

Estimating the Causal Effects of Treatments Using Yelp Reviews

Luisa Quispe Ortiz and Varun D N *Mentor: Prof David Sontag*
Center for Data Science – New York University

Motivation

- Analysis of how features of text affect individuals' decision involves specifying features beforehand
- Automated text analysis methods identify features of a text, but cannot be used to estimate causal effects of features
- Fong et al. [1] propose a methodology to simultaneously discover treatments and estimate their causal effects
- We use their methodology to estimate the causal effects of treatments on user rating on Yelp using user reviews

Experimental Framework

The experimental protocol proposed by Fong et al. [1] to learn treatments from sIBP and estimate their causal effects is summarized in the flow chart in Figure 1.

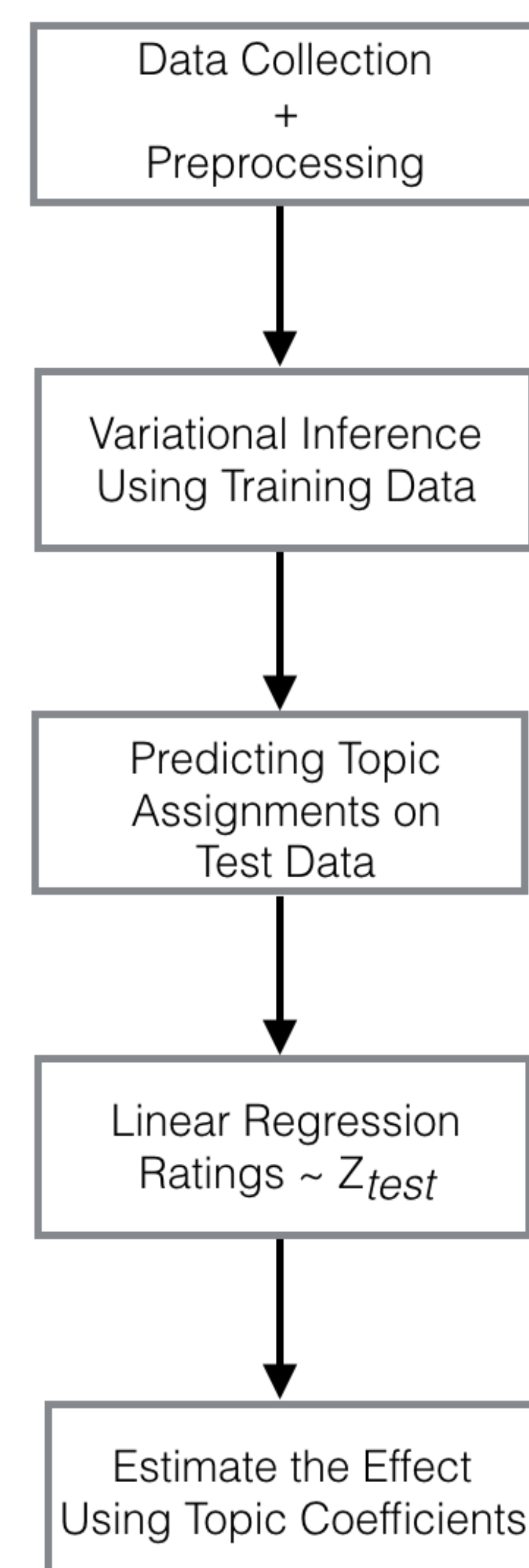


Figure 1: Flow Chart of the Experimental Framework

Supervised Indian Buffet Process

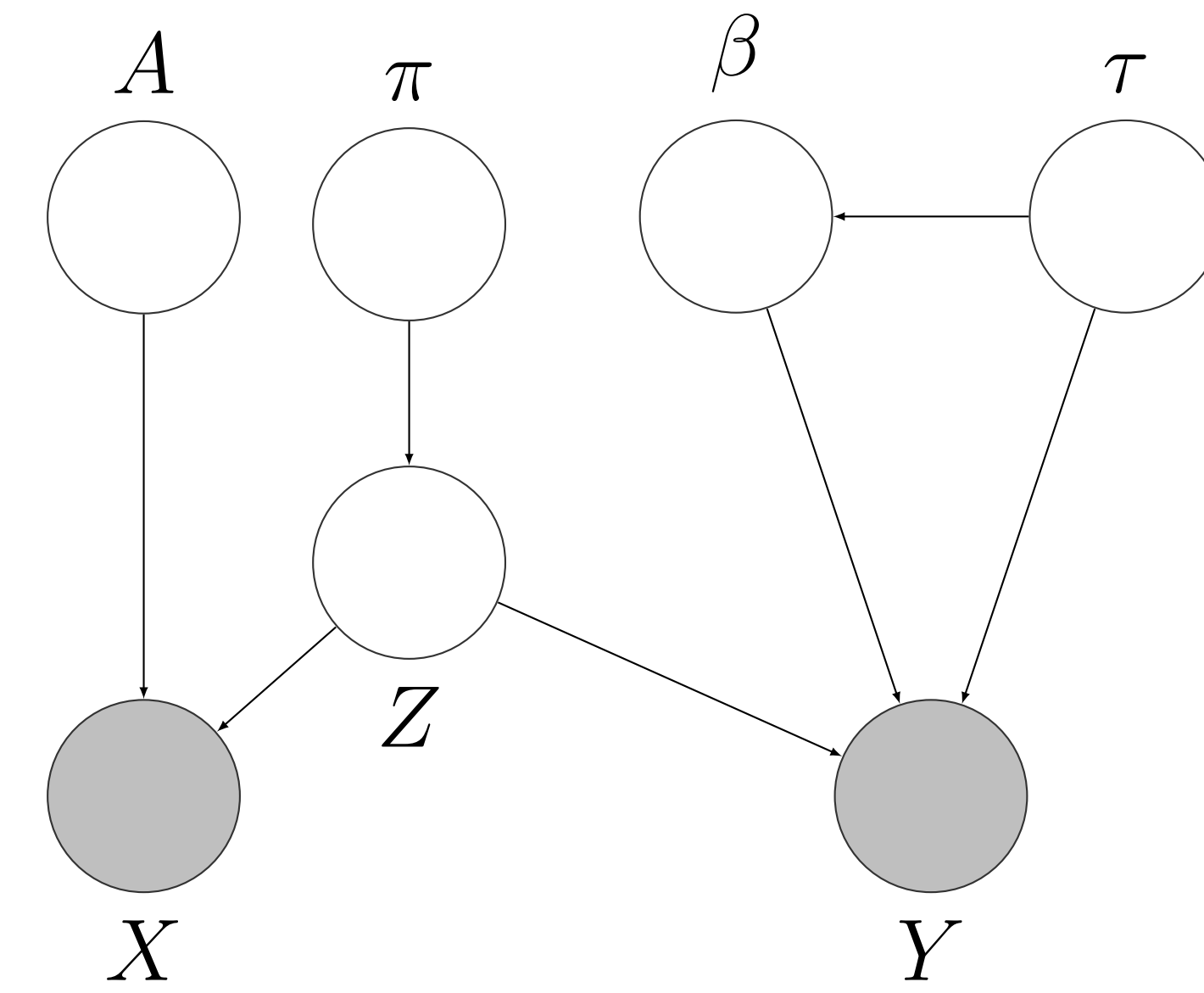


Figure 2: Graphical Model for the Supervised Indian Buffet Process [Source: Fong et al. [1]]

