Text2Data: Low-Resource Data Generation with Textual Control

Shiyu Wang, Yihao Feng, Tian Lan, Ning Yu, Yu Bai, Ran Xu, Huan Wang, Caiming Xiong, Silvio Savarese

Salesforce AI Research

Motivation

Natural language serves as a common and straightforward signal for humans to interact seamlessly with machines. Recognizing the importance of this interface, the machine learning community is investing considerable effort in generating data that is semantically coherent with textual instructions. While strides have been made in text-to-data generation spanning image editing, audio synthesis, video creation, and beyond**, low-resource areas characterized by expensive annotations or complex data structures, such as molecules, motion dynamics, and time series, often lack textual labels**. This deficiency impedes supervised learning, thereby constraining the application of advanced generative models for text-to-data tasks.

We derive our work on diffusion models, but it can be seamlessly adapted to any other generative models.

Method

Overview of Text2Data

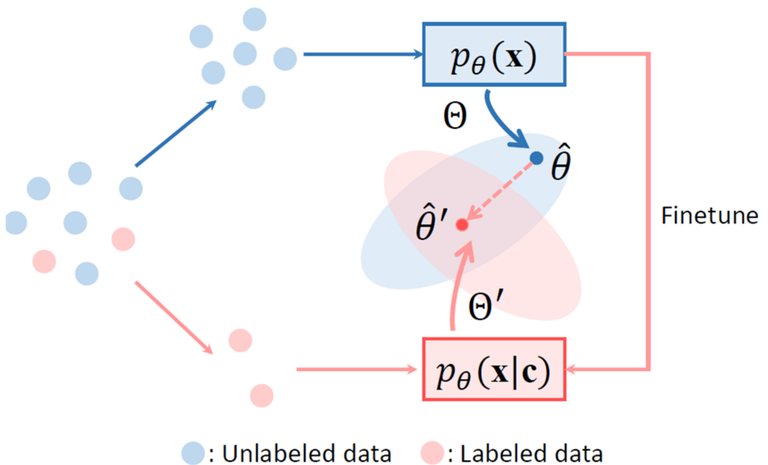


Figure: Overview of Text2Data.

1. The model leverages unlabeled data (i.e., blue) to discern the overall data distribution while the optimal set of model parameters Θ is obtained.
2. The model is finetuned on labeled data (i.e., red) by constraint optimization that gives the optimal set of parameters as $\Theta \cap \Theta'$, where Θ' is the optimal set of parameters if finetune model without constraint.

Low-resource Text-to-data generation

Notations:

- **x**: data samples such as molecules, motions, etc.
- **c**: textual description of **x**.
- $\mathcal{D} = \{\mathbf{x}, \mathbf{c}\}$: dataset with N independent data samples.
- $\mathcal{D}_p \subseteq \mathcal{D}$: data samples with text descriptions.

Generative learning objective:

$$\min_\theta \ \mathbb{E}_{\mathbf{x},\mathbf{c}\sim p_{\mathcal{D}_p}(\mathbf{x},\mathbf{c})}[-\log p_\theta(\mathbf{x}|\mathbf{c})] \quad \leftarrow \text{2. finetune on labeled data}$$

$$\text{s.t. } \mathbb{E}_{\mathbf{x}\sim p_{\mathcal{D}_p}(\mathbf{x})}[-\log p_\theta(\mathbf{x})] \le \xi,$$

$$\xi = \inf_{\theta\in\Theta} \mathbb{E}_{\mathbf{x}\sim p_{\mathcal{D}}(\mathbf{x})}[-\log p_\theta(\mathbf{x})] \quad \leftarrow \text{1. discern overall data distrib.}$$

Generative learning objective on empirical samples:

$$\min_\theta \ \hat{\mathcal{L}}_2(\theta) \quad \text{s.t. } \hat{\mathcal{L}}_1'(\theta) \le \hat{\xi},\ \hat{\xi} = \inf_{\theta\in\hat{\Theta}} \hat{\mathcal{L}}_1(\theta)$$

where $\hat{\mathcal{L}}_1(\theta) = \mathbb{E}_{x\sim\hat{p}_{\mathcal{D}}(x),t}[\|\epsilon_\theta(x^{(t)},t) - \epsilon\|^2]$, $\hat{\mathcal{L}}'_1(\theta) = \mathbb{E}_{x\sim\hat{p}_{\mathcal{D}_p}(x),t}[\|\epsilon_\theta(x^{(t)},t) - \epsilon\|^2]$, $\hat{\mathcal{L}}_2(\theta) = \mathbb{E}_{x,c\sim\hat{p}_{\mathcal{D}_p}(x,c),t}[\|\epsilon_\theta(x^{(t)},c,t) - \epsilon\|^2]$.

We illustrate our method using diffusion learning objective, but it can be seamlessly adapted to other generative models.

