International Journal of Research Publication and Reviews

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

Recognizing Learning Disabilities Using Learning Vector Quantization

Siva Nagaiah Velpula

Vignan University,India

ABSTRACT:

Children and adults with learning disabilities see, hear, and understand things differently. This may result a trouble in gain an understanding of new information and skills, and putting them to use. The most frequent types of learning disabilities involve problems with reading, writing, math, reasoning, listening, and speaking. In today's competitive world, these can be a hurdle for the victims in developing their career. A learning disability can't be cured or fixed; it is a lifelong issue. By the adequate support and intervention, people, especially children with learning disabilities can succeed in school and go on to successful, often distinguished careers later in life. It is often presumed that learning disabilities may be caused by hereditary, teratogenic factors (for instance, alcohol or cocaine use during pregnancy), medical factors such as, premature birth, diabetes, meningitis of mother or offspring), and/or environmental factors (malnutrition, poor prenatal healthcare). The aim of this project is to identify the learning disability within a person. A series of tests can be conducted based on some parameters to identify the abnormality in the learning process. Using a soft computing approach called Learning Vector Quantization, we implemented a model that can classify whether a person has learning disability or not. Further a rule-based approach is used to identify the class of learning disability of a person. This model predicts if a person is suffering with learning disorder or not.

1.INTRODUCTION

1.1 Background and Motivation:-

Learning disability or a learning disorder is a condition in the brain that causes difficulty and struggle to comprehend or process information and can be result of several different factors. However, "difficulty in learning in a typical manner", does not go past the ability to learn in a different manner. Hence , some people can be more precisely described as having a "learning difference", thus avoiding any misunderstanding of being disabled with a lack of ability to learn. In the UK, the phrase "learning disability" generally refers to an intellectual disability, while difficulties such as learning(Dyslexia), writing(Dysgraphia), and mathematics(Dyscalculia) are usually referred to as "learning difficulties". Even though learning disability, learning disorder and learning difficulty are often indistinguishable, they differ in many ways. Disorder may lead to significant learning problems in an academic area. However,these problems, may not be sufficient to cater an official diagnosis. Learning disabilityis an official clinical diagnosis, wherethe individual needs to meet certain criteria, as fixed by a professional. The variation is in level, prevalence, and intensity of described symptoms and problems, and henceboth of them should not be confused. The term "learning disorder" describes a group of disorders catered by insufficient development of certain academic, language, and speech skills. Categories of learning disorders include reading, mathematics and writing.

1.2 Problem Statement: -

We implemented a Soft computing technique to detect learning disability. In computer science, soft computing means the use of approximate solutions to computationally difficult tasks like the result of NP-complete problems, for which there is no conventional algorithm that can compute ancalculate solution in polynomial time. So, this project proposes a model for diagnosis and classification of LD. Introduces and explains the proposed model which works by Learning Vector Quantization which come under artificial neural network for diagnosis and gives the system results based on accuracy.

2 REQUIREMENT ELIITATION AND ANALYSIS

2.1 Existing System:

The existing system used in this project is decision stump algorithm. A decision stump is a machine learning model consisting of a one-level decision tree. That means, it is a decision tree that consists of one internal node /the root which is right away connected to the terminal nodes/leaves. It makes a prediction through the value of just a single input feature. At times they are also called 1-rules.s

2.2 Proposed System:

This project runs using LVQ algorithm. LVQ is a special case of artificial neural network (ANN) which applies a winner-take all learning based approach. Thus, vector quantization can also be seen as a mapping from a dimensional Euclidean space into the finite set also mentioned to as the codebook

Advantages:

• LVQ Neural network has many advantages, for in- stance, a good network performance, fast training speed, a lesser number of neurons, high recognition rate, simple network structure
• It does not require the in- put vectors are normalized, orthogonal, only need to direct calculation between the input vector and the competitive layer distance, which can complete very complex classification through the interaction of internal unit.
• It is also very easy to design conditions of various com- plex dispersed domains converges to the conclusion. LVQ network, so to speak, therefore, has the very good model classification features. Designers do not need to construct complex in the process, even nonlinear processing function to construct.
• In addition, the LVQ network show stronger than Network in fault tolerance and robustness,not easy to lead to the collapse of the system.
• Learning Vector Quantization is that creates prototypes that are easy to interpret.

