Bags *and* Words: Mining Product Features and Customer Sentiment from eBags.com Reviews

Abstract

Product features, along with customer sentiment for those features, are highly desirable to specialty e-commerce retailers so they may provide expertise and value to customers in the areas of product development and curation, shopping search and discovery, personalized recommendations, and effective and efficient marketing. Success in these areas drive retail sales, growth, and customer loyalty in a highly competitive e-commerce marketplace. Written product reviews provided by customers are an authentic and rich source of valuable features and sentiment, but extracting and refining valuable information from reviews is difficult. The goal is to extract product features from raw customer reviews, and attempt to discern customer sentiment for the specific product features. The approach to achieving this goal combines NLP data cleansing and text processing, subjectivity/objectivity partitioning, part-of-speech (POS) tagging, and sentiment classification using deep learning with minimal supervision to lessen the need for costly and time consuming data labeling.

Introduction

E-commerce is highly competitive with Amazon.com dominating nearly 50% of all online spend in the US in 2018[1]. In order to compete as a specialty retailer, one proven strategy is to demonstrate category expertise by understanding and communicating product attributes and features - and their specific benefits - to prospective customers. Product attributes and features generally are provided by product manufacturers but they are usually just basic, rarely comprehensive, and/or they do not convey valuable benefits that customers may want or discover. Product features may also be further discerned by retailers however this can be time consuming and inconsistent. Further, product features

1 https://techcrunch.com/2018/07/13/amazons-share-of-the-us-e-commerce-market-is-now-49-or-5-of-all-retail-spend/

that are problematic or invoke negative sentiment by customers and users are almost never revealed in these activities.

It is a common practice for leading e-commerce retailers to solicit and display textual customer reviews that contain product features along with customer sentiment for these features - good or bad. An example of a product feature and customer sentiment from an actual review is:

"I love the bright color inside so you can always find things."

Where the targeted product feature is "bright color inside" and the target sentiment is extremely positive ("love"). The benefit of the feature is also revealed with "you can always find things."

eBags, LLC. is a successful specialty e-retailer featuring the world's top brands in the luggage, bag, and travel accessories categories. eBags has amassed over 3.5 million independent product ratings and reviews. There is potentially very rich product information within these reviews in the form of product features and related sentiment. It would be impossibly time consuming and cost prohibitive to manually read and extract product features and sentiment from textual reviews.

This project proposes successful mining of eBags.com product reviews for product features and benefits, along with an understanding of sentiment of the product features and benefits. The approach to achieving this goal combines NLP data cleansing and text processing, subjectivity/objectivity partitioning, part-of-speech (POS) tagging, and sentiment classification using deep learning with minimal supervision to lessen the need for costly and time consuming data labeling. If successful, this rich product information can inform product discovery and navigation, product recommender systems, drive targeted marketing campaigns, and even drive product development of valuable features into bags; all helping eBags' customers find their perfect bags and accessories for all of their life journeys and adventures thus enabling eBags.com to successfully compete as a specialty retailer in an Amazon online world.

Related Work and Motivation

Accurately extracting relevant product features from customer reviews while also discerning the customer sentiment toward these product features poses several difficult challenges beyond the common challenges dealing with raw and noisy data. While many reviews contain valuable product features and related sentiment text, they are often buried within large bodies of other text that is irrelevant to the goal. Once sentences are identified that are more subjective, and thus more likely to have details and sentiment specific to the product, they must be further analyzed to increase the likelihood that they do contain product features and related sentiment, so that information can be further refined and extracted and analyzed. Finally, these refined extractions must be accurately assessed and classified for sentiment. The approach taken to address these challenges relies and builds on work in three key areas; subjectivity/objectivity partitioning, POS tagging, and weakly-supervised deep learning for sentiment classification.

Subjectivity/Objectivity Partitioning

Subjectivity/objectivity partitioning is used to identify sentences that are more subjective and thus more likely to have product features and related sentiment. The technique of (Yeh, 2006) was employed to partition review text into subjective and objective sentences which built on the idea of TexTiling and discourse segmentation (Hearst, 1997). This work proposes that reviews are typically composed of sequences of subjective and objective chunks, and demonstrates a technique of partitioning using a language model based on a subjectivity/objectivity ratio from a corpus of subjective and objective sentences for training that was leveraged by (Pang and Lee, 2004) in their work on subjectivity summarization.

Parts-of-Speech (POS) Tagging

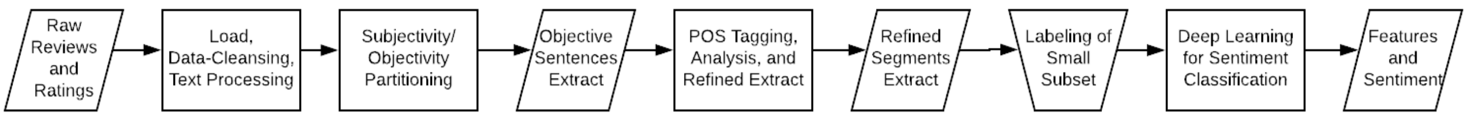
POS tagging and POS pattern analysis is used to further refine and extract text that is more likely to contain product features and related sentiment. This builds on work by (Wang and Ren, 2002) that uses POS patterns to identify interesting feature/attribute and sentiment segments. This is also based on the idea of (Turney, 2002) who also developed POS techniques for extracting phrases for review classification.

