Comparison of Autoencoding Techniques in Few-Shot Learning for Text Classification

Bartosz Chrostowski, Bogdan Jastrzębski, Jakub Drak Sbahi

June, 2022

Content

1. Objective of the project
2. Datasets
3. Architecture
4. Results
5. Deployment

Objective of the project

Comparison of autoencoding techniques for few-shot learning, in particular Variational Autoencoder (VAE) and Denoising Autoencoder (DAE).

Datasets

- AG News - articles belonging to 4 classes
- IMDB - movies reviews categorised as positive or negative

Architecture (1)

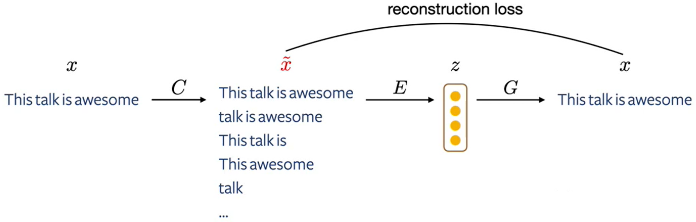
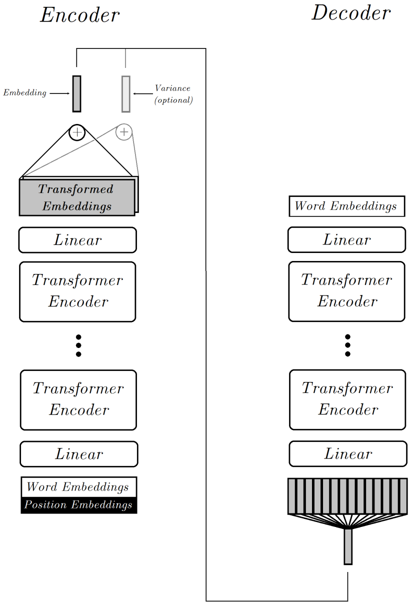


Figure: DAE architecture proposal. The denoising autoencoder augments observations and restores the original ones. Additionally, it could be improved using additional disentanglement loss. Architecture graph based on: [14].

Architecture (2)



Synthetic Example (3)

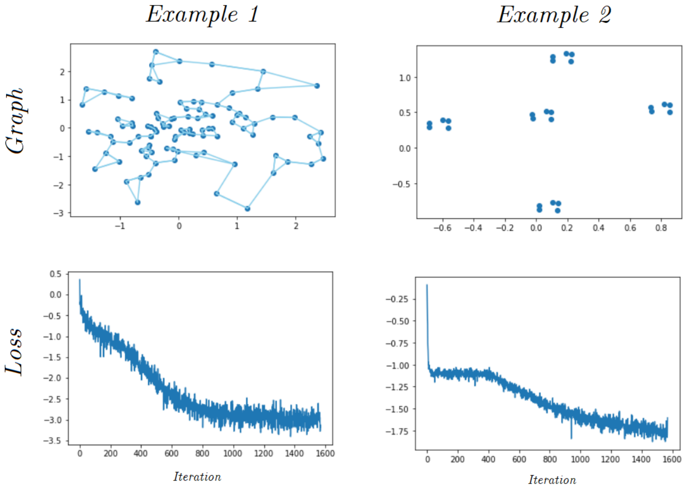


Figure: Random Walks Embeddings.

Synthetic Example (2)

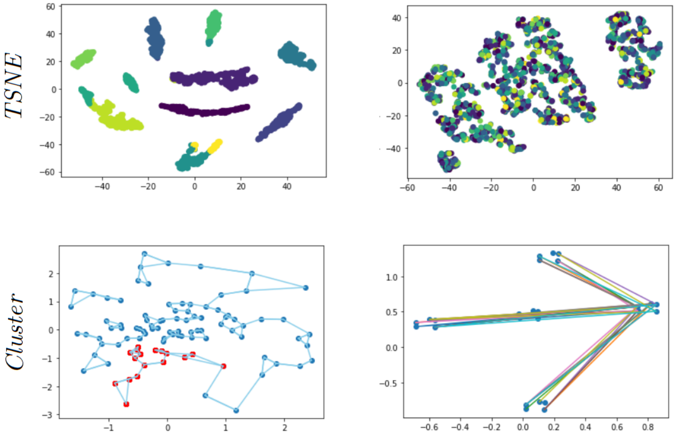


Figure: Random Walks Embeddings.

