Artificial Intelligence for Social Media Marketing: Data, Methods, and Insights

Liu Liu

University of Colorado Boulder – Leeds School of Business

MSI Accelerator, September 15th 2022
Brand Manager
@oldnavy
Brand Manager
@oldnavy

Consumer
#oldnavy
How is my brand portrayed in consumer photos?
How is my brand portrayed in consumer photos?
Increasing Amount of Unstructured Social Conversations

Text

Image

Video

How can we extract insights and get value?
Insights/Problems

Product Design
- Understand customers needs

Branding
- Measure brand perceptions

Advertising and social media
- Generate social media posts and ads

Influencer marketing
- What makes a good influencer video ads?
  ...

Data

- Text
- Image
- Video
Agenda

Artificial Intelligence (AI): A brief intro and recent breakthroughs with Deep learning

AI for Social Media Marketing: Cases studies
- Visual Listening In: Extracting Brand Image Portrayed on Social Media (Liu et al. 2020)
- Identify Customer Needs from User-Generated Content (Timoshenko and Hauser 2019)
- Supporting Content Marketing with Natural Language Generation (Reisenbichler et al. 2022)

Brainstorming and group work

Sharing group work and reflections (e.g., pitfalls)
AI and Recent Breakthroughs with Deep Learning
AI History Outline

1. - 1956  Prehistory of Artificial Intelligence
2. 1956 – 1974 First Artificial Intelligence Spring
3. Break for Activities
4. 1980 - 1987 Second Artificial Intelligence Spring
5. Break for Activities
6. 2011 - Present Third Artificial Intelligence Spring
7. What is next?
ARTIFICIAL INTELLIGENCE IS NOT NEW

ARTIFICIAL INTELLIGENCE
Any technique which enables computers to mimic human behavior

MACHINE LEARNING
AI techniques that give computers the ability to learn without being explicitly programmed to do so

DEEP LEARNING
A subset of ML which make the computation of multi-layer neural networks feasible

Deep learning are taking over

Deep learning have become one of the main approaches to AI

They have been successfully applied to various fields

They have established the state of the art

- Often exceeding previous benchmarks by large margin
- Sometimes solving problems you couldn't solve using earlier ML methods
IMAGENET

22,000 categories : 15,000,000 images

The Image Classification Challenge:
1,000 object classes
1,431,167 images

Output:
- Scale
- T-shirt
- Steel drum
- Drumstick
- Mud turtle

Output:
- Scale
- T-shirt
- Giant panda
- Drumstick
- Mud turtle

Russakovsky et al. IJCV 2015
Pre-deep learning

Deep learning, e.g., Convolutional Neural Network

Top-5 error rate on ImageNet

![Graph showing top-5 error rate on ImageNet from 2010 to 2015, including human error and ArXiv 2015 results.](image)
Microsoft's Artificial Intelligence and Research Unit earlier this week reported that its speech recognition technology had surpassed the performance of human transcriptionists.
Figure 1: Training AlphaZero for 700,000 steps. Elo ratings were computed from evaluation games between different players when given one second per move. **a** Performance of AlphaZero in chess, compared to 2016 TCEC world-champion program Stockfish. **b** Performance of AlphaZero in shogi, compared to 2017 CSA world-champion program Elmo. **c** Performance of AlphaZero in Go, compared to AlphaGo Lee and AlphaGo Zero (20 block / 3 day) (29).
Alphafold: A Solution to a 50-Year-Old Grand Challenge in Biology

This is one of the most significant discoveries in the history of Biology - DeepMind announced that they have solved the protein-folding problem!

#DSOTD

ThisPersonDoesNotExist.com uses AI to generate endless fake faces

Hit refresh to lock eyes with another imaginary stranger

By James Vincent | Feb 15, 2019, 7:38am EST

OpenAI’s new language generator GPT-3 is shockingly good — and completely mindless

The AI is the largest language model ever created and can generate amazing human-like text on demand but won’t bring us closer to true intelligence.

By Will Douglas Heaven

July 20, 2020

Meet GPT-3. It Has Learned to Code (and Blog and Argue).

The latest natural-language system generates tweets, pens poetry, summarizes emails, answers trivia questions, translates languages and even writes its own computer programs.

