Social Media Sentiment Analysis in Marketing

CASE STUDY ON SAMSUNG

Feng Ruili (DSI 13)

Overview

Problem Statement: Traditional metrics focus on statistics (number of likes, views, clicks, comments, shares, etc) which may not reflect what consumers truly think of the product.

The project aims to gather consumers' **sentiments** and **feedback** on Samsung based on latest Twitter and YouTube comments to guide **marketing strategy**.

Part 1: Sentiment analysis to gather marketing intel:

1. Brand (Samsung) vs competitors (Apple & Huawei)
2. Advertisement (Samsung's YouTube videos)

Part 2: Topic modeling to collect feedback and provide better customer support:

1. LatentDirichletAllocation (LDA)
2. Gibbs Sampling Dirichlet Mixture Model (GSDMM)
3. Biterm Topic Modeling (BTM)

Additional part: Samsung's stock price prediction based on sentiment analysis

Reference: https://www.forbes.com/sites/jiawertz/2018/11/30/why-sentiment-analysis-could-be-your-best-kept-marketing-secret/#76f0c3d12bbe

Part 1: Sentiment analysis on Samsung tweet

Dataset: Latest 2500 tweets that contain that word “Samsung”, “Apple” and “Huawei” (using Twitter API)

Sentiment analyzer: VADER (Valence Aware Dictionary for Sentiment Reasoning) from NLTK library

Preprocessing: Almost none

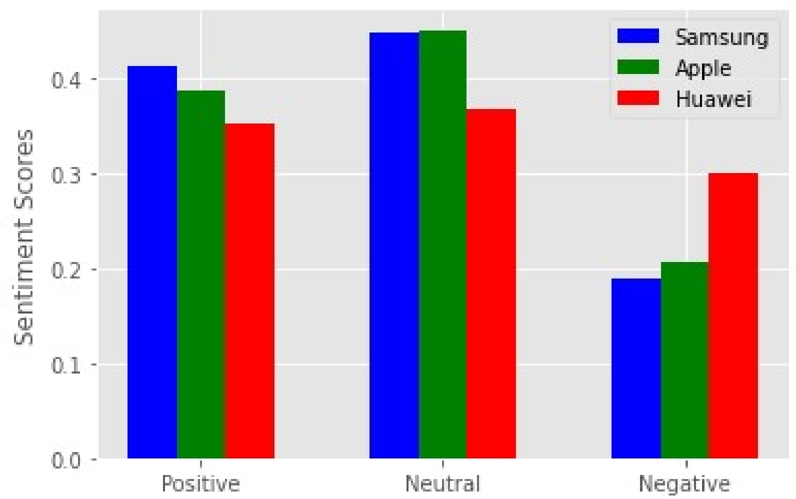
Most Negative Tweets

@Samsung ***why does your device file transfer SUCK SO BAD?? Even @Apple auto exported notes to @gmail. Wtf.*** 😆😆😆

@huvi321 @XcloudTimdog @amirjavadnia ***We need more options when our cell phones die. Apple and Samsung need to get on that shit***

Ich kaufe mir ein iphone aus einem Grund. Weil Samsung in Europa nicht die guten Chips verbaut wie in Amerika sondern die scheisse die dein Akku frisst wie ein Monster. Genus deswegen und dan noch der ueberteuerte preis

