

Computational Molecular Biology

An integration of experimental molecular and genome biology with computational technology

A T C G T A G C C G A T G C
T A C G C G A T G C A T T A

G C C G G C T A G C C G A T
T A C G C G C G A T T A A T
A T G C T A C G T A G C A T
A T T A A T A T C G G C A T



2014
Vol.4

Publisher

Sophia Publishing Group

Editor-in-Chief

Dr. Henry Smith

Edited by

Editorial Team of Computational Molecular Biology

Email: edit@cmb.biopublisher.ca

Website: <http://cmb.biopublisher.ca>

Address:

11388 Stevenston Hwy,

PO Box 96016,

Richmond, V7A 5J5, British Columbia

Canada

Computational Molecular Biology (Online, ISSN 1927-5587), indexed by international database ProQuest, is an open access, peer reviewed journal publishing original research papers of general interest involving in the computational biology at the molecular level.

The Journal is publishing all the latest and outstanding research articles, letters, methods, and reviews in all areas of Computational Molecular Biology, covering new discoveries in molecular biology, from genes to genomes, using statistical, mathematical, and computational methods as well as new development of computational methods and databases in molecular and genome biology.

The papers published in the journal are expected to be of interests to computational scientists, biologists, and teachers/students/researchers engaged in biology, as well as are appropriate for R & D personnel and general readers interested in computational technology and biology.



Computational Molecular Biology is published independently by BioPublisher. It is committed to publishing and disseminating significant original achievements in the related research fields of molecular biology.

Open Access

All the articles published by BioPublisher are Open Access, and are distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



BioPublisher uses CrossCheck service to identify academic plagiarism through the world's leading anti-plagiarism tool, iThenticate, and to protect the original authors' work.

Latest Content

- [PlantSecKB: the Plant Secretome and Subcellular Proteome KnowledgeBase](#) 1-17
Gengkon Lum, John Meinken, Jessica Orr, Stephanie Frazier, Xiang Min
Computational Molecular Biology, 2014, Vol.4, No.1
- [GC2 Biology Dictates Gene Expressivity in *Camellia sinensis*](#) 18-25
Supriyo Chakraborty, Prosenjit Paul
Computational Molecular Biology, 2014, Vol.4, No.2
- [Association Rules for Diagnosis of Hiv-Aids](#) 26-33
Anubha Dubey
Computational Molecular Biology, 2014, Vol.4, No.3

Association Rules for Diagnosis of Hiv-Aids

Anubha Dubey✉

Department of Bioinformatics, MANIT, BHOPAL, India

✉ Corresponding Author email: anubhadubey@rediffmail.com; ✉ Author
Computational Molecular Biology, 2014, Vol.4, No.3 doi: 10.5376/cmb.2014.04.0003

Copyright © 2014 Dubey. This is an open access article published under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract Association rule mining is an active area of research in data mining. Data mining is a process of finding patterns from very large volumes of data. These patterns are important in making association rules and correlations among them. Recent years have witnessed many efforts on discovering associations for genes, proteins, enzymes, networks. In this study, association rules for HIV disease diagnosis is tried to generate. It describes the concept of different stages of HIV progression which are associated with other infections. As huge patient data is available, there is a need to develop some interesting patterns, associations, correlations for proper treatment and disease diagnosis. The efficiency and advantages of these rules has been used by medical practioners to diagnose the disease or recommend the suitable treatment.

Keywords Associations; Correlations; Pattern; HIV; Treatment

1 Introduction

HIV causes AIDS the life threatening opportunistic infection which leads to death of an individual. HIV infections are considered pandemic by the World Health Organization (WHO). As of 2010 approximately 34 million people have HIV globally. Of these approximately 16.8 million are women and 3.4 million are less than 15 years old. It resulted in about 1.8 million deaths in 2010, down from 3.1 million in 2001 (UNAIDS, 2010).

HIV infects primarily vital cells in the human immune system such as helper T cells (specifically, CD4⁺ T cells), macrophages, and dendritic cells. HIV infection leads to low levels of CD4⁺ T cells through three main mechanisms: (a) direct viral killing of infected cells; (b) increased rates of apoptosis in infected cells; and (c) killing of infected CD4⁺ T cells by CD8 cytotoxic lymphocytes that recognize infected cells. When CD4⁺ T cell numbers decline below a critical level, cell-mediated immunity is lost, and the body becomes progressively more susceptible to opportunistic infections.

