Deep Learning for Semantic Search in E-commerce

Somnath Banerjee

Head of Search Algorithms at Walmart Labs

https://www.linkedin.com/in/somnath-banerjee/

March 19, 2019

Walmart E-commerce search problem

Store Associate

E-commerce Search

provides the functionality of a human but at scale

100M+ Customers × 100M+ Queries × 100M+ Items

Flash Drive
USB Drive
Thumb Drive
Jump Drive
Pen Drive
Zip Drive
Memory Stick
USB Stick
USB Flash Drive
USB Memory
USB Storage Device

Misspelled Queries

Flush Drive
USC Drive
Thamb Drive
Jmp Drive
Pin Drive
Zap Drive
Memory *Steak*
USB *Stock*
USB Flash *Drve*

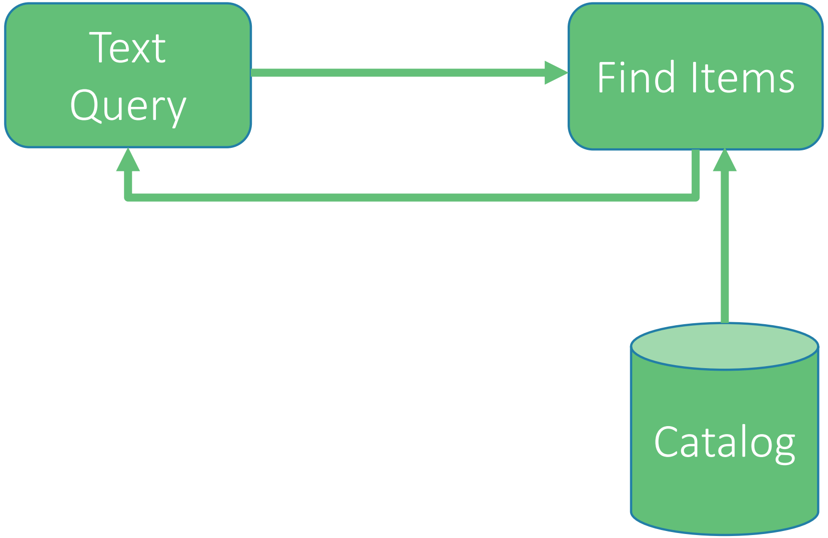
Horse Saddle

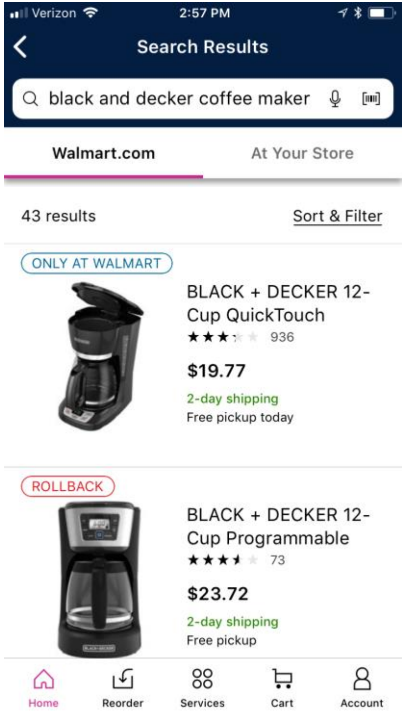
Bike Saddle

Outline

- Core problems of e-commerce search
- Semantic search in e-commerce
- Deep Learning for semantic search
  - Query classification
  - Query token tagging
  - Neural IR
  - Image understanding (sneak peek)

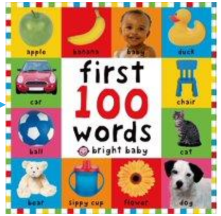
Core of E-commerce Search





Core problems of E-commerce search

Learning book



Tide 100 oz
Tide 100 fl oz
Tide 100 ounce

Levi's
Levi Strauss
Signature by Levi Strauss and Co.

Open vocabulary in query and catalog

Ziploc

product

brand

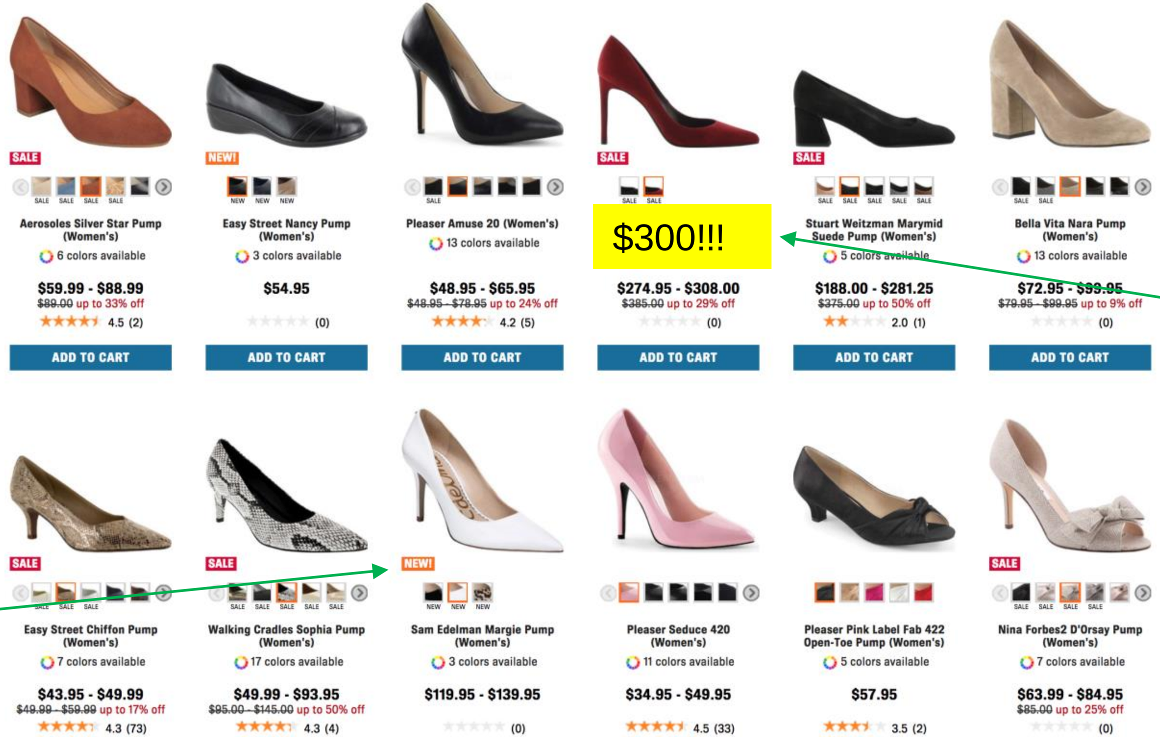
Ambiguity

Neck style?
Fabric?
No. of pockets?

Missing catalog values

Buying decision is influenced by item attractiveness

Image quality



$300!!!

