

# UNCOVERING THE MESSAGE FROM THE MESS OF BIG DATA

Neil Bendle & Shane (Xin) Wang  
Ivey Business School, Western  
University, London Ontario, Canada



Uncovering The Message From The Mess Of Big Data

# Summary

- Consumer generated content proliferates at incredible speed
- This “big data” contains incredible detail on consumers’ preferences
- But many firms can’t use it
- We suggest a non-proprietary technique *Latent Dirichlet Allocation (LDA)*
- LDA can uncover the message in the mess of big data

Consumers generate big data  
e.g., online reviews, blogs, tweets



Firms can analyze unstructured text in  
consumer generated big data using  
LDA



Extracts message from consumers  
e.g., What consumers care about  
How they think about market  
What they want

# Structured & Unstructured Data

- Market research often relies on structured data, e.g., a survey with a set number of response options
  - Can be slow & expensive
  - Only generates information on what is asked
  - Consumers compress nuanced opinions into response options
- Recent proliferation of unstructured data, e.g., online reviews



Great External Drive

November 4, 2015

I have used it for 3 USBs that were full and I have not even tapped the memory it has! Great investment. Small and compact!

✓ I would recommend this to a friend!



*"great location!"*

Reviewed 10 July 2015

Was a very busy hotel, as Wimbledon had started. We were staying due to a Aerosmith concert on Clapton common, this hotel was a couple of tube stops away and was much cheaper than hotels nearer the common. Staff were very friendly and breakfast was great.

34 of 39 people found the following review helpful

★★★★★ **Some Pig Indeed: Moving Story of Charlotte and Wilbur**

By [Amazon Customer](#) on December 30, 2006

First I must confess that I have never read the original book 'Charlotte's Web' by E. B. White or seen the 1973 animated film version (with Debbie Reynolds' voice). So please read my review of this new live-action version as such, though probably this kind of confession is really unnecessary.



Uncovering The Message From The Mess Of Big Data

# Uncovering Consumer Messages

- Consumers often aren't shy about sharing their thoughts
- Clear **benefits to analyzing this data**
  - Allows managers real-time access to feedback
  - Consumers decide what topics discuss
  - Reveals how consumers think
- But data is too large to manually scour
- And is often messy making it hard for traditional analysis
  - Review comments can meander erratically between topics
  - Include poor grammar, misspelt words, and colloquialisms
- Managers often don't know how to extract the trove of information hidden in consumer generated big data
- Need a **way to extract the message** from the mess
- We suggest **Latent Dirichlet Allocation (LDA)**

# Method: Latent Dirichlet Allocation

- LDA is a topic modelling approach
- Associates words used in reviews (and other text) with topics
  - E.g., Car's brakes & early warning system may be grouped under safety
- Estimates topics a consumer cares about given what he/she writes
  - E.g., from review a consumer cares 70% about performance & 30% about MPG
- Is flexible, doesn't use a dictionary
  - Copes with misspelling & colloquialisms
- Can assess valence
  - Is topic a strength or weakness?
- See technical details for limitations

