Discovering Surprising Documents with Context-Aware Word Representations

Michał Dereziński
UC Santa Cruz, CA

Khashayar Rohanimanesh and Aamer Hydrie
eBay Research Labs, San Jose, CA

IUI'18, 3-8-18

Finding Interesting Products

How can we increase user engagement for online shopping?

Idea: Display ads for products that are:

- unique,
- surprising,
- interesting.

Finding Interesting Products

How can we increase user engagement for online shopping?

Idea: Display ads for products that are:

- unique,
- surprising,
- interesting.

Example: Random Phone Cases

Example: Surprising Phone Cases

Unsupervised Discovery of Surprising Products

Can we automatically discover surprising products?

Yes, and all we need is the eBay product titles!

Our hypothesis:
Concept is surprising if it is generated from a *diverse set of topics*.

We identify those topics from words appearing in the *product title*.

Unsupervised Discovery of Surprising Products

Can we automatically discover surprising products?

Yes, and all we need is the eBay product titles!

Our hypothesis:
Concept is surprising if it is generated from a *diverse set of topics*.

We identify those topics from words appearing in the *product title*.

Unsupervised Discovery of Surprising Products

Can we automatically discover surprising products?

Yes, and all we need is the eBay product titles!

Our hypothesis:
Concept is surprising if it is generated from a *diverse set of topics*.

We identify those topics from words appearing in the *product title*.

Examples: Topic Diversity in Product Titles

(a) **Eyeshadow** Palettes for **iPhone** 6 case

(b) White Silicone **Horn** Stand Speaker for Apple **iPhone** 4/4S

(c) **Equation** Wall **Clock** Gifts for Math Gurus

Information-Theoretic Model of Surprisingness

Two main conributions:

1. Context-aware **word representations** via topic modeling
2. **Jensen-Shannon Divergence** as a measure of topic diversity

Information-Theoretic Model of Surprisingness

Two main conributions:

1. Context-aware **word representations** via topic modeling
2. **Jensen-Shannon Divergence** as a measure of topic diversity

Information-Theoretic Model of Surprisingness

Two main conributions:

1. Context-aware **word representations** via topic modeling
2. **Jensen-Shannon Divergence** as a measure of topic diversity

Topic Modeling

Latent Dirichlet Allocation [Blei et al. 2003]

Given a document corpus, annotate all words with a set of topics

Example topics: **Phones**, **Music**, **Food**, **Kitchen supplies**

Topic assignments:

1. "White Silicone **Horn** Stand Speaker for **Apple iPhone**"
2. "New Stainless Steel **Apple Cutter** Tool"

Topic Modeling

Latent Dirichlet Allocation [Blei et al. 2003]

Given a document corpus, annotate all words with a set of topics

Example topics: **Phones**, **Music**, **Food**, **Kitchen supplies**

Topic assignments:

1. "White Silicone **Horn** Stand Speaker for **Apple iPhone**"
2. "New Stainless Steel **Apple Cutter** Tool"

Topic Distributions as Word Representations

Word-topic co-occurrence matrix:

words \ topics	*Phones*	*Computers*	*Food*
Smartphone	111	3	1
Screen	60	40	0
Mouse	0	30	3
Apple	50	25	45

Simple word representation: **Screen** $\rightarrow \underbrace{\left\{ \overset{\text{Phones}}{0.6}, \overset{\text{Computers}}{0.4}, \overset{\text{Food}}{0} \right\}}_{P_{\text{Screen}}}.$

Topic Distributions as Word Representations

Word-topic co-occurrence matrix:

words \ topics	*Phones*	*Computers*	*Food*
Smartphone	111	3	1
Screen	60	40	0
Mouse	0	30	3
Apple	50	25	45

Simple word representation: **Screen** $\rightarrow \underbrace{\left\{ \overset{\text{Phones}}{0.6}, \overset{\text{Computers}}{0.4}, \overset{\text{Food}}{0} \right\}}_{P_{\text{Screen}}}.$

Improving Distributional Word Representations

We enhance those simple word representations through:

- *Laplace smoothing*:
  Mixing with a global prior topic distribution.

- *Context-dependent priors*:
  Introduce a local prior topic distribution for each product

- *Topic similarity*,
  Account for the semantic distance between the topics.
  Example: *Phones* is more similar to *Computers* than to *Food*

Improving Distributional Word Representations

We enhance those simple word representations through:

- *Laplace smoothing*:

  Mixing with a global prior topic distribution.

- *Context-dependent priors*:

  Introduce a local prior topic distribution for each product

- *Topic similarity*,

  Account for the semantic distance between the topics.

  Example: *Phones* is more similar to *Computers* than to *Food*

Improving Distributional Word Representations

We enhance those simple word representations through:

- *Laplace smoothing*:

  Mixing with a global prior topic distribution.

- *Context-dependent priors*:

  Introduce a local prior topic distribution for each product

- *Topic similarity*,

  Account for the semantic distance between the topics.

  Example: *Phones* is more similar to *Computers* than to *Food*

Jensen-Shannon Divergence

Probabilistic notion of diversity [Fuglede and Topsoe, 2004].

Given representations: $\{P_{\text{Screen}}, P_{\text{Mouse}}, P_{\text{Apple}}\}$

and weights: $d_{\text{Screen}} + d_{\text{Mouse}} + d_{\text{Apple}} = 1$

Jensen-Shannon Div. of $T = \{\text{Screen, Mouse, Apple}\}$ is given by

$$D_{JS}(\Sigma d_w P_w) = \sum_{w \in T} d_w D_{KL}(P_w \| M), \quad \text{where} \quad M = \sum_{w \in T} d_w P_w.$$

D_{KL} denotes Kullback-Leibler divergence.

