Analysis of *Lord of the Rings* Scripts

Phillip Yoon ‘20, Hamza Mahmood ‘20, Maharan Murshed ‘21, Diamond Acharya ‘21. Princeton University

Motivation

- Achieve better understanding of characters’ distinctive traits and inter-relationships in *The Lord of the Rings* series.
- Visualization of story arc leads to better understanding of plot
- Compare different methods for text analysis

Prior Work

- Network analysis on Game of Thrones
  - Plots a graph with the characters as the nodes and edges added if two characters have a connection between them. The size of the nodes are proportional to the number of connections of the characters.
- Analysis of emotional arcs on popular literary works
  - Plots and analyzes the progression of emotional arcs for popular novels by using a sentiment dictionary. Finds that the same six emotional arcs appear in most popular novels.
- Sentiment analysis on Harry Potter Texts
  - Analyzes the graphs of different sentiments in the Harry Potter texts.

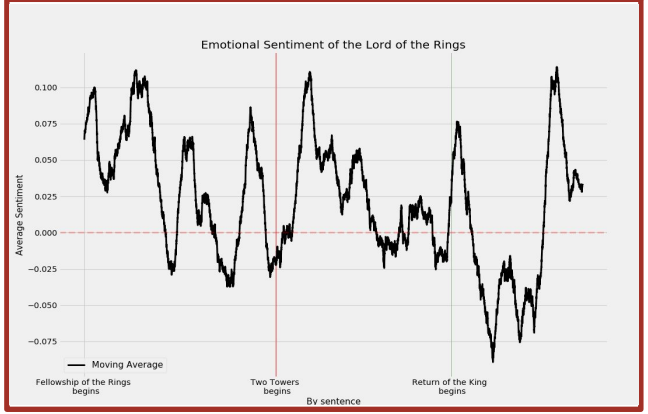
Our Project

1. Predict the character identity given a dialogue
   - Take a random dialogue and find out who the most likely character to have spoken that dialogue is
2. Perform sentiment analysis
   - Analyze emotional arc of the books as a whole
   - Analyze emotional arcs of particular characters
3. Perform network analysis
   - Find out who is connected with who
   - Find out who has the highest / lowest connections

Project Future Work

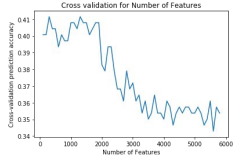
- Extend our model to be applied to other texts and books
- Predict events for ongoing series (e.g. character deaths, major plot points)
- Integrate sentiment analysis with network connections to classify inter-relationships between characters

Sentiment Analysis

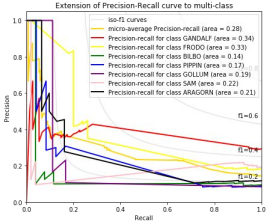


Character Prediction From Dialogue

- Designed a model to classify speaker of dialogue from a sentence of text
- Data Processing: Tokenization, stop words, stemming
- Feature Extraction: K-grams
- Feature Selection: Chi-Squared



Classifier	Accuracy	Precision	Recall	F1-Score
NB	0.3971	0.4978	0.3971	0.3179
LR	0.3827	0.4515	0.3827	0.3174
SVM	0.3791	0.4204	0.3791	0.3250
DT	**0.4116**	**0.5196**	**0.4116**	**0.3676**
K-NN	0.3755	0.4392	0.3755	0.30782
NN	**0.4116**	0.4549	**0.4116**	0.3598

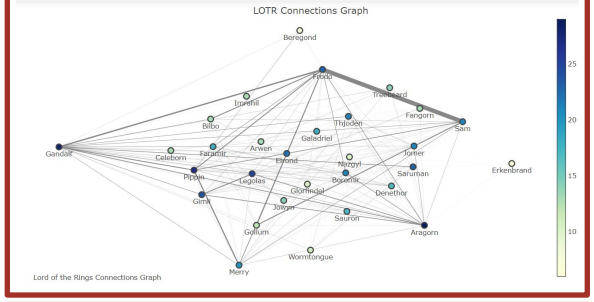


- Gandalf, Frodo, and Sam have highest prediction performance
- Decision Tree Classifier provided highest accuracy, recall, and F1-score

Sentiment Analysis (cont’d)

- VADER (Valence Aware Dictionary and Sentiment Reasoner)
  - Lexicon and rule-based sentiment analysis tool
  - Graph plots the average of negative, positive, and neutral sentiments
  - Positive values indicate positive sentiment, and negative values indicate negative sentiment
- Input consisted of raw individual sentences from across all 3 books
- Moving Average used a window size of 1000 sentences
  - Takes the average sentiment over the previous 1000 sentences

Lord of the Rings Connections Map



Connectivity and Communities

Louvain’s method communities with resolution = 0.6:

The Warriors: Aragorn, Boromir, Fangorn, Gimli, Legolas.
The Elves: Arwen, Celeborn, Elrond, Galadriel, Glorfindel
The Race of Men: Beregond, Denethor, Faramir
The Hobbits+: Bilbo, Frodo, Gollum, Sam
Rohan: Erkenbrand, Jomer, Jowyn, Imrahil, Thjoden
The Baddies and the Person They Hate: Gandalf, Nazgyl, Saruman, Sauron, Wormtongue
The Twins and Treebeard: Merry, Pippin, Treebeard

Top Characters, Degree Centrality (weighted):	Top Characters, Betweeness Centrality (unweighted):	Top Characters, PageRank Centrality (weighted):
1. Frodo	1. Aragorn	1. Frodo
2. Gandalf	2. Gandalf	2. Gandalf
3. Sam	3. Pippin	3. Sam
4. Aragorn	4. Legolas	4.Aragorn
5. Pippin	5. Frodo	5. Pippin
6. Merry	6. Saruman	6. Merry
7. Gimli	7. Gimli	7. Gimli
8. Legolas	8. Jomer	8. Legolas
9. Gollum	9. Thjoden	9. Boromir
10. Boromir	10. Elrond	10. Faramir