Model based Investigation of Pandemic Influenza

Ei Ei Aung; Daw khin Moe San
Computer University(Hinthata),Myanmar
ei.aung3@gmail.com;

Abstract

A pandemic is an epidemic of human disease occurring over a very wide area, crossing international boundaries and affecting a large number of people. Influenza is a virus that causes respiratory disease in humans, with typical symptoms of fever, cough, and muscle ache and pneumonia and death. This system can learn the patterns using Bayesian Analysis and develop a Decision Support System. Bayesian Classifier is based on the theorem of posterior probability. Calculate the probability when the new case comes. Computer-based medical systems are playing an increasing relevant role in assisting both diagnosis and treatments. This paper intends to develop Bayesian Classification method for flu diagnosis based on the symptoms of the patients. This system stores the knowledge of the medical experts and the medical record. Based on the knowledge stored, the system can learn the pattern using Bayesian Analysis and decides the probability when the new case comes. To develop a Decision Support System for automatic classification method for Pandemic Influenza based on symptom of the patients. Decision support system is also used for the patient who tests themselves at home instead of clinical test.

Keywords: Bayes' Theorem, Classifier Accuracy, Decision Support Systems (DSS)

1. Introduction

Mortality due to pandemic influenza is a leading cause for the flu related deaths in the country. The Bayesian's formalism offers a natural way to represent the uncertainties involved in medicine when dealing with diagnosis, treatment selection, planning, and prediction of prognosis. The Bayesian theorem calculates the probability associated with the possible diseases that could account for the clinical picture via the signs, symptoms, and laboratory findings entered by the diagnostician. What was before an isolated database or a laboratory information system is now integrated in a larger scale (departmental, hospital, or community-based) medical information system. The increase in data volume causes difficulties in extracting useful information for decision support. Decision Support System (DSS) is aimed for proliferating in many areas of human endeavor involving complex problem solving including medical diagnosis. Intelligent Data Analysis (IDA) in a clinical setting. IDA of an intelligent assistant that tries to bridge the gap between data gathering and data comprehension, in order to enable the physician to perform the task more efficiently and effectively.

2. Related works

In Bayes' Theorem, knowledge is represented via hypotheses, H_i, each of which is characterized by a subjective probability p (H_i), representing one's confidence in its truth [5]. Early studies of subjective probability assessments in Bayesian decision tasks found consistently that judges adjusted their probabilities less than Bayes' Theorem demands. By failing to adjust their probabilities as much as the objectively should have, the judges were thus 'conservative' ([8]; [6]). Such a strategy could bias the integration of information in ambiguous cases, leading to 'premature closure' whereby possible diagnoses are not considered once a hypothesis has been identified [7]. Hybrid technologies that combine symbolic representations of knowledge with more quantitative representations inspired by biological information processing systems have resulted in more flexible, humanlike behavior [1]. AI ideas also have been adopted by other computer scientists, for example, "data mining", which combine ideas from databases, AI learning, and statistics to yield systems that find interesting patterns in large databases, given only very broad guidelines. AI also owes a debt to the software side of computer science, which has supplied the operating systems, programming languages, and tools needed to write modern programs [4]. As in particular the 'colonisation' variables together with selected antibiotics determine choice of treatment by predicting coverage, time cannot be ignored [3]. One popular diagnostic web-based Decision Support System DSS, Isabel, provides information in the form of additional diagnoses which the practitioner may or may not have considered in the assessment of a given patient (e.g. [9]; [10]; [2]).

3. System overview

The system is used to classify flu using Bayesian classifier and needs the train data. Bayesian classifiers are statistical classifiers. They can predict class membership probabilities, such as the probability that a given sample belongs to a particular class. Bayesian classifiers have also exhibited high accuracy and speed when applied to large databases. Intelligent Data Analysis (IDA) encompasses statistical, pattern recognition, machine learning, data abstraction and visualization tools to support the analysis of data and discovery of principles that are encoded within the data. The useful, operational information/knowledge, which is expressed at the right level of abstraction, is then readily available to support the decision making of the physician in managing a patient.

