

Machine Learning @ Amazon

Rajeev Rastogi

Director, Machine Learning



Numerous ML Applications

- Abuse/Fraud
- Address Quality
- Packaging
- Substitutes Prediction
- Payment Instruments Success Prediction

Operations



- Product Recommendations
- Product Search
- Product Ads
- Customer Targeting
- Lending

Customers



- Demand Forecasting
- Pricing
- Fraud Detection
- Seller Lead Generation

Sellers



- Product Classification
- Duplicate Products Detection
- Product Attributes Mismatch Detection
- Attributes Extraction from Titles/Images

Catalog



- Named-Entity Extraction
- Reviews Summarization
- Reviews Ranking and Insights
- Fraudulent Reviews

Text



- Visual Search
- Product Image Quality
- Celebrity Faces
- Incorrect Postures
- Brand Tracking

Images



- Automatic Speech Recognition
- Natural Language Understanding
- Dialog Management

Speech



- Predicting Resource Requirements
- DDoS Detection
- Reputation Computation of MTurk Workers

AWS



Address Quality

Problem

- Identify and correct poor quality addresses

Fake addresses

- "Pakistan, Sharjah, Lahore"
- "asdf asdf"

Wrong Pincodes

- "28th floor, World Trade Centre, Bangalore, 400001"
- "28th floor, World Trade Centre, Bangalore, 560001"

Incomplete address

- "Near Orion Mall, Malleswaram, Bangalore"
- "Near Bus Stand, Sambhaji Chowk, Nasik"

Spelling errors

- Shop 999 WorliVullage janaata coloniny, 7 1 328
- kodamachalivaristreetpaturutadapalliqu dem

Product Packaging

Problem

- Determine the most cost-effective packaging to use for a product



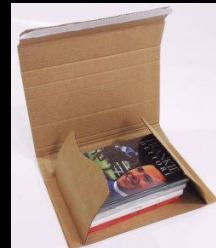
SIOC



Poly bag



Bubble package



Variable
depth box



Corrugated box

Low

Packaging Cost

High

Product Substitutes

Problem

- *Identify substitutes for a product*

amazon.in
prime

Deliver to Jagann...
Bangalore 560032

Shop by Category - Buy Again Jagann... 's Amazon.in Today's Deals Amazon Pay Sell Customer Service

Electronics Mobiles & Accessories Laptops & Accessories TV & Home Entertainment Audio Cameras Computer Peripherals Smart Technology Musical Instruments Office & Stationery

Back to search results for "iphone 7"

Apple iPhone 7 (Black, 2GB RAM, 32GB Storage)

by Apple

★★★★☆ 3,405 customer reviews | 921 answered questions

M.R.P.: ₹ 43,300.00
Price: ₹ 39,480.00
You Save: ₹ 12,890.00 (25%)
Inclusive of all taxes

✓prime FREE Delivery by Tomorrow 9pm
Order within 14 hrs 6 mins. Details

EMI starts at ₹1,858 per month. Options -
Available at a lower price without free Prime shipping from other sellers.

In stock.

Sold by Appario Retail Private Ltd (4.7 out of 5 | 113,938 ratings) and Fulfilled by Amazon. Gift-wrap available.

10 offers from ₹ 39,470.00

Colour: Black

Stop on page 1770

Share

100% Purchase Protection
Original Products | Secure Payments

With Exchange
Up to ₹ 16,000.00 off

Without Exchange
₹ 39,480.00 ₹ 42,370.00

Include OnsiteGo 15 Months
Total Protection Plan for
Mobiles (Rs. 30,001 to Rs. ...
for ₹ 4,479.00

Deliver to Jagannathan -
Bangalore 560032

Quantity: 1

Add to Cart

Customers who viewed this item also viewed

Apple iPhone 7 Plus (Silver, 128GB)
★★★★☆ 917
₹ 61,045.00 online

Apple iPhone 7 Plus (Gold, 128GB)
★★★★☆ 917
₹ 60,999.00 online

Apple iPhone XR (Black, 3GB RAM, 64GB Storage)
★★★★☆ 33
₹ 76,747.00 online

Samsung Galaxy Note 8 (Midnight Black, 6GB RAM, 64GB Storage) with Offers
★★★★☆ 1,400

Samsung Galaxy A4+ (Black, 6GB RAM, 64GB Storage) with Offers
★★★★☆ 7,836

Samsung Galaxy S8 Plus (Orchid Gray, 64 GB) (4 GB RAM)
★★★★☆ 7

OnePlus 5T (Mirror Black, 6GB RAM, 128GB Storage)
★★★★☆ 4,564
₹ 37,999.00 online

Apple iPhone 6S Plus (Gold, 32GB)
★★★★☆ 324
₹ 64,999.00

Page 5 of 5 Start over

Product Recommendations

Problem

- *Recommend products to customers that match their preferences*

The screenshot shows an Amazon India product page for 'Trust basket Round Ribbed Railing Planters (Green, Yellow, Red, Blue, Orange) - Set of 5'. The product is displayed with a large image of the planters and a smaller image of the product details. The price is ₹ 999.00, and it has a 4.5-star rating from 9 customer reviews. The page includes a navigation bar with 'amazon.in prime' and 'amazon prime Exclusive deals'. Below the product image, there is a section titled 'Your recently viewed items and featured recommendations' which is highlighted with a red box. This section displays a carousel of related products, including 'Trust basket Single Pot Railing Planter', 'Perfarmers Seed Starter', 'Perfarmers Plastic Plastic Stack-A-Pot Hanging Set', 'COIR GARDEN Coco Pole', 'Cappi Peace Lily Plant', 'Seedcare Herbs Seeds', 'Minerve Natural Vertical Pot', and 'Kraft Seeds Fragrant White or Tuberosa Flower Bulbs'. The page also features a '100% Purchase Protection' badge and a 'FREE Delivery by Wednesday 3pm' badge.

amazon.in prime | amazon prime Exclusive deals

Deliver to Jagan... Bangalore 560032 | Shop by Category | Buy Again | Jagan... % Amazon.in | Today's Deals | Amazon Pay | Sell | Customer Service

Amazon Home | Kitchen & Home Appliances | Large Appliances | Kitchen & Dining | Furniture | Home Furnishing | Home Decor | Home Improvement | Garden & Outdoor | Storage & Organisation | Lighting

Back to search results for "planters"

Trust basket Round Ribbed Railing Planters (Green, Yellow, Red, Blue, Orange) - Set of 5

by Trust basket

★★★★☆ 9 customer reviews

M.R.P.: ₹ 4,759.00
Price: ₹ 999.00
You Save: ₹ 760.00 (43%)
Inclusive of all taxes

prime FREE Delivery by Wednesday 3pm
Order within 19 hrs 5 mins. Details

Pay on Delivery (Cash/Card) eligible What is this?
In stock.

Sold by TrustBasket® (4.5 out of 5 | 692 ratings) and Fulfilled by Amazon. Gift-wrap available.

2 offers from ₹ 999.00

- Colour: Green, Yellow, Red, Blue, Orange
- Product Dimensions: Length: 15 cm, Width: 15 cm, Height: 15 cm
- Material: Metal
- Package contents: 5 planters
- Plants and flowers are not included

Share | 100% Purchase Protection | Original Products | Secure Payments

Deliver to Jagan... Bangalore 560032

Quantity: 1

Add to Cart | Buy Now

Add to Wish List

Other Sellers on Amazon

2 offers from ₹ 999.00

Your recently viewed items and featured recommendations
Inspired by your browsing history

Page 1 of 7

Trust basket Single Pot Railing Planter, Set of 5 (Red, Yellow, Blue, Orange, Green) ★★★★★ 412 ₹ 149.00 /prime

Perfarmers Seed Starter 250ml Organic Liquid For All Seeds Maximum Sprout ★★★★★ 56 ₹ 149.00 /prime

Perfarmers Plastic Plastic Stack-A-Pot Hanging Set (Dark Green, 5-Pieces) ★★★★★ 93 ₹ 1,200.00 /prime

COIR GARDEN Coco Pole Moss and Coir Stick for Money Plant Support, 3ft, Brown - Pack of 4 Pieces ★★★★★ 4 ₹ 349.00

