Constructing character profiles and character networks in Gladiator using an unsupervised approach.

Term Project, SNLP 2015, IIT Kharagpur

Mid-Term Report, Group No. 3

Arkanath Pathak, Akshay Gupta, Buddha Prakash, Gaurav Sharma, Sanyam Agarwal, Utpal

Introduction

Although there have been approaches to analyze the characters and their relationships in other contexts like novels, there have been almost no successful attempt to analyze the characters in a movie based on the transcript. In an unsupervised manner, our project is to construct and study the character profiles, relationship structure, sentiment dynamics and other interesting aspects of the academy award winning movie Gladiator (2000) using its transcript and subtitles.

We chose Gladiator as our use case because of a few major reasons as described in the following paragraphs. The movie is a historical drama film staring Russell Crowe in the lead role. Crowe portrays the fictional character, loyal Roman general Maximus Decimus Meridius, who is betrayed when Commodus, the ambitious son of Emperor Marcus Aurelius, murders his father and seizes the throne. Reduced to slavery, Maximus rises through the ranks of the gladiatorial arena to avenge the murders of his family and his emperor.

The movie's plot itself is quite complex. Based in the era of the Roman empire, the movie involves a non-modern english language full of circuitous dialogues which makes our task more intriguing. It includes a number of diverse characters who share complex relationship with each other. Gladiator also allows us to study temporal changes in the relationships between characters which we will explain more later in this report.

In this report we present some basic processing and try out some of the proposed approaches to the data available to us. We analyze the complexity of the problems involved. We raise the major challenges to be solved to make the result convincing and non-trivial.

Objectives

The broad objectives of our work are:

1. **Character modelling**: Extract the names of all characters in the movies along with their character attributes.
2. **Relationship modelling**: Obtain the list of main characters and relationships critical to the movie.
3. **Relationship build-up analysis**: Model the temporal variation of relationships between characters.
4. **Flow of emotions**: Examine the overall flow of emotions throughout the movie.

Datasets Used

Since we plan to achieve our goal in a completely unsupervised manner w.r.t. the movie, we perform our preliminary analysis on a single movie script of the movie “Gladiator”. We obtained the transcript from the The Internet Movie Script Database(IMSDb).

We used a python script to parse the transcript and differentiate the various parts of the movies into the types scene informations and dialogues. We differentiated each line in the script as “sceneInfo” or as “dialogue” and then attributed features like last-speaker (name of person who spoke last before it), content and scene number (incremented on each occurrence of “SCENE Change” in the transcript) to them.

Thus we obtain a list of ordered scene-info and dialogue entities with their attributes last speaker, content and scene number. We also use the subtitles for the movie to attribute time to each dialogue in the script. To process the subtitle and matching to the dialogues, we tokenized each dialogue in the script as well as the subtitle and used the *Edit Distance* measure w.r.t. tokens to find the closest match.

The table below shows a sample of dialogues and scene-info in an initial scene of the movie:

typ	speaker	content	scenenum	inde	time
dialogue	commodus	Senator Gaius, Senator Falco. Beware of Gaius, he will pour honeyed potion in y…	5	111	00:18:33
dialogue	gaius	Well, Rome WAS founded as a Republic.	5	112	00:18:36
dialogue	commodus	Yes and in a republic, the senate has the power. But Senator Gaius is not influenc…	5	113	00:18:39
dialogue	falco	Where do you stand General? Emperor or Senate?	5	114	00:18:44
dialogue	maximus	A soldier has the advantage of being able to look his enemy in the eye, Senator.	5	115	00:18:47
dialogue	gaius	You know, with an army behind you, you could be extremely political.	5	116	00:18:52
dialogue	commodus	I warned you, but I shall save you. Senators.	5	117	00:18:57
sceneInfo	commodus	He pulls Maximus away and they leave the Senators. Commodus continues his co…	5	118	NULL
dialogue	commodus	I m going to need good men like you.	5	119	00:19:12
sceneInfo	maximus	Cautiously asking,	5	120	NULL
dialogue	maximus	How may I be of service, Highness?	5	121	00:19:16
dialogue	commodus	You are a man who knows what it is to command. You give your orders, the order…	5	122	00:19:22
sceneInfo	commodus	Glancing over to Marcus and back to Maximus.	5	123	NULL
dialogue	commodus	Can I count on you, when the time comes?	5	124	00:19:34
sceneInfo	maximus	Following Commodus glance, he looks at Marcus and back to Commodus.	5	125	NULL
dialogue	maximus	Highness, when your father releases me I intend to return home.	5	126	00:19:43
dialogue	commodus	Home, well no one has earned it more. Don t get too comfortable – I may call on…	5	127	00:19:51
sceneInfo	commodus	A look of concern over Commodus remarks about future service, followed with s…	5	128	NULL
dialogue	commodus	She has not forgotten you. And now you are the great man.	5	129	00:20:03
sceneInfo	commodus	SCENE CHANGE – Lucilla and her maid are seen peering out from the tent to wher…	6	130	NULL
dialogue	marcus	If only you had been born a man. What a Caesar you would have made.	6	131	00:20:17
dialogue	lucilla	Father	6	132	00:20:25

