UNSUPERVISED STYLE EXTRACTION

FROM SENTENCE EMBEDDINGS

Kim et al. 2023

Contents

Unsupervised Style Extraction

Data & Model

Experiments

Discussion & Comments

References

What is Style?

Given two corpora, the invariance between them is the content,

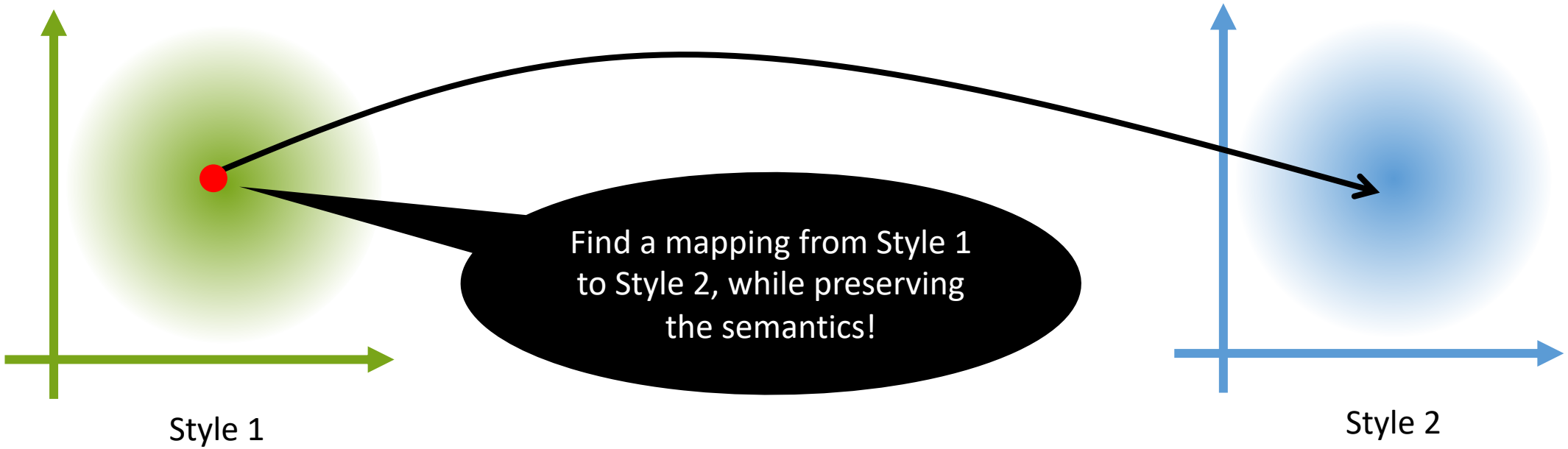
whereas the variance is the style.

Style Attribute	*Source Attribute / Sentence*	*Target Attribute / Sentence*
Politeness	Polite: *"**Could you please** send me the data**?**"*	Impolite: *"send me the data**!!**"*
Toxicity	Offensive: *"I hope they pay out the *******."*	Non-offensive: *"I hope they pay **what they deserve**."*
Simplicity	Expert: "*Many cause **dyspnea**, pleuritic chest pain, or both."*	Layman: *"The most common symptoms, regardless of the type of fluid in the pleural space or its cause, are **shortness of breath** and chest pain."*
Biasedness	Biased: *"A new downtown is being developed **which will bring back**..."*	Neutral: *"A new downtown is being developed **which its promoters hope** will bring back..."*
Authorship	Shakespearean: *"My lord, the queen **would** speak with you**, and presently.**"*	Contemporary: *"My lord, the queen **wants to** speak with you **right away.**"*

https://blog.fastforwardlabs.com/2022/03/22/an-introduction-to-text-style-transfer.html

