COMS4701 Artificial Intelligence

jz2862 Jialin Zhao

Question 1: Association Rules

1. For a minimum support of 50%, use the Apriori algorithm to find all frequent itemsets in the transaction table.

C_1
Itemset
1
2
3
4
5

Scan of D

C_1	
Itemset	**Support**
1	4/6
2	5/6
3	2/6
4	4/6
5	6/6

F_1	
Itemset	**Support**
1	4/6
2	5/6
4	4/6
5	6/6

C_2
Itemset
12
14
15
24
25
45

Scan of D

C_2	
Itemset	**Support**
12	4/6
14	3/6
15	4/6
24	3/6
25	5/6
45	4/6

F_2	
Itemset	**Support**
12	4/6
14	3/6
15	4/6
24	3/6
25	5/6
45	4/6

C_3
Itemset
124
125
145
245

Scan of D

C_3	
Itemset	**Support**
124	3/6
125	4/6
145	3/6
245	3/6

F_3	
Itemset	**Support**
124	3/6
125	4/6
145	3/6
245	3/6

C_4
Itemset
1245

Scan of D

C_4	
Itemset	**Support**
1245	3/6

F_4	
Itemset	**Support**
1245	3/6

The frequent itemset is $F_1+F_2+F_3+F_4$

2. How many scans of the dataset were needed to find all frequent itemsets? What does this number represent?

4.

There are exactly 4 level(1,2,3,4) of frequent itemsets and the maximum size is a 4-itemset.

3. For a minimum confidence of 80%, use the Apriori algorithm to find all strong association rules (report support and confidence) of the form:

Item 1 → Item 2 (support, confidence)

Item 1 and item 2 → item 3 (support, confidence)

Itemset	Rule#	Rule	Confidence	Strong?
12	1	$1 \rightarrow 2$	1	Yes
	2	$2 \rightarrow 1$	4/5	Yes
14	3	$1 \rightarrow 4$	3/4	No
	4	$4 \rightarrow 1$	3/4	No
15	5	$1 \rightarrow 5$	1	Yes
	6	$5 \rightarrow 1$	2/3	No
24	7	$2 \rightarrow 4$	3/5	No
	8	$4 \rightarrow 2$	3/4	No
25	9	$2 \rightarrow 5$	1	Yes
	10	$5 \rightarrow 2$	5/6	Yes
45	11	$4 \rightarrow 5$	1	Yes
	12	$5 \rightarrow 4$	2/3	No

Itemset	Rule#	Rule	Confidence	Strong?
124	13	$1 \rightarrow 24$	3/4	No
	14	$2 \rightarrow 14$	3/5	No
	15	$4 \rightarrow 12$	3/4	No
	16	$24 \rightarrow 1$	1	Yes
	17	$14 \rightarrow 2$	1	Yes
	18	$12 \rightarrow 4$	3/4	No
125	19	$1 \rightarrow 25$	1	Yes
	20	$2 \rightarrow 15$	4/5	Yes
	21	$5 \rightarrow 12$	2/3	No
	22	$25 \rightarrow 1$	4/5	Yes
	23	$15 \rightarrow 2$	1	Yes
	24	$12 \rightarrow 5$	1	Yes

Itemset	Rule#	Rule	Confidence	Strong?
145	25	$1 \rightarrow 45$	3/4	No
	26	$4 \rightarrow 15$	3/4	No
	27	$5 \rightarrow 14$	1/2	No
	28	$45 \rightarrow 1$	3/4	No
	29	$15 \rightarrow 4$	3/4	No
	30	$14 \rightarrow 5$	1	Yes
245	31	$2 \rightarrow 45$	3/5	No
	32	$4 \rightarrow 25$	3/4	No
	33	$5 \rightarrow 24$	1/2	No
	34	$45 \rightarrow 2$	3/4	No
	35	$25 \rightarrow 4$	3/5	No
	36	$24 \rightarrow 5$	1	Yes

Itemset	Rule#	Rule	Confidence	Strong?
1245	37	$1 \rightarrow 245$	3/4	No
	38	$2 \rightarrow 145$	3/5	No
	39	$4 \rightarrow 125$	3/4	No
	40	$5 \rightarrow 124$	1/2	No
	41	$12 \rightarrow 45$	3/4	No
	42	$14 \rightarrow 25$	1	Yes
	43	$15 \rightarrow 24$	3/4	No
	44	$45 \rightarrow 12$	3/4	No
	45	$25 \rightarrow 14$	3/5	No
	46	$24 \rightarrow 15$	1	Yes
	47	$245 \rightarrow 1$	1	Yes
	48	$145 \rightarrow 2$	1	Yes
	49	$125 \rightarrow 4$	3/4	No
	50	$124 \rightarrow 5$	1	Yes