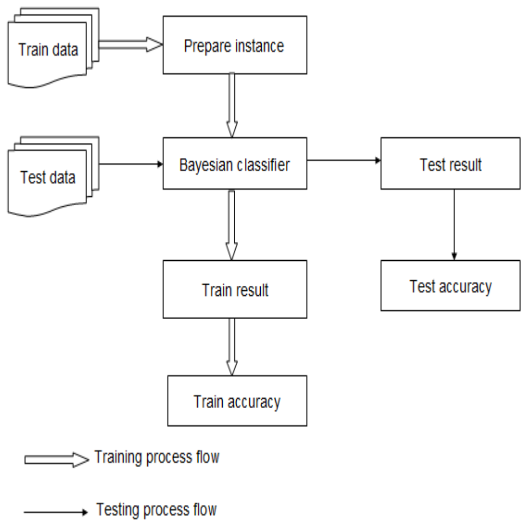

Figure 1: Overview of the System

3.1. Bayesian Classification

Bayesian classifier is based on Bayes' theorem describe

Let D be a training set of tuples and their associated class labels. As usual, each tuple is represented by an n-dimensional attribute vector X=(x1,x2,x3,…..,xn), depicting n measurements made on the tuple from n attribute , respectively, A1 , A2 , A3 , ….., An).

Support that there are m classes C1, C2, ….., Cm. Given a tuples, X, the classifier will predict that X belongs to the class having the highest posterior probability, conditioned on X. The nave Bayesian classifier predicts that tuple X belongs to the class Ci if and only if

$P(Ci / X) > (Cj / X)$ for $i <= J <= m , j != i$. ---(1)

Thus we maximize P (Ci/ X). The class Ci for which P (Ci/ X) is maximized is called the maximum posteriori hypothesis. By Bayes' theorem

$P (Ci/ X) = P (X/ Ci) P (Ci) / P(X)$ ------ (2)

3.2 Classifier Accuracy

Using training data derive a classifier or predictor and than to estimate the accuracy of the resulting learned model can result in misleading overoptimistic estimates due to overspecialization of the learning algorithm to the data. The accuracy of a classifier on a given test set is the percentage of test set tuples that are correctly classified by the classifier.

The system uses precision to access the percentage of tuples labeled as 'flu' that actually are 'flu' tuples. These measure are defined as

$$\text{Sensitivity} = \frac{t\_pos}{pos} \quad \text{-------------- (3)}$$

$$\text{Specificity} = \frac{t\_neg}{neg} \quad \text{-------------- (4)}$$

$$\text{Precision} = \frac{t\_pos}{(t\_pos + f\_pos)} \quad \text{--------------- (5)}$$

Where,

t_pos is the number of true positive ('flu' tuples that were correctly classified).

pos is the number of positive ('flu') tuples.

t_neg is the number of true negative ('non-flu' tuples that were incorrectly labeled as 'flu').

Show that accuracy is a function of sensitivity and specificity:

$$accuracy=sensitivity \frac{pos}{(pos+neg)} + specificity \frac{neg}{(pos+neg)} \quad \text{--- (6)}$$

The true positives, true negatives, false positives and false negatives are also useful in assessing the cost and benefits associated with a classification model.

3.3 Intelligent Data Analysis (IDA)

Investing in the development of appropriate IDA methods, techniques and tools for the analysis of clinical data is thoroughly justified and this research ought to form a main thrust of activity by the relevant research communities. Numerous intelligent data analysis methods have already been applied for supporting decision making in medicine. These methods can be classified into two main categories: data abstraction and data mining.

- Data abstraction is concerned with the intelligent interpretation of patient data in a context-sensitive manner and the presentation of such interpretations in a visual or symbolic form, where the temporal dimension in the representation and intelligent interpretation of patient data is primary importance.

- Data mining is concerned with the analysis and extraction (discovery) of medical knowledge from data, aimed at supporting diagnostic, screening, prognostic, monitoring, thereby support or overall patient management tasks.

The majority of data mining methods belong to machine learning and the majority of data abstraction methods perform temporal abstraction. This is the main reason for machine learning and temporal abstraction being the focus of investigation in this system.

3.4 Decision Support System(DSS)

DSS provides psychological e-health services to clients in a variety of clinical settings and skill-training programs. The system applies Bayesian decision models; the main purpose of this system is to show that Bayes' Theorem can offer a viable and flexible approach to the design of DSSs. Bayesian models can be applied usefully to different kinds of decision problems, and that psychological research can successfully contribute to the background justification, the definition, design, and evaluation of DSSs in the medical and psychological arenas, regardless of whether the system is intended to support human- or machine learning. To develop a DSS for automatic classification method for pandemic influenza diagnosis based on the symptoms of patients. DSS based on lessons learned from empirical research into certain judgmental biases that were found systematically to impact diagnostic decisions in a series of controlled laboratory experiments.