Presence of expensive items

Tags

Walmart Labs

Core technical problems of e-commerce search

Matching query to items

Ranking items

Pump shoes

Position 1

Position 2

Text matching is not enough

Lemon

Lemon Fruit

Lemon Balm

Nivea 16oz

Nivea 15.5oz

Tire sealant 16oz

Sematic Search

Query understanding
- Attribute understanding

Matching query and item
- Text matching
- Attribute matching

Ranking Items

Deep learning for semantic search

Deep Learning for Query understanding

Matching query and item

- Text matching
- Attribute matching

Ranking Items

Deep learning for semantic search

Deep Learning for Query understanding

Matching query and item

- Text matching
- Attribute matching

Ranking Items

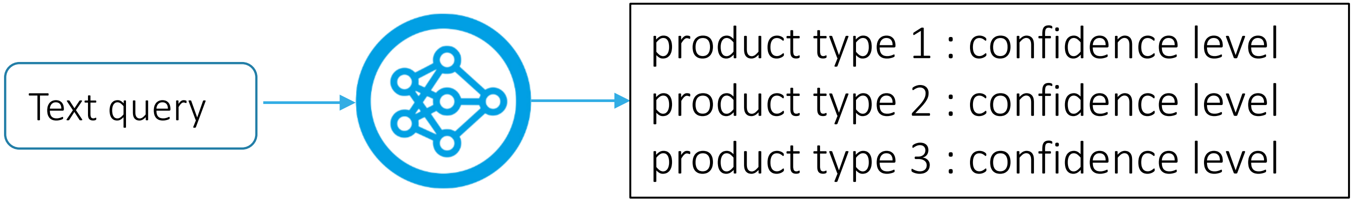
Neural IR
End-to-end matching and ranking

Image understanding
Not just text search

Outline

- Core problems of e-commerce search
- Semantic search in e-commerce
- Deep Learning for semantic search
  - Query classification
  - Query token tagging
  - Neural IR
  - Image understanding (sneak peek)

Query Classification



Product Type

- A predefined list
- Indicates a specific product in the catalog
- Every item in the catalog is tagged with a product type

Query classification examples

ziploc bags → Food Storage Bags: 1.0

nvidia gpu → Computer Video Cards : 0.85
Laptop Computers: 0.08
Desktop Computers: 0.06

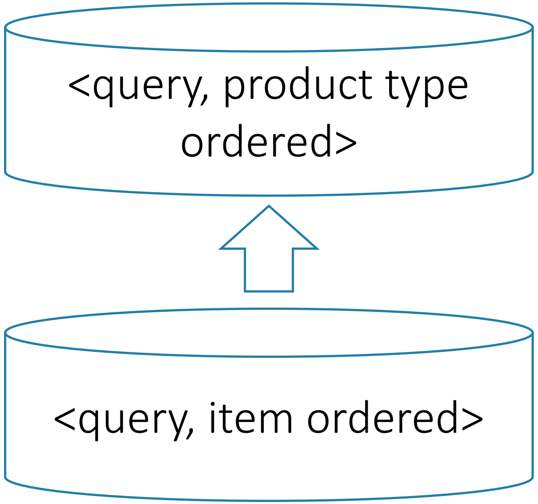
bedroom furniture → Large number of product types

Hard to balance precision vs recall

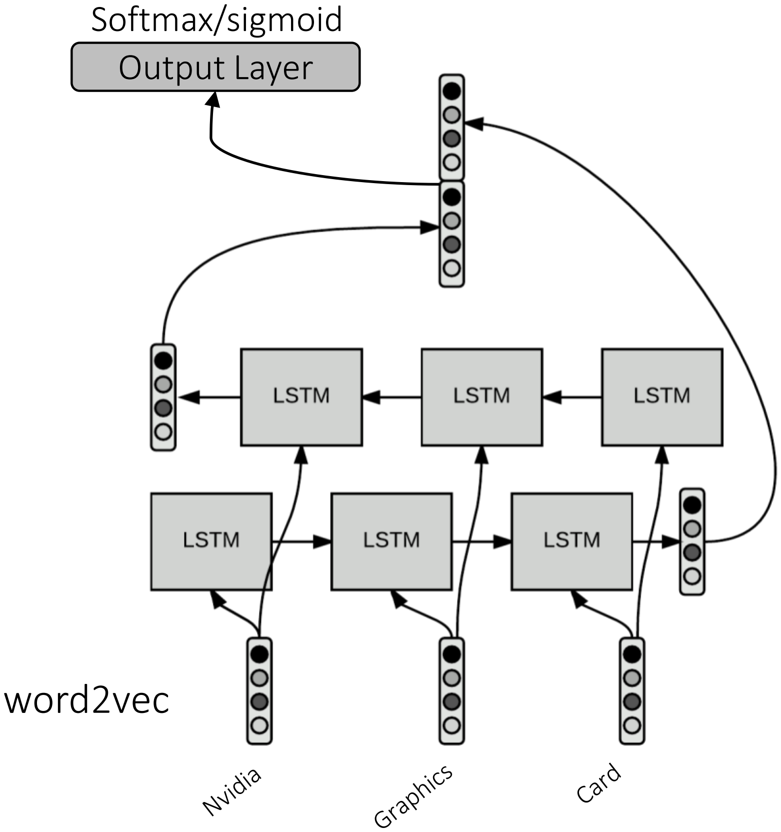
Query classification challenges

Short text	Large scale classification	Multi-class, multi-label problem	Needs to respond in few milliseconds	Unbalanced class distribution
• Queries are of 2-3 tokens	• Thousands of product types (classes)	• Same query can have multiple product types	• Classifies queries at runtime	• Some product types are much more popular

Data and Model



Historical Search Log

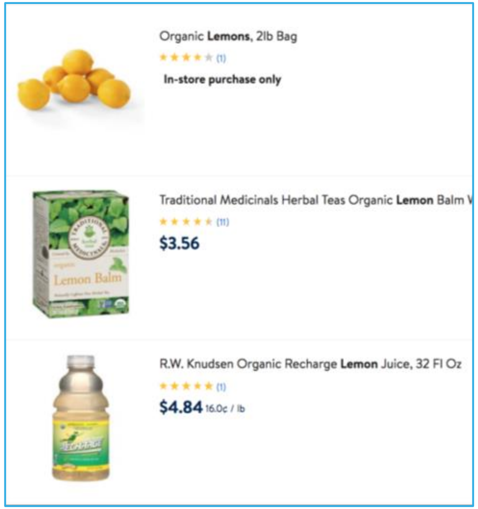


BiLSTM

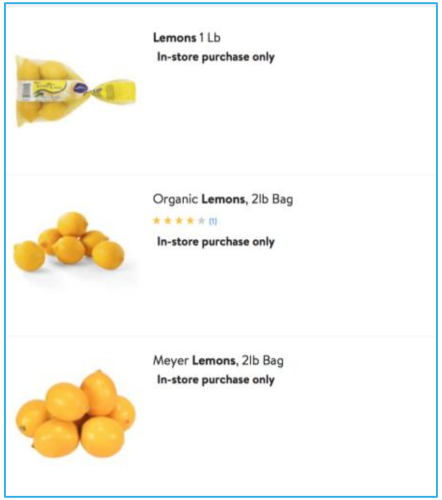
https://guillaumegenthial.github.io/sequence-tagging-with-tensorflow.html

