

A Framework for Agent-Based Data Mining Classification on Allergy-Induced Asthma Patients

T. Rekha¹, K. Shyamala²

¹Research Scholar, PG & Research Department of Computer Science, Dr. Ambedkar Government Arts College, Chennai, India

²Associate Professor, PG & Research Department of Computer Science, Dr. Ambedkar Government Arts College, Chennai, India

Abstract: “Health is wealth”. In human life health plays a main role. Recognizing the importance of healthy life, we should learn about how to stay healthy as well as prevent environment pollution. To create an awareness of allergy-induced asthma, this study aims to identify and analyse the prediction of asthma from allergy patients’ history. Implementing this is in the field of computer science is the first of its kind. To attain this objective, the concept of agent and data mining can be used. Data mining is the process of collecting, analysing and discovering patterns in large datasets. An intelligent agent is software that assists people and act on the behalf. The data used for this study is collected and analysed by means of agent paradigm along with the clustering techniques of Data mining technology.

Keywords: Intelligent Agent, data mining, Clustering, Allergy-induced Asthma, Agent mining

1. Introduction

A. Intelligent agent and data mining

Health promotion is very relevant today. There is a global acceptance that health and social wellbeing are determined by many factors outside the health system which include socio economic conditions, patterns of consumption associated with food and communication, demographic patterns, learning environments, family patterns, the cultural and social fabric of societies. Now a day many of the people suffered by allergy and asthma because of their living style, here we take some example of patient details how the allergy turns over asthma depending upon the sex and age criteria. Based on the analysis, the following predictions are made. This study uses the agent along with the clustering techniques of Data mining technology. The intelligent agents help in predicting the clusters automatically. The visualizing agent is used to view the results in multi-dimensional ways [2]. Visual agent is used to generate various reports and indicates the patterns to be displayed. The reports based on the cluster analysis are understandable by the user.

Kaufman describes the pathophysiology of asthma as a pathologic condition which affects the lower respiratory tract by narrowing the airways as a result of epithelial damage,

excessive mucus production, oedema, bronchoconstriction and muscle damage [1].

A software agent can initiate monitor events communication, filtering information, alarm generation work flow management, personal assistance and perform task without the direct intervention of humans or others. A software agent can sense its environment and acts autonomously [3].

Agent is software that assists people and acts on their behalf. Intelligent agents work by allowing people to delicate work that they could have done to the agent [2].

Data mining is the process of analysing data from different perspective and summarizing it into useful information. Data mining is knowledge based systems, and information can be retrieved Data mining is used to find correlation or pattern [4].

Data mining and intelligent agents can make people to help in the decision making process elaborate decisional models and take good decisions in real time [5].

B. Cluster analysis

Cluster is a group of objects that belongs to the same class. In other words, similar objects are grouped in one cluster and dissimilar objects are grouped in another cluster.

Clustering is the process of making a group of abstract objects into classes of similar objects.

- A cluster of data objects can be treated as one group.
- While doing cluster analysis, we first partition the set of data into groups based on data similarity and then assign the labels to the groups.
- The main advantage of clustering over classification is that, it is adaptable to changes and helps single out useful features that distinguish different groups.

C. Applications of cluster analysis

Clustering analysis is broadly used in many applications such as market research, pattern recognition, data analysis, and image processing.

- Clustering can also help marketers discover distinct groups in their customer base. And they can characterize

their customer groups based on the purchasing patterns.

- In the field of biology, it can be used to derive plant and animal taxonomies, categorize genes with similar functionalities and gain insight into structures inherent to populations.
- Clustering also helps in identification of areas of similar land use in an earth observation database. It also helps in the identification of groups of houses in a city according to house type, value, and geographic location.
- Clustering also helps in classifying documents on the web for information discovery.
- Clustering is also used in outlier detection applications such as detection of credit card fraud.
- As a data mining function, cluster analysis serves as a tool to gain insight into the distribution of data to observe characteristics of each cluster.

D. Clustering methods

Clustering methods can be classified into the following categories:

- Partitioning Method
- Hierarchical Method
- Density-based Method
- Grid-Based Method
- Model-Based Method
- Constraint-based Method

Density-based Method

This method is based on the notion of density. The basic idea is to continue growing the given cluster as long as the density in the neighbourhood exceeds some threshold, i.e., for each data point within a given cluster, the radius of a given cluster has to contain at least a minimum number of points.

2. Literature review

The study presented here, mainly deals with the concept of Allergy-induced Asthma, as we all know, what is allergy and how it affects us.