Proposed Approach

We divide our work in this project into 3 modules. Module 1 focuses on the character to character relationship modelling and build up analysis, whereas in the second module and third module we perform the character modelling and analyze the global flow of emotions respectively. We have tried some approaches to Module 1 and Modules 2 and 3 are explained in the future work section.

Module 1 : Character Relationship Modelling

Two facets of relationships among characters are analysed, namely

- Strength of relationships
- Relationship Modelling and Analysis of Build-Up (Temporal Variations)

Relationship Strength Analysis

In this simple yet effective approach, the number of times characters have interacted or mentioned in a scene is used as a measure to determine how strongly characters are related among each other. Following steps were performed to find the number of interactions :

1. A character character 2-d matrix S is initialized with a score of 0.
2. For each scene, set of all the active characters (V) is constructed. It contains all the characters mentioned in the scene info, all the characters having dialogues in that scene (A) and all the characters mentioned in the dialogues. A simple pseudo code is written below:

```
For each character a in A :
     For each character v in V:
          S(a,v) is incremented by 1/|V|
```

3. A graph is constructed where nodes are characters in the movie and edges are weighted by the scores obtained from the matrix S.

Relationship Modelling and Analysis of Build-Up

Emotions prevalent between the two characters are analysed by scoring the words used in the scene info concerning those two characters on the basis of their semantic similarity with the broad relationships “love” and “hate”. These are broad relationships since “love” also takes into account friendliness and other similar relationships. The steps carried out in the analysis are :

1. Two character character 2-d matrices, L and H respectively are initialized with a score of 0.
2. The sentences in the scene info are tokenized and the stopwords are removed.

3. The two characters mentioned in the sentence are extracted : If pronouns are used then the last mentioned nouns are assumed to be the characters corresponding to the pronouns (in future we plan to use *dependency parsing* or *coreference resolution* to obtain the characters from the pronouns).

4. The semantic similarity scores for all the tokens in the sentence is obtained using the UMBC Semantic Similarity Service - Swoogle. The scores are added to the fields corresponding to the characters in the matrices L and H.

5. The snapshots of the matrices L and H are taken after specific time frames (giving us the cumulative relationship from beginning till that time frame) to obtain the temporal change in the emotional behaviour of the relationships.

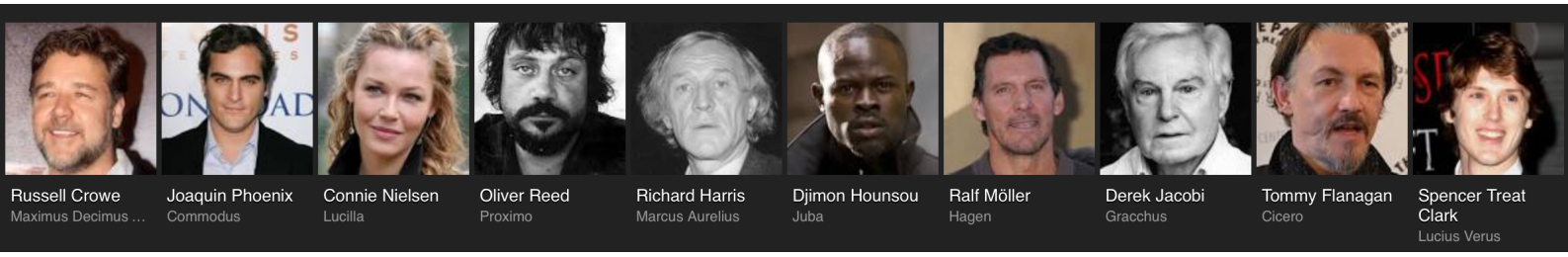
Plots for few sample pair of characters to visualize these temporal changes are shown in the results section of this report.