Usage of query classification

lemon



Without Query Classification

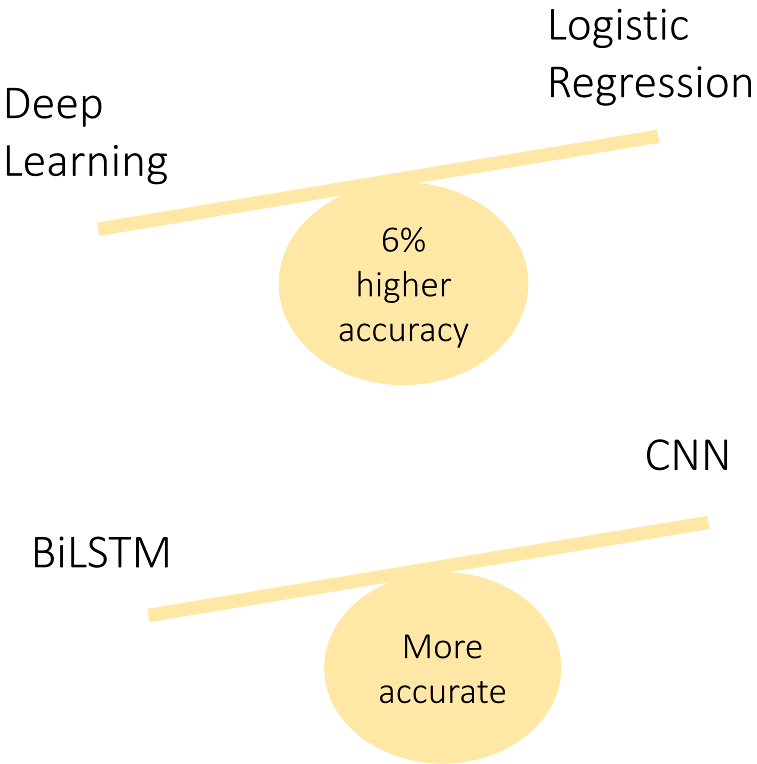


After we understand the query “lemon” as a fruit

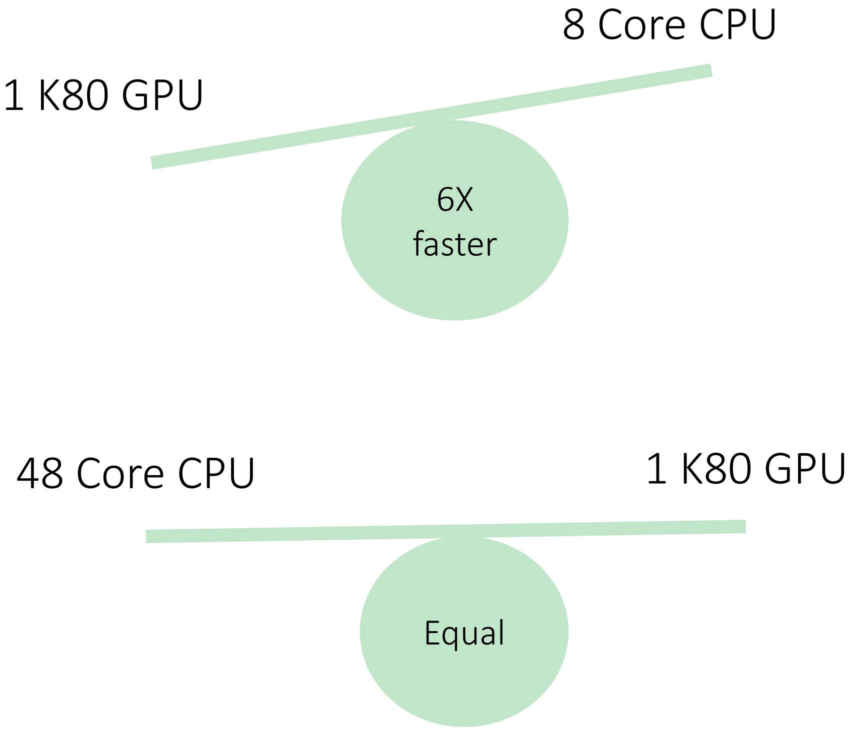
20% reduction of irrelevant items in certain query segments

