

# Evaluating Representation Learning with Reference Games

Anonymous ACL submission

## Abstract

TODO

## A Appendices

| Model          | 100% | 50% | 25% | 10% | 5% | 1% |
|----------------|------|-----|-----|-----|----|----|
| RNN            |      |     |     |     |    |    |
| GloVe          |      |     |     |     |    |    |
| Word2Vec       |      |     |     |     |    |    |
| SkipThought    |      |     |     |     |    |    |
| InferSent      |      |     |     |     |    |    |
| BERT           |      |     |     |     |    |    |
| GPT (OpenAI)   |      |     |     |     |    |    |
| GPT2           |      |     |     |     |    |    |
| CTRL           |      |     |     |     |    |    |
| Transformer-XL |      |     |     |     |    |    |
| XLNet          |      |     |     |     |    |    |
| XLM            |      |     |     |     |    |    |
| DistilBERT     |      |     |     |     |    |    |
| RoBERTa        |      |     |     |     |    |    |

Table 1: Evaluation of several language representations on the Colors in Context dataset.

| Model          | 100% | 50% | 25% | 10% |
|----------------|------|-----|-----|-----|
| RNN            |      |     |     |     |
| GloVe          |      |     |     |     |
| Word2Vec       |      |     |     |     |
| SkipThought    |      |     |     |     |
| InferSent      |      |     |     |     |
| BERT           |      |     |     |     |
| GPT (OpenAI)   |      |     |     |     |
| GPT2           |      |     |     |     |
| CTRL           |      |     |     |     |
| Transformer-XL |      |     |     |     |
| XLNet          |      |     |     |     |
| XLM            |      |     |     |     |
| DistilBERT     |      |     |     |     |
| RoBERTa        |      |     |     |     |

Table 2: Evaluation of several language representations on the ColorGrids in Context dataset.

| Model    | Text           | 100% | 50% | 25% | 10% | 5% | 1% |
|----------|----------------|------|-----|-----|-----|----|----|
| Vanilla  | RNN            |      |     |     |     |    |    |
| Vanilla  | GloVe          |      |     |     |     |    |    |
| Vanilla  | Word2Vec       |      |     |     |     |    |    |
| Vanilla  | SkipThought    |      |     |     |     |    |    |
| Vanilla  | InferSent      |      |     |     |     |    |    |
| Vanilla  | BERT           |      |     |     |     |    |    |
| Vanilla  | GPT (OpenAI)   |      |     |     |     |    |    |
| Vanilla  | GPT2           |      |     |     |     |    |    |
| Vanilla  | CTRL           |      |     |     |     |    |    |
| Vanilla  | Transformer-XL |      |     |     |     |    |    |
| Vanilla  | XLNet          |      |     |     |     |    |    |
| Vanilla  | XLM            |      |     |     |     |    |    |
| Vanilla  | DistilBERT     |      |     |     |     |    |    |
| Vanilla  | RoBERTa        |      |     |     |     |    |    |
| VGG19    | RNN            |      |     |     |     |    |    |
| VGG19    | GloVe          |      |     |     |     |    |    |
| VGG19    | Word2Vec       |      |     |     |     |    |    |
| VGG19    | SkipThought    |      |     |     |     |    |    |
| VGG19    | InferSent      |      |     |     |     |    |    |
| VGG19    | BERT           |      |     |     |     |    |    |
| VGG19    | GPT (OpenAI)   |      |     |     |     |    |    |
| VGG19    | GPT2           |      |     |     |     |    |    |
| VGG19    | CTRL           |      |     |     |     |    |    |
| VGG19    | Transformer-XL |      |     |     |     |    |    |
| VGG19    | XLNet          |      |     |     |     |    |    |
| VGG19    | XLM            |      |     |     |     |    |    |
| VGG19    | DistilBERT     |      |     |     |     |    |    |
| VGG19    | RoBERTa        |      |     |     |     |    |    |
| ResNet34 | RNN            |      |     |     |     |    |    |
| ResNet34 | GloVe          |      |     |     |     |    |    |
| ResNet34 | Word2Vec       |      |     |     |     |    |    |
| ResNet34 | SkipThought    |      |     |     |     |    |    |
| ResNet34 | InferSent      |      |     |     |     |    |    |
| ResNet34 | BERT           |      |     |     |     |    |    |
| ResNet34 | GPT (OpenAI)   |      |     |     |     |    |    |
| ResNet34 | GPT2           |      |     |     |     |    |    |
| ResNet34 | CTRL           |      |     |     |     |    |    |
| ResNet34 | Transformer-XL |      |     |     |     |    |    |
| ResNet34 | XLNet          |      |     |     |     |    |    |
| ResNet34 | XLM            |      |     |     |     |    |    |
| ResNet34 | DistilBERT     |      |     |     |     |    |    |
| ResNet34 | RoBERTa        |      |     |     |     |    |    |