One of his experiments involved a pop psychologist, Scott Barry Kaufman. The system took in Mr. Kaufman’s name and a topic for discussion: creativity. Then, when asked “How do we become more creative?” GPT-3 responded instantly:

I think creative expression is a natural byproduct of growing up in a diverse world. The more diverse the world is, the more you get exposed to different people, to different opportunities, to different places and to different challenges. And the more diverse that is, the more likely you’ll be to be able to put the dots together to form something new. And in many ways, I think if you want to be creative, you have to go for it. If you want to be a writer, you have to write, if you want to be a musician, you have to create music, if you want to be a comedian, you have to create comedy, and the more you create, the more likely it is that you’ll stumble onto some interesting stuff.

Later, when Mr. Wrigley posted the paragraph on Twitter, somebody looped in the real Scott Barry Kaufman. He was stunned. “It definitely sounds like something I would say,” the real Mr. Kaufman tweeted, later adding, “Crazy accurate A.I.”
Are robots with AI going to be better at making art than humans?
Are they already better at making art now?
Stay tuned to find out.

Welcome to GAN NFT
What is a GAN?
A Generative Adversarial Network is a class of machine learning frameworks. Essentially, two large datasets (images, audio, video) etc, are left unsupervised to “learn” or “train” off of each other. After a training session, a file is created that contains the style of what it learned during training. This file could be considered to be the “brain” of the ai or GAN, because if directed, it can generate new images or video or songs with the same style as the training set, usually using a modern gpu with high vram, or eth mining rig hashpower to render.
An AI-Generated Artwork Won First Place at a State Fair Fine Arts Competition, and Artists Are Pissed

Jason Allen's AI-generated work "Théâtre D'opéra Spatial" took first place in the digital category at the Colorado State Fair.

A man came in first at the Colorado State Fair's fine art competition using an AI generated artwork on Monday. "I won first place," a user going by Sincarnate said in a Discord post above photos of the AI-generated canvases hanging at the fair.

DALL·E 2 is a new AI system that can create realistic images and art from a description in natural language.

https://openai.com/dall-e-2/
What are deep neural networks?

Fully-Connected Network

Convolutional Neural Network (CNN)

Recurrent Neural Network (RNN)

Transformer based neural network, e.g., Bert, GPT
What are deep neural networks?

Voice signal $\rightarrow$ N.Net $\rightarrow$ Transcription

Image $\rightarrow$ N.Net $\rightarrow$ Text caption

Game State $\rightarrow$ N.Net $\rightarrow$ Next move

What are these boxes?
What are deep neural networks?

What are these boxes?
Each of these boxes is actually a function
What are deep neural networks?

What are these boxes?
Each of these boxes is actually a function. It can be approximated by a neural network.
Agenda

Artificial Intelligence (AI): A brief intro and recent breakthroughs with Deep learning

AI for Social Media Marketing: Cases studies
- Visual Listening In: Extracting Brand Image Portrayed on Social Media (Liu et al. 2020)
- Identify Customer Needs from User-Generated Content (Timoshenko and Hauser 2019)
- Supporting Content Marketing with Natural Language Generation (Reisenbichler et al. 2022)

Brainstorming and group work

Sharing group work and conclude
Insights/Problems

Product Design
- Understand customers needs

Branding
- Measure brand perceptions

Advertising and social media
- Generate social media posts and ads

Influencer marketing
- What makes a good influencer video ads?
  ...

Data

Text  Image  Video
Identify important problems
- opportunities to use data and analytics to create value for consumers or firms

What is the ideal data to solve this problem?
- e.g., internal data, external data
Training data vs. Production/Application data

What is the right method?
How to evaluate the method and interpret results?
Visual Listening In: Extracting Brand Image Portrayed on Social Media

Liu Liu, Daria Dzyabura, Natalie Mizik
How is my brand portrayed in consumer photos?
Photos Are the New Social Conversation

Image-based social media platforms are on the rise
- E.g., Instagram has 700 million monthly active users
- 95 million photos/videos uploaded daily\(^1\)

   Photo credit: Tom McGrath@Crimson Hexagon
Consumers Associate Brands with Contexts

Consumers hashtag brands and depict interactions with brands
- E.g., 53 million posts on Instagram with #nike

Link brands with usage context and experiences

#eddiebauer  #prada
Social Media Marketing

“Listen in” on consumer conversations

- E.g., Archak, Ghose, & Ipeirotis, 2011; Lee & Bradlow, 2011; Netzer et al., 2012; Tirunillai & Tellis, 2012, 2014; Liu, Lee, & Srinivasan, 2019; Timoshenko and Hauser, 2019

→ Text mining

→ Functional attributes of products
This Research

Measure brand image portrayed on consumer photos
This Research

Measure brand image portrayed on consumer photos

“How are brands portrayed along *intangible brand attributes*?”