Key Learnings

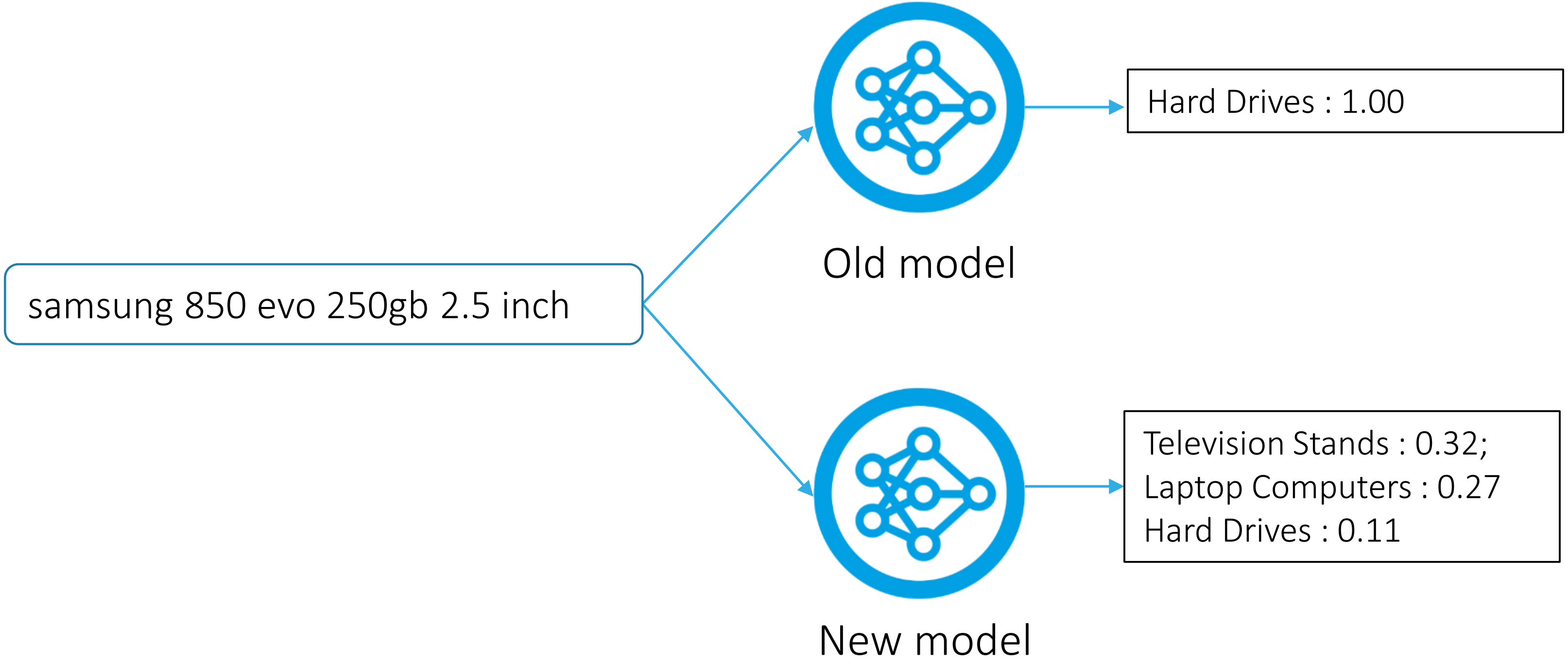
Accuracy



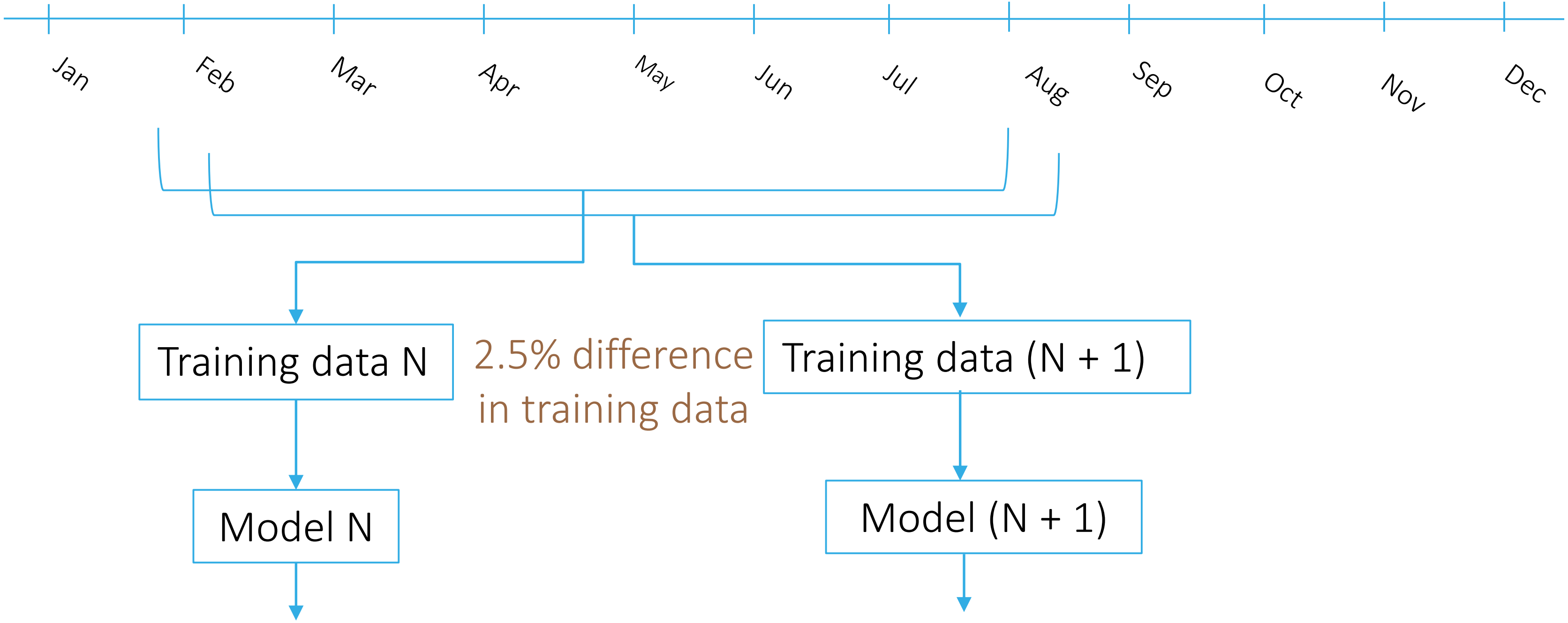
Training Time



Key Learnings - instability in Prediction

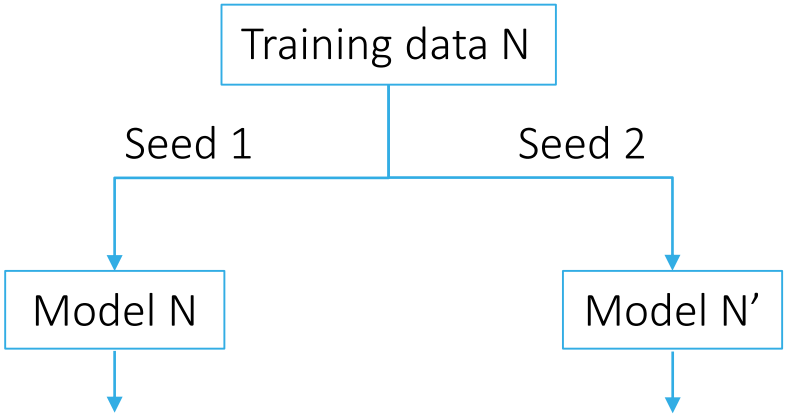


Instability in Prediction



Top predicted class is different for 10% of the test set

Instability in prediction – different seeds



Top predicted class is different for 7% of the test set

Different tensorflow and numpy seeds

Walmart Labs

Sources of Instability

Overfitting

- Deep Learning model has high variance, particularly on the low traffic queries
- Simpler models could be more stable but less accurate

Sigmoid (1-vs-all) classifier is more unstable

- Softmax scores are interdependent across classes and less stable

Noisy training data

- Item order data is less noisy than click

Rounding errors in the arithmetic operations

- CPU is more stable than GPU

Reduction of Instability

Softmax
Clicks
CNN

Sigmoid
Orders
BiLSTM

40% reduction
of instability

Attributes to match

- Product Type
- Brand
- Color
- Gender
- Age Group
- Size (value & unit)
  - Pack Size
  - Screen Size
  - Shoe Size
  - …
- Character
- Style
- Material
- …

Not Feasible – Separate classifier for each attribute

- Too many classes (e.g. 100K+ brand values)
- Sparse attributes; most attribute prediction should be NA
- Creating training data of <query, attribute> is more noisy and inaccurate

Query token tagging

Query → → Query tokens tagged with Attribute Names

faded glory long sleeve shirts for women → →

Faded Glory	Brand
Long sleeve	Sleeve length
shirts	Product type
for	NULL
women	Gender