Few-shot Learning

- SVM
- Random Forest

Embedding Comparison (1)

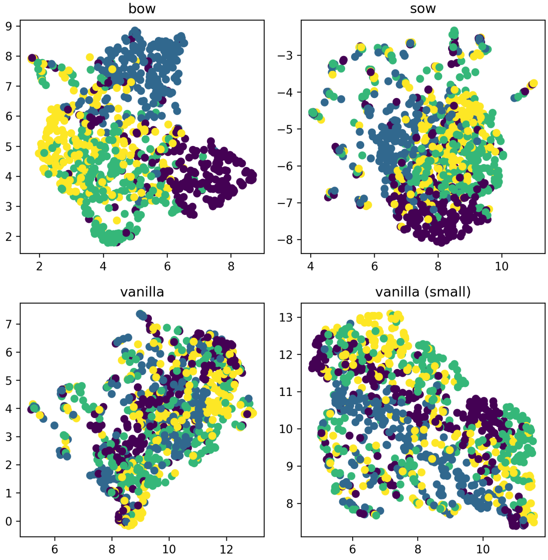


Figure: Comparison of embeddings for the AG-News dataset.

Embedding Comparison (2)

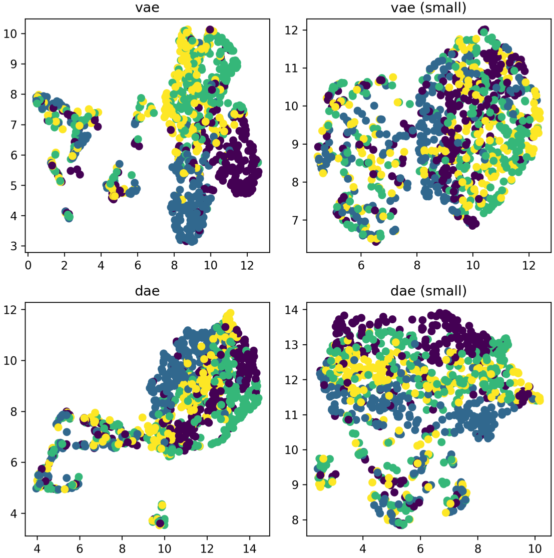


Figure: Comparison of embeddings for the AG-News dataset.

Embedding Comparison (3)

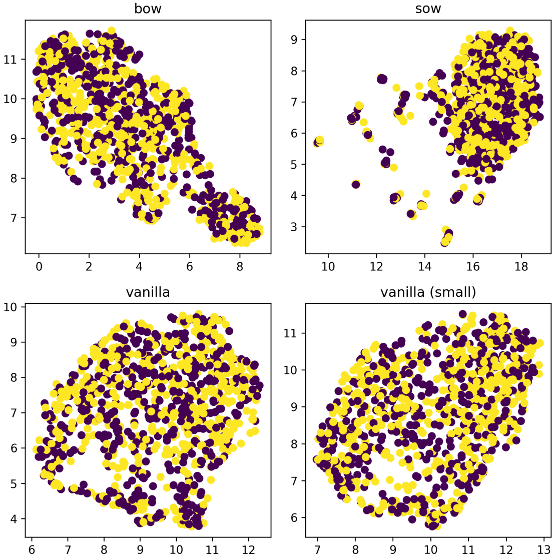


Figure: Comparison of embeddings for the IMDB-News dataset

Embedding Comparison (4)

