

October 11th, 2017

Deep Learning @WalmartLabs

Alessandro Magnani

@WalmartLabs, Walmart Global eCommerce
California, USA

Save
money.
Live
better.

Outline

- @WalmartLabs overview
- Deep Learning @WalmartLabs
 - Product visual search
 - Product recommendation
 - Spell correction and query interpretation
- Deeper technical dive:
 - Product Matching
 - Challenges and insight
 - Product Classification
 - Challenges and insight
- Questions?

Walmart

11.5k

Stores in 28
Countries

12

eCommerce
Websites

Scale to build technology. Scale to consume it. FAST.

12 websites across the globe ~3k Technologists

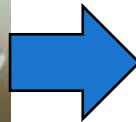
Silicon Valley . Portland . Carlsbad . Bangalore .
Leeds . Ontario . Sao Paulo

Data @WalmartLabs

- User interaction with website and apps
 - Search queries
 - Product views, clicks, purchases, ...
- Product catalog (millions of products)
 - Title, descriptions
 - Product attributes
 - Images, videos
 - User Reviews
 - Prices
 - Curated product tags

- ML and deep learning used extensively both in stores and online
- Many teams from product ingestion to last mile delivery use ML
- Multiple teams have embedded data scientists
- This presentation doesn't cover everything ML

Visual Search @WalmartLabs



Walmart

All Departments

Halloween My local store Pick it up TODAY Tips & Ideas

L'ORÉAL PARIS PURE-CLAY MASKS INSTA-DETOX FOR PURE, BRIGHT, REFINED & COMFORTED SKIN

Showing 1-40 of 74,973 results

Refine Price Store availability

Sort Best match

Departments

Clothing

Bags & Accessories

Handbags

Wallets

Tote Bags

Hobo Bags

Women

Women's Shirts & Blouses

See All Women's Tops & T...

Women's Dresses

Clearance

\$11.50

List price \$22.99

Save \$11.49

Drape Pattern Stars Detail Red Wallet Pouch Purse Bag for Lady

Sold & Shipped by Tasharina Corp

Free shipping

\$19.99

Monte Vista Concealed Carry Handbag Satchel Purse Red

★★★★★ (1)

Sold & Shipped by The Uber Shop

Free shipping

Clearance

\$17.43

List price \$34.99

Save \$17.56

Unique Bargains Women's Zip Closed Quilted Top-Handle PU Crossbody E...

