LT2326 Project

Shopee product matching

Xiumei Xue
Oct 28, 2024

CONTENT

Motivation and Background

Why Product matching?

- **Product matching** allows a company to offer products at rates that are competitive to the same product sold by another retailer.
- Two different images of similar wares may represent the same product or two completely different items. Retailers want to avoid misrepresentations and other issues that could come from conflating two dissimilar products.
- Currently, a combination of deep learning and traditional machine learning analyzes image and text information to compare similarity. But major differences in images, titles, and product descriptions prevent these methods from being entirely effective.

Shopee

a leading e-commerce online shopping platform

Dataset Overview

Dataset

- Download data using Kaggle api

```
kaggle competitions download -c shopee-product-matching
```

- File structure

[train/test].csv - the training set metadata. Each row contains the data for a single posting. Multiple postings might have the exact same image ID, but with different titles or vice versa.

`posting_id` - the ID code for the posting.
`image` - the image id/md5sum.
`image_phash` - a perceptual hash of the image.
`title` - the product description for the posting.
`label_group` - ID code for all postings that map to the same product. **Not** provided for the test set.

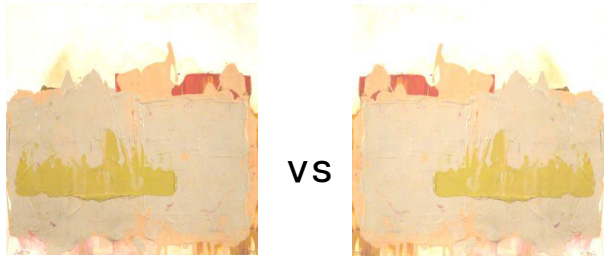
[train/test]images - the images associated with the postings.

sample_submission.csv - a sample submission file in the correct format.

`posting_id` - the ID code for the posting.
`matches` - Space delimited list of all posting IDs that match this posting. Here we focus on label matches.

Dataset Overview

Example



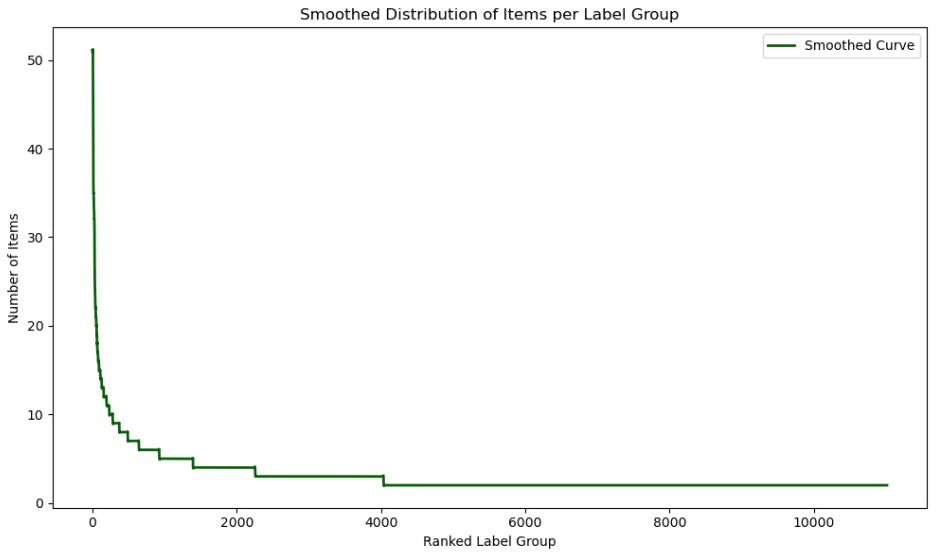
É

- Postings

posting_id	image	image_phash	title	label_group
train_3369186413	00136d1cf4ededе0203f32f05f660588.jpg	a6f319f924ad708c	Nescafe \xc3\x89clair Latte 220ml	3648931069

- Images

00136d1cf4ededе0203f32f05f660588.jpg



Number of postings in train set: 34,250
Number of label groups in train set: 11,014
Number of postings in test set: 3 (70,000+ postings unpublished)

Approaches and Model Selection

How to utilize multimodal data?

