Multimodal Search in E-commerce

Mariya Hendriksen

August 1, 2019

AIRLab, ICAI, University of Amsterdam

Speaker Information

Mariya Hendriksen

1st-year Ph.D. candidate at AIRLab (bol.com & University of Amsterdam)

Topic: Multimodal Search in E-commerce

Contact information: mhendriksen@bol.com

Outline

1. Introduction
2. eBay challenge
3. Conclusions

Introduction

What is Multimodal Search?

Problem

Current e-commerce search is rather limited and time-consuming

Solution

- Multimodal search
- Modalities: text, image, audio, location

Advantage

Better grasp of customer needs

"Find similar dress but short and with lace"

Multimodal Search

Idea

Represent modalities in a common space.

Major approaches

- Real-valued representation learning: projection into real-valued common space.
  Examples: subspace learning, topic modelling, etc.
- Binary representation learning (hashing): projection into common Hamming space.
  Examples: linear and nonlinear modelling.

eBay challenge

Challenge Introduction

Idea

Given: 150 textual queries, 900k multimodal (text + image) documents
Goal: identify which documents are relevant for every query.

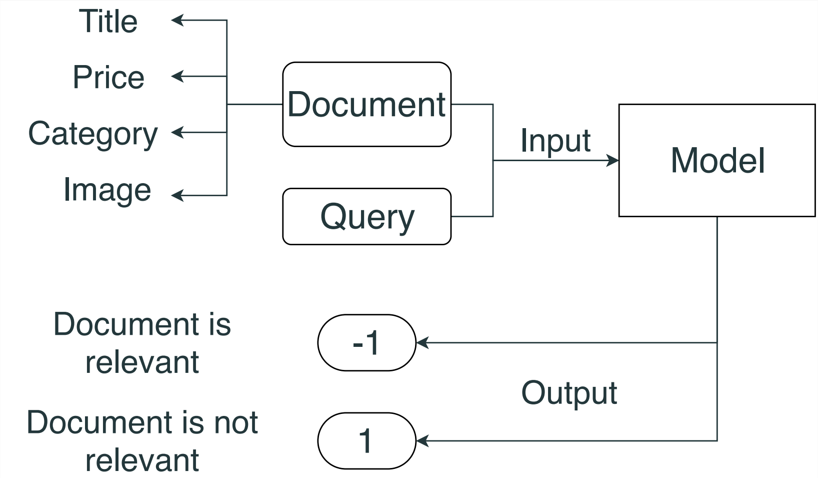
Data Set

Partially labelled documents and queries from eBay.

Document Example

1. Title: Bally Twilight Zone Pinball Machine
2. Price: 3995.00
3. Category: Collectibles > Arcade, Jukeboxes & Pinball > Pinball > Machines
4. image URL: `https://i.ebayimg.com/...`

Task Overview



Text Representation

Pretrained Models

- word2vec pretrained on Google News
- Global Vectors pretrained on Common Crawl

Sequence representation approaches

- Average of tokens embeddings comprising the sequence
- Encoding with long short-term memory encoder

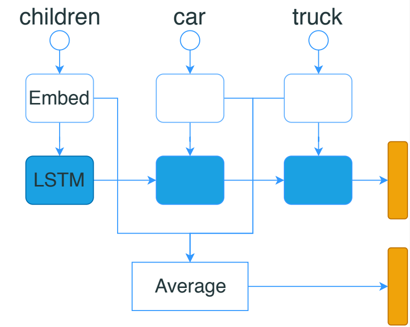


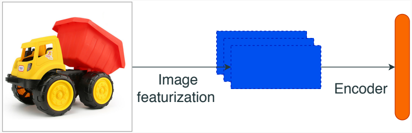
Image Representation

Pretrained Model

- ResNet-50 pretrained on ImageNet

Image representation

1. Image feature extraction with ResNet50
2. Feature dimensionality reduction with encoder



Model Architecture

Considered Models

- textual matching: BM25, n-gram
- single model classifier: support vector machines, neural network with mixed input
- ensemble classifier: gradient boosting classifier

Selected Models

- Baseline: BM25
- Neural network with mixed input
    - Titles: average of token embeddings comprising the title
    - Categories: average of token embeddings comprising the category breadcrumbs, excluding the first category
    - Prices: represented on the log scale

Results & Challenge Overview

Majority of teams utilized only textual data (titles + categories)

Models which outperformed ours:

- Relied on feature engineering: query expansion, length and digits counts, category hashing, price binning, etc.
- ensemble of BERT models

Model Performance Example

Query: 'roman replica'

True Positive

Title: 'Medieval Armour Roman Legionary's Belt For Rome's Legion Collectible Replica'

Price: 73.5

Category: 'Collectibles > Militaria > Pre-1700' > Reenactment & Reproductions

Image:

False Positive

Title: 'Medieval Armor Middle Age Knights Tasset Battle Plated Steel Waist Replica Item'

Price: 92.11

Category: 'Collectibles > Knives, Swords & Blades > Armor & Shields'

Image:

Conclusions

Challenge Conclusions & Future Work

Challenge limitations

- limited amount of queries
- only textual queries
- small labelled dataset

Future Work

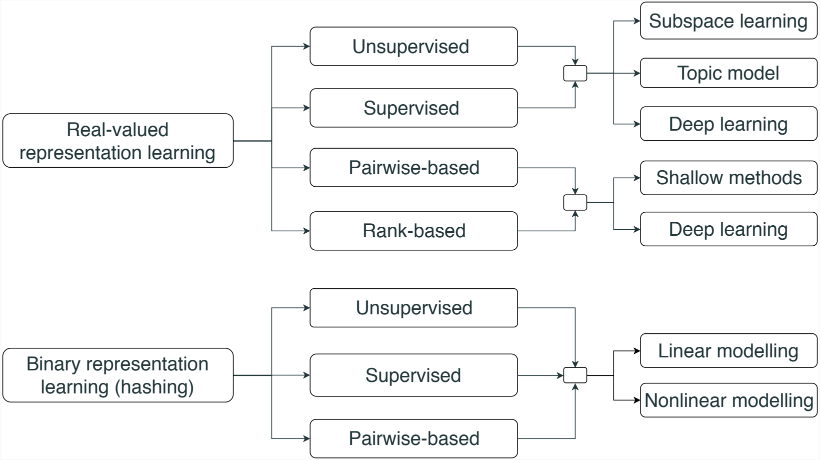
- collect a bigger, more complex dataset from bol.com
- refactor the model so that it can efficiently learn multimodal representations

Summary

- Introduction of **multimodal search**, a type of search which utilizes multiple modalities such as image, text, video, etc.
- Overview of the **eBay challenge**, a challenge where the participants are to predict whether the given document is relevant for the given query
- A run-through of **AIRLab's entry** of the challenge
- Summary of challenge **limitations** and **future work**

Appendix

Approaches


Real-valued representation learning
Unsupervised
Supervised
Pairwise-based
Rank-based
Subspace learning
Topic model
Deep learning
Shallow methods
Deep learning
Binary representation learning (hashing)
Unsupervised
Supervised
Pairwise-based
Linear modelling
Nonlinear modelling

Data Set Statistics

	Dev Set	Test Set
Queries	150	150
Documents	65061	899287
(query, document) pairs	66053	134893050

Baseline vs. Our Model

Model	Avg. Precision	Avg. Recall	Avg. F_1	F_1
Baseline	0.56	0.04	0.06	0.06
Our model	0.73	0.39	0.34	0.60