

A black clapperboard with white text and a red background. The text includes 'PROD.NO.', 'SCENE', 'TAKE', and 'SOU'.

SCRIPT BASED MOVIE RECOMMENDATION SYSTEM

FLETCHER NLP PROJECT



The Internet Movie Script Database (IMDb)

IMDb

The web's largest
movie script resource!

INT. NORTHUP HOUSE - NIGHT

Now alone together, we see Anne and Solomon wrapped in each other's arms. Beyond being physically close, emotionally close, they are just so very comfortable with one another. They are the very representation of a couple who are made for each other.

They look at each other for a prolonged time.

SOLOMON

(COMICALLY FORLORN)

Three weeks. Two days.

ANNE

It is the custom. I wonder what you'll do without me?

LORI: How about before we met?

BOONE: Even then.

They kiss again. We move past them into the darkness outside the window.

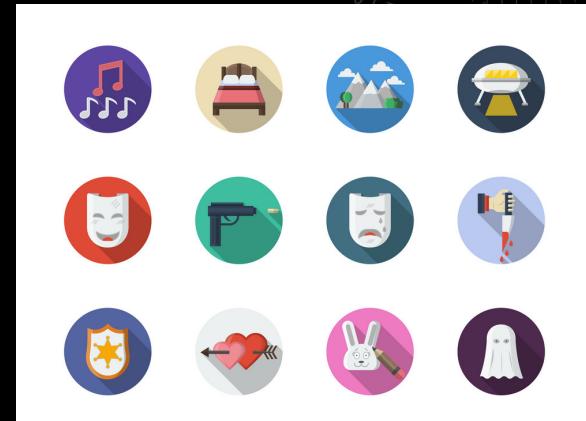
Dissolve to the moon, clearing cloud

SCENES 3-15 DELETED

16. EXT. RICKMAN HOUSE NIGHT

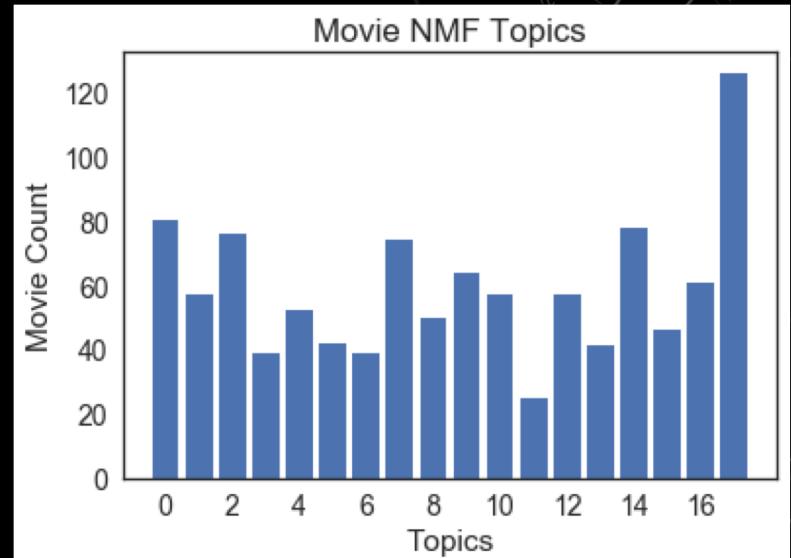


Narration



MODEL

- Using whole script (setting/shot specific content removed) as document
 - Tokenize + Lemmatize(on selected tags
 - TFIDF (min_df=0.1, max_df=0.9)
 - NMF topics (n=18)

