mukherjee_2020_read_what_you_need_controllable_aspect_based_opinion_summarization_of_tourist_reviews

Year

2020

Author(s)

Mukherjee, Rajdeep and Peruri, Hari Chandana and Vishnu, Uppada and Goyal, Pawan and Bhattacharya, Sourangshu and Ganguly, Niloy

Title

Read what you need: Controllable Aspect-based Opinion Summarization of Tourist Reviews

Venue

SIGIR'20

Topic labeling

Manual

Focus

Secondary

Type of contribution

Established approach

Underlying technique

Two-level manual labeling process (fine-grained labels for topics, coarse-grained labels for topic classes)

Topic labeling parameters

\

Label generation

- "each extracted topic is manually interpreted by looking at its representative words and assigned a genuine aspect label"
Individual topics are assigned an initial fine-grained label, and then are grouped together into classes which are assigned a coarse-grained category label.

Aspect Classes	Inferred Topics
Attractions	Architecture, Monuments, Surroundings, Events
Access	Travel infrastructure, Long-distance modes of transport, Entrance, Time of visiting
Activities	Photography, Events, Shopping
Amenities	Tour Guides, Hospitality, Vendors, Services, Information, Food, Accommodation
Culture	Hospitality, Events, Climate, History, Dress
Cost	Cost of visiting, Tickets
Negatives	Bad Experiences, Vendors
Miscellaneous	Nationality, Adjectives

Motivation

Simplify the selection of aspects (topics) for which a summary needs to be generated:
- "In our proposed solution for controllable summarisation, readers have the flexibility of selecting the aspects they are interested in"

Topic modeling

Attention-based Aspect Extraction (ABAE) (He et al., 2017)

Topic modeling parameters

Nr of aspects/topics to be extracted from the corpus (K): 25 (best scoring, experimented with values between 10-50)

Other parameters as appearing in He et al., 2017:
- Embedding size: 200
- Window size: 10
- Negative sample size: 5
- Learning rate: 0.001

- Nr of epochs: 15
- Nr of negative samples per input sample (m): 20
- Orthogonality penalty weight (λ): 1

Nr. of topics

25

Label

Single or multi-world (humanly generated) label for topics and topic classes

Label selection

\

Label quality evaluation

No formal evaluation is given:

- "Since aspect extraction is not a major contribution of this work, we do not evaluate our obtained results against a gold-standard annotated corpus."

Still, a shallow evaluation of coarse-grained topic classes is made against Marrese-Taylor et. al, 2014:

- "We […] categorize the obtained topics into coarse-grained aspect classes and find them to be well-aligned with the tourism literature "

Assessors

\

Domain

Tourism and Hospitality

Problem statement

Generation of personalised aspect-based (topic-based) opinion summaries from large

collections of online tourist reviews.

Proposing an unsupervised approach to extract coherent aspects (topics) from tourist reviews posted on TripAdvisor.

Proposing an opinion summarisation approach that after finding relevant aspects about the entity being reviewed and identifying sentences which discuss them, generates a concise and digestible summary of opinions.

Corpus

Origin: TripAdvisor
Nr. of documents: 7000
Details: Top 1000 most liked reviews for each of the New Seven Wonders of the World

Document

An english TripAdvisor review
Fields:
- id
- text
- user rating (1-5)
- likes (no. of likes received)
- username and reviewer's location