Cappi Peace Lily Plant (Spathiphyllum) Green ★★★★★ 47 ₹ 349.00

Seedcare Herbs Seeds Combo (Pack of 40) ★★★★★ 15 ₹ 471.00 /prime

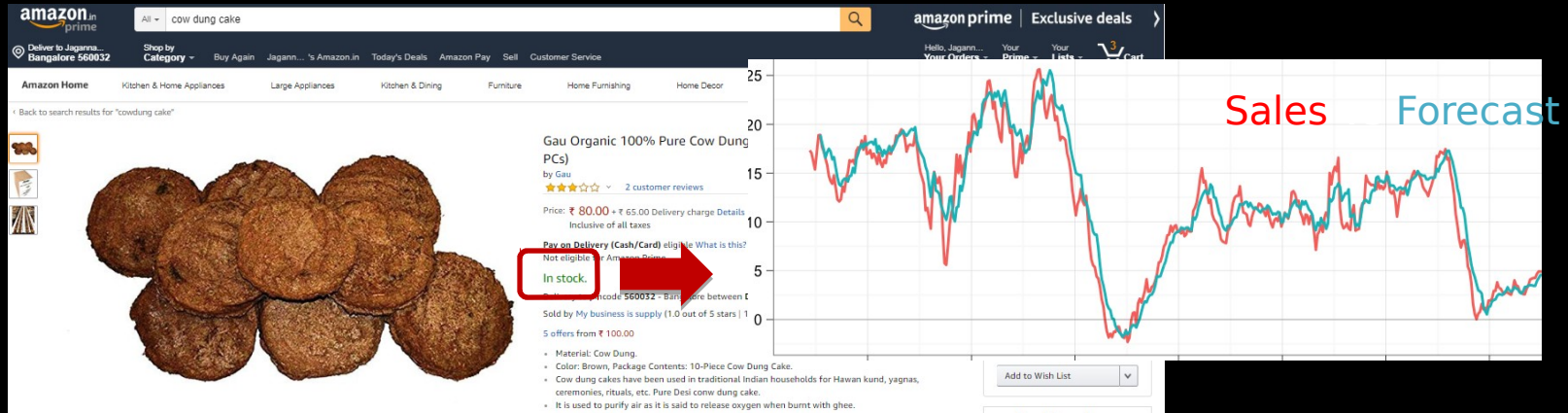
Minerve Natural Vertical Pot (Set Of 15 - 3 Colors) (Green/Brown) ★★★★★ 20 ₹ 799.00 /prime

Kraft Seeds Fragrant White or Tuberosa Flower Bulbs (White, Pack of 8 Bulbs) ★★★★★ 144 ₹ 161.00 /prime

Product Demand Forecasting

Problem

- Given past sales of a product in every region, predict regional demand up to one year into the future



Product Classification

Problem

- *Classify each product into the appropriate leaf node in product taxonomy*

amazon prime

Deliver to Jagann...
Bangalore 560032

Shop by
Category

Buy Again Jagann... 's Amazon.in Today's Deals Amazon Pay Sell Customer Service

Amazon App Contest
Win a OnePlus 6T*

Hello, Jagann...
Your Orders

Your
Prime


Your
Lists


Cart

Amazon Fashion

WOMEN MEN KIDS BAGS & LUGGAGE SPORTSWEAR BRANDS SALES & DEALS 30 DAY RETURNS Restrictions Apply

Jewellery > Women > Bangles & Bracelets > Sasitrends Antique Gold Plated Pearl Studded Bangles For Women





Sasitrends

Sasitrends Antique Gold Plated Pearl Studded Bangles For Women

★★★★☆ 4 customer reviews

M.R.P.: ₹ 3,000.00

Price: ₹ 899.00 FREE Delivery. Details

You Save: ₹ 2,101.00 (70%)

Inclusive of all taxes

Not eligible for Amazon Prime.

In stock.

Delivery to pincode 560032 - Bangalore between Dec 1 - 3. Details

Sold by Sasitrends™ (4.5 out of 5 stars | 62 ratings) and Delivered by Amazon.

Length:

2.4 inches

- Package content- One pair of bangle.
- Base material type- Brass material.
- Trend Style/ Fashion- Easy to wear, light in weight & gives you a rich look. Designer pair of bangles that will make you feel Royal. Jewellery itself has no value, but the jewels which can make women shining. Sasitrends is committed to provide the best jewellery and the best customer services to our customers; your feedback is our motivations to improve.
- Occasion Type- This must-have design can be worn on any Occasion to boost up your style factor and enhance the ultra beauty factor. Indian, or Western, this Bangle gives a glamorous looks. Perfect Gift for Any Occasion for Yourself and Your Dear Ones
- Care instructions- No direct perfume, deo, water, and spray on the jewellery. First wear your

Share

₹ 899.00 + FREE Delivery

Deliver to Jagannathan -
Bangalore 560032

Quantity: 1

Add to Cart

Buy Now

Add to Wish List

Have one to sell? Sell on Amazon

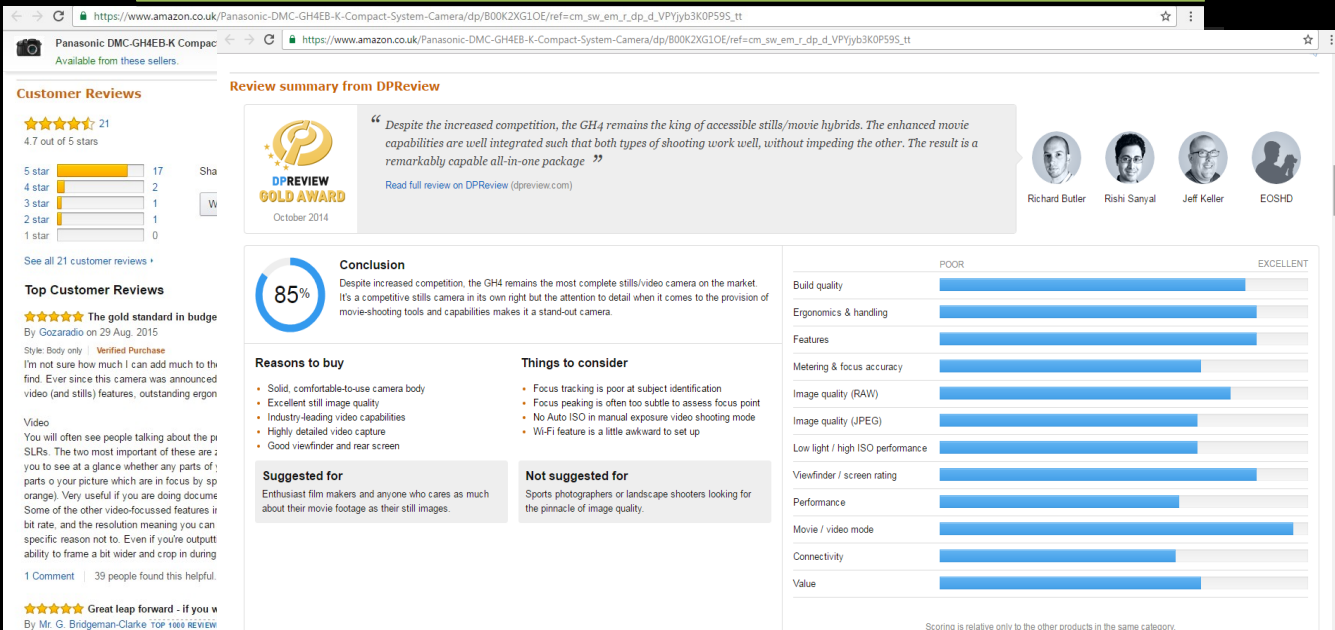
- *Identify duplicate product listings in Amazon catalog*



Insights Extraction from Reviews

Problem

- Extract fine-grained attribute ratings from product reviews



Outline

- Question Answering
- Catalog Quality
- Product Size Recommendations

Outline

- **Question Answering**
- Catalog Quality
- Product Size Recommendations

Amazon Product Pages

- Amazon product pages contain a wealth of information

The image displays four sequential screenshots of an Amazon product page for a GE A730 7MP Digital Camera. The first screenshot shows the product image, a 4.8x zoom, and a 95% rating. The second screenshot shows the 'About this item' section with features like 7MP, 3X optical zoom, and 2.5" LCD. The third screenshot shows the 'Product Information Black' table with dimensions, weight, and manufacturer. The fourth screenshot shows the 'Top reviews' section with three reviews and a 'Customer Questions' section with three questions and answers.

