

# Experimenting with Drugs (and Topic Models)



Michael Paul and Mark Dredze

Johns Hopkins University



# Online Drug Communities

- **Drugs-Forum.com**
  - “Drugs-forum is an information hub of high-standards and a platform where people can freely discuss recreational drugs in a mature, intelligent manner. Drugs-Forum offers a wealth of quality information and discussion of drug-related politics, in addition to assistance for members struggling with addiction.”
- We analyzed 100,000 messages pertaining to personal drug experiences

# Web-Based Drug Research

- For new and emerging drugs, information can be difficult to obtain through traditional means
  - e.g. lab experiments, surveys
- Modern source of information: Internet forums
  - Always curated manually by humans
- A step toward automation: **topic modeling**
  - Corpus exploration
  - Coarse-grained information extraction

# Factorial LDA (f-LDA)

- **Multi-dimensional** topic model
  - M.J. Paul and M. Dredze. Factorial LDA: Sparse Multidimensional Models of Text. NIPS 2012.
- Word tokens are associated with a *vector* of latent variables instead of a single topic variable
  - Can jointly model pairs of concepts like topic and perspective or sentiment
- Instead of a distribution over topics, each document has distribution over *tuples*
- Each tuple is associated with its own word distribution

# Performance Enhanced Factorial LDA

- Joint model of 3 dimensions:
  - Drug type
  - Method of intake (delivery)
  - Aspect
- Learn word distributions for triples such as:  
(Cocaine, Snorting, Health) (Cocaine, Snorting, Usage)

nose  
pain  
damage  
blood  
cocaine  
problem

coke  
line  
lines  
nose  
small  
cut

# Model Parameters

- Why should the word distributions for triples make any sense?
- Parameters are tied across the priors of each word distribution
  - The prior for (Cocaine, Snorting, Effects) shares parameters with (Cocaine, Smoking, Effects) which shares parameters with the prior for (Cannabis, Smoking, Effects)

## Cannabis


weed  
cannabis  
thc  
marijuana  
stoned  
bowl  
bud  
joint  
blunt  
herb  
bong  
pot  
sativa  
blaze  
indica  
smoking  
blunts  
strains  
hemp  
...

## Oral

capsules  
consumes  
toast  
stomach  
chewing  
ambien  
digestion  
juice  
absorbed  
ingestion  
meal  
tiredness  
chew  
juices  
gelatin  
yogurt  
fruit  
oj  
digest  
...

## Chemistry

solvent  
extraction  
evaporate  
evaporated  
solvents  
evaporation  
yield  
chloride  
alkaloids  
tek  
compounds  
evaporating  
atom  
aromatic  
non-polar  
purified  
jar  
methyl  
ethanol  
....



Each dimension  
has a weight vector  
over the vocabulary

## Cannabis

weed  
cannabis  
thc  
marijuana  
stoned  
bowl  
bud  
joint  
blunt  
herb  
bong  
pot  
sativa  
blaze  
indica  
smoking  
blunts  
strains  
hemp  
...

## Oral

capsules  
consumes  
toast  
stomach  
chewing  
ambien  
digestion  
juice  
absorbed  
ingestion  
meal  
tiredness  
chew  
juices  
gelatin  
yogurt  
fruit  
oj  
digest  
...

## Chemistry

solvent  
extraction  
evaporate  
evaporated  
solvents  
evaporation  
yield  
chloride  
alkaloids  
tek  
compounds  
evaporating  
atom  
aromatic  
non-polar  
purified  
jar  
methyl  
ethanol  
....

thc  
method  
extraction  
plant  
material  
cannabis  
simple  
coffee  
oil  
contains  
jar  
dried  
process  
dry  
water  
extract  
results  
salt  
available  
...

exp(





word distribution for triple

( Cannabis  
Oral  
Chemistry )

## Posterior

oil  
water  
butter  
thc  
weed  
hash  
cannabis  
alcohol  
make  
milk  
high  
marijuana  
add  
cup  
extract  
...  
mixture  
hours  
try  
brownies

multinomial parameters  
sampled from Dirichlet



## Prior

thc  
method  
extraction  
plant  
material  
cannabis  
simple  
coffee  
oil  
contains  
jar  
dried  
process  
dry  
water  
extract  
results  
salt  
available  
...

word distribution for triple

( Cannabis  
Oral  
Chemistry )

=



## Posterior

oil  
water  
butter  
thc  
weed  
hash  
cannabis  
alcohol  
make  
milk  
high  
marijuana  
add  
cup  
extract  
...  
mixture  
hours  
try  
brownies

multinomial parameters  
sampled from Dirichlet



## Prior

thc  
method  
extraction  
plant  
material  
cannabis  
simple  
coffee  
oil  
contains  
jar  
dried  
process  
dry  
water  
extract  
results  
salt  
available  
...

# What can we learn from this?

- We can use the model to bring attention to relevant messages and snippets of text
  - Extractive summarization

# Questions

- How are people using Salvia?



(  
Salvia  
Smoking  
Usage  
)

» "Best way is to use a torch lighter, bong or pipe (bong recommended) and hold in each hit 20-40 seconds."

(  
Salvia  
Oral  
Usage  
)

» "A dose of Salvia leaves is 2 grams. A dose of Salvia 5X extract is 0.4 grams (400mg) A dose of salvia 10X extract is 0.2 grams (200mg) A dose of Salvia 25X extract is 0.08 grams (80mg) A dose of Salvia 50X extract is 0.04 grams (40mg) A dose of Salvia 250X extract is impossible."

# Questions



- What are the effects of Salvia?

## Salvia Smoking Effects

- » “He then took one large hit and held it in and laid back and began to feel his body getting heavy and his vision started to get this dim orange brownish light to it and he closed his eyes and moved his body around and it shook both of the feelings off.”

## Salvia Oral Effects

- » “When chewed, the first effects are felt after about 15 minutes. After about 30 minutes, the full effects should be realized. Typical Salvia experiences last 5 to 10 minutes on average, with noticeable after-effects lasting up to 1/2 hour.”

# Conclusion

- Online communities yield a large amount of candid data on a subject that is traditionally difficult to study
- In the future, we plan to consider many analyses beyond the topic model experiments shown here