CSCI 5302 - Final Project

Patrick Connelly Aneesh Khole Uttara Ketkar

2024-02-20

Table of contents

1 Introduction

1.1 Concept and Motivation

Customers, when searching for products with specific features and aspects, need sufficient information to make a decision as to whether to procure a specific product. *(Information on search products vs. experience vs. mixed).* When a product is more in the directon of experience vs. search-based, other customers' experiences can shed light on its features and return on investment than information directly from the vendor can. Having reviews from reliable sources with sufficiently detailed information can enable greater confidence in a purchase, improved customer satisfaction, and smooth the process of ecommerce for customers.

We seek to expound upon the research of (**?**) to explore additional recommended research areas to improve upon and increase the general applicability of the model.

1.2 What / Where

(**?**) provided the following areas for recommended additional research at the conclusion of their paper:

1. Expand the number of products beyond 3 items (one search, one experience, one mixed) to better generalize the model.
2. Explore customer or reviewer metadata for classifying reviewer types to enhance model performance.

We seek to explore Item #1 and #2 above, and to explore the possibility of assessing a scale for products to determine the extent to which they are a search or experience-based product. We've explored work from other research teams to identify potential methods we can leverage to pursue these ends.

- Determine the polarity of a customer review by employing a classifier such as Naive Bayes.
- Use Kansei engineering approaches to convert unstructured product-related texts into feature–affective opinions.
- Attempt to assess the reliability of a customer's review based on star-rating and a 'sentiment score' of their textual feedback.

Exploring combinations of these research methods, we will pursue potential improvements on the models outlined in (**?**). We will examine additional products and product types between multiple e-commerce websites (BestBuy, Target, Amazon).

1.3 Why It Matters

Feedback from customers is beneficial, but it is not always ordered by the most informative or beneficial feedback first. Certain features of data such as... can impact the usefulness of the feedback on a customer-by-customer basis. Level of detail, star-rating, and number of votes that support the review as being useful to a customer can all help determine its usefulness to other customers. Were e-commerce

Examining additional product types can enable the generalization of the authors' methodology to other products. Furthermore, the exploration of a sliding scale for search vs. experience-based products can further support generalization and business goals. Producing a reliable scale and methods for classifying a products' degree of being experienced-based can inform vendors on:

- How to best sort product reviews
- Examine what are the most helpful reviews to know the performance of the product alongside customer experience and sentiment
- Adjust the product, its marketing, or future production based upon market efficacy.
- Understanding the emotions a customer wants to express through a review is crucial as it will affect the "recommendation score" of that particular product or a different one from a similar category.
  - To contribute in determining this recommendation score, we can use a probabilistic machine learning algorithm like Naive Bayes to determine the polarity (positive, negative, or neutral) of customer reviews.
  - Typically used for amending product design, Kansei Engineering can be used to incorporate human emotional responses into evaluation of a customer review.
- Determine which customer is trustworthy, meaning who has actually purchased the product versus a customer who gave a false review. Based on the 'customer reputation score', our aim is to classify customers into groups to judge reviewer reliability. This has two main aspects:
  - Star-rating score which is a discrete scale that tells the inclination of a customer.
  - Text review 'sentiment score' using NLP that explains customer opinions based on words.

1.4 Literature Survey

- **Additional commentary on original paper here**
- **Paper(s) on product classification (search-experience-mixed)**
- **Paper(s) on user/consumer/reviewer classification**
- **All papers you've found, provide a summary of what they did and any key results**
  - Hu, W., Gong, Z., & Guo, J. (2010). Mining Product Features from Online Reviews. 2010 IEEE 7th International Conference on E-Business Engineering. doi:10.1109/icebe.2010.51
    * The proposed system employs a two-step process for opinion mining: identifying opinion sentences using a SentiWordNet-based algorithm and extracting product features from all reviews in the database. This feature extraction function focuses on identifying commonly expressed positive or negative opinions before extracting explicit and implicit product features.
  - Rajeev, P. V., & Rekha, V. S. (2015). Recommending products to customers using opinion mining of online product reviews and features. 2015 International Conference on Circuits, Power and Computing Technologies [ICCPCT-2015]. doi:10.1109/iccpct.2015.7159433

    * This paper presents techniques like Opinion mining, feature extraction and Naives Bayes classification for review polarity determination. The authors suggest performing both Objective and Subjective analysis of features by considering qualitative and quantitative features of the data respectively.

