V-Elliot: Speeding up Visual Recommendation via a GPU-powered Data Input Pipeline

Vito Walter Anelli [1], Alejandro Bellogín [2], Antonio Ferrara [1], Daniele Malitesta [1], Felice Antonio Merra [3], Claudio Pomo [1], Francesco Maria Donini [4], Tommaso Di Noia [1]

[1] Politecnico di Bari, [2] Universidad Autónoma de Madrid, [3] Amazon Science, [4] Università degli Studi della Tuscia

Visual Recommendation

- **Visual recommendation** is the task of recommending items when their visual appearance may influence users in their decision making process (e.g., fashion, food, tourism).
- Most of the state-of-the-art algorithms extract visual features from product images through **pretrained Convolutional Neural Networks** (CNNs).
- Existing frameworks for visual recommendation leverage the performance **boosting of GPUs** for model training.
- **None of them** seeks to speed up the data injection process, which usually involves high-dimensional visual embeddings and/or product images themselves.

V-Elliot

- **V-Elliot is** a comprehensive framework to train and evaluate visual recommender systems (VRSs), from data preprocessing to results visualization. **Six state-of-the-art models** (the largest set of VRSs in a complete recommendation pipeline).
- **V-Elliot** leverages GPUs for models **training** and **testing**, and **parallelizes** data preprocessing/feeding and model training/testing.
- **Uses a Producer/consumer** paradigm, i.e., while the current visual features are used for training/testing on the GPU, the next features are preprocessed through the CPU.
- Inputs: both CNN-extracted **visual features**, and product **images** to extract visual features on-the-fly.
- High-dimensional visual item features are **injected only when they are needed**, without uselessly and inefficiently filling in the memory with all items content data in advance.

V-Elliot

Loading, Prefiltering, and Splitting

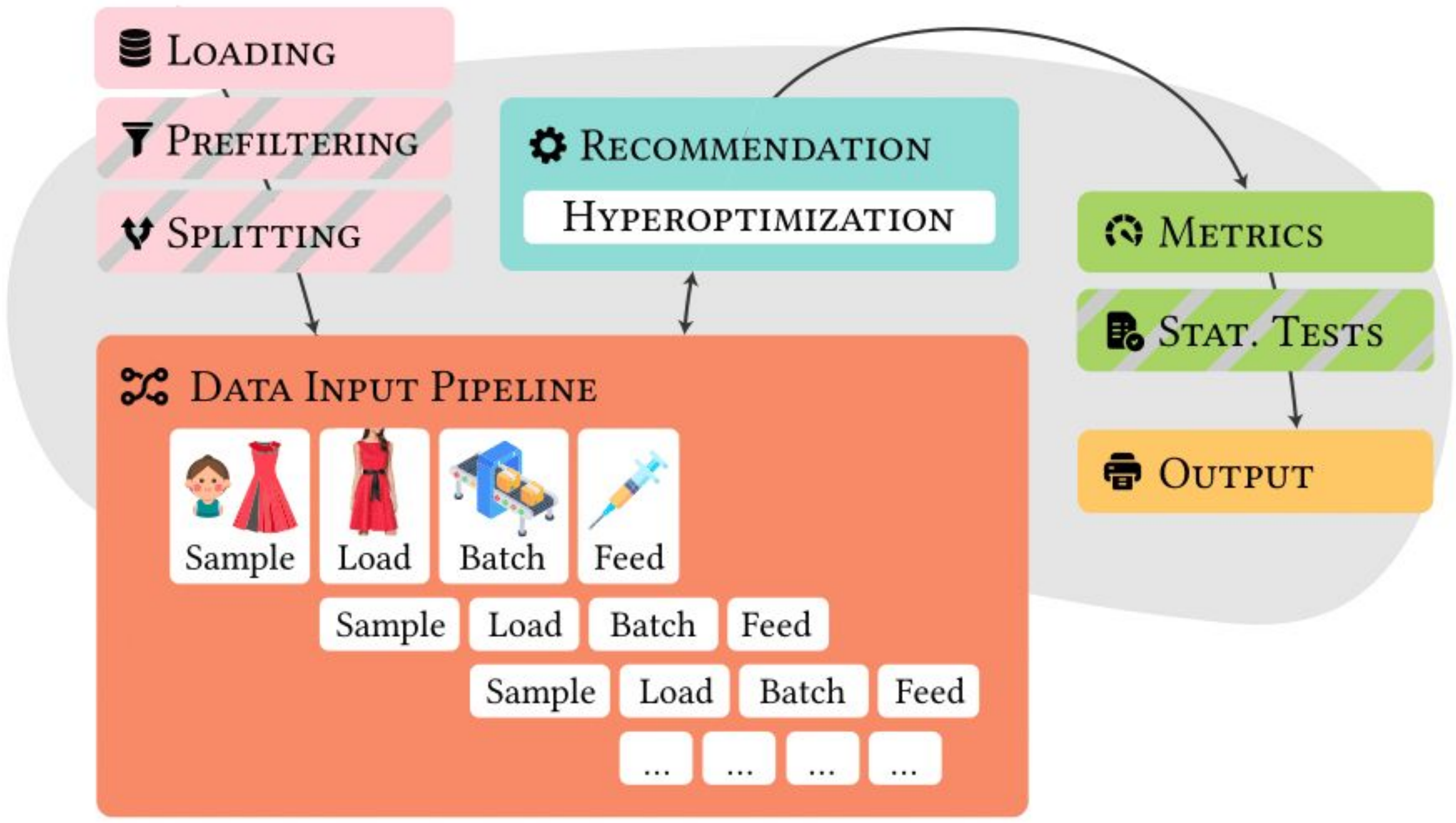
- Item features, semantic information, visual embeddings, and images
- Folder path for visual features and images
- Filter-by-rating and k-core
- Temporal- and random-based splitting

