

Sentiment Analysis Symposium

March 5, 2014

Sequence Package Analysis & Sentimetrics

Amy Neustein, Ph.D.

Founder and CEO

Linguistic Technology Systems (LTS)

www.lingtechsys.com

Sequence Package Analysis – a new natural language method for performing Social Analytic Text Mining

WHY DO WE NEED A NEW NATURAL LANGUAGE TEXT MINING METHOD FOR SOCIAL MEDIA?

- ▶ Because Tweets and Online Customer Reviews are NOT monolithic
- ▶ Because Social Media Posts Contain Contradictions, Mixed-Ratings, Digressions, and Rambling
- ▶ Because Social Media displays inextricable psycho-social dynamics (e.g., bragging, seeking assurance, or trying to convince oneself about a purchase that was already made)

A Corpus of Annotated Training Data to Measure Dense,* Complex and Convoluted Social Media Content

Sequence Package Analysis (SPA) uses a corpus of annotated training data comprised of a unique set of feature extractions (drawn from an SPA-designed BNF table consisting of sequentially-implicative, as opposed to syntactic, parsing structures) that comport with the wide range of emotions found in social media posts.

*I borrowed this term from Seth Grimes' fascinating article "Social Sentiment's Missing Measures" (*Social Media Explorer*, 1/29/14) to refer to those online reviews w/ mixed messages in which a reviewer simultaneously offers a positive rating on some aspect of a product/service but a negative ratings on other aspects.

Social Action – not Syntax or Semantics – the focus of Sequence Package Analysis

Starting Point:

SPA in building its corpus of annotated training data begins with *Social Action*, rather than Syntactic POS (Parts of Speech) or Semantics Tags (named entities consisting of predefined categories such as the names of persons, organizations and locations, or expressions of time, quantities or monetary currencies)

Result:

The computer, in performing pattern recognition of sequence packages, would be able to recognize social action in online reviews arguably in the same way as humans do:

e.g., detecting a subtle complaint in a positively-rated product/service review, or distinguishing between a review that is more of a boast about one's purchase decision than actual praise of the product or brand itself

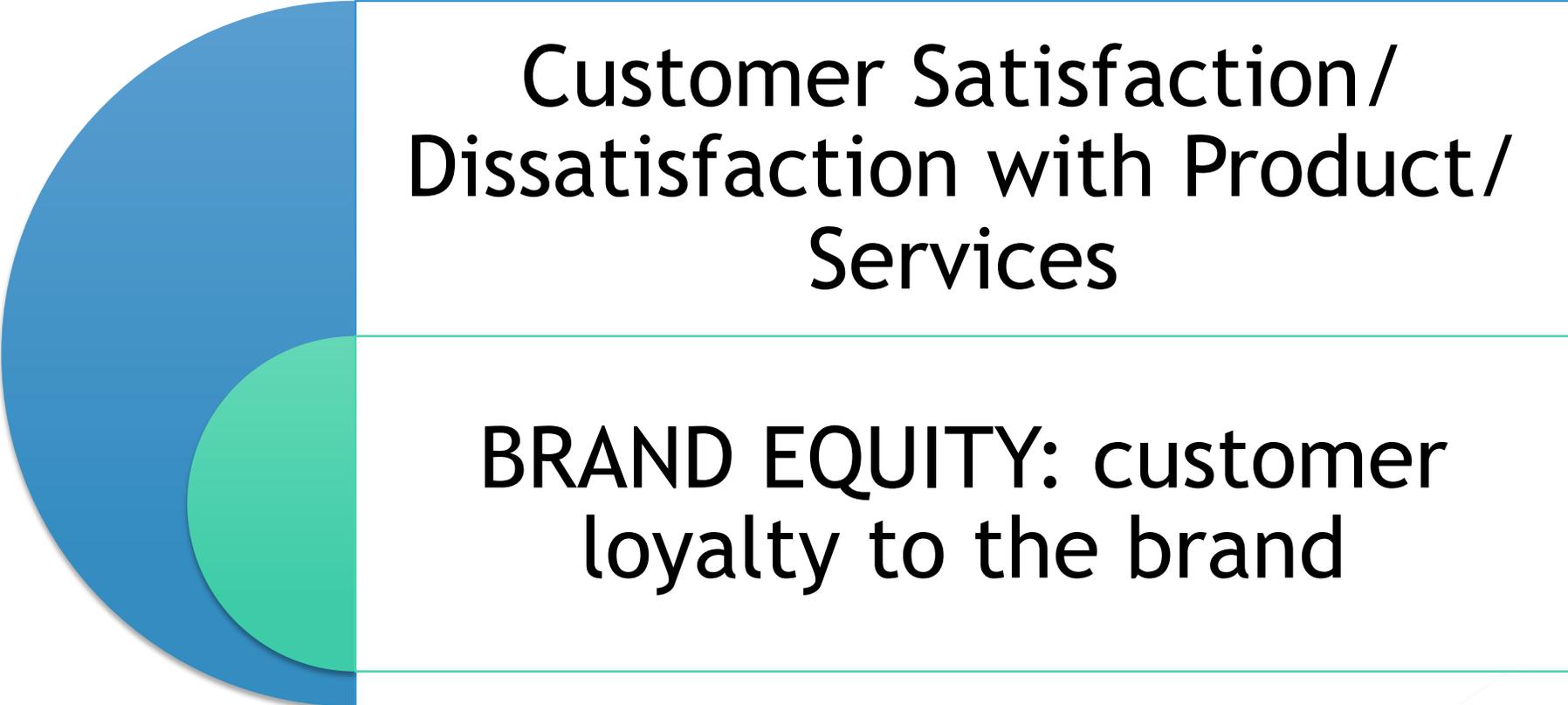
SPA as a Machine Learning Technique for Performing Sentimetrics

