Argument Clustering in Debate Format: Game of Thrones

Authors

B. Sc. Ivana Hacajová, B. Sc. Omar Abugharbiyeh, B. Sc. Volkan Tatlikazan

Affiliations

M.A. M. Sc. Tobias Eder

Social Computing Research Group, TUM

INTRODUCTION

Argument Clustering is a task, in which we try to group together people with the same or similar opinion, usually on some specific topic. In our case, these topics have to be selected automatically, using topic modeling techniques.

Topic modeling is an unsupervised machine learning technique that scans a set of documents, detects word and phrase patterns within them, and automatically clusters word groups and similar expressions that best characterize a set of documents.

We chose the fantasy series Game of Thrones as our debate medium since it is a heavily political series. We get the arguments for each character by using topic modeling to assign a topic to each line.

DATASET

The Dataset is a collection of all the lines spoken thoughout every season in Game of Thrones. It is a table with each row containing a single line, the character that spoke it, the episode number, as well as the season number. All of these lines are ordered in the order they were spoken in during the series.

PREPROCESSING

- **EXAMPLE SENTENCE**
- **TOKENIZING**
- **CLEANING TOKENS***
- **LEMMATIZING**
- **POS TAGGING****
- **NOUNS**
- **NOUNS AND VERBS**
- **NOUNS, ADJECTIVES AND VERBS**
- **NOUNS, ADJECTIVES, ADVERBS AND VERBS**

** Extracted after tokenizing

Never forget what you are. The rest of the world will not. Wear it like armor, and it can never be used to hurt you.
'Never', 'forget', 'what', 'you', 'are.', 'The', 'rest', 'of', 'the', 'world', 'will', 'not.',
'Wear', 'it', 'like', 'armor', ',', 'and', 'it', 'can', 'never', 'be', 'used', 'to', 'hurt', 'you', '.'
'forget', 'hurt', 'never', 'like', 'wear', 'armor', 'used', 'rest', 'world'
'forget', 'hurt', 'never', 'like', 'wear', 'armor', 'used', 'rest', 'world'
('forget', 'VB'), ('hurt', 'VB'), ('never', 'RB'), ('like', 'IN'), ('wear', 'VB'), ('armor', 'NN'), ('used', 'VBN'), ('rest', 'JJ'), ('world', 'NN')
'armor', 'world'
'forget', 'hurt', 'wear', 'armor', 'world'
'forget', 'hurt', 'wear', 'armor', 'rest', 'world'

'forget', 'hurt', 'never', 'wear', 'armor', 'rest', 'world'

***Cleaning Steps:**

- **Deleting foreign language uses (e.g. Dothraki)**
- **Deleting stopwords (e.g. 'ah', 'and', 'it')**
- **Deleting Punctuations**

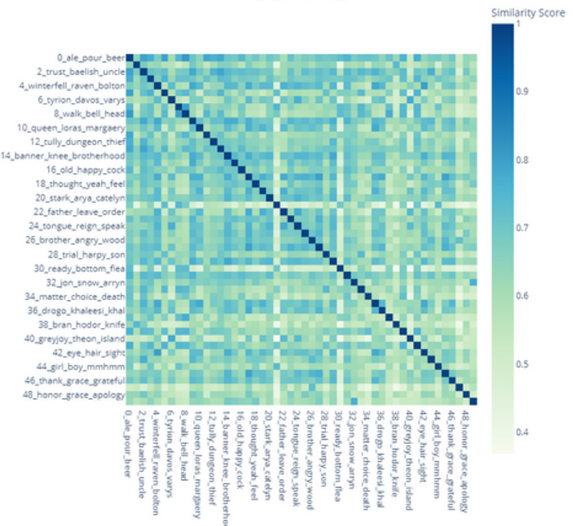
TOPIC MODELLING

For our purposes, we chose BERTopic in order to assign a topic to each of the sentences. The process of properly assigning a topic to each sentence in our dataset goes as follows:

