WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

AN OVERVIEW OF PATTERN RECOGNITION

Haritha C V

Pursuing bachelor's in Computer Science at Presidency University.
Presidency University, haritháreddy1314@gmail.com

DOI: https://doi.org/10.55248/gengpi.2022.3.7.49

ABSTRACT

The operation and design of systems that can recognize a pattern in data are the subjects of the Al research field of pattern recognition. Pattern recognition has a wide range of practical uses. These consist of fraud detection, fraud prevention, handwriting recognition, facial recognition, etc. Face recognition uses a person's face to unlock a device rather than a keyboard to enter a password. As a result, there are fewer chances that the passwords will be forgotten or leaked to an unauthorized person. Since their faces are unlikely to resemble the original user's saved face model, this also prohibits other people from using the computer.

The usage of handwriting in Al is handwriting recognition, one of the most promising ways to interface with small portable computing devices like personal digital assistants. They suggested visualizing the writing process on regular paper and automatically recovering the pen trajectory from numerical tablet sequences to make this communication technique appear more natural.

1. INTRODUCTION

Establishing a tight match between a new stimulus and previously-stored sensory patterns is the process of recognition. Continuously, this procedure is carried out All living creatures go through a process of recognition throughout their lifetimes. Higher animals have a variety of conscious and unconscious manifestations of this ability, which applies to both physical and abstract items. We can recognize a wide variety of unique objects through visual perception and recognition, including faces of people, handwriting, printed words, and places like homes, offices, schools, and restaurants. We recognize recognizable voices, songs, and musical compositions, as well as bird and other animal, sounds through aural perception and recognition. We distinguish between tangible objects like pens, cups, car controls, and food by touch. We also recognize foods using our other senses, At higher levels of cognition, we are able to recognize or distinguish things like ideas (such as the electromagnetic radiation phenomenon, the model of the atom, and world peace), concepts (such as generosity, beauty, and complexity), procedures (such as playing games or depositing money in a bank), plans, stale arguments, metaphors, and more. Our dependence on and widespread usage of pattern recognition has spurred extensive study into the development of mechanical or artificial techniques that are analogous to those employed by intelligent creatures. Numerous applications have resulted from these efforts, and the outcomes to date have been outstanding. A variety of medical and system diagnoses, resource identification and evaluation (geological, forestry, hydrological, crop disease), character and speech recognition, fingerprint and photograph identification, electroencephalogram (EEG), electrocardiogram (ECG), oil log-well, and other graphical pattern analyses, as well as the detection of explosive and hostile threats (submarine, aircraft, missile), have all been made possible by the development of reliable systems.

Recognition and object classification go hand in hand. Class recognition is the ability to categorize or group items based on some often occurring shared characteristics. Making decisions, learning, and doing numerous other cognitive tasks all depend on classification. Similar to recognition, classification relies on the capacity to identify shared traits across items. This capability, in turn, needs to be learned in some way. It is necessary to identify, generalize, and store prominent feature patterns that define classes of objects for later recall and comparison.

Although scientists do not fully understand how humans learn to recognize or categorize objects, it seems the following stages occur:

Humans are exposed to new objects through the stimulation of their senses. The sensors' sensitivity to particular characteristics that serve to identify the objects varies based on their physical characteristics, and the sensor output tends to be proportionate to the more noticeable characteristics. After a new object is perceived, a cognitive model is created from the stimulus patterns and is then memorized. The similarity patterns are strengthened and improved by repeated experiences with perceiving the same or similar objects. The development of generalized or archetypal models of object classes as a result of repeated perception makes them effective for matching. resulting in the identification of related objects.

2. THE RECOGNITION AND CLASSIFICATION PROCESS

Practically the same actions outlined above must be taken in artificial or mechanical recognition. The same steps as previously described are illustrated and are summarized below:

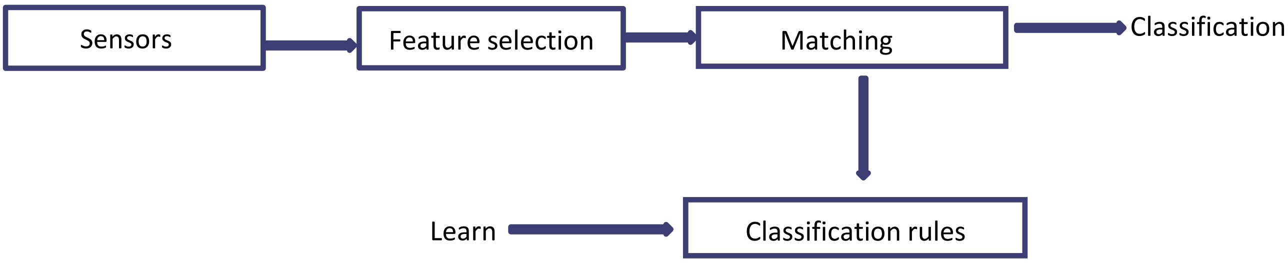
STEP 1: Sensory devices pick up on the stimuli that objects emit. the most powerful stimulation A pattern vector X, a string produced by a grammar, a classification tree, a description graph, or some other kind of representation are all used to characterize an item based on the values of these characteristics and their relationships. The measuring space M is the range of values for the characteristic property.

STEP 2: A collection of attributes is chosen whose values support coherent object grouping or clustering and are in line with specified objectives related to the object classifications. The characteristics used to result in large intraclass and low interclass groups. This subset reduces the dimensionality of the attribute space and hence makes the classification process simpler. The feature space F is the range of the subset of attribute values.

STEP 3: Using the chosen attribute values, generalized prototype descriptions, classification rules, or decision functions are learned to characterize an item or a class. For later recognition, these models are saved. The decision space D is the range of the categorization rules or decision function values.

STEP 4: By comparing and matching object features with the models that have been saved, the rules that were learned in STEP 3 are used to recognize familiar things. After that, improvements and tweaks can be made continuously to enhance the accuracy and speed of recognition.

The recognition problem can be approached from one of two primary angles: decision theory or syntactic analysis.



3. DECISION THEORETIC CLASSIFICATION

The decision-theoretic method bases object classification on the usage of decision functions. Pattern vectors X are translated into decision regions of D by a decision function.

This issue can be defined more precisely as follows.

1. Given a universe of objects 0= (01, 02, ...n), suppose each o has k observable properties and relations that are expressed as a vector V (U2).

2. Find (a) subsets of MK of the 1, say X= (1. 32.), whose values uniquely characterize the o, and (b) c 2 groupings or classifications of the o with high intraclass and low interclass similarities, such that a decision function d(X) can be discovered that partitions D into e disjoint parts. The regions are used to assign each o to one of the c classes at most.

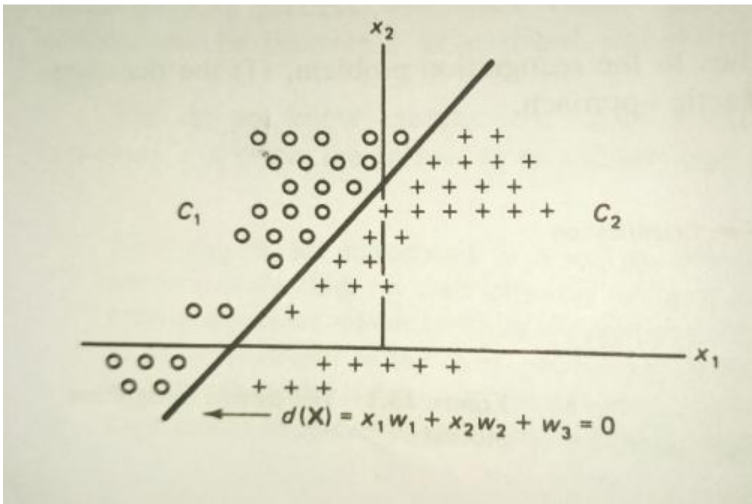
Determining the feature attributes and decision regions requires stipulating or learning mappings from the measurement space M to the feature space F and then a mapping from F to the classification or decision space D

$$M \longrightarrow F \longrightarrow D$$

When there are just two classes, say C and C2, the pattern vector values of the object may tend to cluster into two distinct groups. In this scenario, a linear decision function d(X) is frequently employed to determine the class of an object. When the classes are clustered, as shown in Figure 13.2, a linear decision function d is sufficient to categorize unknown objects as belonging to C1 or C2,

Where d(X) = w1x1 + W2x2 + Ws

The constants w in Dare parameters or weights that are modified to find a class dividing line. When a function such as d is applied, an item is classed as belonging to C if its pattern vector is such that d(X) = 0, and otherwise as belonging to D.



When d(X) > 0, you belong to class C. When d(X)=0, the classification is uncertain, and either (or both) classes can be chosen. When class reference vectors R.J=1..e are available. The distance of the X from the reference vectors can be used to build R1. j decision functions.