1.1 Classification of HIV

Two types of HIV have been characterized: HIV-1 and HIV-2. HIV-1 is the virus that was initially discovered and termed both LAV and HTLV-III. It is more virulent, more infective and is the cause of the majority of HIV infections globally (Gilbert et al., 2003). HIV-1 is originated from Common Chimpanzee and HIV is originated from Sooty Mangabey (Smm). The lower infectivity of HIV-2 compared to HIV-1 implies that fewer of those exposed to HIV-2 will be infected per exposure. Because of its relatively poor capacity for transmission, HIV-2 is largely confined to West Africa (Reeves and Doms, 2002).

Both HIV-1 and HIV-2 are believed to have originated in West-Central Africa and to have jumped species (a process known as zoonosis) from non-human primates to humans (Sharp and Hahn, 2011). AIDS was first clinically observed in 1981 in the United States (Kaiser, 2008). In 1983, two separate research groups led by Robert Gallo and Luc Montagnier independently declared that a novel

Preferred citation for this article:

Anubha Dubey, 2014, Association Rules for Diagnosis of Hiv-Aids, Computational Molecular Biology, Vol.4, No.3 26-33 (doi: 10.5376/cmb.2014.04.0003)

Received: 04 Mar., 2014 | Accepted: 15 Apr., 2014 | Published: 01 May., 2014

retrovirus may have been infecting AIDS patients, and published their findings in the same issue of the journal *Science* (Barre-Sinoussi et al., 1983; Gallo et al., 1983). As the findings of these two research groups LAV (Lymph Adeno-virus) and HTLV-III (Human T-lymphotropic virus-III) were renamed HIV (Aldrich, 2001).

1.2 Stages of HIV infection

HIV infection has four basic stages: incubation period, acute infection, latency stage and AIDS.

Stage I: The initial incubation period upon infection is asymptomatic (if a patient is a carrier for a disease or infection but experiences no symptoms) or clinically silent with a CD4⁺ T cell count (also known as CD4 count) greater than 500/uL. It may include generalized lymph node enlargement and usually lasts between two and four weeks.

Stage II: This is a stage of acute infection (as shown in Figure 1), in which mild symptoms like minor mucocutaneous manifestations and recurrent upper respiratory tract infections, fever, lymphadenopathy (swollen lymph nodes), pharyngitis (sore throat), rash, myalgia (muscle pain), malaise, and mouth and oesophageal sores occurs. A CD4 count of less than 500/uL lasts an average of 28 days.

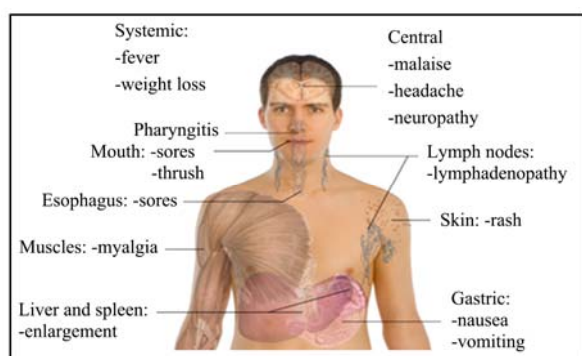


Figure 1 Main symptoms of acute HIV infection

Stage III: The latency stage, which shows advanced symptoms may include unexplained chronic diarrhoea for longer than a month, severe bacterial infections including tuberculosis of the lung and CD4 count of a person is found to be less than 350/uL and can last anywhere from two weeks to twenty years and beyond.

Stage IV: The final stage of HIV infection is AIDS, this shows the symptoms of various opportunistic infections. Severe symptoms which includes toxoplasmosis of the brain, candidiasis of the oesophagus, trachea, bronchi or lungs and Kaposi's sarcoma. A CD4 count of less than 200/uL (WHO case 2007) and viral load increases to millions (Weiss, 1993).

Today there is a plenty of patient data available in the databases they need to be analysed and further knowledge is needed for formulating the drugs available for HIV-AIDS. Recent studies show that association rule mining is used to discover frequent patterns, correlations of genes/ proteins, protein networks. But this study focus on the development of association rules to diagnose disease on the basis of symptoms, medical tests, associated infections etc. At an earlier stage so anti retroviral therapy has been started by medical practitioners and life of a patient could be increased. These are possible due to data mining approaches. Earlier some efforts have been made by some scientists to develop such association rules in medical databases (Abdullah et al., 2010).