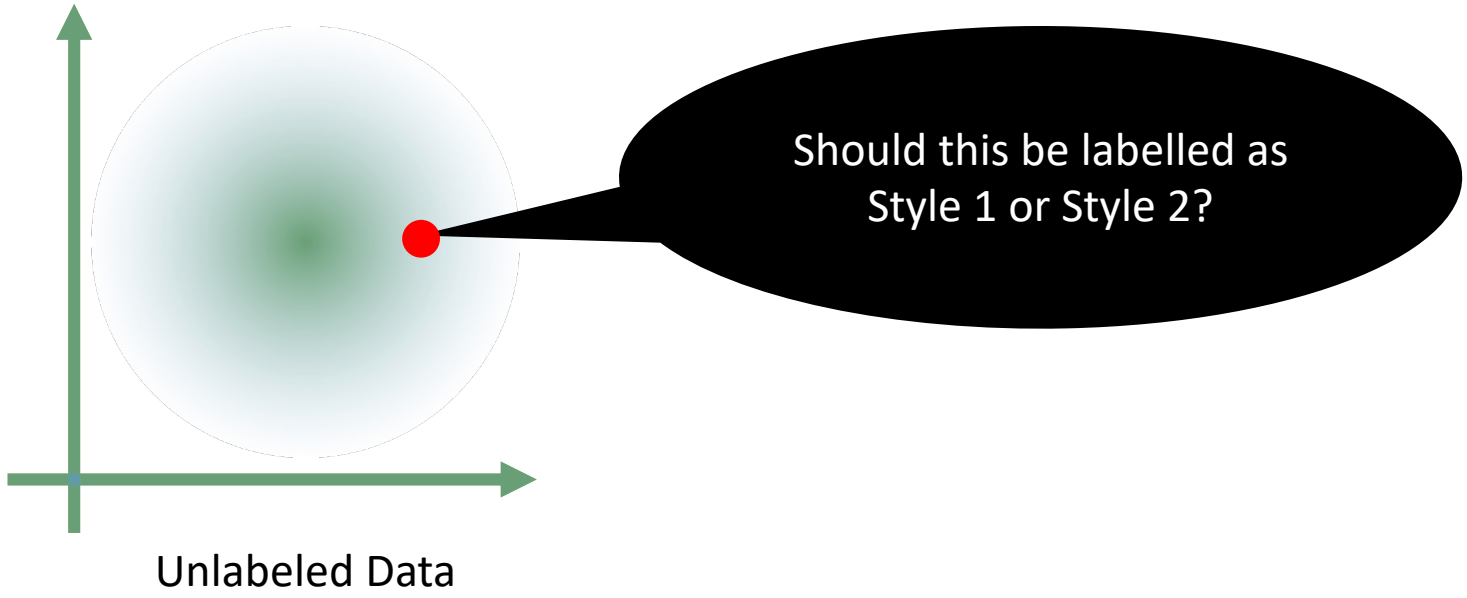
Text Style Transfer

Previous work on text style transfer deals with labelled data.



Unsupervised Style Extraction

But what if the data is not labelled with respective styles?



Data

Data: Kaggle Bible Corpus (Genesis)

Invariant content, variant style!

Genesis 1:1

⋮

1533 sentences

ASV: *In the beginning God created the heavens and the earth.*

BBE: *At the first God made the heaven and the earth.*

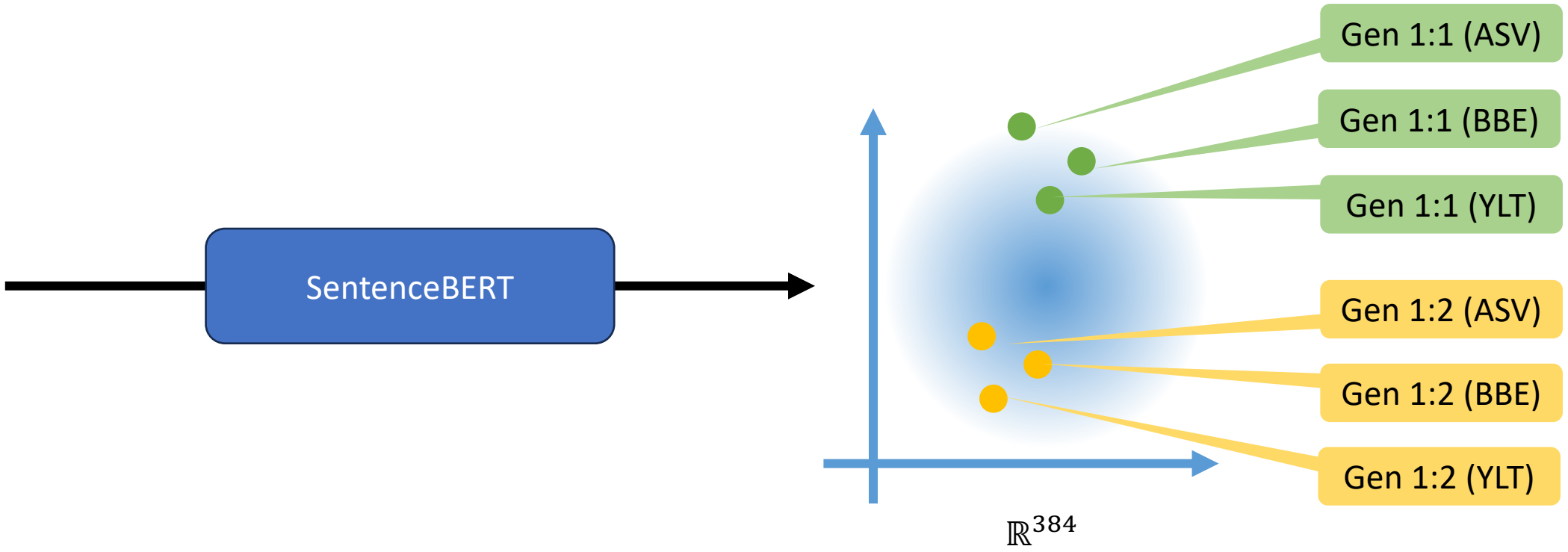
YLT: *In the beginning of God's preparing the heavens and the earth--*

