Semantic Overlap Summarization using Sentence Autoencoders

Naman Bansal, John Salvador and **Shubhra Kanti Karmaker ("Santu")**
Big Data Intelligence (BDI) Lab
Department of Computer Science and Software Engineering
College of Engineering, Auburn University
{nbansal, jms0256, sks0086}@auburn.edu

Abstract

Text autoencoders have found applications in conditional generation tasks such as style transfer. Following (Mai et al., 2020), we introduce a plug and play approach in this study. Our approach facilitates the utilization of any pre-trained sentence autoencoders for the purpose of generating semantic overlap summaries (Bansal et al., 2022b). Importantly, proposed operator learns a mapping solely within the embedding space, thereby enhancing the efficiency of the training procedure.

1 Introduction

Bansal et al. (2022b) introduced a novel NLP task termed Semantic Overlap Summarization (SOS) which focuses on distilling overlapping information from multiple alternate narratives. Subsequently, Bansal et al. (2022a) proposed a novel data augmentation technique, enabling the generation of a substantial volume of synthetic data tailored for training a seq-to-seq model capable of addressing the SOS task effectively. In this study, we adopt a distinct approach by employing an autoencoder-based methodology. This approach facilitates a plug-and-play framework, allowing for the utilization of pre-training on unlabeled data. Diverging from conventional methods, our framework operates within the low-dimensional continuous embedding space, leveraging the manifold of a pre-trained text autoencoder.

2 Proposed Framework

The key idea of our framework is to transform the discrete seq-to-seq task into a continuous embedding-to-embedding regression problem, as illustrated in Figure 1. The initial stage involves training an autoencoder model, depicted on the left of figure 1. Subsequently, we introduce the SOS operator, which learns to map the embeddings of the two inputs x_1 and x_2 and to output sentence embedding, as depicted in the center of figure 1. The loss function consists of the task specific loss components alongside an adversarial term, inspired by (Goodfellow et al., 2014), which serves to incentivizes the output vector $z_{\hat{y}}$ to remain within the autoencoder manifold. During the inference phase, depicted on the right side of Figure 1, the decoder component of the autoencoder is utilized to generate the discrete output sentence ($\hat{y}$), leveraging the output generated by the SOS operator. This process ensures that the generated summaries maintain fidelity to the input sequences while adhering to the learned semantic representations encoded within the autoencoder.

Autoencoders	R1	R2	R3
1D	86.30	69.17	79.57
3D	89.58	77.28	85.71

Table 1: Autoencoder reconstruction performance using ROUGE metric for autoencoders with 1-layer (**1D**) and 3-layer deep decoders.

2.1 Text Autoencoder

The foundational component of our framework is an autoencoder, which is trained to map an input sentence to itself. To achieve this, we follow the architecture as described by Montero et al. (2021) with minor changes. Specifically, we employ the RoBERTa-base model (Liu et al., 2019) as the encoder component and conduct experiments with decoder architectures comprising 1 and 3 layers of depth. Evaluation of the autoencoder's performance is conducted using the ROUGE metric (Lin, 2004), with the results presented in Table 1. Analysis reveals notable disparities in reconstruction performance between the different decoder depths. Specifically, deeper decoder architectures exhibit superior reconstruction capabilities compared to their shallower counterparts. However, experimen-

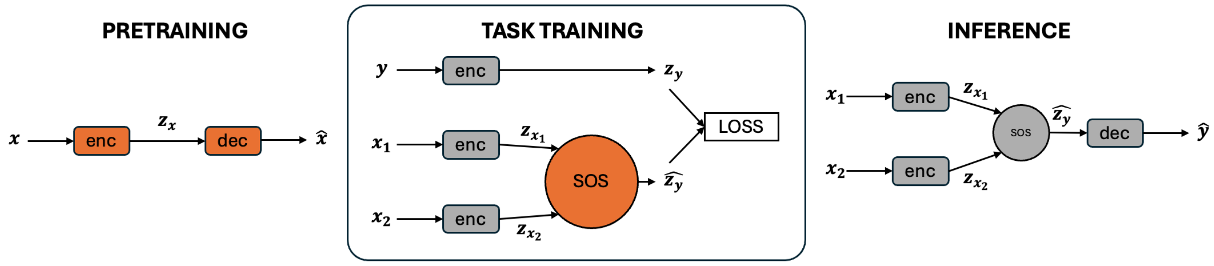

Figure 1: Schematic representation of our framework, inspired by (Mai et al., 2020). **Left:** Pre-training phase involves an autoencoder trained on unlabeled text, converting input sequences x into an embedding z_x and predicting reconstructions $\hat{x}$. **Center**: Two input sentences x_1 and x_2 are processed through the frozen encoder (depicted in gray) of the autoencoder. The Semantic Overlap Summarization (SOS) operator (trained, depicted in orange) learns to map the embeddings z_{x_1} and z_{x_2} to the output sentence embedding z_y of the output sentence y. **Right:** During inference, the SOS operator transforms input embeddings of x_1 and x_2 to the output embedding $\hat{z}_y$, utilized to generate the output sentence $\hat{y}$ using the decoder.

tation indicates diminishing returns beyond a depth of 3 layers, suggesting a saturation point in performance enhancement.

