Relational Algebra

DSC 301: Lecture 3

January 29, 2021

Abstract

Relational databases are based on the mathematical relation, which is the primary topic of this lecture. We begin with elementary set theory notions, including basic operations. The Cartesian product is of utmost importance as it provides the foundation of relational database theory. Finally we examine some algebra on relations.

1 Lecture Objectives

- Set theory (background of relations and relational algebras)
  - Cartesian product
- Define relation
- Define relational algebras and some examples

2 Set Theory

- A set X is a collection of "things". For example, $X = \{1, 2, 3\}$, or $Y = \{\text{cat}, \text{dog}\}$.
- The universe of discourse, or **universal set**, U, is a collection of all objects under consideration. For example, in first-semester calculus, $U = \mathbb{R} =$ set of real numbers.
- A set A is a **subset** of B, written as $A \subseteq B$ if every element of A is an element of B.

  If $A = \{1, 2\}$ and $B = \{1, 2, 3, 4\}$, then $A \subseteq B$.

Set Operations

Let A and B be sets.

- The **union** of A and B is all the elements in A or B,

$$A \cup B = \{x : x \in A \text{ or } x \in B\}$$

  Example: $A = \{1, 2, 3\}$ and $B = \{3, 4\}$, then $A \cup B = \{1, 2, 3, 4\}$.

- The **intersection** of A and B is all elements contained in both A and B.

$$A \cap B = \{x : x \in A \text{ and } x \in B\}$$

  Example: $A = \{1, 2, 3\}$ and $B = \{3, 4\}$, then $A \cap B = \{3\}$.

- The **difference** of A and B is the set with elements in A but not in B.

$$A - B = \{x : x \in A \text{ and } x \notin B\}$$

  Example: $A = \{1, 2, 3\}$ and $B = \{3, 4\}$, then $A - B = \{1, 2\}$.
  Example: $A = \{1, 2, 3\}$ and $B = \{4, 5\}$, then $A - B = \{1, 2, 3\}$.

- The **compliment** of A is the set $A^c = U - A$.
  Example: $U = \{1, 2, 3, 4, 5\}$ and $A = \{2, 5\}$, and so $A^c = \{1, 3, 4\}$

- Let A_1 and A_2 be sets. The set

$$A_1 \times A_2 = \{(a_1, a_2) \mid a_1 \in A_1 \ \wedge \ a_2 \in A_2\}$$

  is the **Cartesian product** of A_1 and A_2.
  Example: If $A_1 = \{1, 2, 3\}$ and $A_2 = \{3, 4\}$, then,

$$A_1 \times A_2 = \{(1, 3), (1, 4), (2, 3), (2, 4), (3, 3), (3, 4)\}.$$

Note: * Letter A was chosen to represent “attribute”. If D was selected, for “domain”. However, any letter will due.

Note: * The Cartesian product can be generalized to the product of n sets, i.e.,

$$A_1 \times A_2 \times \cdots \times A_n$$

n-ary product or degree n.

---

3 Relational Algebra

Definition 1. *A **relation** is a subset of a Cartesian product. Relations are said to be degree n.*

Recall: Relations provide the framework for relational databases (such as MySQL, Oracle, MS SQL, etc.).

Question: What is algebra?

Question: Are there more than one algebra? Yes.

Definition 2. *A (mathematical) structure consisting of a set with operations that follow (obey) some rules (or constraints).*

Example: the set could be the integers, operations addition and multiplication, and rules associative rule, commutative rule, distributive property etc.
Example: the set of equilateral triangles, operations rotation and reflection, and rules associative.

Perhaps the most important rule (property) is **closure** - the operation on two elements in the set must be another element of the set.

Definition 3. *A **relational algebra** defines a set of operations on relations (i.e., tuples).*

Given a set of tuples (i.e., a relation), we will perform operations on those tuples and we will again get a set of tuples.

Let R and S be relations, then $T = R * S$ and get another relation.

Relational Operations

- **Rename**: $\rho_S(R) \to S$, where we rename relation R to S. Also, $\rho_{R(A_1,A_2,\ldots,A_n)}(R)$ renames the attributes of R to $A_1, A_2, \ldots, A_n$.

Table 1: Rename relation R to S, $\rho_S(R)$

S =

A	B	C	D
1	1	1	1
2	1	3	1
3	2	3	4
4	1	3	1

Table 2: Rename Relation R, $\rho_{S(A_1,A_2,A_3,A_4)}(R)$

S =

A_1	A_2	A_3	A_4
1	1	1	1
2	1	3	1
3	2	3	4
4	1	3	1

- **Select**: $\sigma_p(R)$
  - p is a selection predicate and R is a relation
  - Defined as:
    $$\sigma_p(R) = \{t \mid t \in R \land p(t) \text{ is true}\}$$
  - Select output row (records), basically a filter
  - Subset of records

Table 3: Select: $\sigma_{A_1>2}(R)$

R =

A_1	A_2	A_3	A_4
1	1	1	1
2	1	3	1
3	2	3	4
4	1	3	1

$\sigma_{A_1>2}(R) =$

A_1	A_2	A_3	A_4
3	2	3	4
4	1	3	1

- **Projection**: $\Pi_{A_1,A_2,\ldots,A_k}(R)$
  - A Subset of columns
  - Removes duplicate tuples from the output
    - Note how duplicate tuples could arises

Table 4: Projection: $\Pi_{A_2,A_3}(R)$

R =

A_1	A_2	A_3	A_4
1	1	1	1
2	1	3	1
3	2	3	4
4	1	3	1

$\Pi_{A_2,A_3}(R) =$

A_2	A_3
1	1
1	3
2	3