Chapter 3

MASICC: Multi-Agent System for Infant and Child Care

3.1 Introduction to the childcare problem

The situation of health care in rural or remote areas of India is poor because of traditional ways of the treatments and a lack of physical infrastructure available to the doctors. It is observed that at some rural places, even less qualified health care professionals are somehow managing the situation at the rural dispensaries. They lack proper knowledge to diagnose and treat the patients. Whenever they face any critical problem related to childhood diseases, they refer these patients to the CHC located at urban areas. This increases the number of patients for the child specialist at the CHC. The main diseases that young infants and children suffer from include: Pneumonia, Diarrhea, Measles, and Malaria etc. These are often curable at local levels if proper diagnosis and treatment plans are administered timely. Clearly, in such a scenario, conventional techniques of providing health care would not be sufficient. So to alleviate health care issues in the rural areas, a MAS is required that demonstrates the intelligent behavior, by imitating knowledge of a pediatrician. In the subsequent section, a MAS model for infant and child care is presented at abstract level.
3.2 MASICC: Multi-agent model for infant and child care

The MAS is a system in which the agents are connected in an intelligent fashion to achieve their pre-defined objectives [58]. These agents interact with each other, according to their capabilities, and try to produce the desired result. Moreover, there may or may not be any human interference as the intelligence is supposed to be embedded in them (as a software code). A multi-agent based system, called MASICC (Multi-agent model for infant and child care) is proposed in this section. The overall picture of the model is shown in Figure 3.1. The childhood disease ontology is shared by all the agents. It designs aspects are discussed in section 5.1. The model also depicts the agents at the three levels that mirror the real world healthcare model followed in India.

Nomenclature and location of the agents are:

- The agent that supports the doctor at the rural site is termed as User Agent (UA). There can be more than one UA seeking advice.
- The agent that behaves as a pediatrician and is located at the CHC is named as Intelligent Pediatric Agent (IPA),
- The agent that acts as a super specialist is named as the Super Specialist Agent (SSA).

The UA supplies the sign-symptoms; the IPA using its knowledge tries to provide the assistance by suggesting treatment plan(s). In certain cases, the IPA finds it difficult to diagnose a disease, so it consults the appropriate SSA. The IPA needs to demonstrate intelligent behavior for the selection of a suitable SSA. It can utilize various decision making techniques. Before elaborating on other issues, it is essential to understand the behavior, capabilities or responsibilities of these agents. In the subsequent subsection, roles and capabilities of these agents are discussed.
Figure 3.1 Abstract view of the MAS System.

Capabilities of Agents

- **User Agent**

  This agent is designed to interact with the rural doctor on one end and the IPA on the other. It also provides a graphical user interface to the doctor. The main tasks needs to be performed by it in a chronological order are discussed below:

  - Register the rural doctor with the IPA.
Multi-agent model for infant and child care

- Pose some basic questions related to the condition of the patient to the general doctor,
- Receive information filled by the rural doctor and pass that to the IPA,
- Display further queries asked by the IPA,
- Again receive the inputs and pass the information to the IPA,
- Finally display the treatment plan suggested by the IPA.

- **Intelligent Pediatric Agent**
  
  This agent is supposed to be fully automated, i.e. no human intervention is required. It is to imitate the behavior of a pediatrician. The pediatrician serves as a referral physician to rural doctor and helps in diagnosing the disease and suggests the treatment plan. This agent utilizes the diagnostic protocol discussed in section 2.1. All these functionalities will be kept while designing this agent. The actions it performs are discussed below:
  
  - To interact with the UA on one side and the SSA on the other.
  - To suggest the treatment plan as per the sign-symptoms provided by the UA.
  - In a complicated case, use decision making technique(s) to decide which SSA is to be contacted.
  - To pass the values of sign-symptoms to the SSA for diagnosing the disease and the treatment plan.
  - To contact the UA with the treatment plan as suggested by the SSA.

- **Super Specialist Agent**
  
  This agent is to be designed to accept sign-symptoms from the IPA and display them to the super specialist physician. The specialist decides the disease by asking more questions to the IPA and finally it passes the diagnosed disease and its treatment plan to IPA. The capabilities and roles performed by it are listed below:
  
  - Interact with the IPA on one side and super specialist physician on the other.
Display the sign-symptoms provided by the IPA.

