



Mark Turner
Sr. Analytic Data Scientist

Mining Text to Pinpoint Customer Reactions to Products

Mining Online Customer Comments

Online customer comments can tell us something about:

- Customer priorities
- The terms that contribute to positive and negative ratings

Examples

Good:

i've always had pretty good skin and i love love love this moisturizer!!! it isn't too heavy yet delivers moisture like you wouldn't believe. my skin is beautiful and flawless. i never use foundation, never!!!! and i get compliments all the time, my skin glows. . .

Examples

Not So Good:

This is my 3rd jar of this cream and I am going to return it ,it smells like a nasty wet dog,all your products do not smell the way they used too,no perfume smell at all and I have just about all of them. I don't like the lotion either because it don't have a good clean smell . . .

Some Features of Customer Comments

- “Informal reporting”
- Generally first-person with the product and its features as focus, but. . . plenty of cross-talk directed at other reviewers (third person)
- Major elements:
 - Narrative (“I was eager to try this product. . .”)
 - Experience (“ . . . and it was a great feeling.”)
 - Recommendation (“It’s great for anyone with dry skin.”)
 - **Sentiment** (“I was just so pleased.”)

Goals

- Find association of terms with **ratings**: positive and negative
- Find **association** of terms with other terms, including product features, emotions, actions, . . .

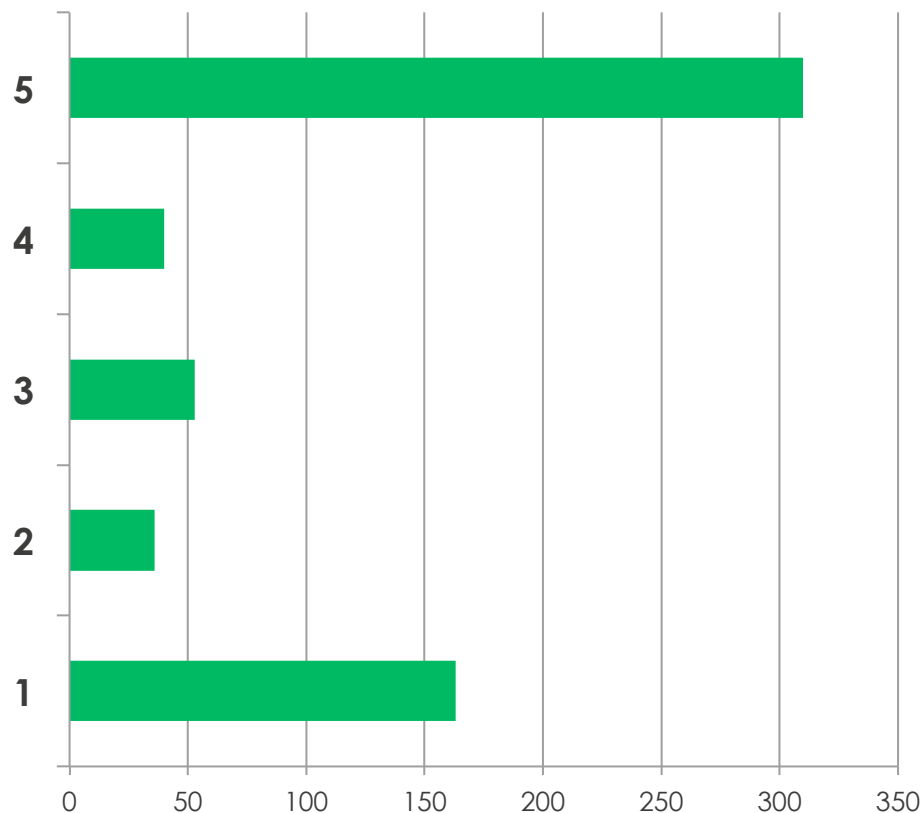
Approach

- Start with the sentiment metric that the customers give us: the star rating of ★ through ★ ★ ★ ★ ★
- We want to find out what text elements in the comment **are associated with** the star rating.

Data

- Online customer comments from a large TV and Internet retailer
- Product: moisturizer
- We chose the product because it had both positive *and* negative ratings over a long period: 8 years.
- The average review was 59 words.
- About 16,000 unique words were used in the reviews.

Responses



Technical Approach

- Used all the words, including misspellings.
- Generated “n-grams”: all sequences of 1 or two words.
 - This captures many of the multi-word phrases.
- Did not stem the words:
 - Worth preserving differences between word forms to study association between words.
 - “I was **disappointed** in. . .” / “This **disappointing** product”
- Ran on Aster parallel (“big data”) parallel processing platform.

Terms as Sentiment Indicators

Which terms go with good reviews? with bad reviews?