#eddiebauer → **rugged**

#prada → **glamorous**
Why Do We Care?

Consumers are co-creating brand image on social media

Allow firms to correct, or leverage, or identify new opportunities for brand positioning and differentiation

```
glamorous

Consumer photos

Perception survey

Firm

rugged
```
How to Measure Brand Attributes from Photos?

Challenging: Data is very unstructured.

Require new methods!
## Contributions

<table>
<thead>
<tr>
<th>Do consumers’ photos contain brand image/perception info?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes - Create a new photo-based metric about consumers’ brand perception</td>
</tr>
</tbody>
</table>

### How to measure?

Measure brand attributes from photos using Deep learning
- Convolutional neural networks for brand attributes

### What insights are generated from social media?

Application to Instagram brand photos
- Consumer photos vs. Firm photos vs. Brand Perception survey
  - Brand image on social media reflects perception survey
  - Identify gaps in positioning strategy
Steps

Measure brand attributes from photos
- Data: Photos labeled with brand attributes
  - glamorous, rugged, healthy, fun
- Algorithms: multi-label image classification

Application to Instagram brand photos
- Data: Apparel and beverage photos on Instagram
- Metrics: Consumer photos vs. Firm photos vs. Perception survey
- Empirical studies
  i. Product category level consistency
  ii. Brand maps
  iii. Gaps in positioning strategy
Steps

Measure brand attributes from photos
- Data: Photos labeled with brand attributes
  - glamorous, rugged, healthy, fun
- Algorithms: multi-label image classification

Application to Instagram brand photos
- Data: Apparel and beverage photos on Instagram
- Metrics: Consumer photos vs. Firm photos vs. Perception survey
- Empirical studies
  i. Product category level consistency
  ii. Brand maps
  iii. Gaps in positioning strategy
Steps

Measure brand attributes from photos
- Data: Photos labeled with brand attributes
  - glamorous, rugged, healthy, fun
- Algorithms: multi-label image classification

Application to Instagram brand photos
- Data: Apparel and beverage photos on Instagram
- Metrics: Consumer photos vs. Firm photos vs. Perception survey
- Empirical studies
  i. Product category level consistency
  ii. Brand maps
  iii. Gaps in positioning strategy
Steps

Measure brand attributes from photos
- Data: Photos labeled with brand attributes
  - glamorous, rugged, healthy, fun
- Algorithms: multi-label image classification

Application to Instagram brand photos
- Data: Apparel and beverage photos on Instagram
- Metrics: Consumer photos vs. Firm photos vs. Perception survey
- Empirical studies
  i. Product category level consistency
  ii. Brand maps
  iii. Gaps in positioning strategy
Multi-label Image Classification on Brand Attributes

1. Images labeled with brand attributes
2. Algorithms that learn mapping between images and attributes

Does it convey ruggedness? Does it convey glamour? Does it convey healthiness? Does it convey fun?
Collect Images Labeled with Brand Attributes

Need positive and negative instances for each brand attribute

Collect data using Flickr search engine
Query attributes (e.g., healthy) and antonyms (e.g., unhealthy)
Example Images

- glamorous
- drab
- rugged
- gentle
- healthy
- unhealthy
- fun
- dull

16,368 images in total
Example Images

- glamorous
- drab
- rugged
- gentle
- healthy
- unhealthy
- fun
- dull

16,368 images in total
Algorithms: Mapping Between Images and Attributes

For example, for “rugged”

Very unstructured
Deep Learning

Automatic feature extraction
Represent hierarchy of concepts (Bengio et al. 2015)
Deep Learning: How Does It Work?