Training data

blue	women	levis	jeans
Color	Gender	Brand	Product Type

toys	for	girls	3 – 6	years
Product Type		Gender	Age Value	Age Unit

Human curated data

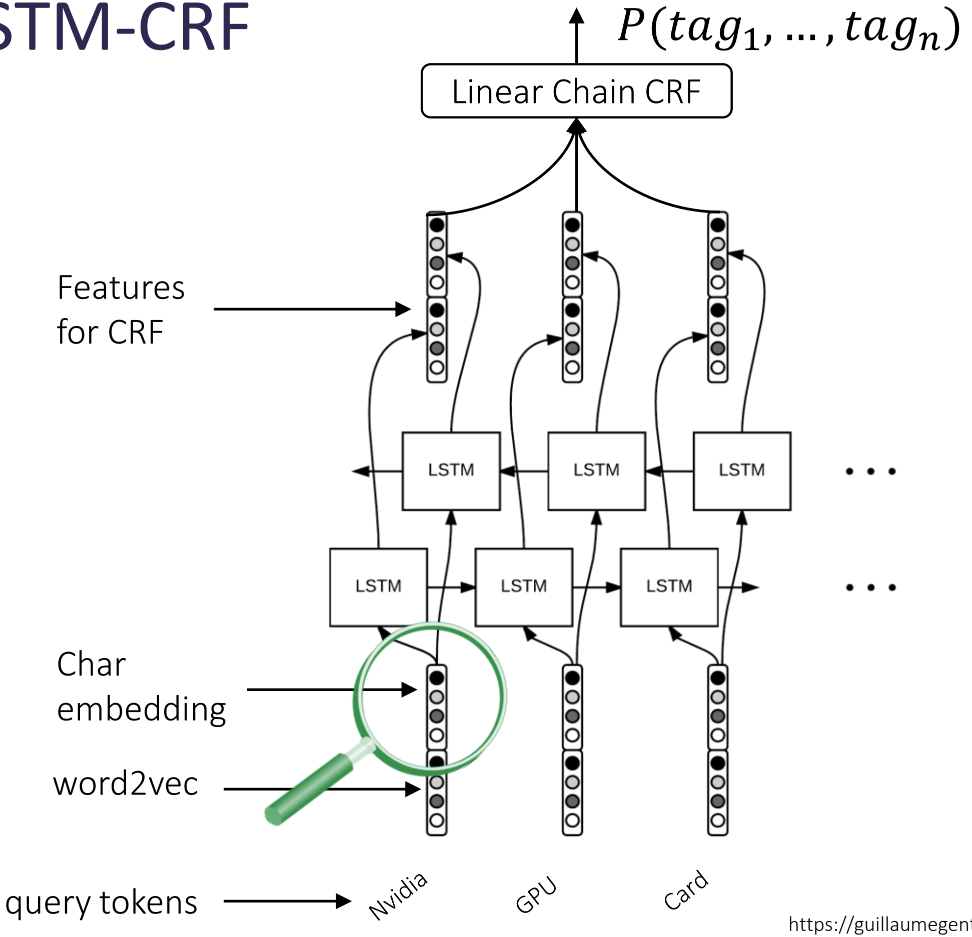
It is a hard task for human

- Is "outside” a product type token in the query, “canopy tents for outside”?

Disagreement between taggers are high (~30%)

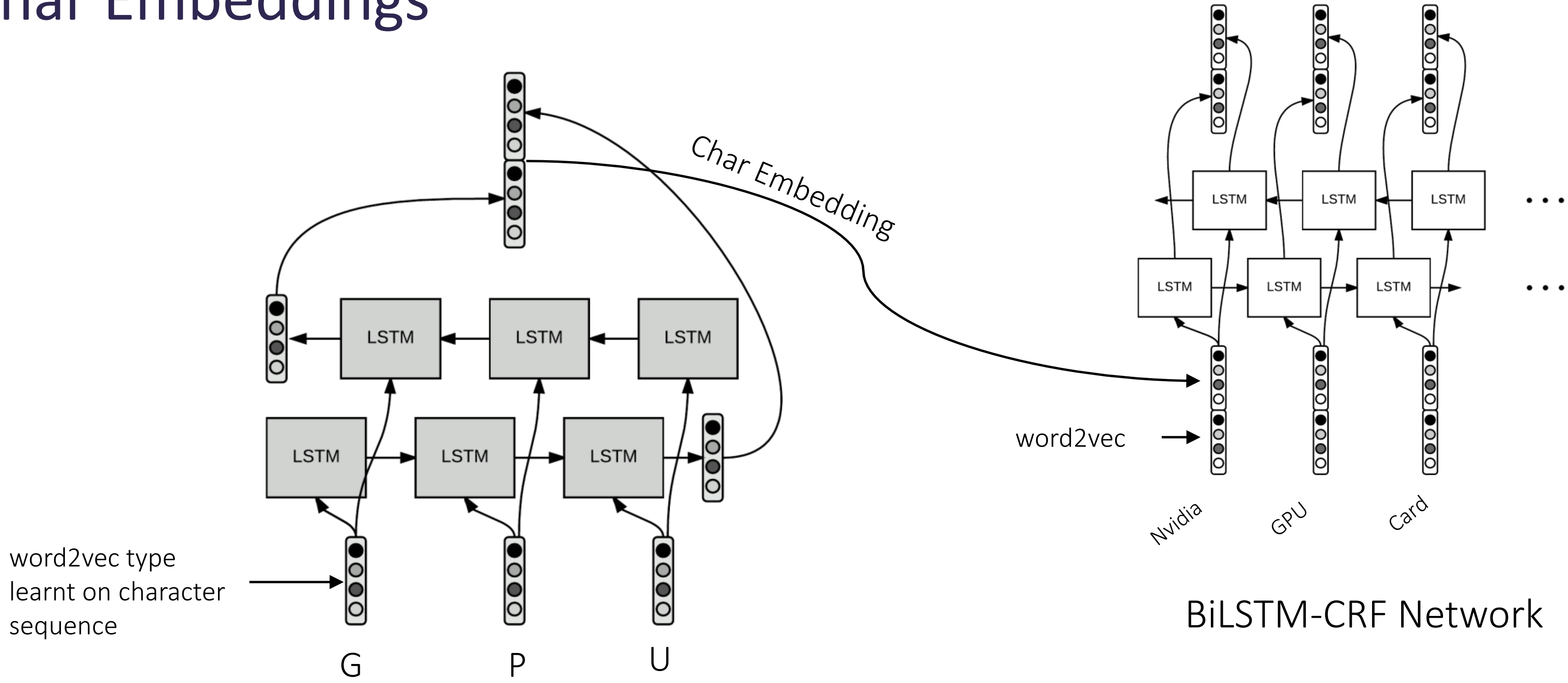
Fortunately 10K training data is a good start

