

karami_2018_characterizing_diabetes_diet_exercise_and_obesity_comments_on_twitter

Year

2018

Author(s)

Amir Karami and Alicia A. Dahl and Gabrielle Turner-McGrievy and Hadi Kharrazi and George Shaw

Title

Characterizing diabetes, diet, exercise, and obesity comments on Twitter

Venue

International Journal of Information Management

Topic labeling

Partially automated

Focus

Secondary

Type of contribution

Established approach

Underlying technique

Manual category assignment; manual labeling

Topic labeling parameters

\

Label generation

Additionally, we labeled topics based on the availability of DDEO words. For example, if a topic had “diet”, we labeled it as a diet-related topic.

(Frequency identifies the nr of associated topics)

Table 2
DDEO **topics** and subtopics – diabetes, diet, exercise, and obesity are shown with italic and underline styles in subtopics.

Topics	Frequency	Subtopics	Distributions (%)	Topics	Frequency	Subtopics	Distributions (%)
Diabetes	21	Diabetes type 2	42.87%	Diet	63	<i>Obesity</i>	39.69%
		<i>Obesity</i>	14.29%			<i>Exercise</i>	15.87%
		<i>Diet</i>	9.52%			Weight loss	12.71%
		<i>Exercise</i>	9.52%			Celebrities	9.52%
		Blood pressure	9.52%			Vegetarian	9.52%
		Heart attack	4.76%			<i>Diabetes</i>	3.17%
		Yoga	4.76%			Religious diet	3.17%
		Alzheimer	4.76%			Weight loss medicine	3.17%
Exercise	80			Obesity	58	Pregnancy	1.59%
						Mental health	1.59%
		Fitness	32.5%			<i>Diet</i>	43.11%
		<i>Obesity</i>	22.5%			<i>Exercise</i>	31.04%
		Daily plan	21.25%			Children	17.24%
		<i>Diet</i>	11.25%			<i>Diabetes</i>	5.17%
		Brain	8.75%			Alzheimer	1.72%
		<i>Diabetes</i>	2.50%			Cancer	1.72%
		Computer games	1.25%				

Table 3
Topics examples.

Blood pressure	Heart attack	Diabetes type 2	Yoga	Alzheimer	Obesity	Diet and exercise	Obesity
risk blood high diabetes pressure	heart diabetes cardiovascular attack stroke	change diabetes #lifestyle type ii	diabetes #yogafightsdiabetes yoga control life	medicine diseases common drugs Alzheimer	diabetes surgery treatment obesity cure	helps diabetes children exercise diet	health diet obesity immune syndrome
Vegetarian	Pregnancy diet	Celebrities diet	Weight loss diet	Weight loss medicine	Religious diet	Mental health	Exercise and diabetes
diet eat fruits vegetables fresh	pregnancy motherhood diet baby motherhood	diet beyonce tips fatloss #angelinajolie	weightlose effective morning dieting banana	diet #weightloss slimming pills #fatburners	burning #weightloss fasting Ramadan diets	health nutrition benefits healing #mentalhealth	helps diabetes children exercise diet
Diet	Daily plan	Computer games	Brain	Fitness	Diet and diabetes	Obesity	Exercise
diet exercise protein beauty muscle	food exercise calorie goal completed	exercise finding pokemon #pokemongo hour	exercise brain improve memory performance	fitness #gymlife bodybuilding gym workout	helps diabetes children exercise diet	workout burning exercise fatburn obesity	bellyfat losing exercise ways effective
Diet		Alzheimer		Cancer		Children	Diabetes
health diet obesity immune syndrome		study link Alzheimer obesity research		cancer breast study risk obesity		obesity kids childhood rates problem	diabetes surgery treatment obesity cure

Motivation

\

Topic modeling

LDA

Topic modeling parameters

Nr of topics: 425

Nr. of topics

222 (LIWC is used to filter the detected 425 topics. 222 health-related topics are found)

Label

One of four DDEO categories (diet, diabetes, exercise, obesity) + a subtopic label

Label selection

\

Label quality evaluation

\

Assessors

\

Domain

Paper: Health

Dataset: Social media (Twitter)

Problem statement

Social media provide a platform for users to express their opinions and share information. Understanding public health opinions on social media, such as Twitter, offers a unique approach to characterizing common health issues such as diabetes, diet, exercise, and obesity (DDEO); however, collecting and analyzing a large scale conversational public health data set is a challenging research task.