Training the Computer to Find the Sentiment Index Score (SIS) in Social Media Posts is a Two-Step Process:

Step One: Identify the Sequence Package in Unstructured Data

Step Two: Add up the Various Sequence Packages to Assign an SIS to the Post by Adding or Subtracting from Reviewer's Own Star Rating

Variables of Online Reviews Measured by SPA-Designed Sentimetrics



Customer Satisfaction/
Dissatisfaction with Product/
Services

BRAND EQUITY: customer
loyalty to the brand

Detecting Subtle Complaints in a 5-star Review

http://www.tripadvisor.com/ShowUserReviews-g33020-d342103-r137943344-Falafel_Drive_In-San_Jose_California.html

“Falafel Drive-In” (San Jose, CA), posted August 4, 2012

Review Heading: “Excellent Falafels and Shakes” (5-star rating)

<Opening Endorsement> *“I’ve been here 4-5 times at least and I never leave disappointed”*

<Opinion Review> *“The falafel is excellent. I always ask for a side of their hot sauce because it’s that good! The falafel combo deal is great because it is cheap and it comes with their fantastic banana shake! The banana shake is the best I’ve ever had!”*

Four Pairs of Subtle Complaints Inserted between Opening and Closing of Online Review

<Closing Endorsement> *“This place is a must if you leave [sic] in San Jose! Excellent, just excellent!”*

Sequence Package Design of Inserted Complaints in Five-Star Review Posting

Four Complaint/Disclaimer Contrastive Pairs:

<Complaint/Disclaimer (Parking)> “Parking can be tough during the lunch crowd but it is totally worth it”

<Complaint/Disclaimer (Waiting)> “There is typically a line - a good sign in my opinion!”

<Complaint/Disclaimer (Seating)> “They have small indoor seating but tons of outdoor seating”

<Complaint/Disclaimer (Payment)> “They do not accept credit cards, only debit and cash so come prepared”

Features of Complaint/Disclaimer Contrastive Pairs

- ▶ First Pair Part: <Complaint>
- ▶ Second Pair Part: <Disclaimer> = <concessive connector(a/k/a contrast marker)> + <idiom/metaphor>
- ▶ In this example, the second part of each of the four contrastive pairs begins with a **concessive connector** (a/k/a **contrast marker**): “but”, “so,” or “n-dash,” followed by an **idiomatic expression or metaphor**.
- ▶ Note: Here idioms are defined rather broadly to include, along with idiomatic expressions, a set of platitudes and clichés that serve as a “shorthand” way of getting the message across.