For example, the distance

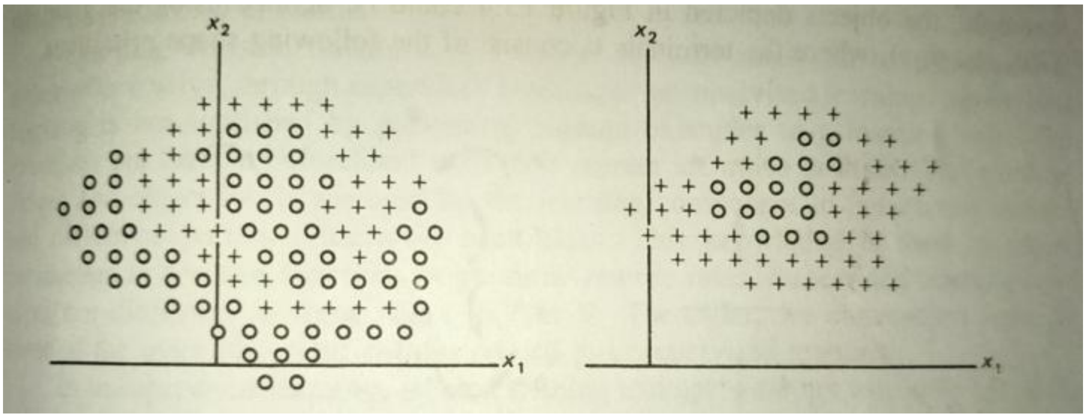
d(X) = (XR)'(X - R.)

could be computed for each class C. and class C, would then be chosen when d = min (C). For the general case of c2 classes. C_1, C_2... C.. a decision function may be defined for each class d. DZ. de A class decision rule, in this case, would be defined to select class C, when

d(X)<d(X) for ij 1.2..... c. and ij.

When a line d (or, more broadly, a hyperplane in n-space) can be found that divides classes into two or more groups, as shown in Figure 13.2, we say the classes are linearly separable. Classes that overlap or surround one another, as shown in Figure 13.3, cannot be classified using simple linear decision functions. More general nonlinear (or piecewise-linear) functions may be necessary for such instances. Alternatively, another selection strategy (such as heuristics) may be required.

Because x is deterministic variable, the decision function approach given above is an example of deterministic recognition. In circumstances where the attribute values are influenced by noise or other random fluctuations, probabilistic decision functions may be more appropriate. The attribute vectors X are treated as random variables in such circumstances, and the decision functions are specified as measures of the likelihood of class inclusion. Using Bayes' rule, for example, one can compute the conditional probability P(CX) that the class of an object o, is C, given the observed value of X for o. This method necessitates knowledge of the prior probability PIC), and the likelihood of the occurrence of samples from C, and P. (XIC).



(Note that the C are handled as random variables in this case.) This is analogous to the assumption made in Bayesian classifi cation, where the distribution parameter is assumed to be a random variable because C can be viewed as a function of 6). In this si tuation, a decision rule is to select class C if

PC, IX)> PC, X) for all I j

A more extensive probabilistic strategy is one based on the use of a loss or risk Bayesian function, with the class chosen ba sed on the lowest loss or risk. Let L represent the loss suffered by wrongly classifying an object that truly belongs to class C as belonging to C. When is a constant for everyone? I . j. I j. Using the likelihood ratio, a decision rule can be developed

$P(X C_2) P(XIC)$

The rule is to choose class C, whenever the relation

P(XIC) P(C) holds for all jk. P(X C) P(C)

Depending on the distribution forms, probabilistic decision rules can be designed as either parametric or nonparametric. For a more in-depth discussion of these strategies, see (Duda and Hart, 1973) or (Tou and Gonzales, 1974)

4. DESIGN PRINCIPLES OF PATTERN RECOGNITION SYSTEM

The goal of pattern recognition is to identify data (patterns) based on a priori knowledge or statistical information retrieved from the patterns. The patterns to be classified are often sets of measurements or observations defined by defining points in a suitable multidimensional space. A full pattern recognition system comprises the following components:

1. **SENSORS:** These are used to collect the observations that will be categorized.
2. **TASKS FOR DESCRIPTION:** It is a mechanism for extracting features. It is a method for calculating quantitative or symbolic information from observations. The aim is to convert the data gathered from the environment into features. It consists of three tasks: data pre-processing, feature extraction, and feature reduction via feature selection. It generates a feature vector, which is a data representation for pattern recognition
3. **A CLASSIFICATION TASKS:** This scheme is used to classify observations based on the extracted features. A classifier is used to map a feature vector to a group. A mapping can be formulated using a training phase to induce the mapping from a collection of feature vectors in the training set and then it can be used to assign an identification to each unlabeled feature vector subsequently presented to the classifier

5. LEARNING CLASSIFICATION PATTERNS

A system must be aware of the distinguishing properties of items before it can recognize them. This means that the system designer must either create or learn the necessary discriminating rules. In the case of a linear decision function, the weights that form class boundaries must be specified or learned. In the case of syntactic recognition, class grammar must be predefined or learned.

