



# Scoring Dimensions of User Reviews

VILANKAR | SAVLA

# Motivation

## Problem

- Ratings are not sufficient
- Too many reviews

## Solution

- Summarize reviews by identifying and scoring features.  
Eg: Features for restaurants - food ,quality, service, ambience
- Support search based on preferences  
Eg: Student prefers worthiness opposed to ambience
- Predict the rating of the restaurant for that user

## Uniqueness

- Crowd based feature identification
- Determining opinion strength instead of opinion orientation

# Visualization

Salient features about Green:



place: Opinion Count - 308 | Sentiment - 6.92

# Algorithm - Feature Scoring

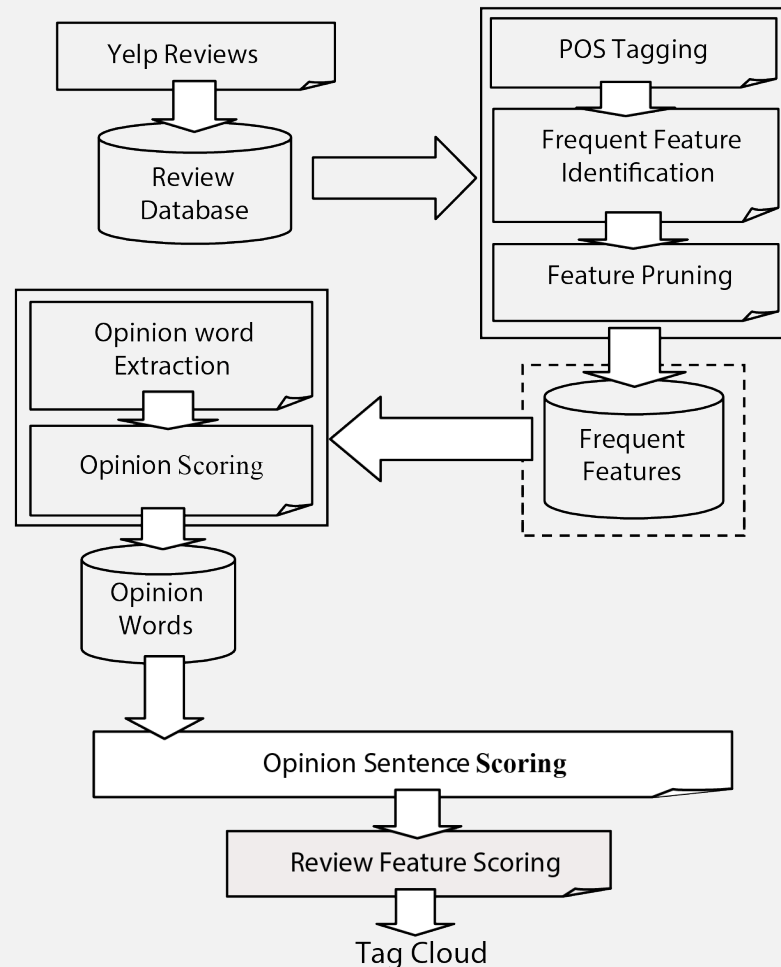
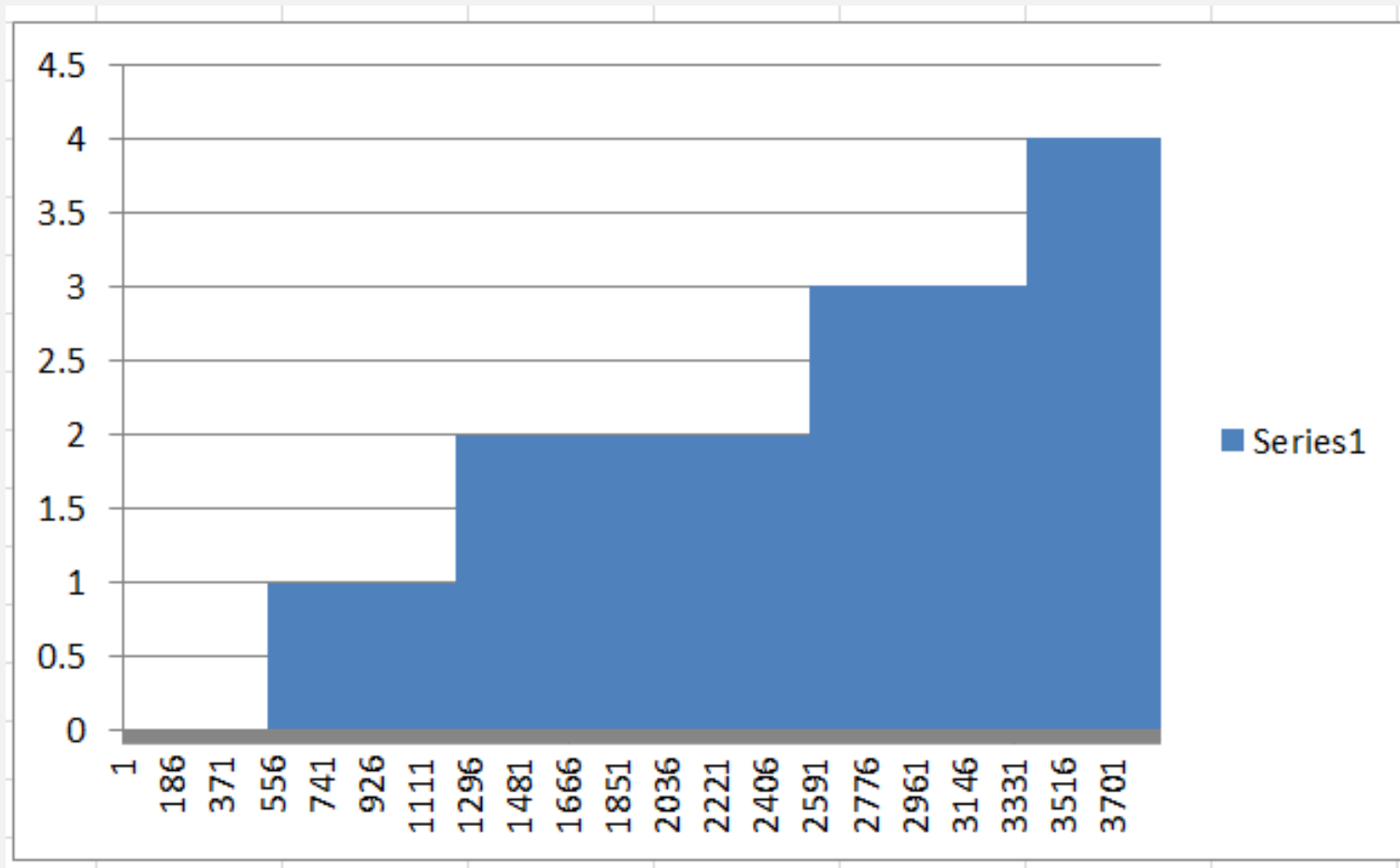


