Sentiment Analysis for correlating AT&T Retails Store performance – Empirical Study on Twitter, Yelp and Google reviews.

Aditya Purandare, Swapnil Phulse, Alok Satpathy, Mohammad Atif Tahir

20th October 2016

Introduction

The year 2016 expects certain insurgency around in what capacity can online networking information be expended in an enhanced way to relate with particular business needs. On the off chance that we think in 2015, we watched parcel of concentration on "Sophistication", "Customization" and "Innovation" inside the web-based social networking space, and it will probably observe the improvements around these territories in 2016 too, alongside the extra focus on "Collaboration", "Integration" and "Automation" to give superior personalized products, services and experiences across industries.

It is imperative that renowned corporations like AT&T jump the social media analytics bandwagon. This competition aimed at harnessing Big data to unravel information about customer sentiments as captured in either structured/unstructured way. This report intends to describe our basic understanding through literature-review, our overall solution design and initial insights obtained from exploratory data analysis.

Literature Review

Sentiment analysis and opinion mapping –Katrekar Ashish. 2009

The authors in this paper mention how feelings of others have a critical impact in our daily decision-making process. These choices extend from purchasing an item, for example, an advanced mobile phone to making speculations to picking a school—all choices that influence different parts of our day by day life. Before the Web, individuals would look for feelings on items and administrations from sources, for example, companions, relatives, or shopper reports. In any case, in the Internet era, it is much simpler to gather assorted assessments from various individuals around the globe. Individuals hope to audit destinations (e.g., CNET, Epinions.com), e-business destinations (e.g., Amazon, eBay), online conclusion locales (e.g., TripAdvisor, Rotten Tomatoes, Yelp) and social media (e.g., Facebook, Twitter) to get criticism on how a specific item or administration might be seen in the advertise.

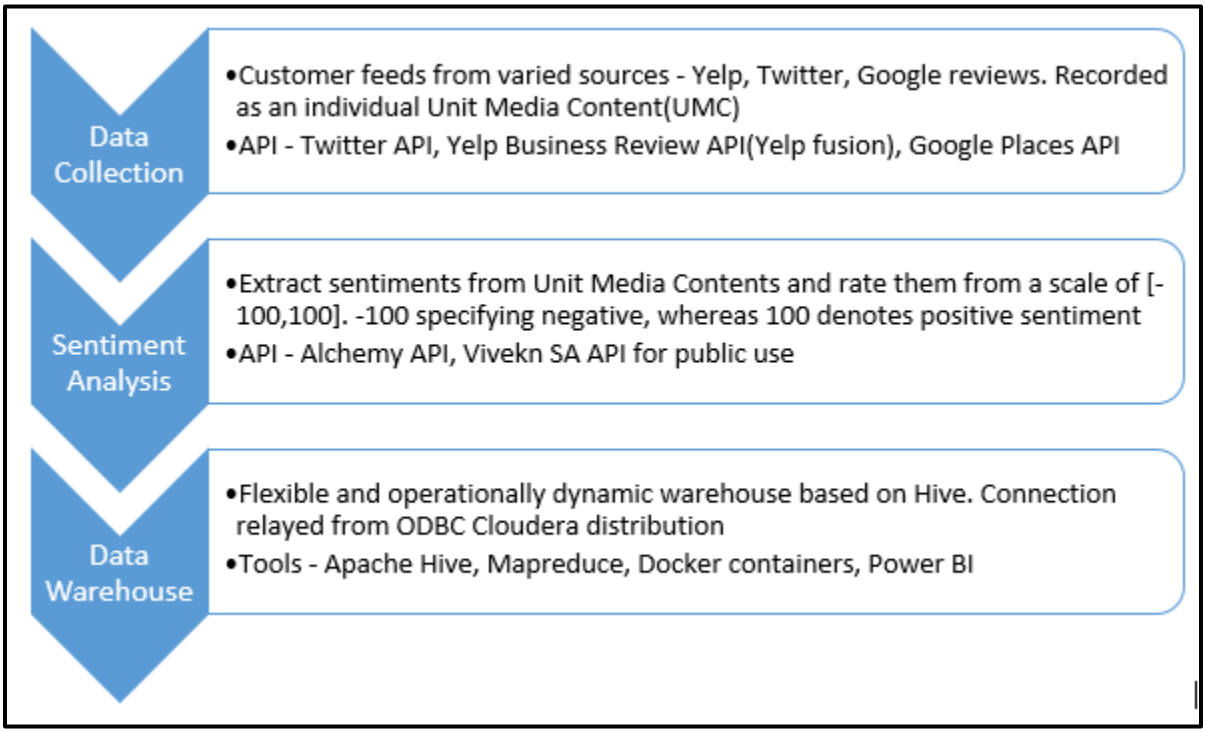
Big Data Stream Analytics for Near Real-Time Sentiment Analysis - Otto K. M. Cheng, Raymond Lau