Unsupervised or supervised learning can be used to learn decision functions, grammar, or other rules. Supervised learning occurs when training examples are presented to a learning unit. The examples are pre-identified with their correct identities or classes. The learning component extracts and derives pattern requirements for each class inductively using attribute values and object labels. This data is used to change parameters in decision functions or grammar rewrite rules. Part V digs deeper into the principles of supervised learning. As a result, in this section, we will concentrate on some of the more important principles connected to unsupervised learning.

It is utilized unsupervised learning. There are no label training examples available, and the object population is unknown beforehand. In such cases, the system must be capable of sensing and extracting relevant features from otherwise unknown items, as well as identifying common patterns among them and producing descriptions or discriminating criteria that are consistent with the goals of the identification process. Clustering is a term used to describe this form of learning. Learning via Clustering is the first stage of any recognition technique in which the distinguishing qualities of objects are unknown in advance.

Clustering is the practice of grouping or categorizing objects based on a close relationship or shared qualities. The objects might be physical or abstract entities, and the characteristics can be attribute values, object relationships, or a combination of the two. The objects might be streets, freeways, and other pathways linking two points in a city, and the classifications could be the pathways that enable quick or slow traversal between the points. At a higher level, the objects could be some abstract idea, such as the quality of the products purchased. In this scenario, classifications could be created based on subjective criteria such as poor, average, or good.

Clustering is simply a discovery learning process in which resemblance patterns among a group of objects are discovered. Pattern discovery is typicall y influenced by the environment or context and motivated by some purpose or ambition (even if only for the economy in cognition). Finding short-cuts between two commonly visited points, for example, is motivated by the need to reduce planning effort and transit time between the points. Similarly, developing a sense of quality is driven by a desire to save time and money or to improve one's looks.

Given various goals, the same collection of objects would be clustered differently. If the above-mentioned purpose for streets, freeways, and the like were adjusted to include safe for bicycle riding, a different object classification would result in general. Finding the most meaningful cluster groupings among a set of unknown items o necessitates the discovery of similarity patterns in the feature space. Clustering is typically done with the goal of capturing any gestalt features of a group of items, rather than merely the similarity of certain attribute values. This is one of the fundamental prerequisites of conceptual clustering (Chapter 19), which groups things together as members of a concept class. Conceptual clustering procedures are based on more than just simple rules . They must also consider the items' context (environment) as well as the clustering's aims or objectives. The clustering problem generates a number of subproblems. Before any implementation may take place, the following issues must be solved.

What traits and relationships are most important, and what weights should be assigned to each? Which traits should be seen or measured first? (If the observation process is sequential, the order of the qualities may affect their ability to discriminate between objects.)

1. **What representation formalism should be used to characterize the objects?**

2. **What representation scheme should be used to describe the cluster groupings or classifications" Usually, some simplification results if the single representation mick can be used (the use of a single representation method for both object and cluster descriptions).**

3. **What clustering criteria are most consistent with and effective in achieving the objectives relative to the context or domain? This requires consideration of an appropriate distance or similarity measure compatible with the description domains noted in 2, above.**

4. **What clustering algorithms can best meet the criteria in 2 within acceptable time and space complexity bounds?**

Questions like these should be familiar by now. They are not insignificant, but they must be considered while developing a system. They are dependent on a number of complex factors, for which the tools from previous chapters are crucial. These issues have been addressed elsewhere, so we will concentrate on the clustering process here.

Except with a small number of objects, the clustering procedure must be performed with a limited set of observations, therefore evaluating all possible object groupings for patterns is not practicable. This is related to the combinatorial explosion, which causes n objects to be arranged into an indeterminate number m of clusters. As a result, approaches that assess only the most promising groupings must be employed. Creating such groups necessitates the usage of some measure of similarity, association, or degree of fit among a set of objects.

When the attribute values are genuine, cluster groupings can be discovered using point-to-point or point-to-set distances, probability measures (such as the covariance matrix between two populations), scatter matrices, and the sum of squared error distance between objects, or other methods . In these cases, an item is classified as belonging to class C if its proximity to other members of the class is within a certain threshold or limiting value.