Using the interface of this agent, the super specialist physician may ask for more values of various sign-symptoms to the IPA.

Pass the diagnosed disease and its treatment plan to the IPA.

The main task of this agent is to provide an interface between the IPA and the super specialist physician.

### 3.3 Uncertainty in the selection of an agent

Firstly, there is a necessity to understand the definition and the source of uncertainty. It can be defined as “the state of being unsure of something”. For instance, a statistician would like to estimate the amount or percentage by which an observed or calculated value may differ from the actual value. The main source of uncertainty is randomness and fuzziness.

Usually it is observed that during an experiment or observation (generally referred to as “test”), even if conducted under exactly the same condition(s), might bring out totally different results. For instance, a dice is thrown; one may get one of the six possible results although it is thrown under the same condition. This phenomenon is called “randomness.” It is characterized by repeatability and uncertainty on the result although we know all possible choices. So there must be some mathematical formula that provides an insight into the situation and helps in predicting the behavior of a system with an acceptable accuracy.

In fuzzy logic, linguistic variable enables its value to be described both qualitatively (name of the fuzzy set) by linguistic term and quantitatively (expresses the meaning of a fuzzy set) by a membership function. For instance, the sentence “The amount of rain is heavy”. *RainAmount* is the linguistic variable where *heavy* is the fuzzy set. Other fuzzy sets for the *RainAmount* variable can be: *low, normal, very heavy* etc.
Uncertainty in the selection of an agent

Such kind of usage does not provide clear boundaries among sets, i.e. there is an overlapping of one membership function on another. Hence it creates uncertainty.

Now with reference to the MASICC, during normal diagnosing procedure, the UA and the IPA communicate with each other and share a common ontology for knowledge sharing. The base of this ontology is derived from [16]. This diagnostic mechanism is limited to some of the diseases. In certain other cases, the UA is to send value(s) of sign-symptoms and the IPA uses decision making techniques to select a SSA. This decision making is very crucial for the effectiveness of the MASICC. Such cases arise only when the IPA seeks help from the Super Specialist Agent (SSA) to diagnose the disease and then to decide the treatment plan. The situation is characterized by uncertainty due to randomness of occurrence of sign-symptoms. So in the next section, the Bayesian network approach and the Neural network approach is elaborated.

3.4 Decision making techniques in multi-agent system

In this section, probability based decision making technique and Neural Network based techniques are discussed. Such a mechanism is required to be embedded in the MAS so that the agent can imitate human behavior while dealing with uncertain circumstances.

3.4.1 Probability theory and Bayesian Network (BN)

Human behavior tends to behave in a random fashion due to the partial knowledge or absence/presence of certain other factors. Probability theory is the main mathematical tool to solve the problem of randomness.

The theoretical aspect of the BN lies in the interpretation of Bayes' Theorem, that is stated below:

\[
P(h | e) = \frac{P(e | h)P(h)}{P(e)}
\]
This is a well defined theorem of the probability calculus. It asserts that the probability of a hypothesis \( h \) conditioned upon some evidence \( e \) is equal to its likelihood \( P(e|h) \) times its probability prior to any evidence \( P(h) \), normalized by dividing by \( P(e) \) (so that the conditional probabilities of all hypotheses sum to 1). A detailed view of this is presented below:

Suppose that the sample space \( S \) is divided into a collection of \( n \) mutually exclusive events (sets) called a partition of \( S \):

\[
S = \{x_1, x_2, \ldots, x_n\}, \quad x_i \cap x_j = 0 \quad i \neq j
\]

Assuming an arbitrary event \( y \) in \( S \), as shown in Figure 3.2 below:

The event \( y \) can be written as the union of the \( n \) disjoint (mutually exclusive) events \( y \cap x_1, y \cap x_2, \ldots, y \cap x_n \) i.e.

\[
y = y \cap x_1 + y \cap x_2 + \ldots + y \cap x_n
\]

