Benchmarking and Improving Compositional Generalization of Multi-aspect Controllable Text Generation

Tianqi Zhong[1*], Zhaoyi Li[1,2*], Quan Wang[3], Linqi Song[2], Ying Wei[4], Defu Lian[1], Zhendong Mao[1†]

[1]University of Science and Technology of China; [2]City University of Hong Kong;
[3]Beijing University of Posts and Telecommunications; [4]Nanyang Technological University

Motivations:

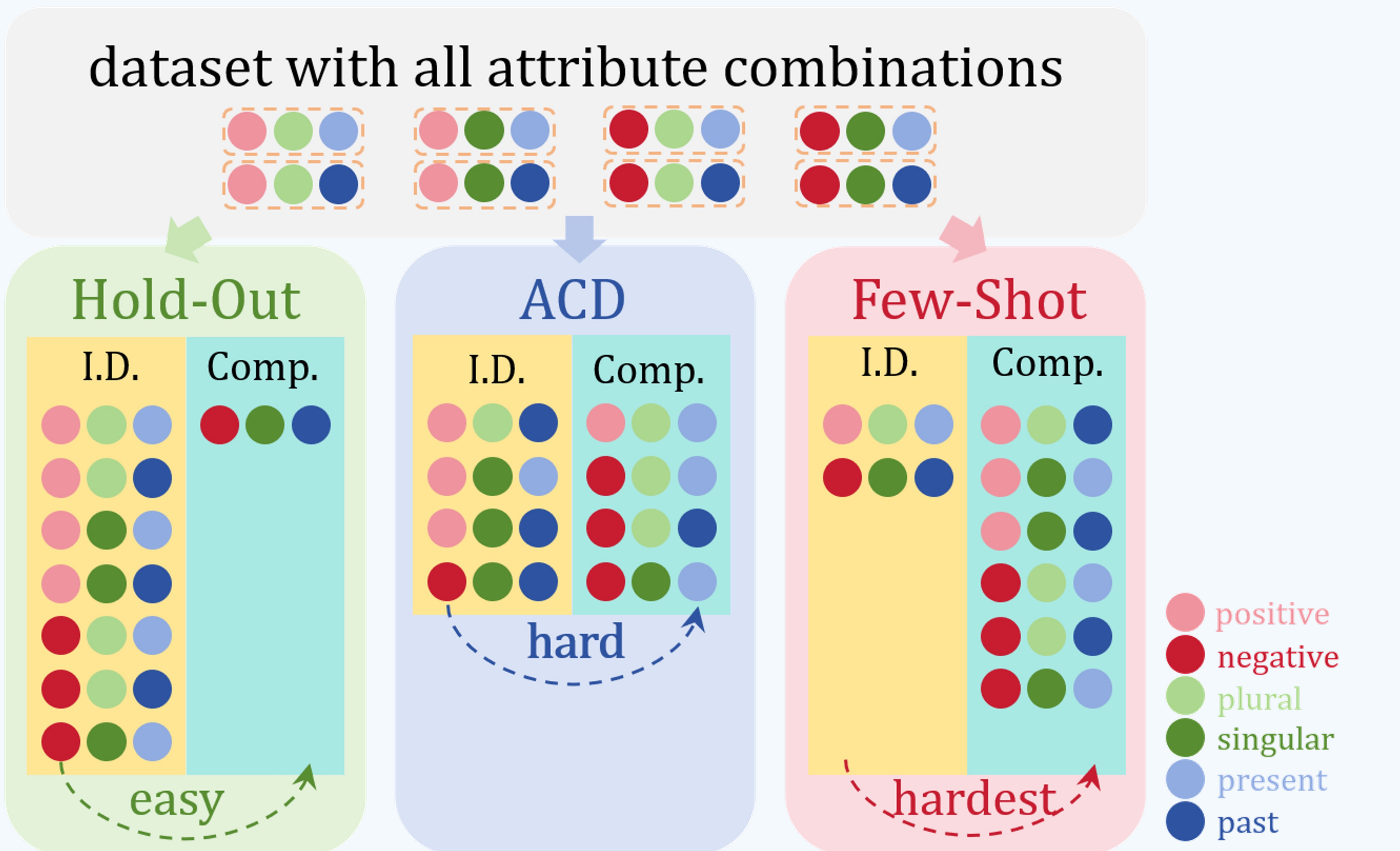
- Compositional generalization is a crucial property of multi-aspect controllable text generation, which refers to the model's ability to generate text with attribute combinations recombined by single attributes from the training data.
- Previous work mainly focused on enhancing the performance of multi-attribute controllable text generation within the training data distribution, while neglecting the model's generalization capabilities outside of the training data distribution[1].