Model complex non-linear relationship via many simple non-linear transformations one after another

\[
y = f(\sum_{j=1}^{K} w_j x_j)
\]

\[
F(x, W) = f(w^K, f(w^{K-1}, f(...f(w^0, x)\ldots))
\]

Estimation: back propagation (gradient descent)
A Multi-label Convolutional Neural Network (ConvNet)

Brand attribute prediction

Image in pixels: 227 x 227 x 3
Train ConvNets with Transfer Learning

Require large amount of data

Transfer learning
  Transfer parameters from one domain to another
  Initialize with parameters from pre-trained models
    ImageNet model (Krizhevsky et al., 2012): object classification
    Flickrstyle model (Karayev et al., 2013): style recognition
  Fine-tune the model on our training data
Models Trained on GPU

Code and train model using Caffe deep learning framework

5000 iterations in a K80 GPU on a university high performance cluster

80% training data, 10% validation, 10% hold-out sample
  - Pick the iteration with the lowest loss in a validation set
Out-of-Sample Predictive Performance

<table>
<thead>
<tr>
<th></th>
<th>Multi-label ConvNet</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Accuracy</td>
</tr>
<tr>
<td>Glamorous</td>
<td>88.6%</td>
</tr>
<tr>
<td>Rugged</td>
<td>91.3%</td>
</tr>
<tr>
<td>Healthy</td>
<td>89.9%</td>
</tr>
<tr>
<td>Fun</td>
<td>89.4%</td>
</tr>
<tr>
<td>Mean</td>
<td>89.8%</td>
</tr>
</tbody>
</table>

*Note: 80% training, 10% validation, 10% hold-out sample*
Steps

Measure brand attributes from photos
- Data: Photos labeled with brand attributes
  - glamorous, rugged, healthy, fun
- Algorithms

Application to Instagram brand photos
- Data: Apparel and beverage photos on Instagram
- Metrics: Consumer photos vs. Firm photos vs. Perception survey
- Empirical studies
  i. Product category level consistency
  ii. Brand maps
  iii. Gaps in positioning strategy
Brand Photos on Instagram

56 brands from Apparel and Beverage categories

Consumer: photos on Instagram (#brand)
- About 2,000 per brand
- 114,367 total

Firms: photos on official Instagram accounts
- 72,089 total
Model Performance on Instagram Data Sample

Data Sample (600 images in total, 150 images per attribute)

<table>
<thead>
<tr>
<th>Group 1</th>
<th>Group 2</th>
<th>Group 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attribute is present</td>
<td>Ambiguous</td>
<td>Attribute is not present</td>
</tr>
<tr>
<td>50 images</td>
<td>50 images</td>
<td>50 images</td>
</tr>
</tbody>
</table>

Human judges

Level-3 US judges on Figure Eight platform\(^1\)
20 judges for each attribute and each image

---

1. Previously Crowdflower
Model vs. Human Judges: Overall Agreement

Note: 600 images (150 images for each attribute), 12,000 total judgments (20 judgments for each image and attribute). The model-based label is equal to 1 if the model-based probability estimate for attribute presence is greater than 50%, and zero otherwise. The human-based label for an image is equal to 1 if the majority of the judges indicate attribute presence, and zero otherwise. Agreement is the percentage of images for which the majority of human judges evaluating an image assign this image the same label as our model. The total cost of data collection is $288.96, with an average cost per judgment of $0.024.

### Table. Aggregate Model Performance According to Human-Based Image Labels

<table>
<thead>
<tr>
<th>Attribute</th>
<th>AUC: Model vs. the Majority Vote of Human Judges</th>
<th>Agreement: Model vs. the Majority Vote of Human Judges</th>
<th>Agreement: A Single Human Judge vs. the Majority Vote of Human Judges, Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>glamorous</td>
<td>0.93</td>
<td>83%</td>
<td>85%</td>
</tr>
<tr>
<td>rugged</td>
<td>0.96</td>
<td>85%</td>
<td>83%</td>
</tr>
<tr>
<td>healthy</td>
<td>0.91</td>
<td>78%</td>
<td>80%</td>
</tr>
<tr>
<td>fun</td>
<td>0.94</td>
<td>84%</td>
<td>80%</td>
</tr>
<tr>
<td><strong>Average:</strong></td>
<td><strong>0.94</strong></td>
<td><strong>82.5%</strong></td>
<td><strong>82%</strong></td>
</tr>
</tbody>
</table>
Image-Based Brand Image (IBBI) Metric

Compute the average probability of brand \( j \) images that express the brand attribute \( p \):

\[
IBBI_{ba} = \frac{\sum_{i=1}^{N(b)} \Pr \left( y_n^{(b)}(a) = 1 \bigg| X_n^{(b)} \right)}{N(b)}.
\]

