Senior Project – Computer Science – 2015

Automatically Determining Review Helpfulness

Hyung Yul Choi

Advisor – Professor Kristina Striegnitz

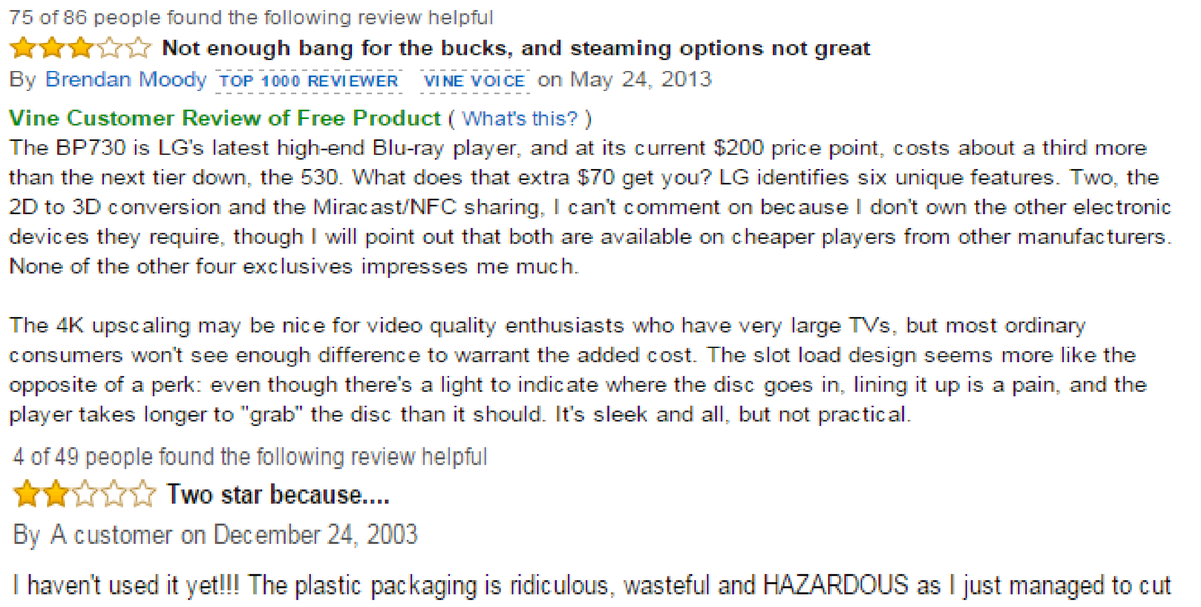
Abstract

Customer reviews from commerce websites have valuable information for online shoppers. They help shoppers gauge whether or not a product is worth the purchase. However, reviews vary considerably in their quality and helpfulness. Most commerce websites have voting systems where shoppers can vote on whether a review was helpful to them or not. This helpfulness score helps other shoppers read just the reviews that are considered most helpful. For popular products however, the number of reviews can be in the thousands. As a result, not all reviews will get enough attention to receive helpfulness votes even though some may contain helpful information for other shoppers. In these scenarios, it would be desirable to be able to automatically extract the most helpful reviews. This research aims to do this by finding features in the review text that are indicative of its helpfulness and training a learning algorithm that predicts a review's helpfulness.

Dataset

The dataset consists of reviews from Amazon. Each review has a helpfulness score, where readers can vote on whether a review was helpful or not. It is displayed as a helpfulness ratio to show how many readers found the review helpful. Though the number of reviews is over 1 million, only the relevant data is used for training. Only the reviews with ≥ 10 helpfulness votes and ≥ 5 sentences are used for training.

Total # of Reviews	1241778
# of Reviews, ≥ 10 Votes	167604
Average Helpfulness Ratio	0.78
Average Length of Review	108 words



75 of 86 people found the following review helpful

Not enough bang for the bucks, and steaming options not great

By Brendan Moody TOP 1000 REVIEWER VINE VOICE on May 24, 2013

Vine Customer Review of Free Product (What's this?)

The BP730 is LG's latest high-end Blu-ray player, and at its current $200 price point, costs about a third more than the next tier down, the 530. What does that extra $70 get you? LG identifies six unique features. Two, the 2D to 3D conversion and the Miracast/NFC sharing, I can't comment on because I don't own the other electronic devices they require, though I will point out that both are available on cheaper players from other manufacturers. None of the other four exclusives impresses me much.

The 4K upscaling may be nice for video quality enthusiasts who have very large TVs, but most ordinary consumers won't see enough difference to warrant the added cost. The slot load design seems more like the opposite of a perk: even though there's a light to indicate where the disc goes in, lining it up is a pain, and the player takes longer to "grab" the disc than it should. It's sleek and all, but not practical.

