ABSTRACT

Background: Understanding patients' concerns and expectations regarding hospitals is crucially important for providers and administrators attempting to provide better service. In the era of big data, as a large and increasing amount of unsolicited feedback is provided on crowd-sourced review websites, the information required for such attempts becomes more accessible and affordable.

Methods: A web crawler was created to retrieve ratings and comments on hospital visits from Google Maps, based on a list of all hospitals in the US obtained from Data.gov. Due to API limitations, only the most recent 5 individual comments and ratings were retrieved. In total, 26,141 individual ratings and comments for 5,888 hospitals in the US were studied. Comments were classified into 3 groups: positive (rating ≥4), neutral (rating 3) and negative (rating ≤2). The top 20 most frequent 2-gram phrases in positive and negative comments were identified and analyzed. A non-supervised machine learning algorithm, K-mean text clustering analysis, was performed on the negative comments. Four clusters were created, and the 10 most representative words of each cluster were identified and analyzed.

Key Findings: Among the top 20 most frequent 2-gram phrases in negative comments, 3 were about waiting time and 3 were about service quality. Among the top 20 most frequent 2-gram phrases in positive comments, 6 referred to the care provided and 6 mentioned hospital staff. Topics of the 4 clusters among negative comments were identified: waiting time and appointments, insurance and billing issues, bad experiences of senior family members, and poor service quality.

Conclusions and Implications: To provide more satisfactory service, hospitals may wish to reduce waiting times, improve billing and payment processes, make facilities more accessible for older patients, and improve service quality. Moreover, hospitals may wish to make patients feel they are receiving excellent and friendly care from hospital staff. Patients' feedback on crowd-sourced review websites is a useful data source for understanding patients' concerns and expectations. Text mining techniques dramatically improve the efficiency of classifying patients' feedback and identifying their underlying meaning. Hospital administrators should value and utilize their patients' feedback on the internet.

INTRODUCTION

- Understanding patients' concerns and expectations regarding hospitals is crucially important for providers and administrators attempting to provide better service.
- As a large and increasing amount of unsolicited feedback is provided on crowd-sourced review websites, the information required for such attempts becomes more accessible and affordable to collect.
- The purpose of this study is to better understand patients' concerns and expectations about hospitals by analyzing feedback data collected on the internet.

METHODS

- A web crawler was created to retrieve ratings and comments on hospital visits from Google Maps, based on a list of all hospitals in the US obtained from Data.gov.
- In total, 26,141 individual ratings and comments for 5,888 hospitals in the US were studied.
- Comments were classified into 3 groups: positive, neutral, and negative.
- Top 20 most frequent 2-gram phrases in positive and negative comments were identified and analyzed.
- K-mean text clustering analysis was performed on negative comments.
- Four clusters were created, and the 10 most representative words of each cluster were identified and analyzed.

RESULTS

Classification of Ratings and Comments

- Individual rating ≥4: Negative rating & comment
- Individual rating =3: Neutral rating & comment
- Individual rating ≤2: Positive rating & comment

Figure 1. Example of Online Hospital Reviews

Figure 2. Distribution of Ratings and Comments (N=26,141)

- Negative: 44.1%
- Neutral: 52.3%
- Positive: 3.6%

Figure 3. Top 20 Most Frequent 2-gram Terms in Positive and Negative Comments and Their Frequencies

- Positive Comments
  - highly recommend: 629
  - emergency room: 553
  - waiting room: 681
  - worst hospital: 470
  - sent home: 392
  - don't know: 418
  - hospital staff: 473
  - year old: 342
  - don't care: 300
  - family member: 300
  - went (to) er: 273
  - treated like: 245
  - family: 174
  - loved ones: 159
  - care (about) patients: 247
  - make sure: 245
  - feel like: 245
  - urgent care: 245
  - worst time: 245
  - customer service: 245

- Negative Comments
  - insurance billing:
  - hospital staff:
  - wait time:
  - appointment:
  - long:
  - finally:
  - blood:

Figure 4. Top Representative Words of 4 Text Clusters

RESULTS (cont.)

- Representative words indicate topics of each cluster.
- Topics were decided mainly based on the words in red

DISCUSSION

- Among the 20 most frequent 2-gram terms in negative comments, 3 were about waiting times and 3 were about service quality.
- Among the 20 most frequent 2-gram terms in positive comments, 6 referred to the care provided and 6 mentioned hospital staff.
- Topics of the 4 clusters among negative comments were identified: waiting time and appointments, insurance and billing issues, bad experiences of senior family members, and poor service quality.

CONCLUSIONS

- To provide more satisfactory service, hospitals may wish to reduce waiting times, improve billing and payment processes, make facilities more accessible for older patients, and improve service quality.
- Hospitals may also wish to make patients feel they are receiving excellent and friendly care from hospital staff.
- Patients' feedback on crowd-sourced review websites is a useful data source for understanding patients' concerns and expectations.
- Text mining techniques dramatically improve the efficiency of classifying patients' feedback and identifying their underlying meaning.
- Hospital administrators should value and utilize their patients' feedback on the internet.