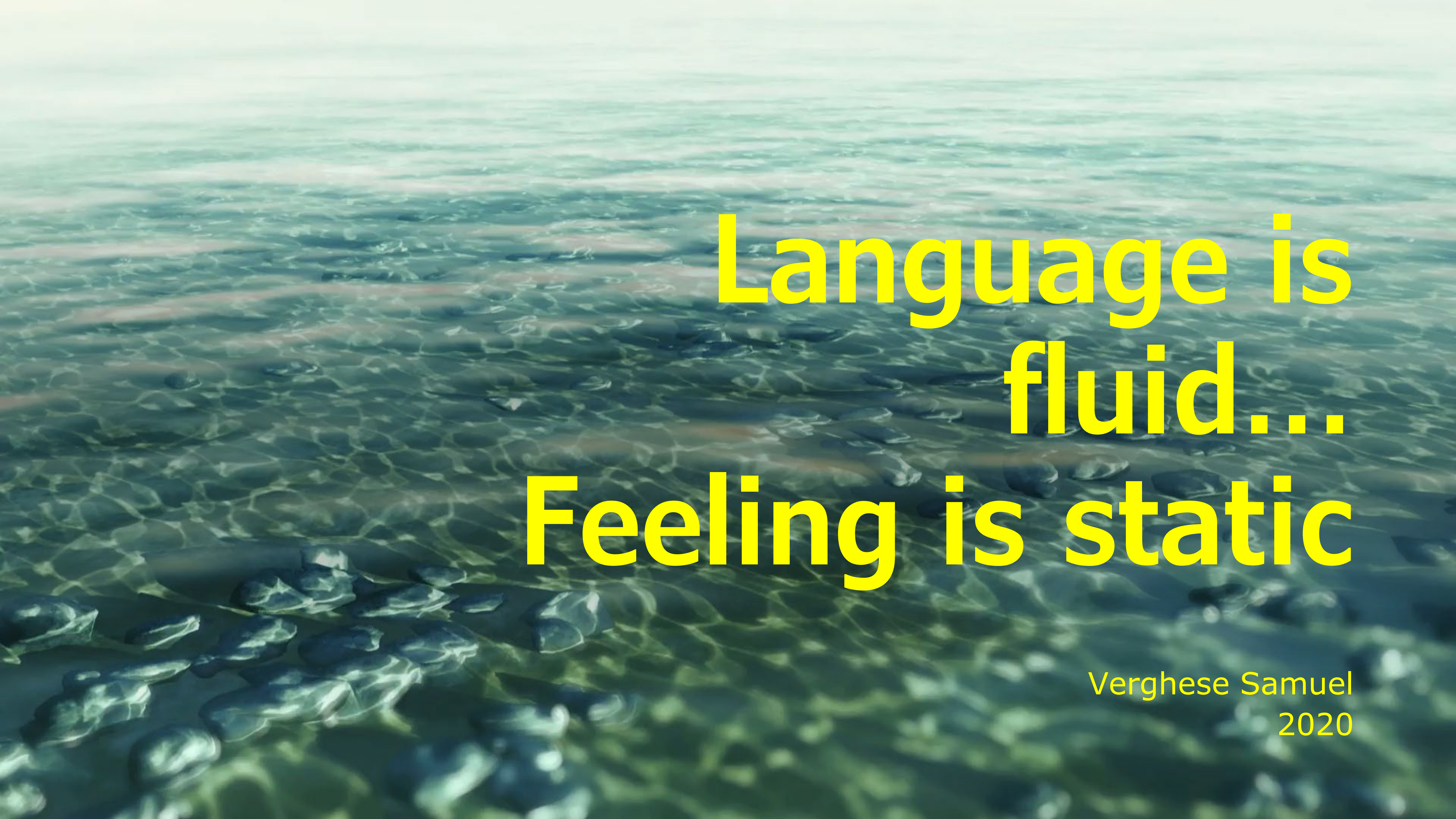


Language is fluid… Feeling is static

Verghese Samuel
2020

Problem Statement

People share their opinions on the social media. These opinions have sentiments of positive, neutral, or negative experience.

