

Chapter 58

A Survey on Supervised and Unsupervised Learning Techniques



K. Sindhu Meena and S. Suriya

Abstract Supervised learning is the popular version of machine learning. It trains the system in the training phase by labeling each of its input with its desired output value. Unsupervised learning is another popular version of machine learning which generates inferences without the concept of labels. The most common supervised learning methods are linear regression, support vector machine, random forest, naïve Bayes, etc. The most common unsupervised learning methods are cluster analysis, K-means, Apriori algorithm, etc. This survey paper gives an overview of supervised algorithms, namely, support vector machine, decision tree, naïve Bayes, KNN, and linear regression, and an overview of unsupervised algorithms, namely, K-means, agglomerative divisive, and neural networks.

Keywords Supervised learning · Unsupervised learning · Support vector machine · Decision tree · Naïve Bayes · KNN · Linear regression · K-means · Agglomerative divisive · Neural networks

Abbreviation

KNN	K-nearest neighbor
WSD	Word-sense disambiguation
CNN	Convolution neural network
DT	Decision tree
NB	Naïve Bayes

K. Sindhu Meena · S. Suriya (✉)
Department of Computer Science and Engineering, PSG College of Technology, Coimbatore,
Tamil Nadu, India

58.1 Introduction (Fig. 58.1)

Supervised Learning Techniques: It involves two phases, namely, training phase and testing phase. This technique trains any system through a training phase, which consists of input data along with target results. After training phase, the system undergoes a testing phase using inputs without target labels. It is a crucial, fast, and accurate technique. Supervised learning is grouped into two divisions: regression and classification. Classification focuses on categorical variables like “black,” “white,” “yes,” and “no.” Regression focuses on variable that possesses real values like “money,” “height,” and “age” (Fig. 58.2)

Unsupervised Learning Techniques: It attempts to classify the input dataset into classes or clusters. The input dataset is a raw data without any class labels or target results. The raw data refers to unstructured dataset without class labels, optimization criteria, and feedback. The goal of unsupervised learning is to find natural partitions in the dataset set. Unsupervised learning techniques can be used for smart preprocessing and feature extraction systems (Fig. 58.3).

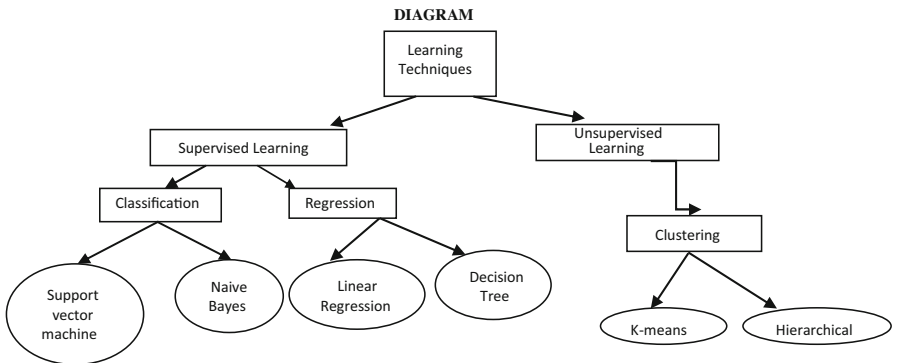


Fig. 58.1 The overall diagrammatic representation of learning techniques

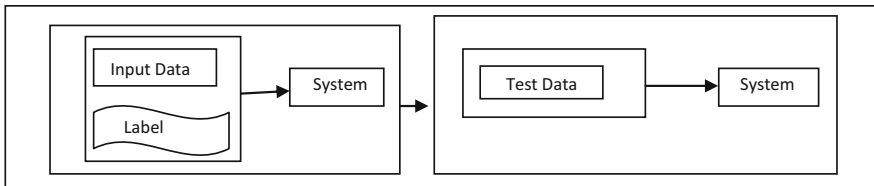


Fig. 58.2 The diagrammatic representation of supervised learning techniques

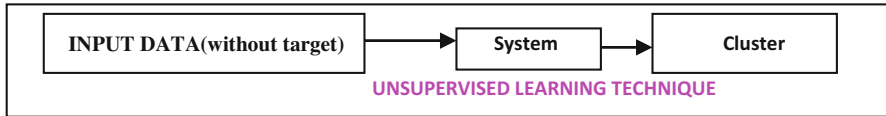


Fig. 58.3 Unsupervised learning techniques

58.2 Literature Survey

58.2.1 Supervised Learning

58.2.1.1 SVM

Paper [1] aims to promote research in sentiment analysis of tweets by providing annotated tweets for training, development, and testing. The objective of the system is to label the sentiment of each tweet as “positive,” “negative,” and “neutral.” They describe a Twitter sentiment analysis system developed by combining a rule-based classifier with supervised learning. Benefits of using this system are to maximize the positive and negative precision and recall; the rule-based classifier is used to correct or verify the neutral SVM predictions. This paper uses a modified version of SVM algorithm. Another paper [2] proposes the conception and implementation of computer-assisted detection (CAD) for the classification of mammographic images. It is to detect the breast cancer early using CAD. The application used here is Breast Cancer Classification. Benefits of using this system are to identify the breast cancer earlier. It classifies the nature of tumors in terms of benignity or malignancy. This paper uses a modified version of SVM algorithm. Paper [3] suggests an extraction of features for classifying the EEG signals into three stages. The first stage includes calculating the empirical mode decomposition (EMD) from the EEG signals and produces a set of intrinsic mode of function. To process further, we can make use of first three IMFs. In the next stage, the features are extracted by calculating the temporal and spectral attributes of IMF. The power spectral density is used for calculating the spectral attributes. The Hilbert-transformed IMF is used for fetching the temporal and spectral components. This transformation helps in removing the DC from the spectral attributes. The spectral signal is one of the main sources of nonstationary signals. The last stage makes use of support vector machine (SVM) for EEG signal classification. The application they have used here is for classifying the seizures and epilepsy in the EEG signals. The benefit of applying SVM classifier is to obtain the best accuracy. The basic algorithm is used here. Paper [4] describes an approach for automatic construction of dictionaries for named-entity recognition (NER). First, we collect a high-recall, low-precision list of candidate phrases from