⋮

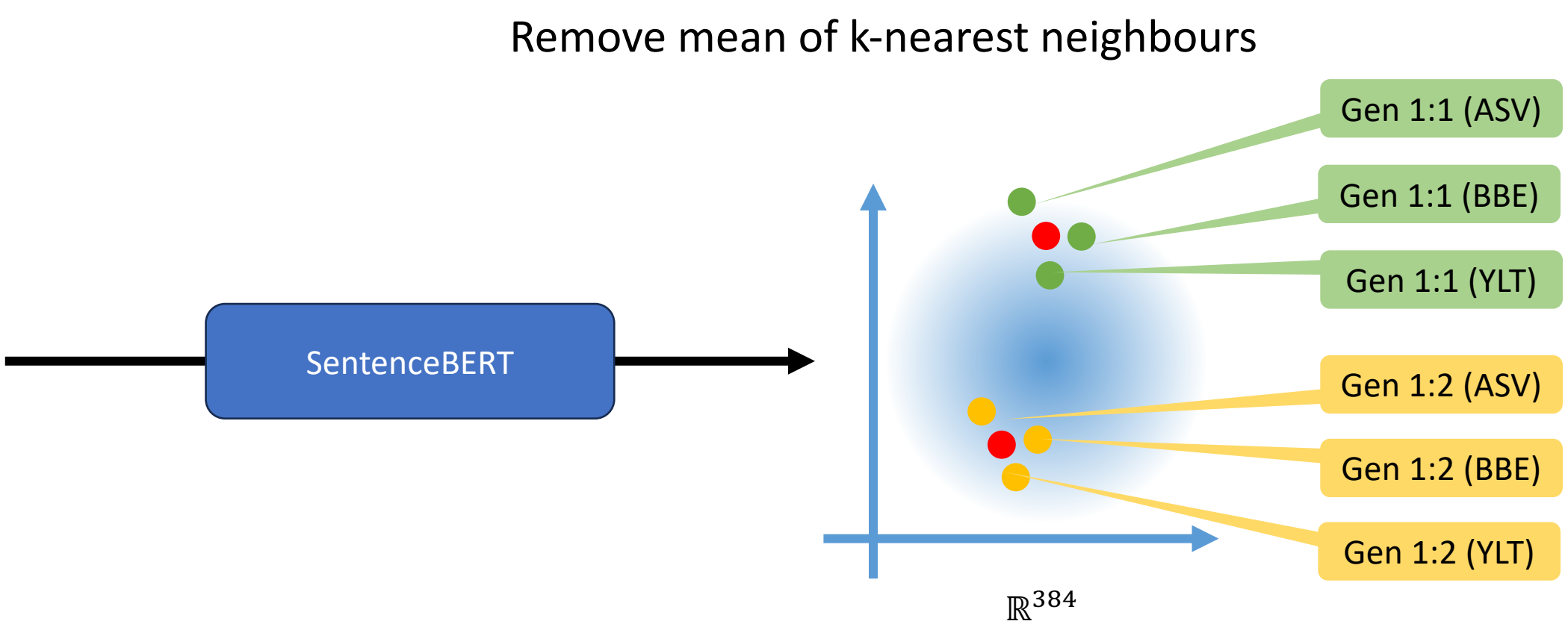
7 versions

https://www.kaggle.com/datasets/oswinrh/bible

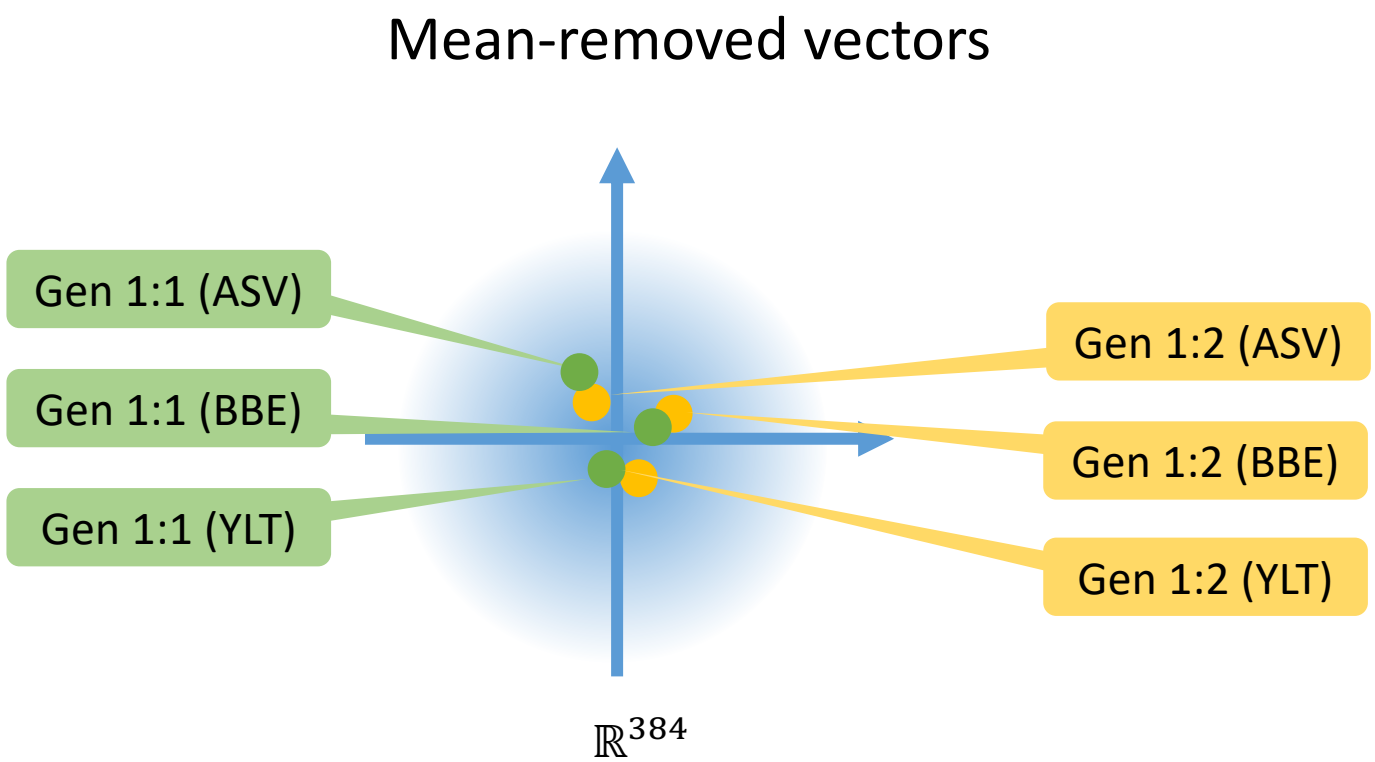
The Model: Sentence Embedding



