zhao_2021_adversarial_learning_of_poisson_factorisation_model_for_gauging_brand_sentiment_in_user_reviews

Year

2021

Author(s)

Zhao, Runcong and Gui, Lin and Pergola, Gabriele and He, Yulan

Title

Adversarial Learning of Poisson Factorisation Model for Gauging Brand Sentiment in User Reviews

Venue

EACL

Topic labeling

Manual

Focus

Secondary

Type of contribution

Established approach

Underlying technique

Manual labeling

Topic labeling parameters

\

Label generation

\

Motivation

\

Topic modeling

Brand-Topic Model built on the Poisson Factorisation model with adversarial learning

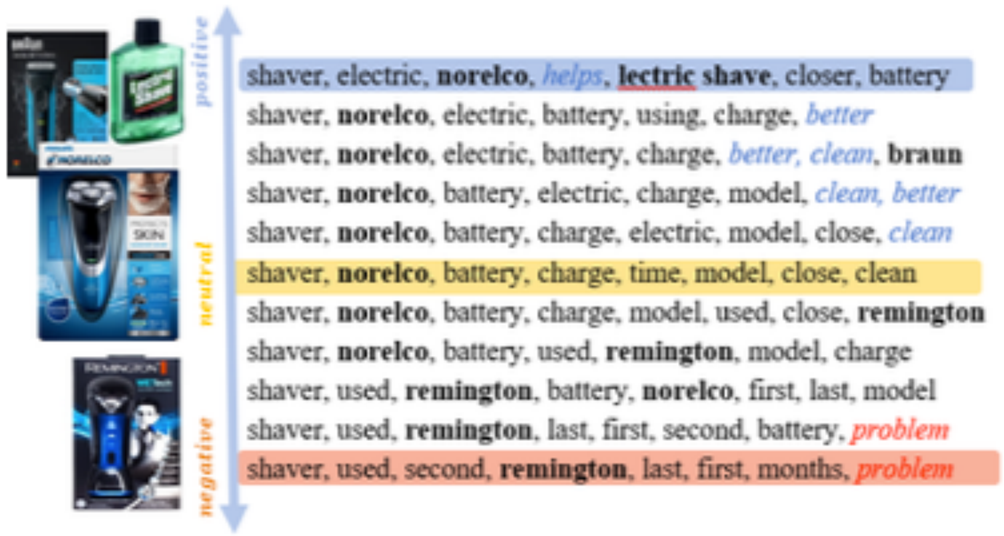

Figure 1: Example topic results generated from proposed Brand-Topic Model. We observe a transition of topics with varying topic polarity scores. Besides the change of sentiment-related words (e.g., '*problem*' in negative topics and '*better*' in positive topics), we could also see a change of their associated brands. Users are more positive about BRAUN, negative about REMINGTON, and have mixed opinions on NORELCO.

Topic modeling parameters

- Train/Test split: 10% reviews (7,826 reviews) as the test set and the remaining (70,436 reviews) as the training set
- Batch size: 1,024

- Maximum training steps: 50,000
- Topic number (K): 30
- Temperature in the Gumbel-Softmax equation: 1

Nr. of topics

30

Label

Single or multi-word manually generated labels

Label selection

\

Label quality evaluation

\

Assessors

\

Domain

Domain (paper): Polarity-bearing topic modeling
Domain (corpus): Online store reviews

Problem statement

Proposing the Brand-Topic Model (BTM), which aims to detect brand- associated polarity-bearing topics from product reviews.
BTM is able to automatically infer real-valued brand-associated sentiment scores and generate fine-grained sentiment-topics in which we can observe continuous changes of words under a certain topic while its associated sentiment gradually varies from negative to positive.

Corpus

Origin: Amazon

Nr. of documents: 78,322 (113,730 in the oversampled dataset)

Details:

- Reviews in the Beauty category from the Amazon review corpus
- Final dataset contains a total of 78,322 reviews from 45 brands

Dataset	**Amazon-Beauty Reviews**
Documents per classes	
Neg / Neu / Pos	9,545 / 5,578 / 63,199
Brands	45
Total #Documents	78,322
Avg. Document Length	9.7
Vocabulary size	$\sim$ 5000

Table 1: Dataset statistics of reviews within the Amazon dataset under the *Beauty* category.

Document

A single amazon review belonging to the Beauty category.

Each review is accompanied with the rating score (between 1 and 5), reviewer name and the product meta-data such as product ID, description, brand and image.

Pre-processing

We use the product meta-data to relate a product with its associated brand.

Reviews with the rating score of 1 and 2 are grouped as negative reviews; those with the score of 3 are neutral reviews; and the remaining are positive reviews.

Documents are represented as the bag-of-words

Tokens, i.e., n-grams (n = {1, 2, 3}), occurred less than twice are filtered.

Oversampled dataset

Since the dataset is highly imbalanced, we balance data in each mini-batch by oversampling.

This results in an increased training set consisting of 113,730 reviews.

```
@inproceedings{zhao_2021_adversarial_learning_of_poisson_factorisation_model_fo
r_gauging_brand_sentiment_in_user_reviews,
    title = "Adversarial Learning of {P}oisson Factorisation Model for Gauging
Brand Sentiment in User Reviews",
    author = "Zhao, Runcong  and
      Gui, Lin  and
      Pergola, Gabriele  and
      He, Yulan",
    booktitle = "Proceedings of the 16th Conference of the European Chapter of
the Association for Computational Linguistics: Main Volume",
    month = apr,
    year = "2021",
    address = "Online",
    publisher = "Association for Computational Linguistics",
    url = "https://aclanthology.org/2021.eacl-main.199",
    doi = "10.18653/v1/2021.eacl-main.199",
    pages = "2341--2351",
    abstract = "In this paper, we propose the Brand-Topic Model (BTM) which
aims to detect brand-associated polarity-bearing topics from product reviews.
Different from existing models for sentiment-topic extraction which assume
topics are grouped under discrete sentiment categories such as {`}positive{'},
{`}negative{'} and {`}neural{'}, BTM is able to automatically infer real-valued
brand-associated sentiment scores and generate fine-grained sentiment-topics in
which we can observe continuous changes of words under a certain topic (e.g.,
{`}shaver{'} or {`}cream{'}) while its associated sentiment gradually varies
from negative to positive. BTM is built on the Poisson factorisation model with
the incorporation of adversarial learning. It has been evaluated on a dataset
constructed from Amazon reviews. Experimental results show that BTM outperforms
a number of competitive baselines in brand ranking, achieving a better balance
of topic coherence and unique-ness, and extracting better-separated polarity-
bearing topics.",
}
```

#Thesis/Papers/Initial