



DIAGNOSIS OF TUBERCULOSIS USING ASSOCIATION RULE METHOD

PAPENDER KUMAR¹, DEEPAK DANGWAL², NIDHI PURI³

Dept. of Computer Science, GBPEC Pauri (Garhwal)

*Corresponding Author: Email- nidhipuri22@gmail.com

Received: December 12, 2011; Accepted: January 15, 2012

Abstract- Tuberculosis is a disease caused by bacteria called Mycobacterium Tuberculosis. It usually spreads through the air & attack lungs and other parts of the body such as bone and brain. There are two categories of Tuberculosis in the body namely Pulmonary and Extra Pulmonary. HIV patient are more likely attacked with Tuberculosis. A correct diagnosis of tuberculosis can be only stated by applying a medical test. It is important health problem in India. Different techniques uses in medical data mining for disease detection. Now in this paper we have applied different association Algorithms. There are 3 types of association algorithms. 1) Priori algorithms 2) A Priori MR3) A Priori PT, and then Find out the result.

Keywords- Tuberculosis, Mycobacterium, Association Rule, Priori Algorithm.

Citation: Papender Kumar, Deepak Dangwal and Nidhi Puri (2012) Diagnosis of Tuberculosis Using Association Rule Method. Journal of Information and Operations Management ISSN: 0976-7754 & E-ISSN: 0976-7762, Volume 3, Issue 1, pp-133-135.

Copyright: Copyright©2012 Nidhi Puri, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Introduction

Tuberculosis is a one type of virus affected disease caused by bacteria called Mycobacterium tuberculosis.[1]

This type of disease is the main cause of high number of death. There are various types of bacteria. Some types of bacteria are beneficial to human and some are harmful.

A person who has infected with tuberculosis disease will have symptoms such as coughing more than 3 weeks, chest pain, night sweat, weight loss. There are 2 types of Tuberculosis. Pulmonary and Extra Pulmonary.

A. **Pulmonary-** when TB bacteria can settle in the lungs and begin to grow. TB in the lungs or throat can be infectious.

B. **Extra Pulmonary-** When TB bacteria can settle the other parts of the body such as bone.

The bacteria are put into the air when a person with active TB disease of the lungs or throat coughs or sneezes. People nearby may breathe in these bacteria and become infected. When a person breathes in TB bacteria, bacteria can settle in the lungs and begin to grow. From there, they can move through the blood to other parts of the body, such as bone, brain. This disease was discovered by Robert Koch in 1882. HIV patient are more likely attacked

with tuberculosis. Mycobacterium tuberculosis is grouped under bacilli rod shaped and its diameter cell ranges from 0.3µm to 0.5µm. Different Techniques uses in medical data mining for disease detection like Tuberculosis. But now this paper we have applied Association Rules. Association Rules are mainly used to show relationship between the data items.[2, 3]

Association Rules

Association rules are frequently used by retail stores to assist in marketing, advertising, and inventory control. They have direct applicability to retail businesses. A database in which an association rule is to be found is viewed as a set of tuples, where each tuple contains a set of items.[4]

In data mining association rules are useful analyzing and predicting customer behaviour.

The goal of this technique is to detect relationship or associations between specific values of categorical variables in large data sets. We have applied 3 types of association algorithms for tuberculosis data. For This Purpose We have taken data from Govt. Combined Hospital Srinagar (Garhwal) from 2008 to 2009 and then find out the result.

Ex-For Tuberculosis Patient Data-

Data source processing	
Computation time	9 MS
Allocated memory	10 KB

Attribute	Category	Information
Date	Discrete	28 values
region	Discrete	3 values
age	Continue	-
sex	Discrete	2 values
X-Ray	Discrete	2 values
Diagnosis	Discrete	2 values
Category	Discrete	3 values

A Priori Algorithm

A Priori algorithm is a popular method for discovering interesting relations between databases. The main purpose of association algorithm finds the logical relation between variables with the same status. In association rule mining, given a set of *item sets* the algorithm attempts to find subsets which are common to at least a minimum number C of the item sets. Apriori uses breath first search and a tree structure to count candidate item sets efficiently. [5]

Parameters

A-Priori parameters	
Support min	0.33
Confidence min	0.75
Max rule length	4
Lift filtering	1.10

Results ITEMS

Transactions	31
Counting items	
All items	12
Filtered items	7
Counting itemsets	
card(itemset) = 2	10
card(itemset) = 3	7
card(itemset) = 4	2
Rules	
Number of rules	6

After Applying a P priori Algorithm for Tuberculosis Data.