- Allergic Asthma is caused by breathing in an allergen, which is a substance that can cause an allergic reaction. Airborne, year-round allergens, such as pet dander or dust mites are commonly called allergic triggers [6]. Allergic asthma is the most common type of asthma.
- There are approximately 26.5 million people in the US with asthma. About 60% (6 out of 10) of people with asthma have allergic asthma. For school-age children, this number is closer to 80%. There are 300 million asthmatics worldwide with 1/10th of those living in India. A recent review analysis of 15 epidemiological studies showed that the mean prevalence of asthma among children was 7.24%. The prevalence of childhood asthma has continued to increase in last 10 years on the Indian subcontinent [7].
- According to India's largest community of verified

doctors Curofy in a poll of 1040 doctors, 82% said that there has been an increased asthma incidence in children due to increased environmental pollution [8].

A. Spirometry

Spirometry helps diagnose all types of asthma including allergy induced, which is shown in figure 1. The test involves taking a deep breath and measuring the amount of air you can inhale and exhale as well as how fast you breathe in and out. Spirometry also helps your doctor determine the severity of your asthma and determine if treatment is effective.

FEV1 is one of the most common indices used to assess airway obstruction. It is automatically calculated during spirometry or pulmonary function testing. It is calculated using a spirometer.

B. How FEV1 is used in asthma treatment

Most commonly forced expiratory volume will be ordered by your doctor as part of complete pulmonary function tests. Your doctor may do this to assess your symptoms before an asthma diagnosis has been made or monitor your asthma control as part of your asthma action plan. Symptoms such as the following may trigger your doctor to order these tests:

- Wheezing
- Chest tightness
- Cough
- Shortness of breath

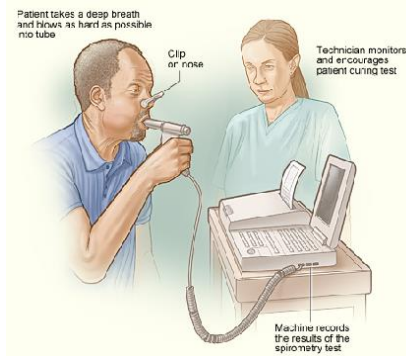


Fig. 1. Spirometry test

Decreasing allergens

Preventing allergy-induced asthma flare-ups is one of the most vital aspects of treatment. Prevention involves identifying and reducing allergens. Depending on what allergen is triggering an asthma attack, there are several ways to decrease exposure [9]. Association rules are if-then statements that help to show the probability of relationships between data items within large data sets in various types of databases. Association rule mining has a number of applications and is widely used to help discover sales correlations in transactional data or in medical data sets. How association rules work.

Association rule mining, at a basic level, involves the use of machine learning models to analyse data for patterns, or co-occurrence, in a database. It identifies frequent if-then

associations, which are called association rules. An association rule has two parts: an antecedent (if) and a consequent (then). An antecedent is an item found within the data. A consequent is an item found in combination with the antecedent.

Association rules are created by searching data for frequent if-then patterns and using the criteria support and confidence to identify the most important relationships. Support is an indication of how frequently the items appear in the data. Confidence indicates the number of times the if-then statements are found true. A third metric, called lift, can be used to compare confidence with expected confidence. Association rules are calculated from item sets, which are made up of two or more items. If rules are built from analysing all the possible item sets, there could be so many rules that the rules hold little meaning. With that, association rules are typically created from rules well-represented in data.

3. Methods and materials

As mention above, data mining techniques are used to discover patterns of data. In the proposed system, the density based clustering technique is used to study the factors that affect patients' healthcare in allergy-induced asthma. To expertise this process, the role of agents is invoked. The proposed framework [14] has the following agents.

A. User interface Agent

The user interface agent interacts with the user to perform data collection activities. The user interface agent observes and monitors the actions taken by the user. The user agents act as an autonomous personal assistant which cooperates with the users accomplishing some task [3]. As for learning, interface agents learn typically to better assist its user in three ways.

- By observing the sex rule
- By observing the age rule
- By observing the season rule

B. Data Mining Agent

A data mining agent is used to discover patterns in an abundance of information. Data mining agents can also "retrieve" relevant data from databases, alerting end – users [15]. In this research work, data mining agent is responsible for generating the results by using clustering technique.

- This agent uses information technology to find trends and patterns in an abundance of information from many different sources. The user can sort through this information in order to find whatever information they are seeking.
- 'Classification' is one of the most common types of data mining, which finds patterns in information and categorizes them into different classes. In this research work, data mining agent is responsible for generating the results using clustering techniques.