2.2 SOS Operator

As a starting point, we employ OffsetNet model, as detailed by Mai et al. (2020), as SOS operator. To adapt it for our task, we integrate Bahdanau Attention (Bahdanau et al., 2014) i.e. we compute a weighted average of z_{x_1} and z_{x_2} which are then fed into the OffsetNet model.

3 Experimental Setup and Results

3.1 Dataset

The absence of an established dataset poses a significant challenge for the SOS task. Addressing this gap, Bansal et al. (2022b) presented the first benchmark dataset in the news domain, sourced from AllSides.com. However, this dataset primarily comprises documents, necessitating the generation of synthetic samples to facilitate granularity at the sentence level. To overcome this limitation, we leverage samples from the CNN-DailyMail dataset (See et al., 2017) dataset. Given a sample document D, we extract three consecutive sentences, S_{prev}, S_O and S_{next}. Subsequently, employing ChatGPT (OpenAI, 2022), we amalgamate information from pairs of sentences (S_{prev}, S_O) and (S_O, S_{next}), resulting in synthesized pairs denoted as S_1 and S_2. Consequently, synthetic samples are constructed in the format of $\{\{S_1, S_2\}, S_O\}$, enabling finer-grained analysis at the sentence level. We further conduct human studies to verify the quality of these synthetic samples.

	R1	**R2**	**R3**
1D	35.33	8.98	24.95
3D	37.94	10.58	26.48
ChatGPT	47.19	25.92	39.50

Table 2: ROUGE scores of baselines models as compared to the proposed autoencoder approach. Here **1D** and **3D** refers to the autoencoders with 1 and 3 layer deep decoders. Autoencoder approach is $\sim 40\%$ as compared to the LLM based upper baseline in terms of ROUGE ($R2$). This is quite promising even though autoencoder model is approximately $\times 1000$ smaller than the LLM model.

3.2 Baseline Models

Due to the absence of dedicated models for the Semantic Overlap Summarization (SOS) task, we utilize large language models (LLMs) as an upper baseline. Specifically, we employ ChatGPT (OpenAI, 2022) with various prompts to address the SOS task and report performance using the most effective prompt.

3.3 Results

Consistent with the reconstruction results presented in Table 2, deeper decoder architectures within the autoencoder demonstrate improved performance. This observation underscores the potential for enhancing SOS performance by leveraging superior autoencoders. Comparatively, our SOS operator exhibits $\sim 40\%$ ROUGE ($R2$) scores of the LLMs, despite our model's significantly smaller size, approximately $\times 1000$ smaller that ChatGPT. Moving forward, we aim to experiment with diverse operators to further optimize the SOS performance.

References

Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2014. Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473*.

Naman Bansal, Mousumi Akter, and Shubhra Kanti Karmaker Santu. 2022a. Learning to generate overlap summaries through noisy synthetic data. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 11765–11777, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.

Naman Bansal, Mousumi Akter, and Shubhra Kanti Karmaker Santu. 2022b. Semantic overlap summarization among multiple alternative narratives: An exploratory study. In *Proceedings of the 29th International Conference on Computational Linguistics*, pages 6195–6207, Gyeongju, Republic of Korea. International Committee on Computational Linguistics.

Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2014. Generative adversarial nets. *Advances in neural information processing systems*, 27.

Chin-Yew Lin. 2004. ROUGE: A package for automatic evaluation of summaries. In *Text Summarization Branches Out*, pages 74–81, Barcelona, Spain. Association for Computational Linguistics.

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*.

Florian Mai, Nikolaos Pappas, Ivan Montero, Noah A. Smith, and James Henderson. 2020. Plug and play autoencoders for conditional text generation. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 6076–6092, Online. Association for Computational Linguistics.

Ivan Montero, Nikolaos Pappas, and Noah A Smith. 2021. Sentence bottleneck autoencoders from transformer language models. *arXiv preprint arXiv:2109.00055*.

OpenAI. 2022. Gpt3-ada-text embedding.

Abigail See, Peter J Liu, and Christopher D Manning. 2017. Get to the point: Summarization with pointer-generator networks. *arXiv preprint arXiv:1704.04368*.