The main theoretical contributions of our research include the design and development of a novel big data stream analytics framework, named BDSASA for the near real-time analysis of consumer sentiments. Another main contribution of this paper is the illustration of a probabilistic inferential language model for analyzing the sentiments embedded in an evolving big data stream generated from online social media. The business implication of the author's research is that business managers and product designers can apply the proposed big data stream analytics framework to more effectively analyze and predict consumers' preferences about products and services. Accordingly, they can take proactive business strategies to streamline the marketing or product design operations.

Business problem

In a worldwide commercial center, where consistent development and client contact is fundamental, AT&T must explore the scene of customary call focuses, retail location collaborations, and now, online presence. As mentioned in case, 500 million tweets are posted in the Twitter universe daily. These tweets go from associations between companions to purchaser protestations. True to its mission " connect people with their world, everywhere they live, work and play … and do it better than anyone else ", AT&T has decided to use data over the plethora of online websites.

Overall architecture and solution design



Data collection

By using API collectors intended to gather data from varied sources, in its native form, we have gathered raw tweets, yelp and google reviews. It helped us capture general sentiments associated with this excerpts that capture customer response.

Big data Sentiment analysis

This, we believe to be the most crucial stage, lying at the heart of the application to derive sentiments through the use of API and Hive processing. Capturing sentiment scores from reviews and rendering a relational database-like appearance through HQL, it should be able to provide answers to most social media presence related issues.

Data warehouse

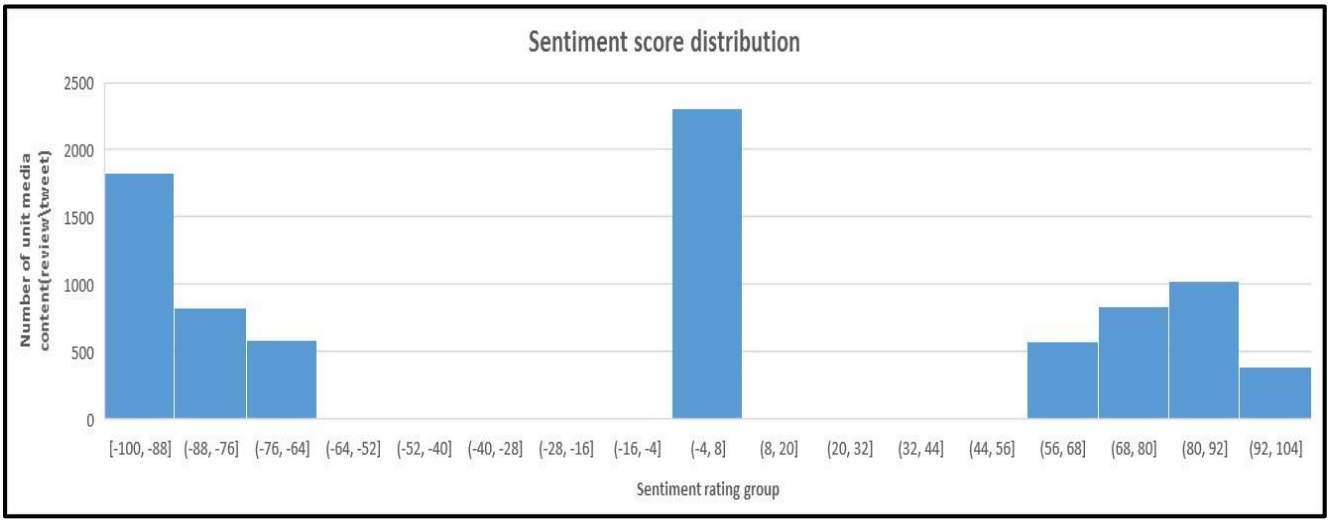
One of the biggest challenges in this project has been data consolidation. We figured a better way enabling dynamicity and operational flexibility (treating this to be a real-world project, which it is) would be through a data-warehouse. A dedicated ETL layer, built with a future vision would be an added advantage over an ad-hoc database solution.