The probability distributions of the random variables defined in the sIBP shown in Figure 2 is as follows:

$$\begin{aligned}
 \pi_k &\sim \text{Stick-Breaking}(\alpha) \\
 z_{i,k} &\sim \text{Bernoulli}(\pi_k) \\
 \mathbf{X}_i | \mathbf{Z}_i, \mathbf{A} &\sim \text{MVN}(\mathbf{Z}_i \mathbf{A}, \sigma_X^2 \mathbf{I}_D) \\
 \mathbf{A}_k &\sim \text{MVN}(\mathbf{0}, \sigma_A^2 \mathbf{I}_D) \\
 \mathbf{Y}_i | \mathbf{Z}_i, \beta &\sim \text{Normal}(\mathbf{Z}_i \beta, \tau^{-1}) \\
 \tau &\sim \text{Gamma}(a, b) \\
 \beta | \tau &\sim \text{MVN}(\mathbf{0}, \tau^{-1} \mathbf{I}_K)
 \end{aligned}$$

Model Details

- The model is described in three steps:
 - Treatment Assignment
 - Document Creation
 - Response to Treatment Vector
- Posterior distribution is approximated by building on Infinite Variational Approach proposed by Doshi-Velez et al. [2]
- Model is run several times for different combinations of α and σ_X to evaluate the output at several different local modes
- Coherence and Exclusivity (CE Score) is calculated for each learned model
- The trained models are sorted by their CE Score and the best model is selected through manual analysis of the treatments

Experiments with Yelp Reviews

- Dataset was obtained from Yelp Academic Dataset Challenge
- 2.7 million reviews from 10 cities around the world
- Information about more than 86,000 business

We designed two experiments using the Yelp data:

- Full Dataset** Sample from the entire dataset which includes reviews of hotels, restaurants, bars, home services and more. Experiment with different vocabulary size
- Mexican Restaurants** Select reviews about Mexican restaurants in Nevada, US. Experiment with different number of treatments.

Results: Full Dataset

Table 1: Top 10 words for 5 treatments (600 features)

X1	X2	X3	X4	X5
sauce	wonderful	told	time	hotel
chicken	food_delicious	asked	ordered	stay
cooked	grab	minutes	order	stayed
ordered	great_place	called	asked	pool
salad	recommend_place	time	table	staying
fried	curry	wait	people	rooms
dish	authentic	manager	told	car
cheese	food_great	left	experience	room
flavor	brunch	phone	wanted	floor
meal	crowd	work	place	buy