(positive, male)
(negative, female)

In-distribution testing: (positive, male) (negative, female)

Compositional testing: (positive, female) (negative, male)

Benchmark: CompMCTG



- Split the whole dataset $\mathcal{C}$ into two disjoints sets: in-distribution set $\mathcal{C}_{i.d.}$ and compositional set $\mathcal{C}_{comp}$. The formal definition of an eligible split $s(\mathcal{C}) = \mathcal{C}_{i.d.}, \mathcal{C}_{comp}$ as following:

$$\mathcal{C} = \mathcal{A}_1 \times \mathcal{A}_2 \times \cdots \times \mathcal{A}_m = \{(A_i^{t_i})_{1 \le i \le m} | 1 \le t_i \le a_i\}$$

$$\boldsymbol{\mathcal{C}} = \boldsymbol{\mathcal{C}_{i.d.}} \cup \boldsymbol{\mathcal{C}_{comp}}, \boldsymbol{\mathcal{C}_{i.d.}} \cap \boldsymbol{\mathcal{C}_{comp}} = \emptyset$$

$$\{\boldsymbol{att} | \exists \boldsymbol{c} \in \boldsymbol{\mathcal{C}_{comp}}, \boldsymbol{att} \in \boldsymbol{c}\} \subseteq \{\boldsymbol{att} | \exists \boldsymbol{c} \in \boldsymbol{C_{i.d.}}, \boldsymbol{att} \in \boldsymbol{c}\}$$

- **Protocol One: Hold-Out**

$$S_{Hold-Out} = \{(\mathcal{C}_{i.d.}, \mathcal{C}_{comp}) | \mathcal{C}_{comp} \in \mathcal{C}, |\mathcal{C}_{comp}| = k, \mathcal{C}_{i.d.} = \mathcal{C} \backslash \mathcal{C}_{comp}\}$$

- **Protocol Two: ACD**.

1. Inspired by Keysers[2]. Calculate frequency density of $\left(A_i^{t_i}, A_j^{t_j}\right)$:

$$f_{\mathcal{C}}\left(\left(A_i^{t_i}, A_j^{t_j}\right)\right) = \frac{\sum_{c \in \mathcal{C}} \mathbb{I}\left(A_i^{t_i} \in c \wedge A_j^{t_j} \in c\right)}{\sum_{c \in \mathcal{C}} \sum_{x \in c, y \in c, x \neq y} \mathbb{I}(1)}, \mathcal{C} \in \{\mathcal{C}_{i.d.}, \mathcal{C}_{comp}\}$$

