Project report 1

Johnny Morgan

03/25/224

IS804 Advanced Quantitative Methods in IS Research

Professor: Ida Ngambeki

Spring 2024

**Project Report Part 1 (20):** This is the first step in your course project.

You will prepare a research document that will include

* an abstract (summarize problem, topic, importance, motivation, questions, procedure, sample, goals, and expected outcomes);
* introduction (historical context, project goals, contribution of study);
* literature review (place work in context, develop hypothesis/ demonstrate gap to be answered by questions); aims and objectives;
* research questions/hypotheses;
* methods (theoretical basis for methods to convince reader of validity of your approach, population and sample with a justification, procedures for data collection with sufficient detail, procedures to address validity and reliability);
* research justification
  + (describe the intellectual merit of the research i.e.
  + the potential to advance knowledge in yours and other disciplines,
  + and the broader impacts i.e.
    - the potential of the research to benefit society –
    - be specific about who will benefit and how,
  + and a plan to disseminate research and leverage research findings to ensure the benefits to discipline and society).

This proposal should be 7-10 pages. Think of this as the first part of a conference paper.

**Introduction**

The purpose of this paper is to examine the use of intent modeling to quantitatively characterize customer reviews of Amazon products. This is a demonstration of several AI modeling capabilities which are combined to synthesize customer intent through the use of fine tuning large language models to perform text classification of reviewer intent. The field of soliciting product reviews and feedback from customers is intended to be informative to product producers and to guide product development in the fields of product features, marketing, distribution, and suitability. The tradecraft of analyzing large corpus review data has not included intent modeling writ large. This research seeks to develop scalable AI based intent processing that can inform quantitative studies of customer interactions. Initial successes are promising and are expected to improve dramatically with fine tuning and human in the loop model training.

Current experimental AI processing code, developed in this research, has resulted in developing 20 intent tags. These intent tags were based on an analysis of a sample of 10,000 sentences to examine and quantify 20 common intents the customers expressed in their reviews. These intent tags were then provided to a facebook/bart-large-mnli model. This methodology has successfully provided a multi classification process that characterizes the intent of the customer when writing the sentences. These intent tags were filtered to include those tags which were over a 80% confidence threshold. Once processing has been completed these intent tags will be analyzed using feature importance methodology to evaluate the relative importance of the intents as they correlate to customer product review ratings. The underlying theory driving this research is the belief that intent tagging will result in a method of text classification that will provide generalizable utility in knowledge extraction from human corpuses.

**Abstract**

**Summary of Problem and Topic:** Quantitative analysis of consumer reviews, historically, have been performed based on content, sentiment and magnitude, customer scores, and reviewer helpfulness. Product feedback is difficult to obtain and difficult to translate into actionable responses. Customer feedback is difficult to parse as it is unstructured, difficult to extract information, and difficult to produce quantitative analysis of large corpus data. There are multitudes of ways to communicate the same information. Whereas customers who provide feedback and product reviews are motivated to communicate useful information either to the producer or to other customers, mining this data requires significant quantities of human resources or computer resources. Human resources are extremely limited and are typically focused on high value information or focused on lower complexity data that is easily obtained and aggregated.

**Importance and Motivation**: With the advent of artificial intelligence, it is increasingly evident that this technology can be developed to interpret human generated data and interpret this data to fit the data into structured data sets. In recent years AI technologies have been developed that reliably predict/characterize sentiment. With this capability, a data set can be analyzed to produce a numerical representation of positivity, negativity, and even sarcasm, and a corresponding magnitude. However nuanced characterization of customer reviews is difficult to process programmatically, as such, reviews are mostly presented for other consumers to assist them in product choice. Developing nuanced extraction of actionable data that can be provided to producers is the goal of this research.

**Research Questions and Procedure**:

**Research Topic:** This research project proposes to provide valuable insights into customer perspectives through quantitative analysis of customer intent in product reviews to produce actionable insights to inform product development, distribution, and marketing.

