Non-Negative Matrix Factorization (NMF)

What is Non-Negative Matrix Factorization (NMF)?

Non-negative matrix factorization (NMF) is a group of algorithms in multivariate analysis and linear algebra where a matrix V is factorized into two matrices W and H, with the property that all three matrices have no negative elements. This non-negativity makes the resulting matrices easier to inspect.

Clustering in NMF

Non-Negative Matrix Factorization (NMF) is an unsupervised technique which means that there are no labeling of clusters that the model will be trained on. The way it works is that, NMF decomposes high-dimensional vectors into a lower-dimensional representation. These lower-dimensional vectors are non-negative which also means their coefficients are non-negative. Using the original matrix V, NMF will give you two matrices (W and H). W gathers the clusters found and H is the coefficients (weights) for those clusters. In other words, V is the CoreNLP Sentiment Score table, H is the texts in those clusters and W is the weights of each cluster.

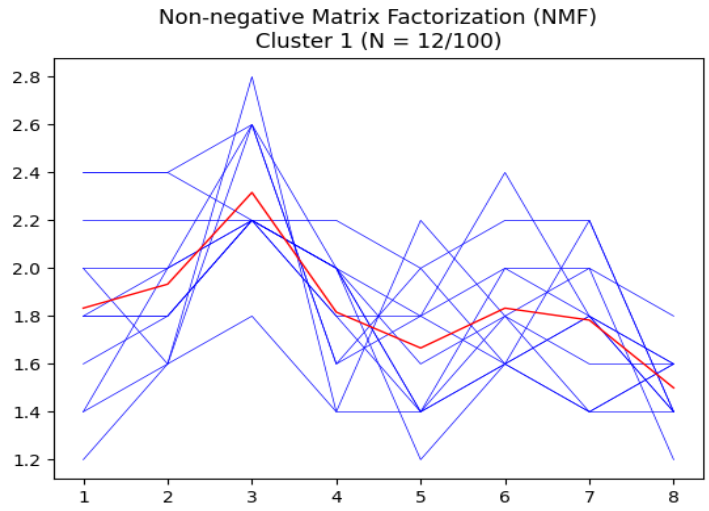
$$\underbrace{\mathrm{V}(:,j)}_{j\text{th document}} \approx \sum_{k=1}^{r} \underbrace{W(:,k)}_{k\text{th cluster}} \underbrace{H(k,j)}_{\substack{\text{importance of } k\text{th cluster} \\ \text{in } j\text{th document}}}, \qquad \text{with } W \geq 0 \text{ and } H \geq 0.$$

Output

The NLP Suite produces 2 different output graphs for each cluster: a main plot showing NMF sentiment score throughout each cluster and a more precise subplot that shows distinct sentiment score for each text in each cluster.

NMF Clusters plots

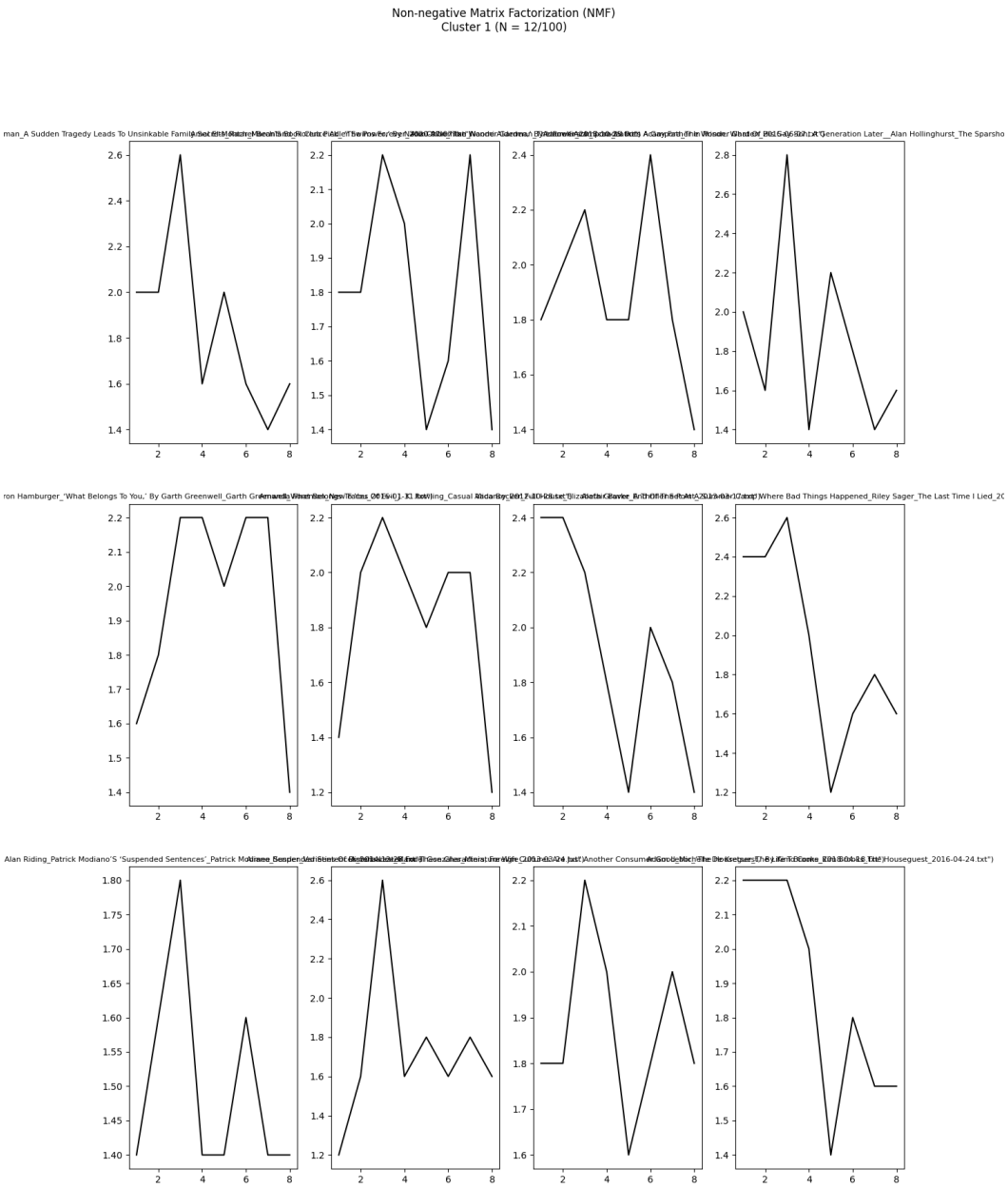
NMF Clusters are displayed in the NLP Suite by line plots. Each Blue line represents a single text while the red line represents the general trend of the cluster.



