I read your true colors
Machine Learning for Natural Language Processing 2020

Antoine Belloir
antoine.belloir@ensae.fr

Charlotte Gallezot
charlotte.gallezot@ensae.fr

1 Problem Framing

Here, we sought to investigate the linguistic relationship between colors and emotions. This is a problem of interest for various fields such as cognitive sciences or literature (Lafourcade et al., 2014; Desikan et al., 2020; Skard, 1946; Hupka et al., 1997). Several studies explored this question with experiments leveraging human judgment (Hupka et al., 1997), however this is costly and time-consuming. Here, we used NLP tools to tackle this question. We decided to restrain our study to a simpler version of the problem by defining emotions as valence (positive/negative), and colors as discrete categories (black, white, blue, green, red, yellow, orange, pink ,purple, brown, grey (Sahlins, 1976)). Hence, our research question can be formulated as: are colors associated with a specific valence in literature? Our project is available as a colab notebook [1].

2 Experiments Protocol

A difficulty is that colors can be either be explicitly expressed or suggested (e.g. herb suggests green). To capture this dimension we created a word-color dictionary based on co-occurences. Then, we built a classifier to predict if a given paragraph is positive or negative. Finally, we used these tools to answer our question.

2.1 Word color dictionary

We built a word-color dictionary by comparing co-occurrences -at the sentence level- of words with each basic color. We restrained our analysis to noun and adjectives. We used literary texts ($\sim$ 2000 books) from the Gutenberg project to list co-occurrences. We derived descriptive statistics of word-color co-occurrences vector. Then we assigned to words the color with which they appeared the most, except if this color co-occurred less than 1/3 of the time with the word -in this case no color was associated with the word-. To get a qualitative appreciation of our dictionary we tested a few keys to see if the results were plausible.

2.2 Valence classification for sentences

We built a model to classify positive or negative sequences. We used the IMDB dataset (Maas et al.) which gathers 50.000 movie reviews labeled as positive or negative. We tried several models :

- Our first model use LASER (Artetxe and Schwenk, 2019) to get embeddings from paragraphs. Then classic classifiers (LDA 1.1 SVM 1.2 RF 1.3 AdaBoost 1.4) [2] are used for the classification task.
- We compare these first approaches with transformer-based models. Namely, we train BERT and RoBERTa models with a final linear layer to perform binary classification.
- In addition, we evaluate the performance of a SentenceTransformer model (Reimers and Gurevych, 2019) on two settings. First, zero-shot performance of a SentenceBERT in a question answering-like setting : we encode spearately the questions *"Is the movie great ?"* and *"Is the movie bad ?"* and the review, and then classify the review based on the highest cosine similarity between the sentence and the questions. Second, SentenceBERT trained on a Semantic Similarity Task, using a siamese setting to enforce sentence embeddings to be close if the sentences share the same sentiment. To classify a review from the validation data set, we sample a positive review and a negative review from the training data set and assign the validation sample

[1] Colab link here

[2] Refer to Appendix for parameters

to its class based on the lowest cosine distance between the validation sample and the two reference samples.

2.3 Analysis of color/sentiment association

We combined word color associations and sentence classification to investigate our problem. We used 106 books [3] from the Gutenberg Project. We split them in paragraphs, predicted their valence with our best classification model and derived their color distribution. We then computed descriptive statistics to explore our problem.

3 Results

3.1 Word-color dictionary

We derived a few statistics of our word-color cooccurences dictionary, reported in Appendix 2. Black and white are the most frequent colors, followed by red, green, and blue. Colors such as pink or orange are quite rare. There are important variations among words. We report in Appendix 3. the colors we found for a few chosen words. Over 34 words, 26 have plausible colors. It is good that our dictionary detects shades of a color ("cerulean", "turquoise"), and colors of obviously colored objects ("chalk", "saphir", "sky"). Interestingly, we see that black and white are strongly correlated with negative and positive abstract words such as optimistic or guilt.