4. System architecture

The system includes two data sets for training and testing. So the new patient can train and test using Bayesian Classification. The system provides the interactive communication with the system user and has training data to train the classifier. Testing has to be done to test the accuracy of the classifier. The Bayesian Classification provides accuracy rate for training and testing data. We use the Bayesian analysis as a model to estimate the accuracy of the system and to diagnosis the new patient.

Diagnosing the patient can be done through the patient diagnosing, where entry of the user symptoms can be filled to get the result. Patient's medical records are stored in the history records database.

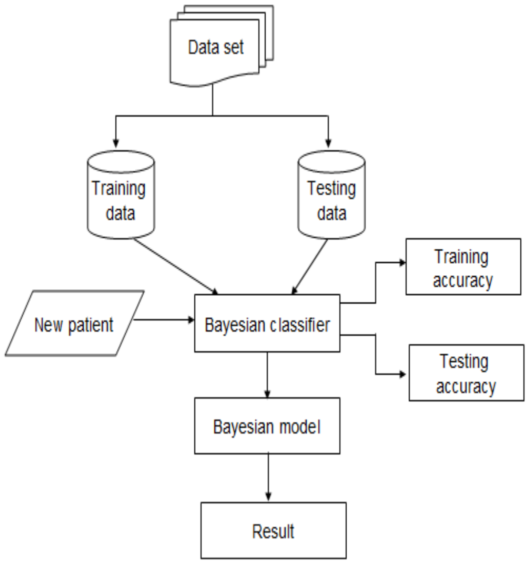

Figure 2: System Architecture

5. Experimental results

Figure 3 show training data for influenza ('no', 'seasonal-flu', 'H1N1' and 'H5N1') entry of the system to classify. The new patient information will be saved in the training data entry.

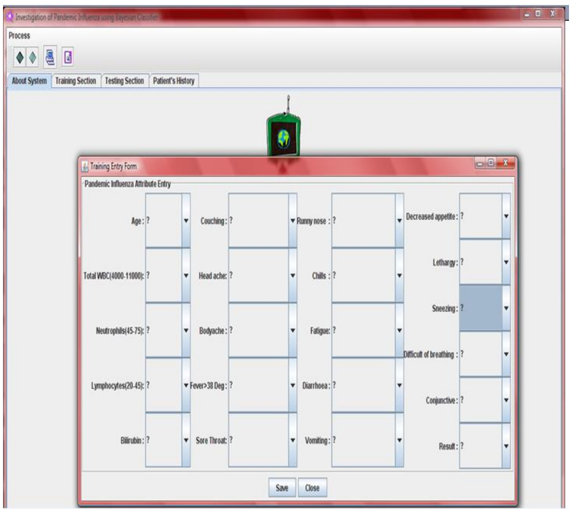

Figure 3: Training Data Entry

Figure 4 show the training data of the system. The training data records are used Bayesian Classification to get the results. There are 20 attributes and one result fields. Each attribute has own attribute value. E.g. the attribute (vomiting) has attribute values such as yes or no.

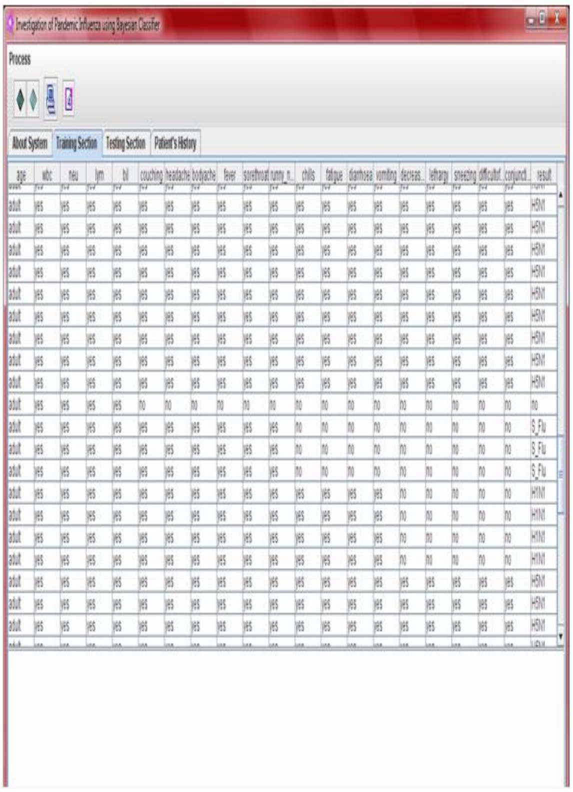

Figure 4: Training the Samples

Figure 5 show training result and accuracy generated from Bayesian analysis for the system.