1. Find and rank the most frequent terms in both good and bad reviews.
2. Eliminate the terms that occur in both.

Top Terms for Rating of 1 – 3 (Worst)

this

my

smell

product

not

have

but

<product line>

face

this product

like

just

i have

jar

the smell

Top Terms for Rating of 4 – 5 (Best)

this

my

skin

have

my skin

but

product

not

hope

jar

love

i have

using

use

like

Top Terms for Rating of 1 – 3 (Worst)

this
my
smell
product
not
have
but
<product line>
face
this product
like
just
i have
jar
the smell

Top Terms for Rating of 4 – 5 (Best)

this
my
skin
have
my skin
but
product
not
hope
jar
love
i have
using
use
like



**Top Differentiating Terms
for Rating of 4 - 5 (Best)**

skin

my skin

not

hope

love

using

use

**Top Differentiating Terms
for Rating of 1 - 3 (Worst)**

smell

not

<product line>

face

this product

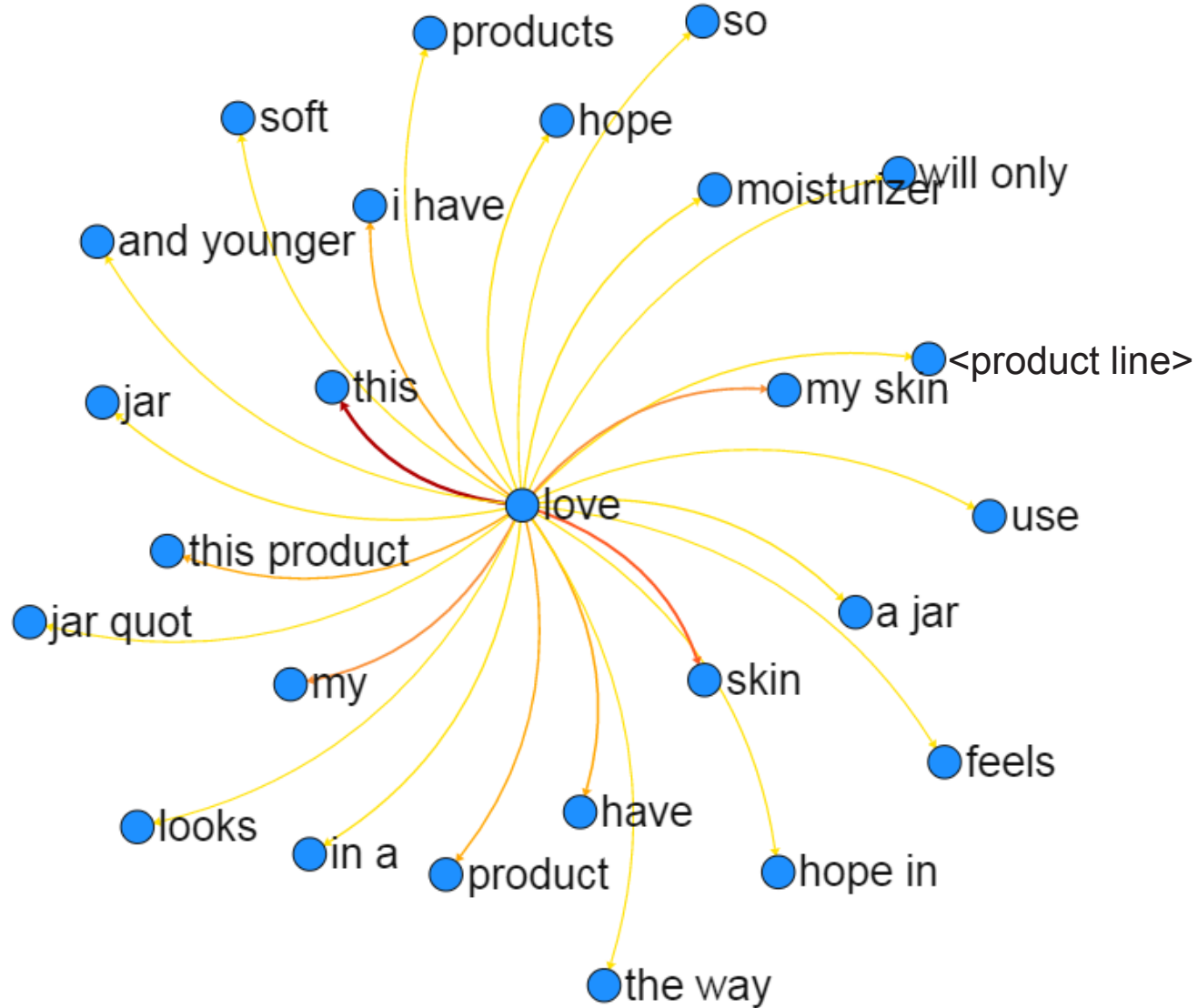
just

the smell

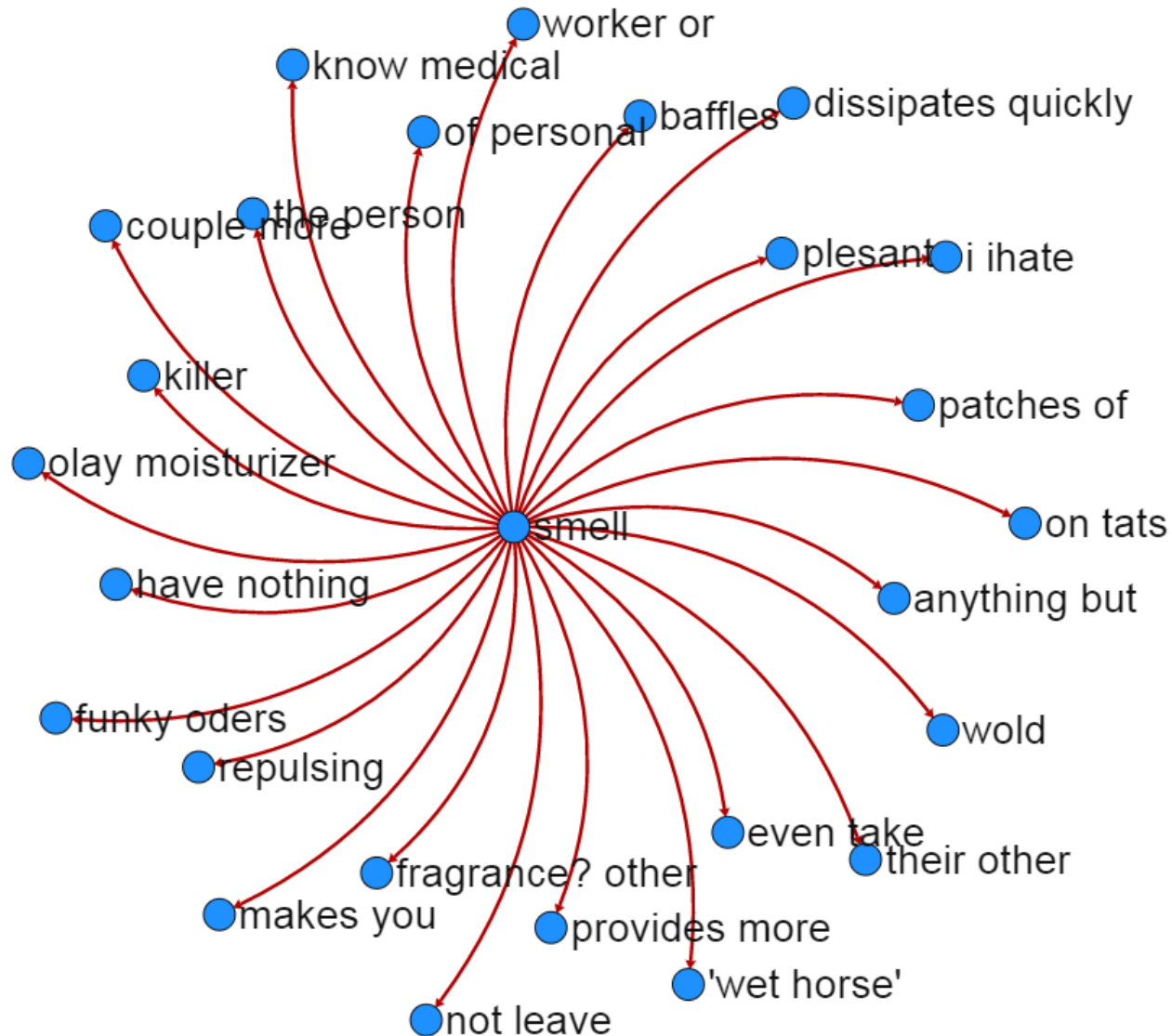
Finding Association of Terms

- We are interested in exploring what terms are **associated with** each other in the reviews.
- In text analytics, **association = co-occurrence**
 - Scope: occurring in the same review
- Approach:
 - Use same set of n-grams (n=1, 2)
 - Find association of **all pairs** with collaborative filter
 - This is a large combinatoric problem (N^2): 4.4 billion combinations
 - Visualize in 2D

Association of Terms with 'Love'



Association of Terms with 'Smell'



Conclusion

- Combining text analytics with product ratings gives insight into product reviews.
- We can see the product features and characteristics contributing to favorable and unfavorable reviews.
- We can explore how concepts are associated in the reviews.