Are Words Commensurate with Actions?
Quantifying Commitment to A Cause from Public Messaging

Zhao Wang, Aron Culotta—Text Analysis in the Public Interest Lab, Illinois Institute of Technology

Abstract

Motivation: Online social networks are increasingly used by public entities such as companies and politicians to speak directly to their constituencies. In addition to typical marketing and campaigning activities, these entities often post messages to foster cause-related associations such as eco-friendliness or public health, which are becoming important components of brand equity. However, due to the low effort and informal nature of such communication, it can be difficult for consumers and voters to determine an entity's commitment to a cause based on their public messaging.

Objectives: 1) quantify how much of commitment an entity express towards a cause. 2) measure the discrepancy between public messages and external ratings to identify "inauthentic" entities whose public messaging show higher commitment to a cause than their actions.

Motivation Example

An entity's commitment to a cause in their public messaging does not always align with their actions in practice.

Example of low and high commitment tweets:

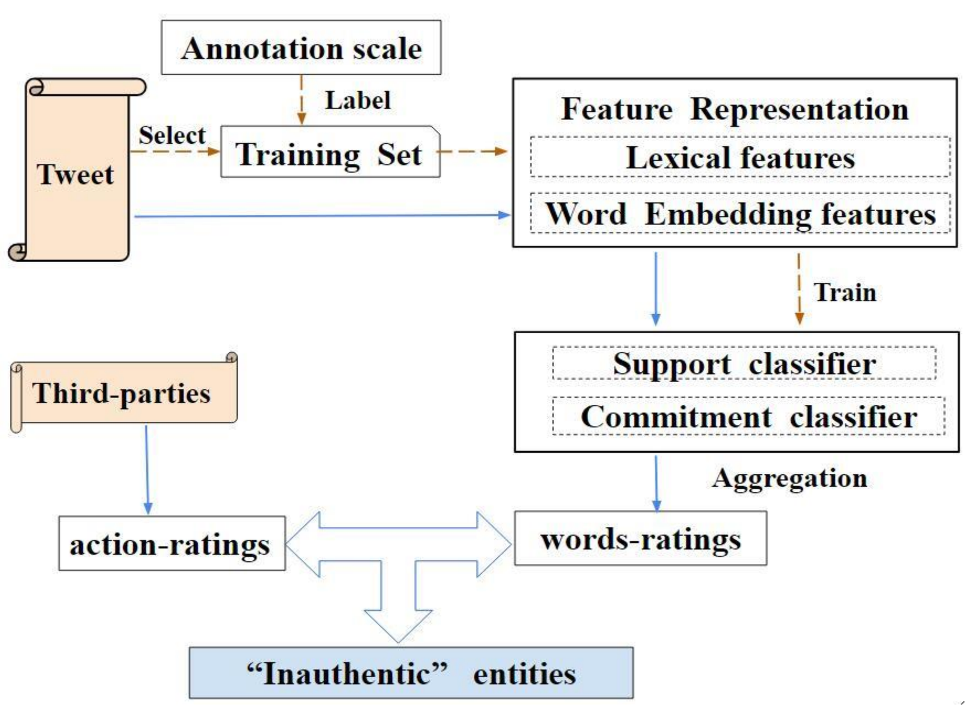
A low-commitment tweet like *"Happy #EarthDay – let's recycle today!"* allows the entity to signal support for a cause without making specific claims about their own actions.

Contrast this with tweets like "*Today I've introduced legislation to support our fisheries and habitat.*" or "*Our products are 100% sustainably sourced.*" which indicate a high-commitment to a cause by showing entities' efforts and contribution.

If an entity express high-commitment towards a cause in their public messaging but have low-commitment in their action, we call it "inauthentic".

Method

1. Build 2 binary **classifiers** to categorize messages according to commitment levels: support classifier and commitment classifier.
2. Classify all entities' historical messages and apply **aggregation** methods to quantify messages assigned to each category of commitment.
3. Measure the **discrepancy** between classified messages and third-party ratings to identify "**inauthentic**" entities who express high commitment in public messaging but have low commitment in action.



Data

We choose **Twitter** as the public messaging platform.

2 **issues**: environmental protection and personal health.

