The Proposal Investigating Dynamic Relationships between Customer Sentiments in Social Media and Retail Store Performance at AT&T

Project Members: Guopeng Yin & Hongpeng Jin

The University of Texas at Dallas, Jindal School of Management

Abstract:

We think there are dynamic relationships of customer sentiments in social media and retail store performance. Therefore, we design a longitudinal study to collect social media or retail store data and employ relevant econometric models to deal with time series data. Also, our proposal has two unique features. One is to consider the inequality of different social media message, and find the influential messages; then, we think about how to aggregate the attributes of an individual message into the metrics of each retail store. To provide more reasonable suggestions, we plan to cluster retail stores in term of a few metrics rather than a single indicator. At last, we believe that the AT&T should also pay close attention to the sentiments in social media of its competitors.

1. Data Collecting Plan

Most of social media platforms provide APIs to help developers access their data or build applications. Therefore, we can develop a few software tools targeting for different platforms, and thus collect the social media data regarding the AT&T stores in Dallas.

Due to the time limit, we do not present the technical details about data collecting process. However, we put our efforts on finding required fields (variables) for our analytical goal. The main parts we will collect from the social media platform are listed as follows (Table 1). Because of differences across social media platforms, we only approximately describe core variables we need to capture. Here, we still want to focus on four mainstream social media platforms: Yelp, Google reviews, Twitter and Foursquare, all of that provide geo-location information more or less. Considering the dynamic relationship between social media and retail store performance, we will collect the relevant data every data in a six-month period.

Table 1 Key Variables in Social Media Content

Main Parts	Key Variables (differ across Platforms)
Basic Infos	Posting Time
	Text
	Numerical rating (if available)
Influential Capability	Numbers of retweet/forward/comments/likes/useful votes
	Retweet/forward time lags
User Characteristics	UserID
	Numbers of followers/friends
	Numbers of tweets/reviews
Geo-location	Available Structural Geo-location Info (tweets, store address, check in, been here times and etc.)
	Use software or develop program to extract info from the social media content

2. Analysis Customer Sentiments and Identify Influential Messages

The previous studies have indicated all the consumer's messages at social media are not equal in term of the influential power on consuming decisions. So, we need to find what customer sentiments make an influential message, which is the uniqueness and value of our proposal. In this first phase, this unit of our analysis is the individual message.

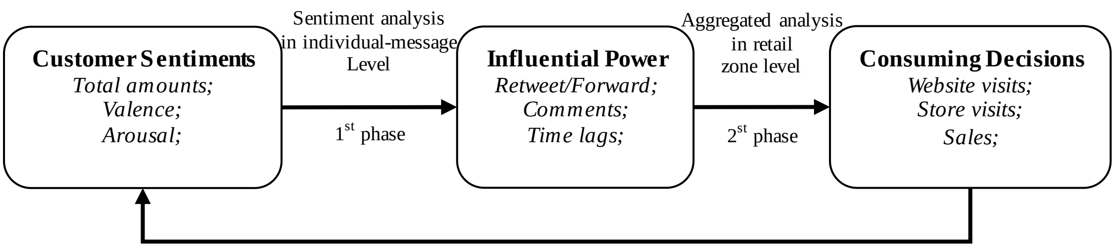

Fig.1 The Process of Our Analysis

2.1 Investigate Customer Sentiments using the Circumplex Model of Emotion

To identify customer sentiments within the social media content, we use the circumplex model of emotion developed by James Russell as the framework. As the figure 1 shows, this model suggests that emotions are distributed in a two-dimensional circular space, containing arousal and valence dimensions (Russell 2003).

