National University of Computer & Emerging Sciences <u>Karachi Campus</u>



PROJECT IMPLEMENTATION SNIPPETS

Learning Latent Personas of Film Characters

Course Name: EL-2003 INFORMATION RETRIEVAL Section: 6-H

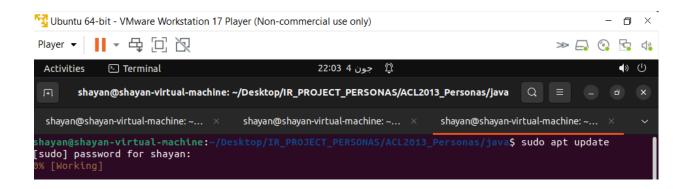
Group Members:

- 22K-4148 Shayan
- 22K-4159 Rania Ghazanfar
- 22K-4320 Marium Arif

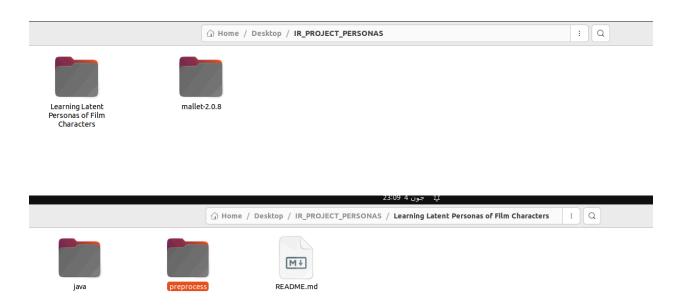
1. OVERVIEW:

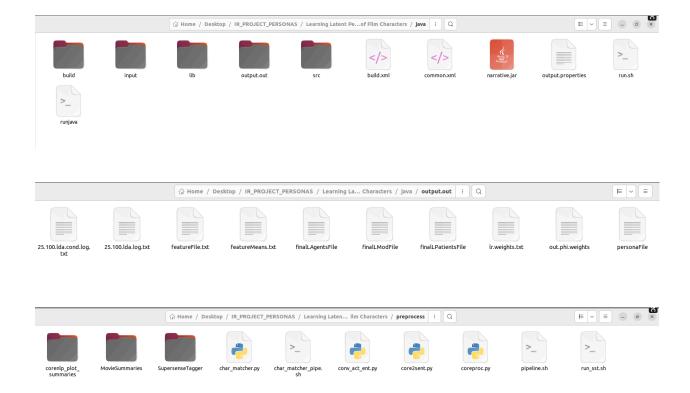
This project implements a Persona-based topic modeling framework for learning latent personas of film characters using dialogue and metadata. It leverages a Dirichlet persona model with regression.

