VISUM 2021

- Project Component -

INESCTEC

VCMI VISUAL COMPUTING & MACHINE INTELLIGENCE RESEARCH GROUP

FARFETCH

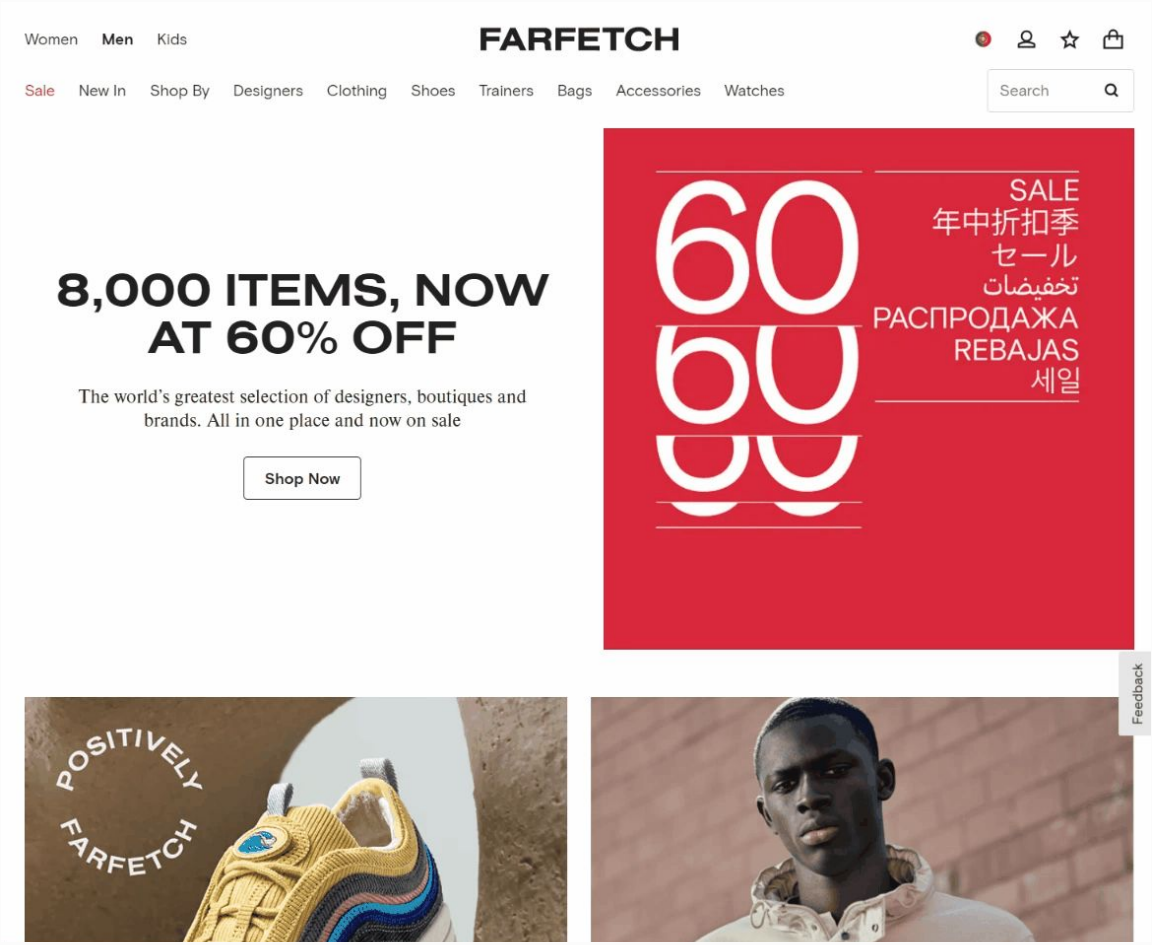
powered by: Google Cloud Platform

Index

- Problem Presentation and Motivation
- Baseline Solution
- Framework
- Participation
- First steps

Problem & Motivation

FARFETCH


Women
Men
Kids
FARFETCH
Sale
New In
Shop By
Designers
Clothing
Shoes
Trainers
Bags
Accessories
Watches
Search
8,000 ITEMS, NOW AT 60% OFF
The world's greatest selection of designers, boutiques and brands. All in one place and now on sale
Shop Now
60
SALE
年中折扣季
セール
تخفيضات
РАСПРОДАЖА
REBAJAS
세일
POSITIVELY
FARFETCH
Feedback

