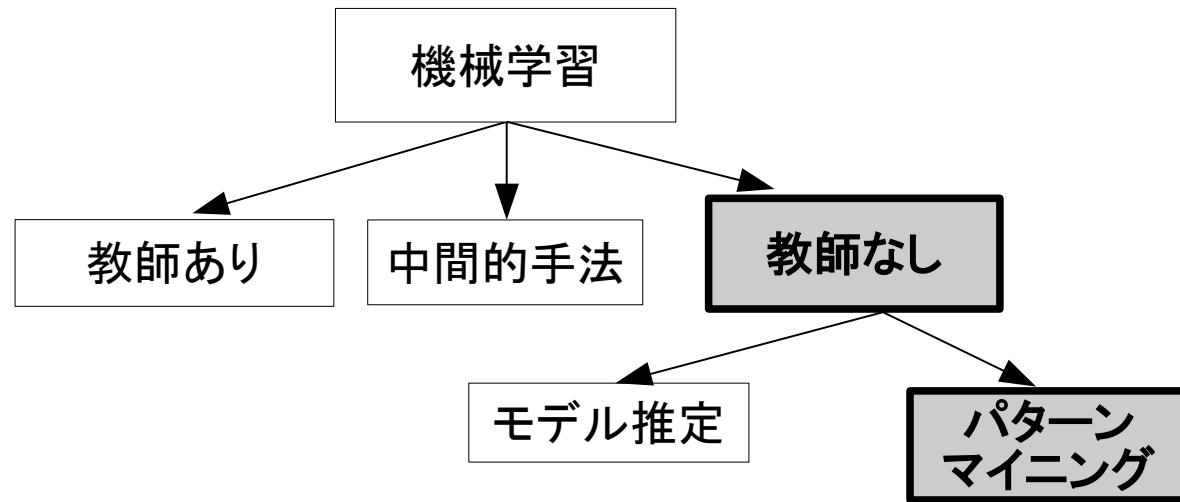


11 章 パターンマイニング

- パターンマイニングの問題設定
 - 入力：ラベル特徴の教師なしデータ
 - 出力：頻出項目、連想規則



No.	ミルク	パン	バター	雑誌
1	t	t		
2		t		
3				t
4		t	t	
5	t	t	t	
6	t	t		

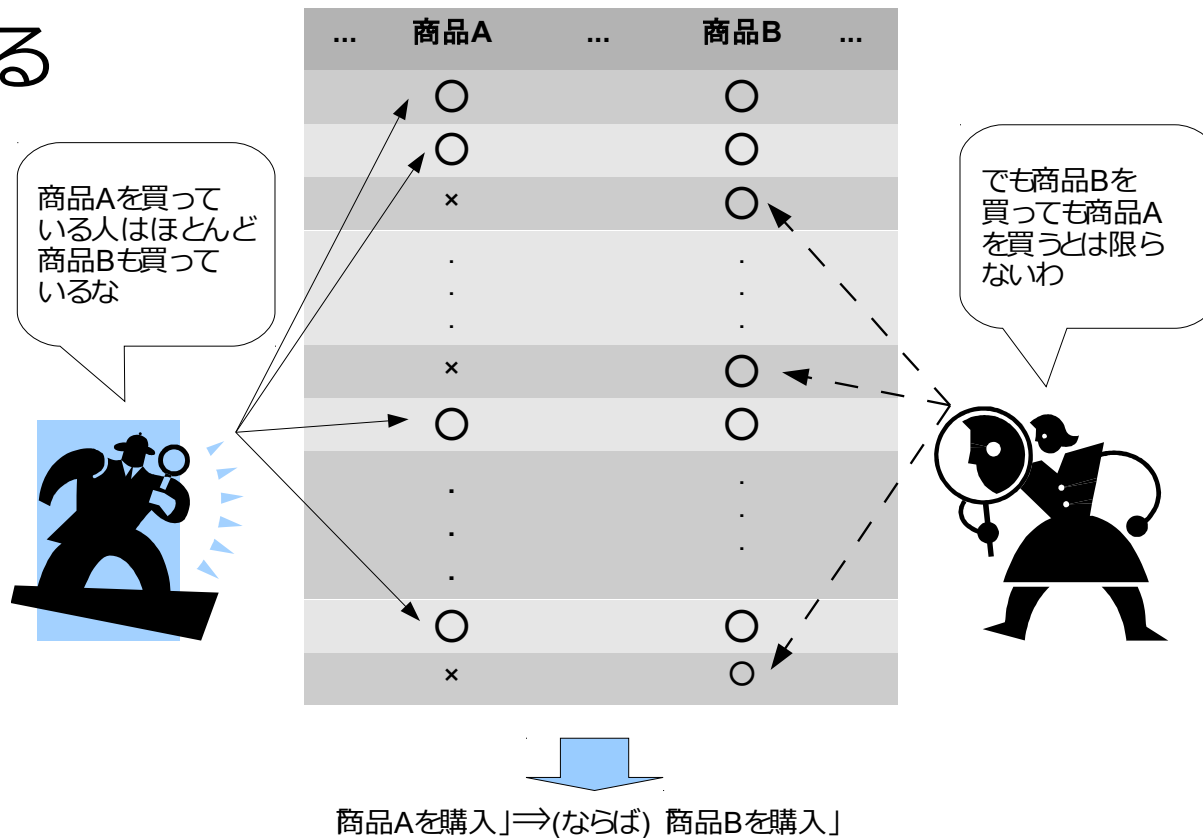
問題の定義

- 学習データ

$$\{\mathbf{x}^{(i)}\} \quad i = 1, \dots, N$$

- 問題設定

- データ集合中で、一定頻度以上で現れるパターンを抽出する



11.2 Apriori アルゴリズムによる頻出項目抽出

- 例題：バスケット分析

No.	ミルク	パン	バター	雑誌
1	t	t		
2		t		
3				t
4		t	t	
5	t	t	t	
6	t	t		

- 支持度
 - 全データ T に対して、ある項目集合が出現する割合
 - 支持度が閾値以上の項目集合を抽出したい

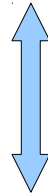
$$\text{support}(\text{items}) = \frac{T_{\text{items}}}{T}$$