2. Introduce the Chernoff Coefficient $S(P,Q)$

$$P = (p_1, p_2, \ldots, p_n), Q = (q_1, q_2, \ldots, q_n)$$

$$\Rightarrow S(P,Q) = \sum_{i=1}^{n} p_i^{\alpha} q_i^{1-\alpha} \in [0,1]$$

3. Define the ACD: $D(P_{i.d.}, P_{comp}) = 1 - S(P_{i.d.}, P_{comp}) \in [0,1]$

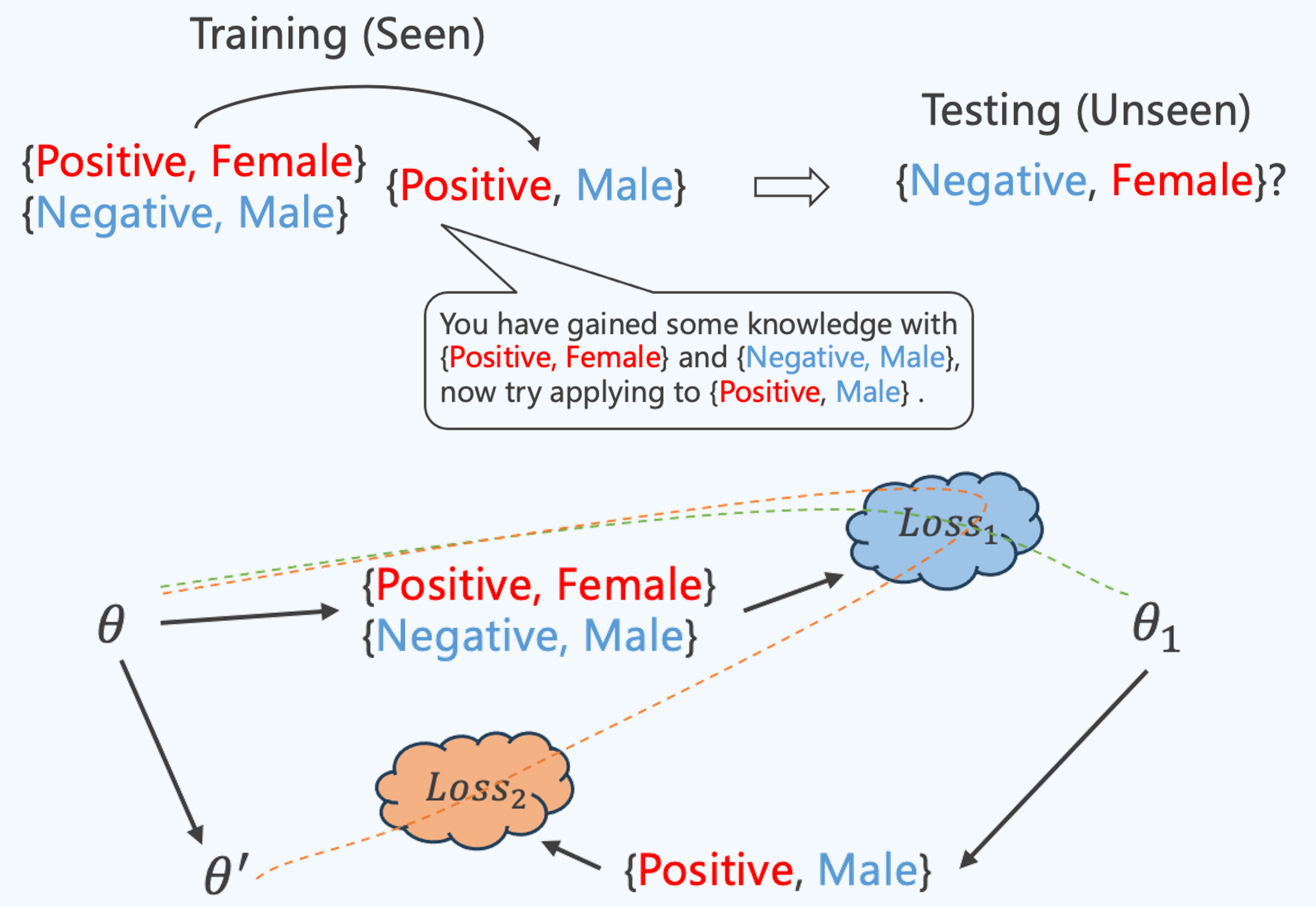
$$S_{ACD} = \{(\mathcal{C}_{i.d.}, \mathcal{C}_{comp}) | \max_{\mathcal{C}_{i.d.}, \mathcal{C}_{comp}} D(P_{i.d.}, P_{comp}) \wedge |\mathcal{C}_{i.d.}| = |\mathcal{C}_{comp}|\}$$