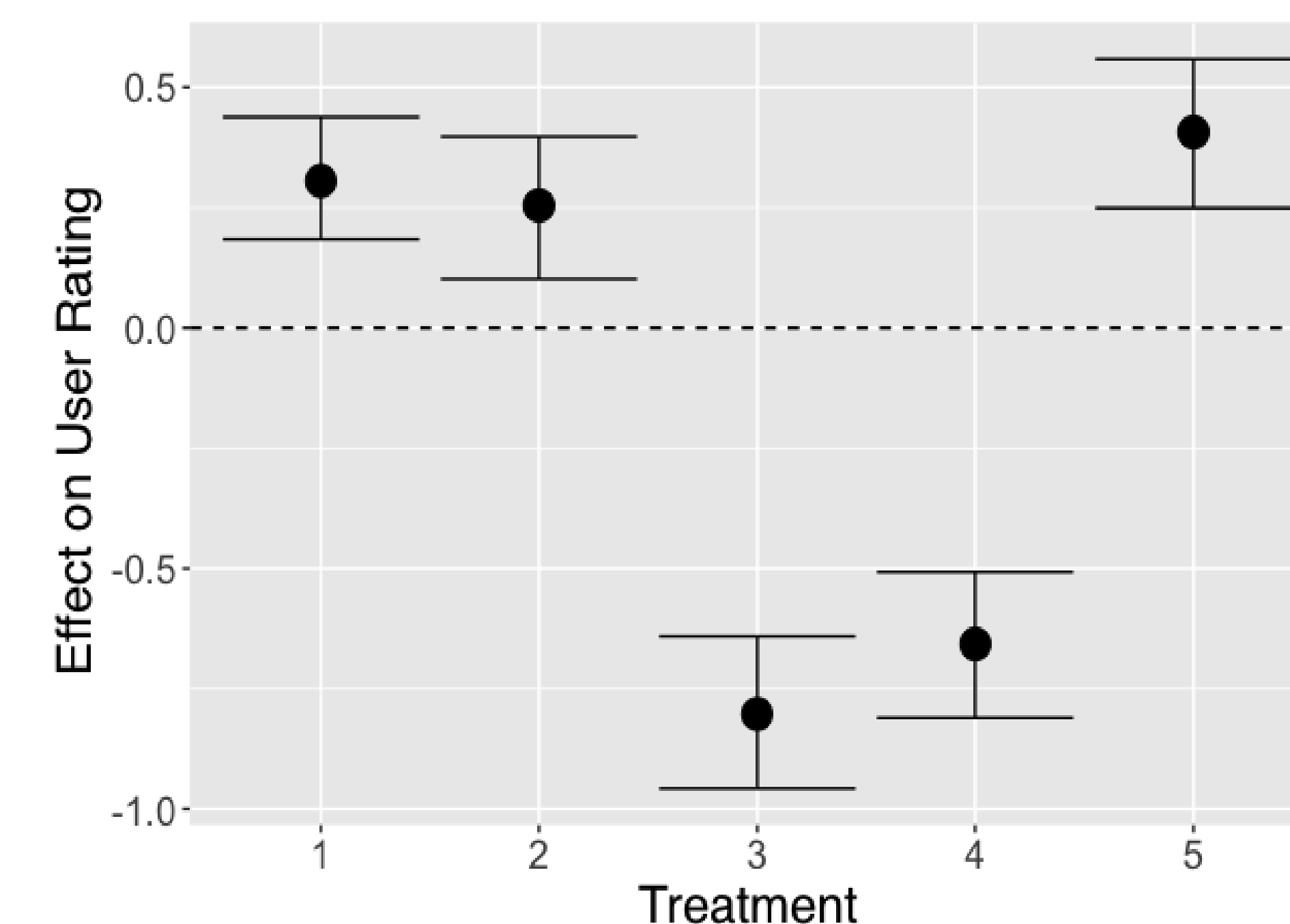


Figure 3: 95% CI - Full Dataset (600 features, 5 treatments)

Results: Mexican Restaurants

Table 2: Top 10 words for 10 treatments (1300 features)

X1	X2	X3	X4	X5	X6	X7	X8	X9	X10
taco	table	meat	tacos_el	table	location	chips_salsa	bowl	worst	minutes
mexican	seated	cheese	el_gordo	place	business	appetizer	fresh	terrible	table
small	server	tortilla	lines	minutes	patient	great	fried_ice	attitude	food
choose	brought	taco	fries	asked	dine	dish	bean_dip	waste	asked
tacos_el	asked	tacos_el	meats	finally	told	bean_dip	great_experience	store	order
delicious	minutes	el_gordo	pollo	order	ago	sweet	chipotle	overcooked	offered
tacos	chips	tacos	sushi	night	reviews	mandalay	rice	girl	manager
came_asada	water	came_asada	taco_place	wait	sushi	margarita	pico_de	standing	water
el_gordo	manager	asked	bowls	ordered	management	tequila	de_gallo	food	left
lengua	empty	onions	best_tacos	half	read	delicious	seaweed	sat	time