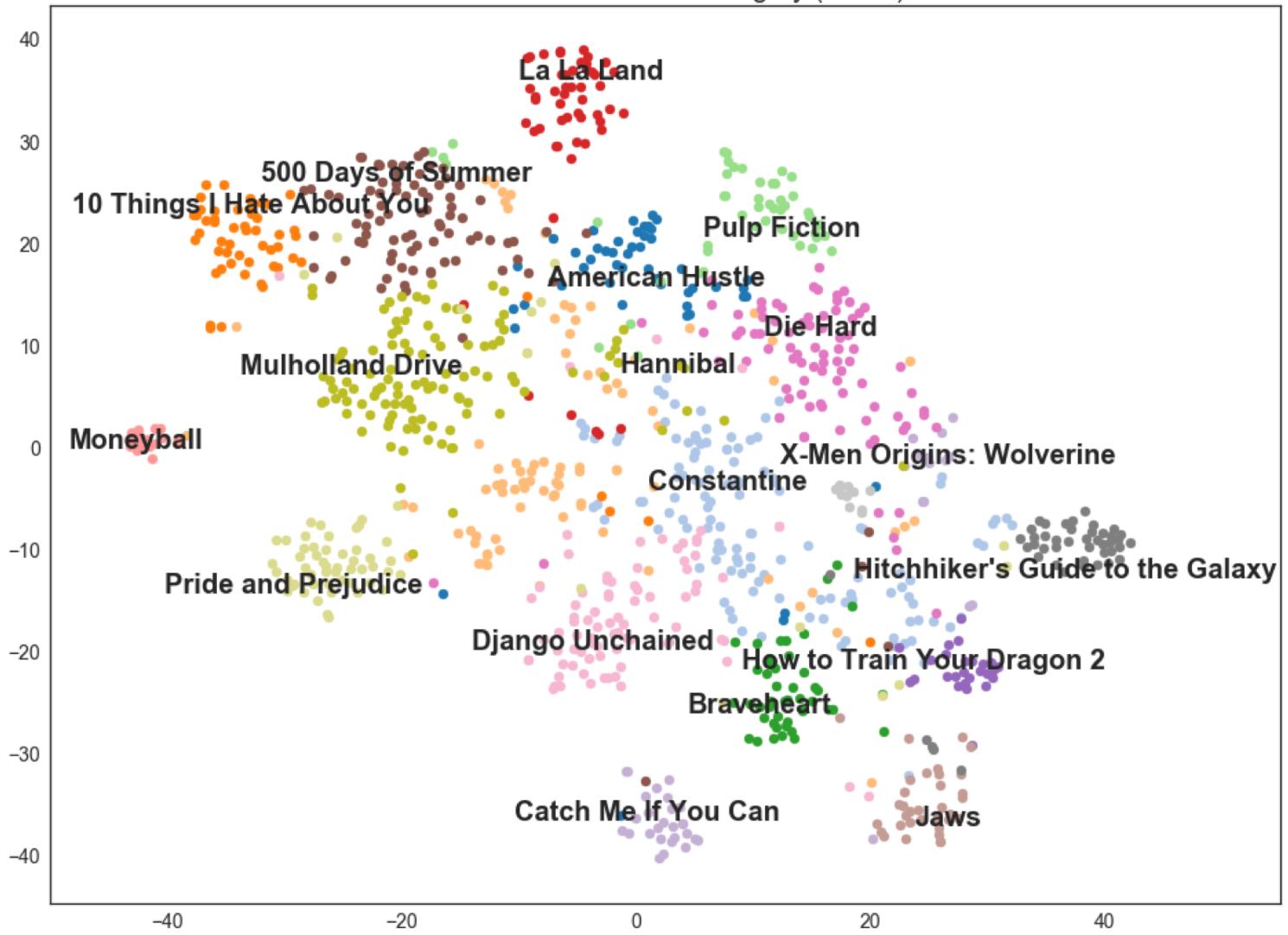


Topic #	Topic Theme	Words
1	WAR	soldier, general, sergeant, army, war, lieutenant, rifle, major, officer, prisoner
4	SPACE	ship, data, planet, crew, computer, control, space, console, pilot, deck
5	CURSES	█████, shit, money, alright, ██████, jesus, ██████████████████, asshole, gun, brother
6	FANTASY	sword, dragon, warrior, battle, merry, arrow, forest, tree, blood, armor
7	ROMANCE	baby, party, wedding, date, sex, marry, wow, dog, happy, thanks
8	SECRETE SERVICE	agent, fbi, computer, monitor, security, federal, tank, surveillance, tech, pilot
12	ACTING	stage, audience, movie, actor, theater, script, music, crowd, director, studio
13	OCEAN	boat, deck, shark, sea, island, dock, beach, swim, plane, ship

