The 56th Annual Meeting of the Association for Computational Linguistics

ACL 2018

Author Response

Title: Learning Distributions of Meant Color
Authors: Lyndon White, Roberto Togneri, Wei Liu and Mohammed Bennamoun

Instructions

The author response period has begun. The reviews for your submission are displayed on this page. If you want to respond to the points raised in the reviews, you may do so in the boxes provided below.

We encourage you to explicitly respond to every point raised in **Weaknesses** and **Questions to Authors** sections in the reviews. Use the template provided in the response box as far as possible, but this is not obligatory.

Please note: *you are not obligated to respond to the reviews.*

For reference, you may see the review form that reviewers used to evaluate your submission. If you do not see some of the filled-in fields in the reviews below, it means that they were intended to be seen only by the committee. See the review form HERE.

Review #1

Appropriateness: Appropriate
Adhere to ACL 2018 Guidelines: Yes
Adhere to ACL Author Guidelines: Yes
Handling of Data / Resources: Yes
Handling of Human Participants: N/A

Summary and Contributions

The paper describes a model that maps from color names (sequences of one or more word tokens) to probability distributions in HSV color space. The key features of the model are that it (1) applies a simple morphological decomposition to the word sequence to separate out morphs such as 'ish' and '-' and (2) uses a recurrent neural network (specifically, a GRU network) to learn a compositional model of color names. This permits the model to extrapolate to color names that were not seen during training (provided that all tokens comprising the name were observed). Experiments show that the proposed model works as well as a noncompositional baseline for terms observed in the training data and is able to extrapolate to sequences of descriptors that were not observed in the training data, unlike the baseline.

Contribution 1: A compositional model that maps color names to probability distributions in HSV space.

Contribution 2: A sensible, non-compositional baseline model that maps color names to probability distributions in HSV space.

Contribution 3: Code implementing both models, and (I believe) the extrapolation dataset described in Section 4.2.

Strengths

1. The research appears to be completely reproducible: the authors provide code and (I think) their new extrapolation dataset. The training, development, and test data are available at http://mcmahan.io/lux/
2. The experiments are well designed and executed, so the results are convincing.
3. The paper contains a thorough and solid comparison to prior work in the area.

Weaknesses

1. The "discretize and then blur" framework described in Section 3.2 seems unnecessarily complicated. Why not first perform kernel density estimation using a Gaussian kernel (with the bandwidth optimized on the development data) and then discretize? Also, what criterion was optimized on the development data in the parameter sweep used to select the value of Ω?
2. There appears to be an error in the description of the discretization process. Because the Gaussian blurring kernel has support on the full real line, there should not be any problems with the baseline model predicting zero probabilities as described in Section 3.3. This should only happen if the Gaussian was truncated, which is not mentioned in the paper.
3. The training of the CDEST model is not well specified in the paper. I was able to determine that Adam is used as the optimizer by looking at the supplied code, but I'm still not sure how the weights are initialized. This information should be provided in the paper itself.

NLP Tasks / Applications: Moderate contribution
Methods / Algorithms: Moderate contribution
Theoretical / Algorithmic Results: N/A
Empirical Results: Moderate contribution
Data / Resources: Moderate contribution
Software / Systems: Moderate contribution
Evaluation Methods / Metrics: N/A
Other Contributions: N/A
Originality (1-5): 3
Soundness/Correctness (1-5): 4
Substance (1-5): 4
Replicability (1-5): 5
Meaningful Comparison (1-5): 5
Readability (1-5): 5
Overall Score (1-6): 4

Additional Comments (Optional)

The software would benefit from a little more documentation -- a more extensive README would suffice.

The papers by "Oord et al." and "van den Oord et al." mentioned at the bottom of page 4 have the same first author. It looks like there is a BiBTeX problem here.

to the color described, allows for knowledge → to the color described allows for knowledge

improves predicative capacity → improves predictive capacity

likelihoods of which colors are be intended → likelihoods of which colors are intended

To qualify our estimate of the distribution → To quantify our estimate of the distribution

overcoming data sparsity programs (Bengio et al., 2003). Including, the extreme case → overcoming data sparsity problems (Bengio et al., 2003), including the extreme case

which while lacking the generalisation capacity → which while lacking generalisation capacity

should be a corresponds field → should be a corresponding field

which propose methods for the automatic mapping of colors → propose methods for the automatic mapping of colors

a much courser level → a much coarser level

function with much larger number of colors → function with a much larger number of colors

and larger pools of respondents. In particular making uses of → and larger pools of respondents, in particular making use of

They extends beyond, all prior color naming systems → They extend beyond all prior color naming systems

direct inverse of their conditional language model, CDEST use a RNN to map → direct inverse of their conditional language model: CDEST uses an RNN to map

current work proposes as a distribution estimation → current work proposes a distribution estimation

and the color distribution; and is trained → and the color distribution, and is trained

Superficial checks were carried out on the accuracy of this assumption. → Superficial checks on the accuracy of this assumption were carried out.

on the degree of correctness → of the degree of correctness

to estimate a continuous probability distributions → to estimate a continuous probability distribution

Estimating a discrete conditional distributions → Estimating a discrete conditional distribution

Hue can elegantly be handled → Hue can be elegantly handled

but some will be shared with the bins either side, and further. → but some will be shared with the remaining bins in the space.