2 Methodology

Data Mining refers to extracting or mining knowledge from large amount of data. Data mining has been around for several years for exploration of interesting knowledge or information from a large amount of data. Association rule mining is the discovery of association rules showing attribute value conditions that occur frequently together in a given set of data. These algorithms search for interesting frequent patterns, associations, correlations, or causal relationship among sets of items or objects. Such relationships are usually represented by association rules, rules that are produced by association mining (Han and Kamber, 2011).

2.1 Association rule

The term was coined by Agrawal et al. (1993) and amazingly it stills becomes an active area of research in knowledge database discovery. Suppose the viruses as the set of items causes disease, and then each virus has a Boolean variable representing the presence or absence of that disease. Each patient can then be

represented by a Boolean vector of values (checking of particular virus infection) assigned to these variables. The Boolean vectors can be analyzed for symptoms shown by patients of viruses that reflect particular symptoms that are frequently associated together to diagnose the particular virus infection. These patterns can be represented in the form of association rules. For example, the information that patients told to medical practitioners also tend to test for certain tests based on symptoms. By this the association rule is generated as: In association rule, support and confidence are two measures of rule interestingness. They respectively reflect the usefulness and certainty of discovered rules. A support of 2% for Association Rule (1) means that 2% of all the symptoms under analysis show that patient is suffering from disease which might be infected by certain viruses. A confidence of 60% means that 60% of the patients who checked specific virus test have HIV-AIDS. Typically, association rules are considered interesting if they satisfy both a minimum support threshold and a minimum confidence threshold. Such thresholds can be set by users or domain experts. Additional analysis can be performed to uncover interesting statistical correlations between associated items.

In case of HIV, association rule 1 can be written as:

Virus \rightarrow Disease [support = 2%, confidence = 60%] (1)

A support of 2% for association rule (1) means that 2% of all viruses undergo analyses shows that it might be HIV which causes AIDS. A confidence of 60% means that 60% of the patients who recommended for virus analyses test might be test for HIV which causes AIDS.

Let $HIV=I$, $HIV=\{I_1, I_2, \dots, I_m\}$ be a set of symptoms. Let D , be HIV stages which has set of symptoms according to each stage. Where each symptom T is a set of symptoms such as $T \subseteq I$. Each symptom is associated with a stage called TID . Let A be a set of symptoms. A symptom T is said to contain A if and only if $A \subseteq T$. An association rule is an implication of the form $A \Rightarrow B$, where $A \subset I$, $B \subset I$, and $A \cap B = \emptyset$. The rule $A \Rightarrow B$ holds in the stage set D with supports,

where s is the percentages of stages in D that contain $A \cup B$ (i.e. the union of set A and set B). This is taken to be the probability, $P(A \cup B)$. The rule $A \Rightarrow B$ has confidence c in the stage set D , where c is the percentages in D containing A that also contain B . This is taken to be conditional probability, $P(B/A)$. That is,

$$\text{support}(A \Rightarrow B) = P(A \cup B) \quad (2)$$

$$\text{confidence}(A \Rightarrow B) = P(B|A) \quad (3)$$

From equation 3 we have:

$$\text{confidence}(A \Rightarrow B) = P(B/A) = \frac{\text{support}(A \cup B)}{\text{support}(A)} = \frac{\text{support_count}(A \cup B)}{\text{support_count}(A)} \quad (4)$$

The equation 4 shows that confidence of rule $A \Rightarrow B$ can be easily derived from the support counts of A and $A \cup B$. That is, once the support count of A , B , $A \cup B$ are found, it derives corresponding association rules $A \Rightarrow B$ and $B \Rightarrow A$ and check whether they are strong. Rules that satisfy both a minimum support threshold (min_sup) and a minimum confidence threshold (min_conf) are called strong. By convention it might be written as support and confidence values so as to occur between 0% and 100%, rather than 0 to 1.0 (Han and Kamber, 2011).