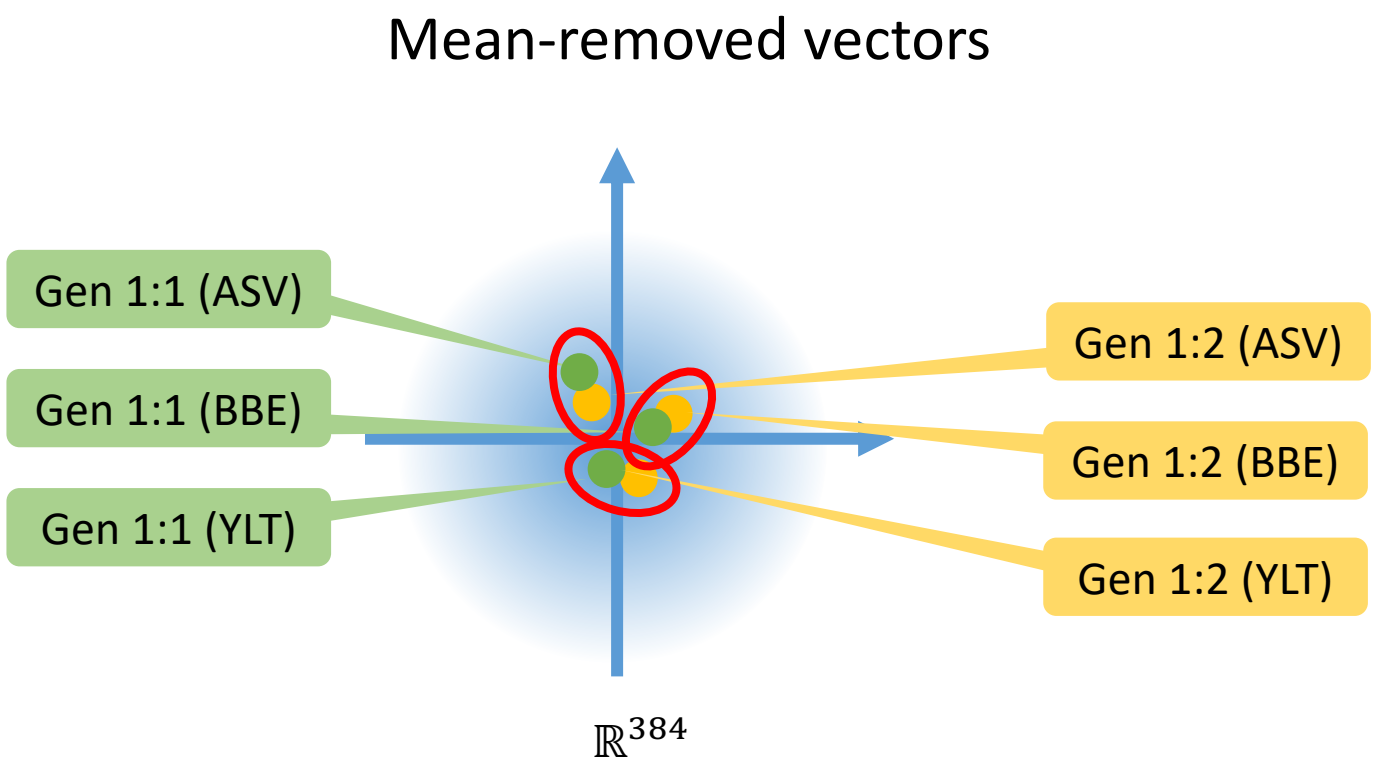
The Model: Sentence Embedding



The Model: Sentence Embedding

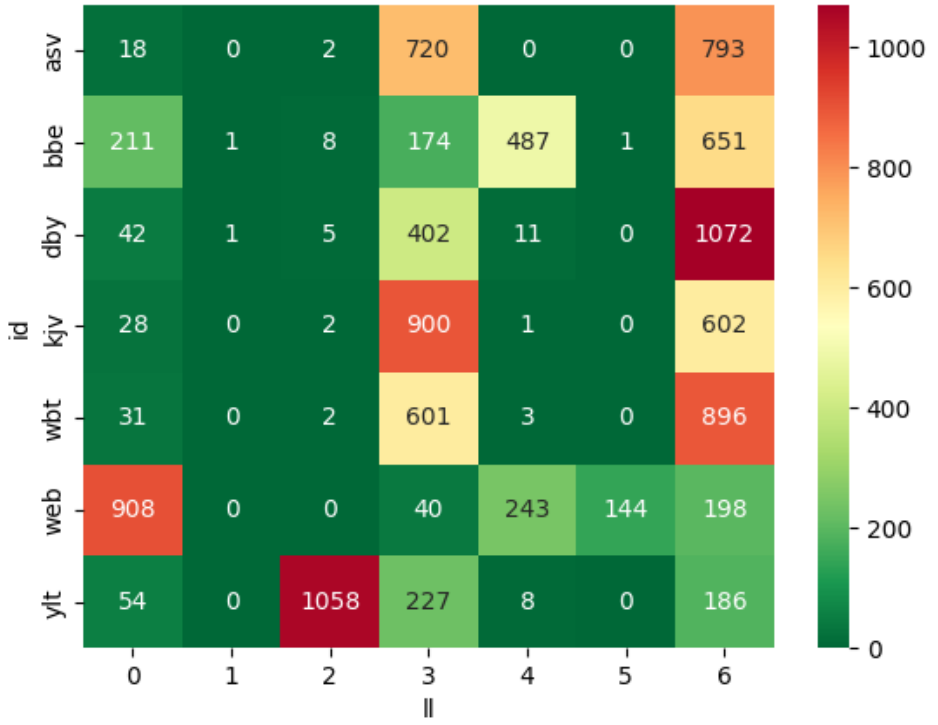