For example

\[
IBBI_{Prada, \text{ rugged}} = 0.07
\]

\[
IBBI_{Eddie Bauer, \text{ rugged}} = 0.17
\]
Three Different but Related Brand Metrics

**Consumer-created brand photos**
- Brand image conveyed from consumer photos
- Context, usage, consumption experience

**Firm-created brand photos**
- Part of firms’ marketing effort to create brand identity

**Brand perception survey**
- Young and Rubicam’s Brand Asset Valuator (BAV)
- How do consumers perceive brands?
Empirical Studies and Insights

1. Product category level
   - Check consistency between different brand metrics

2. Brand maps
   - Compare brand maps created from different brand metrics

3. Identify Gaps in positioning strategy
   - A case study of underwear brands
Empirical Studies and Insights

1. Product category level
   - Check consistency between different brand metrics

2. Brand maps
   - Compare brand maps created from different brand metrics

3. Identify Gaps in positioning strategy
   - A case study of underwear brands
# Consistency Between Different Brand Metrics

Table. Correlation Analyses of Model Predictions for Consumer and Firm-Created Images on Instagram and the BAV Survey-Based Measures of Brand Perceptions

<table>
<thead>
<tr>
<th>Apparel</th>
<th>Consumer Images vs. Firm Images</th>
<th>Consumer Images vs. BAV</th>
<th>Firm Images vs. BAV</th>
</tr>
</thead>
<tbody>
<tr>
<td>GLAMOROUS</td>
<td>0.7838***</td>
<td>0.5519***</td>
<td>0.6100***</td>
</tr>
<tr>
<td>RUGGED</td>
<td>0.9122***</td>
<td>0.5467**</td>
<td>0.5035**</td>
</tr>
<tr>
<td>HEALTHY</td>
<td>0.4680**</td>
<td>0.1794</td>
<td>0.3225*</td>
</tr>
<tr>
<td>FUN</td>
<td>0.6061***</td>
<td>0.3583*</td>
<td>0.2883</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Beverages</th>
<th>Consumer Images vs. Firm Images</th>
<th>Consumer Images vs. BAV</th>
<th>Firm Images vs. BAV</th>
</tr>
</thead>
<tbody>
<tr>
<td>GLAMOROUS</td>
<td>0.5518**</td>
<td>0.4568**</td>
<td>0.6582***</td>
</tr>
<tr>
<td>RUGGED</td>
<td>0.8259***</td>
<td>0.3596*</td>
<td>0.4708*</td>
</tr>
<tr>
<td>HEALTHY</td>
<td>0.7370***</td>
<td>0.6976***</td>
<td>0.4766**</td>
</tr>
<tr>
<td>FUN</td>
<td>0.3775*</td>
<td>0.1791</td>
<td>0.2584</td>
</tr>
</tbody>
</table>

**Note:** We see further increases in correlations between consumer and firm-image based data and survey-based measures of brand perceptions when we substitute BAV measures with our survey data from Instagram users. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.**
Table. Correlations between Model Predictions for Consumer and Firm-Created Images on Instagram and Survey-Based Measures of Brand Perceptions from Instagram Users

<table>
<thead>
<tr>
<th></th>
<th>Apparel BAV vs. Survey</th>
<th>Consumer Images vs. Survey</th>
<th>Firm Images vs. Survey</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>GLAMOROUS</strong></td>
<td>0.9503***</td>
<td>0.5824***</td>
<td>0.6325***</td>
</tr>
<tr>
<td><strong>RUGGED</strong></td>
<td>0.9338***</td>
<td>0.6831***</td>
<td>0.6630***</td>
</tr>
<tr>
<td><strong>HEALTHY</strong></td>
<td>0.8600***</td>
<td>0.0842</td>
<td>0.1941</td>
</tr>
<tr>
<td><strong>FUN</strong></td>
<td>0.6486***</td>
<td>0.5672***</td>
<td>0.4914**</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Beverages BAV vs. Survey</th>
<th>Consumer Images vs. Survey</th>
<th>Firm Images vs. Survey</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>GLAMOROUS</strong></td>
<td>0.9238***</td>
<td>0.5001**</td>
<td>0.5743**</td>
</tr>
<tr>
<td><strong>RUGGED</strong></td>
<td>0.5485***</td>
<td>0.7899***</td>
<td>0.6645***</td>
</tr>
<tr>
<td><strong>HEALTHY</strong></td>
<td>0.9482***</td>
<td>0.7127***</td>
<td>0.5350**</td>
</tr>
<tr>
<td><strong>FUN</strong></td>
<td>0.8714***</td>
<td>0.2130</td>
<td>0.3648*</td>
</tr>
</tbody>
</table>