2.2 Frequent itemsets for association rule mining

A set of items is referred to as an itemset. An itemset that contains k items is a k -itemset. The set {virus, disease} is a 2-itemset. The occurrence frequency of an itemset is the number of transactions that contain the itemset. This is also known, simply, as the frequency, support count, or count of the itemset. Note that the itemset support defined in Equation (4) is sometimes referred to as relative support, whereas the occurrence frequency is called the absolute support. If the relative support of an itemset I satisfies a prespecified minimum support threshold (i.e., the absolute support of I satisfies the corresponding minimum support count threshold), then I is a frequent itemset. In general, association rule mining can be viewed as a two-step process:

- (i). find all frequent itemsets: By definition, each of these itemsets will occur at least as frequently as a predetermined minimum support count, min sup .

(ii). Generate strong association rules from the frequent itemsets: By definition, these rules must satisfy minimum support and minimum confidence.

2.3 From association analysis to correlation analysis

Correlation relationships between associated items: Most association rule mining algorithms employ a support-confidence framework. Often, many interesting rules can be found using low support thresholds. Although minimum support and confidence thresholds help weed out or exclude the exploration. It is said that the support and confidence measures are insufficient at filtering out uninteresting association rules. To tackle this weakness, a correlation measure can be used to augment the support-confidence framework for association rules. This leads to correlation rules of the form,

$$A \Rightarrow B [\text{support, confidence, correlation}] \quad (5)$$

That is, a correlation rule is measured not only by its support and confidence but also by the correlation between itemsets A and B. There are many different correlation measures from which to choose. Let us choose a simple example. HIV is a simple correlation measure that is given as follows. The occurrence of itemset A is independent of the occurrence of itemset B if $P(A \cup B) = P(A)P(B)$; otherwise, item sets A and B are dependent and correlated as events. This definition can easily be extended to more than two itemsets. The HIV between the occurrence of A and B can be measured by computing (lift equation):

$$\text{HIV}(A, B) = P(A \cup B) / P(A)P(B) \quad (6)$$

If the resulting value of Equation (6) is less than 1, then the occurrence of A is negatively correlated with the occurrence of B. If the resulting value is greater than 1, then A and B are positively correlated, meaning that the occurrence of one implies the occurrence of the other. If the resulting value is equal to 1, then A and B are independent and there is no correlation between them. Equation (6) is equivalent to $P(B/A)/P(B)$, or $\text{conf}(A \Rightarrow B) / \text{sup}(B)$, which is also referred as the HIV of the association (or correlation) rule $A \Rightarrow B$. In other words, it assesses the degree to which the occurrence of one “HIV” the occurrence of

the other tuberculosis. For example, if A corresponds to the HIV and B corresponds to the tuberculosis, then given the current relation of HIV with tuberculosis, which can be measured by 5.23. As it was studied that HIV infected patients also suffer from tuberculosis.

Example 1 Correlation analysis: To help filter out misleading “strong” associations of the form $A \Rightarrow B$ from the data, we need to study how the two itemsets, A and B, are correlated. Let we have the data of any hospital. Of the 10 000 patients analyzed the data showed that 6000 of the patients with HIV positive, while 7 500 were infected with tuberculosis and 4 000 included patients are infected with HIV and tuberculosis. Suppose that a data mining program for discovering association rules is run on the data, using a minimum support of say 30% and a minimum confidence of 60%. The following association rule is discovered:

$$\text{Disease (virus, HIV)} \Rightarrow \text{buys (X, "Other")} \\ [\text{support}=40\%, \text{confidence}=66\%] \quad (7)$$

Rule (7) is a strong association rule and would therefore be reported, since its support value of $4000/10,000 = 40\%$ and confidence value of $4000/6000 = 66\%$ satisfy the minimum support and minimum confidence thresholds, respectively. However, rule 5.21 is misleading because the probability of purchasing videos is 75%, which is even larger than 66%. In fact, virus and HIV are negatively associated because the infection of one of these diseases actually decreases the likelihood of infection of other.

The above example also illustrates that the confidence of rule $A \Rightarrow B$ can be deceiving in that it is only an estimate of the conditional probability of itemset B given itemset A. It does not measure the real strength of the correlation and implication between A and B. Hence alternatives to the support-confidence framework can be useful in mining data relationships. From above example we need to study how the two itemsets A and B are correlated. Followed this example let the disease refer to the patients that don't have HIV and tuberculosis refer to those that do not contain tuberculosis. The patients data can be

summarized in a contingency table, as shown in table 1. From the table, it is illustrated that the probability of HIV infected patients are $P(\{HIV\}) = 0:60$, the probability of tuberculosis infected patients are $P(\{Tuberculosis\}) = 0:75$, and the probability of infecting both is $P(HIV,Tuberculosis) = 0.40$. By Equation (6) and (7) is $P(HIV, Tuberculosis) = (P(\{HIV\}) \times P(\{Tuberculosis\})) = 0.40 = (0.60 \times 0.75) = 0.89$. Because this value is less than 1, there is a negative correlation between the occurrence of $\{HIV\}$ and $\{Tuberculosis\}$. The numerator is the likelihood of a customer purchasing both, while the denominator is what the likelihood would have been if the two purchases were completely independent. Such a negative correlation cannot be identified by a support confidence framework.