This implies the total probability as:

\[
P(y) = P(y \mid x_1)P(x_1) + P(y \mid x_2)P(x_2) + \ldots + P(y \mid x_n)P(x_n)
\]

The total probability theorem and the definition of the conditional probability may be used to derive Bayes theorem:
Decision making techniques in multi-agent system

\[ P(x_i | y) = \frac{P(y | x_i)P(x_i)}{P(y)} = \frac{P(y | x_i)P(x_i)}{P(y | x_1)P(x_1) + P(y | x_2)P(x_2) + \ldots + P(y | x_n)P(x_n)} \]

The Bayes' rule updates \( P(x_i) \), given the information on the probabilities of obtaining 'y' when outcomes are \( x_1, x_2, \ldots, x_n \).

Having presented the theoretical reasons for artificial intelligence to use probabilistic reasoning, now the key computer technology for dealing with probabilities, namely BNs is presented. A BN is a graphical structure, shown in Figure 3.3, representing an uncertain domain. The symbol ‘?’ provides the mechanism of setting evidence for a particular node. Let us consider an example, a dental patient complains about the ‘food lodgment’ and informs about his high consumption of sweets. Now food lodgment can be a symptom of many other dental problems. A dentist would like to conduct X-ray to diagnose dental caries as a potent problem.

![Figure 3.3 Bayesian Network representing a dental caries problem](image)

The nodes in a BN represent a set of random variables from the domain. A set of directed arcs (or links) connects pairs of the nodes, representing the direct dependencies among variables. Assuming discrete variables, the strength of the relationship among variables is quantified by conditional probability distributions associated with each node. There is a constraint on the arcs that there must not be any directed cycles, i.e. one cannot return to a node simply by following directed arcs. In other terms, BN is constructed as a directed acyclic graph.
In short, a BN is composed of a qualitative as well as a quantitative part. The qualitative part is an acyclic directed graph reflecting typically the causal structure of the domain; the quantitative part represents the joint probability distribution over its variables/nodes. Every variable consists of a Conditional Probability Table (CPT) representing the probabilities of each state, given the state of the parent variable. If a variable does not have any parent variable in the graph, the CPT represents the prior probability distribution over the variable. A BN is capable of calculating the posterior probability distribution over an uncertain variable, given some evidence obtained from the related variables.

After representing a domain and its uncertainty, let us look how to use the BN to reason about the domain. During observation, particular variable(s) stand true; it means that they become the evidence for the network. Now, how these evidences affect the reasoning is vital to understand. In simple words, how information flows in the network. It is worth observing that this information flow is not limited to the directions of the arcs. A brief description about the various forms of reasoning using an example, wherein DC stands for dental caries, FL stands for Food Lodgment, S stands for high intake of Sweets, BOH stands for Bad Oral Hygiene, X stands for X-rays, is as follows:

**Diagnostic reasoning**: Reasoning from symptoms to cause, such as when a dentist knows about food lodgment and then he updates his belief about Caries and queries about the oral hygiene of the patient. This reasoning occurs in the opposite direction to the network arcs, Figure 3.4 (a).

**Predictive reasoning**: Reasoning from new information about ‘causes’ to the new beliefs about ‘effects’, following the directions of the network arcs. For example, the patient may tell his dentist that his ‘intake of sweet is high’; even before any symptoms have been assessed, the dentist knows this will increase the chances of the patient carious teeth, Figure 3.4 (b). This structure is created with the help of a tool called GeNIe, [63].
Intercausal Reasoning: Reasoning that involves reasoning about the mutual causes of a common effect; this is termed as intercausal reasoning. It is also known as explaining away. For instance, the effect of the patient having ‘high intake of sweet’ and having ‘bad oral hygiene’ have same amount of effect on dental caries. This means that there is no connection between these two causes. Given that tooth is carious and patient does not possess good oral hygiene lower the probability of patient consuming high quantity of sweets. So, even though the two causes are initially independent, with knowledge of the effect the presence of one explanatory cause renders an alternative cause less likely.
Combined reasoning: Since any node may be a query node and any other may be evidence node, so this type of reasoning is termed as combined reasoning. This is shown in Figure 3.4 (d). It combines the diagnostic and predictive reasoning.