FARFETCH

a platform with over

2.0M

active consumers

a fast growing database of customers spending an average of

$608

per order

orders shipped to more than

190

countries worldwide

Farfetch partners with over

1,200

of the world's best brands, boutiques and department stores

approximately

50%

of traffic comes from iOS devices

over

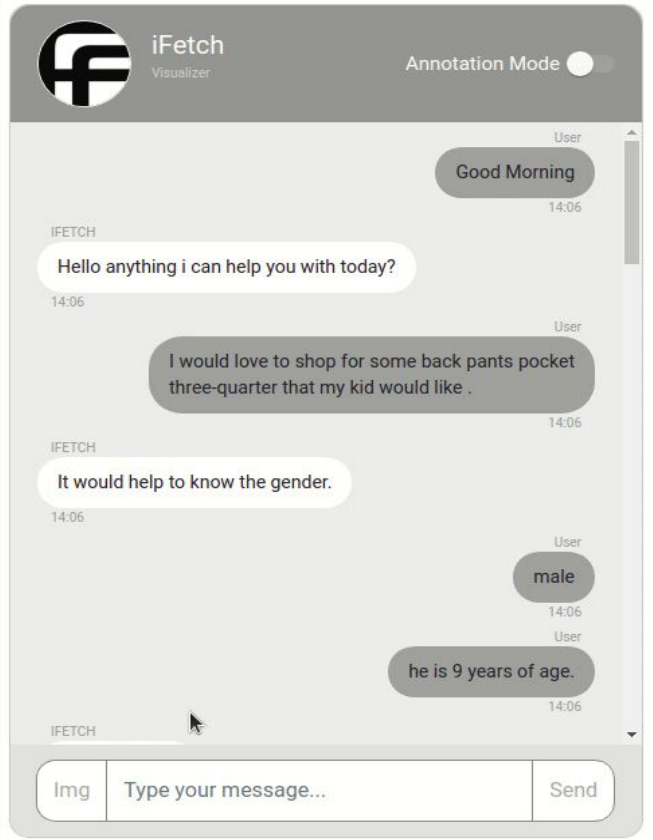
4532

highly qualified and multilingual staff in 14 locations worldwide

High-end customers expect a **high-end experience**

iFetch

Multimodal conversational agent for the online fashion marketplace



Vision

To make a step change in the online high-fashion marketplace by **advancing conversational AI** with multimodal capabilities.

Goal

Through conversation, iFetch aims to **provide targeted advice and "physical store-like" experience** while maintaining user engagement.

iFetch

by FARFETCH x CMU x TÉCNICO LISBOA x NOVA

iFetch

Multimodal conversational agent for the online fashion marketplace

iFetch will have the capability of **suggesting products that "*goes well with*"**.

USER
Can you show me pants that goes well with this jacket?

IFETCH

...

The challenge

Fashion outfit complementary product retrieval

Fill in the Blank (FITB) task

Dataset

NOTE that the VISUM challenge dataset is a subset of this FARFETCH dataset!

	FARFETCH dataset
Product representation	Image; Category; Name; Description
# outfits	128398
# unique products	149944
min/max/avg # products per outfit	2 / 14 / 4.5
# categories	133

An outfit example:

Id	productname	category	description
1	cashmere knit trackpants	Trousers	Black cashmere cashmere knit trackpants from BARRIE featuring drawstring fastening waist, slip pockets to the sides, tonal stitching, ribbed detailing and elasticated cuffs.
2	floral embroidered V-neck jumper	Knitwear	Grey cashmere floral embroidered V-neck jumper from BARRIE featuring V-neck, floral embroidery, long sleeves and ribbed detailing.
3	silk shirt	Tops	White silk shirt silk shirt from SAINT LAURENT.

Baseline Solution

Paradigm for Complementary Product Retrieval

- Follow a deep metric learning approach

To **learn a "style" embedding space**, where complementary products are close to each other and products that do not fit are far apart.

From Vasileva *et al* in Learning Type-Aware Embeddings for Fashion Compatibility

Baseline model

Image encoder:

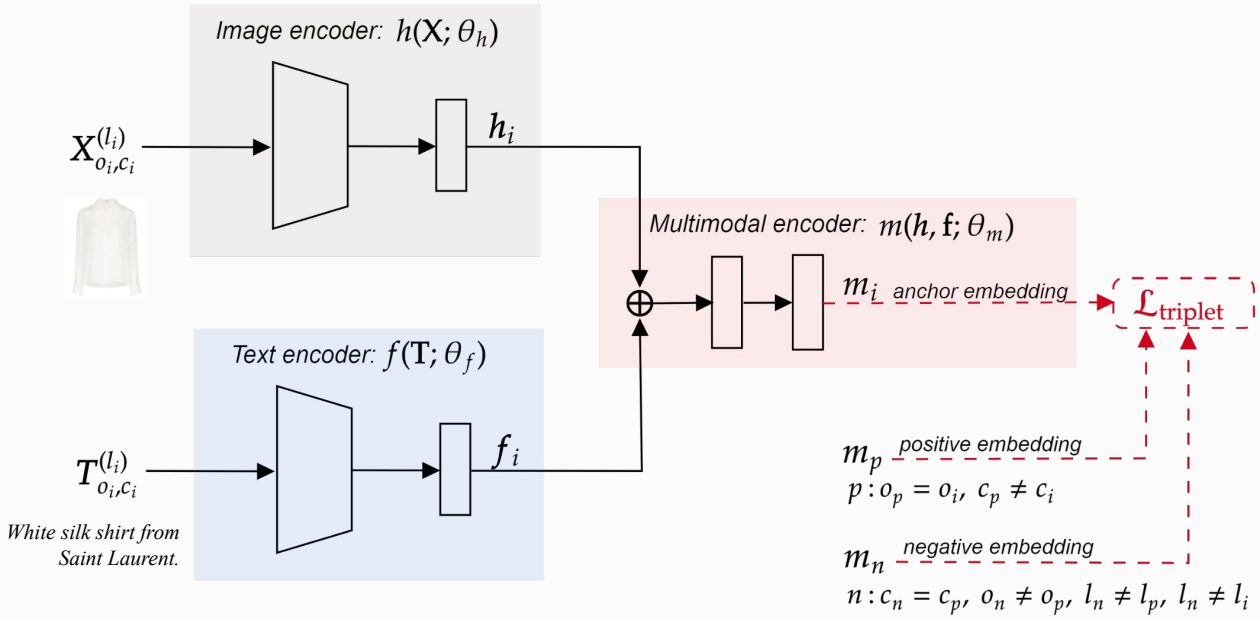
- Maps from a product image $\boldsymbol{X}$ to a feature representation $\boldsymbol{h}$.
- ResNet-50 + Projection block (with additional trainable layers).