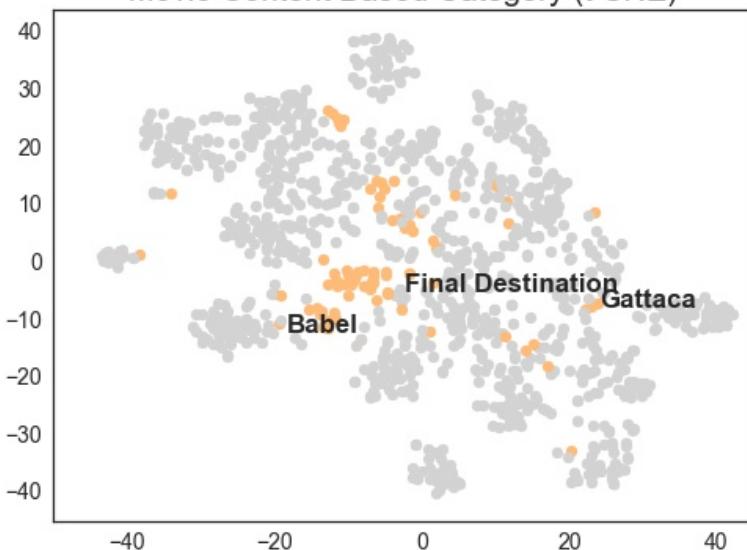
CLUSTERING

- A movie can have more than one topics
- Use KMeans($k=18$) to classify the movies
- To visualize clusters of movies (t-SNE + Tableau)

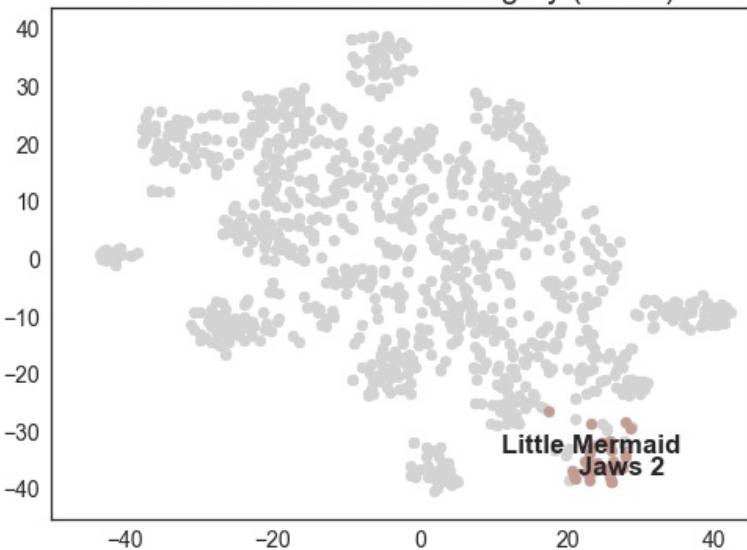
Movie Content Based Category (t-SNE)