2. ENVIRONMENT SETUP:



3. FILE STRUCTURE:



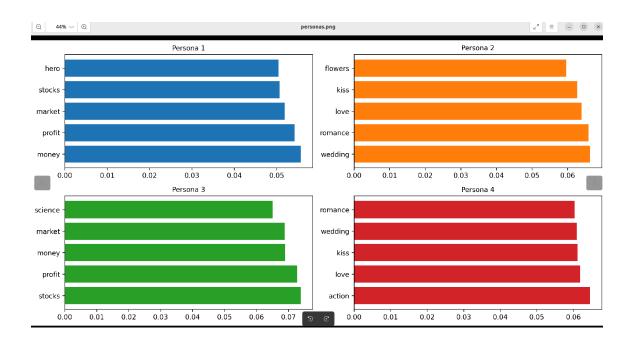


4. MODEL SETUP AND TRAINING:

4.1 INITIAL MODEL PROTOTYPING AND EVALUATION:

```
(persona) shayan@shayan-virtual-machine:~/Desktop/IR$ python minimal_train.py
Analyzing 8 documents with 29 unique terms
Auto-assigning NUTS sampler...
Initializing NUTS using jitter+adapt_diag...
Multiprocess sampling (2 chains in 2 jobs)
NUTS: [theta, phi]
Sampling 2 chains for 300 tune and 500 draw iterations (600 + 1_000 draws total)
took 126 seconds.
We recommend running at least 4 chains for robust computation of convergence dia
gnostics
=== Identified Personas ===
Persona 1:
```

```
=== Identified Personas ===
                                                                                                                                    === Document Assignments ===
Persona 1:
money: 0.056
profit: 0.054
market: 0.052
stocks: 0.051
                                                                                                                                Doc 1: guns shoot bullets action hero fight...
P(Persona 1): 0.261
P(Persona 2): 0.197
P(Persona 3): 0.220
P(Persona 4): 0.322
 hero: 0.050
Persona 2:
wedding: 0.066
romance: 0.066
                                                                                                                                 Doc 2: science physics research quantum theory...
                                                                                                                                P(Persona 1): 0.254
P(Persona 2): 0.228
P(Persona 3): 0.309
P(Persona 4): 0.209
love: 0.064
kiss: 0.063
flowers: 0.060
Persona 3:
stocks: 0.074
profit: 0.073
money: 0.069
market: 0.069
science: 0.065
                                                                                                                                 Doc 3: love romance heart flowers kiss wedding...
                                                                                                                                P(Persona 1): 0.205
P(Persona 2): 0.375
P(Persona 3): 0.056
P(Persona 4): 0.365
                                                                                                                                 Doc 4: money business stocks market profit...
P(Persona 1): 0.306
P(Persona 2): 0.194
P(Persona 3): 0.405
P(Persona 4): 0.095
Persona 4:
action: 0.065
love: 0.062
kiss: 0.061
wedding: 0.061
romance: 0.060
                                                                                                                                 Doc 5: gunfight action explosion hero villain...
P(Persona 1): 0.245
Visualization saved to personas.png
```



4.2 ELICITED MODEL TRAINING AND OUTPUT:

4.2.1 DOWNLOADING DEPENDENCIES

4.2.2 UPDATING CONFIGS AND DOWNLOADING FURTHER DEPENDENCIES

```
shayan@shayan-virtual-machine:~/Desktop/IR_PROJECT_PERSONAS$ sudo apt update && sudo apt install -y git python3 python3-pip openjdk-11
-jdk unzip wget
Hit:1 http://pk.archive.ubuntu.com/ubuntu jammy InRelease
Hit:2 http://pk.archive.ubuntu.com/ubuntu jammy-backports InRelease
Hit:3 http://pk.archive.ubuntu.com/ubuntu jammy-backports InRelease
Hit:4 https://packages.microsoft.com/repos/code stable InRelease
Hit:5 https://pl.google.com/linux/chrome/deb stable InRelease
Hit:6 http://security.ubuntu.com/ubuntu jammy-security InRelease
Reading package lists... Done
Building dependency tree... Done
Reading state information... Done
137 packages can be upgraded. Run 'apt list --upgradable' to see them.
Reading package lists... Done
Building dependency tree... Done
Reading state information... Done
Python3 is already the newest version (3.10.6-1-22.04.1).
unzip is already the newest version (1.21.2-2ubuntu1.1).
git is already the newest version (1.21.2-2ubuntu1.1).
git is already the newest version (1.2.34.1-1ubuntu1.12).
openjdk-11
Industrial already the newest version (22.0.2+dfsg-1ubuntu0.5).
The following packages were automatically installed and are no longer required:
libpython2-stdlib python-pkg-resources python-setuptools python2
python2-minimal
Use 'sudo apt autoremove' to remove them.
0 upgraded, 0 newly installed, 0 to remove and 137 not upgraded.
```