Table 1 A 2x2 contingency table summarizing the transactions with respect to HIV and Tuberculosis infection

	HIV	\overline{HIV}	Σ ROW
Tuberculosis	4000	3500	7500
$\overline{Tuberculosis}$	2000	500	2500
Σ col	6000	4000	10000

Example II Correlation analysis using chi square measure: To compute the correlation using χ^2 analysis, we need the observed value and expected value (displayed in parenthesis) for each slot of the contingency table, as shown in Table 2 from the table we compute the χ^2 value as follows:

$$\chi^2 = (4000-4500)^2/4500 + (3500-3000)^2/3000 + (2500-1500)^2/1500 + (500-1000)^2/1000 = 555.6$$

Table 2 The above contingency table, now shown with the expected values

	HIV	\overline{HIV}	Σ row
Tuberculosis	4000 (4500)	3500 (3000)	7500
$\overline{Tuberculosis}$	2000 (1500)	500 (1000)	2500
Σ col	6000	4000	10000

Because the χ^2 value is greater than one, and the observed value of the slot (HIV, Tuberculosis) =4000, which is less than the expected value 4500, having HIV and Tuberculosis are negatively correlated. This is consistent with the conclusion derived from the analysis of the Example 1 and 2.

Let's examine two other correlation measures, all confidence and cosine, as defined below.

Given an itemset $x = \{i_1, i_2, \dots, i_k\}$ the all_confidence of X is defined as:

$$all_conf(x) = \frac{sup(x)}{\max_item_sup(x)} = \frac{sup(x)}{\max\{sup(i_j) / \forall i_j \in x\}} \tag{8}$$

Where $\max\{sup(i_j) / \forall i_j \in x\}$ is the maximum (single) item support of all the items in x, and hence is called the max_item_sup of the itemset x. The all confidence of x is the minimal confidence among the set of rules $i_j \rightarrow x \rightarrow i_j$, where $i_j \in x$.

Given two itemsets A and B, the cosine measure of A and B is defined as:

$$cosine(A, B) = \frac{P(A \cup B)}{\sqrt{P(A) \times P(B)}} = \frac{sup(A \cup B)}{\sqrt{sup(A) \times sup(B)}} \tag{9}$$

The cosine measure can be viewed as a harmonized HIV measure: the two formulae are similar except that for cosine, the square root is taken on the product of the probabilities of A and B. This is an important difference however because by taking the square root the cosine value is only influenced by the supports of A, B, and AUB, and not by the total number of patients.

Lift and chi-square are poor indicators of the other relationships, whereas all_confidence and cosine are good indicators. In between all confidence and cosine are good indicators. This is because cosine considers the supports of both A and B, whereas all_confidence considers only the maximal support. The lift and chi-square are poor correlations because we do not consider null patients data. A null patient data that does not contain any of the disease patient data being examined. Typically, the number of null-transactions can outweigh the number of individual infected with diseases, because many patients may neither infected with HIV nor Tuberculosis. On the other hand, all confidence and cosine values are good indicators of correlation because they are not influenced by the number of null patient data. A measure is null-invariant if its value is free from the influence of null-data. Null-invariance is an important property for

measuring correlations in large patient databases. To prove all confidence and cosine the best at assessing correlation in all cases. Let's examine the HIV and Tuberculosis examples again.

Example 2 Comparison of four correlation measures on HIV and Tuberculosis data. We revisit example 1.