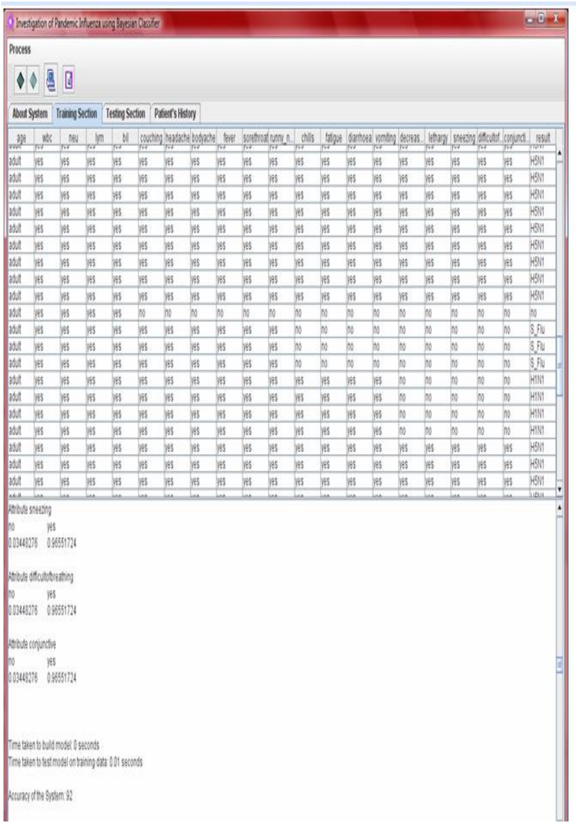

Figure 5: Accuracy Rate of Training Result

Figure 6 show the testing section of the system. When we fill the information needed by the system, we can see the result of the patient. The system can save testing data to train.

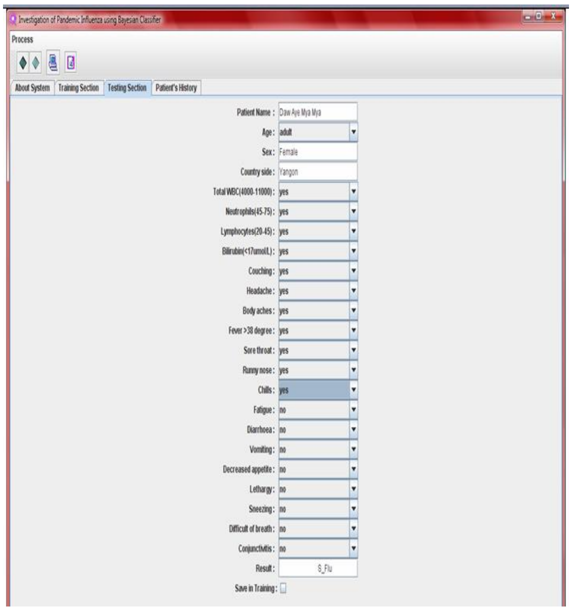

Figure 6: Result for detection of Pandemic Influenza

Figure 7 show testing result and accuracy using Bayesian classifier. The decision support for influenza classification is obtained by using 20 attributes and 4 classes for various kind of result.

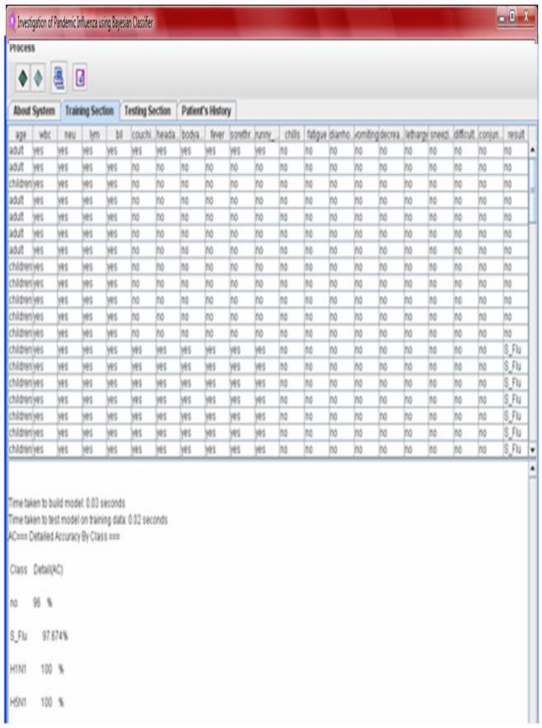

Figure 7: Accuracy Rate for Each Class of Influenza

Figure 8 shows the history of the patient. So we can see the patient list of the system.