- **Protocol Three: Few-Shot**

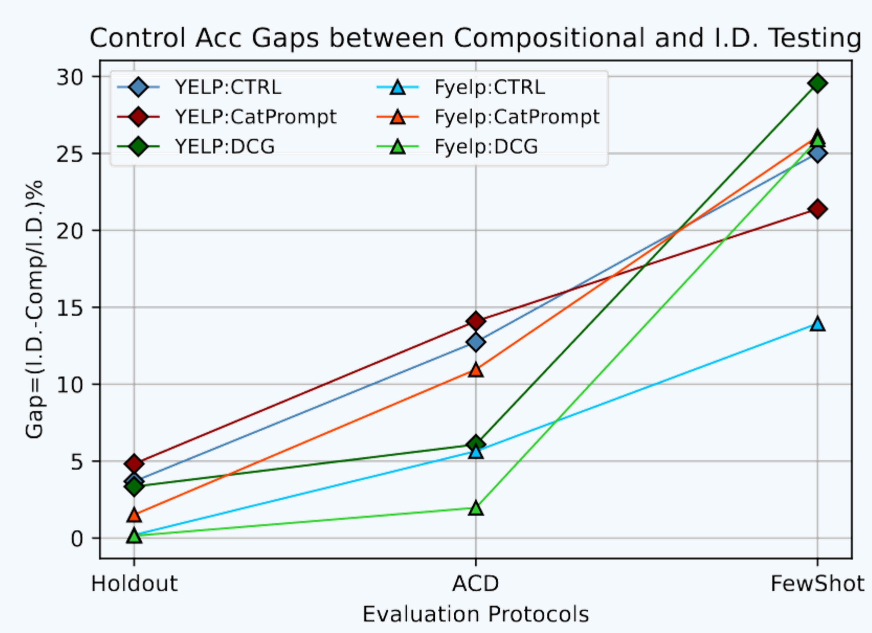
$$S_{Few-Shot} = \{(\mathcal{C}_{i.d.}, \mathcal{C}_{comp}) | (A_i^{t_i})_{1 \le i \le m}^{1 \le t_i \le a_i} \text{ in } \mathcal{C}_{i.d.} \wedge \min |\mathcal{C}_{i.d.}|\}$$

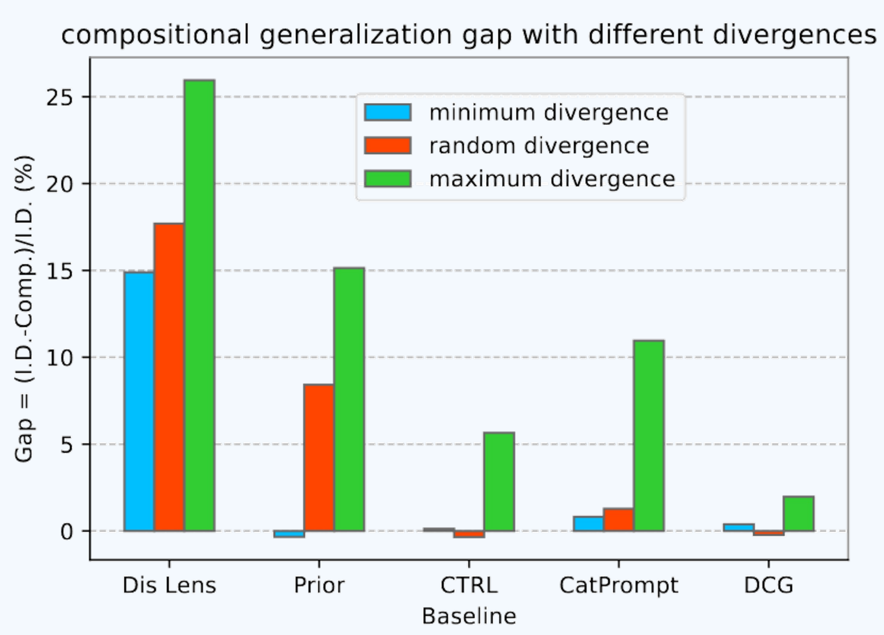
Meta-MCTG Algorithm:

- Inspired by previous meta-learning works[3] targeting generalization, we aim to leverage Model-Agnostic Meta Learning (MAML[4]) to mitigate the overfitting problem in join-training-based MCTG.