Text encoder:

- Maps from a text description $\boldsymbol{T}$ to a latent textual representation $\boldsymbol{f}$.
- DistilBERT + Projection block.
- CLS hidden representation is used to summarize the entire product description.

Multimodal encoder:

- Maps from both $\boldsymbol{h}$ and $\boldsymbol{f}$ to a multimodal complementary feature space.
- Merge layer + Projection block.



Training

Training objective

- Model parameters are optimized via a ranking loss, i.e., triplet loss $\mathcal{L}_{triplet}$:

$$\mathcal{L}_{triplet} = \frac{1}{N}\sum_{i=1}^{N}\left[\left\|m_{o_i,c_i}^{(l_i)} - m_{o_p,c_p}^{(l_p)}\right\|_2^2 - \left\|m_{o_i,c_i}^{(l_i)} - m_{o_n,c_n}^{(l_n)}\right\|_2^2 + \alpha\right],$$

Anchor · Positive · Positive · Negative · Margin

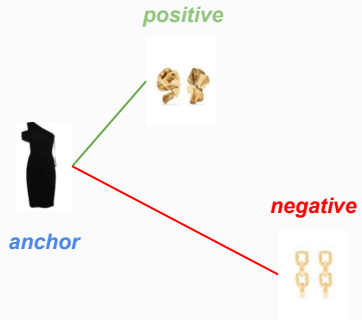
where:

$o_p = o_i \quad c_p \neq c_i$

$c_n = c_p,\ o_n \neq o_p,\ l_n \neq l_p,\ l_n \neq l_i$

Triplet mining constraints

Forces the **distance between non-complementary product embeddings** (*anchor* and *negative* samples) to be **larger than** the **distance of complementary product embeddings** (*anchor* and *positive* samples) by a margin.



Training

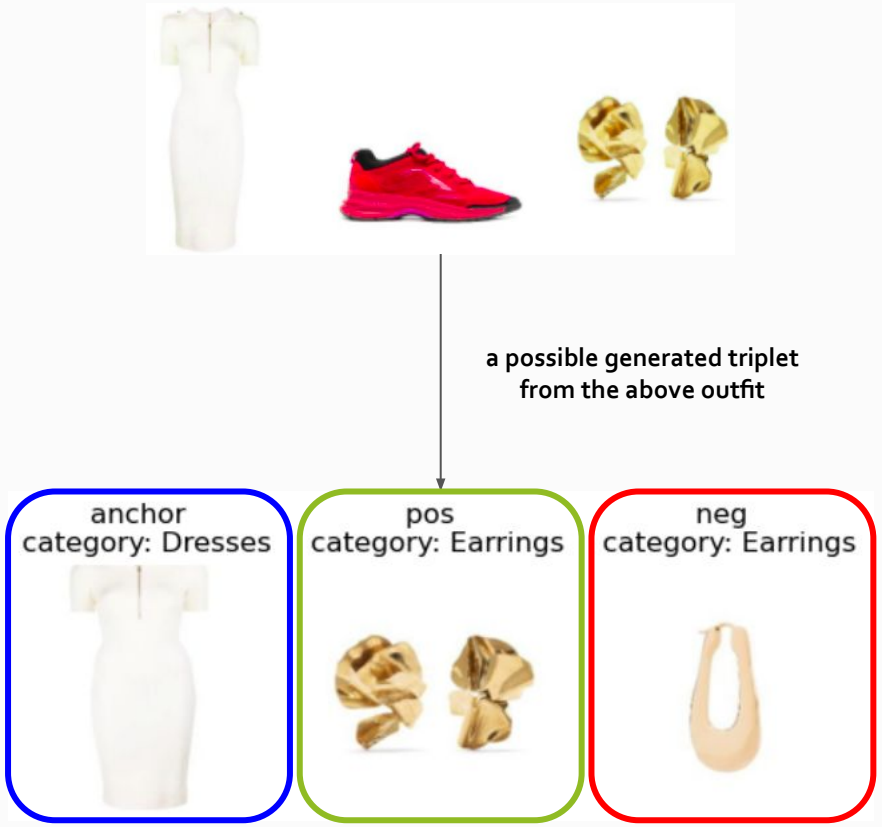
Triplet mining process

***Anchor* and *positive* pairs sampling**:

- Randomly selected products from the same outfit (o), but from a different category (c), i.e. $o_p = o_i\,,\ c_p \neq c_i$

***Negative* sampling**:

- A randomly selected product of the same semantic category of the *positive*, but from a different outfit, i.e. $c_n = c_p,\ o_n \neq o_p,$



However, this may lead to a **large number of false *negative* products**, especially when the outfits in the training dataset share a large number of products.