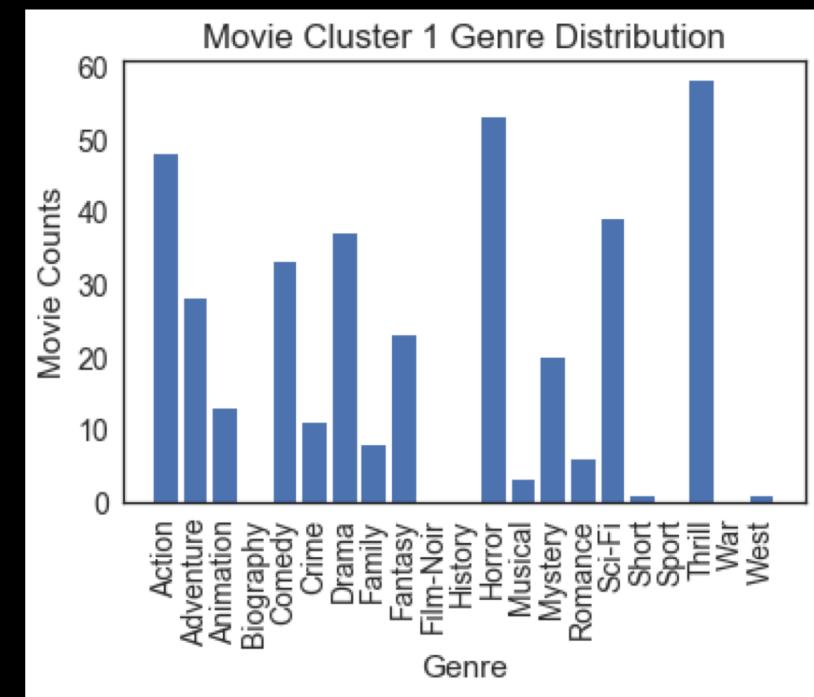
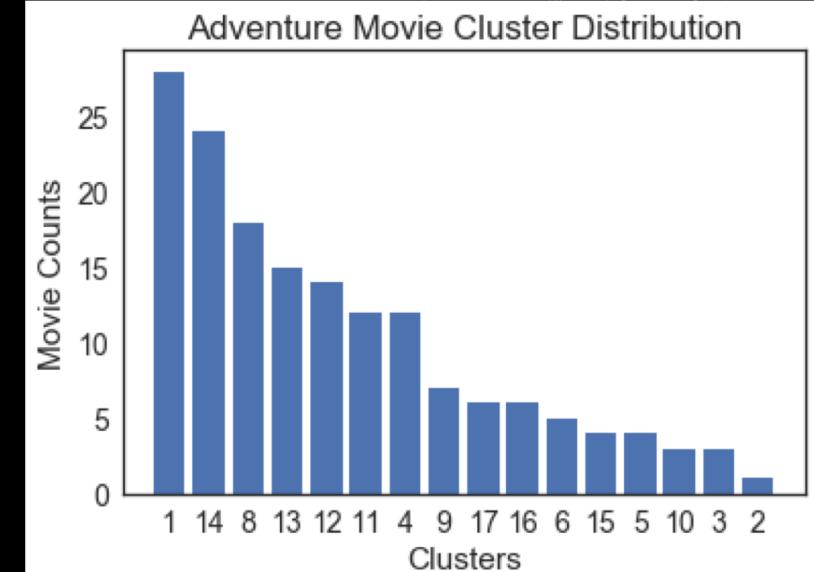