Failed approaches to Relationship Modelling

Initially we tried using a rule based approached that is commonly used which ascribes a set of causal words for each category of relationship. We chose 6 broad categories as "fight", "love", "respect", "action", “trust" and “mistrust”. We associated around 3-7 causal words to them and for each of tokens we cumulated the WordNet path_similarity (associating to the distance w.r.t. hyponyms/hypernyms) to find out the weights for the categories involved in the relationship corresponding to the characters mentioned in the sentence. The results were not interesting and usually favored towards fight, love and action since the dialogues and scene information very frequently mentions words in that context. *Because of these difficulties we reduced the number of categories to only 2 broad ones as mentioned before.* We also tried a second approach to use word2vec deep learning approach (by Google) to get a better similarity measure directly to the categories without using WordNet similarities. This also fails since word2vec gives >0.5 similarity to hate and love since they appear in similar contexts.

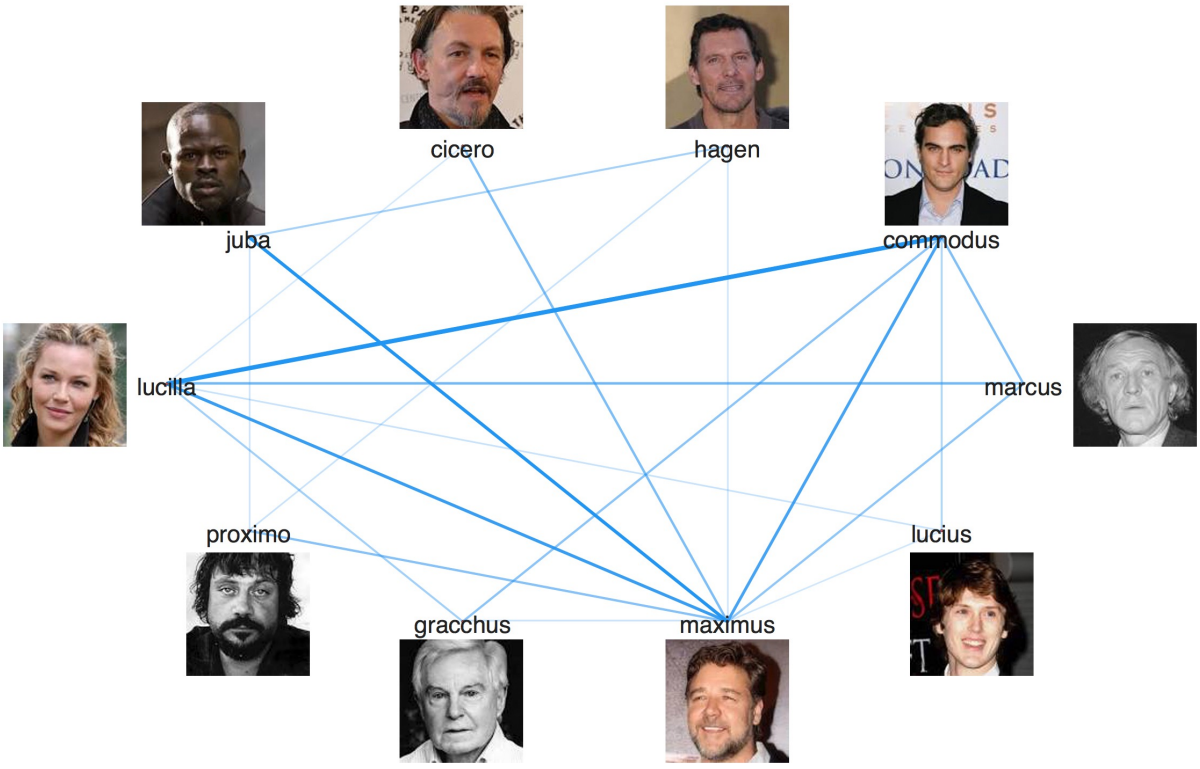
Results

Relationship Strength Analysis



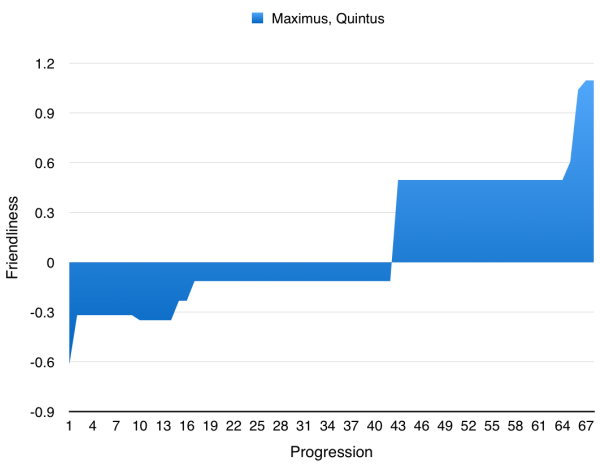
For evaluating the algorithm described in the previous section, we plotted a graph having the top 10 common characters (as shown in the above image) as the nodes and the strengths of the relationships as the weights. The edge widths and opacity were made

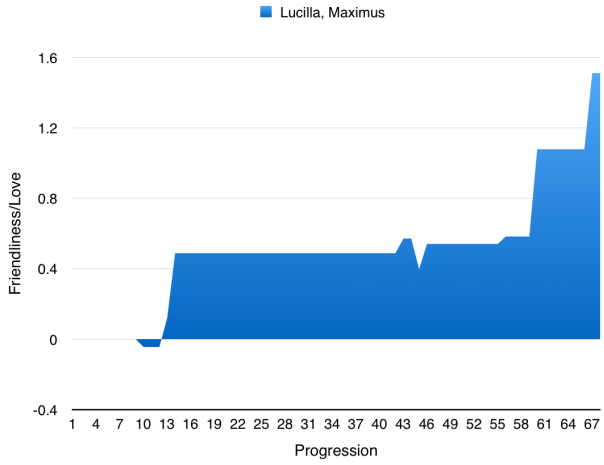
proportional to the edge weight (relationship strength) to give a better visualization of the results.



Relationship Modelling and Build-Up Analysis

For evaluating this part of the work, we plotted a chart showing the cumulative **weight[love]-weight[hate]** score over different time portions of the movie to get a sense of build-up. While this was not very accurate over all the relationships, it was quite insightful for some relationships shown below. For example, the relationship between Lucilla and Maximum is slightly bitter in the initial part of the movie and the love (or friendliness) increases as movie progresses.





Work Plan

Module 2: Character Modelling or Profiling