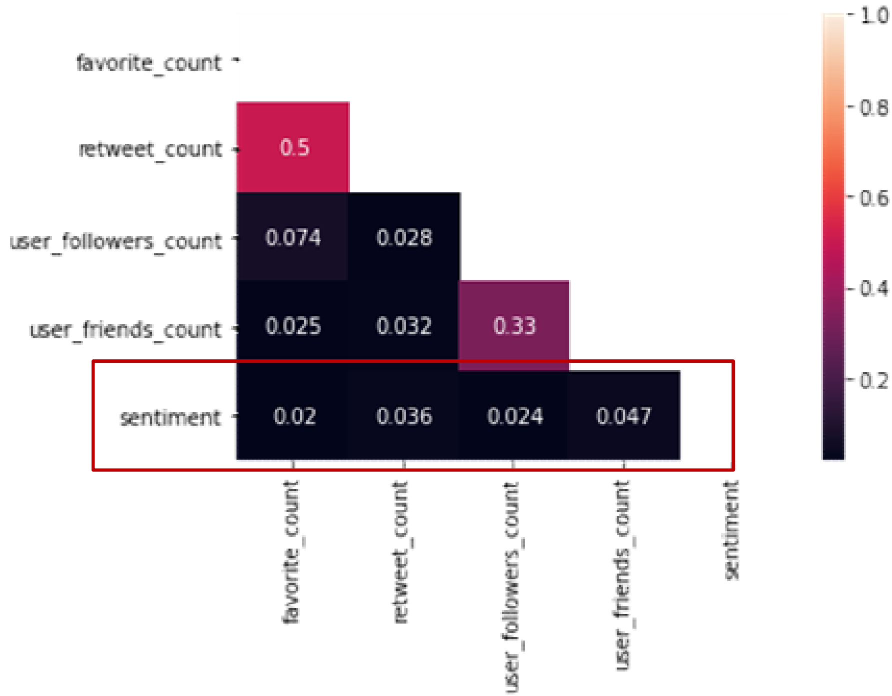
Sentiment analysis on competitors’ tweets



Dataset: Latest 2500 tweets (8-9 Apr 2020) from “Samsung”, “Apple” and “Huawei” respectively

Part 1: Sentiment provides marketing intel

1. Sentiment analysis on tweets provides invaluable marketing intel beyond tweet statistics
2. Sentiment gathered from tweets is not correlated to tweet statistics (number of retweets, favourite counts, etc.)
3. There is no way to tell whether a tweet is positive or negative based on its statistics

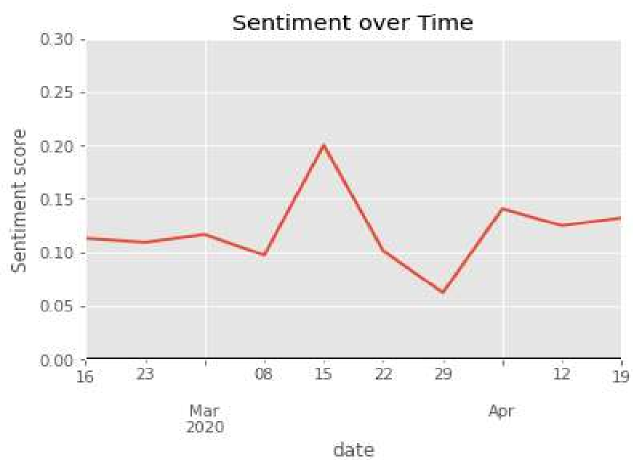


Part 1: Sentiment analysis in advertisements

Dataset: 7331 YouTube comments on Samsung's latest S20 Ultra YouTube video and YouTube comments on 50 YouTube videos from SamsungSingapore YouTube channel

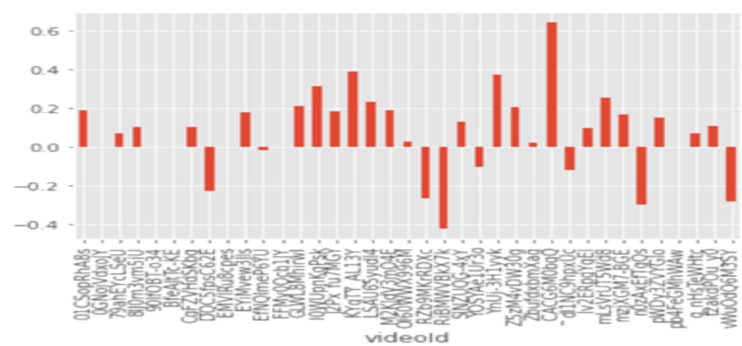
Track the 'freshness' of an advertisement

Sentiment score on Samsung S20 Ultra video (7331 comments)



Compare the popularity of different Samsung YouTube video advertisements

Sentiment score on SamsungSingapore channel's YouTube videos (50 videos)



Part 2: Topic modeling on tweets using LDA

1. Model: sklearn's LDA – most popular for unsupervised analysis of text
2. Performance metric: Perplexity (lower value is better), Likelihood (less negative value is better)
3. Preprocessing: Cleaning and CountVectorizer (compared with TFID vectorizer)
4. Optimised parameters (through gridsearch): Number of topics and learning decay