Sold & Shipped by Tasharina Corp

Free shipping

\$29.49

3pc Red Purse Set - LUPU700RD

Sold & Shipped by BigFrys Discount

Free shipping

Clearance

\$21.04

List price \$42.99

Save \$21.95

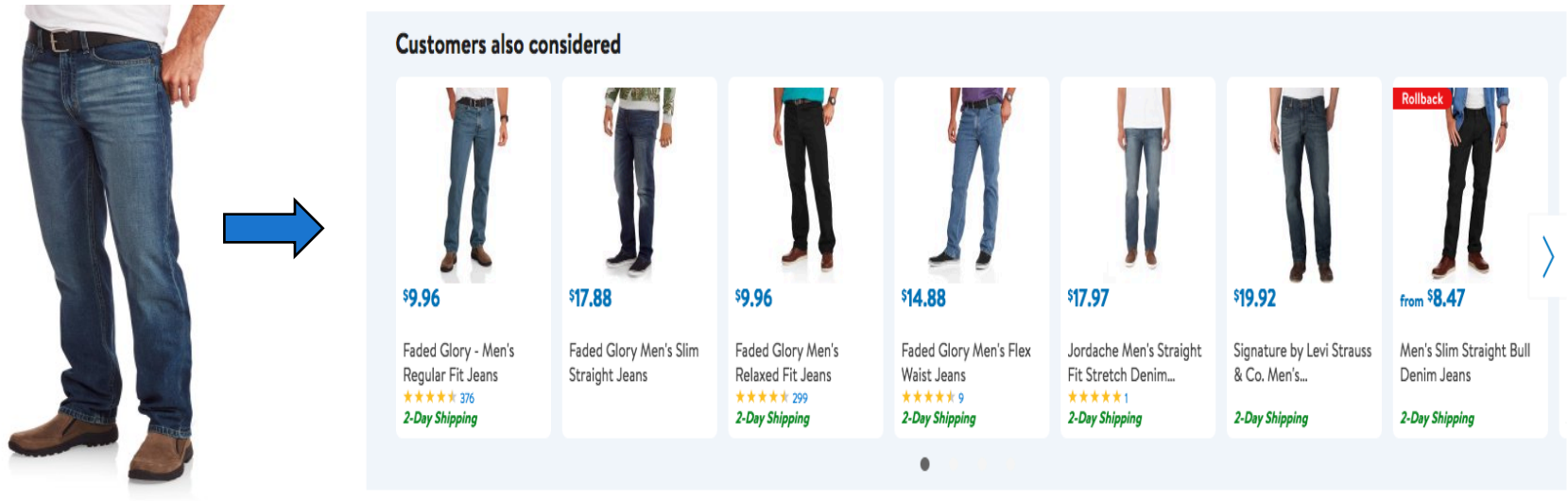
Unique Bargains Women's Imitation Leather Detachable Strap ...

Sold & Shipped by Tasharina Corp

Free shipping

- Working on a **new user experience** for shoppers
- Extract attributes style, brand, material, from user taken images
- Retrieve **visually “similar”** results
- Deep learning at the core of this new experience
- Using: CNN, GAN networks

Product recommendation @WalmartLabs



- Products recommended based on user history
- Each product can be abstracted using an embedding (Word2Vec)
- Can suggest products that don't have a “big” history
- Can use image similarities (CNN embedding)

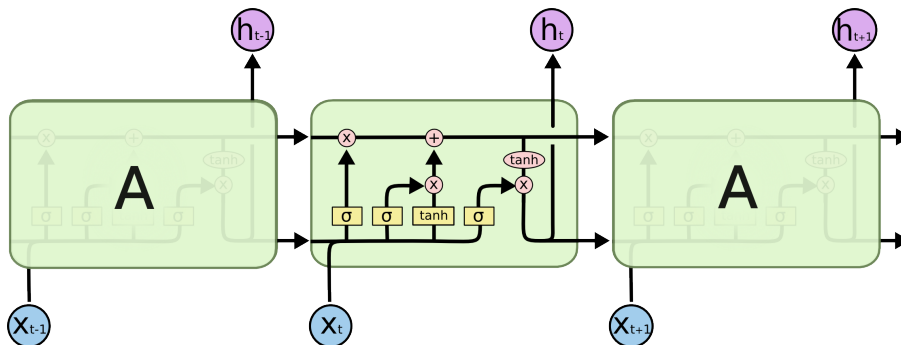
Black and decker coffee maker under 50\$

Brand

Item


Price

- Understand the user intent
- Suggest query refinement
- Corrects user misspelling
- Using: Language model for user queries (RNN models)




Product Matching @WalmartLabs





- Walmart sells same product from multiple vendors
- Products have to be matched
- Errors can be costly and have big impact on user satisfaction



Hippo Head Adult T-Shirt by The Mountain - 10-3384
ADULT M



Researched. Trusted.
Choose from more Every Day Low Prices and get the same great customer care. Shop one of our [trusted sellers](#).

