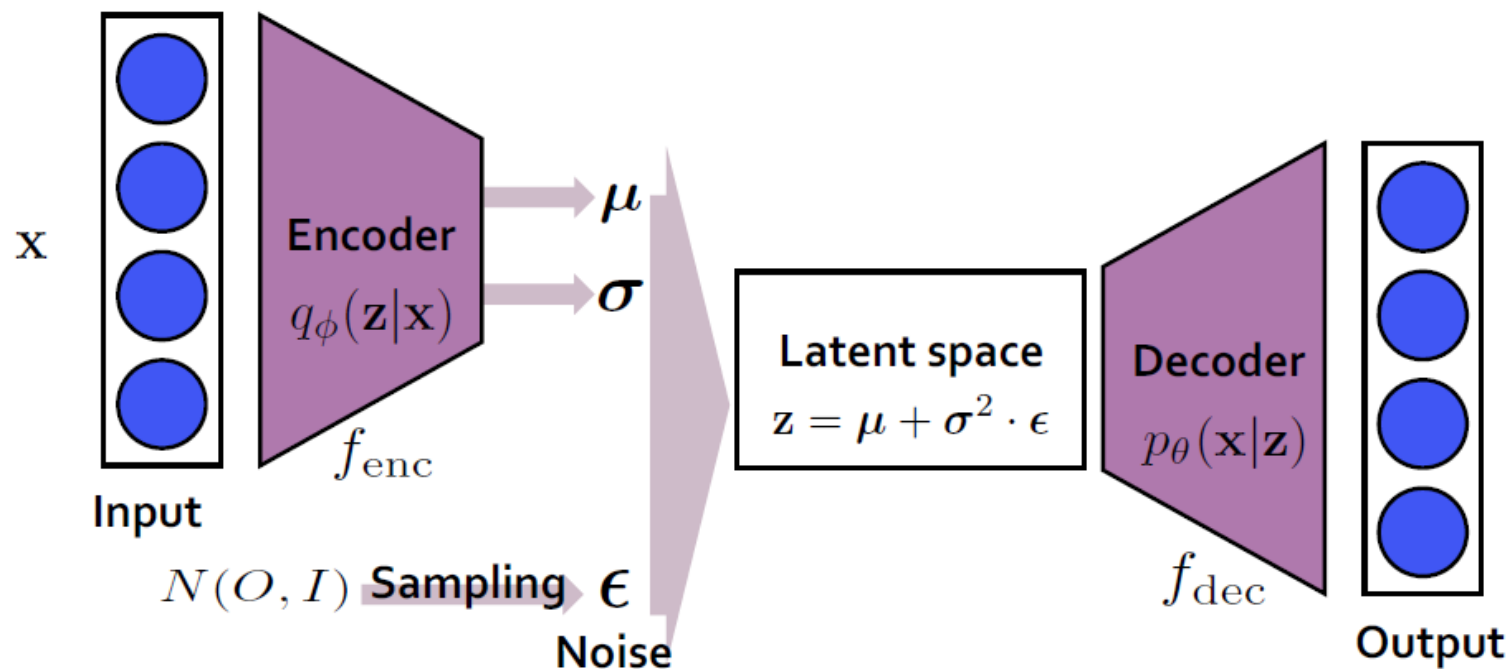


# Text VAE를 이용한 문장 잠재공간 분석 및 의미적 보간 시각화

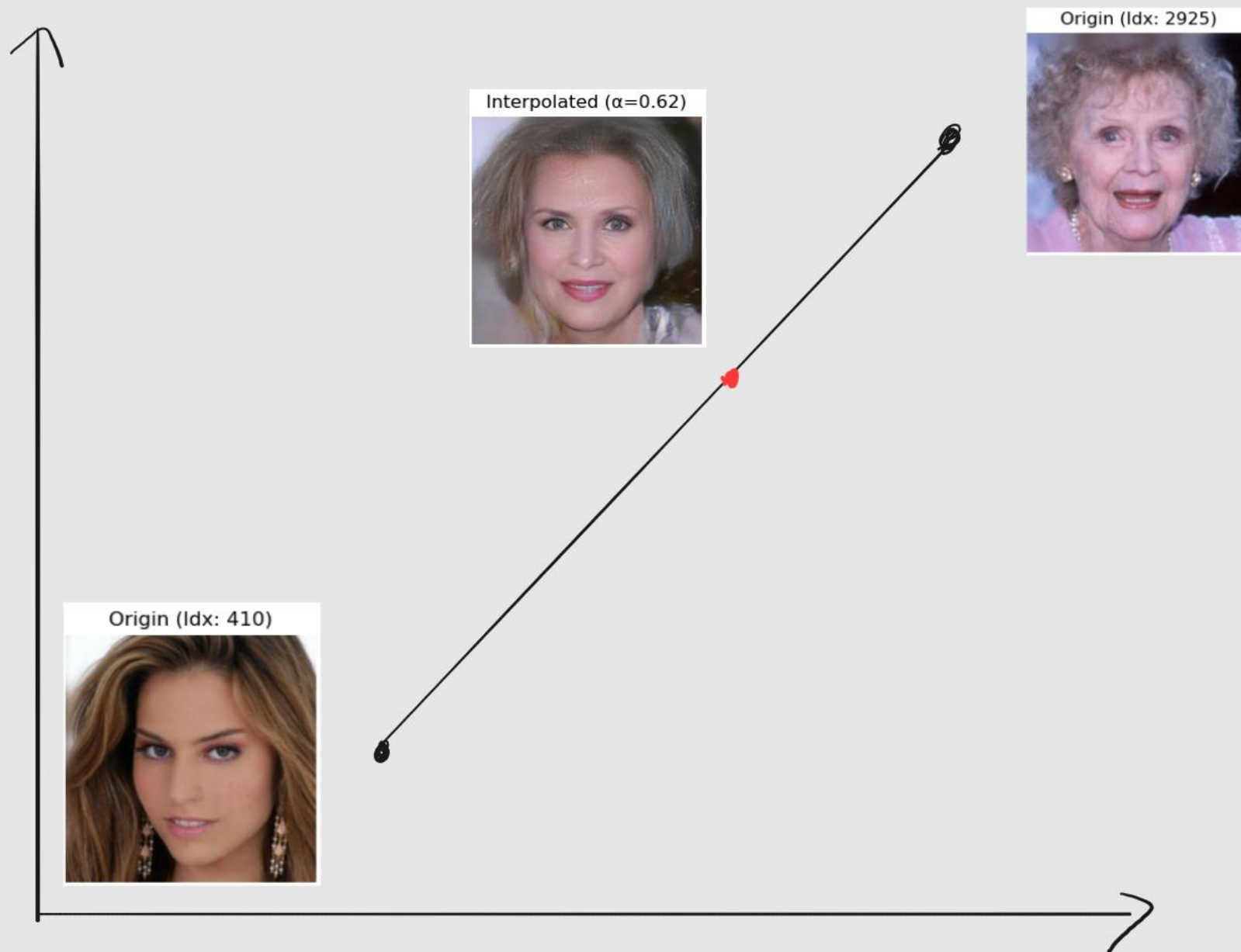
202192022 여명구

# 주제선정배경

- VAE, GAN, Diffusion과 같은 최신 생성 모델들이 고차원의 복잡한 데이터를 보다 작은 저차원으로 압축하고, 이 그 잠재공간 (Latent Space)으로부터 새로운 데이터를 생성하는 공통점을 가짐.



# Latent Space



# Latent Space



# 필요성

- 모델이 이 잠재공간에서 데이터를 단순히 압축, 복원하는 것을 넘어 데이터의 의미적 구조를 실제로 학습했는지 검증할 필요가 있음.

# 목적

- 텍스트 VAE를 구현하여, 모델이 학습한 잠재 공간이 실제로 문장의 의미에 따라 군집화되는지 시각적으로 확인(1), 공간 내에서 의미가 연속적으로 변화를 시각적으로 검증.(2)
- 이를 통해 생성모델의 텍스트 '이해' 방식을 직관적으로 분석하고 해석 가능성을 높이는 것을 목표로 함.

# 목표 결과물

- RNN기반 인코더,디코더를 사용해 VAE 모델 구현.(추후 pre-trained된 Bert나 GPT모델로 변경할 수 있음.)
- 특정 한국어 텍스트 데이터셋으로 학습된 모델의 잠재공간에 데이터들이 어떤 군집을 가지는지 파악(긍/부정 데이터가 서로 나뉘었는지 등)