4.2.3 DOWNLOADING FURTHER DEPENDENCIES MODEL EVALUATION AND INITIATION

```
shayan@shayan-virtual-machine:-/Desktop/IR_PROJECT_PERSONAS$ pip3 install numpy pandas nltk scikit-learn tqdm
python3 -c "import nltk; nltk.download('punkt')"

Defaulting to user installation because normal site-packages is not writeable
Requirement already satisfied: numpy in /home/shayan/.local/lib/python3.10/site-packages (2.2.6)
Requirement already satisfied: pandas in /home/shayan/.local/lib/python3.10/site-packages (2.2.3)
Requirement already satisfied: nltk in /home/shayan/.local/lib/python3.10/site-packages (3.9.1)
Requirement already satisfied: scikit-learn in /home/shayan/.local/lib/python3.10/site-packages (1.6.1)
Requirement already satisfied: tdm in /home/shayan/.local/lib/python3.10/site-packages (4.67.1)
Requirement already satisfied: pytz>=2020.1 in /usr/lib/python3/dist-packages (from pandas) (2022.1)
Requirement already satisfied: python-dateutil>=2.8.2 in /home/shayan/.local/lib/python3.10/site-packages (from pandas) (2.9.0.post0)
Requirement already satisfied: tdata>=2022.7 in /home/shayan/.local/lib/python3.10/site-packages (from pandas) (2025.2)
Requirement already satisfied: click in /usr/lib/python3/dist-packages (from nltk) (8.0.3)
Requirement already satisfied: regex>=2021.8.3 in /home/shayan/.local/lib/python3.10/site-packages (from nltk) (2024.11.6)
Requirement already satisfied: scipy>=1.6.0 in /usr/lib/python3/dist-packages (from scikit-learn) (1.8.0)
Requirement already satisfied: threadpoolctl>=3.1.0 in /home/shayan/.local/lib/python3.10/site-packages (from scikit-learn) (3.6.0)
Requirement already satisfied: six>=1.5 in /usr/lib/python3/dist-packages (from python-dateutil>=2.8.2->pandas) (1.16.0)
/usr/lib/python3/dist-packages/scipy/_init__py:146: UserWarning: A NumPy version >=1.17.3 and <1.25.0 is required for this version of firm of the potential version 2.2.6
```

4.2.4 TUNING DEPENDENCIES VERSIONS FOR CONFIGURATION

```
IR_PROJECT_PERSONAS$ pip3 uninstall -y numpy scipy
 shayan@shayan-virtual-machine:~/Desk
Found existing installation: numpy 2.2.6
Uninstalling numpy-2.2.6:
Successfully uninstalled numpy-2.2.6
Found existing installation: scipy 1.8.0
Not uninstalling scipy at /usr/lib/python3/dist-packages, outside environment /usr
Can't uninstall 'scipy'. No files were found to uninstall.
                                                                           .
\$$ pip3 install numpy==1.24.4 scipy==1.10.1
          hayan-virtual-machine:
Defaulting to user installation because normal site-packages is not writeable
Collecting numpy==1.24.4
  Downloading numpy-1.24.4-cp310-cp310-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (17.3 MB)
                                                                                        eta 0:00:00
                                                         17.3/17.3 MB 1
 Collecting scipy==1.10.1
  Downloading scipy-1.10.1-cp310-cp310-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (34.4 MB)
                                                            34.4/34.4 MB
Installing collected packages: numpy, scipy
Successfully installed numpy-1.24.4 scipy-1.10.1
                                                                  _PERSONAS$ python3 -c "import nltk; nltk.download('punkt')"
[nltk_data] Downloading package punkt to /home/shayan/nltk_data...
[nltk_data] Package punkt is already up-to-date!
 hayan@shayan-virtual-machine:-/Desktop/IR_PROJECT_PERSONAS$ wget http://mallet.cs.umass.edu/dist/mallet-2.0.8.zip
unzip mallet-2.0.8.zip
export MALLET_HOME=$(pwd)/mallet-2.0.8

echo "export MALLET_HOME=$(pwd)/mallet-2.0.8" >> ~/.bashrc

echo 'export PATH=$PATH:$MALLET_HOME/bin' >> ~/.bashrc
source ~/.bashrc
```