Weakly-Supervised Deep Learning for Sentiment Classification

Weakly-supervised deep learning for sentiment classification is used to identify the semantic orientation of the extracted product feature and related sentiment segments with minimal training data. This builds on the ideas and techniques used by (Guan, et. al., 2016).

Methodology

A csv dataset of 369,888 reviews was obtained with permission from eBags.com. These reviews are all in the public domain on the eBags.com website and were filtered from the total 3.5 million set by only including active products and also where at least a single rating value was present (in addition to reviews eBags also asks customers to optionally rate their product on a 1-10 scale across the dimensions of durability, price-value, appearance, organization, and overall). The following flowchart illustrates the overall flow and architecture with details for each major step presented below.



Load, Data Cleansing, and Text Processing

Raw customer reviews can be very noisy sources of natural language and information. Grammatical and spelling errors are common, and html and character codes often litter the data, which make parsing, sentence boundary detection, and identifying basic units of meaning and words difficult. The following techniques were used to clean and process the raw customer reviews; spell check and correction using Peter Norvig's NLP Spell Corrector in Python[2], html and character code search and replace through regular expressions, and Splitta sentence boundary detection by Dan Gillick[3].

[Milestone update: still completing this very time consuming step. I have been able to load and parse the 369,888 reviews and build regular expression logic to replace identified html character codes. I have also been able to implement Peter Norvig's NLP Spell Corrector but I am working on fine tuning it to ignore certain industry specific named entities I have identified in initial runs and statistics on spell correction. I have also been able to install Splitta and related components, convert some of the remaining code to Python 3, and successfully run on the provided test data. I estimate several more days to complete these tasks. The raw data as well as the python notebook for load, data cleansing, and text processing can be found in the github project directory: https://github.com/mfrazzini-MIDS/w266_Project/tree/master/data]

2 http://norvig.com/spell-correct.html
3 https://code.google.com/archive/p/splitta

Subjectivity/Objectivity Partitioning

[Milestone update: will move on to this task when load, data cleansing, and text processing is complete. Task will entail building a n-gram model to be used for subjective sentence probability analysis from a subjective / objective tagged dataset, and then to apply the model to the cleansed review data and extract more subjective sentences for further processing. I estimate this task will take about a day to complete and possibly another day to fine-tune. I have obtained the dataset and it can be found in the github project directory: https://github.com/mfrazzini-MIDS/w266_Project/tree/master/data]

Part-of-speech tagging

[Milestone update: will move on to this task when subjectivity/objectivity partitioning is complete. I intend to use Google Syntaxnet and Parsey McParseface libraries and I have successfully installed from various web tutorials, and am working through the provided tutorial notebook. The goal of this task is to further analyze the subjective sentences extracted in the prior task to extract specific product features and related sentiment words using POS patterns. I intend to explore POS patterns presented in the above referenced works as well as pattern analysis from a dataset of manufacturer provided features for bags. I will time box this task at no more than two days.]

Weakly-Supervised Deep Learning for Sentiment Classification

[Milestone update: will move onto this task after the POS task has been explored and desirable outputs are produced. I expect to label a small subset of data (approximately 1% of the extract from the previous tasks) as positive or negative and train and explore various Neural Network models and parameters to achieve the best results. I estimate this will take a week to complete.]

Experimentation and Results

[To be completed]

Conclusion and Next Steps

[To be completed]

References

[Guan et. al., 2016] Ziyu Guan, Long Chen, Wei Zhao, Yi Zheng, Shulong Tan, and Deng Cai. 2016. "Weakly-Supervised Deep Learning for Customer Review Sentiment Classification." In Proceedings of the 25th International Joint Conference on Artificial Intelligence (IJCAI'16). 3719--3725.

[Hearst, 1992] M.A. Hearst. Direction-based text interpretation as an information access refinement. In P. Jacobs (Ed.), Text-Based Intelligent Systems: Current Research and Practice in Information Extraction and Retrieval. Mahwah, NJ: Lawrence Erlbaum Associates.

[Pang and Lee, 2004] Bo Pang and Lillian Lee, A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts, Proceedings of ACL 2004.

[Turney, 2002] Peter D Turney. Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews. In ACL, pages 417–424, 2002.

[Wang and Ren, 2002] J. Wang and H.Ren, "Feature-based Customer Review Mining," System p.1-9 (2002).

[Yeh, 2006] Eric Yeh. CS224N/Ling237 Final Project Picking the Fresh from the Rotten: Quote and Sentiment Extraction from Rotten Tomatoes Movie Reviews.