**Research question**: How can Intent Modeling of Consumer Product reviews, decoding nuanced consumer intentions, provide comprehensive product value scoring to improve product features and provide better alignment with consumer expectations?

**Data encoding (tagging):** The overall intent modeling process is generated through the use of large language model embeddings which capture broad scope semantic relationships between words and concepts as they are understood in general pretrained models and can be fine-tuned by training models on the local corpus. Semantic similarity is calculated based on these embeddings and intent tagging is propagated throughout the corpus using semantic similarity. Finally intent tags are scrutinized through LLM analysis of clusters and human in the loop validation of the tags finalizes the tagging process.

**Quantitative analysis**: Intent encoding will be evaluated using quantitative methods to produce classes of intents and generating weighted scores of intent classes. Product scores provide comparisons between products and provide customer derived valuation of product classes.

**Importance**: The methodology proposed in this project differs significantly from popular customer feedback analysis in that it devises a dynamic intent tagging scheme derived from semantic clustering of customer product reviews. This provides a generalizable methodology that can be used across product classes, focused on characterization of product utility, customer recommendation of the product, criticism of product, criticism of marketing, criticism of packaging, characterization of suitability, etc derived semantically from customer content.

This is in contrast to current customer review methodologies:

* Customer reviews and feedback can be evaluated for sentiment providing a range of positivity/negativity using Valence Aware Dictionary and Sentiment Reasoner (VADER)
* Customer Value Score Method (CVSM) that focuses on Profit, revenue, and positive reviews (Huang et al., 2022)
* Semantic analysis of functional, behavioral, and structural domains (Xuanyu et al., 2022)

**Goals**:

The work I am doing is to develop a method of extracting and classifying the intent of the customer when writing the review. This data will then be correlated with the score that the customer provided in order to better understand not just the review but why the customer provided the review. Thus, the qualitative intent can be quantitatively evaluated.

**Literature Review**

"Online Reviews and Product Sales: The Role of Review Visibility" by Alzate et al. (2021)

**Contextual Placement**: Alzate et al. (2021) investigate the relationships between online review visibility and product sales. This diverges from typical sentiment focused review analysis. The paper’s contextual placement is situated in e-commerce and consumer behavior studies. This spotlights the impact of placement of how reviews are presented to consumers and the impact this has on the purchase intent of the consumer.

**Hypothesis Development**: Authors propose that review visibility, affected by sorting methods such as helpfulness and recency, play a critical role in their influence on sales. This perspective challenges prevailing focus on sentiment analysis in customer feedback analysis.

**Gap Identification**: This research pinpoints a neglected area- the effect of review visibility on sales. They note the oversite in existing literature that assumes a uniform influence of all reviews, highlighting the necessity of exploring consumer interaction with review presentation beyond sentiment, to fully grasp impact on sales.

“A study on the influence of online reviews of new products on consumers’ purchase decisions: An empirical study on JD.com” by "Kang et al. (2022)

**Contextual Placement:** Kang et al. (2022) examine the effect of online reviews on consumer purchase decisions for new products on JD.com. They expand the scope beyond sentiment analysis. Their research underscores the significance of review content and volume in shaping buyer choices.

**Hypothesis Development:** This research suggests that various dimensions of online reviews, including volume and sentiment polarity, significantly impact consumer purchasing behavior. This view shifts the analytical lens from mere sentiment evaluation to a multifaceted analysis of reviews, proposing a more comprehensive approach to understanding consumer interactions with online feedback.

**Gap Identification:** Kang et al. point out that while some research has applied topic modeling to discern consumer concerns in reviews, there's a significant gap in combining these insights with reviews' various characteristics to study their effect on consumer behavior. This indicates a need for deeper content analysis within reviews to better understand how specific feedback aspects influence purchase decisions.

“Sentiment Analysis of Customer Reviews of Food Delivery Services Using Deep Learning and Explainable Artificial Intelligence: Systematic Review” Adak et al. (2022)

**Contextual Placement:** Adak et al. examine the importance of online customer reviews in the food delivery service industry in the context of the covid-19 pandemic. This research examines sentiment analysis using AI/ML technology. This research explores the role of customer feedback in service improvement and customer satisfaction and the trend to digital customer engagement.