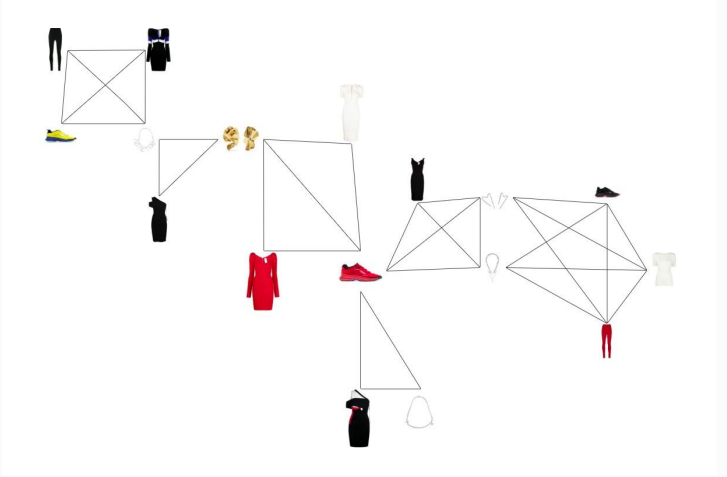
Training

Triplet mining process

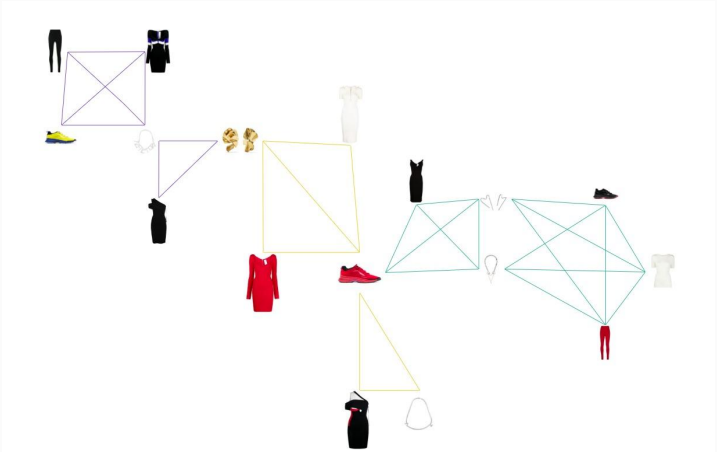
***Negative* sampling with the Louvain constrain:**

- We further constrained the *negative* sampling process on the product communities (*l*), obtained by applying the Louvain method on the product's graph.

- A *negative* product should be not only from a different outfit but also from a different community graph than the *anchor* and the *positive*, i.e. $l_n \neq l_p,\ l_n \neq l_i$



Product graph for 7 outfits



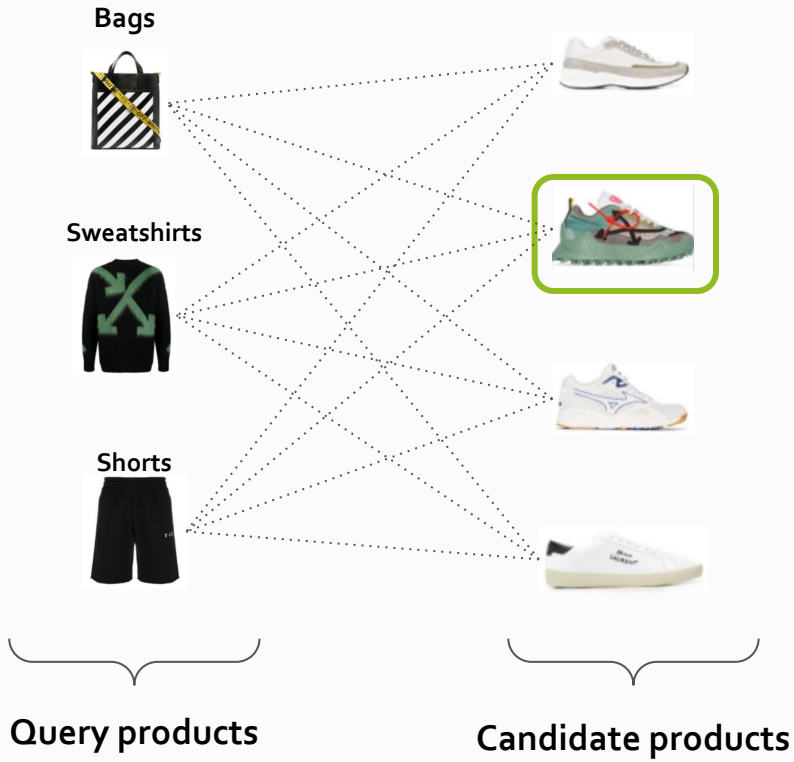
Generated Louvain communities
(each color represents a community)

Inference

Given a FITB test query*:

1. Compute the embedding distance between each candidate and every query product;

2. Return the candidate product with the lowest sum of the distances.



***comprises of a set of query products and a set of candidate products:**

- *1 Positive*: a randomly selected product from a given test outfit;
- *3 Negatives*: randomly sampled products of the same category as the positive from other test outfits.