Data exploration and preliminary insights

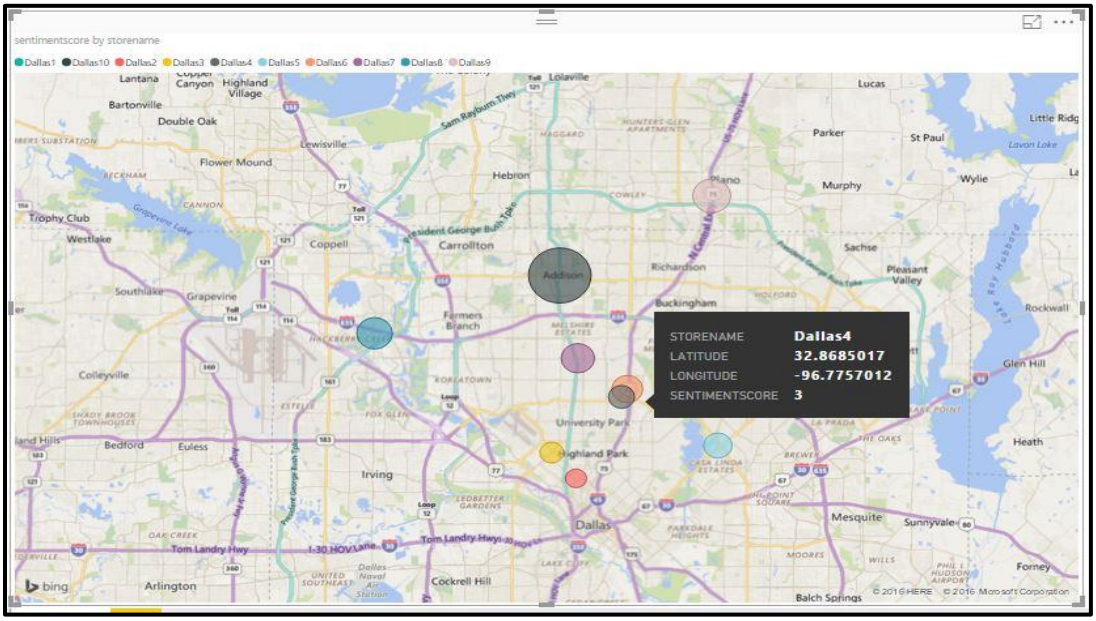
This exploratory analysis is based on a random sample of 3000 UMC (unit media content – our case observation

Sentiment distribution

As can be seen from figure below, we found that sentiment ratings were mostly polarized. If you look at the figure below showing an absolute frequency histogram for sentiment ratings, most of those concentrated towards the tails/ends of the graph with a relatively fewer neutrals. Please note that cumulative counts for positive or negatively rated sentiments in a range of -100 to 100. for analysis).



Location wise scoring – The following graph shows overall sentiment scores generated in the 10 AT&T stores in and around Dallas. The ones with bigger size are major stores to concentrate on for performance wrt customer satisfaction



Word Cloud Creation

From the word cloud created from recorded UMC, we found the following dominant sentiments/expressions. The ones with maximum concerns are as follows

- Issue
- Service
- Still
- Time
- Since

These are more often than not indicators of customer dis-satisfaction and should be vigilantly monitored in real time.

Additional thoughts

If we were given more time, the correlation of following items would be our focus of interest

Aging/Depreciation factor

We found that a particular product gets lesser reviews as it grows old in the market and is impacted by other competitive products. A predictor input that can capture this depreciation of interest would be really helpful.

Capturing triggers

Within events like protests, strikes, elections, a particular marketing campaign or special occasions (Christmas, Thanksgiving, NBA season finals etc.) tend to have an impact on the number of tweets. The gradient of its influence on our outcome would be interesting to explore

Location-specific preference

Certain tech-hubs/ sophisticated locations tend to get much more network traffic of value in this case. It is customary to factor in location favorite indexes where AT&T team would like to focus their efforts on.

Deep intuit

There is plethora of techniques developed that can help us dig deeper. Continuous investigation of the services on offer, and reviews/remarks posted about it, could help the association refine the message and how it's conveyed to enhance the reaction rate and the online input. Investigation of that sort sustains into a more intricate layer: taking a gander at how diverse

members in online networking groups communicate with each other. Doing as such can distinguish compelling individuals inside a group - for instance, Twitters mentioning blog links – that generate content on another website and how one influences conclusions of users on that website. Such individuals could then be singled out for nearer observation of specific patterns that AT&T should monitor, either to fortify positive remarks or to answer negative ones.