Model – BiLSTM-CRF



https://guillaumegenthial.github.io/sequence-tagging-with-tensorflow.html

Char Embeddings



Character embedding network

Char Embedding

- Maps a sequence of characters to a fixed size vector
- Handles out of vocabulary words
- Handles misspellings

sansung	tv
NULL	Product Type

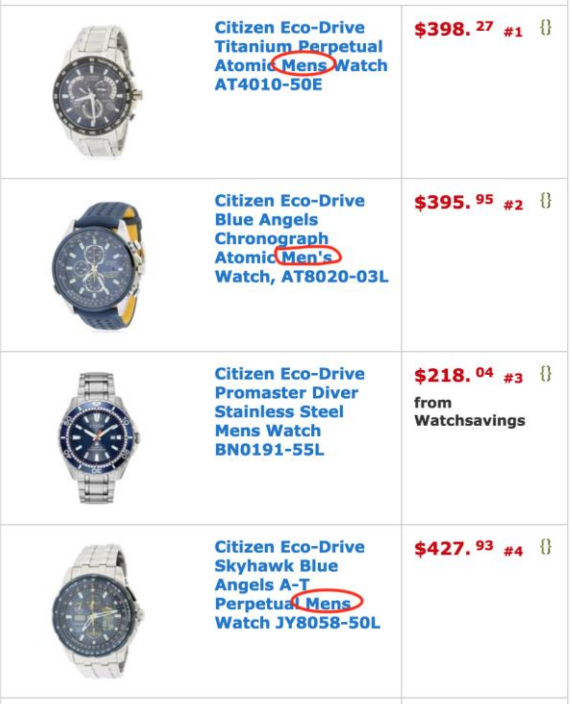
Without Char Embedding

sansung	tv
Brand	Product Type

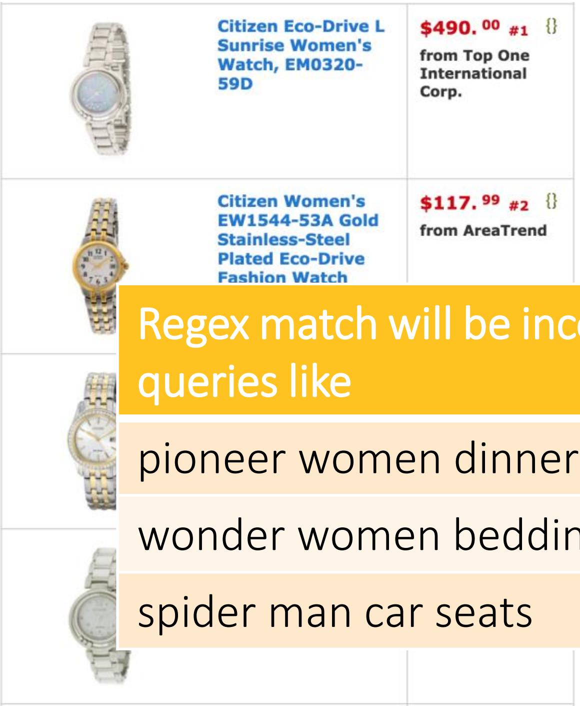
With Char Embedding

Improving search results using query tagging

Women citizen eco drive watch



Before



Regex match will be incorrect for queries like

- pioneer women dinnerware
- wonder women bedding
- spider man car seats

After understanding the Gender token

Walmart Labs

Other use cases of query tagging

TV queries / Not TV queries	Search query log
TV queries	samsung tv 32 in
	32 in vizio tv
	sanyo flat screen tv
	led tv sony 55”
Not TV queries	samsung tv stand
	sony tv remote

Search query log

Customer Demand Analysis

- Most searched brand of TV

Attribute filter suggestion

- Suggest top attributes (e.g. brand, screen size) that customers look for for in a product type query (e.g. TV)

Neural IR

Traditional IR → Semantic Search → Neural IR

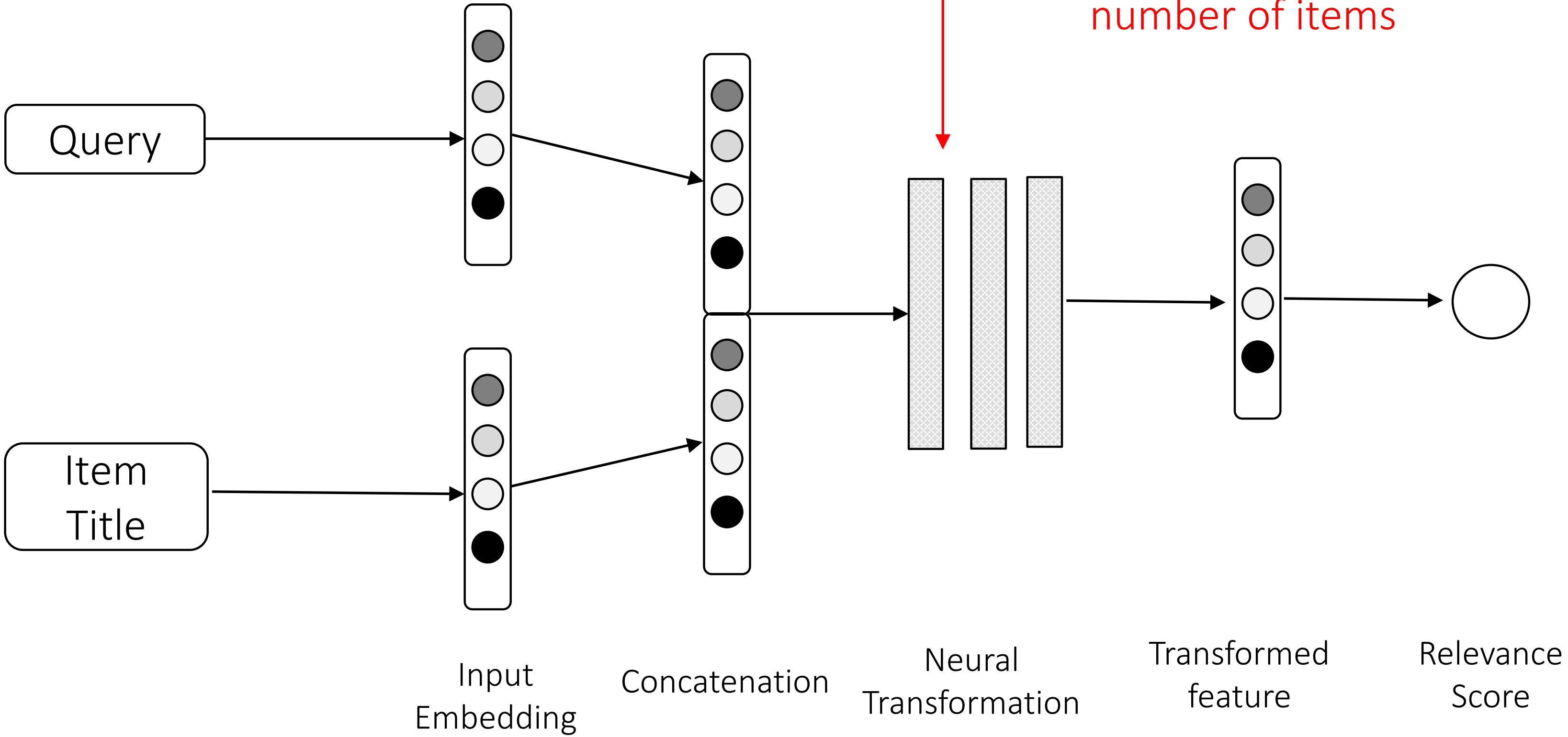
Token and synonym match
Learning to Rank