Results

The impact of different modalities and the Louvain constraint

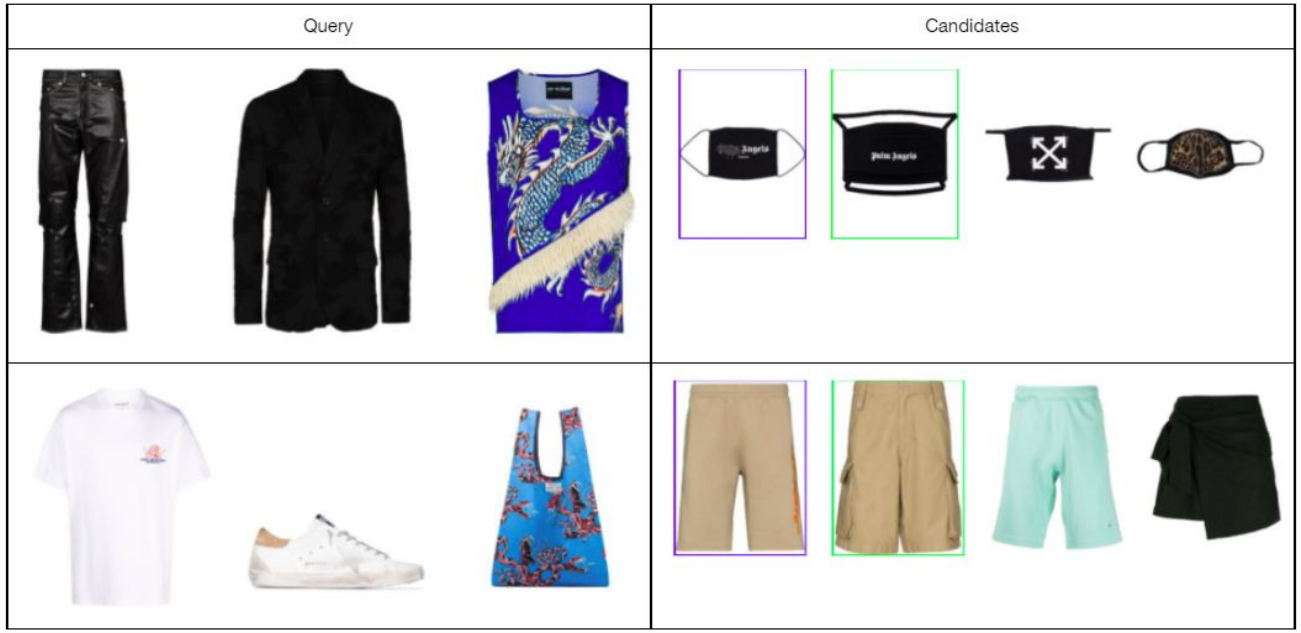
Modality	Best Valid loss		Test FITB acc	
	w/o Louvain	w/ Louvain	w/o Louvain	w/ Louvain
Image	0.791	0.760	0.442	0.454
Text	0.831	0.792	0.433	0.471
Multimodal	**0.751**	**0.715**	**0.472**	**0.483**

Results

Correct predictions (violet box)

Results

Wrong predictions (violet box) along with ground-truth (green box)

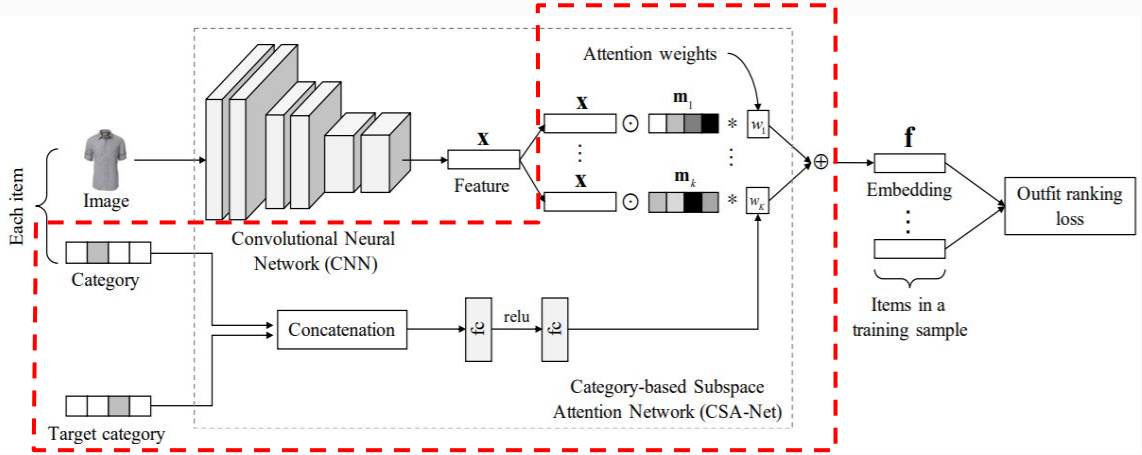


[Some ideas for improvement]

- Learn multiple embedding subspaces;
- Capture interactions of all products in an outfit.