the large unlabeled data collection for every named entity type using simple rules. The second step is to build the accurate dictionary of named entities by eliminating the noisy information from the list in the previous step. By using virus (GENIA) and disease (Dogan), the dictionaries are evaluated. It can be applied directly to the dictionary-based taggers. The advantage of this approach is different from binding belief about the data. SVD can be solved and better performance is achieved. It makes use of seed examples that are compiled manually. The modified version of SVM is used. Paper [5] aims to assess the neural predictors of long-term treatment outcome in participants with social anxiety disorder (SAD) 1 year after the completion of Internet-delivered cognitive behavioral therapy (CBT). SVM allows the predictions at the individual level, where it is trained to separate long-term treatment responders from nonresponders based on blood oxygen level-dependent (BOLD) responses to self-referential criticism. It can be predicted by using fMRI and SVM classification. The advantage of using SVM is that it provides accurate prediction of treatment outcome at the individual level. The modified version of SVM is used here. Paper [6] presents an effective method for gene classification using SVM. It solves the complex classification problems. Mutual information (MI) between the genes and the class labels is used for identifying the informative genes. The selected genes are utilized for training the SVM classifier, and the testing ability is evaluated using the leave-one-out cross validation (LOOCV) method. It reduces the dimensions of the input features by identifying the most informative gene subset and improves classification accuracy. This paper uses a basic SVM model. Paper [7] concentrates on the accuracy of least square support vector machine (LSSVM) and multivariate adaptive regression splines (MARS) and M5 model tree in identifying the river water pollution. The LSSVM and MARS provide the equal accuracy, and better performance is acquired compared to the M5 model tree. The monthly water pollution level from the river is obtained by using ammonia (AMM), water temperature (WT), and total Kjeldahl nitrogen (TKN) attributes. The application they have used is long-term prediction river water pollution. The modified version of SVM is used.

58.2.1.2 Decision Tree

Paper [8] explains the medical care quality and the status, development, and variation related to the quality by data. Data mining helps to find out the problems of management, various possible causes of the problem, and the corresponding solution strategies. There are two main methods: first one is using statistical methods to analyze a certain medical quality indicator, and another one is multidimensional comprehensive evaluation methods like rank sum ratio method and synthetic index method. The C4.5 algorithm helps to select hospital's inpatient homepage as training set and test set to realize the predictive parsing of hospital's medical quality indicator vividly by graphical means and finally calculates the accuracy of prediction by test set. The decision tree is the algorithm used. Paper [9] intended to demonstrate the data mining technique in the disease prediction systems in medical domain. The

heart disease-based data is selected for analysis and prediction. It gives patients a warning about the probable presence of heart disease even before he visits a hospital or tends toward a costly medical checkup. Benefits of using a decision tree are that it is more accurate than other methods and it classifies and predicts the heart disease. The decision tree algorithm is used here. Paper [10] uses the sentiment analysis to study the people's opinions, attitudes, and emotions toward an entity. Opinion helps to collect information about the positive and negative aspects of a particular topic. Finally the positive and highly scored opinions obtained about a particular product are recommended to the user. There are several challenges in sentiment analysis. The first is an opinion word that is considered to be positive in one situation may be considered negative in another situation. The second challenge is that people do not always express opinions in the same way. The basic model of decision tree is used here. Paper [11] focuses on the issues of leakage of information which impacts on both financial organization and customers. And it proposes a new approach that uses combined supervised learning techniques to classify the information in order to avoid releasing information that can be harmful for either financial services providers or customers. They entitled this approach as Supervised Learning-Based Security Information Classification (SEB-SIC). The goal will be achieved by using the supervised learning techniques to predict whether the information sharing will be hazardous for any relevant parties. The prediction process will be executed by decision tree-based risk prediction (DTRP). The modified version of DT is used here. Paper [12] describes the network traffic in the Internet, either based on the user demand for network resources, QoS scheduling, or development trend of network applications for expansion transformation of the existing network; here various network applications need to be classified and identified accurately. The benefit of using Hadoop platform is that it will improve the C4.5 decision tree algorithm parallelization. It improves parallel algorithm and has faster speed and higher accuracy. The modified version of decision tree is used here. Paper [13] attempts to propose a supervised approach decision tree for word-sense disambiguation task in Assamese language. This paper aims to disambiguate the words which have multiple sense in a context automatically. Some challenges like sense inventory along with their senses were discovered, and sense-annotated data as a training sample was manually prepared. Precision, recall, and F-measure were used as metrics for WSD. Assamese is a less computational-aware language, and WSD task using a supervised approach DT with cross-validation evaluation was the first initiative toward Assamese language. This will provide a remarkable contribution to NLP fields. The basic model of DT is used here. Paper [14] aspires an effective spam filtering technique related to the decision tree in the data mining technique. The spam is analyzed by using the association rules. The rules are applied for developing the systematized spam filtering method. Benefits are as follows: the revision information learned from the analysis of misjudged emails incrementally gave feedback to their method and its ability to identify spam would be improved. The basic model of DT is used here. In paper [15], a new supervised one-side sampling technique to preprocess the imbalanced dataset is used to overcome three main difficulties. First, the customer churn dataset is substantially imbalanced in reality. Second, the samples in

feature space are relatively scattering. Third, the dimension of feature space is high and dimension reduction is necessary for algorithm efficiency. C5.0 decision tree is the classifier applied in this study to predict customer churn in two or three months based on current information. Benefit of this method is that it retains the potential lost customer. The modified version of decision tree is used here.