Given the time and frequency of the user driven content sharing, it is humanly not possible to evaluate the sentiments of each of the content.

How do organizations understand which of these expressed sentiments impact their brand?


Hurricane Hit…

Search Posts | united states | SENTIMENT | EXCLUDE RETWEETS | RESET

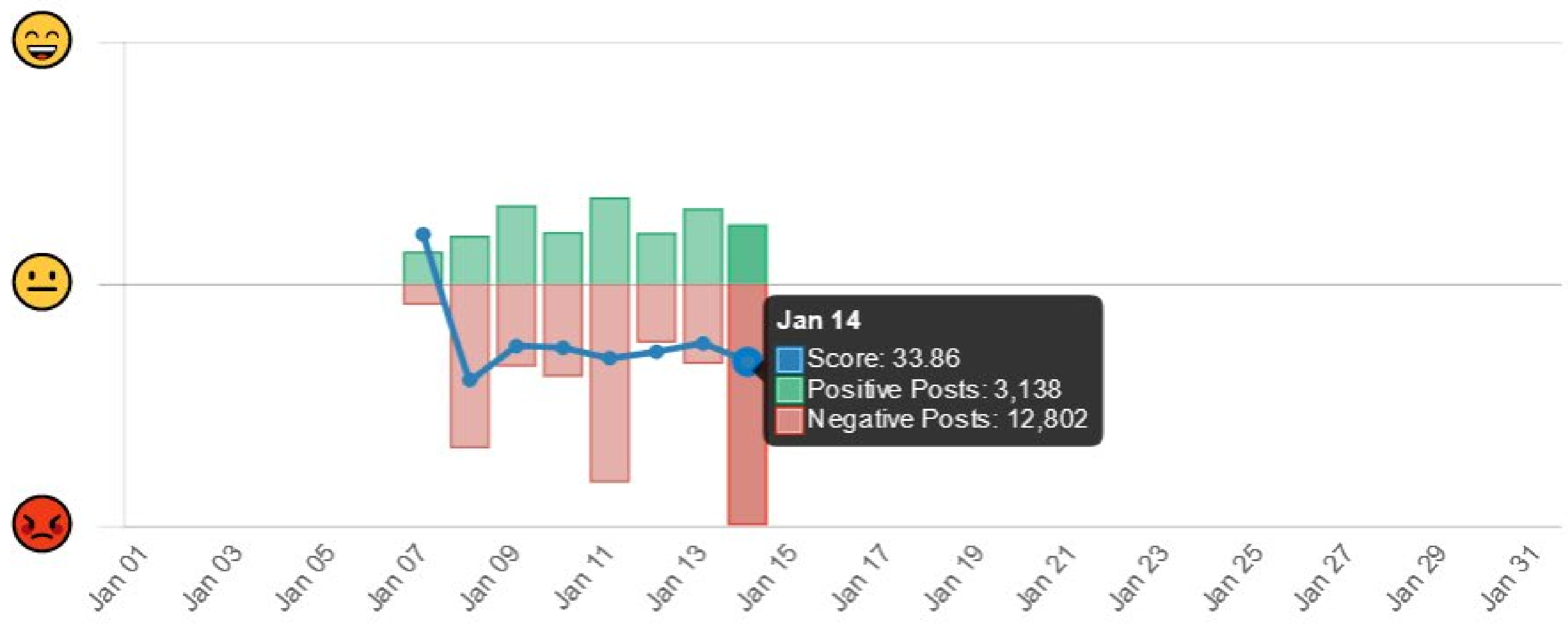
USER	POST	REGION	COUNTRY	ENGAGEMENTS
Chuck Schumer @SenSchumer	After one of the worst disasters in U.S. history, we're demanding answers. America deserves to know the truth about the Trump administration's inadequate preparedness and response to Hurricanes Irma and Maria so we can ensure it does not happen...	NY	United States	4,716
SteelerNation @SteelerNation	People died and lost their homes in that hurricane and Eagles fans think making a meme out of that is okay. Saints by a million	PA	United States	2,238
Joy Reid @JoyAnnReid	California, Texas, Florida and Puerto Rico hurricane victims will pay for the wall?	NY	United States	2,130
The Dodo @dodo	This baby cow was rescued from a hurricane and became best friends with a little boy 🐄❤️ https://t.co/NAsgrtZxgd	NY	United States	1,961
Adam Best	This right-wing smear is getting out of hand. Democrats were in Puerto Rico to meet with the governor about post-			

