Homework 1

Tien Thuy Ho

April 20, 2018

Problem 1

1. Symmetric part R^*, where R is a binary relation:

   (a) Relational calculus
   $$\{(a, b) : R(a, b) \wedge R(b, a)\}$$

   (b) Relational algebra
   $$\pi_{1,2}(\sigma_{1=4 \wedge 2=3}(R \times R))$$
   or
   $$R \cap \pi_{2,1}(R)$$

2. Symmetric difference $R \Delta S$, where R and S are two ternary relations:

   (a) Relational calculus
   $$\{(x, y, z) : (R(x, y, z) \wedge \neg S(x, y, z)) \vee (S(x, y, z) \wedge \neg R(x, y, z))\}$$

   (b) Relational algebra
   $$(R \cup S) - (R \cap S)$$

3. Composition $R_1 \circ R_2$ where R_1 and R_2 are two binary relations:

   (a) Relational calculus
   $$\{(a, c) : \exists b\, (R_1(a, b) \wedge R_2(b, c))\}$$

   (b) Relational algebra
   $$\pi_{R_1.1, R_2.2}(\sigma_{R_1.2=R_2.1}(R_1 \times R_2))$$

Problem 2

Given a DIRECT database with two attributes FROM and TO:

1. Relational calculus
$$\begin{aligned}\{(c, d) : DIRECT(c, d) \\ &\vee (\exists x\, (DIRECT(c, x) \wedge DIRECT(x, d))) \\ &\vee (\exists x, y\, (DIRECT(c, x) \wedge DIRECT(x, y) \wedge DIRECT(y, d)))\}\end{aligned}$$

2. Relational algebra
$$\begin{aligned}DIRECT \vee \pi_{1,4}(\sigma_{2=3}(DIRECT \times DIRECT)) \\ &\vee \pi_{1,6}(\sigma_{2=3 \wedge 4=5}(DIRECT \times DIRECT \times DIRECT))\end{aligned}$$

Problem 3

Semijoin $R \rtimes S$ of two ternary relations R and S

1. Relational algebra
$$\pi_{R.A,R.B,R.C}(\sigma_{R.B=S.B \wedge R.C=S.C}(R \times S))$$

2. SQL

```
SELECT R.A, R.B, R.C
FROM R, S
WHERE R.B = S.B AND R.C = S.C
```

Problem 4

1. $\pi_{1,2}(R) \cup \pi_{1,2}(S) = \pi_{1,2}(R \cup S)$
   (a) ($\Rightarrow$) Let $x \in \pi_{1,2}(R) \cup \pi_{1,2}(S)$, we need to show that $x \in \pi_{1,2}(R \cup S)$. Suppose $(a,b) \in \pi_{1,2}(R) \cup \pi_{1,2}(S)$, then $(a,b) \in \pi_{1,2}(R)$ or $(a,b) \in \pi_{1,2}(S)$. Without loss of generality, let's assume that R and S are ternary relations. So $\exists c\ (a,b,c) \in R$ or $(a,b,c) \in S$. Then $(a,b,c) \in (R \cup S)$, and it follows that $\pi_{1,2}(a,b,c) \in \pi_{1,2}(R \cup S)$. Therefore, $(a,b) \in \pi_{1,2}(R \cup S)$.
   (b) ($\Leftarrow$) Let $x \in \pi_{1,2}(R \cup S)$, we need to show that $x \in \pi_{1,2}(R) \cup \pi_{1,2}(S)$. Suppose $(a,b) \in \pi_{1,2}(R \cup S)$. Without loss of generality, let's assume that R and S are ternary relations. Then, $R \cup S$ is also a ternary relation, and $\exists c\,(a,b,c) \in (R \cup S)$. It follows that $(a,b,c) \in R$ or $(a,b,c) \in S$. So $(a,b) \in \pi_{1,2}(R)$ or $(a,b) \in \pi_{1,2}(S)$ by the projection definition of $\pi_{1,2}(a,b,c) = (a,b)$. This proves that $(a,b) \in \pi_{1,2}(R) \cup \pi_{1,2}(S)$.

2. $\pi_{1,2}(R) \cap \pi_{1,2}(S) \neq \pi_{1,2}(R \cap S)$
   Counter-example: Suppose $R = \{(a,b,c),(m,n,h)\}$ and $S = \{(a,b,l),(m,n,h)\}$. Both relations have the same arity of three. We have:
$$\pi_{1,2}(R) \cap \pi_{1,2}(S) = \{(a,b),(m,n)\} \cap \{(a,b),(m,n)\} = \{(a,b),(m,n)\}$$
   whereas
$$\pi_{1,2}(R \cap S) = \pi_{1,2}\{(m,n,h)\} = \{(m,n)\}$$
   The two sides are not equal.