We aim to classify a character into multiple categories like Hero, Villain, etc. We are in the process of working it out

Task: Use Dependency Structure parsing for dialogue to improve identification of which character is carrying the action and who is the victim

Target Benefit: Improving accuracy for bi-directional relationship and character profiling

Task: We have used scene information for approximating the Character-Character Relation Matrix due to the fact that scene informations are narrative in style and generally tends to have Character-Action-Character format. Further, we aim to include the the dialogues exchanged between two character.

Target Benefit: Increase knowledge about the relationship

Task: Include more categories (possibly LIWC categories) other than love or hate.

Target Benefit: Improve accuracy for emotion tagging

Module 3: Overall Flow of Emotions

A similar approach to sentiment analysis between the relationships cumulated over all the scenes. We are yet to explore the specifics of the implementation.

Evaluation Plan

Till now, the evaluation has been manual based on the human analysis of the results. This analysis is carried out by the students who have watched the movie more than once so they know a fair amount of details about the relationships and emotions in the movie.

However, we do plan to build up a mathematical model for evaluation based on existing databases. The result obtained will be Character-Character Pair denoting the relationship shared between the character.

We will compare it with the actual relationship verified through IMDB database.

CR = Correctly classified relationships

TR = Total number of relationships

Accuracy = (CR/ TR)%

References

1. Baird, Eric T. Nalisnick Henry S. "Character-to-Character Sentiment Analysis in Shakespeare's Plays." Proceedings of the Conference. 2013.
2. Janakiraman, Krishna. "Extracting Character Relationships From Stories."
3. Valls-Vargas, Josep, Santiago Ontanón, and Jichen Zhu. "Toward character role assignment for natural language stories." Proceedings of the Ninth Artificial Intelligence and Interactive Digital Entertainment Conference. 2013.
4. Laptev, Ivan, et al. "Learning realistic human actions from movies." Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on. IEEE, 2008.