Product Information Black

| | |
|--------------------|------------------------|
| Product Dimensions | 5.6 x 4.1 x 2.6 inches |
| Item Weight | 1.1 pounds |
| Shipping Weight | 1.2 pounds |
| Manufacturer | General Imaging |
| ASIN | B000QGBCVY |

Top reviews

★★★★☆ I bought the GE A730 because it has by far the best feature set of any camera in its class at an unbelievable price. It's pocket-sized, with image stabilization, manual...
By Joe W

★★★★☆ I have had this camera for about 2 weeks and I LOVE IT!! I previously had a Nikon L11 and returned it due to the known issue of it just freezing up when it wants. I purchased...
By DXTAC

★★★★☆ I bought this camera to replace a \$350 first generation SONY digital camera that I've had for over 5 years. This camera works great. Large LCD screen and easy to use...
By Troy Kennedy

Customer Questions

Q: does it record videos well?
A: It does shoot video, kind of well, the lighting is very bad on video I'd say, and for it t...
dianna decoste | 3 years ago

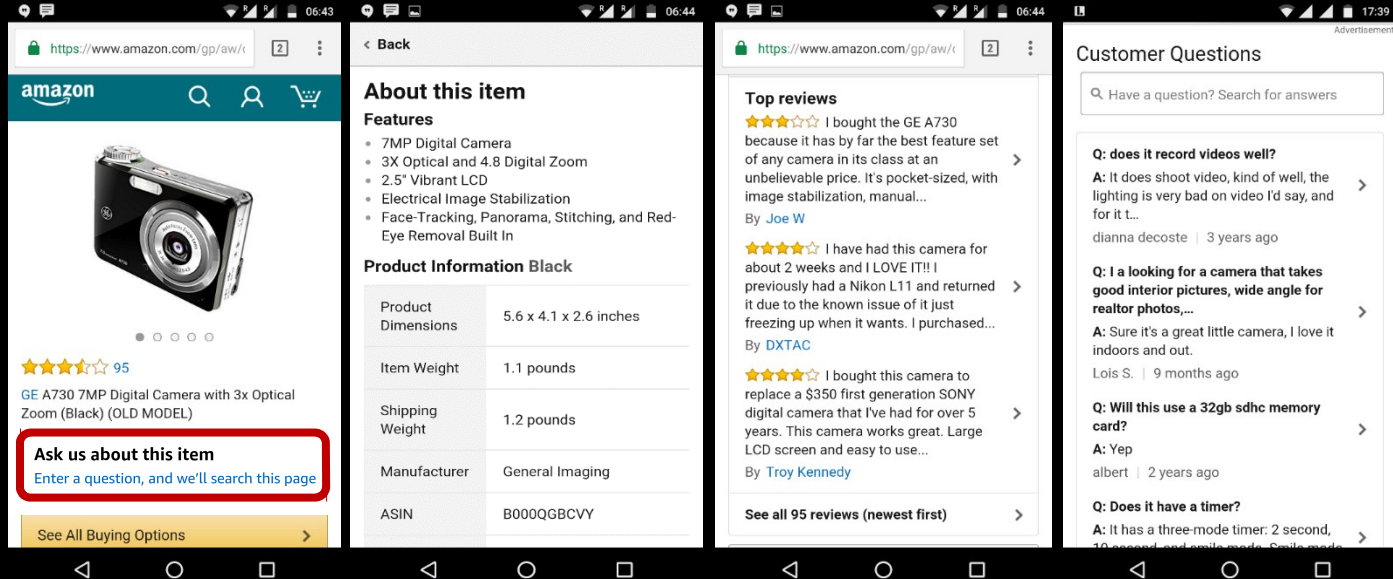
Q: I a looking for a camera that takes good interior pictures, wide angle for realtor photos,...
A: Sure it's a great little camera, I love it indoors and out.
Lois S. | 9 months ago

Q: Will this use a 32gb sdhc memory card?
A: Yep
albert | 2 years ago

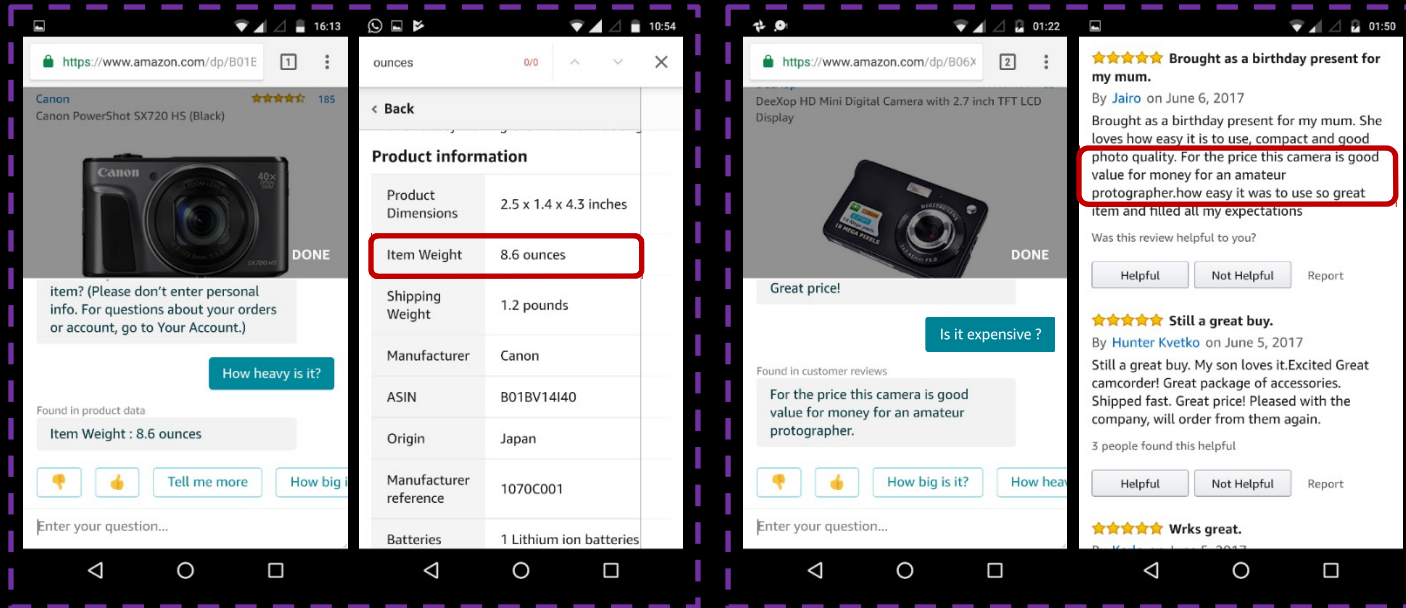
Q: Does it have a timer?
A: It has a three-mode timer: 2 second, 10 second and smile mode. Smile mode...

Question & Answering Bot

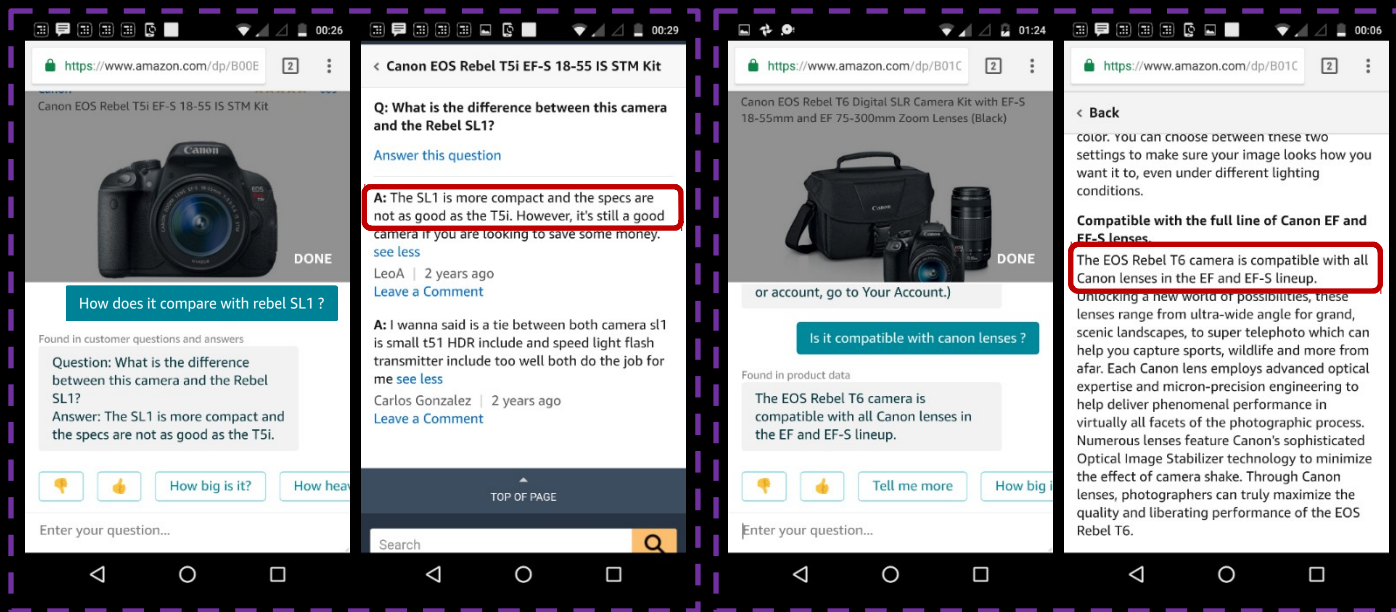
- Question answering interface to make it easy for users to find information on product page