Ideas for improvement

1) Learn multiple embedding subspaces



Lin et al, ***Fashion Outfit Complementary Item Retrieval***, CVPR 2020.

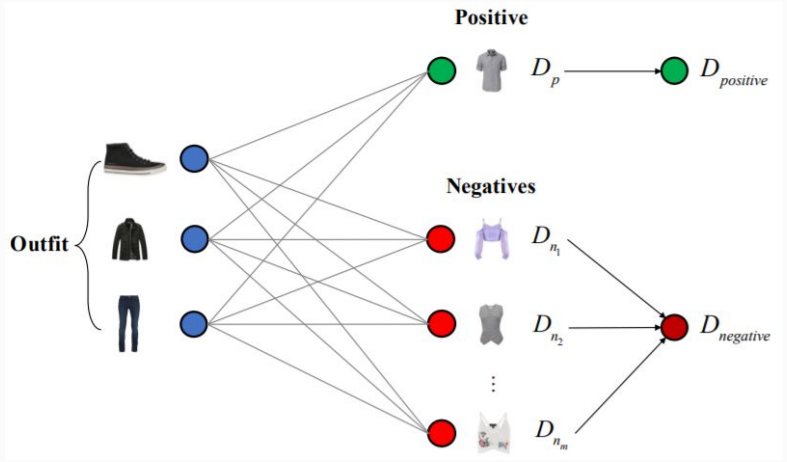
- To encode different notions of product complementary.
- Learn category-aware subspaces to learn category-specific notions of complementary.

Other important refs:

Tan et al, ***Learning Similarity Conditions Without Explicit Supervision***, ICCV 2019.
Vasileva et al, ***Learning Type-Aware Embeddings for Fashion Compatibility***, ECCV 2018.

Ideas for improvement

2) Capture interactions of all products in an outfit



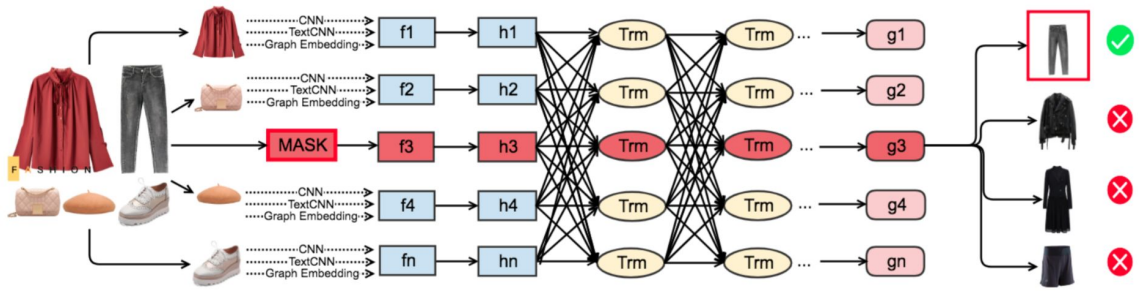
Lin et al, ***Fashion Outfit Complementary Item Retrieval***, CVPR 2020.

2.1 Outfit ranking loss:

- considers the similarities among all existing items in the outfit.

Ideas for improvement

2) Capture interactions of all products in an outfit



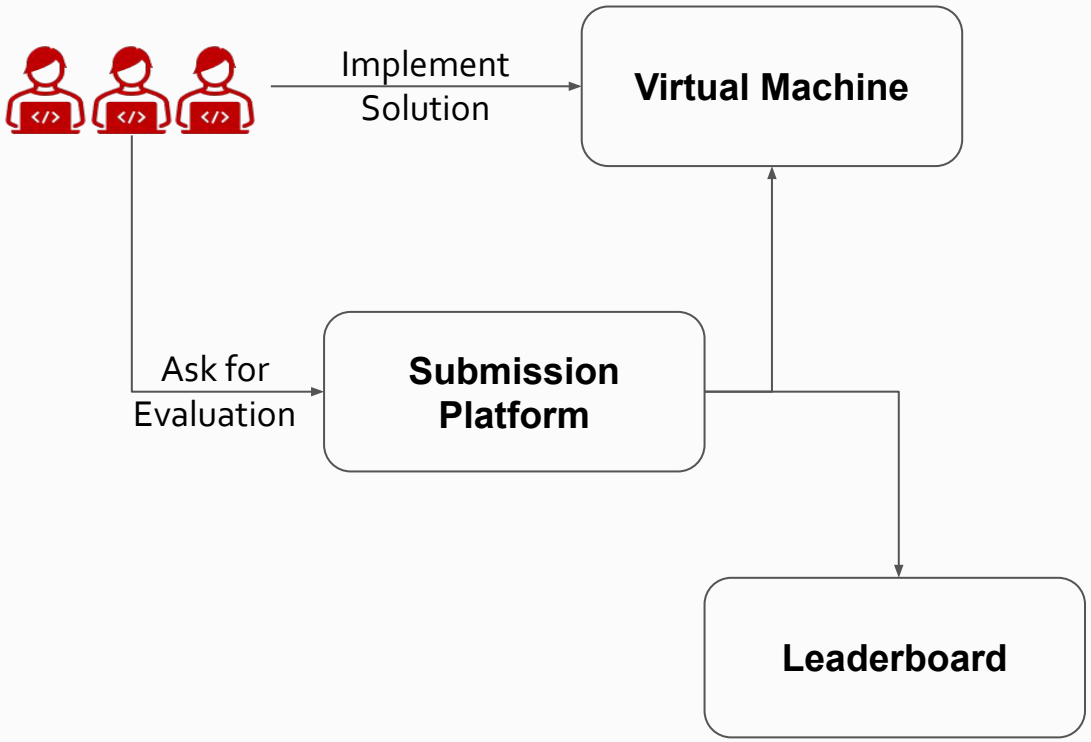
Chen et al, [POG: Personalized Outfit Generation for Fashion Recommendation at Alibaba iFashion](), KDD 2019.