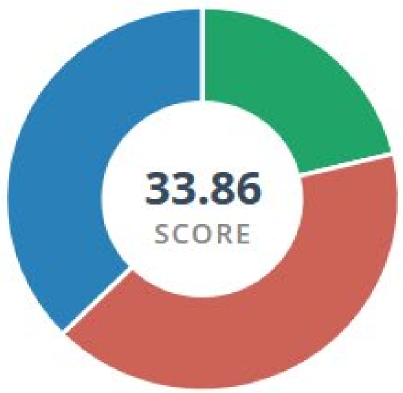
Real-time Tracker: **hurricane**

120,052	84,231	125,480	551,327,342	995,381,46
POSTS	USERS	ENGAGEMENT	REACH	IMPRESSIONS

Sentiment Timeline

hurricane





Real-time Tracker: hurricane

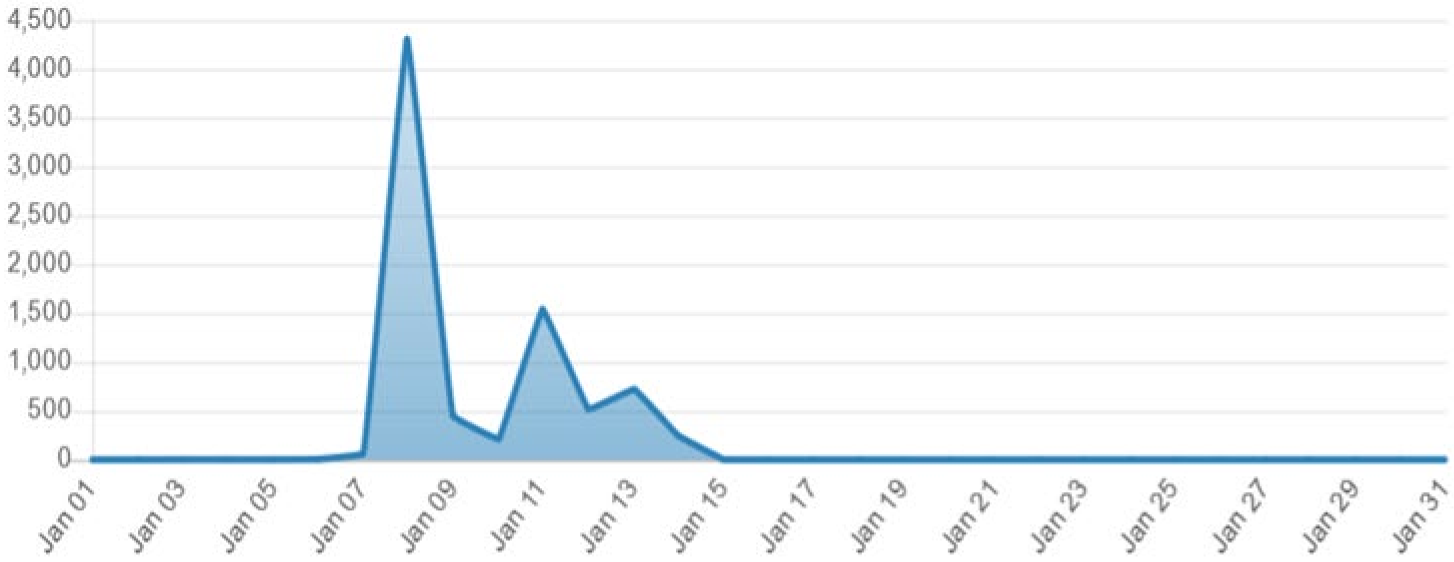
Trending Topics: Insights into conversations around this tracker

TOPICS + Add Subtopic

Top | Rising

▲ puerto	17,574
▲ rico	16,482
▲ wall	14,832
▲ trump	12,159
▲ pay	11,394
▲ america	9,658
▼ like	8,370
▼ disaster	8,056
▲ build	7,771

