Assignment 4

Nathan Cochran

Computer Science, Columbia College

CISS472, Computer Architecture

Professor Rigsby

3/30/2025

Assignment 4

Introduction

Text mining is a subset of data mining that is applied specifically to large amounts of text. Specifically, it is the process of identifying patterns and other valuable insights from unstructured text data. Web mining refers to the application of data mining algorithms to process data from web sources. With the amount of data that is generated on the web, web mining is a prominent player in the world of data mining and text mining. In this article, we will be examining a case study of how a company was able to successfully use text mining and web mining applications in order to gain a better understanding of how certain features in Bluetooth speakers were associated with positive customer ratings.

Description of Case Study: Tool Selection and Implementation

The company we are analyzing in this article is Amazon. In this report by Amazon, there were 5 Bluetooth speakers that were used for analysis: Beats Pill 2.0 Portable Speaker (Black), JBL Charge 3 JBLCHARGE3BLKAM Waterproof Portable Bluetooth Speaker (Black), Ultimate Ears MEGABOOM Charcoal Wireless Mobile Bluetooth Speaker (Waterproof and Shockproof), Bose Sound Link Color Bluetooth Speaker (Black), and Altec Lansing iM227 Orbit MP3 Speaker.

In regards to data preparation, it is important to note that each “product in Amazon is assigned to a unique 10 digits Hexadecimal ASIN code.” (Mellachervu & Minukuri, 2018) Each product page (associated with a ASIN code) can have multiple subpages of reviews, with each subpage consistings of a maximum of 10 reviews.

In regards to what tools Amazon used, there are several worth mentioning. To start, Amazon used Python to extract reviews from the products using the ASIN code and range of pages. Data was then temporarily stored in an Excel workbook before being converted into a SATA data set where further analysis will be performed.

After using Python to extract data, Amazon included a few variables in their data set, but the only "variables used for analysis are only User Rating and Comment." (Mellachervu & Minukuri, 2018) The SAS data set generated for each product is then "imported into SAS Enterprise Miner 14.2, which is then partitioned into two data sets using the filter node, one for 1 & 2 ratings and other for 4 & 5 ratings." (Mellachervu & Minukuri, 2018)

Using the data from the data sets above, Amazon then used text parsing in order to generate a term by document matrix which was used to identify the frequency of words and the number of comments that used each word. What the researchers found is that the "most commonly used words by reviewers in the comments are speaker, not, charge, sound, quality, which is expected as these words relate to the features of speakers." (Mellachervu & Minukuri, 2018) Next, the Amazon researchers used a text filter to remove words that did not meet a certain threshold (the minimum number of documents a word must occur in to be retained is set to 4).

Using concept link analysis, Amazon researchers were able to determine what words were associated with what other words. For example, one concept link diagram that was generated showed that the word "PRICE is highly associated with words QUALITY, BASS, SPEAKER, SOUND which suggests that customers relate these parameters to the price of the product." (Mellachervu & Minukuri, 2018)

After filtering the data, Amazon used the Text topic feature from the SAS Enterprise Miner, which enables SAS to "combine terms into topics for obtaining valuable insights from data." (Mellachervu & Minukuri, 2018) By looking at the topics, the researchers were able to see key insights such as that reviewers expressed concerns "that battery discharges quickly" or that the "pricing could be better." (Mellachervu & Minukuri, 2018)

SAS Enterprise miner allows for grouping terms together which helps the researchers gain insights into what features were associated with positive or negative opinions/ratings/reviews For example, one cluster of data showed that a speaker was viewed as portable because of its small form factor.

In the summary of the report the researchers listed some of the biggest takeaways, and the gains from each. For example, they reported that the "best selling product in the market is JBL with average rating 4.5," which will help "us understand the features that we need to incorporate while launching a similar product in the market to withstand the competitor strategies." (Mellachervu & Minukuri, 2018) Then, in regards to customer satisfaction the researchers noted that noise and price were the biggest factors determining lower ratings, leading them to conclude that "we should try to overcome these lower rating features that avoid customer dissatisfaction on the new launches." (Mellachervu & Minukuri, 2018) The success of the text/web mining can be seen by all of the valuable insights that were gained. In short, the resulting analytics from the text mining "benefits the amazon retailers to constantly view and evaluate the customer's reviews and increase their brand loyalty by following the above recommendations." (Mellachervu & Minukuri, 2018)

Analysis of Case Study: Success Factors and Room for Improvement

Looking back at the insights gained from the case study by Amazon, there are a couple different success factors we could highlight. One of the biggest ones that stands out is the choice of smart tools that work together to provide the flexibility to properly deal with the data, both in the data preprocessing phase as well as the actual analysis of the data. To reiterate, some of the tools used were Python, SAS® Text Miner, and SAS® Sentiment Analysis. Another important factor leading to the success of the study was the quality of data collected. Specifically, they collected text from the reviews on the products they were selling, which is a reliable source for determining how people view the product.

As far as room for improvement, there are a couple things that come to my mind. First, the sample size of products is relatively small. While examining customers' views regarding the 5 products definitely gave valuable insights, for a more comprehensive view of customer perception, it would be helpful to expand the data collection to include a larger variety of products. Secondly, the sentiment analysis could probably be refined a bit more to better understand customer motivations for buying the product.

Conclusion

In conclusion, this case study did an excellent job demonstrating how text and web mining can provide valuable insight to businesses who are trying to optimize their profit by meeting the needs of their customers. In this case specifically, Amazon was able to identify what factors led to positive reviews, and what factors led to negative reviews, which collectively helped provide direction as to how they could improve their products and marketing.

References

Mellachervu, M., & Minukuri, A. (2018). *Analyzing Amazon's Customer Reviews using SAS® Text Miner for Devising Successful Product Launch Strategies*. MidWest SAS Users Group (MWSUG). https://www.mwsug.org/proceedings/2018/SB/MWSUG-2018-SB-89.pdf

Sharda, R., Delen, D., Turban, E. (2019-01-04). *Analytics, Data Science, & Artificial Intelligence, 11th Edition*. [[VitalSource Bookshelf version]]. Retrieved from vbk://9780135192382