Jensen-Shannon Divergence

Probabilistic notion of diversity [Fuglede and Topsoe, 2004].

Given representations: $\{P_{\text{Screen}}, P_{\text{Mouse}}, P_{\text{Apple}}\}$

and weights: $d_{\text{Screen}} + d_{\text{Mouse}} + d_{\text{Apple}} = 1$

Jensen-Shannon Div. of $T = \{\text{Screen, Mouse, Apple}\}$ is given by

$$D_{JS}(\Sigma d_w P_w) = \sum_{w \in T} d_w D_{KL}(P_w \| M), \quad \text{where} \quad M = \sum_{w \in T} d_w P_w.$$

D_{KL} denotes Kullback-Leibler divergence.

Evaluating Surprisingness - Baselines

Our surprisingness metric is compared against:

- *Rao* text diversity metric [Bache et al, 2013],
- Shannon entropy as a measure of diversity,
- Determinantal Point Processes (*DPP*) [Kulesza and Taskar, 2012]

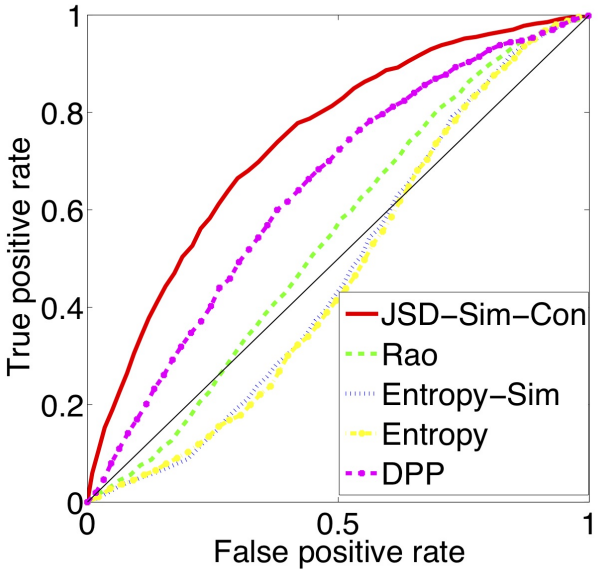
Evaluating Surprisingness - Data Sets

1. Phone cases on eBay (product titles, 8-12 words each)
   - We hired workers from AMT to label a collection of phone cases found on Pinterest and eBay.
   - 2179 positive and 9770 negative instances.
2. NSF proposal abstracts (61,902 total, 200-300 words each).
   - We used this set for training a topic model.
   - To get labeled data, we had to generate 5000 artificial labeled examples, by mixing random abstracts.

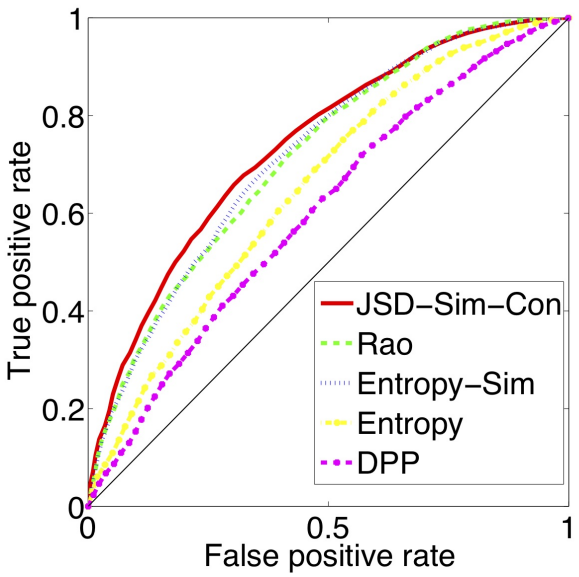
Evaluating Surprisingness - Data Sets

1. Phone cases on eBay (product titles, 8-12 words each)
   - We hired workers from AMT to label a collection of phone cases found on Pinterest and eBay.
   - 2179 positive and 9770 negative instances.
2. NSF proposal abstracts (61,902 total, 200-300 words each).
   - We used this set for training a topic model.
   - To get labeled data, we had to generate 5000 artificial labeled examples, by mixing random abstracts.

Evaluating Surprisingness - ROC Plots



(a) eBay phone cases dataset



(b) NSF abstracts dataset

eBay dataset is harder for the baselines, due to shorter text length.

Our method JSD-Sim-Con performs equally well on both datasets.

Takeaways

- Fully unsupervised system for discovering surprising products
- Text surprisingness is correlated with topic diversity
- Product title (8-12 words) is enough to produce a surprisingness score
- We employ two key ideas:
  1. Using topic distributions as word representations,
  2. Jensen-Shannon Divergence as a measure of diversity.

Takeaways

- Fully unsupervised system for discovering surprising products
- Text surprisingness is correlated with topic diversity
- Product title (8-12 words) is enough to produce a surprisingness score
- We employ two key ideas:
  1. Using topic distributions as word representations,
  2. Jensen-Shannon Divergence as a measure of diversity.

Takeaways

- Fully unsupervised system for discovering surprising products
- Text surprisingness is correlated with topic diversity
- Product title (8-12 words) is enough to produce a surprisingness score
- We employ two key ideas:
  1. Using topic distributions as word representations,
  2. Jensen-Shannon Divergence as a measure of diversity.

Takeaways

- Fully unsupervised system for discovering surprising products
- Text surprisingness is correlated with topic diversity
- Product title (8-12 words) is enough to produce a surprisingness score
- We employ two key ideas:
  1. Using topic distributions as word representations,
  2. Jensen-Shannon Divergence as a measure of diversity.

Thank you