**Apriori Algorithm**

Support:

We go to a shopping mall to purchase goods (items) and those items show on即your list to indicate the frequency. We use Support(X) to represent the behavior of transaction situations, X means item set. For example, if we have 100 transactions and among them have 40 times for buying milk, so the support of item set is 40/100=40%.

Confidence:

Confidence refers to the degree of trust between two project groups, that is, the conditional probability that Y will also appear under the probability that X appears. Usually we use Support(X∩Y) /Support(X) to calculate the Confidence. X and Y are the item sets.

We have four transactions and set up Min Support (threshold) = 2; Min Confidence (threshold) = 80%

|  |
| --- |
| **Transaction** |
| **ABD** |
| **BCE** |
| **ABCE** |
| **BE** |

1 Item Set

|  |  |
| --- | --- |
| A | 2 |
| B | 4 |
| C | 2 |
| D | 1 |
| E | 3 |

2 Item Set

|  |  |
| --- | --- |
| AB | 2 |
| BC | 2 |
| BE | 3 |
| CE | 2 |
| AC | 1 |
| EA | 1 |

3 Item Set

|  |  |
| --- | --- |
| BCE | 2 |

Calculating 2 Item Set Confidence.

For example: A 🡪 C = P(C|A) = Support(A∩C) /Support(A)

|  |  |  |  |
| --- | --- | --- | --- |
| AB | P(B|A) | 2/2 | 100% |
| BA | P(A|B) | 2/4 | 50% |
| BC | P(C|B) | 2/4 | 50% |
| CB | P(B|C) | 2/4 | 100% |
| BE | P(E|B) | 3/4 | 75% |
| EB | P(B|E) | 3/3 | 100% |
| CE | P(E|C) | 2/2 | 100% |
| EC | P(C|E) | 2/3 | 66.7% |

3 Item Set Confidence

|  |  |  |  |
| --- | --- | --- | --- |
| B 🡪 CE | P(CE|B) | 2/4 | 50% |
| C 🡪 BE | P(BE|C) | 2/2 | 100% |
| E 🡪 BC | P(BC|E) | 2/3 | 66.7% |
| CE 🡪 B | P(B|CE) | 2/2 | 100% |
| BE 🡪 C | P(C|BE) | 2/3 | 66.7% |
| BC 🡪 E | P(E|BC) | 2/2 | 100% |

|  |
| --- |
| **Transaction** |
| **ABD** |
| **BCE** |
| **ABCE** |
| **BE** |

Answer:

A🡪B；C🡪B；E🡪B；C🡪E；C🡪BE；CE 🡪 B；BC 🡪 E