InferenceQL: an SQL-like probabilistic programming language

Ulrich Schaechtle                     Zane Shelby               Vikash K. Mansingka

MIT Probabilistic Computing Project

1. InferenceQL platform

Scientific experiment Measurements

Bayesian Program Synthesis

"Virtual Wetlab"

Sum Product Network

Bayesian fine-tuning of structure & parameters

"Virtual Wetlab" = InferenceQL

InferenceQL Query

1. Bayesian synthesis of generative programs [1]
   Structure learning for generative programs scales to 100K+ rows, 5K columns
   Fine-tuning via mini-batches of ~100 columns

2. Fast exact inference via Sum-Product Probabilistic Language (SPPL) [2]
   Milliseconds per query
   Complex events combining multiple genes & genetic design variables

Key innovations:

3. Example InferenceQL application: genetic circuit design

Experimental data available for genetic circuit design

What if we had a program that simulated virtual experiments?

Bayesian program synthesis produces a generative program that can serve as a "virtual wetlab"

Experimental condition Circuit parts ~ 4095 Genes that are not part of the circuit

Hidden variables inside host organisms

We lab environment

4. Example query and accuracy results: predicting more than 4000 genes accurately

SQL-like probabilistic query

Accuracy results: Only 6.51% of predicted gene expression have a fold change error larger than 2x

Qualitative assessment: InferenceQL models the dynamics in the data more accurately than stand approaches

Regression mischaracterizes relationships (4 examples)

Virtual data from the whole-genome simulator matches real data

---

[1] Bayesian synthesis of probabilistic programs for automated data modeling
   Sead, F. A.; Cosmiano-Towner, M.; Schaechtle, U.; Riniard, M. C.; and Mansinghka, V. K. (POPL 2019)

   Sead, F. A.; Riniard, M. C.; and Mansinghka, V. K. (in review for POPL 2020)