Product Feature Questions



Product Comparison/Compatibility Questions

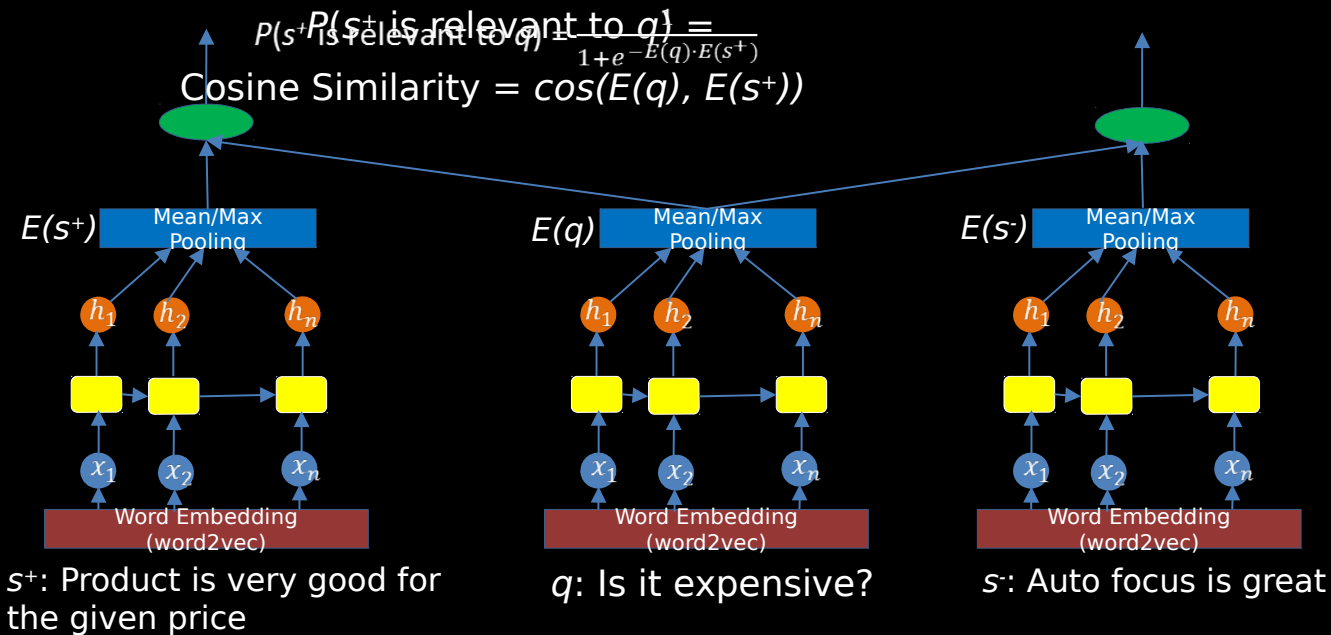


Key Challenges

- Question understanding
 - “What is ISO?” vs “What is the ISO [of this camera]?”
- Semantic matching
 - “cost”, “price”, “bang for buck”, “expensive”, “cheap”
- Natural language answer generation
 - e.g. “This is great value for money” for question “Is this expensive?”
- High precision (>90%) requirement
- Data availability
 - “Will this suitcase fit in the overhead of an airplane?”
- Data quality
 - “Dimensions: 1x1x1 inches” for the question “How big is it” on a “Tripod page”

Learning Semantically Rich Representations

- Training examples: <question (q), relevant snippet (s+), irrelevant snippet (s-)> triples
- **Triplet network**



Results for Different Loss Functions

- Learn question and snippet representations to minimize the following loss functions:
 - Log loss [EMNLP 2015]
 - Siamese loss [Wang et al. 2014]
 $\max\{0, M - (\cos(E(q), E(s^+)) - \cos(E(q), E(s^-)))\}$
 - Twin loss
- Metric: Precision at rank 1 (P@r1)
- Results:

| Loss function | Baseline | Log loss | Siamese loss | Twin loss |
|---|----------|----------|--------------|-----------|
| $-\cos(E(q), E(s^+)) + \max\{0, \cos(E(q), E(s^-))\}$ | | | | |
| P@r1 | 56.8% | 84.6% | 96.1% | 97.04% |

[Wang et al 2014] Learning fine-grained image similarity with deep ranking, CVPR 2014.

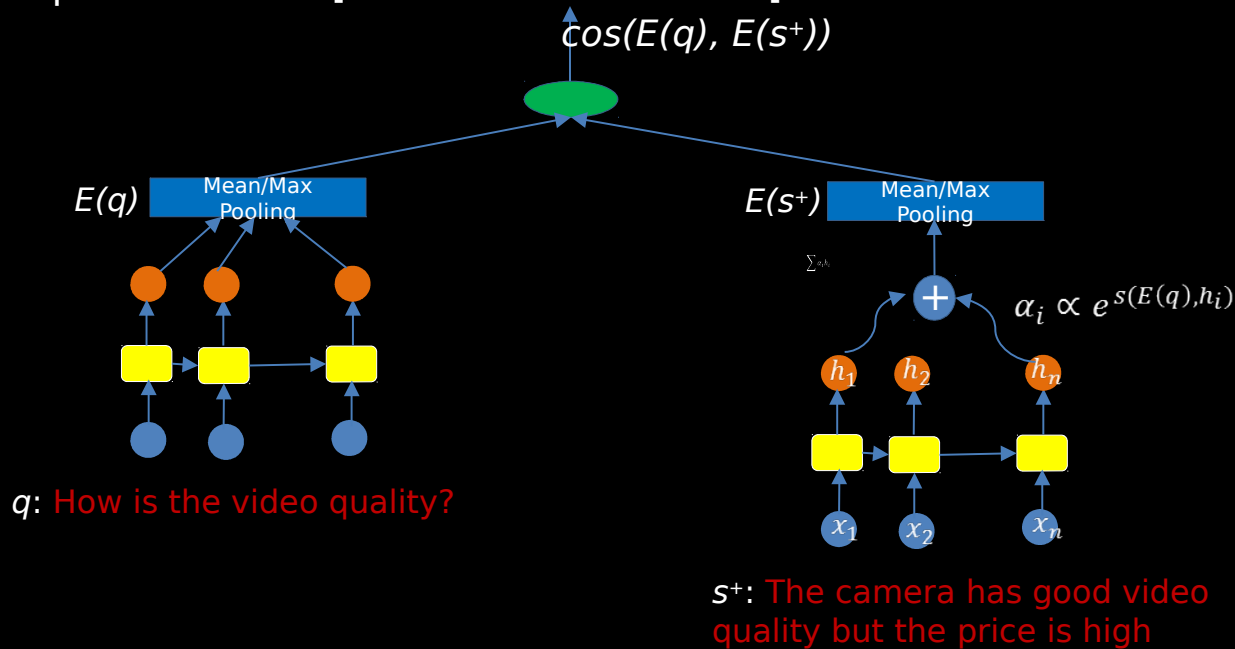


Qualitative Results

| Question | Matching Snippet |
|---|--|
| Is this camera good for pictures at a basketball game? | Works great for sports photography |
| What is the price ? | This item costs \$100.00. To see tax and shipping, add to cart |
| How big is it? | Item dimensions : 3 x 3.28 x 4.37 inches |
| How good is stabilization? | EVERY image came out blurry (and I held the camera still in a well-lit room). |
| Will it fit on Olympus air? | Fits very well the Olympus Air OA-01 |
| How much weight can it hold? | Item weight : 2.2 pounds |
| What is the color of the paper on which the photo is printed? | the color of the camera and the pictures are great. |