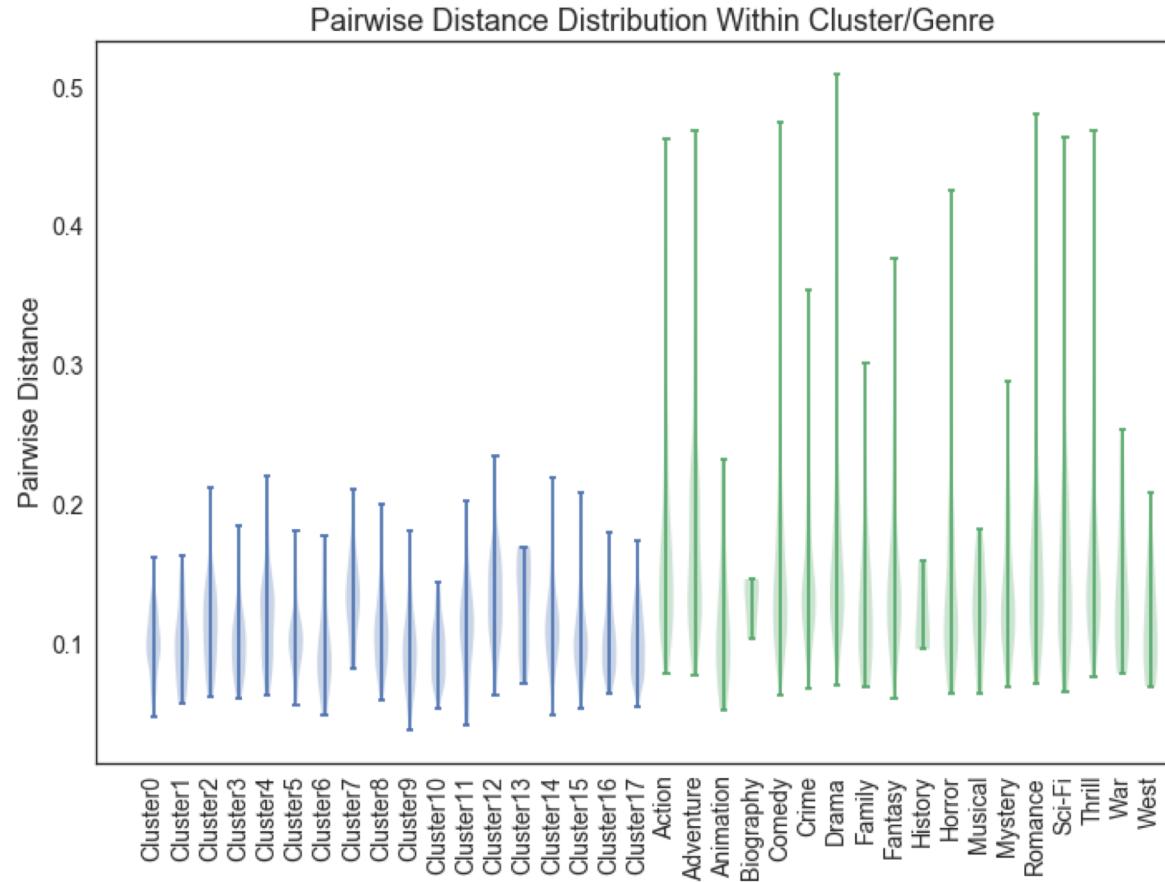
Movie Content Based Category (t-SNE)



