chen_2020_a_structural_topic_modeling_based_bibliometric_study_of_sentiment_analysis_literature

Year

2020

Author(s)

Chen, Xieling and Xie, Haoran

Title

A Structural Topic Modeling-Based Bibliometric Study of Sentiment Analysis Literature

Venue

Cognitive Computation

Topic labeling

Manual

Focus

Secondary

Type of contribution

Established approach

Underlying technique

Manual labeling assisted by associated documents

Topic labeling parameters

Label generation

Furthermore, with a topic-term distribution matrix estimated by STM, we identified the representative terms for each topic. Topical labels were then assigned to each topic on the basis of a review of representative terms and articles by do- main experts with prior knowledge of sentiment analysis.

Table 7 STM analysis results with 16 identified topics

Discriminating terms	%	Suggested labels	Trend	*p*
Phrase, urdu, semantic, syntactic, orientation, dependency, rule, collocation, crf, sentence, parsing, relation, target, extraction, adjective	9.38	*Sentiment lexicons and knowledge bases*	↓↓	0.0235
Aspect-based, myanmar, aspect, product, review, absa, e-commerce, merchant, implicit, ate, ranking, aspect-opinion, aspect-level, explicit, feature-opinion	9.09	*Aspect-based sentiment analysis*	↑	0.9671
Fan, retweet, hashtag, tweet, twitter, soccer, leader, stream, trending, facebook, networking, bitcoin, football, sn, event	8.79	*Social network analysis*	↑↑↑↑	0.0001
Domain-specific, cross-lingual, malay, self-training, lexicon, multilingual, cross-domain, meta-level, disambiguation, supervision, emoticon, immune, semi-supervised, adaptation, domain, co-training	8.43	*Multiple domain and cross-domain adaption*	↑	0.7108
Naive, bayes, ensemble, k-nearest, swarm, particle, selection, weighting, multi-class, svm, preprocessing, indonesian, maximum, stopword, entropy, knn, tfidf, k-means	7.40	*Conventional machine learning and optimization methods*	↑	0.9016
Valence-arousal, va, gmm, circumplex, multi-label, time-frequency, music, expressivity, human-machine, temperature, eeg, affective, arousal, emotion, signal, emotinet	6.84	*Bio-signals and emotion models*	↓↓	0.0290
Deep, convolutional, cnn, lstm, recurrent, rnn, bidirectional, convolution, autoencoder, pre-trained, dbn, bilstm, gru	6.83	*Deep learning for natural language processing*	↑↑↑	0.0020
Ewom, mapreduce, big, tourist, cloud, saas, hadoop, spark, airline, disaster, sale, nuclear, intelligence, satisfaction	6.50	*Web services*	↑↑↑↑	0.0006
Dirichlet, lda, weibo, sentiment-topic, multi-feature, chinese, topic-sentiment, jst, multi-grain, latent, hot, topic, allocation, joint, sentimental	6.05	*Topic model*	↓	0.5923
Negation, sarcasm, spam, irony, fake, email, figurative, sarcastic, spammer, deceptive, ironic, detection, satire, satirical	5.46	*Spam and sarcasm detection*	↓	0.3031
Stock, financial, investor, trading, volatility, portfolio, trader, bankruptcy, price, news, guba, sp., forecasting, return	5.33	*Financial market*	↑	0.0638
Recommendation, recommender, app, star, helpfulness, cf., fuzzy, rating, mobile, filtering, collaborative, item, travel, recommend, explainable, uninorm	4.79	*Recommender systems and personalization*	↑↑	0.0151
Blogger, subtopic, ontology, image, flickr, extractive, retrieval, query, visualization, selfie, underground, visual, multimedia, retweeting, video	4.26	*Multimedia and multi-modality*	↑	0.8368
Stance, voter, echo, arguing, contentious, trump, political, donald, referendum, election, republican, presidential, nostalgia, clinton, parliamentary, electoral	3.93	*Political and media issues*	↑	0.1494
Health, student, drug, teaching, portuguese, cancer, surveillance, teacher, tobacco, forum, education, spanish, writing, educational, university	3.70	*Education and social issues*	↑↑	0.0120
Cognition, deficit, empathy, impairment, schizophrenia, cortex, oxytocin, parkinson, prefrontal, human-agent, epilepsy, amygdala, asd, injury, mdma	3.23	*Emotion-related disease*	↓	0.3031

