Privacy-Preserving Domain Adaptation of Semantic Parser

Fatemehsadat Mireshghallah, Yu Su, Tatsunori Hashimoto, Jason Eisner, Richard Shin

Task-oriented dialogue systems often assist users with personal or confidential matters. So:
- Data is private and practitioners are not allowed to look at it.
- How can we know where the system is failing and needs more training data or new functionality?

"Could you tell me what the weather is gonna be like today in New York?"

Email everyone who declined the invitation.

We simulated a situation where users are asking about the weather but the original semantic parser was not trained on weather-related functions:
- We created the original semantic parser by training on $\frac{1}{10}$ of our data (SMCalFlow), excluding any examples that use weather-related functions.
- We treated the other $\frac{9}{10}$ of the data as private user utterances, including those requesting weather. We created approximate private annotations for the private utterances, using the original semantic parser.
- We apply the baseline and proposed methods to create public synthesized datasets, which include weather functions.
- We simulated high-quality human annotation of the public synthetic utterances. We re-train the parser with this additional annotated data.

We created the original semantic parser by training on

We model $p(x)$, where $x$ is a private utterance: 1-stage baseline approach of fine-tuning a pre-trained generative auto-regressive language model on private user utterances using differentially private SGD. To create the synthesized dataset we take samples from the fine-tuned model.

"Could you tell me what the weather is gonna be like today in New York?"

We simulated a situation where users are asking about the weather but the original semantic parser was not trained on weather-related functions:
- We created the original semantic parser by training on $\frac{1}{10}$ of our data (SMCalFlow), excluding any examples that use weather-related functions.
- We treated the other $\frac{9}{10}$ of the data as private user utterances, including those requesting weather. We created approximate private annotations for the private utterances, using the original semantic parser.
- We apply the baseline and proposed methods to create public synthesized datasets, which include weather functions.
- We simulated high-quality human annotation of the public synthetic utterances. We re-train the parser with this additional annotated data.

We simulated a situation where users are asking about the weather but the original semantic parser was not trained on weather-related functions:
- We created the original semantic parser by training on $\frac{1}{10}$ of our data (SMCalFlow), excluding any examples that use weather-related functions.
- We treated the other $\frac{9}{10}$ of the data as private user utterances, including those requesting weather. We created approximate private annotations for the private utterances, using the original semantic parser.
- We apply the baseline and proposed methods to create public synthesized datasets, which include weather functions.
- We simulated high-quality human annotation of the public synthetic utterances. We re-train the parser with this additional annotated data.

We simulated a situation where users are asking about the weather but the original semantic parser was not trained on weather-related functions:
- We created the original semantic parser by training on $\frac{1}{10}$ of our data (SMCalFlow), excluding any examples that use weather-related functions.
- We treated the other $\frac{9}{10}$ of the data as private user utterances, including those requesting weather. We created approximate private annotations for the private utterances, using the original semantic parser.
- We apply the baseline and proposed methods to create public synthesized datasets, which include weather functions.
- We simulated high-quality human annotation of the public synthetic utterances. We re-train the parser with this additional annotated data.

We simulated a situation where users are asking about the weather but the original semantic parser was not trained on weather-related functions:
- We created the original semantic parser by training on $\frac{1}{10}$ of our data (SMCalFlow), excluding any examples that use weather-related functions.
- We treated the other $\frac{9}{10}$ of the data as private user utterances, including those requesting weather. We created approximate private annotations for the private utterances, using the original semantic parser.
- We apply the baseline and proposed methods to create public synthesized datasets, which include weather functions.
- We simulated high-quality human annotation of the public synthetic utterances. We re-train the parser with this additional annotated data.