58.2.1.3 Naïve Bayes

Paper [16] proposes a scheme for adopting the Laplace smoothing technique with binarized naïve Bayes classifier for enhancing the accuracy and employing SparkR for speed up via distributed and parallel processing. Binarized NBC (BNBC) was developed to solve this problem by not counting the frequency of each word but checking its existence. Their proposed approach effectively reflects this notion in the calculation of the probabilities of the classifier. The benefit of using this application is that if a word not existing in the training data appears, the probability of 0 occurs during the test of new document. This results in decreased accuracy of NBC. Laplace smoothing has thus been adopted to resolve this problem. In sentiment analysis, the presence of a word is more important than its frequency. Paper [17] proposes a simple, efficient, and effective feature weighting approach, called deep feature weighting (DFW), which estimates the conditional probabilities of naïve Bayes by deeply computing feature weighted frequencies from training data. The authors focus their attention on feature weighting approaches and find that all of the existing feature weighting approaches only incorporate the learned feature weights into the classification of formula of naïve Bayes and do not incorporate the learned feature weights into its conditional probability estimates at all. Naïve Bayes with our deep feature weighting rarely degrades the quality of the model compared to standard naïve Bayes and, in many cases, improves it dramatically. Finally, the paper applies the deep feature weighting to some state-of-the-art naïve Bayes text classifiers and has achieved remarkable improvements. In paper [18], the goal is to diagnose the diabetes types and the level of risk of the diabetic patients. The data mining techniques are used like clustering and classification. This paper claims in creating the expert clinical system to diagnose diabetes mellitus. The diabetes type and risk of the every patient can be analyzed by using the data mining techniques. Paper [19] aims to effectively detect Android malware based only on requested permissions. The authors explored machine learning techniques used to learn and train application profiles and to detect and predict application status: either malicious or normal. They evaluated and discussed the system with renowned antiviruses. Their accomplishments are good in detecting the accuracy. It is 98% accurate, and it predicts 96% of true positive rates. This means that it is capable of discriminating almost all cases of malware in detection and prediction. Paper [20] focuses on the impact of feature selection and engineering in the classification of handwritten text by identifying and extracting those attributes of the training dataset that will contribute the most toward the classification task using classifiers like J48, naïve Bayes, and sequential minimal optimization (SMO). This results in improved accuracy of the classifiers as

compared to the work reported earlier. The performance evaluation of the classifiers used for OCR and pattern recognition is done. Initial classification performance of all the classifiers listed above was recorded on the raw dataset. Finally, the dataset was transformed after performing relevant feature selection and engineering on its attributes. The same classifiers were again trained on the transformed dataset and their accuracy was recorded. This paper uses the widely used MNIST dataset of handwritten digits for training the classifiers. In paper [21], the goal is to develop techniques to continuously and automatically detect a smartphone user's mobility activities, including walking, running, driving, and using a bus or train, in real-time or near real-time (<5 s). Their application is sensing mobility contexts using smartphone sensors. They demonstrated that, by combining data from GPS, accelerometers, and GIS with existing ML algorithms, one can build a highly performing classifier for detecting mobility contexts of smartphone users. Thus, the computational complexity of the classification algorithms suggested that many of these classifiers can feasibly be implemented in a smartphone.

58.2.1.4 KNN

Paper [22] explores opinion mining using supervised learning algorithms to find the polarity of the student feedback based on predefined features of teaching and learning. It is for predicting the polarity of the student comments based on extracted features like examination, teaching, etc. It uses K-nearest neighbor for improving the performance. KNN classifier employs an indexing mechanism for the training datasets. To classify a document, it calculates the similarity of the document with the training set index and uses the k-nearest neighbor by measuring the similarity by functions such as Euclidean distance. It classifies the positive and negative based on their polarity to analyze the features which need improvement. The basic KNN model is used here. Paper [23] presents a genetic algorithm optimized K-nearest neighbor algorithm (evolutionary KNN imputation) for missing data imputation. This paper addresses the effectiveness of using supervised learning algorithm for missing data imputation. This paper mainly focuses on local approach where the proposed evolutionary K-nearest neighbor imputation algorithm falls in. The evolutionary K-nearest neighbor imputation is the extension of K-NN. The importance of using machine learning algorithm in this paper is that most of the common imputation methods such as case deletion and mean imputation method are showing less effective results by not considering the correlation of data. The mean error from mean imputations is not effective, and the error rate is relatively higher than the supervised algorithms. Paper [24] analyzes the groundwater potential mapping to get better and more accurate groundwater potential mapping (GPMs). The performance of K-NN is excellent. The paper concluded that spring occurrence had a direct relationship with TWI, while a reverse relationship was observed between spring occurrence and two factors including slope length and distance from faults. The importance of groundwater influence factors changes in different places and feature selection needs to be investigated before the modeling process. It can be used in