Table 3 Comparison of the four correlation measures for HIV-and-Tuberculosis data sets

Dataset	HT	\overline{HT}	$H\overline{T}$	\overline{HT}	All_confi	cosine	Lift	χ^2
D_0	4000	3500	2000	0	0.53	0.60	0.84	1477.8
D_1	4000	3500	2000	500	0.53	0.60	0.89	555.6
D_2	4000	3500	2000	10000	0.53	0.60	1.73	2913.0

In Table 3, HT , \overline{HT} and $H\overline{T}$, remain the same in D_0 , D_1 . However, lift and χ^2 change from rather negative to rather positive correlations, whereas all confidence and cosine have the nice null-invariant property, and their values remain the same in all cases. Unfortunately, we cannot precisely assert that a set of items are positively or negatively correlated when the value of all confidence or cosine is around 0.5. Strictly based on whether the value is greater than 0.5, we will say that H and T are positively correlated in D_1 , however, it has been shown that they are negatively correlated by the lift and χ^2 analysis. Therefore, a good strategy is to perform the all confidence or cosine analysis first, and when the result shows that they are weakly positively/negatively correlated, other analyses can be performed to assist in obtaining a more complete picture.

Besides null-invariance, another nice feature of the all confidence measure is that it has the Apriori-like downward closure property. That is, if a pattern is all-confident (i.e., passing a minimal all confidence threshold), so is every one of its sub patterns. In other words, if a pattern is not all-confident, further growth (or specialization) of this pattern will never satisfy the minimal all confidence threshold. This is obvious since according to Equation (8), adding any item into an itemset X will never increase $\text{sup}(X)$, never decrease $\text{max item sup}(X)$, and thus never increase $\text{all conf}(X)$. This property makes Apriori-like pruning possible: we can prune any patterns that cannot satisfy the minimal all confidence threshold during the growth of all-confident patterns in mining.

Let D_1 be the original HIV (H) and Tuberculosis (T) data set from Table 5.7. We add two more data sets, D_0 and D_1 , where D_0 zero null patients' data has, and D_2 has 10 000 null-patients data (instead of only 500 as in D_1). The values of all four correlation measures are shown in Table 3.

Jaccard Similarity Coefficient: It is a statistical index for measuring the similarity and variety of sample sets (Roussinov and Zhao, 2003).

3 Result and Discussion

Tuberculosis (TB) and HIV have been closely linked since the emergence of AIDS. Worldwide, TB is the most common opportunistic infection affecting HIV-seropositive individuals, and it remains the most common cause of death in patients with AIDS (Raviglione et al., 1995). HIV infection has contributed to a significant increase in the worldwide incidence of TB (AIDSCAP, 1996; Raviglione et al., 1992). By producing a progressive decline in cell-mediated immunity, HIV alters the pathogenesis of TB, greatly increasing the risk of disease from TB in HIV-co infected individuals and leading to more frequent extra pulmonary involvement, atypical radiographic manifestations, and paucibacillary disease, which can impede timely diagnosis. Although HIV-related TB is both treatable and preventable, incidence continues to climb in developing nations wherein HIV infection and TB are endemic and resources are limited. Interactions between HIV and TB medications, overlapping medication toxicities, and immune reconstitution inflammatory syndrome (IRIS) complicate the co treatment of HIV and TB. These association rules with correlation of diseases help the bio- medical scientists and medical practitioner for better treatment of diseases.

It was observed that the use of only support and confidence measures to mine associations results in

the generation of a large number of rules, most of which are uninteresting to the user. Instead, it is augmented the support-confidence framework with a correlation measure, resulting in the mining of correlation rules. The added measure substantially reduces the number of rules generated, and leads to the discovery of more meaningful rules. However, there seems to be no single correlation measure that works well for all cases. Unfortunately, most such measures do not have the null invariance property. Because large data sets typically have many null-transactions, it is important to consider the null-invariance property when selecting appropriate interestingness measures in the correlation analysis. Our analysis shows that both all confidence and cosine are good correlation measures for large applications, although it is wise to augment them with additional tests, such as lift, when the test result is not conclusive. There are other correlations measures are also given in literature that can be used for analysis.

4 Conclusion and Future work

The improvement of new technologies raises data collection and accumulation. Without appropriate processing and interpretation this information is useless. There are four standard methods of data mining: Association, classification, clustering techniques and prediction. For most medical applications, the logical rules are not precise but vague and the uncertainty is present both in premise and decision. For this kind of application a good methodology is the rule representation from decision-tree method, which is easily understood by user. Therefore the integration of fuzzy set and data mining methods gives a much better and more exact representation of relationship between symptoms and diagnosis. These associative patterns are useful in classification on the basis of certain properties.