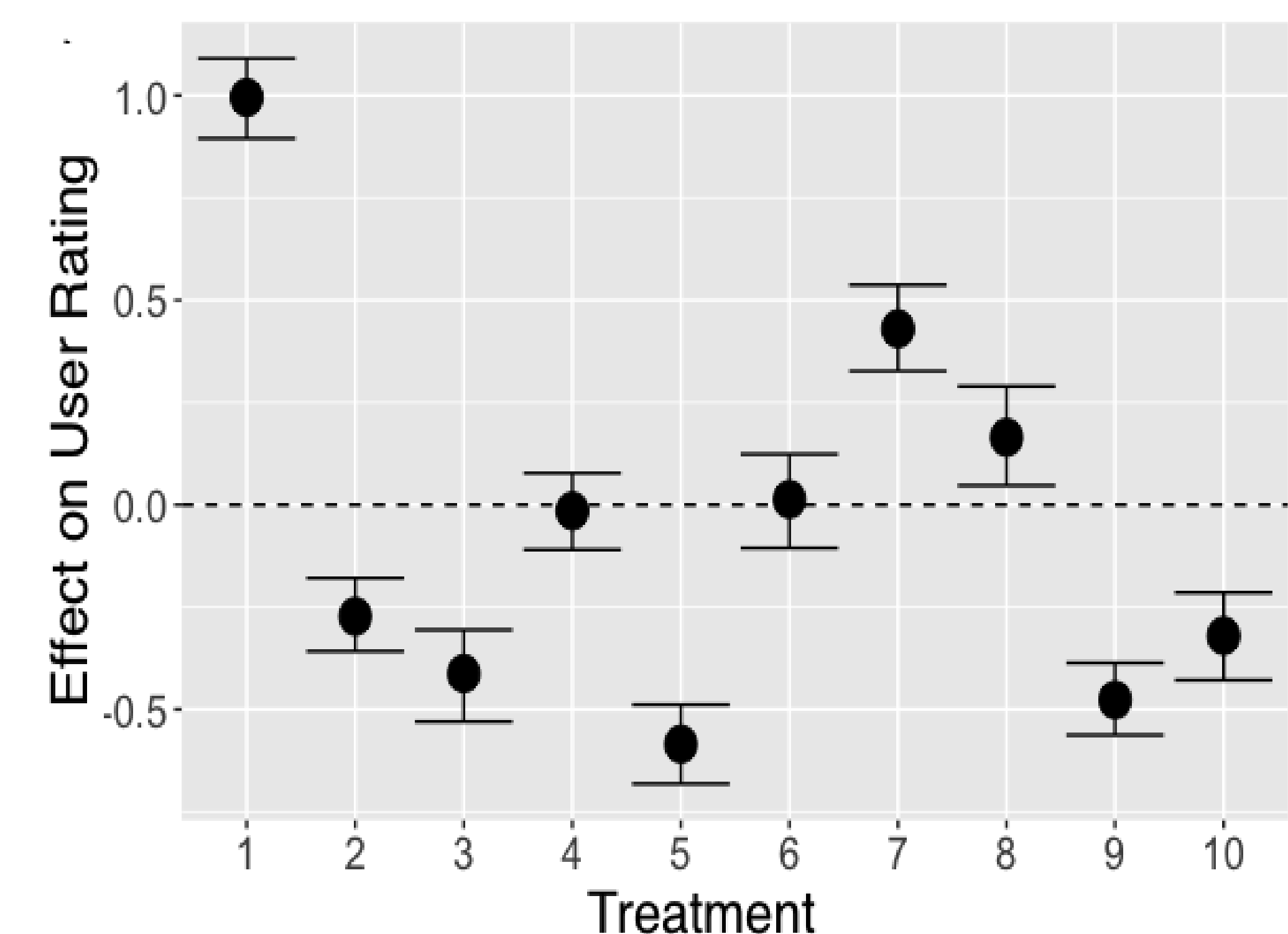


Figure 4: 95% CI - Mexican Restaurants (10 treatments)

Conclusions

- Proposed methodology works on Yelp reviews
- Higher ratings are given for quality of food, variety in toppings and range of options in sides
- Lower ratings are given for bad service

References

- Christian Fong and Justin Grimmer. Discovery of treatments from text corpora. 2016.
- Finale Doshi-Velez, Kurt T Miller, Jurgen Van Gael, Yee Whye Teh, and Gatsby Unit. Variational inference for the indian buffet process.