

Movie Genre identification based on plot summary

GROUP 10

Chandan Basavaraju

Charan Shampur
Sharan Kadagad

November 29, 2015

Overview

- 1 Why? & How?
- 2 Dataset
- 3 Feature extraction
- 4 Classification
- 5 Results

Why? & How?

- Automation
- Mainly based on the description

- 42,306 movie plot summaries
- Aligned metadata
 - Movie box office revenue, genre, release date, runtime, and language
 - Character names and aligned information about the actors who portray them

Movie metadata

- Wikipedia movie ID - 54166
- Freebase movie ID - /m/0f4yh
- Movie name - Indiana Jones and the Raiders of the Lost Ark
- Movie release date - 1981-06-12
- Movie box office revenue - 389,925,971
- Movie runtime - 115.0
- Movie languages - Arabic Language, Nepali Language, Spanish Language, Hebrew Language, English Language, German Language
- Movie countries - United States of America
- Movie genres - Adventure

Character metadata

- Wikipedia movie ID - 54166
- Freebase movie ID - /m/0f4yh
- Character Name - Indiana Jones
- Actor DOB - 1942-07-13
- Actor gender - M
- Actor height - 1.85
- Actor ethnicity - /m/01qhm'
- Actor Name - Harrison Ford
- Actor age at movie release - 38
- Freebase character map - /m/0k294p

Plot summary

- Shlykov, a hard-working taxi driver and Lyosha, a saxophonist, develop a bizarre love-hate relationship, and despite their prejudices, realize they aren't so different after all.
- Dan Hardesty is an escaped murderer, sentenced to hang. In Hong Kong, he meets Joan Ames, a terminally-ill woman, in a bar. They share a drink, then Dan breaks his glass, followed by Joan. Police Sergeant Steve Burke captures Dan when he leaves and escorts his prisoner aboard an ocean liner crossing the Pacific to San Francisco. On board, Dan jumps into the water in a bid to escape, dragging a handcuffed Steve with him, but spots Joan among the passengers and changes his mind. Once the ship is underway, he persuades Steve to remove his handcuffs. Dan and Joan fall in love on the month-long cruise, neither knowing that the other is under the shadow of death. By chance, two of Dan's friends are also aboard, thief Skippy and con artist "Barrel House Betty", masquerading as "Countess Barilhaus". The countess distracts Steve as much as she can to help Dan. Just before the only stop, at Honolulu, Steve has Dan put in the brig, but he escapes with their help and goes ashore. Joan intercepts him and they spend an idyllic day together. When they drive back to the dock, Dan starts to tell her why he cannot return to the ship, only to have her faint. Dan carries her aboard for medical help, forfeiting his chance. Later, Joan's doctor tells Dan about her condition and that the slightest excitement or shock could be fatal. Meanwhile, the "countess" has spent so much time with the policeman that a romance blooms between them. When they near the end of the voyage, he awkwardly proposes to her. She tells him her true identity, but he still wants to marry her. As Steve and Dan get ready to disembark, a steward overhears the grim truth and, when Joan comes looking for Dan, tells her. The two lovers part for the last time without letting on they know each other's secret, and Joan collapses after Dan is out of sight. They had agreed to meet again on New Year's Eve, a month later. At the appointed time and place, a bartender is startled when two glasses on the bar break with no one around.

Plot Summary

- Tokenisation
- Removal of stop words and stemming
- POS tagging
- Term frequency

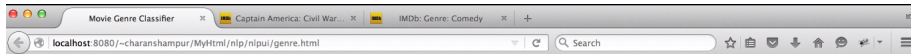
Actors

- Weights for each actor on each genre is based on the number of movies he/she's acted in that particular genre
- Weights for a movie is the sum of weights of each genre of all the actors present in the movie

Multi-class SVM

- Action
- Thriller
- Comedy
- Drama
- Horror
- Fiction

https://www.cs.cornell.edu/people/tj/svm_light/svm_multiclass.html



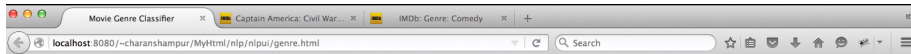
Movie Genre Identifier

Enter Movie Summary here:

Political interference in the Avenger's activities causes a rift between former allies Captain America and Iron Man.

