korfiatis_2019_measuring_service_quality_from_unstructured_data_a_topic_modeling_application_on_airline_passengers_online_reviews

Year

2019

Author(s)

Nikolaos Korfiatis and Panagiotis Stamolampros and Panos Kourouthanassis and Vasileios Sagiadinos

Title

Measuring service quality from unstructured data: A topic modeling application on airline passengers' online reviews

Venue

Expert Systems with Applications

Topic labeling

Manual

Focus

Secondary

Type of contribution

Established approach

Underlying technique

Manual labeling assisted by associated documents

Topic labeling parameters

Nr of inspected reviews: 10
Nr of inspected topic terms: 7

Label generation

We assigned labels to topics by recruiting two experts with airline customer service experience to help us evaluate each topic using a sample of the top 10 loading reviews and the top 7 FREX words.
Both experts agreed that the selected topic solution had a high degree of coherence in terms of the top loading reviews and assigned mutually agreed labels

(FREX is estimated as a weighted harmonic mean of a word's rank in terms of exclusivity and semantic coherence. Semantic coherence uses the frequency of co-occurrence of the most probable words in each topic of the topic solution, while exclusivity considers the mutual appearance of the most probable words in more than one topics)

Table 2
Labels, distribution and FREX score for the top 7 keywords in the topic solution.

#	Topic label	Prop. (%)	Top 7 FREX words
1	Business Class	3.24	flat, lounge, business, class, bed, access, wine
2	Value for Money	6.84	good, food, value, overall, experience, money, airplane
3	Baggage Policy	3.02	bag, carry, charge, line, checked, fee, item
4	Low Cost	4.94	low, budget, price, cheap, cost, fare, carrier
5	Legroom (Critique)	2.02	room, space, enough, tall, foot, extra, amount
6	Delays	11.03	delay, hotel, hour, late, due, minute, connection
7	Staff (Praise)	10.11	friendly, helpful, clean, efficient, professional, courteous, smooth
8	Premium Economy	2.23	economy, premium, comfort, upgrade, difference, section, haul
9	Staff (Critique)	5.50	water, stewardess, toilet, poor, terrible, light, steward
10	Passenger Experience	6.01	best, many, domestic, world, past, travel, frequent
11	Frequent Flyer Status	1.91	flyer, member, traveller, group, point, aircraft
12	Mode of Travel	5.49	trip, return, direct, stop, home, round, non
13	Seating (Critique)	6.70	row, front, uncomfortable, seat, window, aisle, exit
14	Refund/Cancelation	5.42	phone, card, credit, email, call, agent, ticket
15	Food/In-flight entertainment	6.24	entertainment, movie, inflight, selection, screen, meal, average
16	Staff Assistance	5.22	child, holiday, nothing, much, special, cabin, crew
17	Legroom (Praise)	1.71	leg, lot, plenty, journey, second, extra, bit
18	Check-in	4.75	hand, luggage, check, queue, case, baggage, online
19	Airport Experience	4.98	free, snack, board, terminal, early, boarding, WIFI
20	Onboard Service	2.65	short, tea, full, usual, bit, etc, available

Motivation

\

Topic modeling

STM

Topic modeling parameters

Nr of topics (K): 20

Nr. of topics

20

Label

Single for multi-word labels manually assigned by two experts with airline customer service experience

Label selection

\

Label quality evaluation

\

Assessors

\

Domain

Paper: Service quality
Dataset: Airline reviews

Problem statement

In this study, we utilize online reviews to show the information gains from the consideration of factors identified from topic modeling of unstructured data which provide a flexible extension to numerical scores to understand customer satisfaction and subsequently service quality.
When numerical and textual features are combined, the explained variation in overall satisfaction improves significantly.

We further present how such information can be of value for firms for corporate strategy decision-making when incorporated in an expert system that acts as a tool to perform market analysis and assess their competitive performance.
We apply our methodology on airline passengers' online reviews using Structural Topic Models (STM).
This innovation allows us to capture dominant drivers of satisfaction along with their dynamics and interdependencies.

Corpus

Origin: TripAdvisor
Nr. of documents: 557,208 (184,502 after pre-processing)
Details:

- Airline passenger reviews

Document

Airline passenger review with information about flight date, flight distance, name of airline, route (start and destination airport), cabin class (first class, business class, premium economy, and economy class) and reviewers' level of contribution to the platform (computed from the number of review posts)

Passengers provide an overall score for their total experience (in an ordinal categorical scale from 1 to 5), which is accompanied by an individual rating for 8 specific service aspects of the flight namely: (a) Seat Comfort, (b) Customer Service, (c) Cleanliness, (d) Food and Beverage, (e) Legroom, (f) In-Flight Entertainment (g) Value for Money and (h) Check-in and Boarding.

Pre-processing

- Removal of non-English reviews
- Tokenization
- Elimination of numbers and punctuation marks
- Exclusion of language stop-words as well as context-specific stopwords such as names of airlines, airports, and routes, and words with a length under a specific threshold (set to three characters)
- Filtering the remaining words to keep only adverbs adjectives and nouns as these words have information about the product and product quality
- Stemming
- Lemmatization

```
@article{korfiatis_2019_measuring_service_quality_from_unstructured_data_a_topic_modeling_application_on_airline_passengers_online_reviews,
  abstract = {Service quality is a multi-dimensional construct which is not accurately measured by aspects deriving from numerical ratings and their associated weights. Extant literature in the expert and intelligent systems examines this issue by relying mainly on such constrained information sets. In this study, we utilize online reviews to show the information gains from the consideration of factors identified from topic modeling of unstructured data which provide a flexible extension to numerical scores to understand customer satisfaction and subsequently service quality. When numerical and textual features are combined, the explained variation in overall satisfaction improves significantly. We further present how such information can be of value for firms for corporate strategy decision-making when incorporated in an expert system that acts as a tool to perform market analysis and assess their competitive performance. We apply our methodology on airline passengers' online reviews using Structural Topic Models (STM), a recent probabilistic extension to Latent Dirichlet Allocation (LDA) that allows the incorporation of document level covariates. This innovation allows us to capture dominant drivers of satisfaction along with their dynamics and interdependencies. Results unveil the orthogonality of the low-cost aspect of airline competition when all other service quality dimensions are considered, thus explaining the success of low-cost carriers in the airline market.},
  author = {Nikolaos Korfiatis and Panagiotis Stamolampros and Panos Kourouthanassis and Vasileios Sagiadinos},
  date-added = {2023-03-21 17:45:50 +0100},
  date-modified = {2023-03-21 17:45:50 +0100},
  doi = {https://doi.org/10.1016/j.eswa.2018.09.037},
  issn = {0957-4174},
  journal = {Expert Systems with Applications},
  keywords = {Electronic WOM, Unstructured data, Service quality, Correspondence analysis, Structural topic model},
  pages = {472-486},
  title = {Measuring service quality from unstructured data: A topic modeling application on airline passengers' online reviews},
  url = {https://www.sciencedirect.com/science/article/pii/S0957417418306146},
```

```
  volume = {116},
  year = {2019}}
```

#Thesis/Papers/Initial