**Note:** the shaded cells represent instances of improvement over the correlations between consumer- and firm-image based data and BAV reported in Table 2. Average number of respondents per brand is 62. Cost of the survey data collection is $346.10. *p < 0.05, **p < 0.01, ***p < 0.001.
→ Overall convergence

→ Brand image portrayed on social media, i.e., our IBBI measure, reflect consumers’ brand perceptions

→ What I “say” vs. What I “post”
Empirical Studies and Insights

1. Product category level
   - Check consistency between different brand metrics

2. Brand maps
   - Compare brand maps created from different brand metrics

3. Identify Gaps in positioning strategy
   - A case study of underwear brands
Beverage Brand Map from Consumer Photos

Dim 1 (71.6%)

Dim 2 (18.6%)
Beverage Brand Map from Consumer Photos

Dim 1 (71.6%)

Dim 2 (18.6%)
Empirical Studies and Insights

1. Product category level
   - Check consistency between different brand metrics

2. Brand maps
   - Compare brand maps created from different brand metrics

3. Identify Gaps in positioning strategy
   - A case study of underwear brands
Identify Gaps in Positioning

Consumer vs. Firm: e.g., glamorous in Apparel
Identify Gaps in Positioning

Consumer vs. Firm: e.g., glamorous in Apparel
**Validate Differences using Human Judges**

Victoria’s secret, Joe Boxer, and Hanes

500 random consumer images and 500 random firm images

10 human judgments per image

<table>
<thead>
<tr>
<th>Brand</th>
<th>Consumer Images (N=500)</th>
<th>Firm Images (N=500)</th>
<th>Human Judges</th>
</tr>
</thead>
<tbody>
<tr>
<td>Victoria’s Secret</td>
<td>0.39</td>
<td>0.43</td>
<td>0.47</td>
</tr>
<tr>
<td>Joe Boxer</td>
<td>0.20</td>
<td>0.21</td>
<td>0.11</td>
</tr>
<tr>
<td>Hanes</td>
<td>0.16</td>
<td>0.30</td>
<td>0.12</td>
</tr>
</tbody>
</table>
Model Predictions and Human Judgements are Consistent

Consistent difference between consumer and firm portrayal
Consistent difference between brands

<table>
<thead>
<tr>
<th>(b) IBBI Score Differentials between Firm and Consumer Images: Difference (T-stat, p-value)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
</tr>
<tr>
<td>Victoria’s Secret: Firm vs. Consumer images</td>
</tr>
<tr>
<td>Joe Boxer: Firm vs. Consumer images</td>
</tr>
<tr>
<td>Hanes: Firm vs. Consumer images</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>(c) IBBI Score Differentials between Brands: Difference (T-stat, p-value)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
</tr>
<tr>
<td>Victoria’s Secret vs. Joe Boxer</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Victoria’s Secret vs. Hanes</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Joe Boxer vs. Hanes</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>
**Problem**

- Measure brand perceptions from consumer-created brand photos
  - Perceptual map
  - Identify gaps between firm and consumers
  - Track consumers’ brand perceptions in real time

**Data**

- Training data: “labeled” Flicker data
- Production/Application data: Instagram data (consumer vs. firm)

**Method**

- Convolutional neural network
- Evaluation:
  - Out-of-sample prediction performance
  - Compare with traditional measures
Identifying Customer Needs from User-Generated Content

Artem Timoshenko, John Hauser
Problem: Identify Customer Needs

Examples of customer needs for oral care products from interviews and focus groups:

- Able to make my teeth look whiter
- Oral care items I carry around are easy to keep clean
- Oral care items that match my bathroom décor

Slides credit: Artem Timoshenko
Customer Needs for Oral Care

HYGIENE

I am sure that the brush is clean every time I put it in my mouth

- Has germ resistant properties
- A cap that keeps the bristles clean and germs out
- Prevents bacteria from getting from a used brush to other places (e.g., my mouth, my hands, other objects)
- Assured that I am not putting any unknown or harmful chemicals into my mouth (e.g., antibacterial agents)
- Cap won't come off unexpectedly
- Packaging that keeps the IDB fresh and clean