Entity	Issue	Public Message
Brands	Eco	2,624,800 tweets for 966 brands (10 sectors)
Brands	Health	429,009 tweets for 142 brands (food and personal care)
Congress	Eco	1,118,962 tweets for 514 congress members

Third-party ratings:

- GoodGuide: eco and health ratings for brands.
- League of Conservation Voters: eco ratings for politicians.

Annotation Scale

Label	Description	Eco	Health
0	Not about the issue.	Tourism is FL economy's lifeblood, providing 1.2 mil jobs.	Just saying hi, regards to hubby on this very special day!!
1	About the issue, but does not indicate support.	Check out the stunning landscape for our photoshoot: crisp river waters, mountains, fall foliage #NatureIsGreat	Our nutritional information is listed on each package.
2	Indicates support of the issue in words but not actions. (**low commitment**)	#CleanWaterAct protects drinking water, critical habitats, and waterways vital for the economy. #ProtectCleanWater	Here is a list of the top 10 foods to eat for healthy hair.
3	Indicates that the entity has taken actions to support the issue. (**high commitment**)	I've introduced legislation to help conserve our fisheries and habitat.	Bringing 23 new Certified Organic products to our fans in 2016.

Train two separate binary classifiers based on the labels:

- **Support classifier**: distinguish between {0,1} and {2,3}.
- **Commitment classifier**: distinguish between 2 and 3.

Feature Representation

Lexical features:

- Polarity: *negative indicator*
- Pronoun usage: *first person, second person and third person*
- Issue synonyms: *issue relevant terms*
- Context: *left and right word of issue term*
- Self-mention and retweet: *RT, @*

Word embedding features:

- Tweet vector*: average vector for all words in a tweet*.
- Commitment vector: *average vector of top-5 issue relevant terms in each tweet*
- Context vector: *average vector of left and right word of issue terms in each tweet*

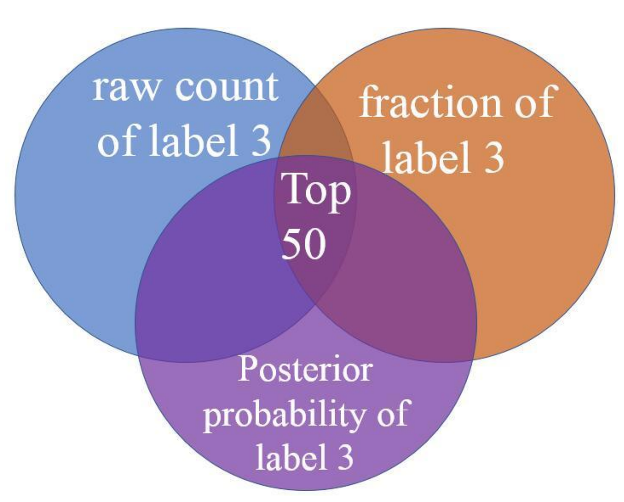
Others:

- *Number of issue synonyms in each tweet.*
- *Commitment score of each synonym, context vector.*
- *etc*

Aggregation

Classify all tweets into different levels of cause commitment.

Sort and select high words-rating entities:



Sort and select those below the mean in third-party ratings.

Cross-validation results for classification

			Support ({0,1} vs. {2,3})			Commitment (2 vs. 3)		
Entity	Issue	Feature	Precision	Recall	F1	Precision	Recall	F1
Brands	Health	Bag-of-Words	0.925	0.893	0.908	0.782	0.765	0.773
		Embeddings	0.933	0.874	0.903	0.712	0.702	0.705
		Combination	0.947	0.931	0.932	0.782	0.765	0.773
Brands	Eco	Bag-of-Words	0.894	0.556	0.678	0.776	0.703	0.732
		Embeddings	0.823	0.666	0.733	0.666	0.656	0.655
		Combination	0.843	0.790	0.812	0.810	0.716	0.755
Congress	Eco	Bag-of-Words	0.867	0.929	0.896	0.712	0.720	0.714
		Embeddings	0.915	0.837	0.873	0.555	0.648	0.594
		Combination	0.882	0.934	0.970	0.740	0.720	0.726