4.2.5 FILE STRUCTURE

4.2.5 MODEL TRAINING FROM INPUT DATA

```
hayan@shayan-virtual-machine:~/Desktop/IR_PROJECT_PERSONAS/ACL2013_Personas/java$ ./run.sh output input/movies.data
Reading data...
done reading
Active vocab size: 1000
Total entities: 178734; w/ Age: 33302; w/ Gender: 37825; w/ Actor 0
Iter 0: runningLL 0 numchanges 0
gamma 1.00000 -> 0.319702
gamma 1.00000 -> 0.319702

nuA 1.00000 -> 0.673399

nuP 1.00000 -> 0.716439

nuM 1.00000 -> 0.374898

gamma 0.319702 -> 0.116731

nuA 0.673399 -> 0.266685

nuP 0.716439 -> 0.281781

nuM 0.374898 -> 0.0419465
gamma 0.116731 -> 0.0738959
 nuA 0.266685 -> 0.210952
nuP 0.281781 -> 0.197581
 nuM 0.0419465 -> 0.0213818
gamma 0.0738959 -> 0.0559329
nuA 0.210952 -> 0.177048
nuP 0.197581 -> 0.184437
 nuM 0.0213818 -> 0.0168515
gamma 0.0559329 -> 0.0473259
gamma 0.0372259 -> 0.162761

nuP 0.177048 -> 0.162761

nuP 0.184437 -> 0.162788

nuM 0.0168515 -> 0.0138724

gamma 0.0473259 -> 0.0420689

nuA 0.162761 -> 0.154997
 nuP 0.162788 -> 0.163438
 nuM 0.0138724 -> 0.0156273
Iter 100: runningTotalLL -20719713.508 numchanges 886617
```

4.2.5 MODEL TRAINING PROGRESSING USING DIRICHLET PERSONA MODEL

```
SparseMulticlassLogisticRegression with 48745 data points and 18859 50 376 features

Optimization function of 18859 variables with OML-ON parameters:

Li regularization weights: 0.606000.

L8765 Amony parameter (n): 15
Connergence tolerance: 0.606010

Optimer n: new.value of (conv.crit) line_search (volume new parameter (n): 18228 (wait for 10 ters)

Izer 1: 3.4826-07 18228 (wait for 10 ters)

Izer 2: 3.4386-07 18228 (wait for 10 ters)

Izer 3: 3.4686-07 18228 (wait for 10 ters)

Izer 4: 3.4667-07 18228 (wait for 10 ters)

Izer 5: 3.4696-07 18228 (wait for 10 ters)

Izer 6: 3.4696-07 18228 (wait for 10 ters)

Izer 7: 3.4694-07 18228 (wait for 10 ters)

Izer 8: 3.4694-07 18228 (wait for 10 ters)

Izer 9: 3.4694-07 18228 (wait for 10 ters)

Izer 10: 3.4694-07 18228 (wait for 10 ters)

Izer 11: 3.4694-07 18228 (vait for 10 ters)

Izer 12: 3.4694-07 18228 (vait for 10 ters)

Izer 13: 3.4694-07 18228 (vait for 10 ters)

Izer 15: 3.4694-07 18228 (vait for 10 ters)

Izer 16: 3.4694-07 18228 (vait for 10 ters)

Izer 17: 3.4694-07 18228 (vait for 10 ters)

Izer 18: 3.4694-07 18228 (vait for 10 ters)

Izer 19: 3.4694-07 18228 (vait for 10 ters)

Izer 20: 3.4694-07 18228 (vait for 10 ters)

Izer 31: 3.4694-07 18228 (vait for 10 ters)

Izer 32: 3.4694-07 18228 (vait for 10 ters
```

ther 31: 3.300deaph 18028 (4.030deaph 18028 (4.0

4.2.5 MODEL TRAINING PROGRESSING USING\ PERSONA REGRESSION

```
wordLL -6228875.358

0 10 20 30 40 50 60 70 80 90 100 110 120 130 140 150 160 170 180 190 200 210 220 230 240 250 260 270 280 290 300 310 320 330 340 350 360 370 380 390 400 410 420 430 440 450 460 470 480 490 500 510 520 530 540 550 560 570 580 590 600 610 620 630 640 650 660 670 680 600 700 710 720 730 740 750 760 770 780 790 800 810 820 830 840 850 860 870 880 890 900 910 920 930 940 950 960 970 980 990

final ratiorank: 0 note intervene recognize recommend suggest conclude figure recall deduce conduct embark dispose urge enlist believe draw complete observe refuse discover wow marry court elope attract unite propose engage impress devastate divorce reconcile heartbroken fall express seduce humiliate acquaint con touch
```