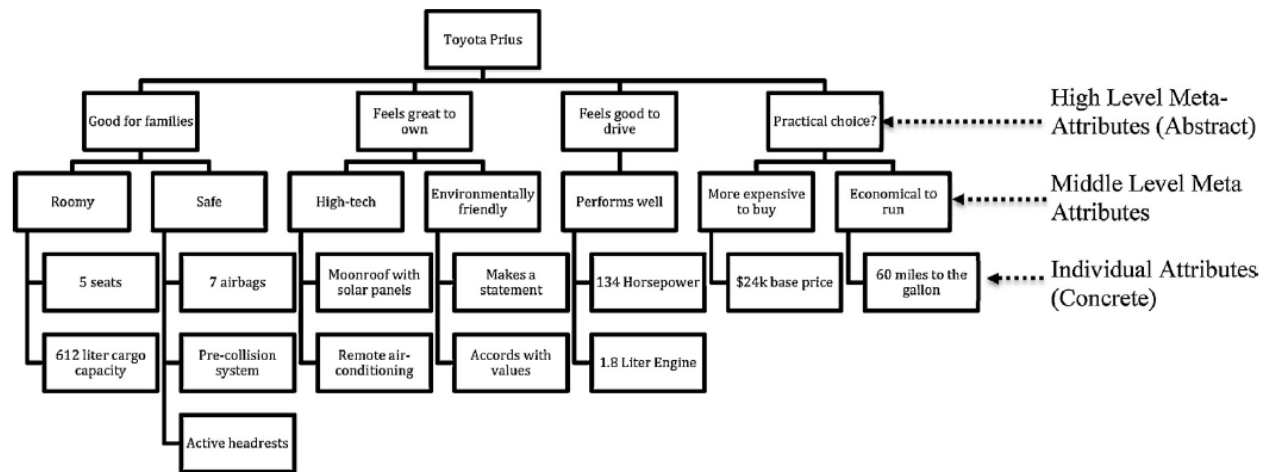
## Technical Details

- *Assumes consumers write in proportion to how much a topics matters to them*
- *"Bag of words": i.e., order of words doesn't matter*
- *Unsupervised: Little human involvement – limits bias but ignores analyst's knowledge*
- *All topics are assumed to be equally dissimilar*
- *Analyst picks topic number. No theory on precise number. Different analysts may generate different results*

# Results: LDA & Your Firm

- Using LDA you can learn what matters to customers in your Industry
  - Can group attributes at various levels of abstraction
  - “Airbags” & “Seats” may link into same topic -- “good for families”

Figure 2. Hierarchy of attributes



- Using LDA you can uncover what customers say about your firm
- You can also find if you perform well on topics that matter

# Results: The Market Structure/Vulnerable Competitors

- Business strategists can benefit greatly from using LDA
- Remember information on your competitors' is there in plain sight
- You can **find the market structure**
  - Which firm's offerings are seen as similar?
  - How do the priorities of firm A's customers differ from those of firm B's customers?
- You can perform **competitor identification**
  - Who competes with you where it matters, in consumers' minds?
- You can then **uncover the weaknesses of your competitors**
  - Where are your competitors performing especially poorly?

# Conclusion: Big Data Can Be Tamed

- Our main aim is not to advocate for LDA against similar techniques ...but that big data can be tamed
  - We can relatively easily analyze unstructured data
- Managers can use LDA to extract messages from messy big data, E.g.,
  1. Uncover topics that consumers are talking about
  2. Uncover connections between the topics
  3. Understand which topics are seen positively or negatively
  4. Reveal structure of industry
  5. Highlight vulnerable competitors

**Big data is intimidating but taming big data allow uncovering the message in the mess**



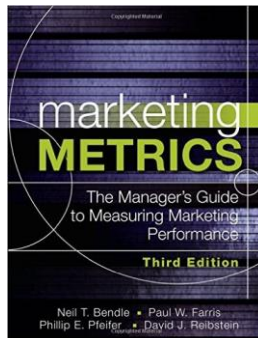
# Next steps/future work

## The *Journal of Consumer Research* at 40: A Historical Analysis

XIN (SHANE) WANG  
NEIL T. BENDLE  
FENG MAI  
JUNE COTTE

This article reviews 40 years of the *Journal of Consumer Research* (JCR). Using text mining, we uncover the key phrases associated with consumer research. We use a topic modeling procedure to uncover 16 topics that have been featured in the journal since its inception and to show the trends in topics over time. For example, we highlight the decline in family decision-making research and the flourishing of social identity and influence research since the journal's inception. A citation analysis shows which JCR articles have had the most impact and compares the topics in top-cited articles with all JCR journal articles. We show that methodological and consumer culture articles tend to be heavily cited. We conclude by investigating the scholars who have been the top contributors to the journal across the four decades of its existence. And to better understand which schools have contributed most to the knowledge of consumer research over this history, we provide an analysis of where these top-performing scholars were trained. Our approach shows that the JCR archives can be an excellent source of data for scholars trying to understand the complicated, challenging, and dynamic field of consumer research.

**Keywords:** topic modeling, *Journal of Consumer Research*, historical analysis, citation analysis



- LDA can be widely applied beyond online user reviews

For example, we extracted topics in consumer research

<http://jcr.oxfordjournals.org/content/42/1/5>

- Techniques advance every day  
Improved variants of LDA and other techniques are developing

- We research/teach big data & marketing metrics

<http://www.ivey.uwo.ca/faculty/directory/xin-wang/>

<http://www.ivey.uwo.ca/faculty/directory/neil-bendle/>

- Visit Neil's Marketing Thought blog

[www.neilbendle.com](http://www.neilbendle.com)

- Or follow him on twitter

@neilbendle



Uncovering The Message From The Mess Of Big Data