  - Wang, W. M., Li, Z., Tian, Z. G., Wang, J. W., & Cheng, M. N. (2018). Extracting and summarizing affective features and responses from online product descriptions and reviews: A Kansei text mining approach. Engineering Applications of Artificial Intelligence, 73, 149–162. doi:10.1016/j.engappai.2018.05.005

    * Authors have proposed a solution by implementing Kansei engineering and text mining simultaneously which will help customers in decision making process. It helps to categorize reviews into multiple sections and perform text mining by NLP techniques like Sentence segmentation, Tokenization, POS tagging

1.5 Research Questions

- Can the model from (**?**) be generalized with
  - larger volume of products and product types from which to mine data?
  - a sliding scalar multiplier representing the degree to which a product is a "search" (0) or "experience" (1) product?
  - Adding modifiers to review content based upon:
    * Customer / Reviewer Reliability?
- Can the polarity of reviews be judged accurately by using a Naive Bayes classification model?
  - What is the impact of different feature extraction methods (e.g., bag-of-words, TF-IDF) on the performance of Naive Bayes classification model?
- Can products be classified on their degree of being search or experience based by examining product variables such as:
  - Degree of specificity in the product description?
  - Whether the product is offered in brand-new condition only, or offered as new, used, or refurbished?
  - Which of the 5 senses the product engages?
  - Item rarity (limited production or unique items vs. bulk-produced items)?
- Can newer natrual language processing libraries provide a better fit for (**?**) 's Review Content metrics?
- How does sentiment in customer reviews correlate with customer satisfaction metrics or sales figures for a particular product?
- Can we categorize customer reviews based on customer experience and sentiment?
- Do specific product star ratings tend to incite more reviews, and if so, how does this impact the overall reputation measurement?
- Are specific quality descriptors in text-based reviews (e.g., 'enthusiastic', 'disappointed') strongly associated with certain rating levels, and how does this association affect product reputation?

1.6 Goals / Definition of Success

- Replicate similar results to (**?**) with similar product types
- Expound upon (**?**) with additional products, including:
  a. Original products from (paper): Digital Music, Video Game, and Grocery Item
  b. Additional products (Amazon and Target): Furniture Items, Clothing Items, Home Appliances, Books, Cosmetics, Cleaning supplies
  c. Additional Proucts (Amazon, Target, BestBuy): Electronics
  d. Verify goodness of fit of original model
- Determing best metrics and/or modifiers for Review Content and Customer Reliability
- Achieving similar or better fit than original paper's modeling; extrapolate to other product types.
- Determining strength of correlation metrics (support, confidence, lift) between Naive Bayes' classifier for review polarity
  - Integrate with the model and test if Naive Bayes shows strong correlation metrics.
  - Compare and contrast the model with and without incorporation.
-

2 Data

Here's our data section.

2.1 Some items to research

- Wikipedia
-

2.2 Collection

In collecting our data, in order to adhere to the model implemented by *(original paper here)*, we require the following data points, at a minimum:

Table 2.1

Variable	Data Type
Star Rating	float
Review Content	string
Useful Votes	integer

To classify products...we need metadata of the product itself. We are seeking to use the following products, and for each, we'll collect the following data:

- Original Products: (update to table later)
  - Video Games (rating I.e. pg, pg-13, R, but from MSRB ratings).
  - Digital/Physical Music (duration, music style / genre, others?)
  - Grocery Products (calories per serving, special markers [gluten-free, fat-free, vegan],) - *potential hypothesis that grocery products with special markings may be more experience-based than search-based.*

Additional Products

- Clothing - likely a mixed product, potentially ranging from search-based for plain t-shirts to experience-based for high fashion products
  - material?
- Laptops / Computers
  - Likely a mix between search and experience
  - RAM, CPU, Storage, OS Included, Graphics Type, ...
- Something else?