different areas for groundwater potential mapping. The modified version of K-NN is used here. Paper [25] uses different seismicity indicators as an input for system to predict earthquakes, which is becoming increasingly popular. The different attribute for the input in supervised learning technique is created by using the new techniques. It relates to training and testing the dataset for calculating the b-value. This method is applied to the four Chilean zones. It can be extended to any location. When the dataset length selection is good, the prediction accuracy is also good. The attributes of some indicators result in better accurate prediction. The basic model of KNN is used here. Paper [26] proposes an approach of multiple-implementation testing to test supervised learning software, a major type of ML software. This paper derives a test input's proxy oracle from the majority-voted output running the test input of multiple implementations of the same algorithm. K-NN algorithm is used here. This approach is highly effective in detecting faults in real-world supervised learning software. The majority-voted oracle has low false positives and can detect more real faults than the benchmark-listed oracle. The modified version of KNN is used. Paper [27] focuses to detect and associate fake profiles on Twitter social network which are employed for defamatory activities to real profile within the same network by analyzing the content of the comments generated by both profiles. It presents a successful real-life use case in which the methodology is applied to detect and stop a cyberbullying situation in a real elementary school. The KNN algorithm helps to check the effectiveness thereof have been $k = 1-5$ to determine if taking account of more neighbors significantly improves the result. To optimize the result, filter the tweets text with stop words. It helps in identifying the real user or users behind a trolling account. The basic model of K-NN is used here. Paper [28] proposes the word embeddings which help to learn the semantic for words from a given sentence. The Word Mover's Distance (WMD) helps in finding the dissimilarity between the two text documents. The minimum distance of the embedded words of a single document need to travel for reaching the embedded word of other document. The WMD metric leads to unprecedented low k-NN document classification error rates. It discovers how quickly the lower bound distance helps to speed up the nearest neighbor. It is done by the collecting and pruning the neighbor. The advantage of using the WMD is to reduce the error rates in the datasets. It obtains the intelligence about the text documents in English language. It reinforces interpretability. The penalty for the terms is added if any two words take place in different selection of the same structured documents. The basic model of K-NN is used here. Paper [29] focuses on classifying the leukemia patients' dataset. In case of abnormal leukemia, it leads to bleeding, anemia, and impaired ability to survive against the infection. Biomarkers give the doctors the necessary information regarding the class of therapy required for the patient. It also provides the information about patients having severe condition of growing the disease. This paper proposes a research work on selecting biomarkers from gene-based microarray dataset and then predicting the subtype of cancer. K-NN is of higher accuracy. The basic model of KNN is used here.

58.2.1.5 Regression

Paper [30] proposes how a universal sentence is trained by using the supervised data of natural language inference datasets that exceed unsupervised methods like skip thought vectors. It is similar to the computer vision which uses the Image Net for acquiring the features that can be transferred to other tasks. It shows the suitable natural language inference for learning other NLP tasks. In a large corpus, the sentence encoder model is trained and transferred to other tasks. They solve the following two questions: Which is the preferable network architecture, and how is the task trained? The best results like accuracy are obtained while training the natural language tasks. In paper [31], the goal is to use two supervised methods like regression and classification. While learning from the crowds, the heterogeneity and biases among different annotators are discovered. The stochastic variational inference algorithm is used to scale a larger dataset. The proposed model joins the words in a document from the mixture of topics. The topics for the document are analyzed. The labels for the different annotators are a noisy version of the latent ground. Regression concentrate is used to handle the multiple annotators with different biases and reliabilities when the target variables are continuous variables. The altered version of regression is used. Paper [32] works toward helping the readers in transforming the motivation and formulating the problems and methods of the strong machine learning techniques for future networks in order to move into the unexplored applications. The Future 5G mobile will access in self-manner. It accesses the spectral band with the help of spectral efficiency learning and inference to control the transmission power and adjusting transmission protocols with the help of quality of service. The study uses regression algorithm. The goal of regression analysis is to predict the value of one or more continuous-valued estimation targets, given the value of a D -dimensional vector x of input variables. The estimation target is a function of the independent variables. Regression models can be used for estimating or predicting radio parameters that are associated with specific users. The basic model of regression is used. Paper [33] proposes a cross-modality consistent regression (CCR) model, which is able to utilize both the state-of-the-art visual and textual sentiment analysis techniques. They first fine-tune a convolutional neural network (CNN) for image sentiment analysis and train a paragraph vector model for textual sentiment analysis. On top of them, they train the multi-modality regression model. The results show that the proposed model can achieve better performance than the state-of-the-art textual and visual sentiment analysis algorithms alone. They would like to learn people's overall sentiment over the same object from different modalities of the object provided by the user. In particular, they focus on inferring people's sentiment according to the available images and the short and informal text. They used a modified version of regression model. In paper [34], trouble with learning the mixture of regression models for the individuals is analyzed. The mixing component is unknown, and it has to be

determined in the unsupervised method using data-driven tools. The novel penalized method is established to pick the mixing components, to know the mixture proportions, and also to know the unknown attributes in the models. It deals with the continuous and discrete response by initial two moment conditions.

58.2.2 *Unsupervised Learning*

58.2.2.1 **K-means**

Paper [35] proposes a simple and effective approach to automatically learn a multilayer image feature for satellite image scene classification. The feature extraction is composed of two layers which are uniformly learned by the K-means algorithm. It can also extract the complex features. In the multilayer approach, the feature from the higher layer has better classification performance than the lower layers. It achieves better performance compared with any single layer. It can achieve the VHR Image scene classification performance compared with the state-of-the-art approaches. K-means have a better performance when compared to SC and S-RBM. The benefit of this method is that it automatically mines the structure information from simple to complex, which meets the hierarchical perceptual mechanism of the human visual cortex. The basic model of K-means is used here. Paper [36] explores the application of the spherical k-means algorithm for feature learning from audio signals. They evaluate the approach on largest public datasets of urban sound sources available for research, and compare it to baseline approach by configuring it to capture the temporal dynamics of urban sources. It mainly focuses on the classification of auditory scene type (e.g., street, park), as opposed to the identification of specific sound sources in scenes such as car horn and bird tweet. It does not outperform the baseline approach. The modified version of K-means is used here. Paper [37] deals with the emotion recognition. It is to identify the emotions and assign one of the seven emotions to the short videos which are obtained from the Hollywood movies. The video clips represent the emotions based on the realistic condition by using the attributes like pose and illumination and explore by combining the feature from multiple labels. Deep learning techniques are used in this paper for focusing on modalities. CNN helps to capture the visual information by detecting the faces. The visual features around the mouth are extracted by using the K-means bag of mouth model. Better accuracy is achieved. Benefits are that it can be used in large-scale mining of imagery from Google image search to train the deep neural network. The basic model of K-NN algorithm is used. Paper [38] focuses on the unsupervised deep learning techniques to know the details of the patient from the electronic health data record (EHR) for clinical predictive model. The problems of supervised methods are overcome by using the unsupervised methods. The unsupervised methods help to identify the patterns automatically and make an overall representation which is easier to extract the needed information automatically while building the classifiers. To preprocess the level of the patient, the unsupervised