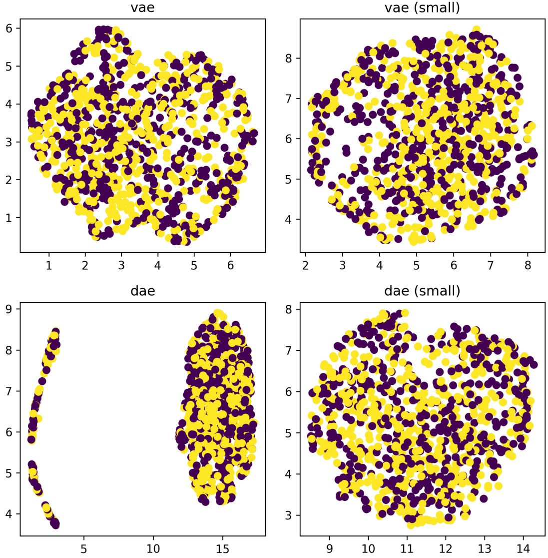
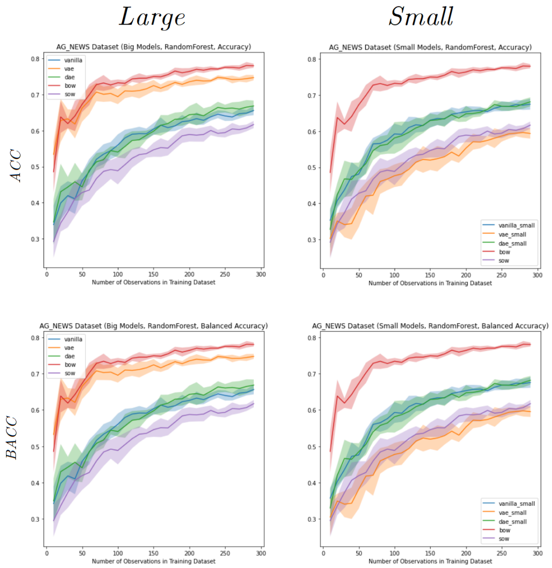
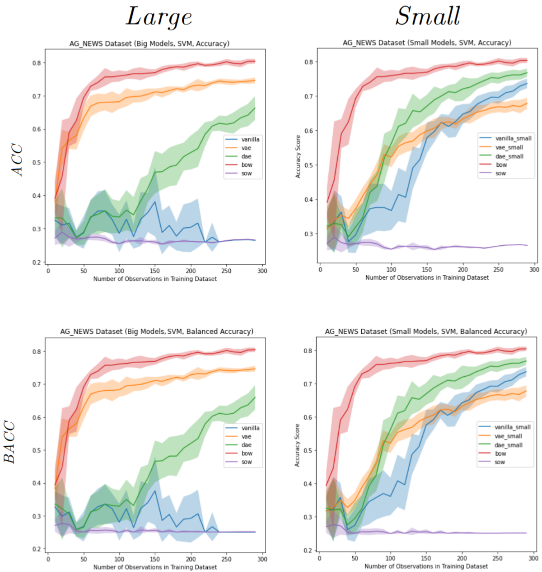


Figure: Comparison of embeddings for the IMDB-News dataset

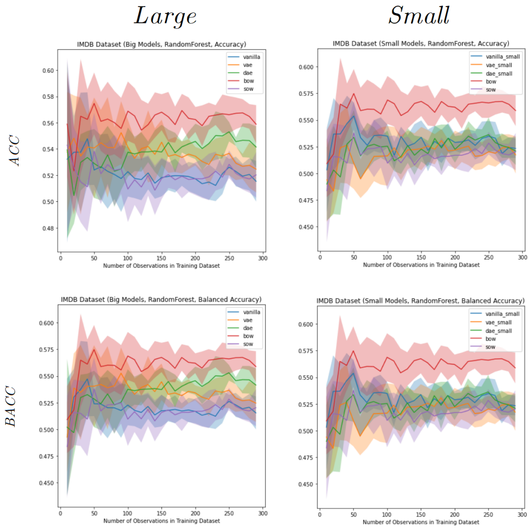
Few-shot Results (AG-NEWS, Random Forest)



Few-shot Results (AG-NEWS, SVM)



Few-shot Results (IMDB, Random Forest)