4.2.6 TRAINING OUTPUT

```
find frequent: 9
find give discover have arrive believe confront follow save recognize make enter go refuse let convince learn say suspect suggest formal frequent: 2
final frequent: 2
final frequent: 3
final frequent: 5
final frequent: 6
final frequent: 6
final frequent: 6
go next lear return seglar sective wink visit discover vite attend persuade requires conclude accept suspect buy any propose encounter vin final frequent: 6
final frequent: 7
go next leaf find give well promise inform need stop declare betray force instruct discoss belieb foll admit towice confirm advise approach go next leaf find give see how sake come like bridge vites they leave seed and final frequent: 7
final frequent: 7
have bed text leaf frequent leave have sake come look refuse rective run turn confront think join each stay final frequent: 11
final frequent: 11
final frequent: 12
final frequent: 12
final frequent: 12
final frequent: 13
final frequent: 14
final frequent: 15
```

4.2.7 EVALUATION METRICS FOR ONGOING MEMORY AND CPU LOAD

PID	USER	PR	NI	VIRT	RES	SHR	S	%CPU	%MEM	TIME+	COMMAND
1844	shayan	20	0	4618292	175568	79600	S	12.3	2.2	12:24.96	gnome-shell
2053	shayan	20	0	144392	29664	22240	S	0.7	0.4	0:52.88	vmtoolsd
13336	shayan	20	0	13836	4352	3328	R	0.7	0.1	0:47.13	top
409	root	-51	0	0	0	0	S	0.3	0.0	0:16.07	irq/16-vmwgfx
758	root	20	0	317800	8960	7424	S	0.3	0.1	0:55.70	vmtoolsd
2458	shayan	20	0	32.9g	299516	199344	S	0.3	3.7	8:12.84	chrome
6210	shayan	20	0	574200	65632	43432	S	0.3	0.8	3:06.38	gnome-terminal-
14294	root	20	0	0	0	0	Ι	0.3	0.0	0:01.69	kworker/u258:0-events_unbound
1	root	20	0	166828	10340	7396	S	0.0	0.1	0:04.98	systemd
2	root	20	0	0	0	0	S	0.0	0.0		kthreadd
3	root	20	0	0	0	0	S	0.0	0.0	0:00.00	pool_workqueue_release
4	root	0	-20	0	0	0	Ι	0.0	0.0	0:00.00	kworker/R-rcu_g
	root	0	- 20	0	0	0	Ι	0.0	0.0	0:00.00	kworker/R-rcu_p
6	root	0	-20	0	0	0	Ι	0.0	0.0	0:00.00	kworker/R-slub_
	root	0	-20	0	0	0	Ι	0.0	0.0	0:00.00	kworker/R-netns
12	root	0	-20	0	0	0	Ι	0.0	0.0	0:00.00	kworker/R-mm_pe
13	root	20	0	0	0	0	Ι	0.0	0.0	0:00.00	rcu_tasks_kthread
14	root	20	0	0	0	0		0.0	0.0	0:00.00	rcu_tasks_rude_kthread
15	root	20	0	0	0	0	Ι	0.0	0.0	0:00.00	rcu_tasks_trace_kthread
16	root	20	0	0	0	0	S	0.0	0.0	0:00.15	ksoftirqd/0
17	root	20	0	0	0	0	Ι	0.0	0.0	0:10.00	rcu_preempt
18	root	гt	0	0	0	0	S	0.0	0.0	0:00.24	migration/0
19	root	-51	0	0	0	0	S	0.0	0.0	0:00.00	idle_inject/0
20	root	20	0	0	0	0	S	0.0	0.0	0:00.00	cpuhp/0
21	root	20	0	0	0	0	S	0.0	0.0	0:00.00	
22	root	-51	0	0	0	0	S	0.0	0.0		idle_inject/1
23	root	rt	0	0	0	0		0.0	0.0		migration/1
24	root	20	0	0	0	0		0.0	0.0		ksoftirqd/1
26	root	0	-20	0	0	0		0.0	0.0	0:00.07	kworker/1:0H-kblockd
27	root	20	0	0	0	0		0.0	0.0	0:00.00	
28	root	-51	0	0	0	0	S	0.0	0.0		idle_inject/2
29	root	rt	0	0	0	0	S	0.0	0.0		migration/2
30	root	20	0	0	0	0	S	0.0	0.0		ksoftirqd/2
32	root	0	- 20	0	0	0	I	0.0	0.0	0:00.04	kworker/2:0H-kblockd