**Hypothesis Development:** This research suggests that there is a gap in employing deep learning models compared to employing lexicon-based sentiment analysis based on the relative lack of explainability of deep learning models.

**Gap Identification:** This research explores the gap in the application of deep learning and explainable AI in sentiment analysis.

Overall, from my literature review search I find that the most prevalent analysis methodology associated with customer feedback and product reviews focus on sentiment. This is mostly due to the relative maturity of sentiment analysis technology and the relative trust associated with sentiment analysis. I find very little relevant research on other methodology for this purpose. This supports my theory that methodology such as intent modeling, if developed and matured, could have a significant impact on customer feedback analysis.

**Aims and Objectives**

Develop a dynamic intent tagging scheme that can generalize across product classes. Produce a product value scoring system informed by quantitatively analyzed customer intents. Demonstrate the methodology’s superiority in deriving actionable product development insights.

**Research Questions/Hypotheses**

**Main Question**:

How can intent modeling of consumer product reviews enhance product value scoring and alignment with consumer expectations?

**Hypotheses**:

It is posited that dynamic intent tagging, leveraging advanced NLP techniques, will provide deeper insights into consumer perspectives than traditional sentiment analysis methods.

**Methods:**

**Theoretical Basis**:

Once intent tagging can be demonstrated it is theorized that the intent tags will provide an avenue to perform knowledge extraction from the review corpus and provide significantly better fidelity in actionable data extracted from the corpus. Quantitative models of the relationship between the intent tags and customer product ratings are expected to reveal a nuanced set of expectations that correspond to the ratings. Feature importance calculations can be performed to interpret the relative meaning of intent tags. It is theorized that important intents can be quantified based on feature importance and relationship to rating to develop a product value score based on consumer intent.

**Sample**:

Data Source - Amazon Product Reviews dataset from Kaggle, containing 568K narrative product reviews. Dataset is 115MB in size and is delivered as a CSV file. Data contains a unique id for each review, ProductId, UserID, ProfileName, HelpfulnessNumerator, HelpfulnessDenominator, Score, Time, Summary, and Text.

Id 6

ProductId B006K2ZZ7K

UserId ADT0SRK1MGOEU

ProfileName Twoapennything

HelpfulnessNumerator 0

HelpfulnessDenominator 0

Score 4

Time 1342051200

Summary Nice Taffy

Paragraph I got a wild hair for taffy and ordered this f...

Paragraph\_vector [-0.08328723, 0.08941264, 0.39326283, 0.106011...

Summary\_vector [0.45653412, -0.095085524, 0.59296936, -0.1120...

I have enriched the dataset to include BERT vector embeddings of the summary and the paragraph.

I pulled the paragraph data into another dataframe where I have broken out each sentence. This data is comprised of 2.8Million sentences.

'Id', 'ProductId', 'UserId', 'ProfileName', 'HelpfulnessNumerator',

'HelpfulnessDenominator', 'Score', 'Time', 'Summary', 'P\_index', 'S\_sentence\_number', 'Sentence', 'Summary\_vector', 'Sentence\_vector',

'Sentence\_Cluster', 'Sentence\_Distance', Sentence\_similarity\_distance',

'Sentence\_similarity\_index'

This dataframe was also enriched with BERT vector embeddings.

I have conducted cosine similarity calculations to identify the 100 most similar paragraphs, summary, and sentences. The sentence index and distance are stored in columns in the df\_sentences and df\_paragraph dataframes.

Id 1

ProductId B001E4KFG0

UserId A3SGXH7AUHU8GW

ProfileName delmartian

HelpfulnessNumerator 1

HelpfulnessDenominator 1

Score 5

Time 1303862400

Summary Good Quality Dog Food

Paragraph I have bought several of the Vitality canned d...

Paragraph\_vector [-0.12606543, 0.062131397, 0.0551221, -0.00242...

Summary\_vector [0.10738717, 0.015432298, -0.28650856, 0.00476...