Figure: Feature-based opinion summarization

# Data

- Yelp Dataset from the greater Phoenix, AZ metropolitan area including:
  - 15,585 businesses
  - 70,817 users
  - 335,022 reviews
- Total Size - 353 MB JSON file
- Total Restaurants - 5556
- Top 50 Restaurants with total 26,746 reviews
- 535 Average Reviews Per Restaurant

# Results - Sentiment Dictionary



# Comparison

Google products

sony camera

Search Products

## Sony Cyber-shot DSC-W370 14.1 MP Digital Camera (Silver)

[Overview](#) - [Online stores](#) - [Nearby stores](#) - [Reviews](#) - [Technical specifications](#) - [Similar items](#) - [Accessories](#)



**\$140 [online](#), \$170 [nearby](#)**



159 reviews



### Reviews

Summary - Based on 159 reviews



### What people are saying

<a href="#">pictures</a>	<div><div></div><div></div></div>	"We use the product to take quickly photos."
<a href="#">features</a>	<div><div></div><div></div></div>	"Impressive panoramic feature."
<a href="#">zoom/lens</a>	<div><div></div><div></div></div>	"It also record better and focus better on sunny days."
<a href="#">design</a>	<div><div></div><div></div></div>	"It has the slightest grip but it's sufficient."
<a href="#">video</a>	<div><div></div><div></div></div>	"Video zoom is choppy."
<a href="#">battery life</a>	<div><div></div><div></div></div>	"Even better, the battery lasts long."
<a href="#">screen</a>	<div><div></div><div></div></div>	"I Love the Sony's 3" screen which I really wanted."



# Comparison



HP printer

ALL RESULTS

Shopping

POPULAR FEATURES

all

Affordability

Speed

Print Quality

Reliability

Ease Of Use

Brand

Installation

Size

Compatibility

SHOPPING

HP LaserJet 1020 - printer - B/W - laser, 15ppm, USB



from \$179 (2 stores)  Bing cashback - 3%


★★★★☆ user reviews (177)

The HP LaserJet 1020 Printer, an excellent laser printer for the cost-co high-quality LaserJet printing in a compact size, and at a price you can

user reviews | product details | expert reviews | compare prices

user reviews

view: positive comments (44)

speed  96%

The quality is as good as any laserjet printer I've used and the speed is fast.  
Love Reading [www.amazon.com](http://www.amazon.com) 3/17/2006 [more...](#)

Quick and fast transaction.  
Arthur L. Taylor [www.amazon.com](http://www.amazon.com) 2/5/2008 [more...](#)

It's small and fast and very reliable.  
Muffinhead's mom [www.amazon.com](http://www.amazon.com) 1/9/2007 [more...](#)