The goal of this research is to analyze the characteristics of the general public's opinions in regard to diabetes, diet, exercise and obesity (DDEO) as expressed on Twitter. A multi-component semantic and linguistic framework was developed to collect Twitter data, discover topics of interest about DDEO, and analyze the topics.

Corpus

Origin: Twitter

Nr. of documents: 4.5M

Details:

- tweets between 06/01/2016 and 06/30/2016

Table 1
DDEO queries.

Health issue	Queries	Number of tweets	Percentage
Diabetes	diabetes OR #diabetes	353,655	8.0%
Diet	diet OR #diet OR dieting	1,045,374	23.7%
Exercise	exercise OR #exercise OR exercising	734,118	16.6%
Obesity	obesity OR #obesity OR fat	2,283,517	51.7%

Document

Pre-processing

Removal of stop words

```
@article{karami_2018_characterizing_diabetes_diet_exercise_and_obesity_comments_on_twitter,
```

```
  abstract = {Social media provide a platform for users to express their opinions and share information. Understanding public health opinions on social media, such as Twitter, offers a unique approach to characterizing common health issues such as diabetes, diet, exercise, and obesity (DDEO); however, collecting and analyzing a large scale conversational public health data set is a challenging research task. The goal of this research is to analyze the
```

characteristics of the general public's opinions in regard to diabetes, diet, exercise and obesity (DDEO) as expressed on Twitter. A multi-component semantic and linguistic framework was developed to collect Twitter data, discover topics of interest about DDEO, and analyze the topics. From the extracted 4.5 million tweets, 8% of tweets discussed diabetes, 23.7% diet, 16.6% exercise, and 51.7% obesity. The strongest correlation among the topics was determined between exercise and obesity ($p < .0002$). Other notable correlations were: diabetes and obesity ($p < .0005$), and diet and obesity ($p < .001$). DDEO terms were also identified as subtopics of each of the DDEO topics. The frequent subtopics discussed along with ``Diabetes'', excluding the DDEO terms themselves, were blood pressure, heart attack, yoga, and Alzheimer. The non-DDEO subtopics for ``Diet'' included vegetarian, pregnancy, celebrities, weight loss, religious, and mental health, while subtopics for ``Exercise'' included computer games, brain, fitness, and daily plan. Non-DDEO subtopics for ``Obesity'' included Alzheimer, cancer, and children. With 2.67 billion social media users in 2016, publicly available data such as Twitter posts can be utilized to support clinical providers, public health experts, and social scientists in better understanding common public opinions in regard to diabetes, diet, exercise, and obesity.},

```
author = {Amir Karami and Alicia A. Dahl and Gabrielle Turner-McGrievy and
Hadi Kharrazi and George Shaw},
date-added = {2023-04-02 12:26:37 +0200},
date-modified = {2023-04-02 12:26:37 +0200},
doi = {https://doi.org/10.1016/j.ijinfomgt.2017.08.002},
issn = {0268-4012},
journal = {International Journal of Information Management},
keywords = {Health, Diabetes, Diet, Obesity, Exercise, Topic model, Text
mining, Twitter},
number = {1},
pages = {1-6},
title = {Characterizing diabetes, diet, exercise, and obesity comments on
Twitter},
url = {https://www.sciencedirect.com/science/article/pii/S0268401217306126},
volume = {38},
year = {2018}}
```