Different association's patterns can be found by the rules discussed and further correlation can be studied by these patterns. These patterns can be useful for medical diagnostics to improve the condition of patients by proper medications. Further associated diseases can be finding by associated symptoms. Like HIV-tuberculosis other associated pattern of disease can be diagnosed and treated simultaneously. These

analyses are useful for biomedical scientists to further improve the formulations of drugs for treatment of diseases.

References

- AIDS Control and Prevention (AIDSCAP) Project of Family Health Internal, 1996, The Francois-Xavier Bagnoud Center for Public Health and Human Rights Of the Harvard School of Public Health, UNAIDS. The Status and Trends of the Global HIV/AIDS Pandemic. Final Report
- Aldrich, ed., Robert, Wotherspoon, Garry, 2001, Who's who in gay and lesbian history, London: Routledge, pp.154
- Barre-Sinoussi F., Chermann J., Rey F., Nugeyre M., Chamaret S., Gruest J., Dauguet C., and Axler-Blin C., 1983, Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS), *Science*, 220(4599): 868-871
<http://dx.doi.org/10.1126/science.6189183>
- D.Roussinov, and J.L. Zhao, 2003, Automatic Discovery of similarity Relationships through Web Mining, *Decision Support Systems*, 25: 149
[http://dx.doi.org/10.1016/S0167-9236\(02\)00102-1](http://dx.doi.org/10.1016/S0167-9236(02)00102-1)
- Doc Kaiser's Microbiology Home Page>IV.VIRUSES>F.ANIMAL VIRUS LIFE CYCLE>3.The Life Cycle of HIV Community College of Baltimore County
- Gallo P.S. Sarin, E.P. Gelmann, M. Robert-Guroff, E. Richardson, V.S. Kalyanaraman, D. Mann, G.D. Sidhu, R.E. Stahl, S. Zolla-Pazner, J. Leibowitch, and M. Popovic, 1983, Isolation of human T-cell leukemia virus in acquired immune deficiency syndrome (AIDS), *Science*, 220(4599): 865-867
<http://dx.doi.org/10.1126/science.6601823>
- Gilbert P.B., McKeague I.W., Eisen G., Mullins C., Guéye-Ndiaye A., Mboup S., and Kanki P.J., 2003, Comparison of HIV-1 and HIV-2 infectivity from a prospective cohort study in Senegal, *Statistics in Medicine*, pp.573-593
<http://dx.doi.org/10.1002/sim.1342>
- J. Han and M. Kamber, 2000, *Data Mining Concepts and Techniques*, The Morgan Kaufmann Series in Data Management Systems Morgan Kaufmann Publishers
- R. Agrawal, T. Imielinski and A.Swami, 1993, Database Mining: A Performance Perspective, *IEEE Transactions on Knowledge and Data Engineering*, 5(6): 914

- <http://dx.doi.org/10.1109/69.250074>
- Raviglione M.C., Narain J.P., and Kochi A., 1992, HIV-associated tuberculosis in developing countries: clinical features, diagnosis, and treatment, *Bull WHO*, 70: 515-526
- Raviglione M.C., Snider D.E., and Kochi A., 1995, Global epidemiology of tuberculosis: morbidity and mortality of a worldwide epidemic, *JAMA*, 273: 220-226
<http://dx.doi.org/10.1001/jama.1995.03520270054031>
- Reeves J.D., and Doms R.W., 2002, Human Immunodeficiency Virus Type 2, *Journal of General Virology*, 83(6): 1253-1265
- Sharp P.M., and Hahn B.H., 2011, Origins of HIV and the AIDS Pandemic, *Cold Spring Harbor perspectives in medicine*, 1(1): a006841
- UNAIDS and WHO, 2010, UNAIDS Report on the global AIDS epidemic, pp.16-34
- Weiss R.A., 1993, How does HIV cause AIDS? *Science*, 260(5112): 1273-1279
<http://dx.doi.org/10.1126/science.8493571>
- Z. Abdullah, T. Herawan, and M.M. Deris, 2010, Detecting Critical Least Association Rules in Medical Databases, *International Journal of Modern Physics: Conference Series*, 1(1): 1-5



Reasons to publish in BioPublisher

A BioScience Publishing Platform

- ★ Peer review quickly and professionally
- ☆ Publish online immediately upon acceptance
- ★ Deposit permanently and track easily
- ☆ Access free and open around the world
- ★ Disseminate multilingual available

Submit your manuscript at: <http://bio.sophiapublisher.com>

