

The background of the slide is a complex, abstract composition. It features a dark, reddish-brown base with a network of thin, light-colored lines forming a mesh or web-like structure. Overlaid on this are various data visualization elements: a grid of small, light-colored plus signs, a series of small, colorful dots (green, blue, yellow) connected by lines, and a large, semi-transparent white triangle that serves as a backdrop for the title. In the bottom left corner, there is a small, rectangular inset image showing a cluster of orange and red dots on a light blue background, with a white grid overlay.

Lecture 5. Mining Diverse Patterns

Lecture 5. Mining Diverse Patterns

- ❑ Mining Multiple-Level Associations
- ❑ Mining Multi-Dimensional Associations
- ❑ Mining Quantitative Associations
- ❑ Mining Negative Correlations
- ❑ Mining Compressed and Redundancy-Aware Patterns
- ❑ Mining Long/Colossal Patterns

The background of the slide is a complex, abstract composition. It features a network of thin, light-colored lines forming a web-like structure. Overlaid on this are various data visualizations: a grid of small plus signs in the top-left, a series of purple arrows pointing left in the top-center, a large network of green nodes connected by red lines in the top-right and bottom, and a cluster of orange and brown nodes on the left side. The overall color palette is muted, with earthy tones and soft pastels.

Session 1: Mining Multi-Level Associations

Mining Multiple-Level Frequent Patterns

- Items often form hierarchies

- Ex.: Dairyland 2% milk;
Wonder wheat bread

- How to set min-support thresholds?

- Uniform min-support across multiple levels (reasonable?)

- Level-reduced min-support: Items at the lower level are expected to have lower support

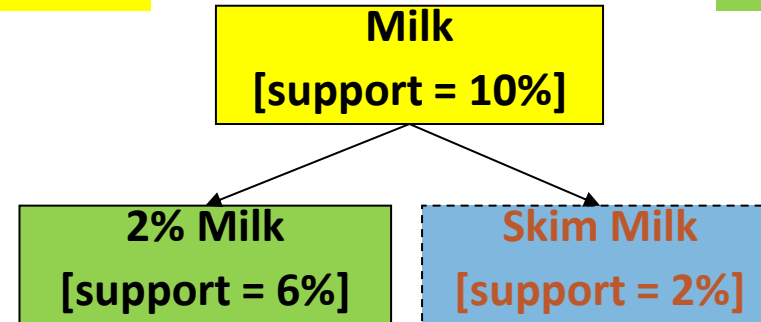
- Efficient mining: *Shared* multi-level mining

- Use the lowest min-support to pass down the set of candidates

Uniform support

Level 1
min_sup = 5%

Level 2
min_sup = 5%



Reduced support

Level 1
min_sup = 5%

Level 2
min_sup = 1%

Redundancy Filtering at Mining Multi-Level Associations

- ❑ Multi-level association mining may generate many redundant rules

- ❑ Redundancy filtering: Some rules may be redundant due to “ancestor” relationships between items

(Suppose the 2% milk sold is about $\frac{1}{4}$ of milk sold in gallons)

- ❑ milk \Rightarrow wheat bread [support = 8%, confidence = 70%] (1)

- ❑ 2% milk \Rightarrow wheat bread [support = 2%, confidence = 72%] (2)

- ❑ A rule is *redundant* if its support is close to the “expected” value, according to its “ancestor” rule, and it has a similar confidence as its “ancestor”

- ❑ Rule (1) is an ancestor of rule (2), which one to prune?

Customized Min-Supports for Different Kinds of Items

- ❑ We have used the same min-support threshold for all the items or item sets to be mined in each association mining
- ❑ In reality, some items (e.g., diamond, watch, ...) are valuable but less frequent
- ❑ It is necessary to have customized min-support settings for different kinds of items
- ❑ One Method: Use **group-based “individualized” min-support**
 - ❑ E.g., {diamond, watch}: 0.05%; {bread, milk}: 5%; ...
 - ❑ How to mine such rules efficiently?
 - ❑ Existing scalable mining algorithms can be easily extended to cover such cases