3.2 Selection of the best classification model

The accuracies of all models are reported in the Appendix.The RoBERTa model achieve the best accuracy (92%) and was used for the statistical analysis of color/sentiment association. An interesting point is the good zero-shot performance of the SentenceTransformer (84%). However, when trained on a Semantic Similarity Task (SST), we achieve the same accuracy. Our best chance to obtain a better model would be to fine-tune the SentenceTransformer trained on SST on the same classification task as BERT and RoBERTa, after having added a classification head.

3.3 Valence-Color analysis

We found that 78% of the texts were classified as positive. We could have expected an even repartition between positive and negative paragraphs. This might be due to the fact that we trained our model on a dataset only constituted of movie review, and hence it does not perfectly generalize. Interestingly there are more occurrences of colors in negative paragraphs (16 vs 7 on average). On average black is more frequent in negative paragraphs than in positive which is consistent with the classic association black/bad but this is not the case for white. Then, we can observe small differences (e.g. red more frequent in positive paragraphs) but it is not sure that they are significant. Exhaustive results can be found in Appendix section 5. Interestingly, when we look at results by book we observe variations. First, it seems that colors are distributed differently between books. Then, depending on books some colors have different valence. For instance in Pride and prejudice by Jane Austen red is more frequent in positive paragraphs whereas it is more frequent in negative ones in Swann's way by Proust. Blue is more frequent in positive paragraphs in Father Goriot by Balzac whereas it is almost even in other books.

4 Discussion/Conclusion

Time and space complexity The color classification at inference is linear in the number words in the sentence, while the transformer architecture used for sentiment analysis is quadratic in the sentence's length and linear in the embedding's dimension. The color classification is a simple dictionary, while BERT-like architecture necessitate to store typically 120M parameters. Overall, the sentiment analysis has a predominant spatial and computational complexity.

Colors emotions and words Our results suggest that the color-valence question should be considered in context (book, author, era). Furthermore, the variation detected in color use between author is worth being explored. A future research question could be: is it possible to determine a color fingerprint for each author?

To pursue our study and explore this new question our methods could be enhanced. First the color/word dictionary could be compared exhaustively to human judgment for external validation. It could also take into account the importance of context in suggested color (e.g. "sea" is blue, "sea at night" is dark blue). One other lead we could explore is the use of compsyn (Desikan et al., 2020), a word embedding based on colors. Then, our classification model could be trained with a data set that is not limited to movie reviews.

[3] from number 100 to 200 plus number 1342, 1212,1237,1715,2701,7178

References

Andrew L Maas, Raymond E Daly, Peter T Pham, Dan Huang, Andrew Y Ng, and Christopher Potts. Learning Word Vectors for Sentiment Analysis. page 9.

Sigmund Skard. 1946. The Use of Color in Literature: A Survey of Research. *Proceedings of the American Philosophical Society*, 90(3):163–249. Publisher: American Philosophical Society.

Marshall Sahlins. 1976. COLORS AND CULTURES. 16(1):1–22. Publisher: De Gruyter Mouton Section: Semiotica.

Ralph B. Hupka, Zbigniew Zaleski, Jurgen Otto, Lucy Reidl, and Nadia V. Tarabrina. 1997. The Colors of Anger, Envy, Fear, and Jealousy: A Cross-Cultural Study. *Journal of Cross-Cultural Psychology*, 28(2):156–171. Publisher: SAGE Publications Inc.

Mathieu Lafourcade, Nathalie Le Brun, and Virginie Zampa. 2014. Crowdsourcing Word-Color Associations. In *Natural Language Processing and Information Systems*, Lecture Notes in Computer Science, pages 39–44, Cham. Springer International Publishing.

Mikel Artetxe and Holger Schwenk. 2019. Massively Multilingual Sentence Embeddings for Zero-Shot Cross-Lingual Transfer and Beyond. *Transactions of the Association for Computational Linguistics*, 7:597–610. ArXiv: 1812.10464.

Nils Reimers and Iryna Gurevych. 2019. Sentence-bert: Sentence embeddings using siamese bert-networks. *CoRR*, abs/1908.10084.