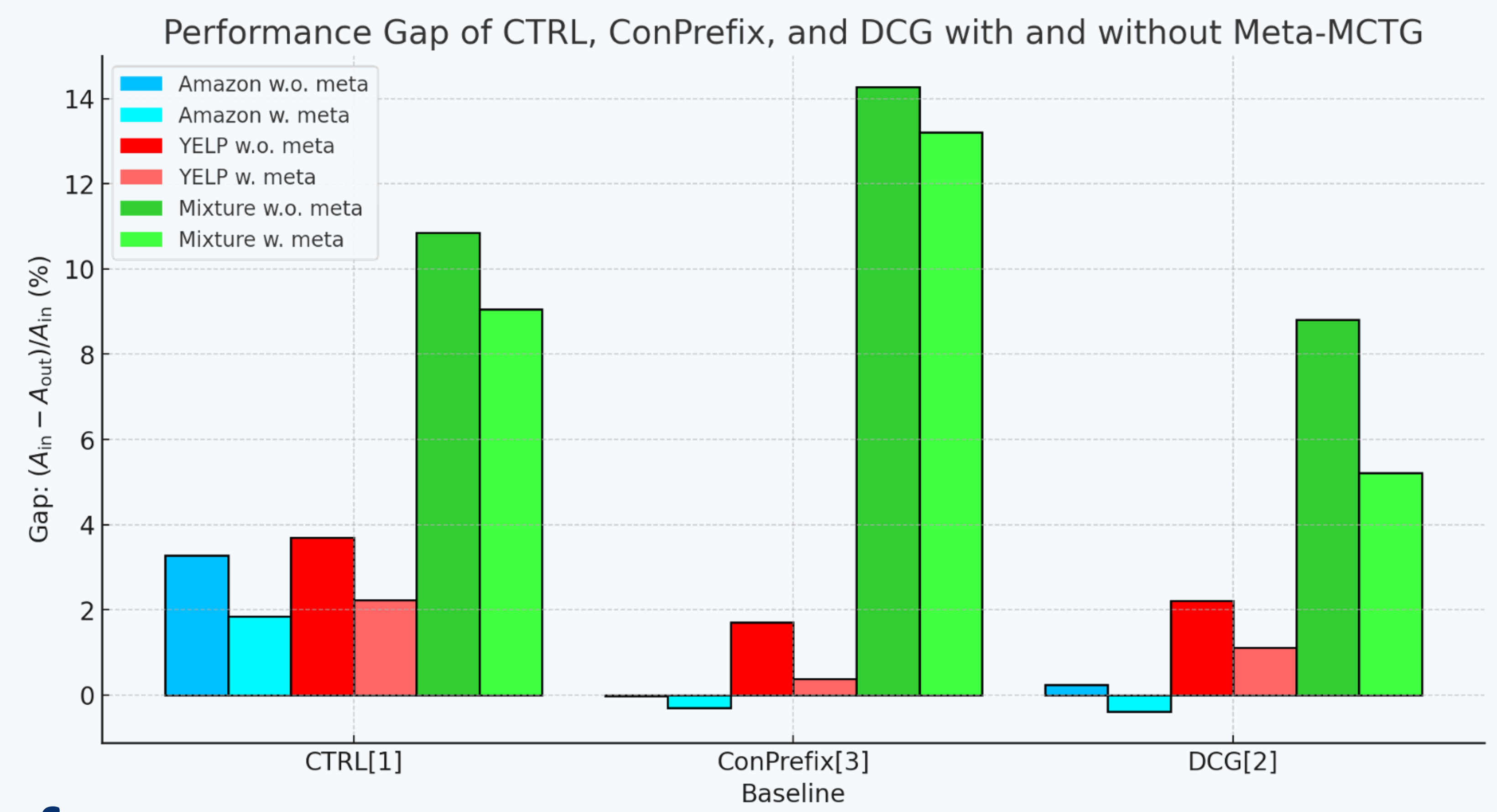


Results in CompMCTG:





Results in Meta-MCTG:



Reference:

[1]Zeng, Weihao, et al. "Seen to Unseen: Exploring Compositional Generalization of Multi-Attribute Controllable Dialogue Generation." Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). 2023.

[2]Li, Da, et al. "Learning to generalize: Meta-learning for domain generalization." Proceedings of the AAAI conference on artificial intelligence. Vol. 32. No. 1. 2018.

[3]Conklin, Henry, et al. "Meta-Learning to Compositionally Generalize." Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers). 2021.

[4]Finn, Chelsea, Pieter Abbeel, and Sergey Levine. "Model-agnostic meta-learning for fast adaptation of deep networks." International conference on machine learning. PMLR, 2017.

ACL 2024
Bangkok, Thailand