Price + Shipping \$18.7 to \$21.18	Delivery Date Jul 14 to 28	5 Sellers
\$18⁷⁰ + Free shipping	Jul 14-20	Sold & Shipped by Best Service Stores  Return policy Qty: 1 <input type="button" value="Add to Cart"/>
\$18⁷⁰ + Free shipping	Jul 21-27	Sold & Shipped by AllPosters.com AllPosters™ Return policy Qty: 1 <input type="button" value="Add to Cart"/>
\$18⁸⁸ + Free shipping	Jul 17-21	Sold & Shipped by Tuff Supplies Corp  Return policy Qty: 1 <input type="button" value="Add to Cart"/>
\$18⁹⁵ + Free shipping	Jul 24-28	Sold & Shipped by 2Bhip  Return policy Qty: 1 <input type="button" value="Add to Cart"/>
\$16¹⁸ + \$5.00 shipping	Jul 24-27 <small>Want it sooner? Choose an earlier date at checkout.</small>	Sold & Shipped by Rockabilia  Return policy Qty: 1 <input type="button" value="Add to Cart"/>

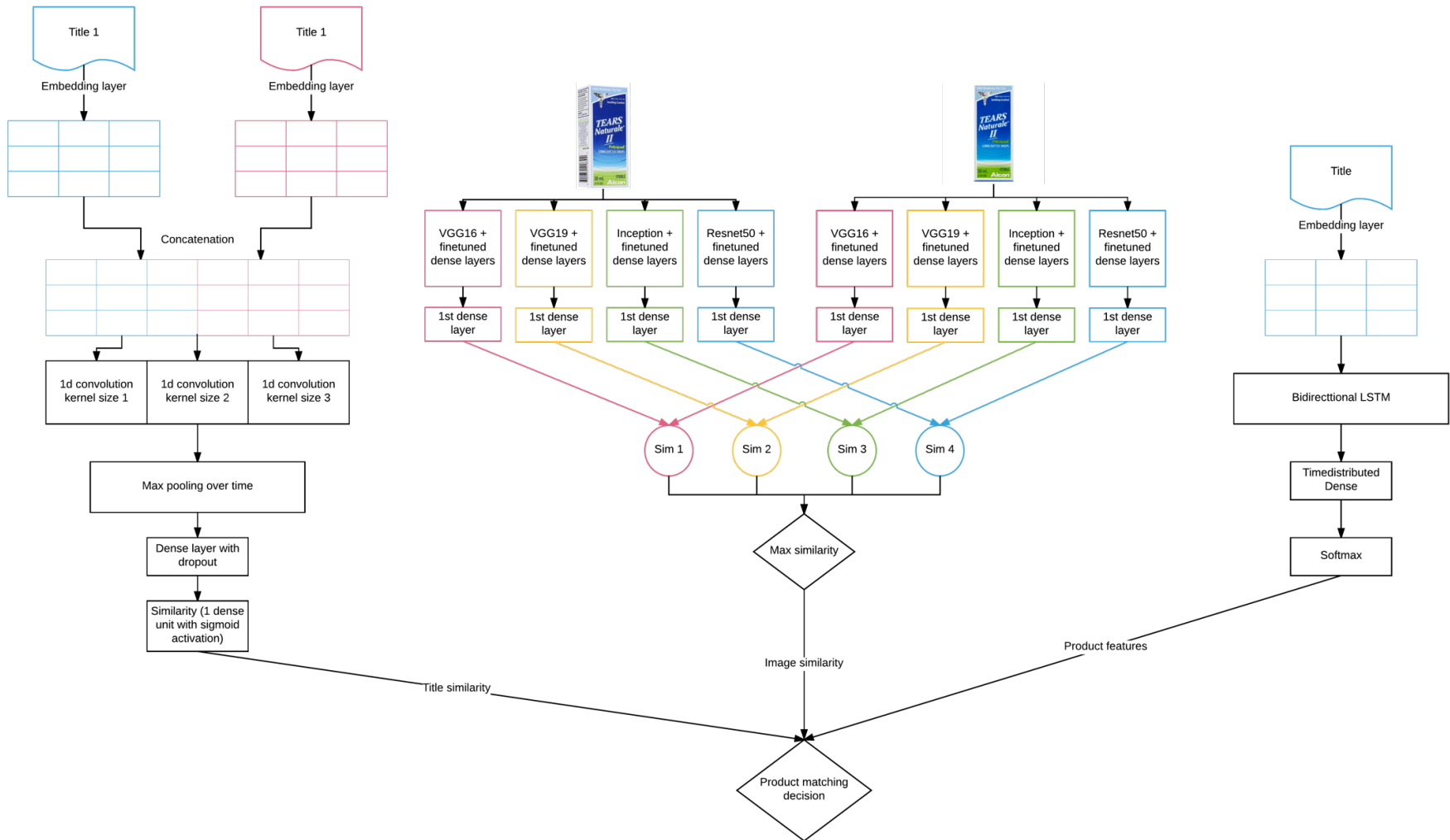
Challenges

- Unique identifiers (UPC, gtin, ...) **don't always work**
- Titles/images can be different