2.2 Transformer-based models (e.g., BERT)

- Product compatibility modeled via self-attention (each product attends to all other products);
- Outfit represented as a set (by removing the positional encoding as well as the next sentence task);
- Trained to predict the masked product in the outfit.

Framework

Framework Overview



- ➔ Best teams in the final leaderboard will be invited to present their work in the last day;
- ➔ Among this a jury will elect the winning team, taking into consideration your solution's accuracy and your presentation;

Virtual Machine - Setup

- Each Team will be assigned a Google Cloud Virtual Machine and its respective access credentials:
  - user: visum
  - pass: ********

- The machine has the following specifications:
  - GPU - 1 X NVIDIA Tesla T4
  - CPU - 4 vcpu's (Intel Broadwell)
  - RAM - 15GB
  - Storage - 60GB

Virtual Machine - Software

- Machines will have useful software installed (CUDA, torch, tensorflow, etc.);

- You may install python packages using **“pip”**. Do not use virtual environments;

Virtual Machine - Files

- Each machine will have the baseline solution at:
  - /home/visum
- Each machine will have a copy of the train data at:
  - /home/master/dataset/train

- Each machine will have a **dummy test folder (same as train)** which allows you to verify your submission is valid;

Virtual Machine - Protocol

- Develop a solution to the FITB problem using python and the provided training data.
  - the main file: “/home/visum/test.py”
  - this file should look at:
    - “/home/master/dataset/test/queries.csv”
    - “/home/master/dataset/test/options.csv”
    - “/home/master/dataset/test/product_images/”
    - “/home/master/dataset/test/df_products.csv”
  - this file should generate:
    - “preds.csv” at the same directory as test.py

Submission Platform

To access the VISUM 2021's submission platform, click the following link: https://visum.inesctec.pt/submissions

Submission Platform - two per day

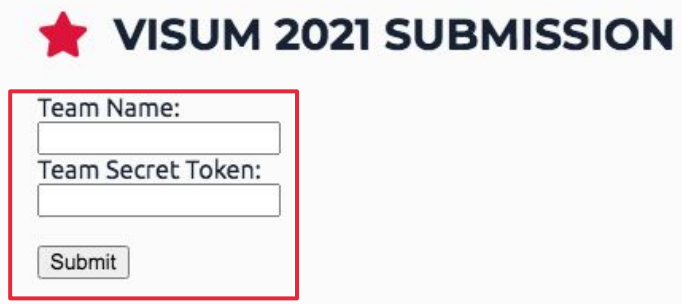
- Each Team will have two daily test submissions:
  - Daily Individual Submission - which consists of an individual request of the Team in the submission platform (from 5.00 am until 11.30 pm, GMT+1)
  - Daily Global Submission - which consists of the automatic execution of the test for all the Teams (at 12.00 Midnight, GMT+1)

- These are run on a **subset** of the **private test data**;

Submission Platform - Individual

- To perform a submission request you need to fill the form with the name of your Team and the Secret Token that will be assigned

- After submission, the system will:
  - copy your /home/visum to our system;
  - run your test.py on private data;
  - evaluate your accuracy and print it on the leaderboard;



Leaderboard

➔ All the results will be updated at our Daily Leaderboard:

- **Last Accuracy** - the accuracy obtained for the last submission;
- **Best Accuracy** - the best accuracy obtained so far;
- **Status** - the submission status for the Team
  - "Waiting for submission"
  - "Running submission"
  - "Daily submission done"

VISUM 2021 DAILY LEADERBOARD

Position	Team	Last Accuracy	Best Accuracy	Status
1st	VISUM	No data	No data	Running submission
1st	Farfetch	No data	No data	Running submission

Participation

Registering your Team

- Each Team should be composed of 3 people
  - Find a Team Name and send the information about the Team to visum@inesctec.pt;
- All team members must have signed the NDA agreement;
- We will provide you with:
  - A VM machine;
  - A secret token for the submission platform;
- We will also create (private) text and voice channels in **Discord** for your Team;

Project Sessions -Discord

- ➔ Mainly autonomous work from the group members on the Discord server;
- ➔ During this time feel free to ask for help either with technical problems, simple questions or to briefly discuss your approach to the problem. For efficiency please post a quick message in **#request-help**:
  - e.g. “hey, I need assistance understanding the baseline”
- ➔ We will make an effort to check in with the groups to make sure everything is running smoothly.