- Approach 1

Produce text matches from text embedding, image matches from image embedding, then **union** text matches & image matches

- Approach 2

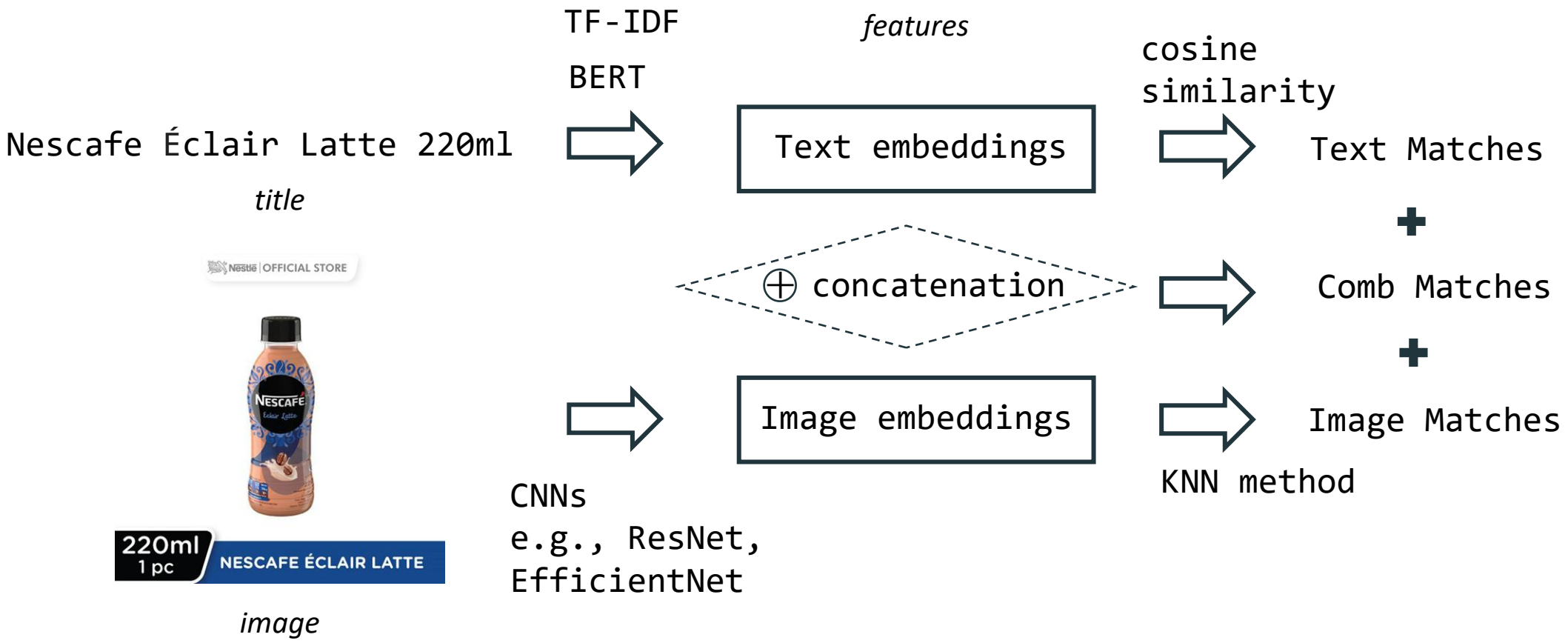
Concatenate text embedding & image embedding, and then produce **comb matches**

- Approach 3

Union comb matches & text matches & image matches

Approaches and Model Selection

Model Selection



Approaches and Model Selection

CLIP & LLaVa

- **CLIP** is a model developed by OpenAI that combines vision and language to understand images and their corresponding textual descriptions. It learns to associate images and text through a *contrastive learning approach*, enabling it to perform a wide range of tasks involving both modalities.
- **LLaVa** is a multi-modal model designed for tasks that require understanding and generating both language and visual content. It aims to create an *intelligent assistant* that can process and respond to inputs in a conversational manner, integrating visual and textual information.

limitation?

Evaluation

Cross-Validation Score (CV Score)

- Cross-validation score is a statistical method used to evaluate the performance of a machine learning model.
- It involves partitioning the dataset into multiple subsets (folds), training the model on a subset, and validating it on another. This process is repeated several times, and the average score (e.g., accuracy, F1 score) is calculated.
- CV score helps to assess model stability, reduce overfitting, and ensure generalization to unseen data.

4-fold example

Limitation and Future Work

- More exploration-based rather than competition-based optimal solution?

- Pre-trained Models
  Allocate too much memory?
  Fune-tuning?

=> Subset dataset...
  EfficientNet with ArcFace

- Error analysis?

=> More preprocessing logic...


Thank you fot
lisening :)