3. $\pi_{1,2}(R) - \pi_{1,2}(S) \neq \pi_{1,2}(R - S)$
   Counter-example: Suppose $R = \{(a,b,c),(m,n,h)\}$ and $S = \{(a,b,l),(m,n,h)\}$. Both relations have the same arity of three. We have:
$$\pi_{1,2}(R) - \pi_{1,2}(S) = \{(a,b),(m,n)\} - \{(a,b),(m,n)\} = \{\}$$
   whereas
$$\pi_{1,2}(R - S) = \pi_{1,2}\{(a,b,c)\} = \{(a,b)\}$$
   The two sides are not equal.

Problem 5

Acknowledgment: Professor Kolaitis

1. **Claim**: Difference operation cannot be expressed in terms of union, cartesian product, projection, and selection.
   Observation: Difference operation is the only one whose output size can decrease if its input size increases. For instance, if we take the difference $R - S$, where R and S have the same arity, and increase the size of S by some x number of tuples, there are two cases:

   (a) x does not add to $|R \cap S|$, then $|R - S|$ remains the same.

   (b) x adds to $|R \cap S|$, then $|R - S| = |R| - |R \cap S|$ decreases.

   We need to show that the output of every relational algebra expression in the other four operations does not decrease if its input size increases.
   Proof: We perform a structural induction proof over the context-free grammar for relational algebra expressions:

   $$E := R, S|(E_1 \cup E_2)|(E_1 \times E_2)|\pi_X(E)|\sigma_\theta(E)$$

   where R, S are k-ary relations, X is a list of attributes, and θ is a condition. By induction on the length of E, we assume that any expression E' smaller than E does not decrease its output size if its input size increases. There are six cases to consider:

   - **Case 1 (Relation)**: $E = R$. It is trivial that $|E|$ increases if $|R|$ increases.
   - **Case 2 (Relation)**: $E = S$. It is trivial that $|E|$ increases if $|S|$ increases.
   - **Case 3 (Union)**: $E = E_1 \cup E_2$
     If we increase the input size of E, then by the induction hypothesis, $|E_1|$ and $|E_2|$ either stay the same or increase. Since $|E_1 \cup E_2| = |E_1| + |E_2| - |E_1 \cap E_2|$. There are four cases to consider:

     (a) $|E_1|$ and $|E_2|$ stay the same, then $|E_1 \cap E_2|$ stays the same and $|E_1 \cup E_2|$ does not change.

     (b) $|E_1|$ increases by some $x > 0$. Then there are three possible consequences:

       i. $|E_1 \cap E_2|$ remains the same. Then $|E_1| + x + |E_2| - |E_1 \cap E_2| > |E_1| + |E_2| - |E_1 \cap E_2|$, and $|E_1 \cup E_2|$ increases by $x > 0$.

       ii. $|E_1 \cap E_2|$ increases by x. Then $|E_1| + x + |E_2| - (|E_1 \cap E_2| + x) = |E_1| + |E_2| - |E_1 \cap E_2|$. Therefore, $|E_1 \cup E_2|$ stays the same.

       iii. $|E_1 \cap E_2|$ increases by some positive $m < x$. Then $|E_1| + x + |E_2| - (|E_1 \cap E_2| + m) = |E_1| + |E_2| - |E_1 \cap E_2| + (x - m) > |E_1| + |E_2| - |E_1 \cap E_2|$. Then $|E_1 \cup E_2|$ increases by $(x - m)$.

     (c) $|E_2|$ increases (same argument as (b)).

     (d) $|E_1|$ increases by $x > 0$ and $|E_2|$ increases by $y > 0$. Again, we have three possible consequences:

       i. $|E_1 \cap E_2|$ remains the same. Then $|E_1| + x + |E_2| + y - |E_1 \cap E_2| > |E_1| + |E_2| - |E_1 \cap E_2|$, and $|E_1 \cup E_2|$ increases by $(x + y)$.

       ii. $|E_1 \cap E_2|$ increases by $(x + y)$. Then $|E_1| + x + |E_2| + y - (|E_1 \cap E_2| + x + y) = |E_1| + |E_2| - |E_1 \cap E_2|$. Therefore $|E_1 \cup E_2|$ stays the same.

       iii. $|E_1 \cap E_2|$ increases by some positive $m < (x + y)$. Then $|E_1| + x + |E_2| + y - (|E_1 \cap E_2| + m) = |E_1| + |E_2| - |E_1 \cap E_2| + (x + y - m) > |E_1| + |E_2| - |E_1 \cup E_2|$. Then $|E_1 \cup E_2|$ increases by $(x + y - m)$.