C. Mobile Agent

Mobile agents are agents because they are autonomous and

they cooperate, with other agents. Mobile agents are well suited to the workflow applications. The nature of workflow applications includes support for the flow of information. Mobile agents provide a degree of autonomy to the workflow item. Data can be analysed using mobile agent tools.

D. Learning Agent

Learning agents have the capacity to adapt or modify their behavior to learn. The value of learning agents is used to create the agent, specify the rules to be used by the agent. Learning agents is also used for creating and managing rules. Learning agents performs tasks, by receiving instructions from the user.

E. Multi-Agent

Agents can communicate, cooperate and negotiate with other agents. In multi agent systems, there is no single designer who stands behind all the agents. Each agent in the system may be working toward different goals, even contradictory ones [16]. Coordination is a key factor in the success of multi agent systems. Multi-agent is used to training the factors which affects the customer preference and it interact with another agent. Coordination can be done by reversing actions, synchronization, structured group mediation and information sharing [17].

F. Visualization Agent

The visualization agent is used to generate results. The results based on the cluster. The results can be viewed in multi-dimensional way. The cluster can be formed by using data mining technique. The visualization might be 1D, 2D, 3D. Visualization of data set can be defined as combination of methods and indicating what patterns should be displayed. Visualization agent will coordinate with data mining agent, based on that it will identify suitable visual method for each specific cluster [14].

From these aspects, an agent based framework is developed for data analysis, cluster detection and visualization for decision making by means of intelligent agent technology. The methodology adopted in this research work is described as follows,

Proposed methodology

Step 1: Start

Step 2: Collect the data using user interface agent

Step 3: Selecting the clustering technique using Data mining agent

Step 4: Creating the workflow in the appropriate tool using Mobile agent

Step 5: Training each factor that affect the customer preference using Multi Agent

Step 6: Selection of appropriate visualization tool

Step 7: Multi-dimensional visualization of result using Visual Agent

Step 8: Stop

Proposed frame work for analysis of allergy-induced asthma

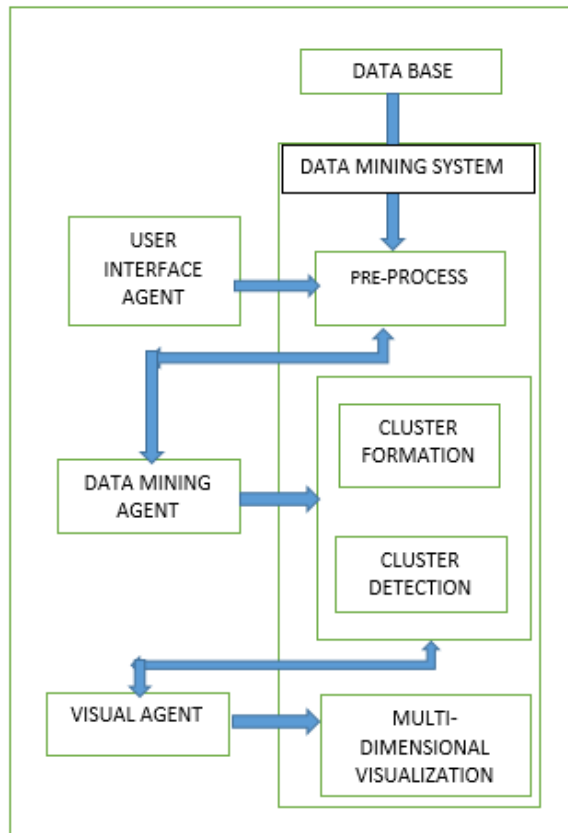


Fig. 2. Proposed frame work for Agent based computing using Data mining technique

The newly developed framework diagrammatically is shown (figure) from this framework, it's clear that the datamining system gets the data from the database or its processing.

Apart from getting data from the given database, it also gets the user specifications through the user interface agent. The data mining agent chooses the appropriate algorithm for the pre-processed data received from the user inter face agent the algorithm used here is the density based clustering algorithm the datamining agent is responsible for the cluster formation and detection for the datasets available. Once the selection of appropriator clustering algorithm is over, the next step is to selected the suitable visualization tool. The visualization agent is responsible for displaying the result in an easy understandable format, which can be used to portray the exact results of the survey. The intelligent agents are used in automated process of attribute section, ranking process, cluster formation, cluster detection and visualization.

