Logging an Egg:
Datalog on E-Graphs

Philip Zucker

Draper Laboratory

June 13, 2022

Datalog : Databases, Logic, and Proofs.

$$\forall x, y.edge(x, y) \implies path(x, y)$$

$$\frac{\text{edge(x,y)}}{\text{path(x,y)}}$$

```
path(X, Y) :- edge(X, Y).
path(X, Y) :- path(X, Z), edge(Z, Y).
```

Search right hand side in database. Insert left side. Repeat.

Datalog: Applications

- Graph Problems
- Worklist Algorithms
- You can program in it. Explicit control flow
- Mutually Recursive Analyses
- Program Analysis
  - Doop [1]
  - DDisasm [2]

[1] https://bitbucket.org/yanniss/doop/src/master/

[2] https://github.com/GrammaTech/ddisasm

Datalog vs Prolog

- Pattern matching vs Unification
- Top down vs Bottom up
- Complete vs Incomplete Search
- Tabling

E-Graphs

- Datastructure for terms and equalities
- Egg [3]: Efficient Rust library
- Term Rewriting $?a + 0 \rightarrow ?a$
- An E-Graph Never Forgets
- Maximize sharing up and down.
- Rule ordering
- $?a + (?b + ?c) = (?a + ?b) + ?c$. $?a + -?a = 0$
- $x + (-x + 10)$

[3] https://egraphs-good.github.io/

Applications of E-Graphs

- SMT and other Theorem Proving
- Compiler Optimization. PEG [4]
- Herbie [5]
- Query Optimization
- Szalinski - CAD
- YOGO [6]

[4] https://rosstate.org/publications/eqsat/
[5] https://herbie.uwplse.org/
[6] https://www.jameskoppel.com/files/papers/yogo-preprint.pdf

Egglog0[7]

- E-Graphs are a Database
- The database holds terms and equality relation
- Supports ordinary datalog with terms
- Pattern variables bind to eclasses
- Rules: query using RHS (e-matching multipattern), instantiate and insert LHS
- Special equality $=$ is E-graph Equality

$$\mathsf{add}(\mathsf{Y},\mathsf{X}) = \mathsf{E} \;:\!-\; \mathsf{add}(\mathsf{X},\mathsf{Y}) = \mathsf{E}.$$

- Queries e-match and return all results.

$$?\!-\; \mathsf{add}(\mathsf{succ}(\mathsf{zero}),\mathsf{succ}(\mathsf{Y})) = \mathsf{Z}.$$

[7] https://www.philipzucker.com/egglog/

Egg Multipatterns

- Upstreamed to egg[8].
- Multipatterns vs Guards
- Threads e-matching compiler env binding between patterns.

[8] https://github.com/egraphs-good/egg/pull/168

Demo

Example: Injectivity

- $\forall a, b. f(a) = f(b) \implies a = b$
- Example: Constructors, Negation, constant addition
- Unification

```
X = Y, Xs = Ys :- cons(X,Xs) = cons(Y,Ys).
X = Y :- X + Z = Y + Z.
```

Example: Memory Simplification[9]

- Alias Analysis + Simplification
- SMTlib theory of arrays
- Many SMT theories are expressible as Horn Clauses (side conditions)

```
//select grabs stored value
V <- select(A ,store(A, V, Mem)).

//select ignores different addresses
select(A1,Mem) = E :- select(A1,store(A2,V,Mem)) = E, neq(A1,A2).

//non aliasing writes commute
store(A2,V2, store(A1,V1,Mem)) = E :- store(A1,V1, store(A2,V2,Mem)) = E, neq(A1,A2).

//Aliasing Writes destroy old value.
store(A, V1, Mem) <- store(A, V1, store(A,V2,Mem)).
```

[9] https://www.philipzucker.com/egglog/?example=mem.pl

Example: Equation Solving

- Isolation
- Extract terms without variables

$$\mathsf{sub}(\mathsf{Z},\mathsf{X}) = \mathsf{Y} :- \mathsf{add}(\mathsf{X},\mathsf{Y}) = \mathsf{Z}$$

Example: Reflection

- Hypothetical reasoning
- Boolean algebraic reasoning

```
A = B :- true = eq(A,B).
true = eq(A,B) :- A = B.
```

Example: Uniqueness Quantification[10]

- Common in universal constructions in category theory
- Skolemize existentials $\forall x, P(x) \implies \exists y, Q(x, y)$ becomes $\forall x, P(x) \implies Q(x, f(x))$
- Uniqueness Property $\forall a, b. P(a) \wedge P(b) \implies a = b$

[10] https://www.philipzucker.com/egglog/?example=cat1.pl

Related Work

- Relational E-Matching [11]
- SMT Multipatterns
- Souffle Egg [12]
- Egg-lite

[11] https://arxiv.org/abs/2108.02290

[12] https://www.hytradboi.com/2022/writing-part-of-a-compiler-in-datalog

Questions?

- Thanks to Yihong Zhang, Remy Yisu Wang, Max Willsey, Zachary Tatlock, Alessandro Cheli, Cody Roux, James Fairbanks, and Evan Patterson for their helpful discussions.
- DARPA Grant