- Price can have large differences for matching products
- Product attribute are missing/wrong

Approach

- Use all available signals (Title, Image, price, ...)
- Extract attributes if not available to compare products
- CNN embedding used for image comparison
- Word embedding + CNN for title comparison
- CNN for attribute extraction

Product Matching @WalmartLabs



Insights

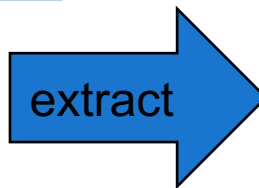
- Matching is a hard problem
 - Unique Ids don't completely solve the problem
- An error can be very costly
- A solution requires careful evaluation of all the data sources

Product classification @WalmartLabs

Goal: given product (title, image) find product category

- Walmart has multiple taxonomies
 - For product placement on website
 - For internal reporting
- Each taxonomy has **thousands** of categories

Title: Top Headwear
Safari Explorer Bucket Hat



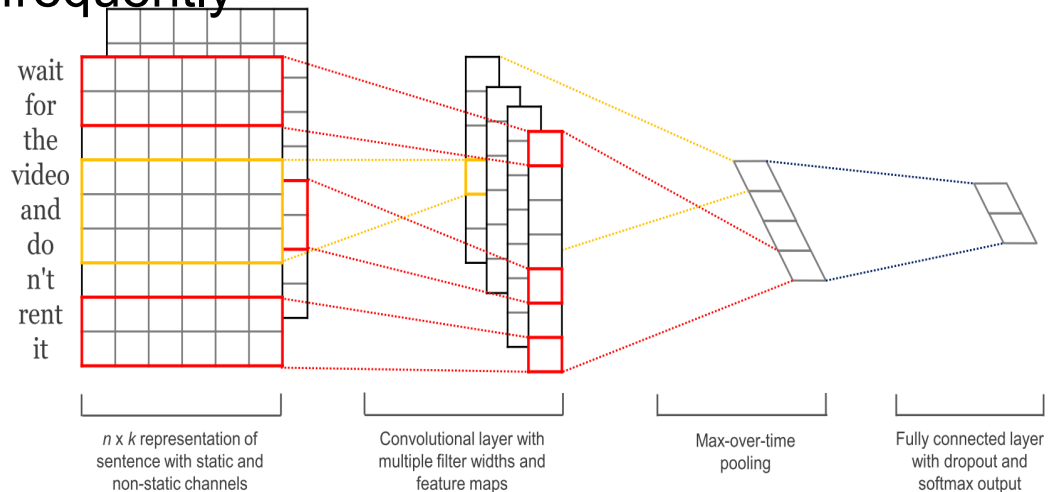
- “Hat” in Taxonomy_1
- “Explorer Hat”: Taxonomy_2
- ...