Bhargav Srinivasa Desikan, Tasker Hull, Ethan O. Nadler, Douglas Guilbeault, Aabir Abubaker Kar, Mark Chu, and Donald Ruggiero Lo Sardo. 2020. comp-syn: Perceptually Grounded Word Embeddings with Color. *arXiv:2010.04292 [cs]*. ArXiv: 2010.04292.

Appendix

1. LASER + Classifiers models - Parameters
Training with 95% of the IMDB dataset, testing with %.

- Model 1.1 (LDA): n components=1
- Model 1.2 (SVM): polynomial kernel
- Model 1.3 (RF): n estimators=200, max depth=2
- Model 1.4 (AdaBoost): n estimators=100

2. Word-color co-occurences

Color	mean	std
White	0.244	0.339
Black	0.209	0.323
Grey	0.036	0.151
Orange	0.016	0.101
Purple	0.027	0.128
Blue	0.087	0.218
Yellow	0.054	0.175
Red	0.139	0.274
Brown	0.064	0.198
Green	0.106	0.245
Pink	0.016	0.099

3. Word-color dictionary

- White: butter, swan, daytime, nighting, chalk, organ, porcelain, eyeball, optimist, immaculate
- Black: sphynx, hell, abyss, darki, guilt, sorceress, lava, nightmare, sleepless
- Blue: sky, cobalt, cerulean, turquoise
- Purple: saphir, ultraviolet
- Red: redder, forrest, ridinghood
- Green: tyger, scars, woomb, grass
- Yellow: fever

4. Classifiers accuracy
Table 1. Accuracy of each model

Model	1.1	1.2	1.3	1.4	2	**3**	4	5
Accuracy	0.84	0.85	0.77	0.80	0.88	**0.92**	0.84	0.84

- Model 2 : BERT
- Model 3 : **RoBERTa**
- Model 4 : Zero-Shot SentenceTransformer
- Model 5 : Semantic Similarity SentenceTransformer

5. Analysis results Table 2. Mean frequence of each color in positive paragraphs over all books .

Color	White	Black	Grey	Orange	Purple	Blue	Yellow	Red	Brown	Green	Pink
Mean	0.455	0.227	0.005	0.001	0.005	0.138	0.004	0.113	0.029	0.019	0.002

Table 3. Mean frequence of each color in negative paragraphs over all books

Color	White	Black	Grey	Orange	Purple	Blue	Yellow	Red	Brown	Green	Pink
Mean	0.461	0.260	0.001	0.001	0.005	0.115	0.004	0.096	0.033	0.021	0.002

Table 4. Mean frequence of each color in positive paragraphs in specific books (1: Jane Austen, Pride and Prejudice, 2: Herman Melville, Moby Dick, 3: Honoré de Balzac, Father Goriot, 4: Marcel Proust, Swann's way)

Color	White	Black	Grey	Orange	Purple	Blue	Yellow	Red	Brown	Green	Pink
Mean - 1	0.485	0.204	0.002	0	0.003	0.125	0.003	0.128	0.027	0.021	0.001
Mean - 2	0.613	0.194	0.002	0.001	0.005	0.054	0.006	0.065	0.034	0.024	0
Mean - 3	0.429	0.215	0.001	0	0.002	0.161	0.004	0.159	0.017	0.010	0
Mean - 4	0.449	0.248	0.003	0	0.006	0.162	0.005	0.086	0.018	0.016	0.006

Table 5. Mean frequence of each color in negative paragraphs over in specific books (1: Jane Austen, Pride and Prejudice, 2: Herman Melville, Moby Dick, 3: Honoré de Balzac, Father Goriot, 4: Marcel Proust, Swann's way)

Color	White	Black	Grey	Orange	Purple	Blue	Yellow	Red	Brown	Green	Pink
Mean - 1	0.458	0.245	0	0	0.005	0.133	0	0.079	0.050	0.028	0
Mean - 2	0.626	0.169	0.001	0	0.008	0.048	0.004	0.056	0.058	0.027	0.002
Mean - 3	0.446	0.241	0	0	0	0.118	0.006	0.152	0.028	0.006	0
Mean - 4	0.404	0.255	0	0	0.003	0.160	0	0.122	0.020	0.033	0