Answering FAQs in CSPs, Probabilistic Graphical Models, Databases, Logic and Matrix Operations (Invited Talk)

Atri Rudra

State University of New York at Buffalo, USA
atri@buffalo.edu

ABSTRACT

In this talk we will discuss a general framework to solve certain sums of products of functions over semi-rings. This captures many well-known problems in disparate areas such as CSPs, Probabilistic Graphical Models, Databases, Logic and Matrix Operations. This talk is based on joint work titled *FAQ: Questions Asked Frequently* with Mahmoud Abo Khamis and Hung Q. Ngo, which appeared in *PODS 2016*.

CCS CONCEPTS

• Theory of computation → Logic and databases;

KEYWORDS

Probabilistic Graphical Models, Database Joins, Semi Rings

ACM Reference format:

Atri Rudra. 2017. Answering FAQs in CSPs, Probabilistic Graphical Models, Databases, Logic and Matrix Operations (Invited Talk). In *Proceedings of 49th Annual ACM SIGACT Symposium on the Theory of Computing, Montreal, Canada, June 2017 (STOC'17-KEY)*, 1 pages.
DOI: 10.1145/3055399.3079073

1 INTRODUCTION

Consider the following four problems that span disparate areas.

PROBLEM 1.1 (EXACT INFERENCE IN PGMs). *We are given a hypergraph $\mathcal{H} = (\mathcal{V}, \mathcal{E})$ where for every edge $e \in \mathcal{E}$ we have a function $f_e : \mathcal{D}^e \rightarrow \mathbb{R}$. Further, given a subset $S \subseteq \mathcal{V}$, we want to compute the function (where $\mathbf{x} = (x_S, \mathbf{x}_{\mathcal{V} \setminus S}) \in \mathcal{D}^{\mathcal{V}}$):*

$$f(\mathbf{x}_S) = \sum_{\mathbf{x}_{\mathcal{V} \setminus S} \in \mathcal{D}^{\mathcal{V} \setminus S}} \prod_{e \in \mathcal{E}} f_e(\mathbf{x}_e),$$

where for any subset $T \subseteq \mathcal{V}$, we use $\mathbf{x}_T$ to denote the vector projected down to indices in T .

PROBLEM 1.2 (DATABASE JOINS). *We are given a hypergraph $\mathcal{H} = (\mathcal{V}, \mathcal{E})$ where for every edge $e \in \mathcal{E}$ we have a function $f_e : \mathcal{D}^e \rightarrow \{0, 1\}$. Then compute the function $f : \mathcal{D}^{\mathcal{V}} \rightarrow \{0, 1\}$ where for every $\mathbf{x} \in \mathcal{D}^{\mathcal{V}}$, we have*

$$f(\mathbf{x}) = \prod_{e \in \mathcal{E}} f_e(\mathbf{x}_e).$$

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

STOC'17-KEY, Montreal, Canada

© 2017 ACM. 978-1-4503-4528-6/17/06...\$15.00
DOI: 10.1145/3055399.3079073

We note that in database systems we represent all the functions involved by only listing the tuples $\mathbf{x}$ that evaluate to 1 (also called the *listing representation*).

PROBLEM 1.3 (COUNTING QUANTIFIED CONJUNCTIVE QUERIES). *We are given a hypergraph $\mathcal{H} = (\mathcal{V}, \mathcal{E})$ where for every edge $e \in \mathcal{E}$ we have a function $f_e : \mathcal{D}^e \rightarrow \{0, 1\}$. Further, we will assume that $\mathcal{V} = \{1, \dots, n\}$. Given a subset $S = \{1, \dots, s\}$, we want to compute the constant function*

$$f = \sum_{(x_1, \dots, x_s) \in \mathcal{D}^s} Q_{s+1} x_{s+1} \cdots Q_n x_n \prod_{e \in \mathcal{E}} f_e(\mathbf{x}_e),$$

where $\mathbf{x} = (x_1, \dots, x_n) \in \mathcal{D}^n$ and each $Q_i \in \{\exists, \forall\}$.

PROBLEM 1.4 (MATRIX-VECTOR MULTIPLICATION). *Given functions $f(X, Y) : \mathcal{D}^2 \rightarrow \mathbb{R}$ and $g(Y) : \mathcal{D} \rightarrow \mathbb{R}$, compute the function*

$$h(x) = \sum_{y \in \mathcal{D}} f(x, y) \cdot g(y).$$

Generally the function f is encoded by a matrix (with the rows and columns indexed by $\mathcal{D}$) and g and h are encoded as vectors indexed by $\mathcal{D}$. Finally, we present a generic “sum of product” problem that captures all of the above problems.

PROBLEM 1.5 (FUNCTIONAL AGGREGATE QUERY (FAQ)). *We are given a hypergraph $\mathcal{H} = (\mathcal{V} = \{1, \dots, n\}, \mathcal{E})$ where for every edge $e \in \mathcal{E}$ we have a function $f_e : \mathcal{D}^e \rightarrow \mathcal{R}$. Given a subset $S = \{1, \dots, s\}$ of $\mathcal{V}$, we want to compute the function:*

$$f(x_1, \dots, x_s) = \oplus_{x_{s+1} \in \mathcal{D}}^{(s+1)} \cdots \oplus_{x_i \in \mathcal{D}}^{(i)} \cdots \oplus_{x_n \in \mathcal{D}}^{(n)} \otimes_{e \in \mathcal{E}} f_e(\mathbf{x}_e),$$

where for each $s+1 \leq i \leq n$, either $\oplus^{(i)} = \otimes$ or $(\mathcal{R}, \oplus^{(i)}, \otimes, 0, 1)$ form a commutative semi-ring¹.

We note that Problems 1.1 and 1.4 are special FAQs where all the semi-rings are $(\mathbb{R}, +, \cdot, 0, 1)$ —i.e. the “usual” arithmetic over reals while Problem 1.2 is a special case where all the semi-rings are $(\{0, 1\}, \vee, \wedge, 0, 1)$. See [1] for a longer listing of problems (including Problem 1.3) that are special cases of Problem 1.5.

Our technical contribution is to present one single algorithm to solve Problem 1.5 using some recent algorithmic developments for Problem 1.2. For a gentle introduction to these ideas, come to the talk or see our survey [2]. Readers interested in the full gory details should see the full version of the paper [1].

REFERENCES

- [1] M. Abo Khamis, H. Q. Ngo, and A. Rudra. FAQ: questions asked frequently. *CoRR*, abs/1504.04044, 2015.
- [2] M. Abo Khamis, H. Q. Ngo, and A. Rudra. Juggling functions inside a database. *CoRR*, abs/1703.03147, 2017.

¹In other words, adding with $\oplus^{(i)}$ and multiplying with $\otimes$ are closed under $\mathcal{R}$ and are commutative. $\otimes^{(i)}$ distributes over $\otimes$ and all semi-rings share the same additive and multiplicative identities: 0 and 1 respectively.