Table 3: Evaluation (Part 1 of 2) of several multimodal representations on the Chairs in Context dataset.

| Model         | Text           | 100% | 50% | 25% | 10% | 5% | 1% |
|---------------|----------------|------|-----|-----|-----|----|----|
| IR (ImageNet) | RNN            |      |     |     |     |    |    |
| IR (ImageNet) | GloVe          |      |     |     |     |    |    |
| IR (ImageNet) | Word2Vec       |      |     |     |     |    |    |
| IR (ImageNet) | SkipThought    |      |     |     |     |    |    |
| IR (ImageNet) | InferSent      |      |     |     |     |    |    |
| IR (ImageNet) | BERT           |      |     |     |     |    |    |
| IR (ImageNet) | GPT (OpenAI)   |      |     |     |     |    |    |
| IR (ImageNet) | GPT2           |      |     |     |     |    |    |
| IR (ImageNet) | CTRL           |      |     |     |     |    |    |
| IR (ImageNet) | Transformer-XL |      |     |     |     |    |    |
| IR (ImageNet) | XLNet          |      |     |     |     |    |    |
| IR (ImageNet) | XLM            |      |     |     |     |    |    |
| IR (ImageNet) | DistilBERT     |      |     |     |     |    |    |
| IR (ImageNet) | RoBERTa        |      |     |     |     |    |    |
| LA (ImageNet) | RNN            |      |     |     |     |    |    |
| LA (ImageNet) | GloVe          |      |     |     |     |    |    |
| LA (ImageNet) | Word2Vec       |      |     |     |     |    |    |
| LA (ImageNet) | SkipThought    |      |     |     |     |    |    |
| LA (ImageNet) | InferSent      |      |     |     |     |    |    |
| LA (ImageNet) | BERT           |      |     |     |     |    |    |
| LA (ImageNet) | GPT (OpenAI)   |      |     |     |     |    |    |
| LA (ImageNet) | GPT2           |      |     |     |     |    |    |
| LA (ImageNet) | CTRL           |      |     |     |     |    |    |
| LA (ImageNet) | Transformer-XL |      |     |     |     |    |    |
| LA (ImageNet) | XLNet          |      |     |     |     |    |    |
| LA (ImageNet) | XLM            |      |     |     |     |    |    |
| LA (ImageNet) | DistilBERT     |      |     |     |     |    |    |
| LA (ImageNet) | RoBERTa        |      |     |     |     |    |    |
| VAE (COCO)    | RNN            |      |     |     |     |    |    |
| VAE (COCO)    | GloVe          |      |     |     |     |    |    |
| VAE (COCO)    | Word2Vec       |      |     |     |     |    |    |
| VAE (COCO)    | SkipThought    |      |     |     |     |    |    |
| VAE (COCO)    | InferSent      |      |     |     |     |    |    |
| VAE (COCO)    | BERT           |      |     |     |     |    |    |
| VAE (COCO)    | GPT (OpenAI)   |      |     |     |     |    |    |
| VAE (COCO)    | GPT2           |      |     |     |     |    |    |
| VAE (COCO)    | CTRL           |      |     |     |     |    |    |
| VAE (COCO)    | Transformer-XL |      |     |     |     |    |    |
| VAE (COCO)    | XLNet          |      |     |     |     |    |    |
| VAE (COCO)    | XLM            |      |     |     |     |    |    |
| VAE (COCO)    | DistilBERT     |      |     |     |     |    |    |
| VAE (COCO)    | RoBERTa        |      |     |     |     |    |    |
| IR (COCO)     | IR (COCO)      |      |     |     |     |    |    |
| VAEVAE (COCO) | VAEVAE (COCO)  |      |     |     |     |    |    |
| VAEGAN (COCO) | VAEGAN (COCO)  |      |     |     |     |    |    |

Table 4: Evaluation (Part 2 of 2) of several multimodal representations on the Chairs in Context dataset.