In all cases, $|E|$ either stays the same or increases when its inputs increase.

- **Case 4 (Cartesian product)**: $E = E_1 \times E_2$
  If we increase the input size of E, then by the induction hypothesis, $|E_1|$ and $|E_2|$ either stay the same or increases. Since $|E_1 \times E_2| = |E_1| \times |E_2|$, $|E|$ either stays the same if both $|E_1|$ and $|E_2|$ do not change or increases if either or both $|E_1|$ and $|E_2|$ increase.

- **Case 5 (Projection)**: $E = \pi_{i_1,\ldots,i_m}(E_1)$
  If we increase the input size of E, then by the induction hypothesis, $|E_1|$ either stays the same or increase. There are four cases to consider:

  (a) $|E_1|$ either stays the same, then $|\pi_{i_1,\ldots,i_m}(E_1)|$ stays the same.
  (b) $|E_1|$ increases with non-duplicate values in the projected attributes, then $|\pi_{i_1,\ldots,i_m}(E_1)|$ increases.
  (c) $|E_1|$ increases with some duplicate values in the projected attributes, then $|\pi_{i_1,\ldots,i_m}(E_1)|$ still increases.
  (d) $|E_1|$ increases with all duplicate values in the projected attributes, then $|\pi_{i_1,\ldots,i_m}(E_1)|$ does not change.

  In all cases, $|E|$ either stays the same or increases when its inputs increase.

- **Case 6 (Selection)**: $E = \sigma_\theta(E_1)$
  If we increase the input size of E, then by the induction hypothesis, $|E_1|$ either stays the same or increases. If $|E_1|$ remains the same, then $|\sigma_\theta(E_1)|$ does not change. On the other hand, if $|E_1|$ increases, then there are more candidates for the tested condition θ. It follows that $|E|$ either remains the same if all the new tuples fail the condition θ or increases if some new tuples satisfy the condition.

2. **Claim**: If relational algebra expressions are built using union, intersection, cartesian product, projection, and selection, then this version of relational algebra is not relationally complete.
   Justification: Let's call this new algebra A. By the definition of relational completeness, every expression E in relational algebra should have an equivalent in A. In other words, the objective is to find the equivalence of the five basic operations (difference, union, cartesian product, projection, selection) in A. We want to represent the difference operation in A. We showed in (1) that the difference operation cannot be expressed in terms of union, cartesant product, projection, and selection. We only need to show that the intersection operation cannot express difference either. As we have argued in (1), one unique property of difference is that its output size might decrease if its input size increases. By induction on the length of E, we assume that any expression E' smaller than E does not decrease its output size if its input size increases. We need to show that all the expressions $E = E_1 \cap E_2$ do not decrease their output size if their input size increases. By the induction hypothesis, $|E_1|$ and $|E_2|$ either stay the same or increase if we increase the input size of E. There are four cases to consider:

   (a) $|E_1|$ and $|E_2|$ stay the same, then $|E_1 \cap E_2|$ stays the same.
   (b) $|E_1|$ increases by some $x > 0$:
      i. Some of x tuples are common tuples between E_1 and E_2, then $|E_1 \cap E_2|$ increases.
      ii. None of x tuples are common tuples between E_1 and E_2, then $|E_1 \cap E_2|$ stays the same.
   (c) $|E_2|$ increases (same argument as (b)).
   (d) $|E_1|$ increases by $x > 0$ and $|E_2|$ increases by $y > 0$:

i. Some of $(x+y)$ tuples are common tuples between E_1 and E_2, then $|E_1 \cap E_2|$ increases.

ii. None of $(x+y)$ tuples are common tuples between E_1 and E_2, then $|E_1 \cap E_2|$ stays the same.

Hence, in no way does the intersection operation decrease its output size when its inputs increase their size. In other words, intersection cannot express the difference operation. Since there does not exist any operation in A that could express the difference operation, this new version of algebra is not relationally complete.