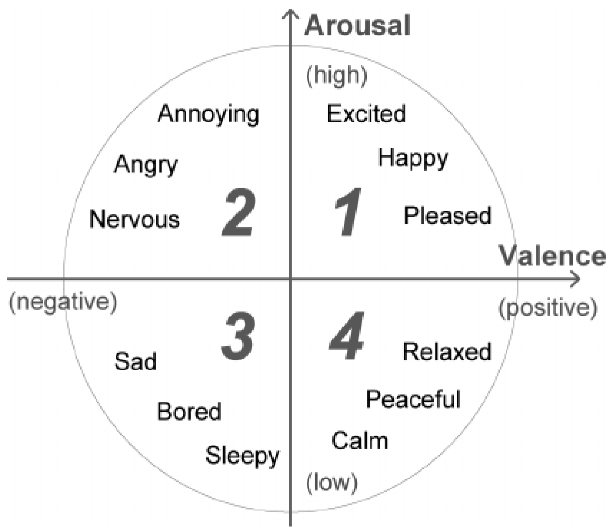

Fig.2 The Circumplex Model of Emotion

Valence describes the extent to which an individual perceives an experience as pleasant or unpleasant and is used to distinguish 'positive' and 'negative' affective states. Arousal describes the extent to which an individual is energized or activated by an experience. Arousal has been shown to be a driver of information sharing (Stieglitz and Xuan 2013). The social media content that evokes high-arousal, positive or negative emotions is more viral. Conversely, content that evokes low-arousal, or deactivating, emotions (e.g., sadness) is less viral.

Specifically, the Revised Dictionary of Affect in Language (RDAL - Whissell 2009), a popular text analysis software, can help us measure emotional valence and arousal in social media content. And, some studies on emotional content in social media have provided substantial evidence supports for the RDAL's reliability and validity (Yin, Bond and Zhang 2016).

2.2 Identify Influential Customer Messages and Develop an Aggregated Metric

After computing the scores for every social media message, we will utilize econometric analysis to model the relationship between customer sentiments and the influential power. This analysis aims to verify the unique role of customer sentiments in shaping the influential messages in social media.

We use total sentiment amounts, emotional valence, and arousal as the focused independent variables, and consider the characteristics of a user who posted this message. Particularly, we are interested in the total amount of sentiments of social media messages regardless of their valence (i.e., positive or negative). It's especially important when a message of social media contain both positive and negative sentiment words.

Furthermore, we can employ the principal component analysis (or other methods) to develop a metric to measure the influential power of social media message, which we call the Message Influence Effect (MIE) metric.

3. Aggregated Analysis of Influential Messages on Every Retail Zone

At last, we focus on the effect of social media on the performance of retail zone. To do this analysis, we need to aggregate the indicators of an individual message into the metrics of retail zone first. In Table 2, we briefly list the corresponding relations between them.

Table 2 Corresponding relations between individual-message and retain zone level

Aggregated Metrics for Every Retail Zone	Computing based on the individual message of every retail zone
Volume	Total number of influential messages for a retail zone
Average Influential Power	Average of Message Influence Effect (MIE) for a retail zone
Average Valence	Average of valence of influential message
Average Arousal	Average of arousal of influential message
Variance of Valence	*Variance of valence of influential message*
Variance of Valence	*Variance of Arousal of influential message*

Then, we model the relationship between aggregated metrics of social media content and the performance of retail zone. Also, we choose three indicators for every retail zone:

web traffic, store visits and sales. On the other hand, to account for dynamic relationships of social media sentiments and the retail zone performance, we collect and organize time series data for each retail zone every week.

In our opinion, we will not rank the retail zones based on a single metrics. We prepare to cluster retail zones based on these metrics and clustering analysis techniques, and then provide corresponding suggestions for different retail zones.

4. Further Analysis along with Other Competitors

To capture the correlation between social media and retail store performance, we also need to pay close attention to the sentiments in social media of the AT&T's competitors. In this way, we can get the whole picture on this problem.

Reference

Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological review*, 110(1), 145.

Whissell, C. (2009). Using the revised dictionary of affect in language to quantify the emotional undertones of samples of natural language. *Psychological reports*, 105(2), 509-521.

Dezhi Yin, Samuel D. Bond, and Han Zhang Keep Your Cool or Let It Out: Nonlinear Effects of Expressed Arousal on Perceptions of Consumer Reviews. *Journal of Marketing Research* In-Press.

Stieglitz, S., & Dang-Xuan, L. (2013). Emotions and information diffusion in social media—sentiment of microblogs and sharing behavior. *Journal of Management Information Systems*, 29(4), 217-248.

Guopeng Yin

Hongpeng Jin