

Dataset Annotation Guideline

Weizhe Yuan

1 Aspect Typology

We define a typology that contains 8 aspects, which are **Summary**, **Motivation/Impact**, **Originality**, **Soundness/Correctness**, **Substance**, **Replicability**, **Meaningful Comparison** and **Clarity**. The detailed explanations for each aspect is shown below.

- **Summary:** What was done in the paper?

Example:

1. *The paper proposes a new memory access scheme based on Lie group actions for NTMs.*

- **Motivation/Impact:** Does the paper address an important problem? Are other people (practitioners or researchers) likely to use these ideas or build on them?

Example:

1. *The issue researched in this work is of significance because understanding the predictive uncertainty of a deep learning model has its both theoretical and practical value.*
2. *The method seems limited in both practical usefulness and enlightenment to the reader.*

- **Originality:** Are there new research topic, technique, methodology, or insight?

Example:

1. *Novel addressing scheme as an extension to NTM.*
2. *The reviewer believes that the idea of the paper is similar to the one in [1].*

- **Soundness/Correctness:** Is the proposed approach sound? Are the claims in the paper convincingly supported?

Example:

1. *Illustrations using simulated data and real data are also very clear and convincing.*
 2. *The proposed method is sensible and technically sound.*
 3. *The experiments are also quite convincing.*
 4. *The required condition is rather implicit, and it is unclear how this condition can be checked in practice.*
 5. *There is not much theory to support the method.*
 6. *Several model designs are not well justified.*
 7. *There is not enough justification to demonstrate improvements.*
- **Substance:** Does the paper contains substantial experiments to demonstrate the effectiveness of proposed methods? Are there detailed result analysis? Does it contain meaningful ablation studies?

Example:

1. *This is a thorough exploration of a mostly under-studied problem.*
 2. *The experiment section shows extensive experiment.*
 3. *There are several modules introduced in the paper, but there isn't much analysis of them during the experiments.*
 4. *This experimental study does not seem to conduct sufficient experiments to demonstrate the advantages.*
 5. *Lack detailed and insightful ablation studies.*
 6. *I would expect the authors to conduct some more analysis of their results besides acc. and distortion levels.*
- **Replicability:** Is it easy to reproduce the results and verify the correctness of the results?
Is the supporting dataset and/or software provided?

Example:

1. *Release of the dataset and code should help with reproducibility.*
 2. *There are some technical ambiguities.*
- **Meaningful Comparison:** Are the comparisons to prior work sufficient given the space constraints? Are the comparisons fair?

Example:

1. *The authors do a good job of positioning their study with respect to related work on black-box adversarial techniques.*
 2. *The comparison with the Caron-Fox approach is very good and useful for the reader.*
 3. *The experimental study can have more comparison on challenging datasets with more classes.*
 4. *Since the attention based aggregation is similar to GAT, a discussion on the difference is important.*
 5. *The paper fails to locate itself in the literature, how it compares itself into other techniques (both analytically and experimentally).*
 6. *The comparison does not seem fair.*
- **Clarity:** For a reasonably well-prepared reader, is it clear what was done and why? Is the paper well-written and well-structured?

Example:

1. *The paper is well-written and easy to follow.*
2. *The presentation of the results is not very clear.*

2 Annotation Tips

We further decompose each aspect (except **Summary**) into a positive one and a negative one. For example, **Motivation** will be decomposed to **Positive Motivation** and **Negative Motivation**. As so, there are in total 15 aspects for you to choose from when annotating reviews. Below are some tips.

- Please annotate the shortest while complete span that indicates a specific aspect. Don't include specific details if the aspect has been stated clearly prior to those details.

Example:

This experimental study does not seem to conduct sufficient experiments to demonstrate the advantages [Negative Substance] *(say, in terms of training efficiency the capability in making the network scalable for more challenging dataset) of the proposed objective function over the existing one.*

- Please be as fine-grained as possible. If a sentence contains multiple aspects, annotate them separately if they can be disentangled.

Example:

The results are new[Positive Originality] and important to this field.[Pos. Motiv.]