2.3 Preparation, Standardization, and Cleaning

- Gathering from Amazon (All Products)
- Gathering from BestBuy (Electronic Products, no grocery or clothing)
- Gathering from Target (All products)

2.4 Visualization

3 Models Implemented

Here's where we'll list out our models

3.1 Examined Models from Literature Research

- Perceived Usefulness of Product Reviews
  - Modeling for search vs. experience goods as "mitigations" on reviews, examining their impact on perceived review usefulness
- Wealth of Data Paper on Customer Evaulation
  - Examination of 'Vine' customers from Amazon
  - Don't think that BestBuy or Target have the same.
  - Scope of this would be limited to Amazon, combining research from Perceived Usefulness paper (i.e. their modeling) with a customer "trustworthiness" score for Amazon Vine customers
    * May have challenges identifying products with Vine customers
    * Should also have a look at "verified purchase" customers, too.
  - Weaknesses (from authors)
    * Their NLP instance wasn't well trained for fully accurate
    * Their weighting system was manually (i.e. arbitrarily) set as opposed to having a labeled dataset
    * Can we fix these problems? If so - what data is needed?

3.2 Potential Model for Search vs. Experience Weight (scale of 0-1)

3.2.1 Jaccard Similarity Measure

https://www.sciencedirect.com/science/article/pii/S1567422318300450

- Can be used to produce a similarity between two items on a scale of 0 to 1
- May be able to use this for evaluation of a novel item against a pure-search good vs. a pure-experience good.
  - May require additional calculation / computation between these two values - maybe its arithmetic or harmonic mean between search and experience
  - Using that, we could potentially produce a weight.
  - Authors models
    * Model 1: Perceived Usefulness $= \beta_0 + \beta_1 \cdot \text{Review Content} + \beta_2 \cdot \text{Review Length} + \beta_3 \cdot \text{Star Rating} + \beta_4 \cdot \text{Total Votes Received} + \epsilon_1$

* Model 2 (product type as a moderator): Perceived Usefulness $= \beta_0 + \beta_1 \cdot$ Review Content $+ \beta_2 \cdot$ Review Length $+ \beta_3 \cdot$ Star Rating $+ \beta_4 \cdot$ Total Votes Received $+ \beta_5 \cdot$ Digital Music $+ \beta_6 \cdot$ Video Game $+ \beta_7 \cdot$ Review Content $\cdot$ Digital Music $+ \beta_8 \cdot$ Review Content $\cdot$ Video Game $+ \epsilon_2$
* Our proposition #1 (may require some modification)

  Perceived Usefulness $= \beta_0 + \beta_1 \cdot$Review Content$+\beta_2\cdot$Review Length$+\beta_3\cdot$Star Rating$+\beta_4\cdot$Total Votes Recei

or

Perceived Usefulness $= \beta_0 + \gamma \cdot \beta_1 \cdot$Review Content$+\beta_2\cdot$Review Length$+\beta_3\cdot$Star Rating$+\beta_4\cdot$Total Votes Received

- Where γ is the Jaccard Similarity Score between a given product and elements we are identifying as "pure" experience and "pure" search good.
- Operationalizaton of Jaccard Similarity score Variable (i.e. inputs)
  - Search good: those with attributes that can be evaluated prior to purchase or consumption. Consumers rely on prior experience, direct product inspection and other information search activities to locate information that assists in the evaluation process. Most products fall into the search goods category (e.g. clothing, office stationery, home furnishings).
  - Number of measurement specifications?
  - comment/review information?
  - What else could we gather that could be considered a "universal" tangible or intangible feature from a product online? They need to be applicable to both search and experience goods.
- Experience good: those that can be accurately evaluated only after the product has been purchased and experienced. Many personal services fall into this category (e.g. restaurant, hairdresser, beauty salon, theme park, travel, holiday).
  - subjective descriptiveness vs.
  - comment/review information?
- Are there other methods aside from Jaccard Similarity?

3.3 Model Comparison

4 Conclusions

Here's our conclusions section