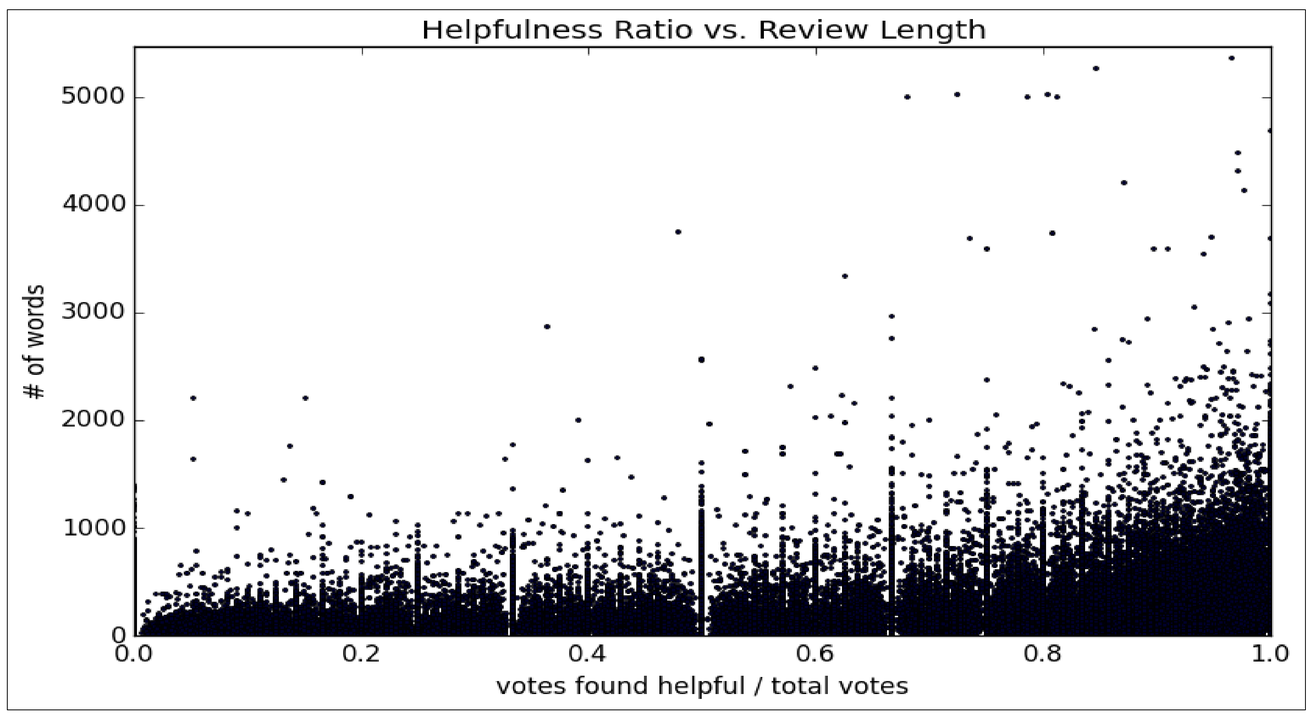
4 of 49 people found the following review helpful

Two star because....

By A customer on December 24, 2003

I haven't used it yet!!! The plastic packaging is ridiculous, wasteful and HAZARDOUS as I just managed to cut myself trying to open it. Seriously, what a waste of resources. I can't imagine the energy and waste used just for the packaging. Why go through the effort of marketing and sitting around a conference room discussing how cool it would be to package a MEMORY STICK with razor sharp plastic with a neat colorful cardboard backround just so it can be thrown away to forever stay on the face of this earth to injure another person!!!? ah!! For a MEMORY STICK? Gimme a break!!

Comment | Was this review helpful to you? Yes No Report abuse



Features

Features from the data are tested for any correlations to its helpfulness ratio using the Pearson Correlation Coefficient. For this research, we judge a review's helpfulness by its helpfulness ratio given by its readers.

Features used have correlations > 0.15. The tested features are: length, number of sentences, readability, sentiment polarity, and extraneous punctuation such as exclamation and question marks.

Results and Future Work

10 fold cross-validation used to generalize data and reduce variability. Confusion matrix used to visualize the performance of classification system. Decision tree's average accuracy is 41.9%. SVM's average accuracy is 40.5%. For evenly distributed frequencies, the classifiers are a little better than the random baseline accuracy of 33.3%.

More possible features can be explored. Advanced features such as identifying word associations with specific aspects of products may help with improving accuracy of the learning algorithm.

Decision Tree Confusion Matrix

	Poor	Neutral	Good
Poor	308	146	133
Neutral	217	201	169
Good	192	187	208

SVM Confusion Matrix

	Poor	Neutral	Good
Poor	98	22	467
Neutral	73	16	498
Good	34	9	544