Challenges:

- Manual classification is too expensive
- Data can be noisy
- Large number of categories
- Manual classification requires domain knowledge
 - Hard to know the entire taxonomy
 - Crowd can produce noisy data
 - Categories created internally can also be noisy
- Taxonomies can change frequently

Approach for text:

- Word embedding
- CNN (shallow)



Product Classification @WalmartLabs

Final prediction



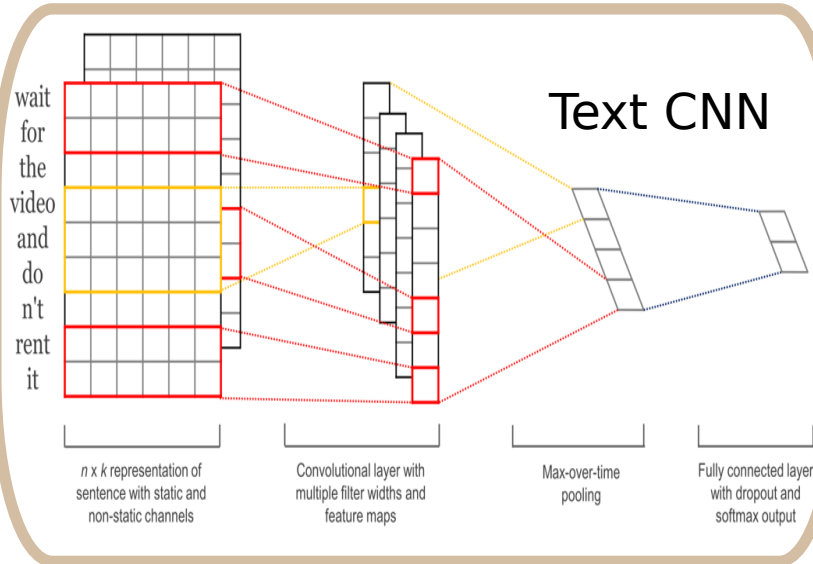
Policy

Prediction ↑

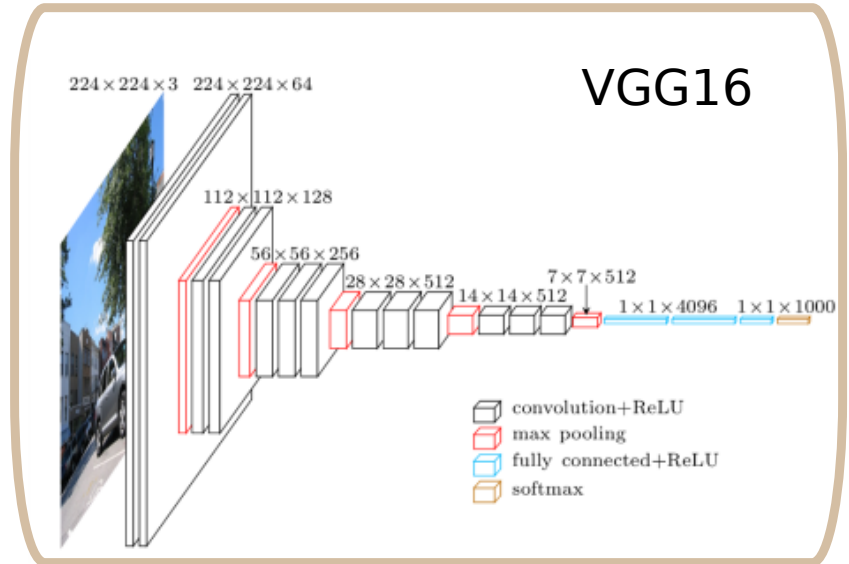
↑ Class probabilities

Class probabilities ↑

Prediction ↑



↑
Text



↑
Image

- Leverage all data from products (image, text, ...)
- Research paper currently under review: "Is a picture worth a thousand words?"

Insights:

- Accuracy is not always the “best” metric
 - Not all errors are the same
 - Classes are not equally important
- Title is the most important signal
- Improving classification accuracy with image is hard
- Data quality is a challenge
 - Change in taxonomy makes this harder
- Leveraging crowd requires careful design of tasks
 - Vetting of people knowledge of space

Much More Deep Learning @WalmartLabs

- Attribute extraction from product images
- Product image recognition scan-less checkout
- Detecting “offending” images
- Detecting other image problems
- Automate recognition of inbound cases in distribution centers

We are hiring!

Any question?