For various tasks, many clustering techniques have been proposed. The ISODATA technique is one of the most common algorithms created at the Stanford Research Institute by G. H. Ball and D. J. Hall (Anderberg, 1973). This approach requires the number of clusters m, as well as the threshold values 4 and 13 be given or determined for use in splitting, merging, or discarding it.

The number of ways in which n objects can be arranged into m groups is an exponential quantity.

When m is unknown, the number of possible arrangements grows as the total of the Sm. That is correct, Sm. When n=25, for example, the number of configurations is greater than 4*10. clusters, and so on. The thresholds are used during the clustering process to evaluate if a cluster should be split into two clusters, integrated with other clusters, or deleted (when too small) The following steps comprise the algorithm.

1. **Choose m samples as the initial cluster centers' seed locations. This can be accomplished by selecting the first m points, random points, or fine points that surpass some mutual minimum separation distance d.**

2. **Associate each sample with the cluster centers that are closest to it.**

3. **After grouping all samples, compute new cluster centers for each group. The centroid (mean value of the Annaba vectors) or some other central metric can be used to define the center.**

4. **If the split threshold for any cluster is surpassed, break it into two parts and recompute new cluster centers.**

5. **Combine the clutter and recompute new cluster centers if the distance between two cluster centers is less than 2.**

6. **If a cluster has fewer than, members, it is discarded and ignored for the rest of the operation.**

7. **Repeat steps 3–6 until no change occurs within-cluster groupings or an iteration limit is reached.**

Distance and center location measurements do not have to be based on co-ordered variates. Probabilistic or fuzzy metrics of similarity between graphs, texts, and even FOPL descriptions exist. In any scenario, each object A is believed to be described by a distinct point or event in the feature space F. We have so far neglected the issue of attribute scaling. In a similarity measure, a few large valued variables may totally dominate the remaining variables. This could happen if one variable is measured in meters and another in millimeters, or if the range and scale of variation for two variables are vastly different. This problem is closely connected to the feature selection problem, in which weights are assigned to feature variables based on their value or relevance. A diagonal weight matrix W can be used to modify the representation vector X to X =WX, which is a straightforward approach for altering the scales of such variables. As a result, for all of the above-mentioned measurements, one should assume that the representation vectors X have been adequately normalized to account for scale variances.

To summarize the preceding approach, a subset of typical attributes representing the o is initially chosen. The attributes used should be excellent discriminators in dividing things into distinct classes, relevant, quantifiable (observable), and affordable. To avoid swamping caused by big valued variables, feature variables should be scaled as described above. Following that, a suitable metric that assesses the degree of association or similarity between objects should be determined, as well as an acceptable clustering technique. Finally, the feature variables may need to be weighted during the clustering process to reflect the relative relevance of the feature in effecting the grouping.