A Priori MR Algorithm

A Priori MR algorithm is used for the extraction of association rule. [6]

It is mainly used for market basket analysis. But infect it can be implemented in various application, where we want to association between different variables. Here we can implemented a P priori MR algorithm in Tanagra

For tuberculosis patient. This data taken for the "Government Combined Hospital Srinagar(Garhwal)".

Parameters

A-Priori parameters	
Support min	0.33
Confidence min	0.75
Max rule length	4
Lift filtering	1.10
Learning set ratio	1.00
Repetition	1

Results ITEMS

Transactions	31
Counting items	
All items	12
Filtered items	7
Counting itemsets	
card(itemset) = 2	10
card(itemset) = 3	7
card(itemset) = 4	2
Rules	
Number of rules	6

RULES

Computation time: 63 ms

Number of rules : 6					
N°	Antecedent	Consequent	Lift	Support (%)	Confidence (%)
1	"Category=I"	"X-Ray=Positive" - "Diagnosis=Pulmonary"	1.25797	45.161	93.333
2	"X-Ray=Positive" - "region= Srinagar" - "Category=I"	"Diagnosis=Pulmonary"	1.24000	32.258	100.000
3	"region=Srinagar" - "Category=I"	"Diagnosis=Pulmonary"	1.24000	35.484	100.000
4	"X-Ray=Positive" - "Category=I"	"Diagnosis=Pulmonary"	1.24000	45.161	100.000
5	"Category=I"	"Diagnosis=Pulmonary"	1.24000	48.387	100.000
6	"region=Srinagar" - "Category=I"	"X-Ray=Positive" - "Diagnosis=Pulmonary"	1.22530	32.258	90.909

A Priori PT Algorithm

We can apply P priori PT algorithm for tuberculosis data, and then find out the result.

A-Priori parameters	
Support min	0.33
Confidence min	0.75
Max rule length	4
Lift filtering	1.10

Results ITEMS

Transactions		31
Counting items		
All items		12
Filtered items		7
Counting itemsets		
card(itemset) = 2		10
card(itemset) = 3		7
card(itemset) = 4		2
Rules		
Number of rules		6

Results

In this paper, by applying 3 types of association algorithms for tuberculosis data and Found different results for different algorithm, and computation time is 63ms.

Conclusion

By applying Association Rule Method in tuberculosis data, find out in uttarakhand tuberculosis patient increase yearly, Andalso find out by histogram maximum patient find out in year 2010.

		Value s	Co unt	Per cent	Histogram	
Diagnosis	0.3	Pulmonary	25	80.65 %		
					Extra Pulmonary	6
		Valu es	Cou nt	Perce nt	Histogram	
Category	0.6	139	II	5	16.13 %	
			I	15	48.39 %	
			III	11	35.48 %	

References

- [1] Kamerbeek J., Schouls L., Van Agterveld M., Van Soolingen D., Kolk A., Kuijper S., Bunschoten A., Shaw R., Goyal M. and Van Embden J. (1997) J. Clin. Microbiol.
- [2] Jiawei Han and Micheline Kamber *Data Mining Concepts and Techniques*.
- [3] Sebban M., Mokrousov I., Rastogi N. and Sola C. (2002) Bioinformatics, No. 18, pp. 235-243.
- [4] Han J. and Kamber M. (2000) *Data Mining Concepts and Techniques*.
- [5] Chen M. S., Han J. and Yu P.S. (1996) *Data mining: An overview from a database perspective. IEEE Trans Knowledge and Data Engineering*.
- [6] Ferdinand S., Val'etudie G., Sola C. and Rastogi N. (2004) , *Research in Microbiology*.
- [7] World Health Organization (2009) *Global Tuberculosis Control*.

Attribute	Gini	Distribution				
		Value s	Co unt	Per cent	Histogram	
region	0.3	975	Lehr	1	3.23 %	
			Srinagar	23	74.19 %	
			Pauri	7	22.58 %	
		Valu es	Cou nt	Perce nt	Histogram	
sex	0.4	745	M	19	61.29 %	
			F	12	38.71 %	
		Valu es	Co unt	Per cent	Histogram	
X-Ray	0.1	748	Positive	28	90.32 %	
			Negative	3	9.68 %	