Note: Topics are ranked by the proportion in descending order; %: topic proportions. Abbreviations of representative terms are shown in Table S1 in the Appendix. ↑(↓): topic showing an increase (decrease) in proportion annually but not significant ($p > 0.05$); ↑↑(↓↓), ↑↑↑(↓↓↓), ↑↑↑↑(↓↓↓↓): topic showing a significant increase (decrease) in proportion annually ($p < 0.05$, $p < 0.01$, and $p < 0.001$, respectively)

Motivation

\

Topic modeling

STM

Topic modeling parameters

We assigned weights (i.e., 0.4, 0.4, and 0.2) to the terms extracted from keywords, titles, and abstracts, respectively
Nr of topics: 15 to 42

Nr. of topics

16

Label

Manually assigned single or multi word labels

Label selection

\

Label quality evaluation

\

Assessors

\

Domain

Paper: Sentiment analysis research (Bibliometric)

Dataset: Sentiment analysis research

Problem statement

Sentiment analysis is an increasingly evolving field of research in computer science. With the considerable number of studies on innovative sentiment analysis available, it is worth the effort to present a review to understand the research on sentiment analysis comprehensively. This study aimed to investigate issues involved in sentiment analysis; for instance, (1) What types of research topics had been covered in sentiment analysis research? (2) How did the research topics evolve with time? (3) What were the topic distributions for major contributors? (4) How did major contributors collaborate in sentiment analysis research?

Corpus

Origin: Web of Science

Nr. of documents: 4373

Details:

- time span from 1999 to 2018
- “sentiment analysis,” “opinion mining,” “sentiment classification,” “opinion analysis,” “semantic orientation,” “opinion classification,” or “sentiment mining,” in titles, abstracts

Document

Abstract, title and keywords of an article

Pre-processing

- Duplicate removal
- Filtered unimportant terms using TF-IDF

```
@article{chen_2020_a_structural_topic_modeling_based_bibliometric_study_of_senti
ment_analysis_literature,
    abstract = {Sentiment analysis is an increasingly evolving field of research
in computer science. With the considerable number of studies on innovative
sentiment analysis available, it is worth the effort to present a review to
understand the research on sentiment analysis comprehensively. This study aimed
to investigate issues involved in sentiment analysis; for instance, (1) What
```

```
types of research topics had been covered in sentiment analysis research? (2)
How did the research topics evolve with time? (3) What were the topic
distributions for major contributors? (4) How did major contributors collaborate
in sentiment analysis research? Based on articles retrieved from the Web of
Science, this study presented a bibliometric review of sentiment analysis with
the basis of a structural topic modeling method to obtain an extensive overview
of the research field. We also utilized methods such as regression analysis,
geographic visualization, social network analysis, and the Mann--Kendal trend
test. Sentiment analysis research had, overall, received a growing interest in
academia. In addition, institutions and authors within the same countries/
regions were liable to collaborate closely. Highly discussed topics were
sentiment lexicons and knowledge bases, aspect-based sentiment analysis, and
social network analysis. Several current and potential future directions, such
as deep learning for natural language processing, web services, recommender
systems and personalization, and education and social issues, were revealed. The
findings provided a thorough understanding of the trends and topics regarding
sentiment analysis, which could help in efficiently monitoring future research
works and projects. Through this study, we proposed a framework for conducting a
comprehensive bibliometric analysis.},
    author = {Chen, Xieling and Xie, Haoran},
    date-added = {2023-04-12 22:31:22 +0200},
    date-modified = {2023-04-12 22:31:22 +0200},
    day = {01},
    doi = {10.1007/s12559-020-09745-1},
    issn = {1866-9964},
    journal = {Cognitive Computation},
    month = {Nov},
    number = {6},
    pages = {1097--1129},
    title = {A Structural Topic Modeling-Based Bibliometric Study of Sentiment
Analysis Literature},
    url = {https://link.springer.com/content/pdf/10.1007/
s12559-020-09745-1.pdf},
    volume = {12},
    year = {2020},
    bdsk-url-1 = {https://link.springer.com/content/pdf/10.1007/
s12559-020-09745-1.pdf},
    bdsk-url-2 = {https://doi.org/10.1007/s12559-020-09745-1}}
```

#Thesis/Papers/FS