methods are used. It clusters the EHR data for understanding better. It uses deep patient for representing the general features of the patient which are obtained from the deep learning technique. K-means is the algorithm used here. K-means help to cluster the unlabeled data into k-clusters so that each point belongs to the cluster which has the closest means. The centroids of the cluster help to produce the features in future. The main focus of this paper is to assess the methods on EHR data warehouse for consolidating the results. A large number of patient results are also evaluated. The basic model of K-means is used here. Paper [32] works toward helping the readers in transforming the motivation, formulating the problems and methods of the strong machine learning techniques for future networks for the unexplored applications. The algorithm is done iteratively, and the object is allocated to the particular cluster where its centroid is close to the object which is based on the Euclidean distance. Clustering is the main problem for 5G network mainly while concentrating on the heterogeneous scenario. It is to reduce the traffic in the wireless system by utilizing the high-capacity optical infrastructure. It uses initial gateway access point which helps to pluck randomly from the set of MAP, and it intelligently determines the meritorious initialization criterion. Each MAP is allocated to the nearest GAP. If many GAPs are in the environment, then the GAP which has the virtual channel will be chosen. By using the classic K-mean clustering, the MAPs are split into k groups with the close GAPs. The basic model of k-means is used.

58.2.2.2 Agglomerative

Paper [39] proposes the unsupervised learning method to discover the groups of molecular system based on the similarity in the structure by using the Wards minimal variance objective function. The minimal variance clustering is applied to the set of tripeptides by using the informatics theory to find how point mutations affect the protein dynamics. Here it focuses on the ability of the unsupervised MVCA algorithm to identify groups of molecular systems. Hierarchical agglomerative clustering is a set of pairwise distances between data points and iteratively merges the two closest clusters or singletons. Here the agglomerative and similarity distance function is used to quantify the similarity between multiple models for related dynamic systems. This analysis is intended to address a knowledge gap specific to supervised ML for chemoinformatic analyses designed to predict the properties of novel molecules. The agglomerative algorithm is used here. Paper [40] focuses on the membrane computing. It uses the fuzzy membrane, computing techniques, and clustering algorithm. Fuzzy clustering algorithm achieves good fuzzy partitioning for a dataset. Each element as a separate cluster and merge them into two larger cluster, or divisive which begin with the whole set and successively divide it into small cluster. The membrane computing aims to abstract computing models from the structure and functioning of living cells. The main aim of this paper is to solve fuzzy clustering problems. It uses an optimal cluster for a dataset and determines good fuzzy portioning. The basic model of HIAL is used here. Paper [41] focuses on the possibility of remote monitoring and screening of Parkinson's disease and

age-related voice impairment for the general public using self-recorded data on readily available or emerging techniques such as smart phones or IoT devices. It uses a sustained Vowel/a/recordings using iPhone. The purpose of finding the number of clusters is not just trying to find the number of optimal K to best reflect the level of impairments, but also to see how well the subjects can be clustered. The AGNES method joins one closest pair of clusters in each iteration by comparing the distance. The AGNES algorithm performs the best. This study has achieved clustering the voice with a data reduction ratio of 518. The basic model of AGNES is used. Paper [42] explains how attacker track mobile user is explained. Because users carry their phone everywhere where some sensitive or private information is stored about individual user. As user interacts with mobile apps, lot of network traffic is generated by sending or receiving request. Here they try to act as an eavesdropper which attacks on network traffic of the device from controlling Wi-Fi access point. Paper [43] explores the use of unsupervised clustering based on passive DNS records and other inherent network information to identify domains that may be part of campaigns but resistant to detection by domain name or time-of-registration analysis. They found that using this method can achieve up to $2.1 \times$ expansion from a seed of known campaign domains with $<4\%$ false positive. It is useful for identifying malicious domains. The agglomerative algorithm shows the best performance. In agglomerative clustering, on average, 94% of the campaign domains are present. Among the 15 clusters, 8 clusters are 100% campaigns domains. The agglomerative algorithm is used here. Paper [44] analyzes the problems like redesigning the algorithms in order to use the distributed computation resources effectively and to come up with a solution for complex hierarchical clustering algorithm CLUBS+ effectively. The high-quality clusters are grouped within their centroids. The accuracy and scalability are achieved using map-reducing algorithms. This is for clustering the big data in a range but still informative version. High accuracy is attained here. Here it uses parallel version of CLUBS+. CLUBS+ helps to combine the benefits of agglomerative and divisive approach. CLUBS+ defines the binary space partition for the domain, leads to the group of cluster that is refined. Agglomerative is a final refinement phase. Outliers are discovered and other points are allocated to the nearest cluster in the refining phase. The clustering quality is improved. It is implemented by using the priority queue. It decreases the number of pairs. It often produces irregular shapes. It improves the quality and identifies the majority of the outlier points. The agglomerative algorithm is used. Paper [45] aims to reduce the human workload, and they are motivated to automate the annotation process. Here the authors consider the online handwriting problems and the Arabic script. They discussed the implementation of word recognition system. Agglomerative clustering is used to produce a Codebook with one stroke per class. In offline handwriting recognition, handwritten documents are described by a scanned image. In online handwriting recognition, handwritten documents are described by a sequence of sample points which represents the writing online data. Their work aims to automate segmentation, labeling, and recognition process. The agglomerative clustering techniques merge at each learning stage, with two closest prototypes to generate a dendrogram. It allows choosing the final number of prototypes at the end of learning.