Click to get Genre

Loading Genre....



Movie Genre Identifier

Enter Movie Summary here:

Political interference in the Avenger's activities causes a rift between former allies Captain America and Iron Man.

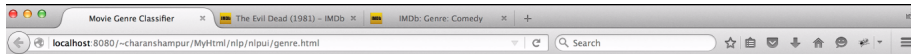
Click to get Genre

Action, Thriller

Similar Movie

Genre

The Hunger Games	Action
Rise of the Planet of the Apes	Action
Total Recall	Action
The Karate Kid	Action



Movie Genre Identifier

Enter Movie Summary here:

After young Riley is uprooted from her Midwest life and moved to San Francisco, her emotions - Joy, Fear, Anger, Disgust and Sadness - conflict on how best to navigate a new city, house, and school.

Click to get Genre

Sci-Fi, Drama

Similar Movie

Genre

Planet of the Apes

Sci-Fi

Terminator 2: Judgment Day

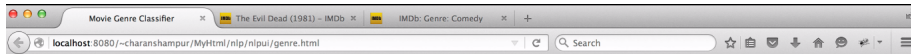
Sci-Fi

Minority Report

Sci-Fi

Guardians of the Galaxy

Sci-Fi



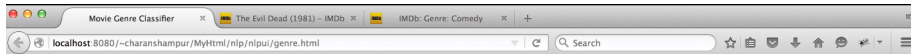
Movie Genre Identifier

Enter Movie Summary here:

After young Riley is uprooted from her Midwest life and moved to San Francisco, her emotions - Joy, Fear, Anger, Disgust and Sadness - conflict on how best to navigate a new city, house, and school. Amy Poehler, Bill Hader, Lewis Black

Click to get Genre

Loading Genre....



Movie Genre Identifier

Enter Movie Summary here:

After young Riley is uprooted from her Midwest life and moved to San Francisco, her emotions - Joy, Fear, Anger, Disgust and Sadness - conflict on how best to navigate a new city, house, and school. Amy Poehler, Bill Hader, Lewis Black

Click to get Genre

Comedy, Drama

Similar Movie

Genre

Bringing Down the House

Comedy

Role Models

Comedy

30 Minutes or Less

Comedy

White Chicks

Comedy

Results

Available data was randomly split into training data and testing data. Evaluation is based on Accuracy, Precision and Recall.

90 % Training & 10% Testing

Class	Accuracy	Precision	Recall
1	0.439215686275	0.506787330317	0.387543252595
2	0.609022556391	0.625965996909	0.592972181552
3	0.599597585513	0.657352941176	0.551171393342
4	0.728922716628	0.686328555678	0.777153558052
5	0.715365239295	0.672985781991	0.763440860215
6	0.58836689038	0.589686098655	0.587053571429

For training on 90% and test on 10%, the correct estimation is 65.04%
loss is 34.96%

80 % Training & 20% Testing

Class	Accuracy	Precision	Recall
1	0.437109723461	0.495951417004	0.390749601276
2	0.601984564498	0.603537214444	0.600439882698
3	0.597818677573	0.65059347181	0.552963430013
4	0.721047930932	0.688068181818	0.757348342714
5	0.733581164808	0.699763593381	0.770833333333
6	0.590422535211	0.584821428571	0.596131968146

For training on 80% and test on 20%, the correct estimation is 64.48%
loss is 35.52%

Confusion Matrix

Table: Confusion Matrix

C.F	Action	Thriller	Comedy	Drama	Horror	Fiction
Action	6	0	0	3	1	0
Thriller	0	6	0	3	0	1
Comedy	0	1	8	1	0	0
Drama	1	2	0	8	0	0
Horror	0	2	0	1	5	2
Fiction	0	2	0	1	0	7

- CMU movie summary corpus:
<http://www.ark.cs.cmu.edu/personas/>
- Multiclass SVM Classifier: https://www.cs.cornell.edu/people/tj/svm_light/svm_multiclass.html
- Python Documentation
- nltk POS tagger: <http://www.nltk.org/book/ch05.html>

Questions?