項目集合の種類数は、
2 の商品数乗

11.2 Apriori アルゴリズムによる頻出項目抽出

- a priori な原理

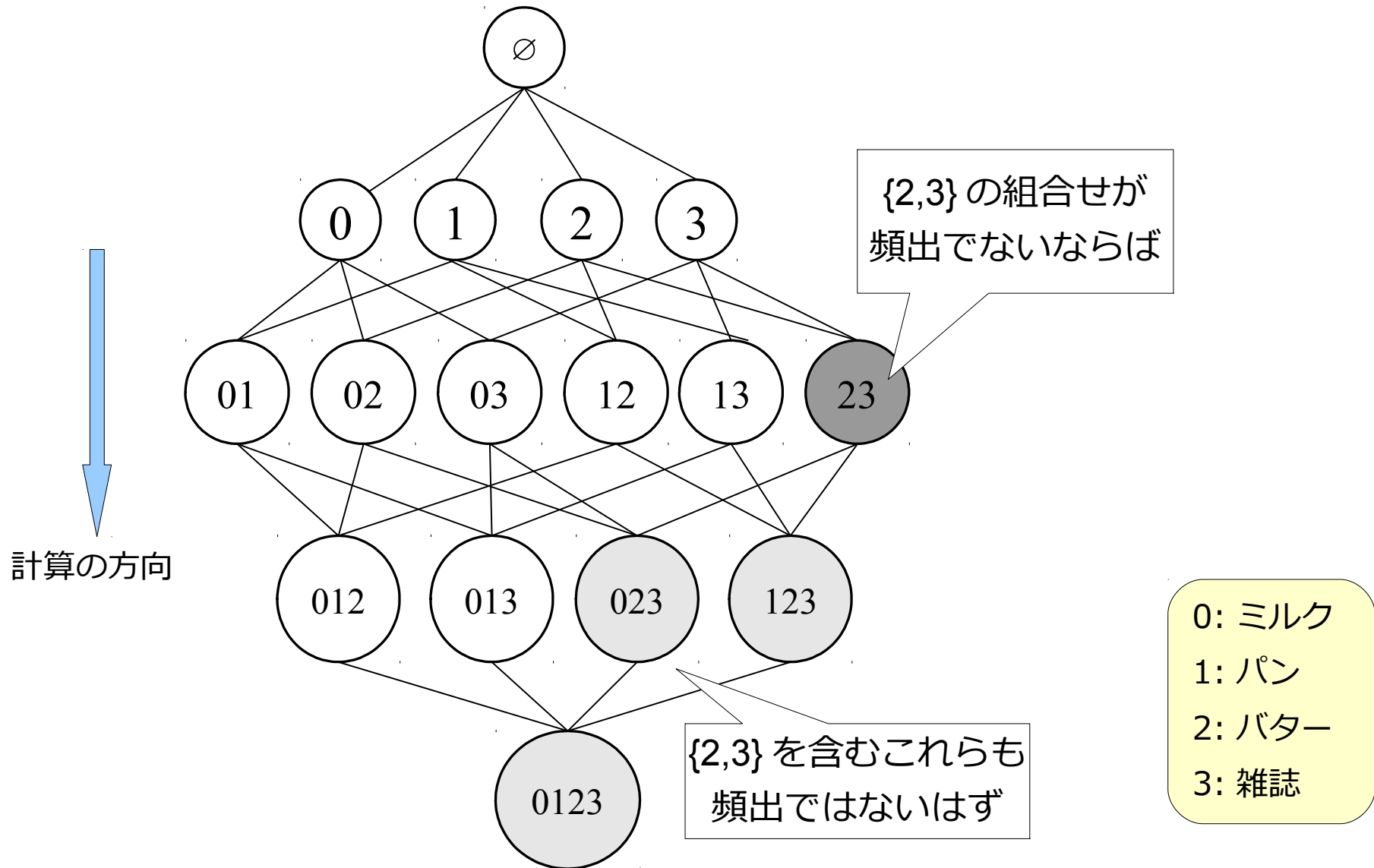
ある項目集合が頻出ならば、その部分集合も頻出である



対偶

ある項目集合が頻出でないならば、
その項目集合を含む集合も頻出でない

11.2 Apriori アルゴリズムによる頻出項目抽出



