



Article Title

Social media analytics: Extracting and visualizing Hilton hotel ratings and reviews from TripAdvisor

Citation

By Chang, Yung-Chun; Ku, Chih-Hao; Chen, Chun-Hung. In International Journal of Information Management. October 2019

Abstract

Analyzing and extracting insights from user-generated data has become a topic of interest among businesses and research groups because such data contains valuable information, e.g., consumers' opinions, ratings, and recommendations of products and services. However, the true value of social media data is rarely discovered due to overloaded information. Existing literature in analyzing online hotel reviews mainly focuses on a single data resource, lexicon, and analysis method and rarely provides marketing insights and decision-making information to improve business' service and quality of products. We propose an integrated framework which includes a data crawler, data preprocessing, sentiment-sensitive tree construction, convolution tree kernel classification, aspect extraction and category detection, and visual analytics to gain insights into hotel ratings and reviews. The empirical findings show that our proposed approach outperforms baseline algorithms as well as well-known sentiment classification methods, and achieves high precision (0.95) and recall (0.96). The visual analytics results reveal that Business travelers tend to give lower ratings, while Couples tend to give higher ratings. In general, users tend to rate lowest in July and highest in December. The Business travelers more frequently use negative keywords, such as "rude," "terrible," "horrible," "broken," and "dirty," to express their dissatisfied emotions toward their hotel stays in July.

Summary

This study reveals the intricacies of users' ratings, sentiments, and types of travelers at different times and locations. It also indicates that a sole source of data, lexicon, and method is not sufficient to discover the genuine value of hotel ratings and reviews. Furthermore, existing research is mostly limited to document-level and sentence-level analysis.

The major contribution of this study includes a proposed, integrated framework to collect and process data, extract and classify aspect-level information, and finally visualize TripAdvisor and Google Trends data.