Learning Representations with Attention

- Only consider relevant portions of snippets when learning representations [Bahdanau et al. 2015]



Highlighting Words with High Attention Weights

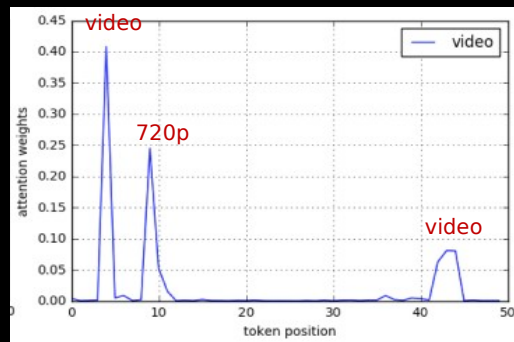
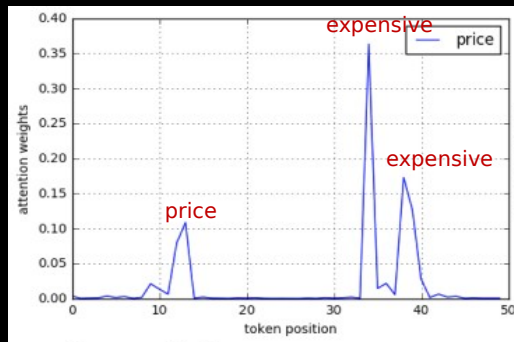
Review Statement:

The package is good video quality is good for 720p but the price is excellent for what you get - especially if you do not want all the whistles and bells of the more expensive gopro 2-4x more expensive and the quality of video is superb and great.

Questions:

Question1: What is the price?

Question2: How good is the video?

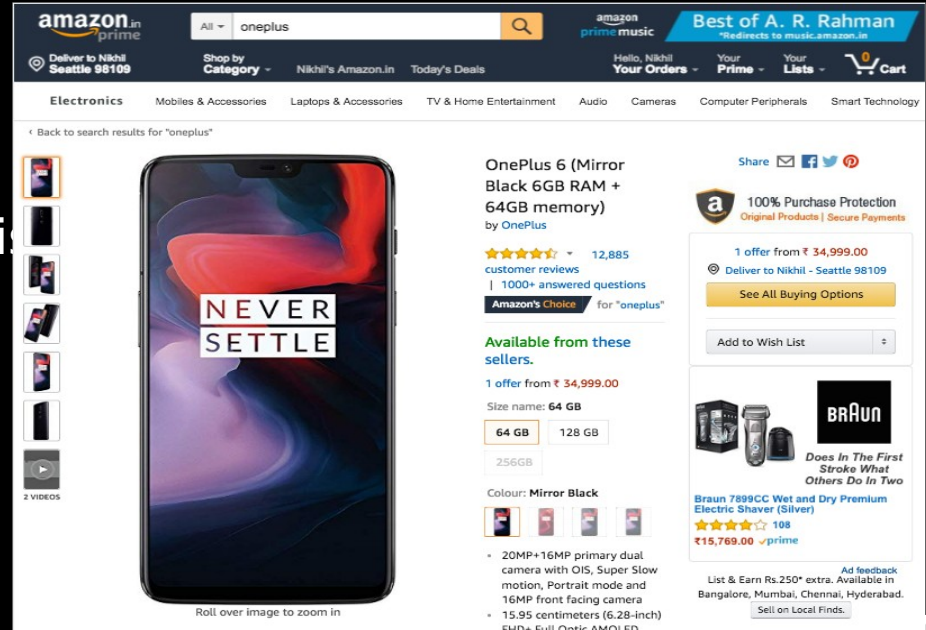


Outline

- Question Answering
- **Catalog Quality**
- Product Size Recommendations

Amazon's Product Catalog

- Product catalog contains rich metadata on Amazon products
 - Title
 - Image
 - Description
- Must ensure that information is
 - Factually correct
 - Consistent
 - Adheres to pre-set guidelines



Title Defects

Title too long, Spam keywords

amazon.in
Clothing & Accessories

Hello, Sign in
Select your address Shop by Category Your Amazon.in Today's Deals Amazon Pay Sell Customer Service Your Orders

Amazon Fashion WOMEN MEN KIDS BAGS & LUGGAGE SPORTSWEAR BRANDS SALE

HDFC BANK 10% Instant Discount* with HDFC Bank Debit/Credit Cards & EMI transactions *T&C apply

Clothing & Accessories > Women > Ethnic Wear > Sarees > Gopinath Feb SAREES FOR WOMEN Latest design for Party Wear Buy in Today Offer in Low

Gopinath Feb
Gopinath Feb SAREES FOR WOMEN Latest design for Party Wear Buy in Today Offer in Low Price Sale, Free Size Ladies Sari, Fancy Material Latest Saree, Designer Beautiful Bollywood Sarees, sarees For Women Party Wear Offer, Designer Sarees, saree With Blouse Piece, New Collection sari, Sarees For Womens, New Party Wear Sarees, Women's Clothing Saree Collection in Multi-Coloured For Women Party Wear, Wedding, Casual sarees Offer Latest Design Wear Sarees With Blouse Piece

Be the first to review this item

Price: ₹4,799.00
Sale: ₹ 499.00 + ₹ 99.00 Delivery charge Details
You Save: ₹ 1,300.00 (72%)
Inclusive of all taxes

Pay on Delivery (Cash/Card) eligible What is this?
In stock.

Delivery to pincode 400001 - Mumbai between Sep 20 - 21. Details
Sold by Gopinath Feb (2.8 out of 5 stars | 8 ratings) and Delivered by Amazon.

Roll over image to zoom in

Email in Title

amazon.in
Clothing & Accessories

Hello, Sign in
Select your address Shop by Category Your Amazon.in Today's Deals Amazon Pay Sell Customer Service Your Orders

Amazon Fashion WOMEN MEN KIDS BAGS & LUGGAGE SPORTSWEAR BRANDS SALE

HDFC BANK 10% Instant Discount* with HDFC Bank Debit/Credit Cards & EMI transactions *T&C apply

GFashion
GFashion family Tshirts-set of 4 -mom dad and kids matching t-shirts (Bista Green, mail the sizes to- sivakm205@gmail.com)

Be the first to review this item

Price: ₹ 1,449.00 + ₹ 50.00 Delivery charge
Inclusive of all taxes

In stock.

Delivery to pincode 400001 - Mumbai within 5 - 9 business days. Details
Sold and fulfilled by gFashion (3.8 out of 5 | 31 ratings).