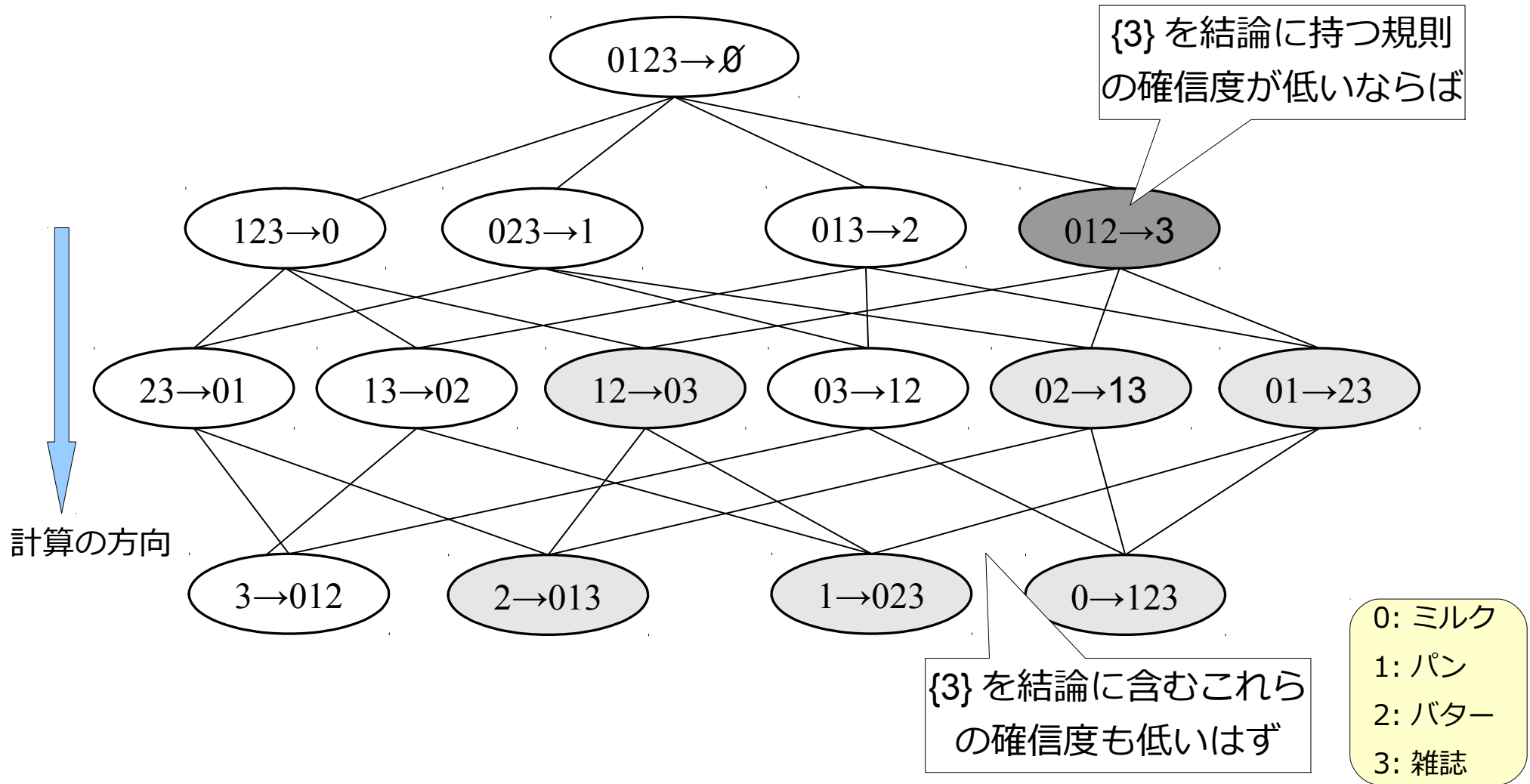
11.3 連想規則抽出

- 連想規則
 - 頻出項目集合を、条件部と結論部に分ける
 - 確信度の高い規則を抽出

$$\text{confidence}(A \rightarrow B) = \frac{\text{support}(A \cup B)}{\text{support}(A)}$$

