xiang_2017_a_comparative_analysis_of_major_online_review_platforms_implications_for_social_media_analytics_in_hospitality_and_tourism

Year

2017

Author(s)

Zheng Xiang and Qianzhou Du and Yufeng Ma and Weiguo Fan

Title

A comparative analysis of major online review platforms: Implications for social media analytics in hospitality and tourism

Venue

Tourism Management

Topic labeling

Manual

Focus

Secondary

Type of contribution
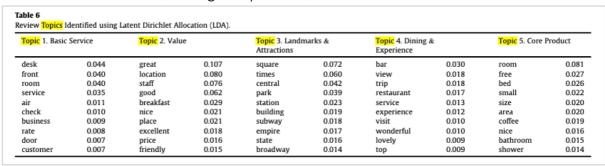
Established approach

Underlying technique

Manual labeling

Topic labeling parameters

\

Label generation

Based upon their meanings these topics were manually labelled as Basic Service, Value, Landmarks & Attractions, Dining & Experience, and Core Product.

Table 6
Review Topics Identified using Latent Dirichlet Allocation (LDA).

Topic 1. Basic Service		Topic 2. Value		Topic 3. Landmarks & Attractions		Topic 4. Dining & Experience		Topic 5. Core Product	
desk	0.044	great	0.107	square	0.072	bar	0.030	room	0.081
front	0.040	location	0.080	times	0.060	view	0.018	free	0.027
room	0.040	staff	0.076	central	0.042	trip	0.018	bed	0.026
service	0.035	good	0.062	park	0.039	restaurant	0.017	small	0.022
air	0.011	breakfast	0.029	station	0.023	service	0.013	size	0.020
check	0.010	nice	0.021	building	0.019	experience	0.012	area	0.020
business	0.009	place	0.021	subway	0.018	visit	0.010	coffee	0.019
rate	0.008	excellent	0.018	empire	0.017	wonderful	0.010	nice	0.016
door	0.007	price	0.016	state	0.016	lovely	0.009	bathroom	0.015
customer	0.007	friendly	0.015	broadway	0.014	top	0.009	shower	0.014

Motivation

To more succinctly: "understand potential differences between the three platforms, we then examined the manifestation of these five topics on each of these websites. "

Topic modeling

LDA

Topic modeling parameters

Nr of topics: 5

Nr. of topics

5

Label

Single or multi-word manually assigned label (one of Basic Service, Value, Landmarks & Attractions, Dining & Experience, and Core Product.)

Label selection

\

Label quality evaluation

\

Assessors

\

Domain

Paper: Tourism and Hospitality
Dataset: Hotels (Reviews)

Problem statement

Online consumer reviews have been studied for various research problems in hospitality and tourism. However, existing studies using review data tend to rely on a single data source and data quality is largely anecdotal. This greatly limits the generalizability and contribution of social media analytics research. Through text analytics this study comparatively examines three major online review platforms, namely TripAdvisor, Expedia, and Yelp, in terms of information quality related to online reviews about the entire hotel population in Manhattan, New York City.

Corpus

Origin: TripAdvisror, Expedia, and Yelp
Nr. of documents: 820778
Details:

- Data collection in late 2015 on all searchable hotel properties in Manhattan, NYC.
- 439 k reviews from TripAdvisor, 481 k from Expedia, and 31 k from Yelp for a total of approx. 500 hotel properties

Table 2
Summary of the main data set.

Review Platform	N of Hotels	N of Reviews	N of English Reviews (percentage)	N of English Reviews per Hotel
TripAdvisor	443	438,890	438,826 (99.99%)	991
Expedia	467	480,589	351,182 (73.07%)	752
Yelp	581	30,816	30,770 (99.85%)	53

Document

Hotel review + name of the hotel property, its address, hotel class, user responses

(usefulness or helpfulness), and the overall rating.

Pre-processing

- tokenization
- stop words removal

```
@article{xiang_2017_a_comparative_analysis_of_major_online_review_platforms_imp
lications_for_social_media_analytics_in_hospitality_and_tourism,
   abstract = {Online consumer reviews have been studied for various research
problems in hospitality and tourism. However, existing studies using review
data tend to rely on a single data source and data quality is largely
anecdotal. This greatly limits the generalizability and contribution of social
media analytics research. Through text analytics this study comparatively
examines three major online review platforms, namely TripAdvisor, Expedia, and
Yelp, in terms of information quality related to online reviews about the
entire hotel population in Manhattan, New York City. The findings show that
there are huge discrepancies in the representation of the hotel industry on
these platforms. Particularly, online reviews vary considerably in terms of
their linguistic characteristics, semantic features, sentiment, rating,
usefulness as well as the relationships between these features. This study
offers a basis for understanding the methodological challenges and identifies
several research directions for social media analytics in hospitality and
tourism.},
   author = {Zheng Xiang and Qianzhou Du and Yufeng Ma and Weiguo Fan},
   date-added = {2023-04-01 19:31:22 +0200},
   date-modified = {2023-04-01 19:31:22 +0200},
   doi = {https://doi.org/10.1016/j.tourman.2016.10.001},
   issn = {0261-5177},
   journal = {Tourism Management},
   keywords = {Online reviews, Hotel industry, Information quality, Social media
analytics, Text analytics, Machine learning},
   pages = {51-65},
   title = {A comparative analysis of major online review platforms:
Implications for social media analytics in hospitality and tourism},
   url = {https://www.sciencedirect.com/science/article/pii/S0261517716301807},
```

```
    volume = {58},
    year = {2017}}
```