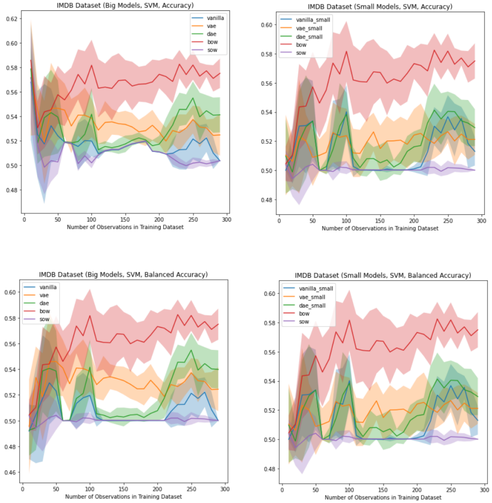


Few-shot Results (IMDB, SVM)

Large *Small*

ACC

BACC



Deployment

- We have deployed our model on managed Kubernetes cluster - AKS
- To achieve so we have written simple API in flask which loads prepared models on startup and serves predictions on request
- Then we have prepared Dockerfile which prepares container by installing all needed dependencies and is configured to serve our API with Waitress WSGI on startup and prepared manifest files for Kubernetes deployment
- We have deployed an infrastructure consisting of Active Directory, Azure Kubernetes Srevice and Azure Container Registry on Azure using free credit
- To automize the deployments we have prepared CD pipeline on Github Actions. On commit container is automatically built, pushed to ACR and then the app is deployed on AKS

Deployment

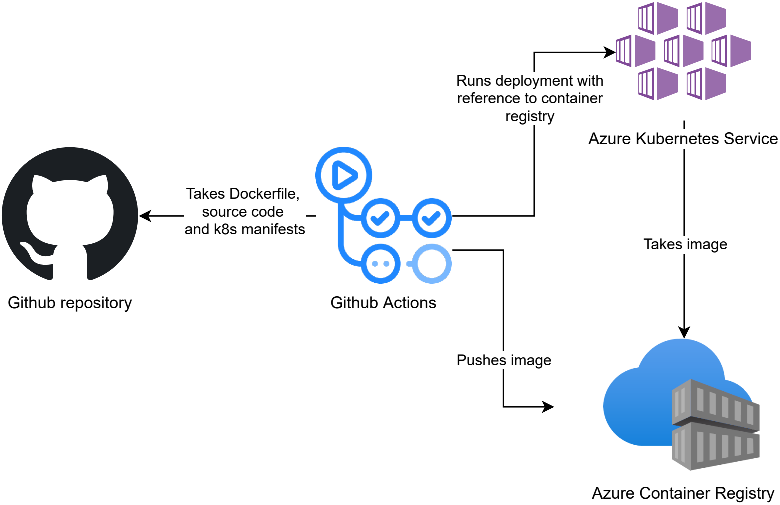


Figure: Deployment architecture. The diagram shows all needed cloud components needed for deployment automatization. The process uses Github and Azure resources.

Summary

Key takeaways from the project:

- the entanglement of features can decrease performance for few-shot learning, hence without a large corpus of data, it might be difficult to improve performance training an autoencoder;
- the VAE can be better than other autoencoders, if the VAE assumptions are valid, because it has a capability to disentangle (to some degree) the representation.
- our proposed DAE performed (in general) worse than VAE, but achieved better results than the Vanilla autoencoder. It is possible, that it too (to some degree) disentangled embedding features.

Questions

References I

[1] Minmin Chen. "Efficient Vector Representation for Documents through Corruption". In: *CoRR* abs/1707.02377 (2017). arXiv: 1707.02377. URL: http://arxiv.org/abs/1707.02377.

[2] Pengcheng He et al. "DeBERTa: Decoding-enhanced BERT with Disentangled Attention". In: *CoRR* abs/2006.03654 (2020). arXiv: 2006.03654. URL: https://arxiv.org/abs/2006.03654.

[3] Irina Higgins et al. "beta-VAE: Learning Basic Visual Concepts with a Constrained Variational Framework". In: *ICLR*. 2017.

[4] Diederik P Kingma and Max Welling. *Auto-Encoding Variational Bayes*. 2014. arXiv: 1312.6114 [stat.ML].