It is important to notice that the x-axis corresponds to the vector size set up by the user before running the algorithm and the y-axis is the sentiment score. One graph will be created for each cluster.

NMF Clusters Subplots

The subplots generated let you have separated graphs for each text in your cluster. These subplots are ideal if you want to do closer reading onto a single text.



CSV

In addition to the 2 above graphs, you can use the csv file generated to find the content of each cluster in *NMF Documents.csv*

Cluster ID	Sentiment Score File Name	Original File Name
Cluster 1	/Users/XXXXX/NLP_CoreNLP_sentiment_Dir_times_reviews.csv	/Users/XXXXX/times_reviews/Erin Somers_The Freelance Life. But With Superheroes_Natalie Zina Walschots_Hench_2020-09-22.txt
Cluster 1	/Users/XXXXX/NLP_CoreNLP_sentiment_Dir_times_reviews.csv	/Users/XXX/times_reviews/Terese Svoboda_A Florida Trailer Park Awash In Lethal Weapons_Jennifer Clement_Gun Love_2018-04-06.txt
Cluster 1	/Users/XXXXX/NLP_CoreNLP_sentiment_Dir_times_reviews.csv	/Users/XXX/times_reviews/Hugo Lindgren_The Back Story_Jane Gardam_Last Friends_2013-04-14.txt
Cluster 1	/Users/XXXXX/NLP_CoreNLP_sentiment_Dir_times_reviews.csv	/Users/XXXXX/times_reviews/Dean Bakopoulos_Adam Ross,Â©S Unsettling Stories_Adam Ross_Ladies And Gentlemen_ Stories_2011-07-24.txt

The CSV file contains the Cluster ID of each text as well as the directory of the original .txt file and the directory of the Sentiment Scores of the .txt file. You can use the csv file generated to find the content of each cluster (the file name and cluster number) for the NMF results.

References

Burrows, J.F. 1987. "Word-Patterns and Story-Shapes: The Statistical Analysis of Narrative Style." *Literary and Linguistic Computing*, Vol. 2, No. 2, pp. 61-70.

Franzosi, Roberto. 2004. *From Words to Numbers*. Cambridge, UK: Cambridge University Press.

Franzosi, Roberto. 2010. *Quantitative Narrative Analysis*. Thousand Oaks, CA: Sage.

Reagan, Andrew J., Lewis Mitchell, Dilan Kiley, Christopher M. Danforth, and Peter Sheridan Dodds. 2016. "The Emotional Arcs of Stories Are Dominated by Six Basic Shapes". *EPJ Data Science*, Vol. 5, No. 31, pp. 1-12.

Vonnegut, Kurt. https://www.youtube.com/watch?v=oP3c1h8v2ZQ

Vonnegut, Kurt. https://www.youtube.com/watch?v=GOGru_4z1Vc

Vonnegut, Kurt. 2005. "Here is a Lesson in Creative Writing." In: pp. 23-28, Kurt Vonnegut, *A Man Without a Country. Edited by Daniel Simon.* New York: Seven Stories Press.

Salgado, R. (2020, April 16). *Topic modeling articles with NMF*. Medium. Retrieved April 17, 2022, from https://towardsdatascience.com/topic-modeling-articles-with-nmf-8c6b2a227a45