- a priori 原理による探索

11.3 連想規則抽出



supermarket データ

Viewer										
Relation: supermarket										
12: baby needs Nominal	13: bread and cake Nominal	14: baking needs Nominal	15: coupons Nominal	16: juice-sat-cord-ms Nominal	17: tea Nominal	18: biscuits Nominal	19: canned fish-meat Nominal	20: canned fruit Nominal		
	t	t			t	t	t		t	▲
	t			t						
	t			t		t		t	t	≡
t	t	t		t		t		t	t	
	t	t		t		t				
				t	t	t		t		
				t		t				
	t	t		t	t	t		t	t	
	t	t		t		t		t	t	
	t	t								
	t				t	t			t	
t				t						
	t	t								
		t			t	t		t		
		t			t	t			t	
	t			t				t	t	
	t	t				t			t	
	t	t			t	t				
	t	t								
	t	t			t	t				
	t									
	t			t					t	
	t	t		t	t	t	t		t	
t	t	t		t		t	t	t	t	
t	t	t				t		t	t	
	t									
	t			t			t			
	t	t		t		t			t	
	t			t		t				
	t	t		t		t	t			
t		t		t		t		t	t	
	t									▼
	t									

◀
|||
▶

Undo
OK
Cancel

Weka の Apriori

The screenshot shows the Weka Explorer application window. The 'Associate' tab is selected, and the 'Apriori' algorithm is chosen. The 'Start' button has been clicked, and the 'Associator output' pane displays the results of the Apriori algorithm. The 'Result list (right-click to open)' pane shows a single entry: '17:53:31 - Apriori'. The 'Status' bar at the bottom indicates 'OK'.

Weka Explorer

Preprocess Classify Cluster **Associate** Select attributes Visualize

Associator

Choose Apriori -N 10 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.1 -S -1.0 -c -1

Start Stop

Result list (right-click to open)

17:53:31 - Apriori

Associator output

Minimum support: 0.15 (694 instances)
Minimum metric <confidence>: 0.9
Number of cycles performed: 17

Generated sets of large itemsets:

Size of set of large itemsets L(1): 44
Size of set of large itemsets L(2): 380
Size of set of large itemsets L(3): 910
Size of set of large itemsets L(4): 633
Size of set of large itemsets L(5): 105
Size of set of large itemsets L(6): 1

Best rules found:

1. biscuits=t frozen foods=t fruit=t total=high 788 ==> bread and cake=t 723 <conf:(0.92)> lift:(1.27) lev:(0.04)
2. baking needs=t biscuits=t fruit=t total=high 760 ==> bread and cake=t 696 <conf:(0.92)> lift:(1.27) lev:(0.04)
3. baking needs=t frozen foods=t fruit=t total=high 770 ==> bread and cake=t 705 <conf:(0.92)> lift:(1.27) lev:(0.04)
4. biscuits=t fruit=t vegetables=t total=high 815 ==> bread and cake=t 746 <conf:(0.92)> lift:(1.27) lev:(0.04)
5. party snack foods=t fruit=t total=high 854 ==> bread and cake=t 779 <conf:(0.91)> lift:(1.27) lev:(0.04)
6. biscuits=t frozen foods=t vegetables=t total=high 797 ==> bread and cake=t 725 <conf:(0.91)> lift:(1.27) lev:(0.04)
7. baking needs=t biscuits=t vegetables=t total=high 772 ==> bread and cake=t 701 <conf:(0.91)> lift:(1.27) lev:(0.04)
8. biscuits=t fruit=t total=high 954 ==> bread and cake=t 866 <conf:(0.91)> lift:(1.26) lev:(0.04)
9. frozen foods=t fruit=t vegetables=t total=high 834 ==> bread and cake=t 757 <conf:(0.91)> lift:(1.26) lev:(0.04)
10. frozen foods=t fruit=t total=high 969 ==> bread and cake=t 877 <conf:(0.91)> lift:(1.26) lev:(0.04)

Status OK

Log x 0