Algorithm 1: Lexicographic optimization

Input: $\hat{\xi} = \inf_{\theta\in\hat{\Theta}} \hat{\mathcal{L}}_1(\theta)$ by pretraining on $\mathcal{D}$; Total diffusion steps T; Scheduled forward variance $\{\beta_t\}_{t=1}^T$; $\{\alpha_t = 1-\beta_t\}_{t=1}^T$; Predefined positive hyperparameters α, β, γ and ω; probability of unconditional training p_{uncond}

while *Not converged* **do**

- Sample $\mathbf{x}, \mathbf{c} \sim \hat{p}_{\mathcal{D}_p}(\mathbf{x},\mathbf{c}), \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I}), t \sim U(1,T)$
- Compute $\bar{\alpha}_t = \prod_{s=1}^t \alpha_t$
- Diffuse $\mathbf{x}^{(t)} = \sqrt{\bar{\alpha}_t}\mathbf{x} + \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}$
- Replace $\mathbf{c}$ with $\emptyset$ with probability p_{uncond}
- Compute $\hat{\mathcal{L}}_2(\theta) = \|\boldsymbol{\epsilon}_\theta(\mathbf{x}^{(t)},\mathbf{c},t) - \boldsymbol{\epsilon}\|^2$, $\hat{\mathcal{L}}'_1(\theta) = \|\boldsymbol{\epsilon}_\theta(\mathbf{x}^{(t)},t) - \boldsymbol{\epsilon}\|^2$
- Compute $\phi(\theta) = \min(\alpha(\hat{\mathcal{L}}'_1(\theta) - \gamma\cdot\hat{\xi}), \beta\|\nabla\hat{\mathcal{L}}'_1(\theta)\|^2)$
- Compute $\lambda = \max(\frac{\phi(\theta) - \nabla\hat{\mathcal{L}}_2(\theta)^T\nabla\hat{\mathcal{L}}'_1(\theta)}{\|\nabla\hat{\mathcal{L}}'_1(\theta)\|^2}, 0)$
- Update θ by $\theta - \omega\cdot(\nabla\hat{\mathcal{L}}_2(\theta) + \lambda\nabla\hat{\mathcal{L}}'_1(\theta))$

Evaluation on text-to-molecule generation

Controllable text-to-molecule generation

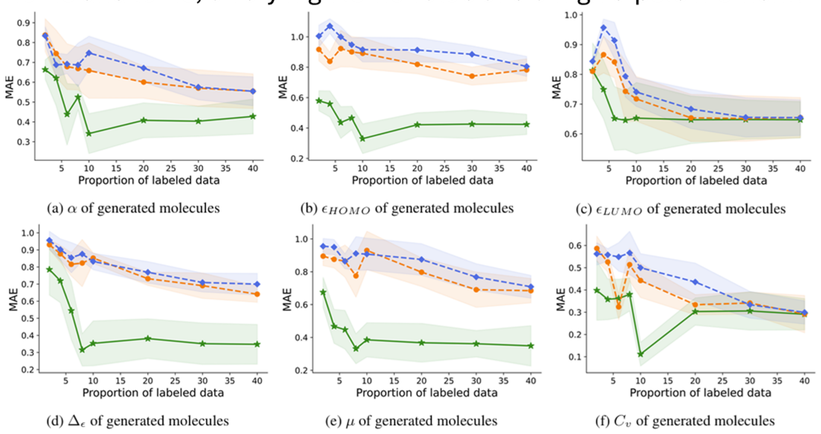
Dataset:

QM9 dataset with six molecular properties: polarizability (α), highest occupied molecular orbital energy (ϵ_{HOMO}), lowest unoccupied molecular orbital energy (ϵ_{LUMO}), the energy difference between HOMO and LUMO ($\triangle_\epsilon$), dipole moment (μ) and heat capacity at 298.15K (C_v).

Text description examples:

(1) **Exact:** A molecule with the heat capacity of-0.11, the lomo of 0.87, the homo of-0.21, the polarizability of 0.95, the dipole moment of-1.61 and the energy gap between homo and lumo as 0.94.

(2) **General:** A molecule with a high homo value, a very low heat capacity, a medium polarizability, a high energy difference between homo and lomo, a very high lomo value and a high dipole moment.



(a) α of generated molecules (b) ϵ_{HOMO} of generated molecules (c) ϵ_{LUMO} of generated molecules

(d) Δ_ϵ of generated molecules (e) μ of generated molecules (f) C_v of generated molecules

Green: EDM+Lexico, Orange: EDM+Finetune, Blue: EDM

Figure: Evaluate controllability on Molecule dataset.

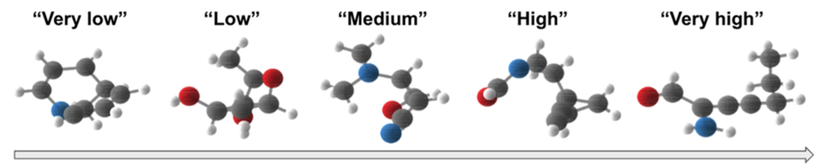


Figure: Visualization of generated molecules when the polarizability increases from "very low" to "very high".

More in paper

- Generalization bound of learning constraint.
- Evaluation results on motion and time series.
- Ablation studies and more comparison models.