Paragraph\_Cluster 97.0

Paragraph\_Distance 8.618866

Summary\_Cluster 72.0

Summary\_Distance 18.906586

Summary\_similarity\_distance [0.0, 0.0, 3.1913757, 3.4827728, 3.4827728, 4....

Summary\_similarity\_index [382666, 0, 399359, 489673, 53168, 221484, 245...

Paragraph\_similarity\_distance [0.0, 9.213089, 9.347366, 9.460304, 9.700714, ...

Paragraph\_similarity\_index [0, 357080, 399351, 118925, 87038, 174270, 779...

Name: 0, dtype: object

Each of these similarity indexes are provided as columns in the dataframe.

I have processed each of the vector columns into Facebook AI Similarity Search (FAISS) indexes. These indexes provide fast access to similarity distance data of the vectors.  
  
I have experimented with numerous huggingface text classification models and have concluded that the classification labels that I require are not represented in any published models. Therefore, I will train my own models to perform this classification.

The data provides reviews of 74258 products. With a range of review counts:

count

[1, 2) 30408

[2, 5) 23435

[5, 10) 9796

[10, 50) 8775

[50, 100) 978

[100, 1000) 866

74258

The distribution of scores:

|  |  |
| --- | --- |
| Score | count |
| 1 | 52268 |
| 2 | 29744 |
| 3 | 42638 |
| 4 | 80655 |
| 5 | 363122 |
|  | 568427 |

**Data Collection Procedures**:

Data collection involves the extraction of review texts from the Amazon customer product review dataset, preprocessing using NLP libraries, and subsequent fine-tuning of large language models for multi label text classification.

Validity and Reliability Procedures:

To ensure validity and reliability of the intent modeling process a quantitative analysis of text similarity and reliability examinations of the intent tags will be performed leveraging the BERT embeddings which have been processed for all text data sets. These embeddings provide the capability of performing similarity analysis based on semantic content. These semantic analyses will provide a calculable, explainable process to validate and fine tune the intent tagging model.

Once the new server level compute platform is received, another round of validation processing will be employed to perform generative large language model processing using a Generative Adversarial Network (GAN) approach to validate and fine tune the intent tagging process.

Currently the compute platform is comprised of a core I7 processor quad core processor running at 2.6 gigahertz, 64 gigabytes ram, running an NVIDIA GTX 980M GPU. The new server hardware has dual Xeon E5-2650 12 core processors running at 2.9Ghz 12 cores each, 256 gigs of ram, and 4 tesla P100 GPUs. This is an increase of 20 CPU cores, 192 gb more ram, and 4 tesla P100 GPUs represents an increase 34 TFLOPS processing capacity. In this compute environment it will be possible to perform significantly more difficult AI processing and model processing such as a GAN

**Methodology**:

I have experimented with Latent Dirichlet Allocation (LDA), Term Frequency Inverse Document Frequency (TFIDF), and word cloud distributions. I have also experimented with kmeans clustering to reduce the cognitive load in developing the intent models.  
  
The most likely path to success will either involve leveraging a generative model to evaluate intent tags from the text, or to calculate similarity distances of a set of intent tags produced by a human agent (myself in this case) and produce a set of intent columns such that for each vector row (summary, paragraph, and sentence) is calculated cosine similarity to intent tags or phrases and the distance below a threshold is considered significant and included as calculated intent.   
  
Exploration of classification models published in Hugging Face reveals that there is significant interest in intent classification modeling.

I have had success in developing intent tagging based on the facebook/bart-large-mnli model. This model was trained to produce zero shot text classification. It is responsive and generates classification tags based on semantic analysis of intent tags.

This being a preliminary exploration of intent processing, a small set of intent tags have been developed that are expected to be useful and representative of actionable intents, expected to be informative to producers.