Brands	Optimised parameters	Perplexity	Log Likelihood	Document-Topic split	Manually assigned topics
SAMSUNG	Number of topics: 2 Learning decay: 0.9	183	-5198	Topic 0: 1554 Topic 1: 884	Samsung phone Samsung phone camera
		157	-3249	Topic 0: 1475 Topic 1: 736	Music app Phone app
HUAWEI		236	-7305	Topic 0: 1533 Topic 1: 871	China communist party US-China

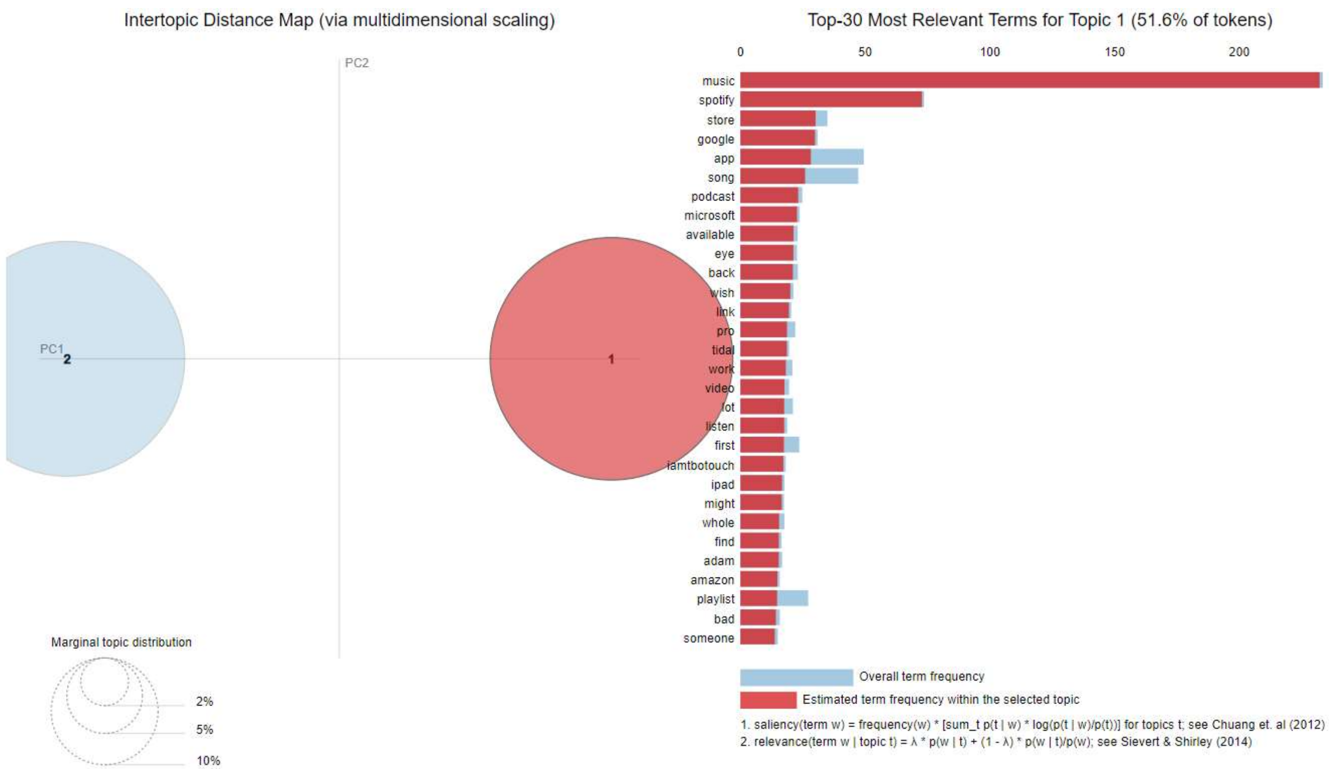
Example on how to derive the manually assigned topics using Apple's top 15 LDA words

Words	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	Topics
Topic 0 (1475)	**music**	**spotify**	**store**	**google**	**app**	**song**	**podcast**	microsoft	available	eye	back	wish	link	pro	tidal	**Music app**
Topic 1 (736)	**iphone**	**phone**	new	people	product	big	best	song	**app**	**android**	help	**buy**	stream	realdonaldtrump	guy	**phone app**

