

Speculative Sampling in Variational Autoencoders for Dialogue Response Generation



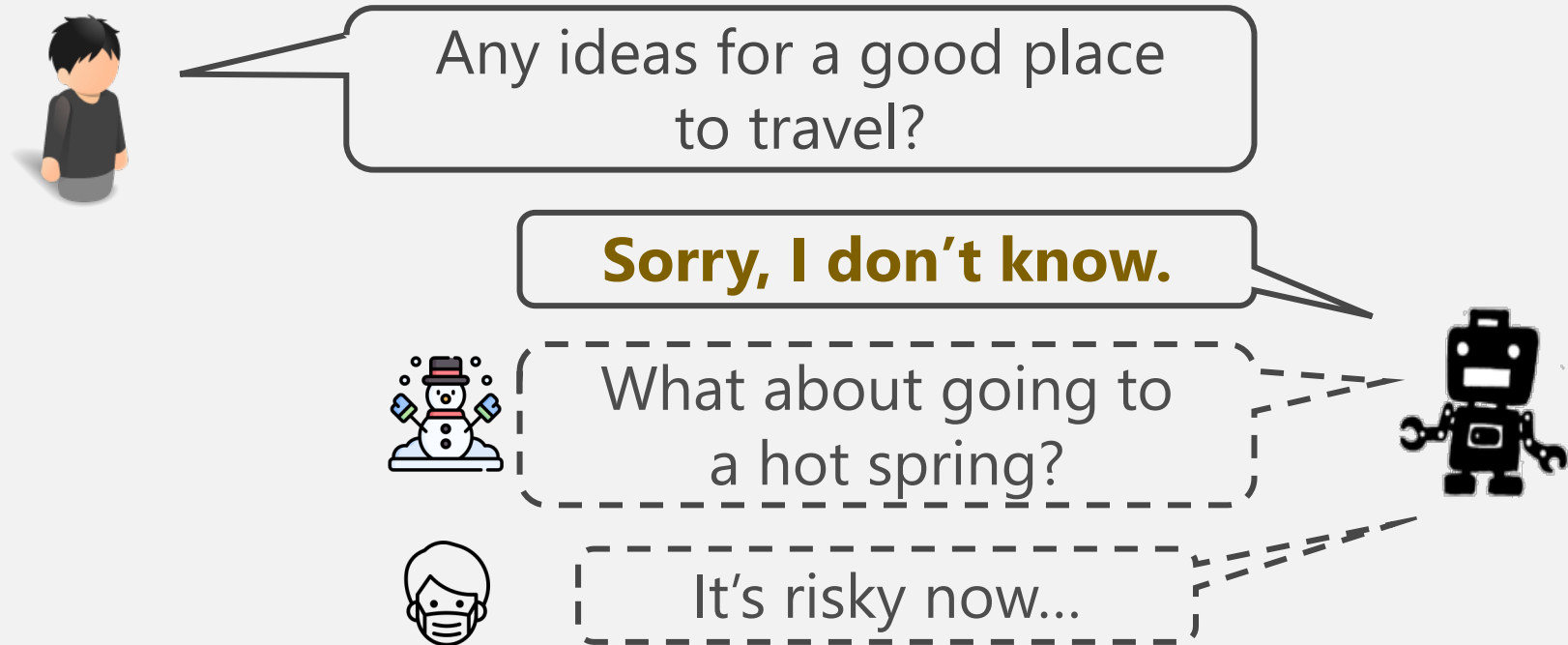
Shoetsu Sato

Institute of Industrial Science,
the University of Tokyo



Safe Response Problem

- Probable responses to a given utterance are affected by countless factors and conditions



Neural-based dialogue models often generate frequent and bland responses



Conditional Variational Autoencoder (CVAE) Is A Promising Solution

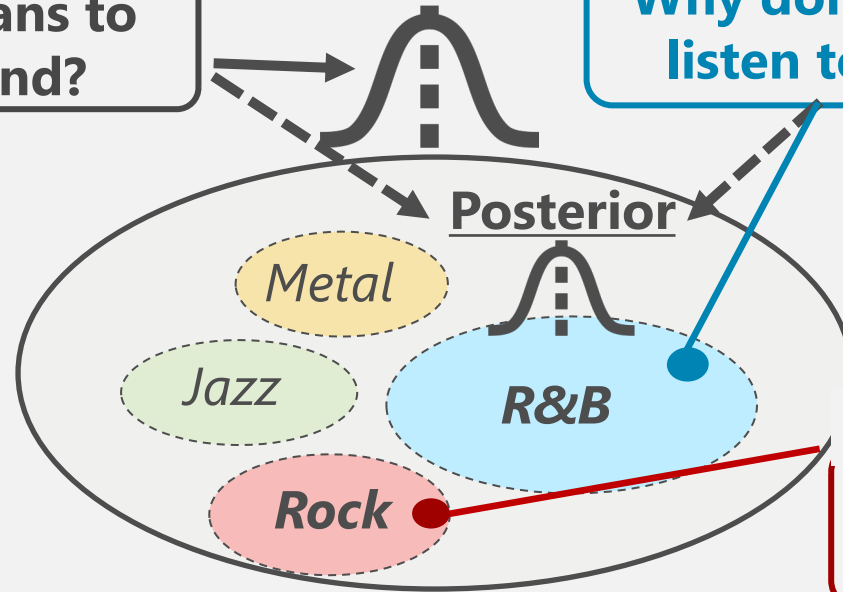
Utterance

Any musicians to recommend?