- Material: Cotton
- Size Chart Mens- S-38,M-40,L-42,XL-44,XXL-46
- Size Chart Womens- S-32,M-34,L-36,XL-8,XXL-40
- Kids - from 6 months to 10yrs available

Amazon Business
Need a GST Invoice on this product? Sign in from/create business account

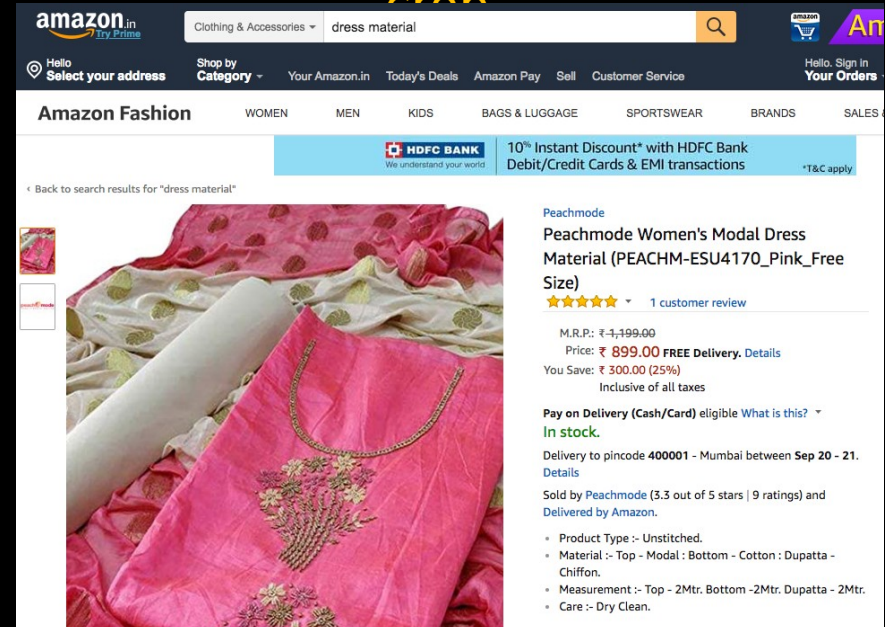
Roll over image to zoom in

Image Defects

Background not white



Incomplete image, Improper



Product Attribute Mismatches


Category Mismatch

amazon.in
Hello, Select your address
Shop by Category
Your Amazon.in
Today's Deals
Amazon Pay
Sell
Hello, Sign in
Your Orders

Amazon Fashion
WOMEN MEN KIDS BAGS & LUGGAGE SPORTSWEAR BRANDS SALE

HDFC BANK
We understand your world
10% Instant Discount* with HDFC Bank Debit/Credit Cards & EMI transactions

Clothing & Accessories › Baby › Soft As A Grape Toddler'S 100% Cotton Short Sleeve T-Shirt (2T, Light Blue)



Soft As A Grape Toddler'S 100% Cotton Short Sleeve T-Shirt (2T, Light Blue)
by BroBag
★★★★★ 18 customer reviews

Available from these sellers.

2 offers from ₹ 6,710.00

- Product Type: Baby Clothing
- Product Brand: Soft As A Grape
- Genuine Imported Products from USA. Prices include Import Custom Duties and Taxes.

LIMITED QUANTITY


Note: The order quantity is limited to 5 units per customer.
Please note that orders which exceed the quantity limit will be auto-canceled. This is applicable across sellers.

Color Mismatch

amazon.in
Hello, Select your address
Shop by Category
Your Amazon.in
Today's Deals
Amazon Pay
Sell
Hello, Sign in
Your Orders

Amazon Fashion
WOMEN MEN KIDS BAGS & LUGGAGE SPORTSWEAR BRANDS

Clothing & Accessories › Women › Ethnic Wear › Dress Material › Anghan Brothers Women's Dress Material(AGB_Dess 4553Q_Blue_Free Size)



Anghan Brothers
Anghan Brothers Women's Dress Material(AGB_Dess 4553Q_Blue_Free Size)
★★★★★ 2 customer reviews

Available from these sellers.

- Length of Top: 2.5 meters (Top Fabric Max Chest Size 43 Inch)
- Size: Free Size | Color: Beige
- Disclaimer :- PRODUCT COLOUR MAY SLIGHTLY VARY DUE TO PHOTOGRAPHIC LIGHTING SOURCES OR YOUR MONITOR SETTINGS.
- Wash Care: Dry Clean Only

Report incorrect product information.

EXPLORE OUR SELECTION OF BEST-SELLING DRESS MATERIALS
50% - 80% OFF
Shop now

Text Attribute Extraction

Objective: Extract attributes like brand, color, size, model from product title and description.

Example:

*Samsung Galaxy A8 Brand New Dustin
hard Silicon Transparent black Color
edge Back CaseCover*

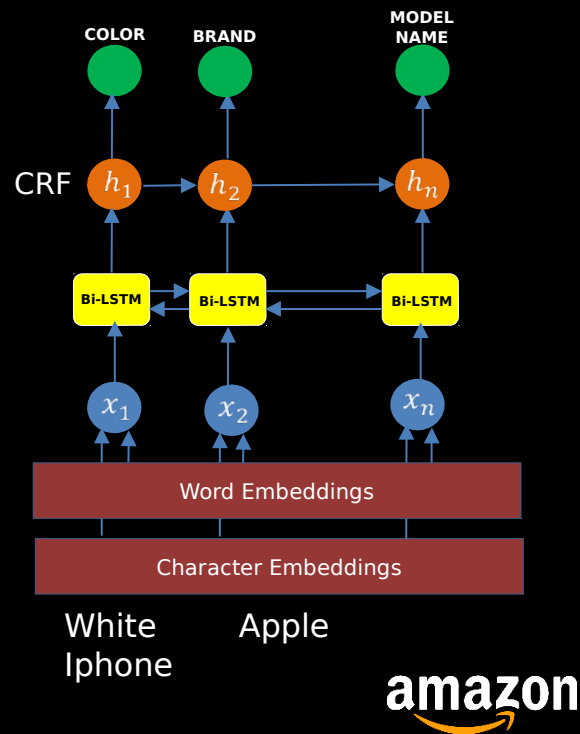
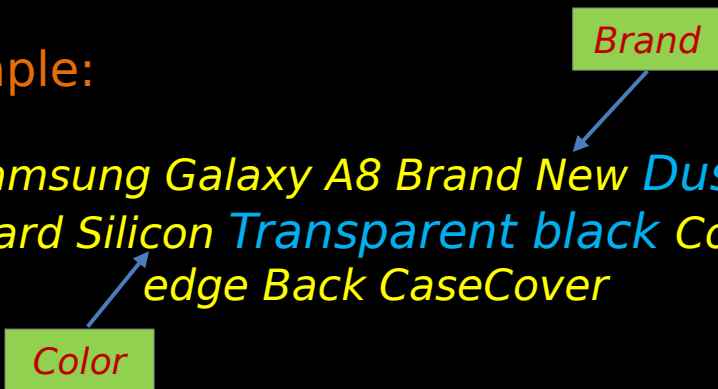
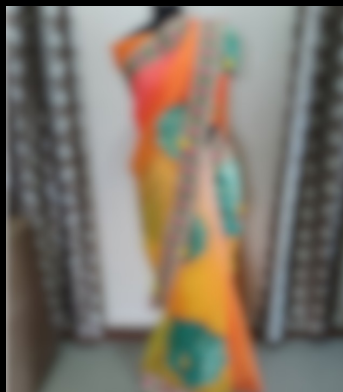


Image Classification/Attribute Extraction

Objective: Detect image defects, extract image attributes such as color, and classify images

Image Defects



Poor
Quality

Attribute
Extraction

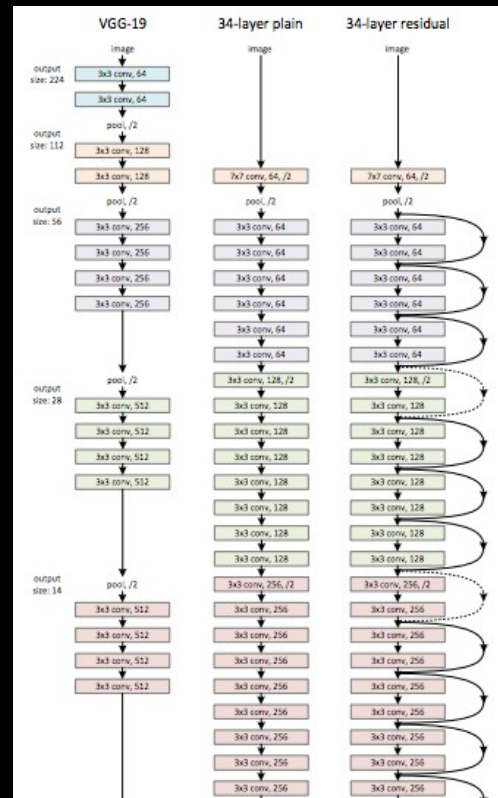


Color:
Red

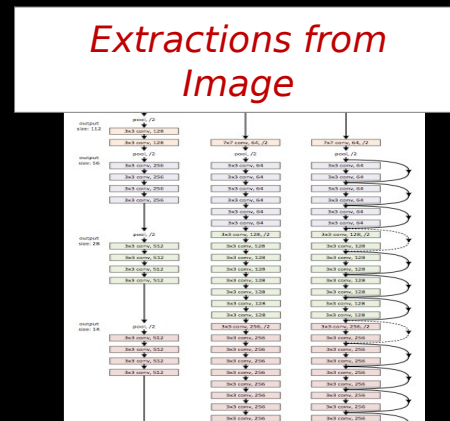
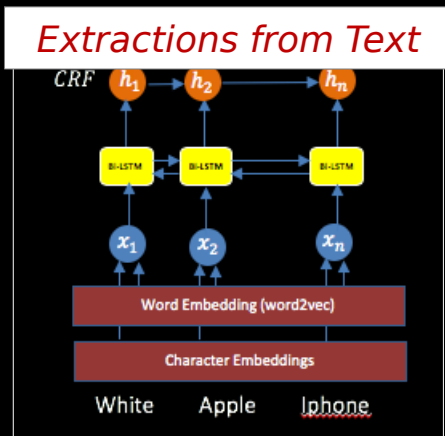
Category
Classification