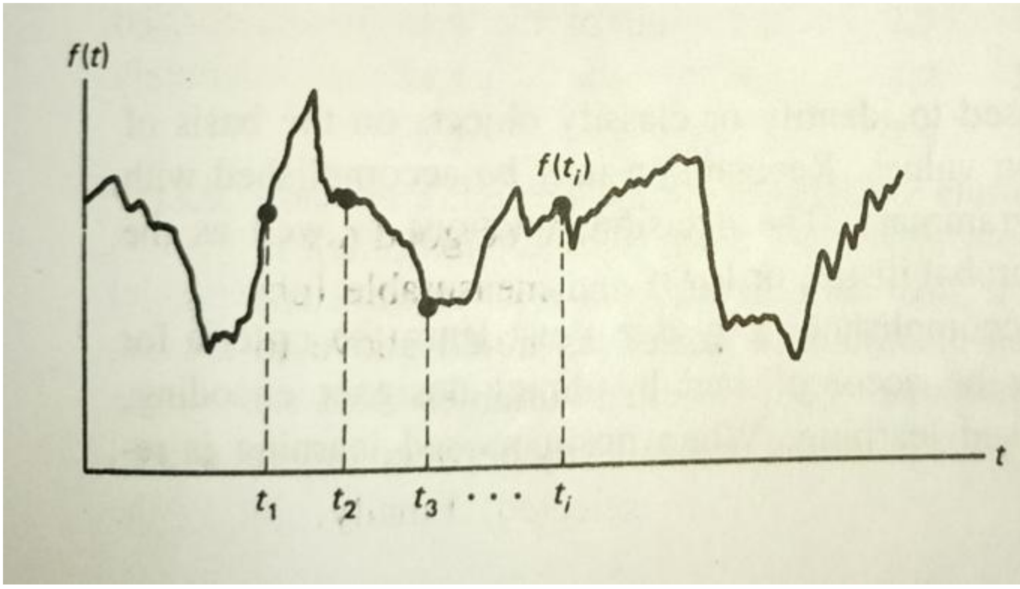
6. RECOGNIZING AND UNDERSTANDING SPEECH

Creating systems that comprehend speech has long been a focus of Al researchers. Speech is one of our most convenient and natural modes of communication, so it's understandable that we'd like Al systems to have it. There are various advantages to being able to interface directly with programs. It eliminates the need for keyboard input and speeds up information exchange between the user and the system. Users can accomplish other jobs concurrently with the computer interchange when speech is used as the communication medium. Finally, more untrained employees would be able to use computers for a variety of purposes.

The recognition of continuous waveform patterns, such as speech, begins with waveform sampling and digitization. The feature values in this situation are the sampled points = f(t) .Information theory tells us that a sampling rate twice the highest speech frequency is required to capture the information content of the speech wave forts. As a result, sampling requirements will often range between 20K and 30K bytes per second. While this velocity of information is not inherently difficult to handle, when combined with subsequent processing, it places significant demands on real-time speech understanding.

After sample digitization, the signals are handled at various abstraction levels. Allophones, which are variations of the phoneme as they actually occur in words, phones (the smallest unit of sound), and syllables are dealt with at the lowest level. Words, phrases, and sentences are the focus of higher-level processing.

Processing might be done from the top, bottom, or a combination of both. When bottom processing is applied, the input signal is divided into fundamental speech units before a search is conducted to compare restored patterns against these units. A lexicon contains information on the phonetic makeup of words so that comparisons can be made. For the top technique, context, semantics, and syntax are used to predict the words the speaker is most likely to have stated



look for recognized patterns with direction. Both approaches have also been used successfully in combination.

The use of early speech recognition research, which focused on the recognition of separate techniques, has been successful. Individual word patterns were restored, and the digitized input patterns were compared to them. These pioneering systems only had modest success. They were extremely sensitive to noise and unable to withstand changes in the speaker’s voice.

Despite being noteworthy, this early research did little to address the general issue of interpreting continuous speech because words that appear as part of a continuous stream differ greatly from isolated words. Words are combined, altered, and abbreviated in a continuous speech to create a wide range of sounds. Speech analysis must therefore be able to identify various sounds that are present in a single word but in various situations. Because of the noise and variability, recognition is best accomplished with some type of fuzzy comparison.

A five-year effort for continuous speech understanding research was financed by the Defense Advanced Research Projects Agency (DARPA) in 1971. (SUR). The goal of this project was to create and put into practice systems that could handle continuous speech from a number of cooperative speakers utilizing a small vocabulary of roughly 1000 words. The systems were predicted to operate at speeds slower than red time. The systems HEARSAY I and II, HARPY, and HWIM are a result of this research. The research created other significant by-products as well, particularly in systems architectures and in the knowledge learned regarding control, despite the systems' only mediocre effectiveness in fulfilling their goals.

The introduction of the blackboard architecture by the HEARSAY system was significant. This architecture is built on the collaborative efforts of many specialized knowledge components interacting over a blackboard to solve a class of issues. Each specialist is an authority in a separate field. For instance, different levels of speech issues may be addressed by various speech analysis experts. Each expert contributes as he is able during the opportunistic solution process. On the whiteboard, a data structure representing the answer to a given problem is built. This data structure is altered by the contributing expert as the solution is created. Barr and Feigenbaum provide a description of the SUR-developed systems (1981).

REFERENCES

[1] Artificial intelligence structures and strategies for complex problem-solving Book by George F. Luger

[2] https://g.co/kgs/sWmY7B

[3] Introduction to artificial intelligence and expert systems dan w. Patterson

[4] https://g.co/kgs/4DR6Qn

[5] Artificial intelligence building intelligent systems Book by PARAG KULKARNI and PRACHI JOSHI

[6] https://g.co/kgs/jyvsxG

[7] https://www.investopedia.com/terms/a/artificial-intelligence-ai.asp#:~:text=Artificial%20intelligence%20(AI)%20refers%20to,as%20learning%20and%20problem%2Dsolving.

[8] https://sloanreview.mit.edu/article/designing-ai-systems-with-human-machine-teams/