- Attribute extraction
- Token, synonym and attribute match
- Learning to rank

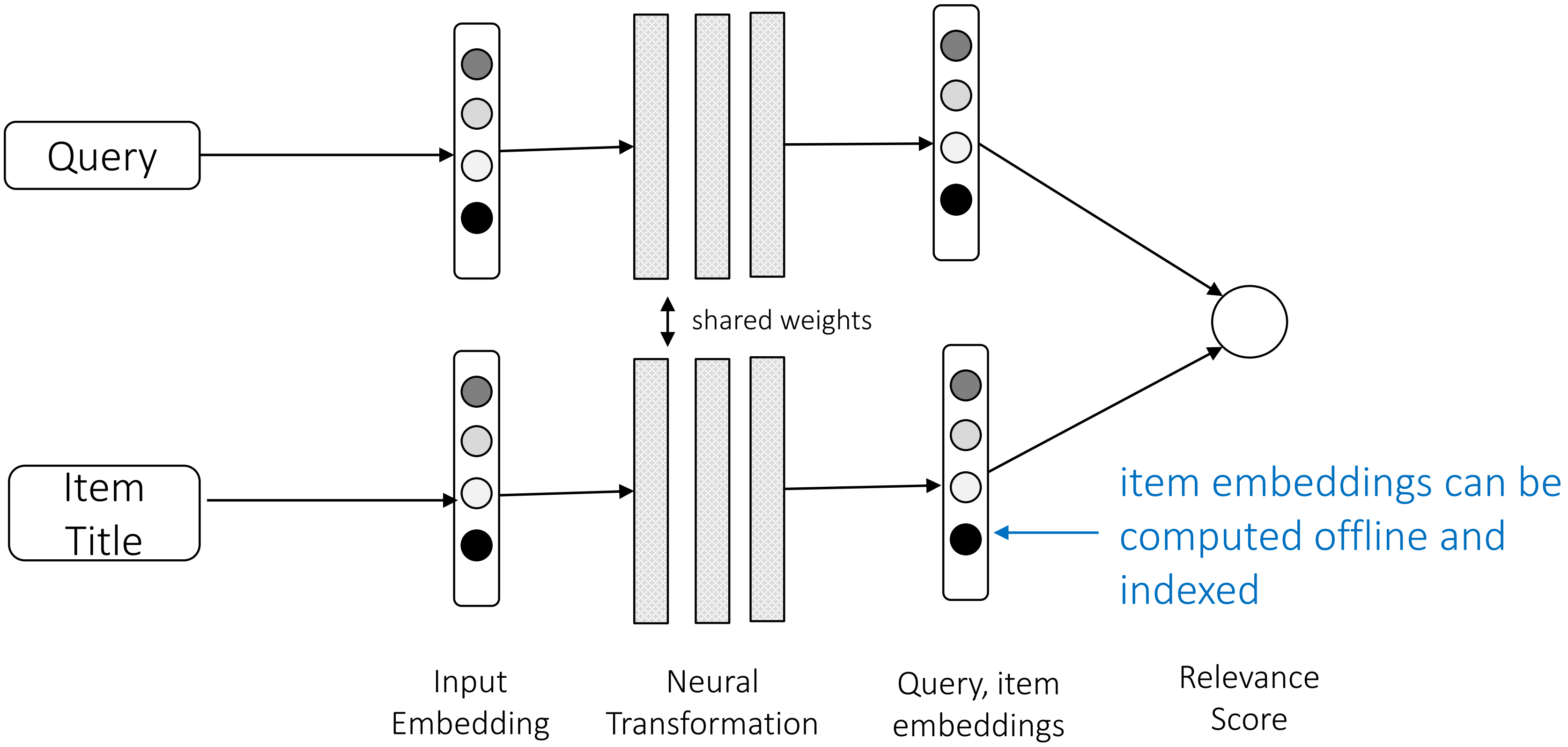
End-to-end matching and ranking

Neural IR – Design 1

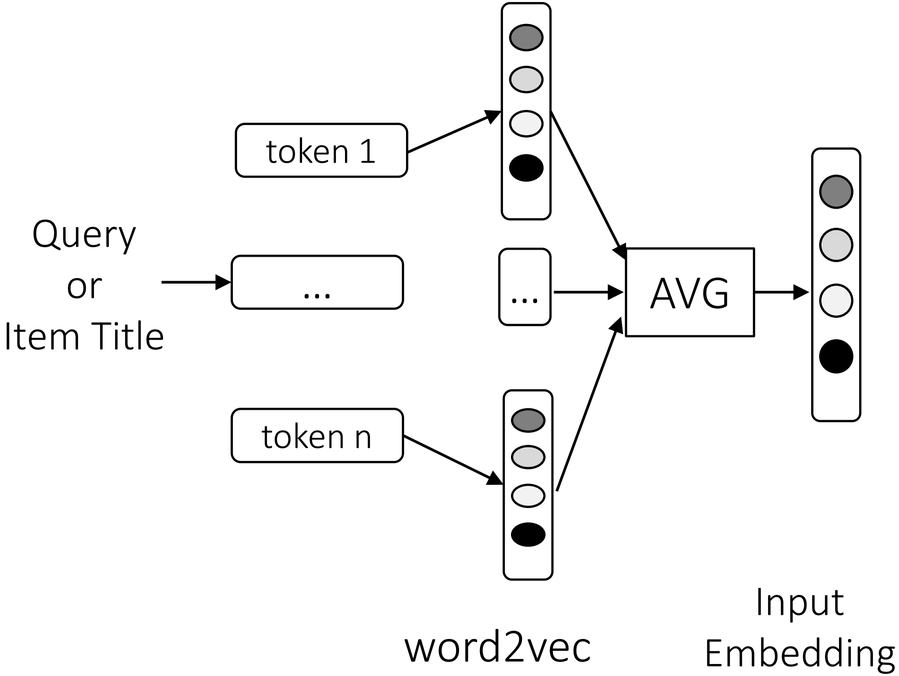
- Runtime computation
- Not scalable for large number of items