They select the samples which minimize the sum of distances to the other samples of the same cluster. Here it is combined between two strokes, and relation is composed of five binary values: right, left, above, below, and intersection. The problem in Arabic script is that it is written from right to left. Some characters are represented in shapes and differ from each other by the existence of dots. To avoid segmentation problems, online handwriting is segmented into strokes as written by the writers. The strokes are grouped into a codebook using hierarchical clustering. The agglomerative clustering is used here. In paper [46], opinion mining is used to analyze and cluster the user-generated data like reviews, blogs, comments, and articles. The main objective of opinion mining is to cluster the tweets into positive and negative clusters. In this paper, the authors are able to collect information from social networking sites like Twitter. The meaningful tweets are clustered into two different clusters, positive and negative, using unsupervised ML technique such as spectral clustering. Here the goal is to determine whether the opinion expressed in a twitter is “thumbs-up” or “thumbs-down.” The hierarchical clustering algorithm groups the data objects to form a tree-shaped structure. It is split into agglomerative and divisive clustering. In the agglomerative approach, each observation starts in its own cluster, and pairs of clusters are merged as one moves up the hierarchy. It achieves the best sentiment accuracy. It can solve the problem of domain dependency and reduce the need of annotated training data. It overcomes the problem of clustering multiple files with unlabeled data and performs sentimental classification. In paper [47], the unsupervised feature learning is applied to the mixed-type data to achieve a sparse representation, which makes it easier for clustering algorithms to separate the data. It uses mixed-type data using fuzzy adaptive resonance theory (ART). The ART obtains a better clustering result by removing the differences in treating categorical and numeric features. Benefit of this is demonstrated with several world datasets. It is also demonstrated on noisy, mixed-type petroleum industry data. The goal is to build a framework that automatically handles the differences in numeric and categorical features in a dataset and group them into a similar cluster. It is to reduce the distinction among three numerical and categorical features. It helps to increase the performance. A good cluster has small standard deviation for each cluster as well as averages that vary greatly from one another. The basic clustering model is used here.

58.2.2.3 Divisive

Paper [48] targets to identify the history of arts by grouping the digital paintings based on their features. It automatically learns without any prior knowledge. The performance of clustering is determined by recovering the original groups, F-score, and reliabilities. The spectral clustering algorithm groups the paintings into distinct style groups. The goal of it is finding the characteristics for each paintings by using the N-dimensional feature vectors, and plotting the distance between the endpoints of the vectors. The endpoints are color coded with the low-cost labels. The plots which are scattered are visually inspected for separating the paints. The separation of color provides the semantic information on relation between the styles. The

clustering had an F-score of 0.212. It enables to extract the character of art without any prior information about the nature of the features or the stylistic designation of the paintings. Paper [43] checks a healthcare database which has the patient report for the knee replacement surgery. The patterns of the pain are detected by using the hierarchical clustering. The result indicates the presence of subgroups of the patients based on the pain characteristics. They explain the problem occurring by using the unlabeled medical data. It recognizes the pain-related patterns in the patient report. The national-level dataset is used. Paper [49] analyzes the problems like redesigning the algorithms in order to use the distributed computation resources effectively and to come up with a solution for complex hierarchical clustering algorithm CLUBS+ effectively. The high-quality clusters are grouped within their centroids. The accuracy and scalability are achieved on map-reducing algorithms. This is for clustering the big data in a range but still informative version. High accuracy is attained. Here it uses parallel version of CLUBS+. CLUBS+ helps to combine the benefits of agglomerative and divisive approaches. CLUBS+ defines the binary space partition. It leads to the set of clusters that are refined. The top-down partition is used to separate the hyper rectangle block where the points are close to each other which is equal to minimize the cluster sums of square. It helps to minimize. It splits and replaces the process. It finds the computation of the best split and evaluating the effectiveness of the split. In paper [50], opinion mining is used to analyze and cluster the user-generated data like reviews, blogs, comments, and articles. The main objective of opinion mining is to cluster the tweets into positive and negative clusters. In this paper, the authors are able to collect the information from social networking sites like Twitter. The meaningful tweets are clustered into two different clusters, positive and negative, using unsupervised ML technique such as spectral clustering. Here the goal is to determine whether the opinion expressed in a twitter is “thumbs-up” or “thumbs-down.” The hierarchical clustering algorithm groups the data objects to form a tree-shaped structure. In the divisive approach, all observation starts in one cluster, and splits are performed recursively as one moves down the hierarchy. It achieves the best sentiment accuracy. It can solve the problem of domain dependency and reduce the need of annotated training data. It overcomes the problem of clustering multiple files with unlabeled data and performs sentimental classification.

58.2.2.4 Neural Network

Paper [51] used unlabeled videos from the web to learn visual representations. The idea here is that visual tracking provides the supervision. The authors used auto-encoder for learning the representation based on the ability to reconstruct the input images. It is used to train the unlabeled visual data on the web to train the CNN. Their idea is that two patches connected by a track should have a similar visual representation in deep feature space since they probably belong to the same object. They track millions of patches and learn an embedding using CNN that keeps patches from the same track closer in the embedding space as compared to any