$$1 \rightarrow 2(\frac{4}{6}, 1)$$

$$2 \rightarrow 1(\frac{4}{6}, \frac{4}{5})$$

$$1 \rightarrow 5(\frac{4}{6}, 1)$$

$$2 \rightarrow 5(\frac{5}{6}, 1)$$

$$5 \rightarrow 2(\frac{5}{6}, \frac{5}{6})$$

$$4 \rightarrow 5(\frac{4}{6}, 1)$$

$$24 \rightarrow 1(\frac{3}{6}, 1)$$

$$14 \rightarrow 2(\frac{3}{6}, 1)$$

$$1 \rightarrow 25(\frac{4}{6} 1)$$

$$2 \rightarrow 15(\frac{4}{6}, \frac{4}{5})$$

$$25 \rightarrow 1(\frac{4}{6}, 4/5)$$
$$15 \rightarrow 2(4/6,1)$$
$$12 \rightarrow 5(4/6,1)$$
$$14 \rightarrow 5(\frac{3}{6}, 1)$$
$$24 \rightarrow 5(\frac{3}{6}, 1)$$
$$14 \rightarrow 25(3/6,1)$$
$$24 \rightarrow 15(\frac{3}{6}, 1)$$
$$245 \rightarrow 1(\frac{3}{6}, 1)$$
$$145 \rightarrow 2(\frac{3}{6}, 1)$$
$$124 \rightarrow 5(\frac{3}{6}, 1)$$

4. a. How can information provided by the TID sets of the most frequent k-itemsets be used to calculate the frequency of the potentially most frequent k+1-itemsets?

The TID that does not contain any frequent k-itemsets definitely doesn't contain any frequent k+1-itemset. So after the scan for k-itemset, the TID sets that contain no frequent k-itemset can be excluded in the next scan for k+1-itemsets.

As in AprioriTid algorithm, Apriori-TID uses sets C_k of the form (TID, $\{c_k\}$) where $\{c_k\}$ contains the list of k-itemsets included in the transaction TID, the TID with empty $\{c_k\}$ is excluded. The k+1 itemset can be obtained by the C_{k+1} which is formed by those TID in C_k and the list of k+1-itemsets included in the transaction TID.

b. Does the Apriori step of generating frequent itemsets still require scanning the entire transaction table? Discuss pros and cons of calculating frequency with this approach.

No.

Pros: The number of TIDs for level k in may be smaller than the number of transactions in the database, especially in the later passes. The algorithm doesn't need to scan the entire transaction after some level to save time.

Cons: Using the information provided by the former level requires additional space to store $\{c_k\}$ which means for large TID sets, the space may be a problem. And for TID sets that few of them can be eliminated during the iteration because they mostly contain a frequent high-level-k-itemset, the algorithm is no more efficient since it need little more time on forming and counting $\{c_k\}$.

Question 2: Local Search Algorithms

1. Describe simulated annealing (3-6 sentences).

Each iteration of SA first randomly generates a new state and check whether the new state can improve the evaluation function representing performance. If the new state is better, move there; else move with a probability based on a certain distribution on the change of evaluation function. By accepting worse case, SA can avoid local optimization in early iteration and get better solutions.

2. Compare simulated annealing to the genetic algorithm and give an example application of each one.

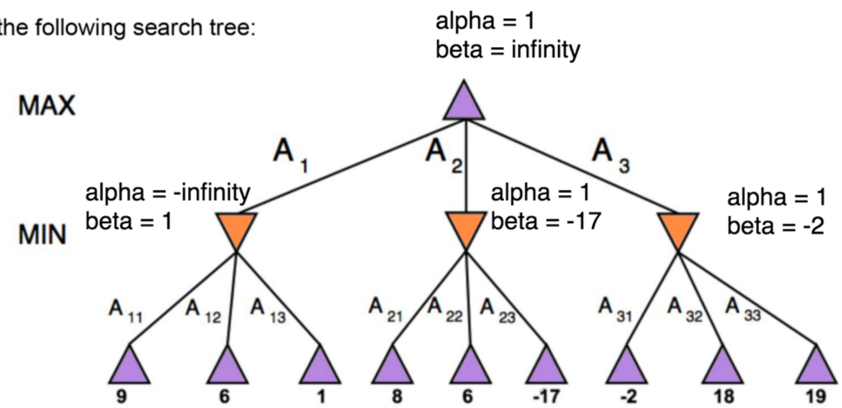
Both of them are used to solve optimization problems and can avoid local optimization.

SA is a single-state method and GA is a population method which means SA and GA differ in how they

generate candidate solutions for evaluation by the cost function.
SA get to next state once a candidate is ‘acceptable’(generate better performance or worse performance but accepted by a probability).
GA get to next population after checking if candidate solutions can survive then mutation and crossover.
The other difference I see is that SA will not abandon a better candidate solution but GA will abandon a better candidate with a surviving probability. Furthermore. SA will generate a candidate randomly while GA generate candidates by mutation and crossover.

Question 3: Minimax and Alpha-Beta Pruning

Consider the following search tree:



a. Using minimax, which of the three possible moves should MAX take at the root node? What is the value of Max at the root? A1,A2,A3. The max at root is 1
b. Using minimax with alpha-beta pruning, compute the value of alpha and beta at each node. Which branches are pruned? A32,A33

Question 4: Iterative Deepening in Adversarial Search

Reason1:
It maximizes the depth of search possible for any fixed time and space restrictions since it minimizes time and space for any given search depth.
Reason2:
So that when there is an unknown time limit on searching and the search ply have to be aborted, the IDS can remember the best node of next shallower depth
Reason3:
IDS can help to improve alpha-beta pruning by providing information from previous iterations.