Pre-processing

\

```
@inproceedings{mukherjee_2020_read_what_you_need_controllable_aspect_based_opin
ion_summarization_of_tourist_reviews,
    author = {Mukherjee, Rajdeep and Peruri, Hari Chandana and Vishnu, Uppada
and Goyal, Pawan and Bhattacharya, Sourangshu and Ganguly, Niloy},
    title = {Read What You Need: Controllable Aspect-Based Opinion
Summarization of Tourist Reviews},
    year = {2020},
    isbn = {9781450380164},
    publisher = {Association for Computing Machinery},
    address = {New York, NY, USA},
```

```
    url = {https://doi.org/10.1145/3397271.3401269},
    doi = {10.1145/3397271.3401269},
    abstract = {Manually extracting relevant aspects and opinions from large
volumes of user-generated text is a time-consuming process. Summaries, on the
other hand, help readers with limited time budgets to quickly consume the key
ideas from the data. State-of-the-art approaches for multi-document
summarization, however, do not consider user preferences while generating
summaries. In this work, we argue the need and propose a solution for
generating personalized aspect-based opinion summaries from large collections
of online tourist reviews. We let our readers decide and control several
attributes of the summary such as the length and specific aspects of interest
among others. Specifically, we take an unsupervised approach to extract
coherent aspects from tourist reviews posted onTripAdvisor. We then propose an
Integer Linear Programming (ILP) based extractive technique to select an
informative subset of opinions around the identified aspects while respecting
the user-specified values for various control parameters. Finally, we evaluate
and compare our summaries using crowdsourcing and ROUGE-based metrics and
obtain competitive results.},
    booktitle = {Proceedings of the 43rd International ACM SIGIR Conference on
Research and Development in Information Retrieval},
    pages = {1825-1828},
    numpages = {4},
    keywords = {tourism, unsupervised extractive opinion summarization, aspect-
based opinion mining, controllable summarization, personalization},
    location = {Virtual Event, China},
    series = {SIGIR '20}
}


@article{marrese_taylor_2014_a_novel_deterministic_approach_for_aspect_based_op
inion_mining_in_tourism_products_reviews,
    abstract = {This work proposes an extension of Bing Liu's aspect-based
opinion mining approach in order to apply it to the tourism domain. The
extension concerns with the fact that users refer differently to different
kinds of products when writing reviews on the Web. Since Liu's approach is
focused on physical product reviews, it could not be directly applied to the
tourism domain, which presents features that are not considered by the model.
Through a detailed study of on-line tourism product reviews, we found these
features and then model them in our extension, proposing the use of new and
more complex NLP-based rules for the tasks of subjective and sentiment
```

classification at the aspect-level. We also entail the task of opinion
visualization and summarization and propose new methods to help users digest
the vast availability of opinions in an easy manner. Our work also included the
development of a generic architecture for an aspect-based opinion mining tool,
which we then used to create a prototype and analyze opinions from TripAdvisor
in the context of the tourism industry in Los Lagos, a Chilean administrative
region also known as the Lake District. Results prove that our extension is
able to perform better than Liu's model in the tourism domain, improving both
Accuracy and Recall for the tasks of subjective and sentiment classification.
Particularly, the approach is very effective in determining the sentiment
orientation of opinions, achieving an F-measure of 92% for the task. However,
on average, the algorithms were only capable of extracting 35% of the explicit
aspect expressions, using a non-extended approach for this task. Finally,
results also showed the effectiveness of our design when applied to solving the
industry's specific issues in the Lake District, since almost 80% of the users
that used our tool considered that our tool adds valuable information to their
business.},
  author = {Edison Marrese-Taylor and Juan D. Vel{\'a}squez and Felipe Bravo-
Marquez},
  date-added = {2023-02-28 11:36:25 +0100},
  date-modified = {2023-02-28 11:36:25 +0100},
  doi = {https://doi.org/10.1016/j.eswa.2014.05.045},
  issn = {0957-4174},
  journal = {Expert Systems with Applications},
  keywords = {Aspect-based, Opinion mining, Tourism, Product reviews},
  number = {17},
  pages = {7764-7775},
  title = {A novel deterministic approach for aspect-based opinion mining in
tourism products reviews},
  url = {https://www.sciencedirect.com/science/article/pii/S0957417414003315},
  volume = {41},
  year = {2014},
  bdsk-url-1 = {https://www.sciencedirect.com/science/article/pii/
S0957417414003315},
  bdsk-url-2 = {https://doi.org/10.1016/j.eswa.2014.05.045}}

#Thesis/Papers/Initial