# Algorithm - Rating Prediction

- Latent Dirichlet Allocation (LDA)
  - Discover hidden dimensions
  - Dimensions are not correlated with ratings
  - Supervised LDA combines both ratings and review
- Matrix Factorization Models
  - Alternating Least Squares
  - Stochastic Gradient Descent
  - Bias-Stochastic Gradient Descent
- Hidden Factors and Topics (HFT)
  - Associate latent rating dimensions with latent review topics
  - Use - Discover features in reviews

$$f(\mathcal{T}|\Theta, \Phi, \kappa, z) = \sum_{r_{u,i} \in \mathcal{T}} \underbrace{(rec(u,i) - r_{u,i})^2}_{\text{rating error}} - \underbrace{\mu l(\mathcal{T}|\theta, \phi, z)}_{\text{corpus likelihood}}.$$

# Data

For Top 50 restaurants

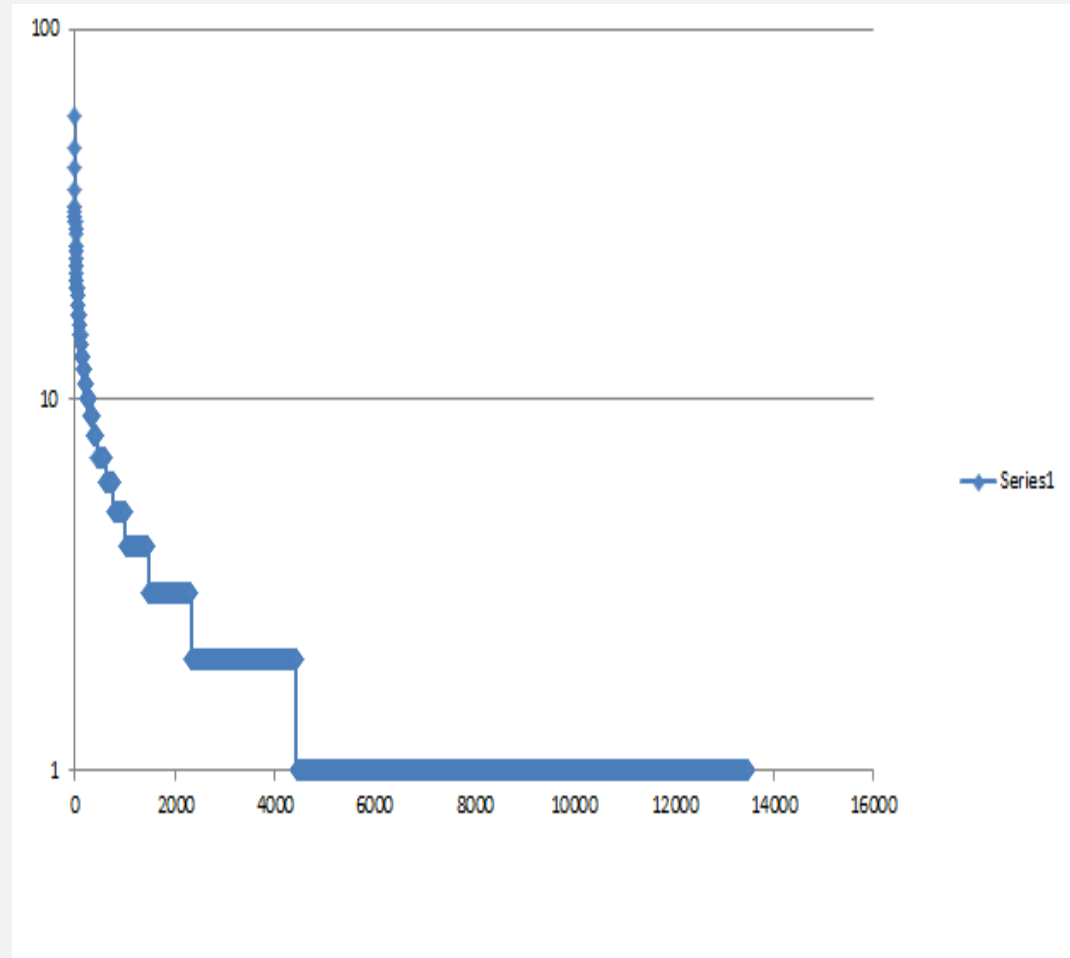
- 13,499 total users
- 26,746 user ratings

Subset

- 294 user with 10 or more reviews
- 4,437 user ratings

Prediction

- 14,700 user ratings



# Prediction Results

Method	Number of Latent Factors	Training RMSE	Validation RMSE
ALS	20	0.220959	1.3471
ALS	40	0.0963635	1.6252
SGD	20	0.423545	1.3556
SGD	40	0.3834435	1.4567
Bias-SGD	20	0.402323	1.1233
Bias-SGD	40	0.39897	1.0933

# Conclusion

## Value

- Business owners
  - Evaluate performance
  - Compare with competitors
- Users
  - Reduce time and effort to find business as per personal preferences

## Future Potential

- Implicit Feature Identification
- Sentiment Analysis of Comparative Opinions