Prior

Given Response

Why don't you listen to MJ?



Posterior

Metal

Jazz

R&B

Rock

Other Response

Jimi Hendrix is a rock guitar legend.

- CVAEs add noises to the input representations for capturing 1-to-N relations
- However, its training is unstable, and it is sometimes ineffective

Problem: Failure of Autoencoding in Training

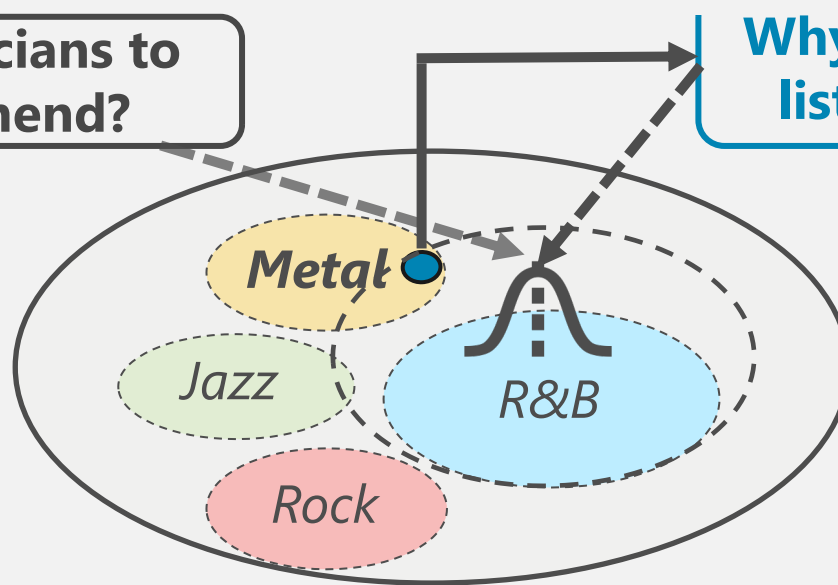
Utterance

Any musicians to
recommend?

Response

Why don't you
listen to MJ?

R&B



- Sampled latent variables can be **inappropriate** for the reconstruction in early stages of training
- As a result, **the decoder distrusts and ignores latent variables (a.k.a., KL vanishing or posterior collapse)**

Our Solution: Sample Several Variables and Use the Most Probable One

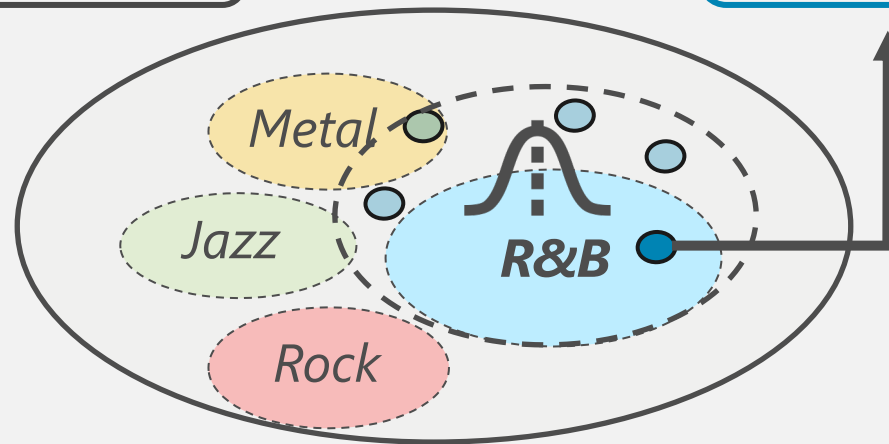
Utterance

Any musicians to recommend?

Response

Why don't you listen to MJ?

R&B



1. Sample several latent variables from the posterior
2. Compute cross-entropy loss for each variable
3. Use the variable w/ the lowest loss for optimization



Summary

- **Goal:** diversification of generated responses
- **Problem:** failure of CVAE training
- **Proposal:**
redundantly sample latent variables and use reliable one to improve the training
- **Experiments:**
 - Response generation test using conversational data constructed from social media
 - The proposed method improved the specificity of generated responses while keeping high sensibleness