The Model: Sentence Embedding



The Model: Baseline

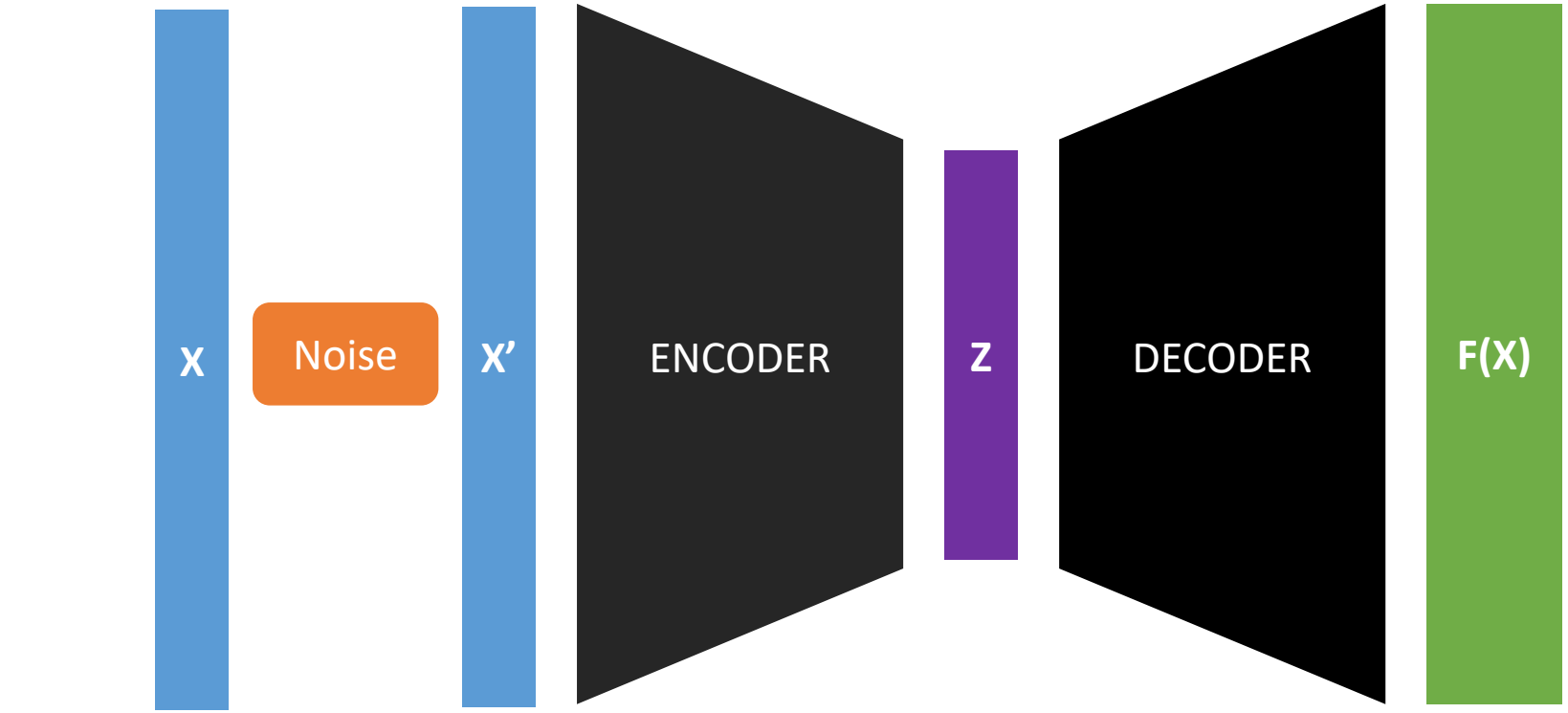
Baseline: Kmeans clustering using only the mean-removed vectors