The user interface agent is responsible for receiving user specifications and delivering results to the data mining agent. The data mining agent mine result based on the parameters given by user inter face agent to perform cluster formation and cluster detection. Clustering algorithms are used to formulate a new cluster, based on the user interface agent with respect to

specific user profile. Then cluster detection techniques are used to detect the cluster quality for farther process. The quality of cluster is identified by various parameters with help of intelligent agent. As result, the best cluster is identified by various parameters with help of intelligent agent. As result the best cluster will be discovered from the known knowledge. The visual agent is used to visualize the identified cluster depends on the nature of data within the detected cluster. After completion of cluster formation and cluster detection, it transfers the result to decision system where the visual agent find suitable representation tools based on the cluster nature. Finally, the end result is visualized in terms of 1D,2D, or 3D by visual agent. Thus, the entire process is monitored as well as executed by automated intelligent system based on the user profile. This makes the less domain knowledge user more convenient and understand ability.

In this research work different types agents are used to perform the operations on behalf of user so that, the data mining result will be productive and knowledgeable less domain knowledge user. Agent used in the framework are for reliable communication, cooperation among the agents, and finally coordination among the other agent within the system of perform some specific tasks [10].

G. Database considered

Health promotion is very relevant today. There is a global acceptance that health and social wellbeing are determined by many factors outside the health system which include socioeconomic conditions, patterns of consumption associated with food and communication, demographic patterns, learning environments, family patterns, the cultural and social fabric of societies; socio-political and economic changes, including commercialization and trade and global environmental change. In such a situation, health issues can be effectively addressed by adopting a holistic approach by empowering individuals and communities to take action for their health, fostering leadership for public health, promoting intersect oral action to build healthy public policies in all sectors and creating sustainable health systems. Now a day many of the people suffered by allergy and asthma because of their living life style, here we take some example of patient details how the allergy turns over asthma depending upon the sex and age criteria. Based on the analysis the following prediction are made

the concept of data mining is used to perform this analysis WEKA 3.8, one of the prominent tools of data mining is used for this study. Its provides a collection of machine learning algorithms for data mining task. WEKA contains tools for data pre-processing, classification, association and clustering rules and visualization. The DBSCAN algorithm of clustering technique is used carry out the study.

4. Result and discussion

To perform the analysis, the test data was collected from the Allergy and Asthma health care centre. Nearly 200 patient data

are taken for this analysis. It contains 18 attributes of various data type. Some of the attributes are as follows: PATIENT ID, ETHNIC GROUP, SMOKING HISTORY, HEIGHT, WEIGHT, AGE, SEX, ADDRESS, OCCUPATION, AND SOME TEST VALUES etc.

A. Based on sex rule

The major aim of this work is to predict the statistical value for asthma patients from the allergy patient details. Here we collect more than 200 patient values from the hospital and the value of the FEV1 test used to divide the patients into two categories of allergy and allergy-induced asthma. From the predicted value of asthma patients, we divide them into two categories depending upon the sex.

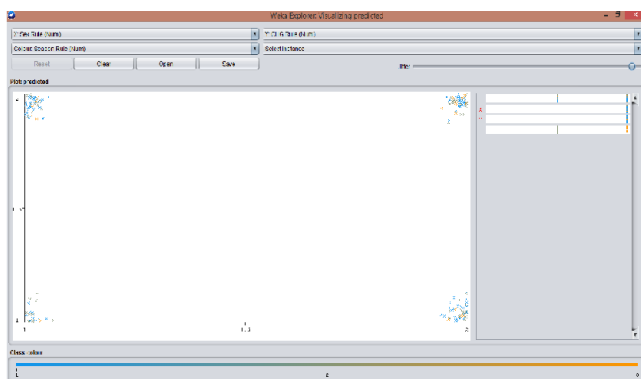


Fig. 3. Clustering value of sex rule

B. Based on age rule

The major aim of this work is to predict the statistical value for asthma patients from the allergy patient details. Here we collect more than 200 patient values from the hospital and the value of the FEV1 test used to divide the patients into two categories of allergy and allergy-induced asthma. From the predicted value of asthma patients, we divide them into three categories depending upon the age for showing which category of patients affect more.

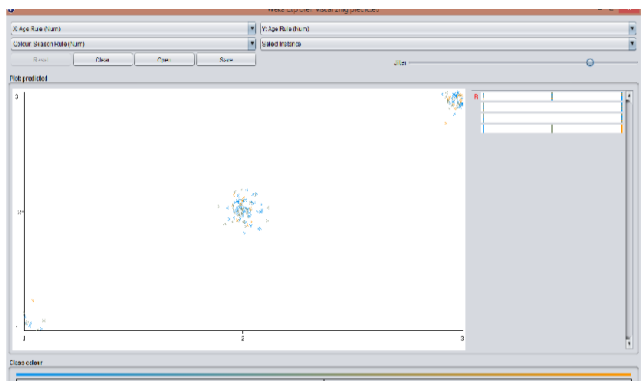


Fig. 4. Clustering value of age rule

C. Based on season rule

The major aim of this work is to predict the statistical value for asthma patients from the allergy patient details. Here we

collect more than 200 patient values from the hospital and the value of the FEV1 test is used to divide the patients into two categories of allergy and allergy-induced asthma. From the predicted value of asthma patients, we divide them into three categories depending upon the season for showing which season affect the patients more.