3. **Claim**: Quotient operation cannot be expressed in terms of union, cartesian product, projection, and selection alone.
Proof: Without loss of generality, let's consider the quotient operation $R \div S$ on a 5-ary relation R and a binary relation S. By the definition of quotient, $R \div S$ only includes $(a_1, a_2, a_3) \in \pi_{1,2,3}(R)$ if $\forall (b_1, b_2) \in S$, $(a_1, a_2, a_3, b_1, b_2) \in R$. In other words, we need to find all the tuples (a_1, a_2, a_3) in R that are paired up with every tuple in S. Suppose $|S| = m$, and there are n tuples in R that satisfy the condition for $R \div S$. Let's consider a new S^* by adding a new tuple (b_1^*, b_2^*) to S. It follows that we only need to consider among the n tuples to see which ones are also matched up with (b_1^*, b_2^*). We do not have to consider the other tuples in R because they already failed to get matched up with the original m tuples in S. In other words, $R - S^*$ can only output at most n tuples resulted from $R - S$. Certainly, there might be tuples (a_1, a_2, a_3) among these n candidates where $(a_1, a_2, a_3, b_1^*, b_2^*) \notin R$ and so are not included in $R - S^*$. Hence, $|R - S^*| \leq n = |R - S|$. In conclusion, the output of the quotient operation may decrease if its input size increases. We have showed in (1) that the output of every relational algebra expression using union, cartesian product, projection, and selection does not decrease if its input size increases. Therefore, none of these operations and their composition can express the quotient operation.

Challenge

Acknowledgment: `https://mathoverflow.net/questions/112004/why-is-a-union-operation-independent-in-relational-algebra`

Claim: Union operation cannot be expressed in terms of union, cartesian product, projection, and selection.
Observation: By definition of union, the two arguments must have the same arity but need not have the same attributes. When taking a union $R \cup S$, where R and S have the same arity of k, there are $2k$ attributes to consider. The only way for $R \cup S$ to result in a new k-ary relation is to combine two attributes, which might be different, into one. We need to show that the output of every relational algebra expression in the other four operations does not combine two separate attributes into one.
Proof: We perform a structural induction proof over the context-free grammar for relational algebra expressions:

$$E := R, S|(E_1 - E_2)|(E_1 \times E_2)|\pi_X(E)|\sigma_\theta(E)$$

where R, S are k-ary relations whose attributes are different, X is a list of attributes, and θ is a condition. By induction on the length of E, we assume that any expression E' smaller than E results in a new relation whose values are associated with the same original attributes of the inputs. There are six cases to consider:

- **Case 1 (Relation)**: $E = R$. It is trivial that all the attributes of E contain the original values.

- **Case 1 (Relation)**: $E = S$. It is trivial that all the attributes of E contain the original values.

- **Case 3 (Difference)**: $E = E_1 - E_2$
By the induction hypothesis, all the values in E_1 and E_2 are associated with the same original attributes of the inputs. It follows that $E_1 - E_2$ only removes tuples in E_1 that are also in E_2. Hence, no two attributes are combined, and all the values in E remain the same attributes that they were in before the operation.

- **Case 4 (Cartesian product)**: $E = E_1 \times E_2$
By the induction hypothesis, all the values in E_1 and E_2 are associated with the same original attributes of the inputs. By the definition of cartesian product, we have:

$$E_1 \times E_2 = \{(a_1, ..., a_m, b_1, ..., b_n) : (a_1, ..., a_m) \in E_1 \wedge (b_1, ..., b_n) \in E_2\}$$

where m is the arity of E_1 and n is the arity of E_2. It follows that $E_1 \times E_2$ results in a $(m+n)$-ary relation, and no two attributes are merged. Hence, all the values in E remain the same attributes that they were in before the operation.

- **Case 5 (Projection)**: $E = \pi_{i_1,...,i_m}(E_1)$
By the induction hypothesis, all the values in E_1 are associated with the same original attributes of the inputs. It follows $\pi_{i_1,...,i_m}(E_1)$ only removes the attributes in E_1 which are not listed in the set $\{i_1, ..., i_m\}$. No two attributes are combined, and hence, all the values in E remain in the same attributes that they were in before the operation.

- **Case 6 (Selection)**: $E = \sigma_\theta(E_1)$
By the induction hypothesis, all the values in E_1 are associated with the same original attributes of the inputs. It follows that $\sigma_\theta(E_1)$ only removes tuples that do not satisfy θ. No two attributes are combined, and hence, all the values in E remain in the same attributes that they were in before the operation.