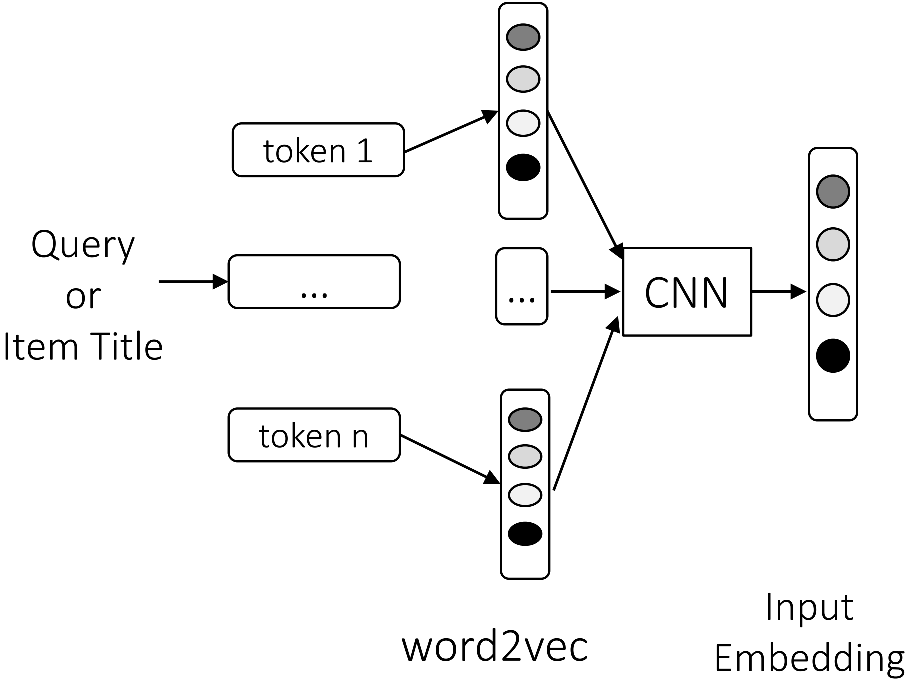


Neural IR – Design 2



Input Embedding





Comparable Accuracy

Training Data

query, item title, click through rate (ctr)*

Historical search log

*Position bias correction for ctr of a query, item pair

$$ctr = \frac{\sum_r clicks\_corrected_r}{\sum_r impressions_r}$$

$$clicks\_corrected_r = clicks_r + (impressions_r - clicks_r) * P(click \mid r)$$

$$r = rank\ at\ which\ the\ item\ was\ displayed$$

Training Loss

Point-wise

$$x_q = query\ features$$
$$x_p = item\ features$$
$$f(x_q, x_p) \rightarrow \text{ctr}$$

Regression problem
Sigmoid cross entropy loss

Pair-wise

x_q x_p x_n

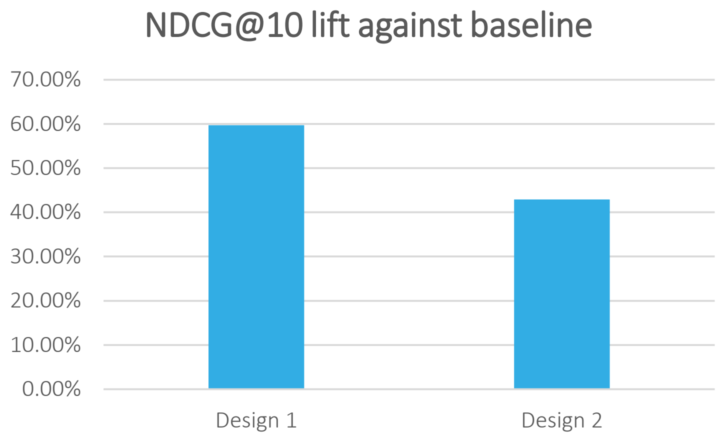
Brooks shoes

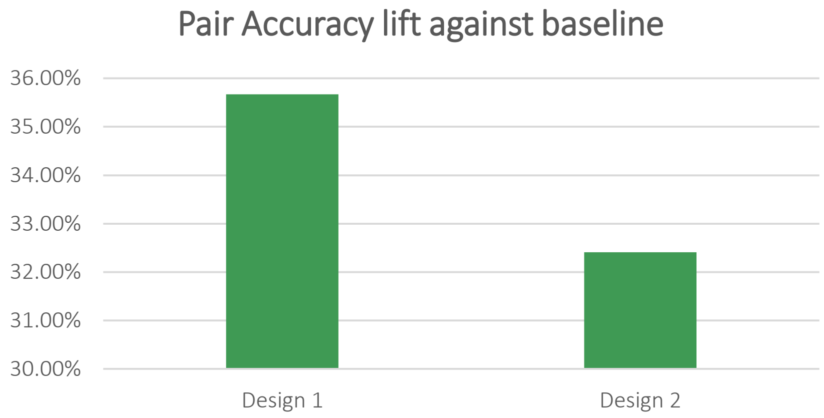
query relevant less relevant

$$f(x_q, x_p) > f(x_q, x_n)$$
$$when\ ctr(x_q, x_p) > ctr(x_q, x_n)$$

Minimize pair inversions
Pair-wise logistic loss

Accuracy on pair-wise loss





NDCG captures quality of overall ranking

Pair accuracy captures if higher ctr (relevant) items ranked above the lower ctr items

Neural IR

Pros

- End to end approach
- Enables Semantic matching implicitly
- Handles different data types (text, image)

Cons

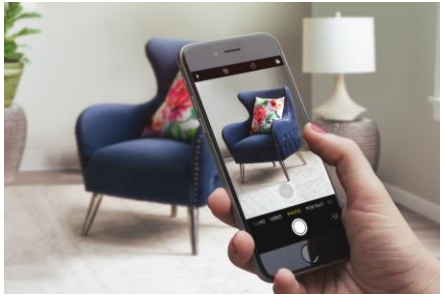
- Not scalable (yet)
- Not so successful (yet)

Image understanding

Attribute Prediction

Predicted Attributes
- Product type
- Style
- Material
- Color

Visual Search



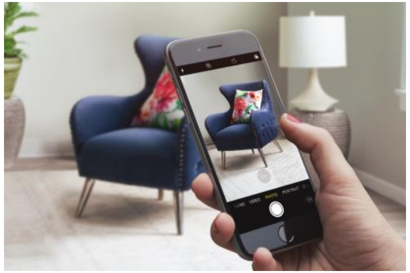
Compatible Outfit

Image understanding key learnings

Attribute Prediction

- Multi-task learning is more accurate
- Predicting style is harder than predicting product type

Visual Search



- A/B test on hayneedle.com
- Comparable results against a well established startup

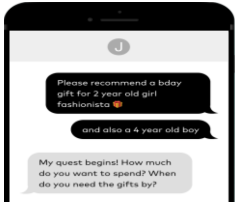
Compatible Outfit

- Under exploration
- Early results beating token based approach

Future

Evolution of mobile phone

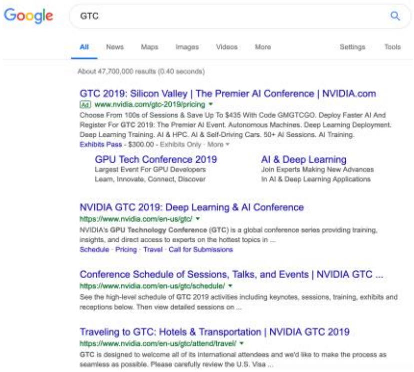
Future



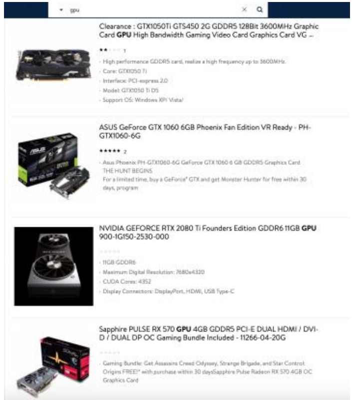
Conversational commerce

V-Commerce

Seamless search and personalized results



Web Search



E-commerce Search

Thank You