Brainstorm Sessions

➔ You will have the opportunity to participate in the following brainstorm sessions regarding the project:

- **6th July (14h) - Learn multiple embedding subspaces**
- **7th July (14h) - Capture interactions of all products in an outfit**
- **7th July (11h30) - AI shot on the topic of the project**

Bring your own ideas to the discussion!

Project Presentation & Prizes

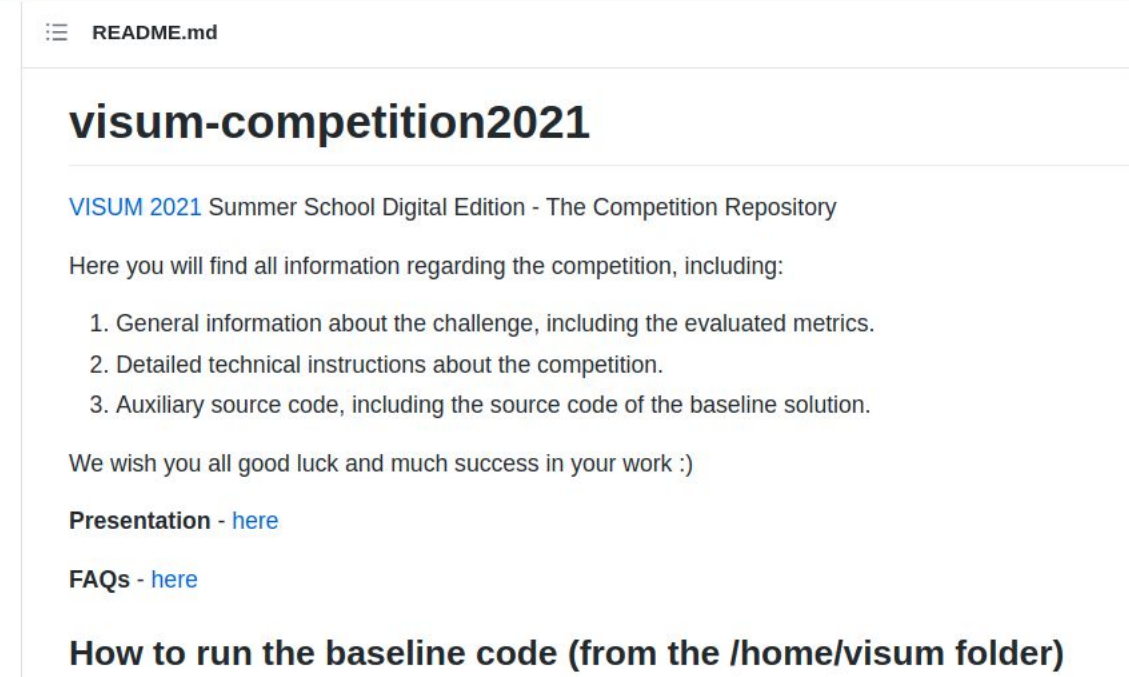
- The best Teams in the Final Leaderboard are invited to present their approach on Friday (9th of July)

- These Teams will be eligible to win the competition along with the following prizes:
  - 3 FARFETCH vouchers (150€ each);
  - 3 VISUM 2022 registrations.

First Steps

The Competition Repository

➔ https://github.com/visum-summerschool/visum-competition2021



README.md

visum-competition2021

VISUM 2021 Summer School Digital Edition - The Competition Repository

Here you will find all information regarding the competition, including:

1. General information about the challenge, including the evaluated metrics.
2. Detailed technical instructions about the competition.
3. Auxiliary source code, including the source code of the baseline solution.

We wish you all good luck and much success in your work :)

Presentation - here

FAQs - here

How to run the baseline code (from the /home/visum folder)

1) Access the baseline source code

→ Accessing the files already present in your machine;

or

→ **Cloning** the project repo into your virtual machine:

```
git clone https://github.com/visum-summerschool/visum-competition2021.git
cp -r visum-competition2021/ .
rm -r visum-competition2021
```

2) Mandatory pre-processing steps

➔ Create the training and validation splits in `/home/visum/processed_data`

```
python3 split_data_nondisjoint.py
```

➔ Generate the louvain communities.

```
python3 generate_community_prods.py
```

3) Train the baseline

➔ Train the baseline

```
python3 train.py
```

➔ Model weights and checkpoints are saved in `/home/visum/results/<timestamp>`

➔ copy your best model (and the tokenizer) into `/home/visum` by doing:

```
cp -r results/<timestamp>/tokenizer results/.
cp results/<timestamp>/best_model_weights.pth results/.
```

4) Evaluate the baseline

1. Generate queries from your validation split:

```
python3 generate_test_queries.py
```

2. Generate predictions with

```
python3 test.py -t processed_data/valid
```

3. Evaluate your model

```
python3 evaluate.py preds.csv processed_data/valid/solutions.csv
```

5) Verify your submission is valid:

➔ Test your model

```
cd /home/visum/
rm preds.csv
python3 test.py
python3 evaluate.py preds.csv /home/master/dataset/test/solutions.csv
```

6) Submit your results

→ Submit your results by accessing: http://vcmi.inesctec.pt/visum/submissions.html

INESCTEC

VCMI VISUAL COMPUTING & MACHINE INTELLIGENCE RESEARCH GROUP

FARFETCH

powered by: Google Cloud Platform