random third patch. The modified version of neural network is used here. In paper [52], sentiment analysis of short texts, such as sentences and Twitter messages, is challenging because of the limited information. To solve it effectively, the authors used some strategies like combining the text content with the prior knowledge. Here they used a deep convolution neural network that exploits from character- to sentence-level information to perform sentiment analysis of short texts. They applied this method for Stanford Sentimental tree bank which has movie reviews and Stanford Twitter Sentimental corpus which contains twitter messages. It achieves 86.4% accuracy. The neural network algorithm is used here. Paper [53] describes the deep learning analysis for sentiment analysis of tweets. Here the authors initialize the parameter weights of CNN, which helps to train an accurate model while avoiding the need to inject any additional features. This method is ranked in the first two positions in both the phrase-level subtask A and on the message-level subtask B. They applied the deep learning model on Semeval2015 Twitter sentiment analysis. The aim of the convolutional layer is to extract patterns. The network initialization process is used to refine the weights of the network passed from unsupervised neural language model. It combines two aspects of IR: unsupervised learning of text representations and learn on weakly supervised data. The basic model of CNN is used here. Paper [54] focuses on supervised technique and uses deep neural network trained on large datasets. Many methods are used such as structured prediction and also multiple pixels are classified by using a network simultaneously. It outperforms the previous algorithm on the area of ROC curve measure. The accuracy is greater than 0.97. The method is also resistant to the phenomenon of central vessel reflex, sensitive in detection of fine vessels (sensitivity >0.87) and fares well on pathological cases. They propose a DL-based method for the problem of detecting blood vessels in fundus imagery, a medical imaging task that has significant diagnostic relevance and was subject to many studies in the past. The proposed approach outperforms previous methods on two major performance indicators, that is, accuracy of classification and area under the ROC curve. Here the authors consider fundus images, that is, pictures of the back of the eye taken in the visible band. A convolutional neural network (CNN) is a composite of multiple elementary processing units, each featuring several weighted inputs and one output, performing convolution of input signals with weights and transforming the outcome with some form of nonlinearity. These results confirm that segmentation of vessels close to the borders of FOVs is more difficult than for the more central pixels. Paper [57] is split into two neural architectures: bidirectional LSTMs and conditional random fields. It constructs and labels the segment by using a transition-based approach and shifts reduce parser. The models depend on two sources of data about the words. The character-based word representations are learned from the supervised corpus and unsupervised word representations are learned from unannotated corpora. The performance in NER in four languages without resorting to any language-specific knowledge or resources such as gazetteers. No language-specific resources or features beyond a small amount of supervised training data and unlabeled corpora are presented by neural architectures for NER. Comparison between the models is made and the scores are reported of the models and with and without the use of

external labeled data such as gazetteers and knowledge bases. The authors obtained an F1 of 91.2 by jointly modeling the NER and entity-linking task. A key aspect of their models is that they model output label dependencies, either via a simple CRF architecture, or by using a transition-based algorithm to explicitly construct and label chunks of the input. Word representations are also crucially important for success; they use both pretrained word representations and “character-based” representations that capture morphological and orthographic information.

58.3 Conclusion

This survey paper gives a clear idea about both supervised and unsupervised learning algorithms. Classification- and regression-based problems can be solved using supervised learning algorithms like support vector machine, KNN, naïve Bayes, and linear regression. Clustering and association mining-based problems can be solved using unsupervised learning algorithms like K-means, divisive, and agglomerative.

References

1. Chikersal P, Poria S, Cambria E. Sentimental analysis of Tweets by combining a rule-based classifier with supervised learning. *International workshop on Semantic Evaluation*, pp 647–651
2. Zemmal N, Azizi N, Dey N, Sellami M (2016) Adaptive semi supervised SVM supervised learning with feature cooperation for breast cancer classification. *Journal* 6:53–62
3. Riaz F, Hassan A, Rehman S, Niazi IK, Dremstrup K (2016) EMD-based temporal and spectral features for the classification of EEG signals using supervised learning. *IEEE Trans* 24(1):28–35
4. Neelakantan A, Collins M (2015) Learning dictionaries for named entity recognition using minimal supervision, 24 April 2015
5. Mansson KNT, Frick A, Boraxbekk C-J, Marquand AF, Williams SCR, Calbring P, Andersson G, Furmark T (2015) Predicting long-term outcome of Internet-delivered cognitive behaviour therapy for social anxiety disorder using fMRI and SVM
6. Devi Arockia Vanitha C, Devaraj D, Venkatesulu M (2015) Gene expression data classification using SVM and mutual information-based gene selection, 13–21, 2015
7. Ozgur Kisi, Kulwinder Singh Parmar (2015) Application of least square SVM and multivariate adaptive regression spline models in long term prediction of river water pollution, 3 August 2015
8. Chen L, Bi L, Si H, Zhang J, Ren Y (2015) Research on prediction method for pivotal indicator of hospital medical quality using decision tree. In: *IEEE international conference*, pp 247–250, 30 November 2015
9. Joshi A, Dangra J, Rawat M (2016) Decision tree based classification technique for accurate heart disease classification and prediction. *Int J* 3(11):1–4
10. Pradhan VM, Vala J, Balani P (2016) A survey on sentimental analysis algorithms for opinion mining. *Int J* 133:7–11