References II

[5] Ryan Kiros et al. "Skip-Thought Vectors". In: *CoRR* abs/1506.06726 (2015). arXiv: 1506.06726. URL: http://arxiv.org/abs/1506.06726.

[6] Quoc V. Le and Tomás Mikolov. "Distributed Representations of Sentences and Documents". In: *CoRR* abs/1405.4053 (2014). arXiv: 1405.4053. URL: http://arxiv.org/abs/1405.4053.

[7] Peirong Ma and Xiao Hu. "A variational autoencoder with deep embedding model for generalized zero-shot learning". In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 34. 07. 2020, pp. 11733–11740.

[8] Alireza Makhzani et al. "Adversarial Autoencoders". In: *CoRR* abs/1511.05644 (2015). arXiv: 1511.05644. URL: http://arxiv.org/abs/1511.05644.

References III

[9] Tomas Mikolov et al. "Efficient estimation of word representations in vector space". In: *arXiv preprint arXiv:1301.3781* (2013).

[10] Mahdi Naser Moghadasi and Yu Zhuang. "Sent2Vec: A New Sentence Embedding Representation With Sentimental Semantic". In: *2020 IEEE International Conference on Big Data (Big Data)*. 2020, pp. 4672–4680. DOI: 10.1109/BigData50022.2020.9378337.

[11] Jeffrey Pennington, Richard Socher, and Christopher Manning. "GloVe: Global Vectors for Word Representation". In: *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Doha, Qatar: Association for Computational Linguistics, Oct. 2014, pp. 1532–1543. DOI: 10.3115/v1/D14-1162. URL: https://aclanthology.org/D14-1162.

References IV

[12] Nils Reimers and Iryna Gurevych. “Sentence-bert: Sentence embeddings using siamese bert-networks”. In: *arXiv preprint arXiv:1908.10084* (2019).

[13] Edgar Schönfeld et al. “Generalized Zero- and Few-Shot Learning via Aligned Variational Autoencoders”. In: *CoRR* abs/1812.01784 (2018). arXiv: 1812.01784. URL: http://arxiv.org/abs/1812.01784.

[14] Tianxiao Shen et al. “Latent Space Secrets of Denoising Text-Autoencoders”. In: *CoRR* abs/1905.12777 (2019). arXiv: 1905.12777. URL: http://arxiv.org/abs/1905.12777.

References V

[15] Kihyuk Sohn, Honglak Lee, and Xinchen Yan. "Learning Structured Output Representation using Deep Conditional Generative Models". In: *Advances in Neural Information Processing Systems*. Ed. by C. Cortes et al. Vol. 28. Curran Associates, Inc., 2015. URL: https://proceedings.neurips.cc/paper/2015/file/8d55a249e6baa5c06772297520da2051-Paper.pdf.

[16] Tailin Wu et al. "Meta-learning autoencoders for few-shot prediction". In: *CoRR* abs/1807.09912 (2018). arXiv: 1807.09912. URL: http://arxiv.org/abs/1807.09912.

[17] Congying Xia et al. "CG-BERT: Conditional Text Generation with BERT for Generalized Few-shot Intent Detection". In: *CoRR* abs/2004.01881 (2020). arXiv: 2004.01881. URL: https://arxiv.org/abs/2004.01881.

References VI

[18] Congying Xia et al. "Low-shot Learning in Natural Language Processing". In: *2020 IEEE Second International Conference on Cognitive Machine Intelligence (CogMI)*. 2020, pp. 185–189. DOI: 10.1109/CogMI50398.2020.00031.

[19] Shengjia Zhao, Jiaming Song, and Stefano Ermon. "InfoVAE: Information Maximizing Variational Autoencoders". In: *CoRR* abs/1706.02262 (2017). arXiv: 1706.02262. URL: http://arxiv.org/abs/1706.02262.

[20] Tiancheng Zhao, Ran Zhao, and Maxine Eskénazi. "Learning Discourse-level Diversity for Neural Dialog Models using Conditional Variational Autoencoders". In: *CoRR* abs/1703.10960 (2017). arXiv: 1703.10960. URL: http://arxiv.org/abs/1703.10960.