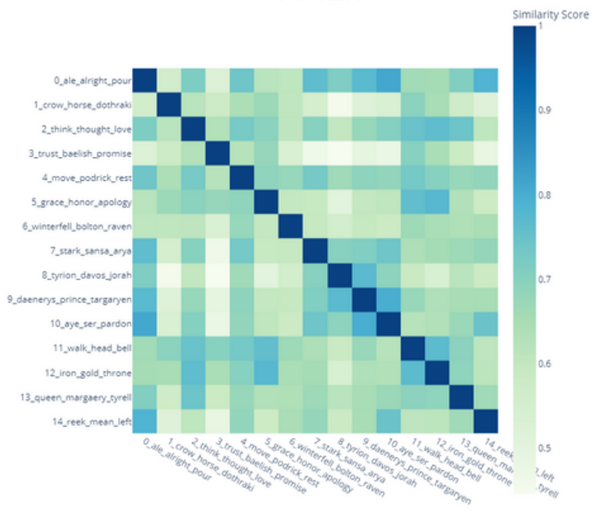
- **Use BERTopic to assign each of the lines to one of 50 separate topics.**
- **Reduce the topics based on their c-TF-IDF similarity down to 15 topics.**
- **Examine the final topics and the top words for those topics in order to choose appropriate topics to utilize in Argument Clustering.**