Intent tags in this experiment:

Quality Appreciation, Product Description, Product Appearance, Preference Expression, Packaging Issue, Comparison, Complaint, Misrepresentation, Historical Mention, Taste/Flavor Comment, Product Ingredients, Value for Money, Purchase Recommendation, Usage Experience, Product Efficacy, Health and Safety Concerns, Customer Service Experience, Repeat Purchase Intention, Emotional Response, Environmental/Sustainability Mention

Compute resource and time constraints have limited the processing of these data sets and limited the intent tags being examined. The Amazon customer review dataset that is being analyzed consists of 568k product reviews. Examining these reviews in a sentence context explodes the data to 2.8m sentences. Processing time on our current compute platform requires 1.7seconds on average to process. Thus, the time required exceeds time allowances in this course. However, a larger compute platform is enroute, so there is potential that the processing to fully demonstrate the utility of this experiment may be possible within time allotted in this course.

**Research Justification**

**Intellectual Merit**:

This research provides an innovative approach that combines intent modeling with quantitative analysis of customer product reviews to extract and classify nuanced consumer intentions. By leveraging large language model embeddings and semantic similarity calculations, this project not only advances the technical capabilities of artificial intelligence in processing unstructured data but also contributes to the theoretical understanding of consumer behavior in digital marketplaces. The utilization of intent tagging and feature importance methodology for evaluating customer review sentiments represents a significant leap in knowledge extractions techniques, pushing forward the boundaries of natural language processing (NLP) and machine learning in real-world applications.

Broader Impacts:

The potential societal benefits of this research are substantial and multi-faceted. First, by providing more accurate and nuanced insights into customer preferences and expectations, businesses can develop products and services that are informed by knowledge extraction from customer review corpus data. This will enable product development to be more closely aligned with customer needs, thus enhancing customer satisfaction. Second, this research contributes to the democratization and explainability of AI technology, enabling smaller enterprises to leverage cutting-edge AI tools for competitive advantage. Further, by improving our understanding of consumer feedback in an accessible manner, this research project has the potential to foster a more responsive and responsible business ecosystem, helping companies to adapt and rectify issues identified in customer reviews more quickly.

**Dissemination and Leverage Plan**:

The findings of this research will be disseminated through multiple channels to ensure a wide-reaching impact.

Academic journals and conferences: Results will be submitted to leading journals in artificial intelligence, data science, and marketing as well as conferences related to AI, ML and consumer research.

Industry workshops and seminars: engage directly with industry and government practitioners through workshops and seminars to demonstrate practical applications of the research findings.

Online platforms: Publishing a series of accessible articles and tutorials on platforms such as medium and GitHub aiming to reach a broader audience including tech enthusiasts, marketing professionals, and fellow researchers.

References:

Adak, C., Pradhan, B., & Shukla, P. K. (2022). Sentiment Analysis of Customer Reviews of Food Delivery Services Using Deep Learning and Explainable Artificial Intelligence: Systematic Review. Foods, 11(11), 1500. https://doi.org/10.3390/foods11111500

Alzate, M., Arce-Urriza, M., & Cebollada, J. (2021). Online reviews and product sales: The role of review visibility. Journal of Theoretical and Applied Electronic Commerce Research, 16(4), 638-669. <https://doi.org/10.3390/jtaer16040038>

*Amazon product Reviews*. (2021, July 11). Kaggle. https://www.kaggle.com/datasets/arhamrumi/amazon-product-reviews

Huang, L., Liu, M. T., Song, X., & Yen, J. (2022). A method of customer valuation score and implementation for marketing strategy. *Asia Pacific Journal of Marketing and Logistics*, *35*(2), 344–363. <https://doi.org/10.1108/apjml-05-2021-0299>

Kang, X., Sun, L., Liang, X., & Mao, W. (2022). A study on the influence of online reviews of new products on consumers’ purchase decisions: An empirical study on JD.com. Frontiers in Psychology, 13, 983060. https://doi.org/10.3389/fpsyg.2022.983060

Xuanyu, W., Zhang, H., Feng, Y., Li, M., Lou, S., & Tan, J. (2022). A semantic analysis-driven customer requirements mining method for product conceptual design. *Scientific Reports*, *12*(1). https://doi.org/10.1038/s41598-022-14396-3