

# Query Flocks: A Generalization of Association-Rule Mining

# Background

- Association Rule Mining, a effective method for Extracting Information.
- Apriori Algorithm a Efficent algorithm for faster Association Rules mining.
- Query Flocks a generalized version of Market Basket Problem
- Query Flocks: parametrized queries with a filter condition to eliminate values of the parameters that are “uninteresting”

# Review of Market Basket Mining

- Is an attempt by a retailstore to learn what items its customers frequently purchase together.
- We shall assume for simplicity that the database is a relation baskets(BID,Item)
- The goal of market basket analysis is to find of items that are associated together.
- Metrics Such as Support and Confidence are Used.

# The Problem With SQL as a Mining Language

```
SELECT i1.Item, i2.Item
FROM baskets i1, baskets i2
WHERE i1.Item < i2.Item AND i1.BID = i2.BID
GROUP BY i1.Item, i2.Item
HAVING 20 <= COUNT(i1.BID)
```

Figure 1: Searching for association rules using SQL

# Query Flocks

- A query flock is a generate-and-test system, in which a family of queries that are identical except for the values of one or more “parameters” are asked simultaneously.
- The answers to these queries are filtered and those that pass the filter test enable their parameters to become part of the answer to the query flock.

# Components Of Query Flock

1. One or more predicates that represent data stored as relations.
2. A set of parameters, which we shall always denote with names beginning with \$.
3. A query expressed in our query language, using the parameters in roles normally reserved for constants.
4. A filter that specifies a condition that the result of the query must satisfy in order for a given assignment of values to the parameters to be acceptable.

# Example Of a Query Flock

We are given a relation baskets(BID,Item) as underlying data,

QUERY:

answer(B) :- baskets(B,\$1) AND baskets(B,\$2)

FILTER:

COUNT(answer.B) >= 20

Figure 2: Market basket association rules as a query flock

1. diagnoses(Patient, Disease): The patient has been diagnosed as having the disease.
2. exhibits(Patient, Symptom): The patient exhibits the symptom.
3. treatments(Patient, Medicine): The medicine has been prescribed for the patient.
4. causes(Disease, Symptom): The disease is known to cause the symptom.

QUERY:

answer(P) :-

exhibits(P,\$s) AND

treatments(P,\$m) AND

diagnoses(P,D) AND

NOT causes(D,\$s)

FILTER:

COUNT(answer.P) >= 20

Figure 3: Mining for side-effects in a medical database