Mentions Over Time: disaster



DATES & TIMES ARE BASED ON YOUR LOCAL TIMEZONE: (**GMT-0800 | PACIFIC STANDARD TIME**)

Framework for Solution

DATA

MODEL
(PROBABILISTIC & CATEGORICAL)

INTERFERENCE ELEMENTS

OUTCOME & TUNING

Markov Model Structure for Information

Report WS-99-11. Compilation copyright © 1999, AAAI (www.aaai.org). All rights

e† Andrew McCallum‡† Ronald Ro

u mccallum@justresearch.com roni@cs.cr

omputer Science ‡Just Research

ellon University 4616 Henry Street

gh, PA 15213 Pittsburgh, PA 15213

ct

techniques, while well ch recognition, are just e information extraction f hidden Markov models ks, specifically focusing ure from data and how ed and unlabeled data. tructed model that con- ction field outperforms a and discuss strategies for utomatically from data. use of distantly-labeled provides a significant im- acy. Our models are ap-

distantly-labeled data consistently tion accuracy.

Hidden Markov models, while rela mation extraction, have enjoyed suc ural language tasks. They have be part-of-speech tagging (Kupiec 199 recently been applied to topic dete (Yamron *et al.* 1998) and dialog act Shriberg, & others 1998). Other sy for information extraction include th who extracts gene names and locat abstracts, and the Nymble system for named-entity extraction. Unlike tems do not consider automatically

A Universal Part-of-Speech Tagset

Slav Petrov
Google Research
New York, NY, USA
slav@google.com

Dipanjan Das
Carnegie Mellon University
Pittsburgh, PA, USA
dipanjan@cs.cmu.edu

Ryan McDonald
Google Research
New York, NY, USA
ryanmcd@google.com

Abstract

To facilitate future research in unsupervised induction of syntactic structure and to standardize best-practices, we propose a tagset that consists of twelve universal part-of-speech categories. In addition to the tagset, we develop a mapping from 25 different treebank tagsets to this universal set. As a result, when combined with the original treebank data, this universal tagset and mapping produce a dataset consisting of common parts-of-speech for 22 different languages. We highlight the use of this resource via two experiments, including one that reports competitive accuracies for unsupervised grammar induction without gold standard part-of-speech tags.

forms across languages. These categories are often called *universals* to represent their cross-lingual nature (Carnie, 2002; Newmeyer, 2005). For example, Naseem et al. (2009) used the Multext-East (Erjavec, 2004) corpus to evaluate their multi-lingual POS induction system, because it uses the same tagset for multiple languages. When corpora with common tagsets are unavailable, a standard approach is to manually define a mapping from language and treebank specific fine-grained tagsets to a predefined universal set. This was the approach taken by Das and Petrov (2011) to evaluate their cross-lingual POS projection system for six different languages.

To facilitate future research and to standardize best-practices, we propose a tagset that consists

Twitter Sentiment Classification using Distant Supervision

Alec Go
Stanford University
Stanford, CA 94305
alecmgo@stanford.edu

Richa Bhayani
Stanford University
Stanford, CA 94305
rbhayani@stanford.edu

Lei Huang
Stanford University
Stanford, CA 94305
leirocky@stanford.edu

ABSTRACT

We introduce a novel approach for automatically classifying the sentiment of Twitter messages. These messages are classified as either positive or negative with respect to a query term. This is useful for consumers who want to research the sentiment of products before purchase, or companies that want to monitor the public sentiment of their brands. There is no previous research on classifying sentiment of messages on microblogging services like Twitter. We present the results of machine learning algorithms for classifying the sentiment of Twitter messages using distant supervision. Our training data consists of Twitter messages

Consumers can use sentiment analysis to research products or services before making a purchase. Marketers can use this to research public opinion of their company and products, or to analyze customer satisfaction. Organizations can also use this to gather critical feedback about problems in newly released products.

There has been a large amount of research in the area of sentiment classification. Traditionally most of it has focused on classifying larger pieces of text, like reviews [9]. Tweets (and microblogs in general) are different from reviews primarily because of their purpose: while reviews represent summa-

Model Foundation