Category: T-
Shirt



Mismatch Detection



Brand, Color,
Category

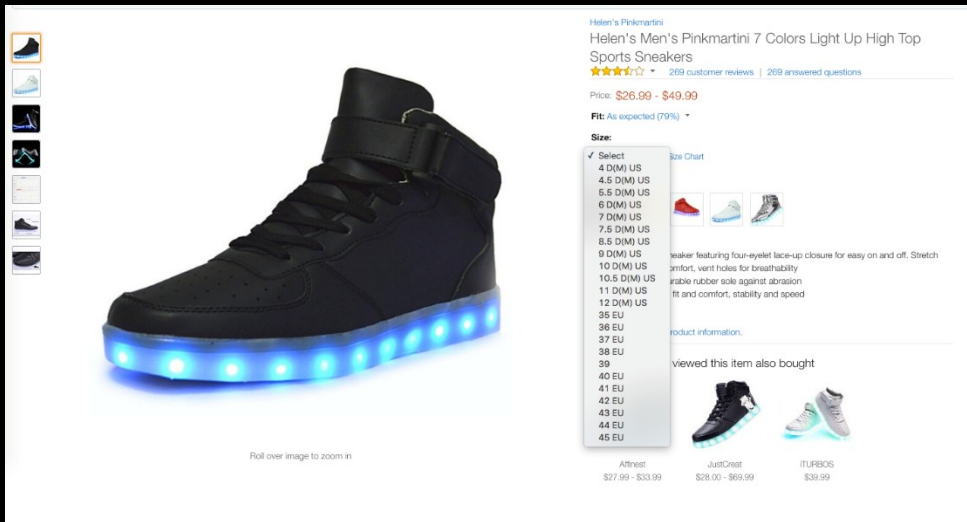
Mismatch Logic

Outline

- Question Answering
- Catalog Quality
- **Product Size
Recommendations**

Size Recommendation Problem

- Given a customer and product, recommend product size that would best fit the customer



Motivation

- No standardization of sizes across brands and locales for product categories such as shoes and apparel
- This leads to users making incorrect purchases, and then returning products
- Products belonging to shoe and apparel categories have high return rates due to fit issues
- Example:
 - Reebok size mapping convention: 6 = 15cm, 7 = 17cm, 8 = 21cm
 - Nike size mapping convention: 6 = 16cm, 7 = 18cm, 8 = 22cm

Key Challenges

- Scale: hundreds of millions of customers and products
- Data sparsity: bulk of users/products have very few purchases
- Cold start scenarios: new customers/products
 - User features: demographics (age, gender), location
 - Product features: catalog size, title, brand, product type
- Multiple personas: each customer i may involve multiple personas
 - E.g., family members sharing an account
 - Personas may have widely varying sizes

Our Approach

- Learn true (latent) size for each customer, product
 - True size for customer corresponds to the physical size of the customer (for shoes, it would be the feet size)
 - True size for product corresponds to it's physical size
- Leverage past customer transactions $T = \{(i, j, y_{ij})\}$
 - y_{ij} takes ordinal values {small, fit, large}

| | Adidas (9) | Nike (8) | Reebok (8) | Nike (9) |
|------------|---------------|-------------|---------------|-------------|
| Customer 1 | large | | fit | ? |
| Customer 2 | | small | | fit |
| Customer 3 | fit | | small | ? |

→ Catalog size

→ Predict fit outcome

Our Approach (Contd)

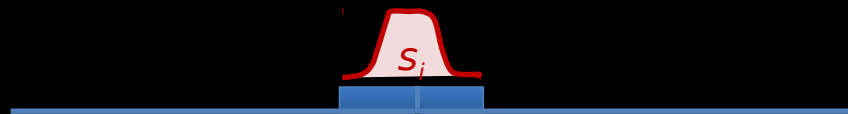
- **Notation**
 - Latent size for customer: s_i
 - Latent size for product: t_j
 - Catalog size for product: c_j
- **Model likelihood** offit as a function of the difference between customer and product latent sizes
$$P(y_{ij} = fit) \propto f(s_i - t_j)$$
- Recommend product j with highest fit likelihood $P(y_{ii} = fit)$ to customer i
- Recommend product j with highest fit likelihood to customer i

Bayesian Modeling Benefits

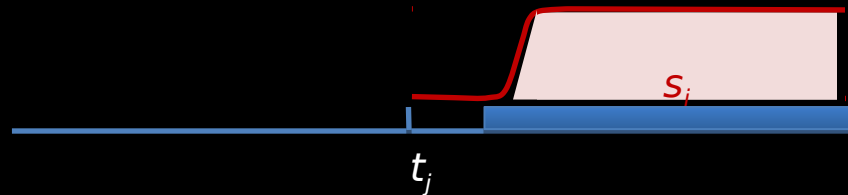
- Handles data sparsity by placing priors on latent size variables
- Models uncertainty in inferred latent sizes
 - Estimates posterior distribution of latent size variables
 - Fit probability is obtained by averaging over posterior size distribution
- Model can capture all the available data
 - Observations: transaction outcomes, customer and product features
 - Hidden variables: latent sizes, customer personas
- Efficient techniques for approximating posterior distributions of latent size variables

Intuition

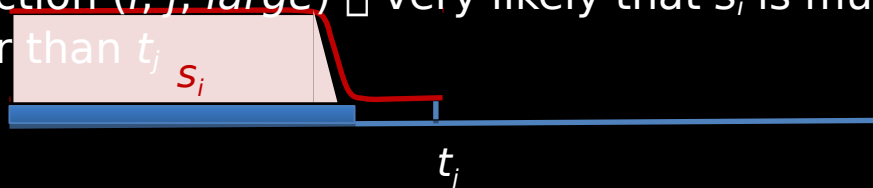
- Transaction (i, j, fit) \Rightarrow very likely that s_i and t_j are close



- Transaction $(i, j, small)$ \Rightarrow very likely that s_i is much larger than t_j



- Transaction $(i, j, large)$ \Rightarrow very likely that s_i is much smaller than t_j