Best results were found for → The best results were found for

of the probably mass → of the probability mass

found their method to outperforming the more complex → found their method to outperform the more complex

requirement to learn the how the multiple terms → requirement to learn how the multiple terms

a Gated Recurrent Unit (GRU) (Cho et al., 2014) → a Gated Recurrent Unit (GRU) (Cho et al., 2014) layer

a Rectified Linear Unit (ReLU) (Dahl et al., 2013) → a Rectified Linear Unit (ReLU) (Dahl et al., 2013) layer

the output of the ReLU is → the output of the ReLU layer is

We chose GRU as the basis → We chose the GRU as the basis

found to preform similarly → found to perform similarly

well to LSTM → well to the LSTM

A component for processing per-term such as the GRU, is essential → A component that processes terms sequentially, such as a GRU layer, is essential

in the test set just well as → in the test set just as well as

demonstrate the benefits of the know-sharing → demonstrate the benefits of knowledge sharing

estimating the probably distribution of colors → estimating the probability distribution of colors

As the it learns how each term → As it learns how each term

Generating and resolving vague color reference → Generating and resolving vague color references

Review #2

Appropriateness	Appropriate
Adhere to ACL 2018 Guidelines	Yes
Adhere to ACL Author Guidelines	Yes
Handling of Data / Resources	Yes
Handling of Human Participants	N/A

Summary and Contributions

Summary: The paper proposes to represent a phrase of color description as a probability distribution over a color-space in order to cope with the problem that different people have different concept of colors. It designs and implements the CDEST model for predicting the color-space probability distribution of color. The CDEST model is evaluated with the Monroe dataset.

Contribution 1: Proposal of understanding a color as a probability distribution over a color-map for flexible manipulation of colors.

Contribution 2: Proposal of mapping of a sequence of color description words to a probability distribution over a color-map, say, HSV.

Contribution 3: Proposal of the CDEST model for predicting the color-space probability distribution of color.

Strengths

Strength argument 1: The CDEST model is designed and implemented.

Strength argument 2: The CDEST model is evaluated on how well it predicts a probability distribution over HSV color-space with the Munroe dataset.

Strength argument 3: Extrapolation is developed to cope with unseen color names and evaluated.

Weaknesses

Weakness argument 1: The limit where the proposed system works well to estimate the color distribution is not clearly described.

Weakness argument 2: The paper describes how a sequence of words is mapped to a probability distribution in HSV color-space. This is one directional and the other direction, that is, how to describe a probability distribution in HSV as a sequence of words, is not mentioned.

Weakness argument 3: The reason why the paper chooses HSV color-space is not clearly described.

Weakness argument 4: The paper mentions differences among individuals, but it does not evaluate the proposed system with respect to such differences.

Weakness argument 5: Embedding representation is not clearly described. It is only mentioned around line 440.

Questions to Authors (Optional)

Question 1: Under what assumptions, does the system work well?

Question 2: "Multimodal" is not defined well. The paper gives some examples such as 'greenish blue'. But what is the difference between multimodal and multi-token descriptions?

Question 3: How do you map a probability distribution in HSV color-space to a sequence of color description words? How do you manipulate personal difference?

Question 4: Figure 3 is confusing because it is a failed case. Anyway, the evaluation is not consistent. With respect to Figure 4 “purplish grey”, Figures 2 and 3 should use "grey" and "purplish" as a standalone color and a modifier, respectively. Of course, "red" is more preferable to "grey" in case. There remains a question: the reason why Figure 4 works well is due to a selection of monaural color "grey". Do the authors have another strong validation?

NLP Tasks / Applications	Strong contribution
Methods / Algorithms	Moderate contribution
Theoretical / Algorithmic Results	Moderate contribution
Empirical Results	Strong contribution
Data / Resources	Moderate contribution
Software / Systems	Moderate contribution
Evaluation Methods / Metrics	Moderate contribution
Other Contributions	N/A
Originality (1-5)	3
Soundness/Correctness (1-5)	3
Substance (1-5)	3
Replicability (1-5)	4
Meaningful Comparison (1-5)	3
Readability (1-5)	3
Overall Score (1-6)	2

Additional Comments (Optional)

Grammatical check should be carefully conducted.
(1) "the" is over used.
L043 the many modifiers, L423 the how the, L686 the more difficult (remove them) (2) some typos L347 is is, L593 is has, L755 the terms are used new descriptions, L798 .. (3) article is missing L018, 021 a color-space

Review #3

Appropriateness	Appropriate
Adhere to ACL 2018 Guidelines	Yes
Adhere to ACL Author Guidelines	No
Handling of Data / Resources	N/A
Handling of Human Participants	N/A

Comments on Preliminary Checks

After I've conducted the paper review and made all needed judgements/decisions for evaluation, I searched the internet to find if the paper adheres to the author guidelines, I found a non-anonomized arxiv version of the paper.