The structure of the BN and the way one can reason out with BN makes it suitable in tackling uncertainty in medical domain. But the main disadvantage of the BN is the amount of probabilities required to fill in CPT. If a node has many parents or the parents can take a large number of states, the CPT can get very large. The size of the CPT is, in fact, exponential to the number of the parents. Thus, for the Boolean networks, a variable with 'n' parents requires a CPT with $2^{n+1}$ probabilities [67].

Even though the BN is suitable technique for building diagnostic models yet there is a need to goad other techniques too. In the next section, the neural network based techniques are discussed.

3.4.2 Introduction to Artificial Neural Networks

Artificial Neural Network or Neural Network is the interconnection of the neurons. The usefulness of the NN has been inspired from the human brain that computes in an entirely different way from the normal digital computers. The human brain is capable of solving highly complex, nonlinear problems in a parallel fashion [65]. So, in computers terminology, the NN is nothing but simulation of Biological Neuron Network (BNN). During such experiment, the emphasis is laid on structural simulation, simulating the network structure of the human biological nervous system. The BNN is in the brain of human beings and helps to process the information from one neuron to another. And hence it develops reasoning capability and other forms of intelligence. The same behavior is imitated by the NN. The NN is an adaptive system that changes itself and it is based on information that flows through it.

The ANN is an abstract mathematical model of the human brain and its activities. It is composed of numerous discrete processing units interconnected through line. The processing element, i.e. neurons, of the neural network is capable of local storage and local operation. It can be regarded as a directed graph with the neurons as the nodes interconnected by a weighted directed arc (link). In the directed graph, processing
element is the simulation of a biological neuron, while directed arc is the simulation of an “axon-synapse-dendrite” and the weight of the directed arc represents the strength of the interrelation between two processing units. Figure 3.5 is a diagram of the ANN, in which the input from another neuron, times the weight, is summed up and compared with the threshold value. When the sum is greater than the threshold value, the output is 1, which corresponds to the excitement; otherwise the output is 0, corresponding to inhibition.

\[
f(x_1, x_2, \ldots, x_n) = \begin{cases} 
1, & \text{if } \sum_{i=0}^{n} w_i x_i > 0 \\
0, & \text{otherwise}
\end{cases}
\]

**Figure 3.5** The mathematical description of artificial neural network

Because of the immense utility of the NN, it is used in applications areas like pattern recognition, classification and function approximation. The usage of the NNs is not only limited to the branch of computing but also widely used in other fields like physics, chemistry, mathematics, economics, psychology, neurophysiology, medicine, and many others. In the proceeding sub section, we discuss the basics of the back propagation NN and the probabilistic NN.
**Back Propagation Neural Network**

A Back Propagation Neural Network is first introduced in [68] and revised in [66]. This NN belongs to the error correction learning type, whose learning process is made up of three phases:

1. the feed-forward of training pattern
2. the back propagation of the associated error, and
3. the adjustment of weights.

The error is propagated backward when it appears between the input and the expected output during the feed-forward process. During the back propagation, the connection weight values among each layer of neurons are corrected and gradually adjusted until the minimum output error is reached. This is shown in Figure 3.6.

![Figure 3.6 Basic learning mechanism of BPNN](image)

The BPNN commences with a random set of connection weights. The network regulates its weights according to some learning rules each time it observes a pair of input-output vector. Every pair of vectors goes through two phases of activation: a forward pass and a backward pass. In forward pass, the sample input is applied to the network and the activation function produces the output that is flown until they reach the output layer. During the backward pass, the network’s actual output is compared with the target output and errors are compared for the output layer neurons. The weights connected to the output neurons can be adjusted in order to reduce those errors. The error
estimates of the output layer neurons are then used to obtain error estimates for the neurons in the hidden layer. After that all errors are back propagated to connections originating from the input layer neurons. After each round of forward–backward passes, the system learns incrementally from the input-output pair and gradually reduces the error between the network’s predicted output and the actual output. One can refer to [69], for an elaborate example.