Data Likelihood

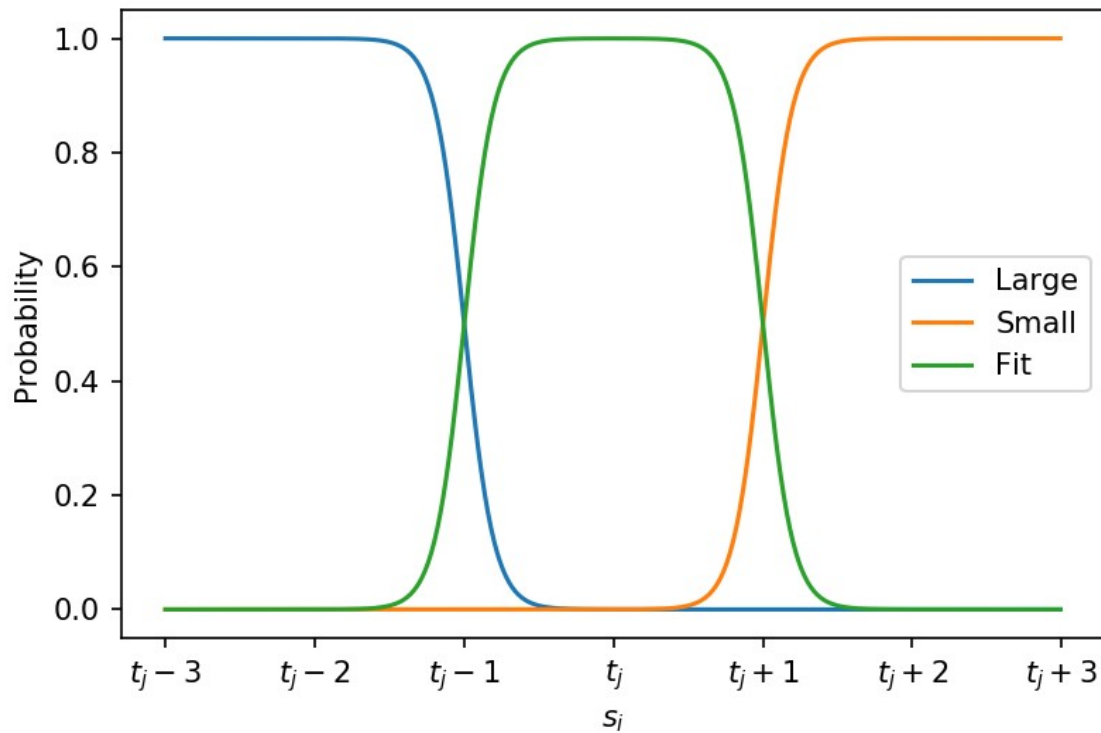
-

$$P(y_{ij} = \textit{small} | s_i, t_j) = \frac{1}{1 + e^{-\alpha(s_i - t_j) + b_1}}$$

$$P(y_{ij} = \textit{fit} | s_i, t_j) = \frac{1}{1 + e^{\alpha(s_i - t_j) - b_1}} \cdot \frac{1}{1 + e^{-\alpha(s_i - t_j) + b_2}}$$

$$P(y_{ij} = \textit{large} | s_i, t_j) = \frac{1}{1 + e^{\alpha(s_i - t_j) - b_1}} \cdot \frac{1}{1 + e^{\alpha(s_i - t_j) - b_2}}$$

Data Likelihood



Generative Model

for each customer i

for each customer i ,

draw latent size $s_i \sim N(\mu_i, \sigma_s^2)$

for each product j ,

draw latent size $t_j \sim N(c_j, \sigma_t^2)$

for each transaction $(i, j, y_{ij}) \in T$,

select $y_{ij} = \text{small}$ with probability $P(y_{ij} = \text{small} | .)$

select $y_{ij} = \text{fit}$ with probability $P(y_{ij} = \text{fit} | .)$

select $y_{ij} = \text{large}$ with probability $P(y_{ij} = \text{large} | .)$

$\text{large} | .)$

Bayesian Inference

- Let β be the vector of latent sizes

$$\beta = (s_1, \dots, s_c, t_1, \dots, t_p, 1)^T$$

- Posterior distribution

$$P(\beta|T) \propto P(T|\beta) \cdot P(\beta)$$

- Not available in closed form due to logistic likelihood terms and Normal priors

0/1

- Not available in closed form due to logistic likelihood terms and Normal priors

Polya-Gamma Augmentation [Polson et al. 2013]

- Introduce Polya-Gamma latent variable w , for every $y \in \mathcal{D}$
- Define the joint likelihood distribution
- Define the joint likelihood distribution
- In [Polson et al. 2013], it is shown that $P(w, y | \beta, x) = \frac{1}{2} e^{-(\beta^T x)} e^{-\frac{(\beta^T x)^2}{2w}} \cdot P(w)$
- In [Polson et al. 2013], it is shown that

$$\int_0^\infty P(w, y | \beta, x) dw = \frac{e^{y\beta^T \cdot x}}{1 + e^{\beta^T \cdot x}}$$

Polya-Gamma Augmentation (Contd)

- Let W be the set of Polya-Gamma variables for D

$$P(\beta|D) \propto \int P(W, D|\beta) \cdot P(\beta) dW$$

- Approximate the augmented joint distribution
Approximate the augmented joint distribution $P(W, D|\beta) \cdot P(\beta)$

$$\prod_{(x,y) \in D} \frac{1}{2} e^{((y-\frac{1}{2}) \cdot (\beta^T \cdot x) - w \cdot \frac{(\beta^T \cdot x)^2}{2})} \cdot P(w) \cdot \prod_i N(s_i | \mu_i, \sigma_s) \cdot \prod_j N(t_j | c_j, \sigma_t)$$

Gibbs Sampling Algorithm

- Conditional distribution of
- Conditional distribution of w_i
 - Conditional distribution for $\max\{0, M - (\cos(E(q), E(s^+)) - \cos(E(q), E(s^-)))\}$
- Conditional distribution for β_j

))

Predictive Distribution

- Let r samples be drawn from the posterior of β
- Let the l^{th} sample be $\beta^l = (s_1^l, \dots, s_c^l, t_1^l, \dots, t_p^l)$

$$P(y_{ij} = fit | \mathbf{D}) = \int P(y_{ij} = fit | \beta) \cdot P(\beta | \mathbf{D}) d\beta$$

$$\approx \frac{1}{r} \sum_{l=1}^r \frac{1}{1 + e^{\alpha(s_i^l - t_j^l) - b_1}} \cdot \frac{1}{1 + e^{-\alpha(s_i^l - t_j^l) + b_2}}$$

Experimental Results

- Consider 6 real-life shoes datasets with between 10M and 33M transactions
- Baseline model
 - Product size $t_j = c_j$
 - Customer size s_i = Average size of products purchased by customer
 - Logistic regression model with feature (s_i, t_j) to predict outcome
- Bayesian Logit model
 - Predict outcome with highest probability
- Performance metric: weighted AUC
- Results (% improvements over baseline)

| Datas et | A | B | C | D | E | F |
|--------------|-------|-------|------|-------|-------|-------|
| Bayesi an | 17.71 | 18.28 | 19.7 | 25.78 | 20.22 | 19.42 |

Leveraging Customer and Product Features

- Means of latent size priors are obtained by performing regression over customer (f_i) and product (g_j) features [AC 2009]

- Generalization:
for each customer i ,

for each customer i ,

draw latent size $s_i \sim N(w_f \cdot f_i, \sigma_s^2)$

for each product j ,

draw latent size $t_j \sim N(w_g \cdot g_j, \sigma_t^2)$

for each transaction $(i, j, y_{ij}) \in T$,

select $y_{ij} = \text{small}$ with probability $P(y_{ij} = \text{small} | .)$

select $y_{ij} = \text{fit}$ with probability $P(y_{ij} = \text{fit} | .)$

select $y_{ij} = \text{large}$ with probability $P(y_{ij} = \text{large} | .)$

$\text{large} | .)$

- Perform least squares regression to learn parameters w_f and w_g from customer and product size samples

Incorporating Customer Personas

- Latent size for persona k of customer i : s_{ik}
- Latent variable containing persona involved in transaction (i, j, y_{ij}) : z_{ij}

- **Generative model:**

for each customer i

for each customer i ,

draw persona distribution $\theta_i \sim \text{Dir}(\alpha)$

for each persona k draw latent size $s_{ik} \sim N(w_f \cdot f_k, \sigma_s^2)$

for each product j ,

draw latent size $t_j \sim N(w_g \cdot g_j, \sigma_t^2)$

for each transaction $(i, j, y_{ij}) \in T$,

draw persona $z_{ij} \sim \text{Mult}(\theta_i)$

select $y_{ij} = \text{small}, \dots$ with probability $P(y_{ij} = \text{small} | z_{ij}), \dots$

$\text{small} | z_{ij}, \dots$

- Gibbs Sampling algorithm can be extended to draw z_{ij} samples

Summary

- Learning semantically rich data representations improves predictive accuracy of models
 - Question answering, catalog quality, product search, product recommendations, ...
- **Deep Learning** to learn embeddings
 - Allows semantic matching between questions and snippets
 - Loss functions like Siamese loss that aim to maximize difference in class scores perform better
- **Probabilistic Graphical Models** to learn latent sizes
 - Priors handle data sparsity, prevent overfitting
 - Posteriors model uncertainty in data
 - Leverage all the available signals

Summary

- AI, ML at the core of Amazon's business
 - Address quality, recommendations, product search, catalog quality, question answering, demand forecasting, ...
- Learning semantically rich data representations improves predictive accuracy of models
 - Question answering, catalog quality, product search, ...
- Deep learning is an effective tool to learn semantically rich data representations