V-Measure	Macro-Avg Acc.	Micro-Avg Acc.
0.312	0.623	0.426

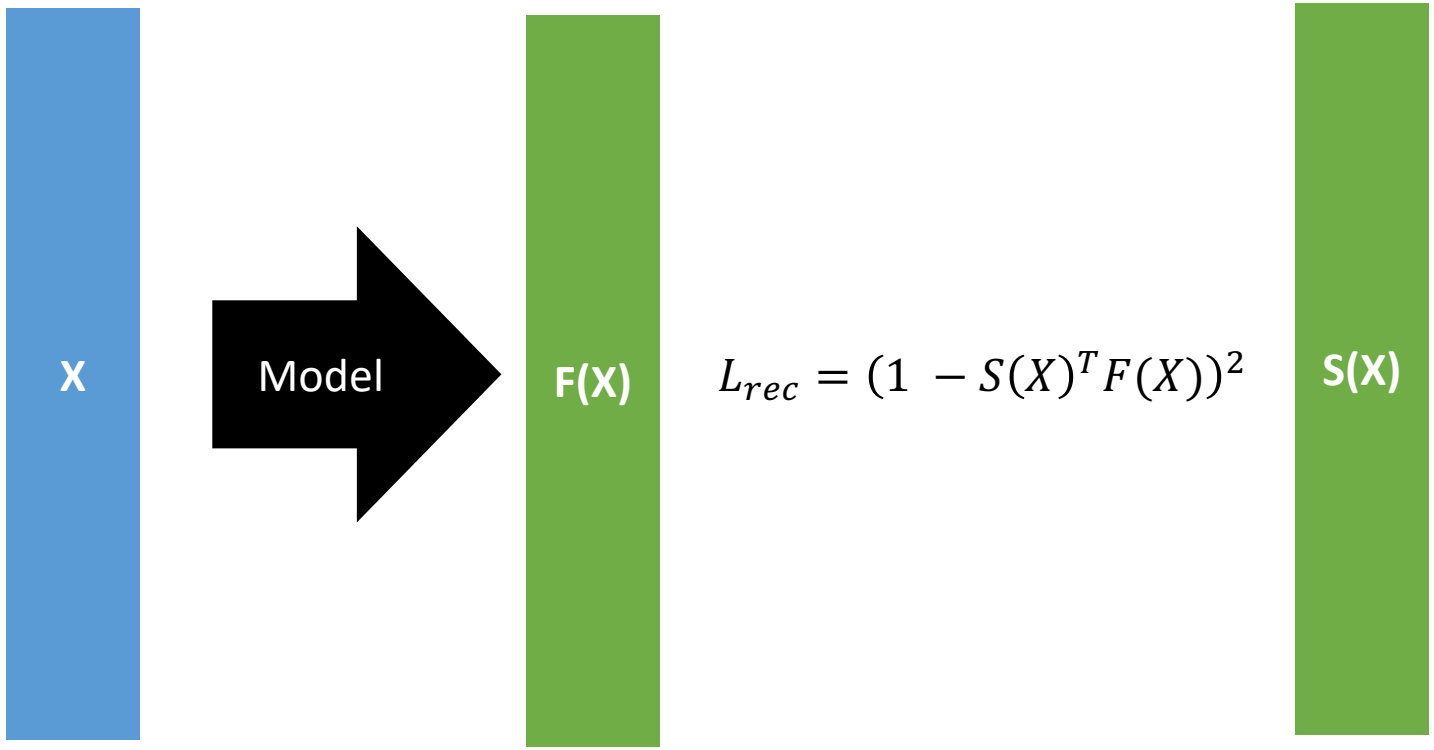
The Model: Denoising AutoEncoder

Model: an Autoencoder to capture the latent style representations



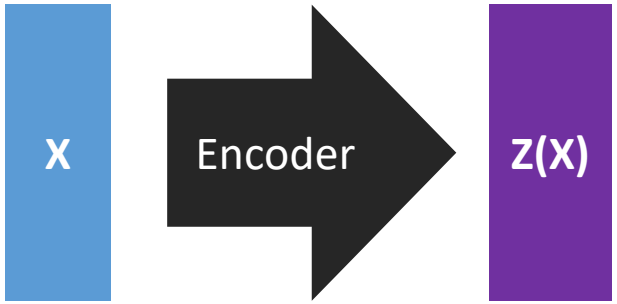
The Model: Loss Function

Reconstruction Loss: F(X) should be similar to mean-removed embedding

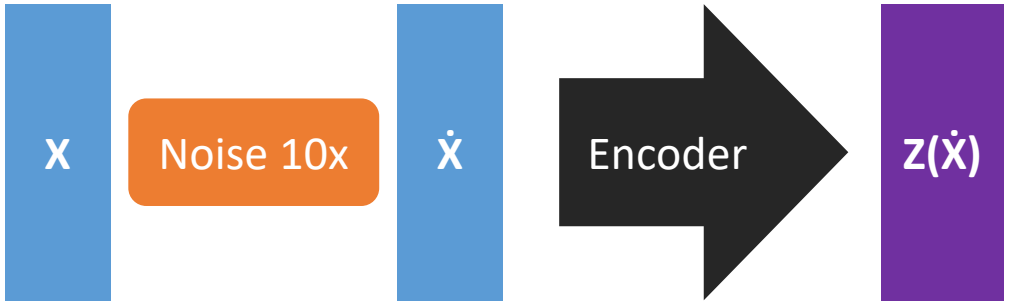


The Model: Loss Function

Positive Example Loss: Z should be simliar to noise-perturbed Z

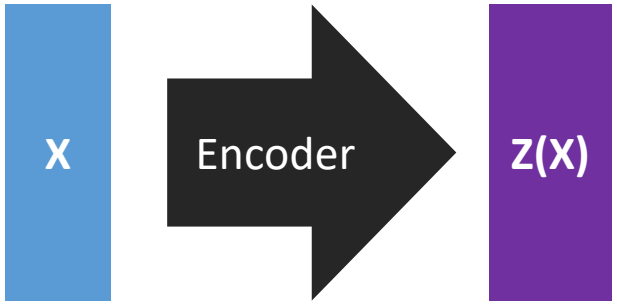


$$L_+ = \left(1 - Z(X)^T Z(\hat{X})\right)^2$$

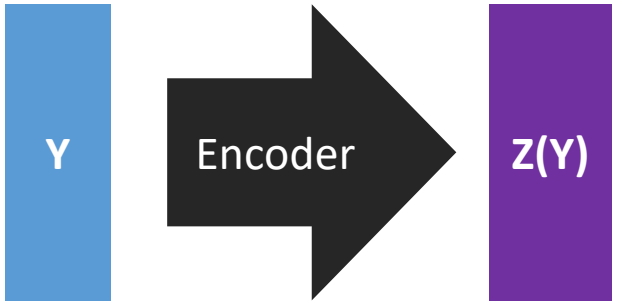


The Model: Loss Function

Negative Example Loss: Sentences different with X should have a different Z

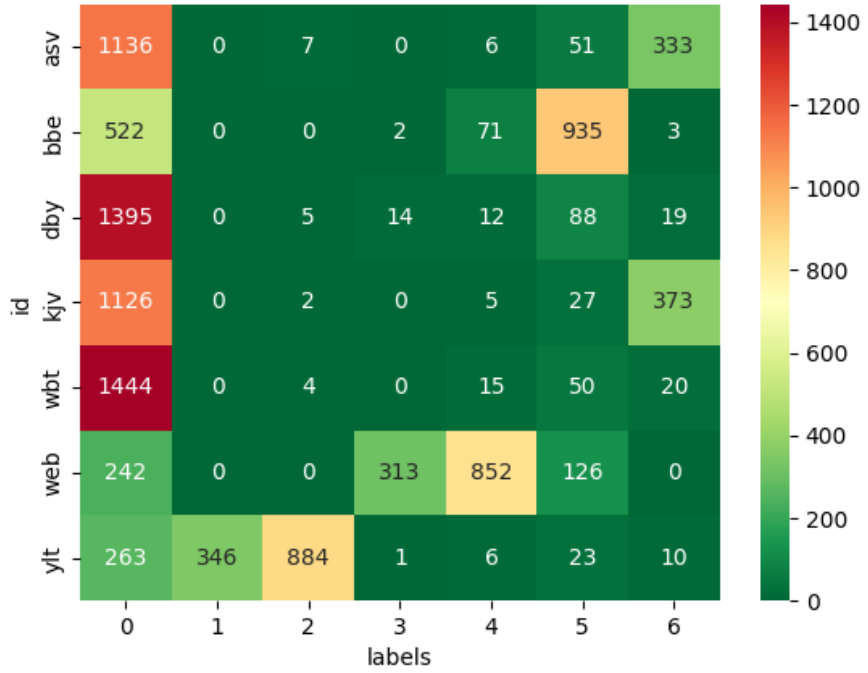


$$L_{-} = (Z(X)^{T} Z(Y))^{2}$$



The Model: Results

Baseline: Kmeans clustering using only the mean-removed vectors