3.IMPLEMENTATION DETAILS

In implementation, the theoretical design is converted into a working system. The most importantphase in getting on a new accomplishing system in leading confidence on the new system for the users that will turn on efficiently and effectively. The system can be executed only after sufficient testing is done and if prone to actin relation to the specification. Itdo requireprecise planning, thorough investigation of the contemporary system and its limitation on implementation, design of methods to achieve the changeover and an analyzation of change over methods to cater the change and an evaluation of this change apart from planning.

Methodology Used: Learning Vector Quantization is used for Predict: Modules:
• Create Dataset
• Preprocessing
• Classification
• Accuracy
• prediction

sModules Description:
Create dataset: A hypothetical dataset is designed with respect with the parameters that mainly important to diagnosis if a person has learning disability or not

INPUT PARAMETERS	MARKS	Category of LD
essays	10	Dysgraphia
readings	10	Dyslexia
Comprehension	10	Dyslexia,Dysgraphia
Spelling	10	Dysgraphia
Perception	10	Dyslexia
Solve	10	Dysgraphia
Word Problem	10	Dyscalculia,Dyslexia
Mental Sums	10	Dyscalculia
Time	10	Dyscalculia
Input parameters	10	Dyscalculia
Money	10	Dyscalculia

TABLE: parameter and marks of the dataset

Table shows the initial 11 inputs corresponding to curriculum-based test. Column 1 represents the name of the parameters, and Column 2 represents the total marks allocated to a particular section of the parameters Currently the Dataset consist of 107 cases, out of which NON-LD are 49 and LD are 58.The system is trained using 70% of the dataset item and the remaining is used for to test the system

Pre-processing: • For achieving better results from the applied model in soft computing projects the format of the data has to be in a proper manner. Some mentioned Machine Learning model needs information in a mentioned format, for instance, algorithm may not hold upvoid values, hence, to execute algorithm,void values have to be controlled from the initial raw dataset.

Classification: Learning Vector Quantization algorithm is used for classification in this project. Learning Vector Quantization (LVQ) , different from Vector quantization (VQ) and Kohonen Self-Organizing Maps (KSOM).LVQ is a competent network that uses supervised learning approach . We can define it as a process of classifying the patterns where each output unit represents a class. Since it uses supervised learning, the network will be given a set of training patterns with familiarcategorization along with an original distribution of the output class. After finishing the training process, LVQ will categorize an inward vector by allocating it to the same class as that of the output unit.

Classification Model Algorithm:

1. Initialize weights vectors (for the system it is $w1\ tow11$).
2. Initialize learning rate (α), (for the system it is 0.8).
3. For each training input vector x (for the system it is x1 to x11), do steps 4 to 5.
4. Compute J using squared Euclidean distance: D(j) = $\sum(wij - xj) * D(j) = \sum(wij - xj)$ Find j when $D(j)$ is minimum. This leads to the classification of the entire dataset into two classes, one pertaining to LD and the other pertaining to the normal class. Cj indicates the class of LD. Cj = 1 indicates that the child is suffering from LD and Cj = 0 indicates the child is normal.
5. Update Wj as follows: If t = Cj, then $Wj\ (new) = Wj\ (old) + \alpha[x - Wj\ (old)\ ]$ If t ≠ Cj, then $Wj\ (new) = Wj\ (old) - \alpha[x - Wj(old)\ ]$
6. Reduce α by a small amount 0.01.
7. System is tested for every new value of α till the result shows redundancy. This is the stopping condition.

Example:

Problem: Construct and test LVQ with four vectors assigned to two classes.

Assume 41 α =0.1. Perform instruction up to α =0.5

Vector 0010 0011 1100 1001

Class 1 2 1 2

Solution:

The first output unit represents class 1; the second output unit represents class 2;

C1 =1 & C2= 2

The first two vectors out of the four vectors are being utilized to initiate the 2 referral vectors. The vectors (1 1 0 0) and (1 0 0 1) are utilizedas training vectors.