Although one can adjust the neural network to lower its errors but can never be sure of lowering of the error. The aim of network training is to find the lowest point in many-dimensional surface. Since the NN error surfaces are very complex so it is impossible to determine analytically where the global minimum of the error surface is, so the NN training is essentially an exploration of the error surface. Typically, the slope of the error surface is calculated at the current point, and is used to make a downhill move. Finally, the algorithm stops in a low point, which may be a local minimum.

**Figure 3.7 Back Propagation Algorithm**

```
Start : begin epoch
  for inp=1 to m
    for each training vector Xp
      Execute Forward Pass
      Execute Backward Pass
    end for
  end for
end epoch

Calculate the error $Err = \sum_{p=1}^{m} E_p$

- if $Err$ is acceptably low then stop
- else goto start

End
```

The algorithm, which is shown in Figure 3.7, progresses iteratively through a number of epochs. On each epoch (an epoch corresponds to a given number of training points) the error, together with the error surface gradient, is used to adjust the weights, and then the process repeats. The initial network configuration is random and training stops when a
given number of the epochs passes by, or when the error reaches at an acceptable level, or when the error stops improving.

More specifically, the training process is terminated if one of the following conditions hold true:

1. In the $N$th cycle of the learning process, when the mean variance of the output is smaller than the prefixed allowable value.

2. When the gradient of the weighted value vector is smaller than the prefixed threshold.

3. At the end of each learning cycle, verify the network to see if the application ability has met the prefixed objectives.

4. Integrate the above three methods and determine one stopping condition so as to avoid the weighting value vector oscillating constantly and being unable to converge or converging slowly.

In the next subsection, probabilistic neural network is presented. This is particularly important to this research work as Bayes’ theory forms the core of this algorithm.

**Probabilistic Neural Network**

Probabilistic Neural Network is a type of the NN, which consists of three layers of nodes; the input layer, the hidden layer and the output layer [52]. More specifically, PNN is a type of the NN which uses probability factor for pattern recognition problems. It uses supervised training to assemble distributed functions contained by a pattern layer. These functions are used to approximate the possibility of an input class vector being a part of the learned group. Now these patterns can be combined with the probability of each category to determine the most likely class for the input vector. A pattern layer is an implementation of Bayes Classifier [53] in which probability functions are approximated, using a Parzen estimator [48]. This is optimum approach of pattern classification and it minimizes the risk of wrong classification. The basic structure of the PNN is shown in Figure 3.8.
PNNs are based on well defined probabilistic methodology which states that an observation \( I \) is placed in a class \( k \), that is different from \( j \) if

\[ \pi_k f_k(I) > \pi_j f_j(I) \quad \text{for all } j \neq k \]

where \( f_k(I) \) is the probability density function of the \( kth \) class, and \( \pi_k \) is the prior probability that the observation belongs to the \( kth \) class.

Usually, \( f_k(I) \) is not known and is estimated using the training set. A nonparametric estimator called Parzen window is employed to calculate \( f_k(I) \) as

\[
\sum_{i=1}^{n_k} \exp\left(-\frac{(I - I_{ki})'(I - I_{ki})}{2\sigma^2}\right)
\]

where \( I_{ki} \) is the \( ith \) sample from the \( kth \) class and \( \sigma \) is the smoothening factor.

![Figure 3.8 Basic structure of Probabilistic Neural Network](image)

The PNN training is much simpler than the BPNN and enjoys the advantages such as, fast training, guaranteed convergence to a Bayes' classifier for large dataset, can be easily
enhanced to accommodate new data, and demonstrate robustness to noise [64]. However, the pattern layer can be fairly huge if the distinction between categories is varied and at the same time quite similar in special areas.

### 3.5 Conclusion

In the beginning of this chapter, the overall structure of the MAS is presented. The model depicts the location of the agents and underlines the roles and responsibilities of the agents. It is observed that, the IPA has to behave intelligently as it has to communicate with the UA on one end and SSA on the other. During communication with the UA, it is to use childhood disease ontology to diagnose diseases and before initiating communication with one of the SSA, it is to decide appropriate SSA. This intelligent mechanism can be probability based or neural network based. So, both the solutions are goaded. These techniques have different requirements and can be effective under specific conditions.