11. Gai K, Qiu M, Elnagdy SA (2016) Security-aware information classifications using supervised learning for cloud-based cyber risk management in financial big data. In: International conference
12. Yuan Z, Wang C (2016) Online analysis and computing science (ICOACS). IEEE international conference:53–56
13. Sarmah J, Sharma SKR (2016) Decision tree based supervised word sense disambiguation for Assamese. *Int J Comput Appl* 141(1):42–48
14. Jyh-Jian Sheu, Yin-Kai Chen, Ko-Tsung Chu, Jin-Hsin Tang, Wei-Pang Yang. An intelligent three phase spam filtering method based on decision tree data mining
15. Hui LI, Deliang Y, Lingling Y, Yao L, Xia LIN (2016) Supervised massive data analysis for telecommunication customer churn prediction. IEEE international conference
16. Jung YG, Kim KT, Lee B, Youn HY (2016) Enhanced naive bayes classifier for real-time sentiment analysis with SparkR. IEEE International Conference
17. Jiang L, Li C, Wang S, Zhang L (2016) Deep feature weighting for naive Bayes and its application to text classification. *Eng Appl Artif Intell* 52:26–39
18. Nagarajan S, Chandrasekaran RM (2015) Design and implementation of expert clinical system for diagnosing diabetes using data mining techniques. *Indian J Sci Technol* 8(8):771–776
19. Tchakounté F, Hayata F (2017) Supervised learning based detection of malware on Android
20. Kaushik A, Gupta H, Latwal DS (2016) Impact of feature selection and engineering in the classification of handwritten text. IEEE international conference
21. Guinness RE (2015) Beyond where to how: a machine learning approach for sensing mobility contexts using smartphone sensors
22. Dhanalakshmi V, Bino D, Saravanan AM (2016) Opinion mining from student feedback data using supervised learning algorithms. International conference
23. de Silva H, Shehan Perera A (2017) Evolutionary K-nearest neighbour imputation algorithm for gene expression data. In: *Int J*
24. Naghibi SA, Moradi Dashtpajardi M (2017) Evaluation of Supervised learning methods for ground water spring potential mapping in khalkhal region using GIS-based features, pp 169–189
25. Asencio-Cortes, Martinez-Alvarez F, Morales-Esteban A, Reyes J (2016) A Sensitivity study of seismicity indicators in supervised learning to improve earthquake prediction
26. Siwakorn Srisakaokul, Zhengkai Wu, Angello Astroga, Oreoluwa Alebiosu, Tao Xie, “Multiple-Implementation Testing of Supervised Learning Software”, 2018
27. Galan-Garcia P, de la Puerta JG, Gomez CL, Santos I, Bringas PG (2016) Supervised ML for the detection of troll profiles in twitter social network. 24(1):42–53
28. Kusner MJ, Sun Y, Kolkin NI, Weinberg KQ (2015) From word embeddings to document distances. International conference, vol:37
29. Beegum S, Chakraborty D (2016) Identifying cancer biomarkers from leukemia data using feature selection and supervised learning. IEEE international conference
30. Conneau A, Kiela D, Schwenk H, Barrault L, Bordes A (2018) Supervised learning of universal sentence representations from natural language inference data, July 2018
31. Rodrigues F, Lourenco M, Ribeiro B, Pereira FC (2017) Learning supervised topic models for classification and regression from crowds. *IEEE Trans Pattern Anal Mach Intell* 39 (12):2409–2422
32. Jiang C, Zhang H, Ren Y, Han Z, Chen K-C, Hanzo L (2017) Machine learning paradigms for next-generation wireless networks. *IEEE Wirel Commun* 24(2):98–105
33. Quanzeng You, Jiebo Luo, Hailin Jin, Jianchao Yang. Cross-modality consistent regression for joint visual-textual sentiment analysis of social multimedia
34. PeirongXu HP, huang T (2018) Unsupervised learning of mixture regression models for longitudinal data. *Comput Stat Data Anal* 125:44–56
35. Li Y, Tao C, Tan Y, Shang K, Tian J (2016) Unsupervised multilayer feature learning for satellite image scene classification. *IEEE Geosci Remote Sens Lett* 13(2):157–161
36. Salamon J, Bello JP (2015) Unsupervised feature learning for urban sound classification

37. Kahou SE, Bouthillier X, Lamblin P, Gulcehre C, Michalski V, Konda K, Jean S, Froumenty P, Dauphin Y, Mirza M, Warde-Farley D, Courville A, Vincent P, bangillo Y (2015) EmoNets: multimodal deep learning approaches for emotion recognition on video. *J Multimod User Interface* 10(2):99–111
38. Miotto R, Li L, Kidd BA, Dudle JT (2016) Deep patient: an unsupervised representation to predict the future of patients from the electronic health records. *Sci Rep* 6:1
39. Husic BE, Pande VS (2017) Unsupervised learning of dynamical and molecular similarity using variance minimization. In: Conference
40. Rueda A, Krishnan S. Clustering Parkinson's and age-related voice impairment signal features for unsupervised learning
41. Kulkarni RA (2018) Scrutinizing action performed by user on mobile app through network using ML techniques. *IEEE international conference*
42. Weber M, Wang J, Zhou Y (2018) Unsupervised clustering for identification of malicious domain campaigns
43. Mazzeo GM, Zaniolo C (2016) The parallelization of a complex hierarchical clustering algorithm: faster unsupervised learning on larger datasets
44. Gargouri M, Touj SM, Ben Amara NE (2015) Towards unsupervised learning and graphical representation for on-line handwriting script. *IEEE international conference*
45. Unnisa M, Ameen A, Raziuddin S (2016) Opinion Mining on Twitter data using Unsupervised Learning Technique. Vol:148
46. Lam D, Wei M, Wunsch D (2015) Clustering data of mixed categorical and numeric type with unsupervised feature learning
47. Gultepe E, Edward T (2018) Predicting and grouping digitized paintings by style using unsupervised feature learning. *J Cult Heritage* 31:13–23
48. Khalid S, Judge A, Pinedo-Villanueva R (2018) An unsupervised learning model for pattern recognition in routinely collected healthcare data. In: *International conference*, vol 5, pp 266–273
49. Unnisa M, Ameen A, Raziuddin S (2016) Opinion mining on Twitter data using unsupervised learning technique. *Int J Comput Appl* 148:12–19
50. Wang X, Gupta A (2015) Unsupervised learning of visual representation using videos. *IEEE*
51. dos Santos CN, Gatti M. Deep CNN for sentimental analysis of short texts
52. Severyn A, Moschitti A (2016) Twitter sentimental analysis with deep CNN
53. Liskowski P, Krawiec K (2016) Segmenting retinal blood vessels with deep neural networks. *IEEE Trans Med Imaging* 35(11):2369–2380
54. Lample G, Ballesteros M, Subramanian S, Kawakami K, Dyer C (2016) Neural architectures for named entity recognition, 17April 2016