Summary and Contributions

Summary:

Contribution 1: generate a distribution over color regions given an input color textual description.

Strengths

Strength argument 1: The proposed technique generates a distribution over a discreteization of the color regions.

Strength argument 2: adding blurring to impose continuity in the output distribution.

Weaknesses

Weakness argument 1: Although the goal of generating an output distribution over color regions is well motivated from the statistical and practical point of view (as different people could sample different color variations differently for the same textual description), the paper lacked a proper system description for which an output distribution would be required as opposed to a point estimate of HSV channels for colors.

Weakness argument 2: The evaluation method using perplexity isn't related to the ground truth point estimate. Having lower pp doesn't necessarily mean better system given the described setup.

Weakness argument 3: Comparison to baseline point estimate systems is lacking.

NLP Tasks / Applications	Marginal contribution
Methods / Algorithms	Marginal contribution
Theoretical / Algorithmic Results	N/A
Empirical Results	Marginal contribution
Data / Resources	N/A
Software / Systems	N/A

Evaluation Methods / Metrics: N/A
Other Contributions: N/A
Originality (1-5): 2
Soundness/Correctness (1-5): 4
Substance (1-5): 3
Replicability (1-5): 4
Meaningful Comparison (1-5): 3
Readability (1-5): 5
Overall Score (1-6): 2

Author Response

ATTENTION: this time, we plan to do some analytics on anonymized reviews and rebuttal statements, upon the agreement of the reviewers and authors, with the purpose of improving the quality of reviews. The data will be compiled into a unique corpus, which we potentially envisage as a great resource for NLP, e.g. for sentiment analysis and argumentation mining, and made available to the community properly anonymized at earliest in 2 years. We hope to provide data on "how to review" to younger researchers, and improve transparency of the reviewing process in ACL in general.

By default, you agree that your anonymised rebuttal statement can be freely used for research purposes and published under an appropriate open-source license within at earliest 2 years from the acceptance deadline.

Select "No" if you would like to opt out of the data collection: Yes

Review Quality Survey

The quality of the reviews significantly influences the quality of the conference. To evaluate the quality of each review and reviewer, we invite the authors to rate the reviews they have received. Note that the survey results will only be presented to Programme Chairs and the corresponding Area Chairs, and will *NOT* be disclosed to the reviewers. Use the guidelines here to answer the questions for each review:

Quality of the Review

Do the reviews address the main strengths/weaknesses of the paper? Does the reviewer have a good understanding of the paper? Do the reviews include some insightful comments?

1. Nonsense: the reviews are mostly confusing and hard to understand
2. Below average: the reviews are largely about minor points, and there are significant misunderstandings of the paper
3. Mediocre: the reviews give some valuable comments, but also ignore/degrade some important strengths/potentials of this paper
4. Above average: the reviews point out some main strengths and weaknesses of the paper and provide good justifications
5. Insightful: the reviews not only grasp the main strengths/weaknesses of the paper, but also give convincing analyses and insightful comments

Helpfulness of the Review

How helpful are the reviews? Can you use the reviews to improve the paper? Do the reviews give some insightful comments that can be helpful to your future research?

1. Poor: the reviews make not much sense and are largely useless
2. Somewhat helpful: the reviews can help me to do some minor changes, but no major improvements can be made
3. Helpful: the reviews give an objective and fairly comprehensive evaluation of the paper, and can help me marginally improve the quality of the paper
4. Very helpful: with the help of these reviews, some major weaknesses of this paper can be strengthened and some strengths can be reinforced, thus the quality of the paper can be significantly improved
5. Out of my expectation: the reviews not only give important comments on how to improve this paper, but also give some insightful analyses and advice that can be helpful to my ongoing research and future works

Submit Response to Reviewers

Use the following boxes to enter your response to the reviews. Please limit the total amount of words in your comments to 1000 words (longer responses will not be accepted by the system).

Response to Review #1:

```
Reply to weakness argument 1:

Reply to weakness argument 2:

Reply to weakness argument 3:

Reply to weakness argument 4:

Reply to weakness argument 5:

Reply to question 1:

Reply to question 2:

Reply to question 3:
```

Quality of Review #1: 1
Helpfulness of Review #1: 1

Response to Review #2:

```
Reply to weakness argument 1:

Reply to weakness argument 2:

Reply to weakness argument 3:

Reply to weakness argument 4:

Reply to weakness argument 5:

Reply to question 1:

Reply to question 2:

Reply to question 3:
```

Quality of Review #2: 1

Helpfulness of Review #2: 1

Response to Review #3:

```
Reply to weakness argument 1:

Reply to weakness argument 2:

Reply to weakness argument 3:

Reply to weakness argument 4:

Reply to weakness argument 5:

Reply to question 1:

Reply to question 2:

Reply to question 3:
```

Quality of Review #3: 1

Helpfulness of Review #3: 1

Response to Chairs

Use this textbox to contact the area chairs directly only when there are serious issues regarding the reviews. Such issues can include reviewers who grossly misunderstood the submission, or have made unfair comparisons or requests in their reviews. Most submissions should not need to use this facility.

Submit