Know how to clean the brush and when to replace it

- Can easily tell how clean the brush is
- Able to know with certainty when it is time to change the brush
- A brush that lasts a long time before needing to be replaced
- My IDB stays clean enough that I can reuse it multiple times
- Always know the proper way to clean my brush

Keeps my teeth white and breath fresh

- Helps me keep my teeth white
- Cleaning tools that do not leave a bitter aftertaste in my mouth
- My IDB can also help freshen my mouth (e.g., contains flavor, toothpaste)

I can avoid having to put my fingers in my mouth or on the bristles

- Can easily clean my teeth without getting my fingers in my mouth
- Prevents my fingers from touching the bristles while cleaning my teeth

PACKAGING

Product and package are environmentally friendly

- An environmentally friendly package
- An environmentally friendly product

I can have an IDB with me whenever I need one, at home or away

- Able to always have a pick when I need it (i.e., away from home, in the car, in my purse)
- An interdental device that I can bring with me wherever I go
- Able to discreetly clean my teeth even when out in public

A package that makes it easy to store and access the tools

- Can easily store IDBs
- The IDBs stay organized once I have opened the packaging (i.e., they're not loose)
- Packaging that is easy to open and re-close
- A container that makes it easy to remove the IDBs one at a time
- A sturdy package that prevents breakage
- IDB can dry while storing it for reuse

Slides credit: Artem Timoshenko
Illustrative Example

Amazon Review:

“I replaced an old brush with a new one BUT the description doesn’t say that this model no longer has a 30 second timer. The brush shuts off after 2 minutes but the 30 second timer is missing. I would not have purchased this product if I had known.”

Slides credit: Artem Timoshenko
Illustrative Example

Amazon Review:

“I replaced an old brush with a new one BUT the description doesn’t say that this model no longer has a 30 second timer. The brush shuts off after 2 minutes but the 30 second timer is missing. I would not have purchased this product if I had known”.

I know the right amount of time to spend on each step of my oral care routine

Slides credit: Artem Timoshenko
Illustrative Example

Amazon Review:

“I replaced an old brush with a new one BUT the description doesn’t say that this model no longer has a 30 second timer. The brush shuts off after 2 minutes but the 30 second timer is missing. I would not have purchased this product if I had known.”

I know the right amount of time to spend on each step of my oral care routine

Slides credit: Artem Timoshenko
Customer Needs from UGC

Amazon Review:

“I replaced an old brush with a new one BUT the description doesn’t say that this model no longer has a 30 second timer. The brush shuts off after 2 minutes but the 30 second timer is missing. I would not have purchased this product if I had known”.

I know the right amount of time to spend on each step of my oral care routine

Slides credit: Artem Timoshenko
Figure A.1. Demonstration of the Application of the Proposed Machine-Learning Hybrid Approach to an Amazon Review

**Input**

**Product Review**

Replaced an old brush with a new one BUT they neglect to say that this model no longer has the 30 second timer. It does shut off after 2 minutes but the 30 second timer is only in more costlier models. I would not have purchased and used this one if I had known.

**Machine Learning**

**Sentences**

1. Replaced an old brush with a new one BUT they neglect to say that this model no longer has the 30 second timer.
2. It does shut off after 2 minutes but the 30 second timer is only in more costlier models.
3. I would not have purchased and used this one if I had known.

**Informative Sentences**

1. Replaced an old brush with a new one BUT they neglect to say that this model no longer has the 30 second timer.
2. It does shut off after 2 minutes but the 30 second timer is only in more costlier models.
3. I would not have purchased and used this one if I had known.

**Informative & Non-Repetitive Sentences**

1. Replaced an old brush with a new one BUT they neglect to say that this model no longer has the 30 second timer.
2. It does shut off after 2 minutes but the 30 second timer is only in more costlier models.
3. I would not have purchased and used this one if I had known.

**Human Judgment**

**Customer Needs**

I know the right amount of time to spend on each step of my oral care routine.

