alp_2019_influence_factorization_for_identifying_authorities_in_twitter

Year

2019

Author(s)

Zeynep Zengin Alp, Şule Gündüz Öğüdücü

Title

Influence Factorization for identifying authorities in Twitter

Venue

Knowledge-Based Systems

Topic labeling

Manual

Focus

Secondary

Type of contribution

Established approach

Underlying technique

Manual labeling

Topic labeling parameters

\

Label generation

"Finally, experts labeled the coherent topics with appropriate topic titles."

Three human experts identified six coherent topics and labeled them with a topical keyword.
After this process, there were six topics formed and named as: ''Politics / Breaking News'', ''Religion'', ''Spiritual'', ''Social Responsibility'', ''Soccer / Sports'', and ''TV /TV Shows''. ''Politics'' and ''Breaking News'' topical words were highly overlapping, thus a single topic is formed for both topics.

Motivation

After assigning topics to all of the tweets, topical networks were formed. For each user–topic pair percentage of topical tweets were calculated. Afterwards, topical labels were assigned to users according to topical contents of their tweets.

Topic modeling

LDA

Topic modeling parameters

Nr of topics (k): 6

Nr. of topics

6

Label

Manually assigned single or multi-word labels identifying the broad categories of the 6 topics

Label selection

\

Label quality evaluation

\

Assessors

\

Domain

Paper: Social media (Twitter)
Dataset: Social media (Twitter)

Problem statement

One of the most intriguing aspects of social networks is to identify influencers who are experts on a specific topic.

In this paper, we aim to identify topic-based experts using a large dataset collected from Twitter. Our proposed approach has three phases:

(1) identification of topics on social media posts (more specifically, tweets),

(2) user modeling, based on a group of user specific features

(3) Influence Factorization to identify topical influencers.

The main advantage of the proposed method is to identify future influencers as well as current ones on Twitter.

Corpus

Origin: Twitter
Nr. of documents: 38M tweets
Details:
- Posts from around 180K public user ids over a 70 days period

Table 2
Statistics about data sets.

	#users	#tweets	#links/follows
Training set	186K	31M	16M
Test set	181K	7M	15.6M

Document

A (pooled) set of Tweets belonging to the same user and published in the same day day

Pre-processing

- Stemming to all of the tweets as the first preprocessing step of topic modeling.
- Afterwards, stop-words, punctuations, and mentions that have no effect for topic modeling were also eliminated.
- Pooling on Tweets belonging to the same user and day is applied

---

```
@article{alp_2019_influence_factorization_for_identifying_authorities_in_twitte
r,
  abstract = {Prevalent usage of social media attracted companies and
researchers to analyze its dynamics and effects on user behavior. One of the
most intriguing aspects of social networks is to identify influencers who are
experts on a specific topic. With the identification of these users within the
network, many applications can be built for user recommendation, information
diffusion modeling, viral marketing, user modeling and many more. In this
paper, we aim to identify topic-based experts using a large dataset collected
from Twitter. Our proposed approach has three phases: (1) identification of
topics on social media posts (more specifically, tweets), (2) user modeling,
based on a group of user specific features, and (3) Influence Factorization to
identify topical influencers. The main advantage of the proposed method is to
identify future influencers as well as current ones on Twitter. Moreover, it is
an easy to implement algorithm using Spark MLlib, which can be easily extended
to include other user specific features, and compare with other methodologies.
The effectiveness of the proposed method is tested on a large dataset that
contains tweets of 180K user over 70 day period. The experimental results show
that our proposed method identifies influencers successfully when used with a
hybrid user specific feature that contains follower count and authenticity
information, and is a highly scalable and extensible algorithm.},
  author = {Zeynep Zengin Alp and {\c S}ule G{\"u}nd{\"u}z {\"O}{\u g}{\"u}
d{\"u}c{\"u}},
  date-added = {2023-03-16 21:02:46 +0100},
  date-modified = {2023-03-16 21:02:46 +0100},
  doi = {https://doi.org/10.1016/j.knosys.2018.10.020},
  issn = {0950-7051},
```

```
    journal = {Knowledge-Based Systems},

    keywords = {Data mining, Influence analysis, Social media analysis, Matrix
Factorization, Collaborative filtering, Influence prediction, Influence
maximization},

    pages = {944-954},

    title = {Influence Factorization for identifying authorities in Twitter},

    url = {https://www.sciencedirect.com/science/article/pii/S0950705118305069},

    volume = {163},

    year = {2019}}
```

#Thesis/Papers/Initial