Movie Content Based Category (t-SNE)



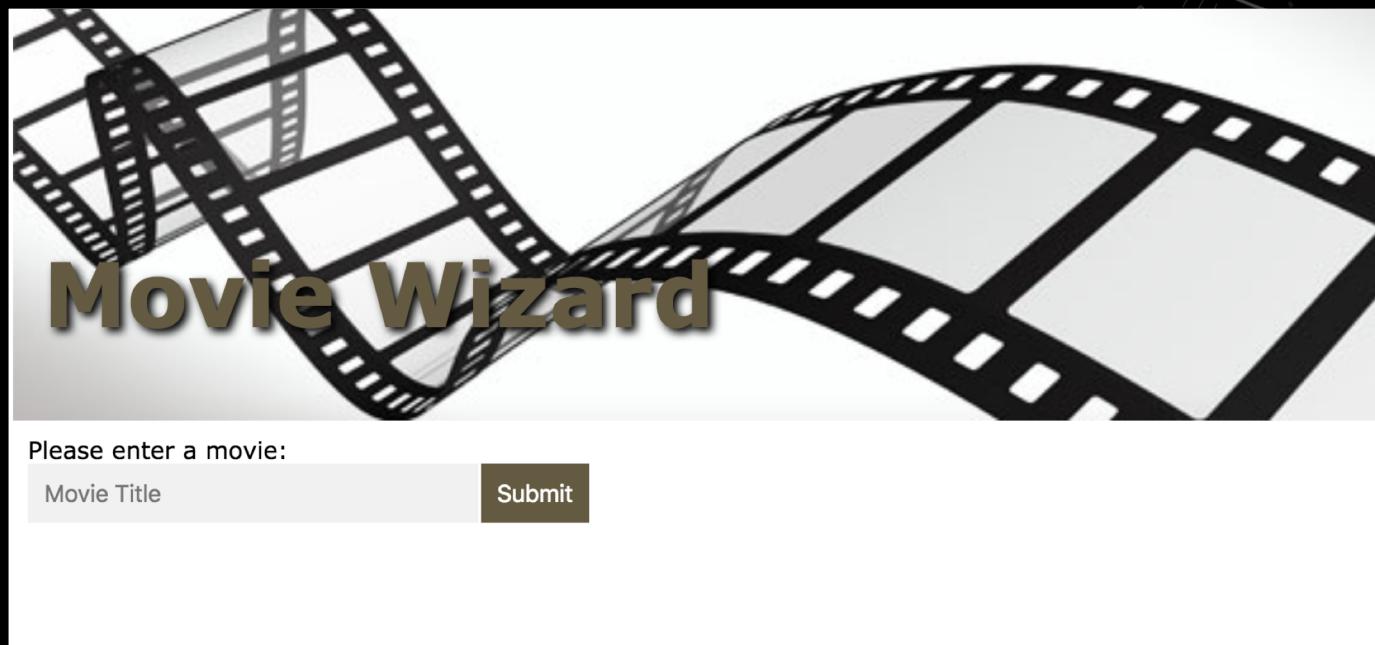
TOPIC ≠ GENRE



RECOMMENDATION SYSTEM

- Euclidean Distance using all 18 NMF topics
- Add genre information and tweak the weight
- A pure content based system + genre info
 - Recommend similar movie by movie title
 - Recommend similar movie by text and genre options

Demo: <http://127.0.0.1:5000>



LIMITATION OF THIS SYSTEM

- Limited numbers of movies
 - only ~1000. Use movie plot summary from wiki to increase the database size
- Movie popularity is not incorporated in the system
 - Incorporate movie reviews or ratings to recommend more popular movies

THANK YOU FOR YOUR ATTENTION



APPENDIX - TOPIC KEY WORDS

Topic #	Topic Theme	Words
0	horror	blood, tree, corridor, human, gaze, spin, skin, wind, flame, edge
1	war	soldier, general, sergeant, army, war, lieutenant, rifle, major, officer, prisoner
2	fight	gun, bullet, weapon, security, pistol, police, thug, building, elevator, alley
3		officer, police, tv, radio, patient, hospital, doctor, reporter, father, monitor
4	space	ship, data, planet, crew, computer, control, space, console, pilot, deck
5	curse	----, shit, money, alright, ----, jesus, -----, asshole, gun, brother
6	fantasy	sword, dragon, warrior, battle, merry, arrow, forest, tree, blood, armor
7	romance	baby, wedding, party, date, sex, marry, wow, thanks, happy, dog
8	secret service	agent, fbi, computer, monitor, security, tank, federal, surveillance, tech, pilot
9	family	mother, father, sister, doctor, family, brother, aunt, daughter, son, wife

APPENDIX - TOPIC KEY WORDS(CON'T)

Topic #	Topic Theme	Words
10	law	lawyer, office, desk, attorney, court, jury, client, file, case, money
11	politics	president, plane, general, pilot, missile, secretary, aide, professor, war, tv
12	acting	stage, audience, movie, actor, theater, script, music, crowd, director, studio
13	ocean	boat, deck, shark, sea, island, dock, beach, swim, plane, ship
14	road	truck, road, trailer, baby, town, deputy, highway, shotgun, goin, hell
15	sports	coach, ball, game, team, field, baseball, stadium, crowd, football, win
16	teenagers	student, school, class, professor, book, teacher, hall, parent, classroom, desk
17	???	pause, apartment, desk, hotel, clerk, elevator, money, office, doctor, bedroom

PLOT MOVIE CLUSTER ON PCA WITH 2 COMPONENTS