Slides credit: Artem Timoshenko
Figure 1. System Architecture for Identifying Customer Needs from UGC

1. Split UGC into sentences
2. Remove stop-words, punctuation, etc.
3. Identify frequent combinations of words

Train Word Embeddings

1. Estimate word embeddings on a large UGC corpus (skip-gram model)

Identify Informative Content

1. Label a small sample of sentences into informative/non-informative
2. Train a machine learning classifier (CNN)
3. Identify informative content in the rest of the corpus

Sample Diverse Content

1. Average word embeddings to create sentence embeddings
2. Cluster sentence embeddings using Ward’s algorithm
3. Sample one sentence from each of Y clusters

Manually Extract Customer Needs

1. Review the Y selected sentences and formulate customer needs

Slides credit: Artem Timoshenko
Support Content Marketing with Natural Language Generation

Martin Reisenbichler, Thomas Reutterer, David A. Schweidel, and Daniel Dan
Problem: How to write search engine (SE) optimized content?

1. Identifying current top performing content
2. Machine learning to capture linguistic patterns (e.g., keyword density, readability)
3. Creation of unique content that mirrors the linguistic patterns of top-performing content
4. Human editing of content before publication

Slide credit: David A. Schweidel
Agenda

Artificial Intelligence (AI): A brief intro and recent breakthroughs with Deep learning

AI for Social Media Marketing: Cases studies
- Visual Listening In: Extracting Brand Image Portrayed on Social Media (Liu et al. 2020)
- Identify Customer Needs from User-Generated Content (Timoshenko and Hauser 2019)
- Supporting Content Marketing with Natural Language Generation (Reisenbichler et al. 2022)

Brainstorming and group work

Sharing group work and conclude
Problem

Identify one important problem
- opportunities to use data and analytics to create value for consumers or firms

Data

What is the ideal data to solve this problem?
- e.g., internal data, external data
  Training data vs. Production/Application data

Method

What is the right method?
How to evaluate the method and interpret results?
Agenda

Artificial Intelligence (AI): A brief intro and recent breakthroughs with Deep learning

AI for Social Media Marketing: Cases studies
- Visual Listening In: Extracting Brand Image Portrayed on Social Media (Liu et al. 2020)
- Identify Customer Needs from User-Generated Content (Timoshenko and Hauser 2019)
- Supporting Content Marketing with Natural Language Generation (Reisenbichler et al. 2022)

Brainstorming and group work

Sharing group work and reflections (e.g., pitfalls)
Causation (vs. Correlation)
Shifting from Model-Centric to Data-Centric AI

Conventional model-centric approach:

\[
\text{AI System} = \text{Code} + \text{Data} \\
\text{(model/algorithm)}
\]

Data-centric approach:

\[
\text{AI System} = \text{Code} + \text{Data}
\]
Data Quality is Important

Data is food to AI

Slide credit: Andrew Ng, “MLOps: From Model-centric to Data-centric AI”
Iguana Detection Example

Labeling instruction:
Use bounding boxes to indicate the position of iguanas

Slide credit: Andrew Ng, “MLOps: From Model-centric to Data-centric AI”
From Big Data to Good Data

- Defined consistently (definition of labels $y$ is unambiguous)

- Cover of important cases (good coverage of inputs $x$)

- Has timely feedback from production data (distribution covers data drift and concept drift)

- Good governance (reasonably free from bias; satisfies privacy; data provenance/lineage, regulatory requirements)

Slide credit: Andrew Ng, “MLOps: From Model-centric to Data-centric AI”
Data-centric AI

Model-centric AI
How can you change the model (code) to improve performance?

Data-centric AI
How can you systematically change your data (inputs x or labels y) to improve performance?

Slide credit: Andrew Ng, “MLOps: From Model-centric to Data-centric AI”
Data-centric AI

Model-centric AI
How can you change the model (code) to improve performance?

Data-centric AI
How can you systematically change your data (inputs x or labels y) to improve performance?

Make high quality data available through all stages of the ML project lifecycle

Scope project
Collect data
Train model
Deploy in production

How do I define and collect my data?
How do I modify data to improve model performance?
What data do I need to track concept/data drift?

Slide credit: Andrew Ng, “MLOps: From Model-centric to Data-centric AI”
Tips for Data-centric AI

1. Make the labels $y$ consistent
2. Use multiple labelers to spot inconsistencies
3. Clarify labeling instructions by tracking down ambiguous examples
4. Toss out noisy examples. More data is not always better
5. Use error analysis to focus on subset of data to improve

Slide credit: Andrew Ng, “MLOps: From Model-centric to Data-centric AI”
Thank you!

Liu Liu (liu.liu-1@colorado.edu)

https://www.colorado.edu/faculty/liu-liu/