- 의미가 다른 두 문장(ex, 긍/부정 리뷰)의 잠재 벡터를 보간하여 생성한 중간 단계 문장 리스트가 보통 느낌의 리뷰일 것을 기대.

# 추진 전략

- 데이터셋 선정(네이버 영화 리뷰 등)
  - 긍/부정 레이블이 명확하여 잠재공간 시각화 시 군집 분석에 용이.
- Okt 형태소 분석기를 사용한 토큰화, 단어 사전 구축, 임베딩 및 패딩
  - 이때 문장의 맥락을 파악해야 하기 때문에 불용어제거는 하지 않음.
- 모델 아키텍처
  - Encoder(RNN) -> 잠재공간 -> Decoder(RNN)
- 분석 및 시각화는 Scikit-learn의 PCA와 Isomap사용.
- 보간은 선형 보간만 진행. 이때 보간계수  $t \sim U(0, 1)$

# 추진 일정

주차	기간	수행 내용	비고
1주차	~10/31	관련 자료, 논문 검색, 데이터셋 확보	데이터 파이프라인 구축
2주차	11/1~11/11	데이터셋 전처리, Baseline 모델링	모델 학습
3주차	11/12~11/22	하이퍼 파라미터 튜닝 및 최종 모델 학습, 잠재공간에서 여러 쌍 보간	잠재 공간 탐색
4주차	11/23~12/4	잠재공간 시각화 및 보간 결과 분석	잠재 공간 시각화
5주차	12/5~12/15	최종 발표 준비	

# 참고자료

- 네이버 영화 리뷰 <https://github.com/e9t/nsmc>
- 핵심 논문

감사합니다.