Detecting Subtle Complaints in Online Posts by Annotating the Complaint Resolution Features

<Complaint/Disclaimer (Parking)> “Parking can be tough during the lunch crowd **but it is totally worth it**”

<Complaint/Disclaimer (Waiting)> “There is typically a line - **a good sign in my opinion!**”

<Complaint/Disclaimer (Seating)> “They have small indoor seating **but tons of outdoor seating**”

<Complaint/Disclaimer (Payment)> “They do not accept credit cards, only debit and cash **so come prepared**”

The SPA-based feature extraction method uncovers the gravamen of the complaint (**parking, waiting, seating, and payments issues**) which was NOT known ahead of time because these attributes are not included in the TripAdvisor rating summary indicators: food, service, value and atmosphere.

Assigning Sentiment Index Score (SIS) to Restaurant Review Post

Starting Point:

- Reviewer's own star rating (5-star)
- Total Number of Reviewer Endorsements and Favorable Opinions Reviews (3) (Second variable used only as a control measure to alter star rating if needed.)



Reduce Star Rating by Complaints:

- One complaint = lower by 1 star
- Two or more complaints = lower by 2 stars
- Four or more complaints = lower by 3 stars



Net Result:

The eateries' 5-star consumer rating is assigned an SIS of 2 stars

Discerning between customer brand loyalty and customer's need to boast about purchase decision

<http://chelz22.livejournal.com/52380.html> (Posted June 16, 2006)

As of tonight, I am the proud owner of a 2003 Honda Civic EX. It is black, has two doors, and a power sun roof. It is pretty much amazing, especially as this is a thirteen year upgrade from my 1990 Toyota Camry. I can hardly believe it's mine. Which I guess it isn't- it pretty much belongs to the bank. But they're letting me drive it.

This car has so many little things that many drivers take for granted, but are amazing to me. Like a radio that works. And air conditioning. Windows that go down. And that little clicky thing to unlock it.

If gas were cheaper I would be cruising 24/7.

I hope everyone is having a night like mine! :)

Annotations of Self-Congratulation in Online Car Review

- ▶ <“As of tonight, I am the proud owner of a 2003 Honda Civic EX”>
- ▶ <“I can hardly believe it's mine”>
- ▶ <“This car has so many little things that many drivers take for granted, but are amazing to me”>
- ▶ <“If gas were cheaper I would be cruising 24/7”>
- ▶ <“I hope everyone is having a night like mine! :)”>

Annotations of Product Descriptions in Online Car Review

- ▶ <“It is black₁, has two doors₁, and a power sun roof₁”>
- ▶ <“Like a radio that works₁. And air conditioning₁. Windows that go down₁. And that little clicky thing to unlock it₁”>

A subscript of “1” (on a scale of 1-5) is designated for each product feature contained in this review because these product features are described as fairly standard and not superlative.

How Does the SPA-Designed Mining Program Measure Sentiment in this Car Review Post?

The SPA-designed sentiment analysis program would mine the unstructured text of this car review posting along two axes:

- ▶ Axis One: Product Descriptions
- ▶ Axis Two: Mitigating Factors (e.g., self-congratulatory boasts made after purchase decision)

Brand Loyalty Analysis:

Five self-congratulatory sequences are inserted in a review that modestly highlights standard/average product features: power sun roof and car remote central lock; radio, air conditioning and windows in serviceable condition. If self congratulatory content were set aside, would this consumer show brand loyalty to Honda? Perhaps not? But let's do the math on next slide to be certain.

Assigning Sentiment Index Score (SIS) to Car Review Post

Starting Point:
Product Rating defined as the total numeric value of all subscripts designated for product descriptions. There were 7 product features, each with a subscript of "1". Total: 7



Reduce Product Rating by Self-Congratulatory Statements:

- One boast = lower by 1 point
- Two or more boasts = lower by 2 points
- Four or more boasts = lower by 3 points



Net Result:
Honda review is given an SIS of 4 on a scale of 1-10