V-Measure	Macro-Avg Acc.	Micro-Avg Acc.
0.402	0.751	0.480

Experiments

Hyperparameter Settings

Epoch	Batch	λ_{rec}	λ_{+}	λ_{-}	ϵ_{in}	ϵ_{+}	d(Z)	L Rate
50	73	1	1	1/6	5e-3	5e-2	64	1e-3

Experiments

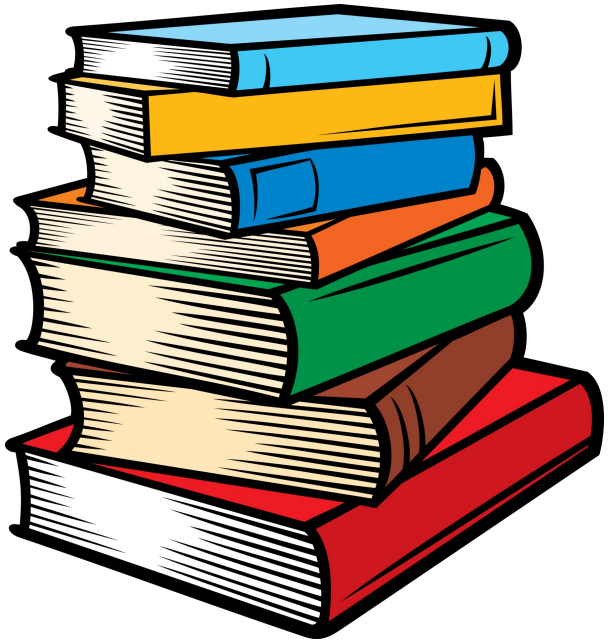
Ablation Study

Reconstruction	Positive	Negative	V-measure
X	O	O	0.310
O	X	O	0.276
O	O	X	0.130

The loss on the negative examples is very important!

Discussion

Testing the Model on Different Data



- Legal Data
- Scientific Papers
- Literary Corpus

Discussion

Extracting More Data from Sentences

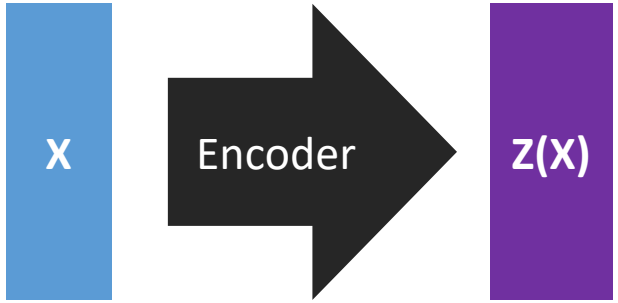
Fine-tune

SentenceBERT

to extract relevant information about style

Discussion

Disentanglement



Conclusion

- Our experiments demonstrate that our approach is able to learn style from sentences without supervision.
- Sentence embeddings with KNN mean-removal was used as the labels.
- An autoencoder was used, and reconstruction loss with positive-negative example loss was employed.
- The results prove significantly better than the baseline.

References

Jin et al., Deep Learning for Text Style Transfer: a Survey https://arxiv.org/pdf/2011.00416.pdf

Toshevska et al., A Review of Text Style Transfer using Deep Learning https://arxiv.org/abs/2109.15144

Reimers et al., Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks https://arxiv.org/abs/1908.10084

Carlson et al., Evaluating prose style transfer with the Bible https://arxiv.org/abs/1711.04731

Hoffer et al., Deep Metric Learning using Triplet network https://arxiv.org/abs/1412.6622

Lample et al., Unsupervised Machine Translation Using Monolingual Corpora Only https://arxiv.org/abs/1711.00043

Our source code: https://github.com/etharthinas/styledetection

Thanks!