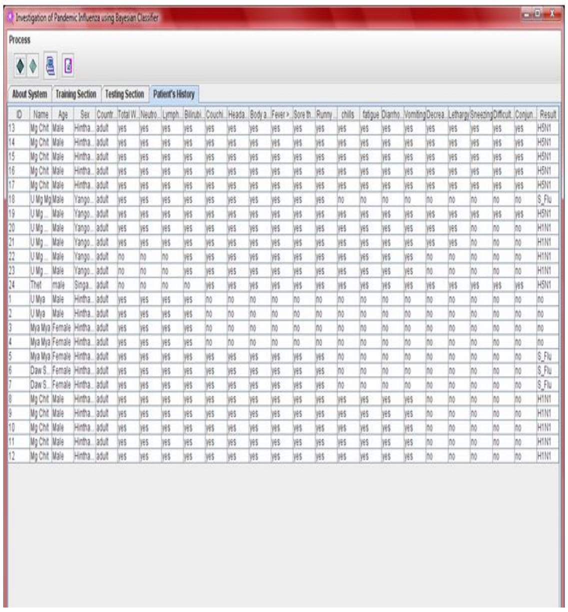

Figure 8: Patient History

5.1 Advantages of the system

This System support user in classifying 'no', 'seasonal_ flu', 'H1N1' and 'H5N1' based on the symptoms of the patients. We can easily test them at home by entering the symptom data. This system reduces opportunities for human infection, morbidity, mortality and social disruption. The system can get treatment early- before it spreads beyond the pandemic influenza and lead our long and healthy life.

6. Conclusion

Naive Bayesian Analysis seems to be a suitable technique for medical knowledge based system. We have presented a Bayesian Analysis system, with the aim of supporting patient's managements by symptom analysis in medical domain. It is a general principle of major incident planning that procedures should not be changed at precisely the moment when the system or institution is under its greatest stress, so planning for pandemic flu needs to make use as much as possible of systems and procedures already in place. The computer system will make the decision instead of a doctor. That means for achieving more accurate and effective medical diagnosis for Pandemic Influenza. At the present time, data mining is the powerful technology in computer fields. The computer will make possible decision instead of a doctor by using this system.

References

[1] Alan Rea, email A.Rea@qub.ac.uk Generated with CERN Web Maker All documents are the responsibility of, and copyright, their authors and do not represent the views of The Parallel Computer Centre, nor of The Queen's University of Belfast.

[2] Bavdekar, S.B. Pawar, M. "Evaluation of an internet [-delivered pediatric diagnosis support system (ISABLE) in a tertiary care center in India', Indian pediatrics, 42, November, (2005), 1086-1091.

[3] N.C de Bruijn, P.J.F. Lucas, K. Schurink, M.Bonten, A.Hoepelman. Temporal probabilistic knowledge for medical decision making, in: AIME 2001. Lecture Notes in Artificial Intelligence 2101, Springer-Verlag, Berlin, 2001, pp. 231-234.

[4] Centers for Disease Control and Prevention: Flu Surge [http://www.cdc.gov/flu/tools/flusurge/]

[5] De Finetti, B.Probability: Beware of falsifications! Scientia, 3, (1976) 283-303.

[6] Edwards, W. Conservatism in human information processing, in B Kleinmuntz (Ed). Formal representation of human judgement, John Wiley & sons, New York, NY (1982)

[7] Grabere, M.L Diagnostic errors in medicine: What do doctors and umpries have in common? Morbidity & mortality, (2007), 1-6, (2007).

[8] Phillips, L.D. Edwards, W. Conservatism in a simple probability inference task', Journal of experimental psychology, 72, 3, (1966), 346-354.

[9] Ramnarayan, P.Decision support systems: A how-to guide to their evaluation, Sage Publications electronic resources, http://c.sagepub.com (2005)

[10] Ramnarayan, P. Cronje, N. Report on Isabel adult medical validation study; Isabel Healthcare report, London, UK (2005) Schiff, G.D., Kim, S, Abrams, R.Cosby, K. Lambert, B. Elstein, A.S. Hasler, S.Krosnjar, N.Odwazny, R.Wisniewski, M.A. & McNutt, R.A. Advances in patient safety, 2, (2006), 255-278.