SIMILARITY INDEX

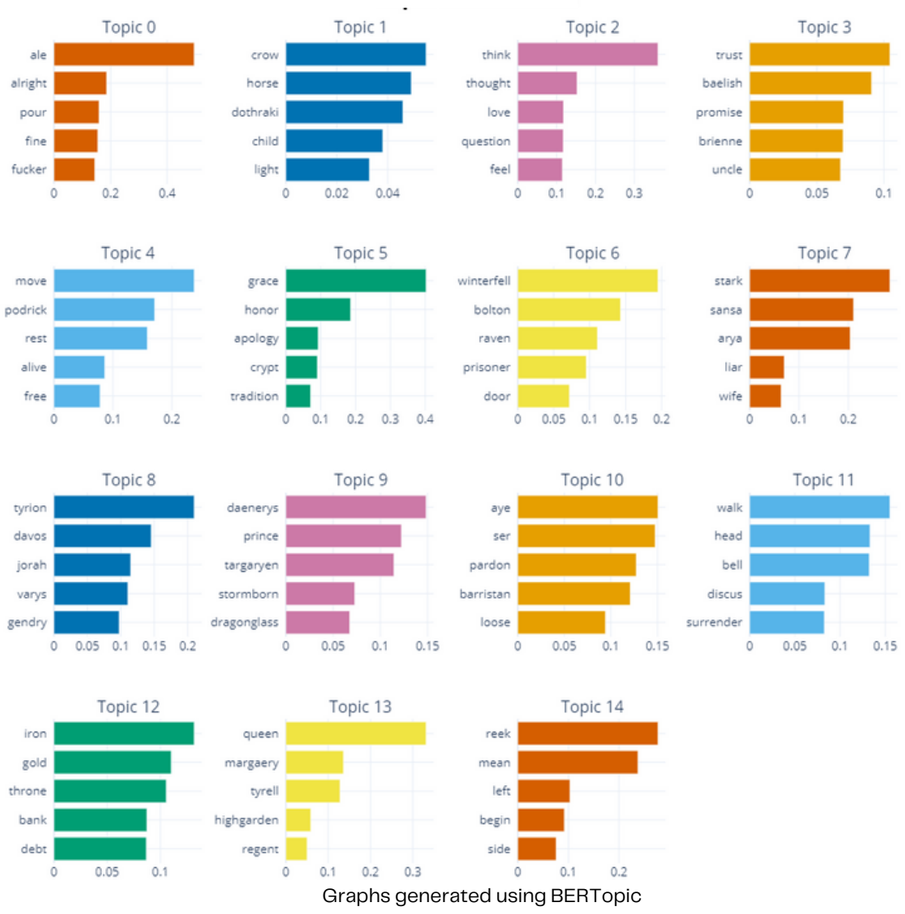
BEFORE



AFTER



WORDS FOR TOPICS

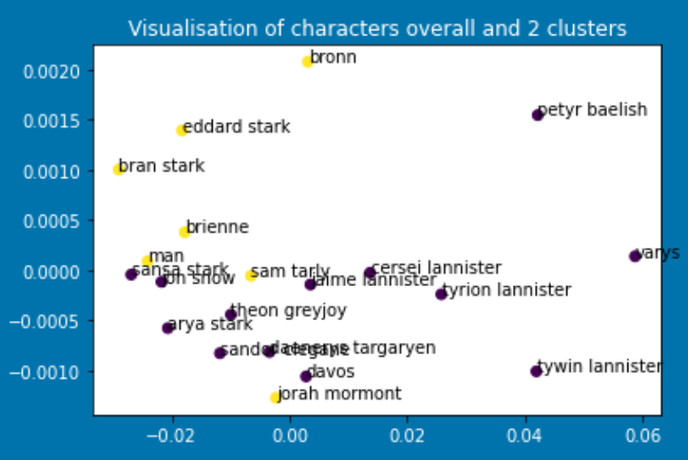


Graphs generated using BERTopic

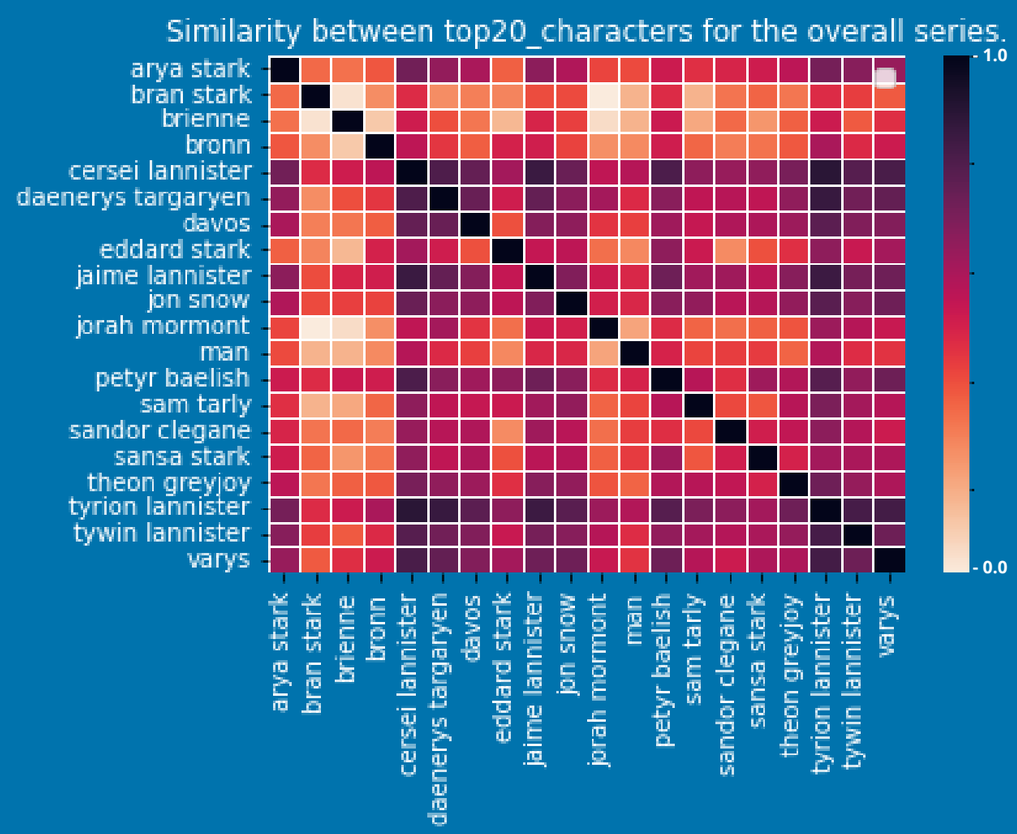
ARGUMENT CLUSTERING

Once we assigned topics to the lines, we computed Doc2Vec embeddings for each one of them. For each topic we got the mean of the embeddings for the selected characters.

The result was a numerical representation of characters, which we used to perform the clustering and similarity analysis with.



CHARACTER SIMILARITY



MOST AND LEAST SIMILAR CHARACTERS FOR THE SELECTED TOPICS

Topic	Most similar characters	Least similar characters
Topic 3: Petyr Baelish	Brienne, Tyrion Lannister	Bronn, Jorah Mormont
Topic 5: Winterfell	Jaime Lannister, Theon Greyjoy	Bronn, Jorah Mormont
Topic 7: The Starks	Arya Stark, Cersei Lannister	Brienne, Petyr Baelish
Topic 8: Daenerys' Advisors	Daenerys Targaryen, Tyrion Lannister	Bran Stark, Petyr Baelish
Topic 9: Castlery Rock	Arya Stark, Tyrion Lannister	Eddard Stark, man
Overall	Cersei Lannister, Tyrion Lannister	Bran Stark, Jorah Mormont