Data Input Pipeline

- Built upon TensorFlow's data input pipeline:
  - the next user-item is sampled
  - visual data is loaded and pre-processed
  - samples optionally grouped into batches
  - batches fed into the algorithm

Recommendation

- Ever-growing implemented models
- Highest number of VRSs
- Extend models with custom and external ones
- Hyperparameter tuning



Metrics and Statistical Tests

- Accuracy/beyond-accuracy metrics
- Evaluation process speeded up by Data Input Pipeline
- Wilcoxon and Paired t-test

Output

- Detailed performance tables
- Model weights
- Recommendation Lists

Try Elliot and reach us out!

github.com/sisinflab/elliot

vitowalter.anelli@poliba.it
alejandro.bellogin@uam.es
antonio.ferrara@poliba.it
daniele.malitesta@poliba.it
felmerra@amazon.de
claudio.pomo@poliba.it
donini@unitus.it
tommaso.dinoia@poliba.it

Tested with:

NVIDIA.

TensorFlow

VRS	Year	Original Code
VBPR	2016	✗
DeepStyle	2017	✗
DVBPR	2017	✓
ACF	2017	✓
VNPR	2018	✗
AMR	2020	✓

Model	Accuracy				Beyond-Accuracy				
	HR	*nDCG*	*Prec*	*MAP*	*EFD*	*EPC*	*Gini*	*SE*	*iCov*
Amazon Baby									
VBPR	.0743	.0160	.0007	.0008	.0075	.0008	.5730	10.0093	1386
DVBPR	.0413	.0082	.0004	.0004	.0039	.0004	.1421	7.7011	370
ACF	**.1221**	**.0384**	**.0012**	**.0022**	**.0169**	**.0018**	<u>.6711</u>	<u>10.1862</u>	<u>1392</u>
DeepStyle	.0561	.0117	.0006	.0005	.0055	.0006	**.7490**	**10.2992**	**1393**
VNPR	.0479	.0112	.0005	.0006	.0052	.0005	.2058	8.5920	965
AMR	<u>.0858</u>	<u>.0192</u>	<u>.0009</u>	<u>.0010</u>	<u>.0092</u>	<u>.0009</u>	.5752	10.0083	1389

bold: best values, <u>underlined</u>: second-best values.