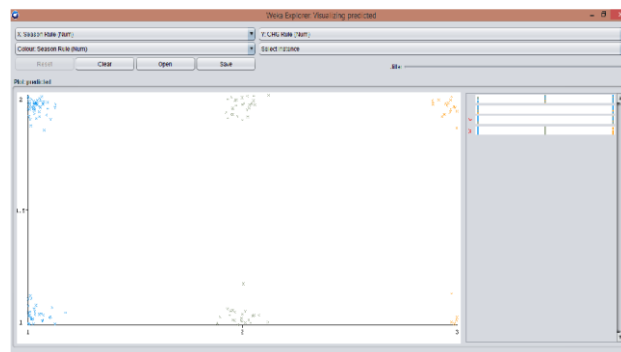


Fig. 5. Clustering value of seasonal rule

5. Conclusion

Data mining is the procedure of mining knowledge from data which can be widely used in the process of decision making. An agent is a computer system located in some environment, and is capable of autonomous action in this environment to meet its design objective. Multi agent systems are systems made up of multiple interacting agents. In this research work, the mining process is carried out by different intelligent agent with the help of the data collected. The results are analysed, evaluated and visualized by forming and detecting the clusters using used interface, data mining and visualization agents.

The major aim of this research work is to predict how many people are suffered from asthma and we classified them in to sex categories, age category and to find which season affect them lot. Therefore, we used 200 data of allergy affected patients, which were all collected from Allergy asthma healthcare centre. From these data we need to predict whether the patients are affected by asthma or only affected by allergy for down immune or climate changing etc. In the patients details there were nearly 18 attributes used and were analysed.

As per our survey we analysed that the issue of Allergy-induced asthma. The classification of Allergy and Asthma patients are done with using some rules like the patient age, sex and season. From this analysis, many people are affected on winter seasons, from sex rule many female patients are predicted with asthma and using age rule the between age 14 to 40 years old are predicted with asthma. From total number of 200 patients 117 people are predicted for asthma and remaining 83 people are suffered by only allergy.

6. Future work

The future scope of this work is to carry out the analysis by using various other suitable methods, and by using other data mining techniques. This model can be extended by

implementing the proposed agent based computing frame work using the various agent tools available. Its scope for future work can also be extended in the part of gathering data by increasing the number of instances from different health care centre.

The data collected in this work is collected from the health care centre point of view and this can be extended by collecting from individual patients. Point of view, this research work focused only on the patients FEV1 test value from the spirometry test. This can be extended by making to take more test values from other test also.

References

- [1] <https://www.ukessays.com/essays/health/asthma-is-one-of-the-major-chronic-respiratory-conditions.php>
- [2] Don Gilbert, "Intelligent Agents: The Right Information at the Right Time," IBM Corporation, Research Tringle Park, NC USA, May 1997.
- [3] P. Maes et. al., "Artificial Intelligence Meets Entertainment: Life like Automous Agents," Communications of the ACM, vol. 38, no. 11. Pp. 108-114, November 1995.
- [4] J. Han, and M. Kamber, "Data Mining: Concepts and Techniques," Morgan Kaufmann Publishers, pp. 5-7, 2001.
- [5] Irina Tudor and liviu Ionida, "Intelligent Agent as Data Mining Techniques used in Academic Environment," 2009.
- [6] [ips-to-fight-breathlessness/articleshow/52088776.cms](https://www.ips-to-fight-breathlessness.com/articleshow/52088776.cms).
- [7] <https://www.news18.com/news/india/india-has-10-of-worlds-asthma-patients-survey-1238186.html>
- [8] <https://www.thehealthsite.com/news/world-asthma-day-3rd-may-4-out-of-5-doctors-say-environmental-pollution-is-the-biggest-cause-of-asthma-ag0516-396386/>
- [9] <https://www.pubfacts.com/detail/29520161/Knowledge-on-self-management-and-levels-of-asthma-control-among-adult-patients-in-Ho-Chi-Minh-City-V>
- [10] R. Jayabrabhu, V. Saravanan, and K. Vivekanandan, "A frame work: cluster detection and multidimensional visualization of automated data mining using intelligent agent," International Journal of Artificial Intelligence & Applications, vol. 3, January 2012.