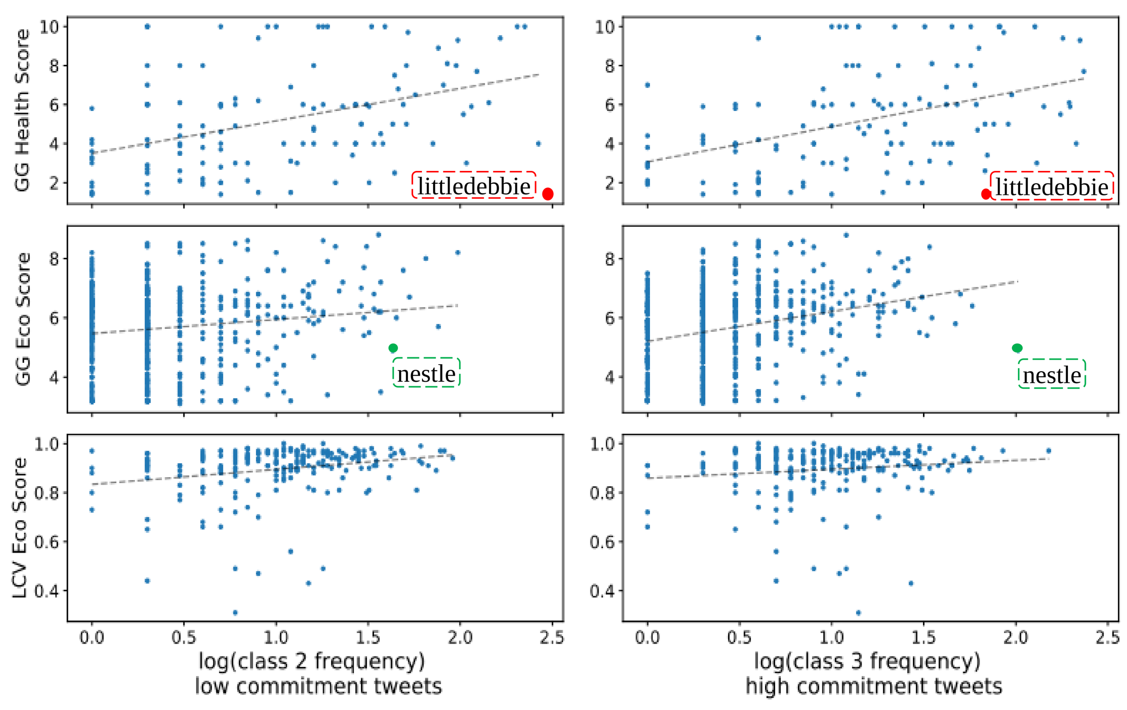
Top coefficient terms in classifiers

Entity	Issue	No Support Labels {0,1}	Support Labels {2,3}	Low Commit Label 2	High Commit Label 3
Brands	Health	flavor chocolate sleek animals cheese	healthy nutritious organic #vegan natrual	foods eat recipes veggies diet	our natural #organic #nongmo certified
Brands	Eco	skin food natural diet fish	sustainable sustainability environment planet endangered	planet day great second_person can	we first_person _self_first_person protect #sustainable
Congress	Eco	lives rural dimension_15 jobs #energy	protect conservation habitats epa dimension_18	rt you epa historic _self_pollution	I my voted bill must

Issue synonyms

stop-words
imperative words
pronouns, RT, @

Correlation with third-party ratings



	GoodGuide Health Score		GoodGuide Eco Score		LCV Eco Score	
	coef	p-value	coef	p-value	coef	p-value
Non-supportive class	-1.2781	0.035	-0.2434	0.112	-0.0923	0.004
class 2	1.0047	0.029	0.1144	0.541	0.1142	6e-5
class 3	1.3009	0.007	1.0750	9e-9	-0.0034	0.898

- ➢ More supporting tweets, higher action ratings.
- ➢ High commitment tweets, stronger signal of high ratings.

Detected "Inauthentic" entities

Entities that show high-commitment in public messaging but low action ratings(0~10):

Littledebbie: health rating 1.5 and example of its high-commitment tweet: *RT @quintanarootri: Do you have a @LittleDebbie* ***nutrition*** *plan for #IMChattanooga?* ***Simple carbs*** *for quick* ***energy*** *on the* ***bike*** *#itspersonal*

Nestle: eco rating 4.9 and example of its high-commitment tweet: *The* ***Eco-Shape*** *bottle with 15%* ***less plastic****, making it easier to* ***recycle****.*