

Assignment 8

Source: <http://jmcauley.ucsd.edu/data/amazon/>

Dataset: Tools and Home Improvement

Reviews: 134,476

Access LDA output on

<http://bl.ocks.org/sdb418/raw/446b23e74d5db6eb7b179aab4a7271f1/#topic=0&lambda=0.5&term=undefined>

Topic 1: At relevance setting 0.5

<http://bl.ocks.org/sdb418/raw/446b23e74d5db6eb7b179aab4a7271f1/#topic=1&lambda=0.5&term=>

In this topic the term “blade” is the most relevant. By hovering the over the term “blade” the topics 1, 2 & 9 have the word “blade” mentioned in it.

By looking at the most relevant words we can say that this topic talks about “table based blade cutting, sanding tool and dusting to make fence”

Topic 2: At relevance setting 0.8

<http://bl.ocks.org/sdb418/raw/446b23e74d5db6eb7b179aab4a7271f1/#topic=2&lambda=0.8&term=>

In this topic the terms “tools” is the most relevant. By hovering the over this term the topics 2,1,9 & 4 have the word “tools” mentioned in it the most.

By looking at the most relevant words we can say that this topic talks about “good quality blade for a knife”

Topic 3: At relevance setting 0.61

<http://bl.ocks.org/sdb418/raw/446b23e74d5db6eb7b179aab4a7271f1/#topic=3&lambda=0.61&term=>

In this topic the terms “light” is the most relevant. By hovering the over this term the topics 3,6 & 9 have the word “light” mentioned in it the most.

By looking at the most relevant words we can say that this topic talks about “flashlight and batteries required to have a bright light in terms of lumens”

Topic 4: At relevance setting 1

<http://bl.ocks.org/sdb418/raw/446b23e74d5db6eb7b179aab4a7271f1/#topic=4&lambda=1&term=>

In this topic the terms “great” is the most relevant. By hovering the over this term all the topics have the word “great” mentioned in it.

By looking at the most relevant words we can say that this topic talks about “great and easy way to nail on a wall using ladder”

Topic 5: At relevance setting 0.5

<http://bl.ocks.org/sdb418/raw/446b23e74d5db6eb7b179aab4a7271f1/#topic=5&lambda=0.5&term=>

In this topic the terms “switch” is the most relevant. By hovering the over this term the topics 5,6,3 & 1 have the word “switch” mentioned in it the most.

By looking at the most relevant words we can say that this topic talks about “switch for a plug and thermostat to control temperature”

Topic 6: At relevance setting 0.71

<http://bl.ocks.org/sdb418/raw/446b23e74d5db6eb7b179aab4a7271f1/#topic=6&lambda=0.7&term=>

In this topic the terms “light” is the most relevant. By hovering the over this term the topics 3,6 & 9 have the word “light” mentioned in it the most.

By looking at the most relevant words we can say that this topic talks about “bright led light bulbs”

Topic 7: At relevance setting 0.86

<http://bl.ocks.org/sdb418/raw/446b23e74d5db6eb7b179aab4a7271f1/#topic=7&lambda=0.86&term=>

In this topic the terms “good” is the most relevant. By hovering the over this term all the topics have the word “good” mentioned in it.

By looking at the most relevant words we can say that this topic talks about “good way to work with paint is with gloves for clean job”

Topic 8: At relevance setting 0.36

<http://bl.ocks.org/sdb418/raw/446b23e74d5db6eb7b179aab4a7271f1/#topic=8&lambda=0.36&term=>

In this topic the terms “door” is the most relevant. By hovering the over this term the topics 8,4 & 6 have the word “door” mentioned in it the most.

By looking at the most relevant words we can say that this topic talks about “smoke detector and alarm alert for locking the door”

Topic 9: At relevance setting 0.78

<http://bl.ocks.org/sdb418/raw/446b23e74d5db6eb7b179aab4a7271f1/#topic=9&lambda=0.78&term=>

By looking at the most relevant words we can say that this topic talks about “charging battery powered bosch drills and drill bits”

<http://bl.ocks.org/sdb418/raw/446b23e74d5db6eb7b179aab4a7271f1/#topic=10&lambda=0.78&term>

By looking at the most relevant words we can say that this topic talks about “water, shower, toilet, sink; in general bathroom fittings”

Amazon review data | Text Mining in R [Sentiment Analysis] | sdb-416/Assignment-8 | NYU Classes : F16 BA MD7 (1) | LDAvis

localhost/Tools/index.html?Topic=0&lambda=1&term=

Selected Topic: 0 | Previous Topic | Next Topic | Clear Topic

Slide to adjust relevance metric: $\lambda = 1$ (0.0 to 1.0)

Intertopic Distance Map (via multidimensional scaling)

PC2 vs PC1

Topics 1-10 are represented as bubbles of varying sizes and positions.

Top-30 Most Salient Terms^[1]

Overall term frequency vs. Estimated term frequency within the selected topic

Term	Estimated term frequency within the selected topic
light	~65,000
drill	~25,000
bulb	~15,000
flashlight	~12,000
water	~12,000
bulbs	~12,000
tool	~30,000
batteries	~25,000
battery	~25,000
led	~25,000
bright	~15,000
blade	~25,000
lights	~15,000
door	~15,000
tools	~25,000
lamp	~15,000
bits	~15,000
knife	~15,000
lucet	~15,000
shower	~15,000
power	~25,000
drill	~15,000
switch	~15,000
table	~15,000
lock	~15,000
router	~15,000
brush	~15,000
hose	~15,000
cut	~15,000

1. $\text{salience}(\text{term } w) = \text{frequency}(w) * (\sum_i \lambda_i \alpha_i^2)^{-1} * \log(p_i) / w(p_i)$ for topics i ; see Chuang et al. (2012)
 2. $\text{relevance}(\text{term } w | \text{topic } t) = \lambda * p(w | t) * (1 - \lambda) * p(w | t) / p(w)$; see Stewart & Shirley (2014)

Marginal topic distribution: 2%, 9%, 10%

When I am trying to share my files on GitHub its showing me 404 error.