Part 2: LDA visualisation

Python package: Gensim or sklearn’s pyLDAvis

Example of pyLDAvis on Apple tweets



Part 2: Comparing LDA, GSDMM & BTM

Model	Advantages	Disadvantages	Parameters
LDA	• Proven model on traditional documents (news and academic papers) • Well maintained model	• Not suitable for short text as LDA assumes that a text is a mixture of topics	Gridsearch to optimise parameters Number of topics: 2 Learning rate: 0.9
GSDMM	• Modification from LDA that assumes 1 topic ↔1 document	• Less used model, lacking performance metrics • Overfitting • Loses flexibility to capture multiple topics in one document	No of topics=2, alpha=0.1, beta=0.1, n_iters=50
BTM	• Use aggregated document (entire corpus) to address sparsity issue in short text • Incorporate context by using word-pair co-occurrence patterns (biterm)	• Less used model, uses coherence score as performance metrics	No of topics=2, alpha=0.01, beta=1.0, n_iters=100

Major limitation: Lack of common performance metrics for comparison (need to edit source code to add own metrics)

Compare LDA, GSDMM and BTM's word cluster using Apple example

Topic 0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
LDA	music	spotify	store	google	app	song	podcast	microsoft	available	eye	back	wish	link	pro	tidal
GSDMM	music	spotify	tv	song	playlist	people	eye	card	app	google	big	stream	tidal	adam	link
BTM ☺	music	microsoft	help	app	download	store	Iamtbotouch	give	iphone	**tim**	google	podcast	**cook**	pro	itunes

BTM seems to capture more relevant words

[Paper on BTM](http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.402.4032&rep=rep1&type=pdf)
http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.402.4032&rep=rep1&type=pdf

Additional: Samsung stock price prediction based on YouTube sentiments

Dataset: 40 days of Samsung Electronics stock price listed on Frankfurt Exchange & 40 days of comments collected on Samsung S20 Ultra YouTube video

Major Limitations:
- Samsung Electronics has many entities (chips, home appliances), not only mobile phone business
- Stock price can be affected by earning release, covid-19, news
- Insufficient data

Model	Parameters	R2 value
Linear regression	Intercept: 828 Gradient: 283	Train data: 0.04 Test data: -16.6
Logistic Regression	-	Train data: -1.5 Test data: -1.9
Random Forest Regressor	Default	Train data: 0.8 (overfitting) Test data: -12
Random Forest Regressor with gridsearch	'max_depth': 4, 'min_samples_leaf': 3, 'n_estimators': 1000	Train data: 0.5 (less overfitting) Test data: -7

Conclusion and Future work

Marketing insights:

- Based on the most negative tweets picked up by VADER, Samsung could improve on customer support by improving battery, file transfer system, etc.
- Further research on well-received Samsung YouTube advertisement video could help guide marketing effort
- Topic modeling revealed Huawei's relatively low sentiment score could be due to association to China's political system and ongoing US-China trade war

Limitations:

- A lot of work to get topic to make sense – human effort
- Lack of common performance metrics to compare different topic modeling models
- No gold standard on the "right" evaluation metric for topic modeling
- A word may have multiple meanings – e.g. APPLE

Future work:

- Edit GSDMM and BTM's source code to include my own performance metric for comparison

Difference between topic modeling and text classification

Topic modeling	Text classification
Unsupervised learning	Supervised learning
Not mutually-exclusive tokens	Mutually-exclusive classes
Metrics: likelihood of words occurring in specific patterns, relative to topics	Metrics: A variety of different methods (confusion matrix, ROC AUC, vectors)

Note: Clustering can use topic modeling but clustering method is NOT topic modeling.

Application example: In preliminary stage, scientists do topic modeling to gain understanding of the major topics. At a mature stage, they will apply text classification to categorise future information. Topic modeling is more ‘dummy-proof’ and useful when there is a lack of good understand of the subject/topic.

Topic modeling evaluation

There are two measures in topic coherence:
1) Intrinsic measure UMass –measures how much, within the words used to describe a topic, a common word is in average a good predictor for a less common word. Likelihood and perplexity.
2) Extrinsic measure UCI - Pointwise mutual information PMI. Probability of seeing two words co-occurring in a random document