LECTURE NOTES

Campus: PCE Course: BTECH in CSE Class/Section: III Yr. Section- A Date: 7/7/21

Name of Faculty: Praveen Kumar Yadav Name of Subject: Machine Learning Code: 6CS4-02

Date (Prep.): 7/3/21 Date (Del.): 10/7/21 Unit No.: III Lect. No: 6-17

OBJECTIVE: To be written before taking the lecture (Pl. write in bullet points the main topics/concepts etc., which will be taught in this lecture)

Apriori Algorithm

IMPORTANT & RELEVANT QUESTIONS:

what are advantages of Apriori Algo^m ?

FEED BACK QUESTIONS (AFTER 20 MINUTES):

what are the disadvantage of Apriori Algo^m

OUTCOME OF THE DELIVERED LECTURE: To be written after taking the lecture (Pl. write in bullet points about students' feedback on this lecture, level of understanding of this lecture by students etc.)

good

REFERENCES: Text/Ref. Book with Page No. and relevant Internet Websites:

scikit with ML.

PAGE NO. 7

Apriori Algorithm:- It works on the principle that "Having prior knowledge of frequent itemsets can generate strong association rules".

↳ Apriori means prior knowledge.

↳ Apriori find the frequent itemset by a process called candidate itemset generation. It is an iterative approach where k-itemset are used to explore (k+1) itemset. So, first the set of frequent 1-itemset is found, then frequent 2-itemset and so on until no more frequent k-itemset can be found.

↳ A k-candidate itemset is an itemset with k items in it.

↳ It follows Apriori property which states that all non-empty subset of a frequent itemset must also be frequent.

An Apriori Algorithm Example:

steps:- Minimum threshold value = 50%
or minimum support count = 2
minimum confidence = 2

TID	Itemset
1	[A A C]
2	[A C]
3	[A B]
4	[A B E F]

step 2:- Generate candidate itemset.

Itemset	Support count
A	3
B	2
C	2
D	1
E	1
F	1

L1: [A], [B], [C].

→ eliminated since SupportCount < minimum support count.

candidate - 1 - itemset

Itemset	Support Count
A, B	1
B, C	1
A, C	2

L2: [A, C]

→ eliminated

candidate - 2 - itemset

step 3. Mine Association Rules - To mine association rules from candidate itemset, a measure called confidance. It is simply depend as an association rule between items

* The confidance measure helps to identify which product drives the sale of which other product.

{ A ⇒ B } is 60% means 60% transaction containing A also contain B together)
↑
implies

$$\text{confidance}(A \Rightarrow B) = \frac{\text{Support Count}(A, B)}{\text{Support Count}(A)}$$

Rules	Support	Confidence
A ⇒ C (A drives the sale of C)	2	confidance = support(AC)/support(A) = 2/3 = 66%
C ⇒ A (C drives the sale A)	2	confidance = support(A,C)/support(C) = 2/2 = 100%

Here both rules are Accepted. Since confidence is 5.

but $C \Rightarrow A$ is strong Association.

Drawbacks of Apriori :-

1. Computationally Expensive - Since it requires repeated scans of the database. This is resource intensive and Time consuming.

2. it can mine misleading patterns - They are known to produce rules that are product of chance.

Campus: PCE. Course: BTECH
Name of Faculty: Praveen Kumar Yadav
Class/Section: IV CSE-A
Name of Subject: MC
Date: 7/4/21
Code: 6CS4-02

A-priori Algorithm:-

eg- Min Support = 50%

Threshhold Confidence = 70%.

TId	Items
100	1 3 4
200	2 3 5
300	1 2 3 5
400	2 5

represent Items distinctly.

(I)

Itemset	Support
1	2/4 → 50%
2	3/4 → 75%
3	3/4 → 75%
4	1/4 → 25% eliminated since < min-support
5	3/4 → 75%

(II)

ItemSet	Support
{1,2}	1/4 → 25% eliminated
{1,3}	2/4 → 50%
{1,5}	1/4 → 25% eliminated
{2,3}	2/4 → 50%
{2,5}	3/4 → 75%
{2,5}	3/4 → 50%

(III) Form Triplete Set

ItemSet	Support
{1,3,5}	1/4 = 25% elimina
{2,3,5}	2/4 = 50%
{1,2,3}	1/4 = 25% eliminated

Item Set → {1, 2, 3, 5}

Rules	Support	Confidence
$(2 \wedge 3) \rightarrow 5$	2	$2/2 = 100\%$
$(3 \wedge 5) \rightarrow 2$	2	$2/2 = 100$
$(2 \wedge 5) \rightarrow 3$	2	$2/3 = 66\%$
$2 \rightarrow (3 \wedge 5)$	2	$2/3 = 66\%$
$5 \rightarrow (2 \wedge 3)$	2	$2/3 = 66\%$
$3 \rightarrow (2 \wedge 5)$	2	$2/3 = 66\%$

eliminated since less than given confidence

Itemset	Support
1	2
2	3
3	3
5	3
{1, 3}	2
{2, 3}	2
{2, 5}	3
{3, 5}	2
{2, 3, 5}	2

TID	Items
100	1 3 4
200	2 3 5
300	1 2 3 5
400	2 5

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

$(2 \wedge 3) \rightarrow 5$

$(A \rightarrow B)$

$$S((2 \wedge 3) \cup 5) / S(2 \wedge 3)$$

$$= 2/2 = 100\%$$

Association

So Here only Two Rules are considered.

i.e. $(2 \wedge 3) \rightarrow 5$

$(3 \wedge 5) \rightarrow 2$