EPOCH-1:

Step 1: Format weights as $w1$= (1 0 1 0) and w2 = (0 0 1 1)

Initiate the learning rate as α =0.1

Step2: Start the training process

Step3:For input vector x=(1 1 0 0) along with T=1 do step 4-5

Step4: Calculate J D(j) = $(wij - xj)$2 D(1) = (1-1)2+ (0-1)2+(1-0)2+(0-0)2 = 2 D(2) = (0-1)2 + (0-1)2+(1-0)2+(1-0)2 = 4 D(j) is min J = 1 =>Cj =1

Step5: Since T = 1 &Cj = 1, then T = Cj$W1(new) = W1(old) + \alpha[x - W1(old)\ ]$ = (1 0 1 0) + 0.1(1 1 0 0) – (1 0 1 0) $W1(new)$ = (1 0.1 0.9 0)

Step3: For, input vector x= (1001) with T=2

do step 4-5

Step4 Calculate J D(j) = $\sum(wij - xj)$ 2 42 D(1) = (1-1)2 + (0.1-0)2+(0.9-0)2+(0-1)2 = 1.82 D(2) = (0-1)2 + (0-0)2+(1-0)2+(1-1)2 = 2 D(j) is min J= 1 =>Cj =1

Step5 Since T=2 and Cj=1,T≠Cj the weight updated is, $W1\ (new) = W1(old) - \alpha[x - W1(old)\ ]$ = (10.10.90) - 0.1(1001) – (10.1090) $W1(new)$ = (10.11 0.99 -0.1)

Step6 One approach of training is done. Deduce the learning rate as α(t+1) = 0.5 α (t) α (t) = 0.5 x 0.1 α (t) = 0.05

Step7: Test the stop condition. If the required learning rate has not yet obtained. Perform second epoch.

EPOCH-2:

Step1: Initialize weights $w1$ = (1 0.11 0.99 − 0.1) $andw2$ = (0 0 1 1). Initialize the learning rate as α =0.05

Step2: start the training process

Step3: As of input vector x = (1100) with T=1 do step 4-5

Step4: Calculate J D (j) = $(\sum wij - xj)$2 D (1) = (1-1)2 + (0.11-1)2 + (0.99-0)2 + (-0.1-0)2 = 1.7822 D (2) = (0-1)2 + (0-1)2 + (1-0 2+ (1-0)2 = 4 D (j) is min J=1 =>Cj =1

Step5: Since T = 1 and Cj = 1, then T ≠ Cj$W1(new) = W1(old) + \alpha\ [x - W1(old)\ ]$ = (1 0.11 0.9 -0.1) + 0.05(1 1 0 0) – (1 0.11 0.99 -0.1) W1 (new) = (1 0.1 5 0.94 -0.095) **Step3**: For input vector x = (1 0 0 1) with T=2 do step 4-5

Step4: Calculate J 43 D(j) = $\sum(wij - xj)$2 D(1) = (1-1)2 + (0.15-0)2 + (0.94-0)2+ (-0.095-1)2 = 1.725 D(2) = (0-1)2 + (0-0)2+(1-0)2+(1-1)2 = 2 D(j) is minimum J = 1 =>Cj = 1

Step5: Since T= 2 and Cj=1, T ≠ Cj the weight updated is, $Wj(new) = Wj(old) - \alpha[x - Wj(old)\ ]$ = (1 0.1 5 0.94 -0.095) - 0.05[(1 0 0 1) – (1 0.15 0.94 -0.095)] $Wj(new)$ = (1 0.1575 0.987 -0.15)

Step 6: Second epoch is finished.

Step7: Test the stop criteria Thus for learning rate 0.05, the process is done. Hence the LVQ net has been constructed and tested

Accuracy: Accuracy of classifier refers to the ability of classifier. It predicts the class label correctly a Classification accuracy is the number of correct predictions divided by the total number of predictions. The fraction of predictions that a classification model got right. In multi-class classification, accuracy is defined as follows: Accuracy=Correct Predictions/Total Number Of examples

Prediction: At the initial stage of prediction data values are entered. After submitting the values data is observed and the report is given regarding the person having disability or not. Once LD has been detected, the system further classifies the case into the type/s of LD. Classification is done using the rule-based approach. It consists of an input stage, a processing stage, and an output stage. The input stage initiate the inputs to appropriate member functions and truth values. The processing stage invokes each suitable rule and bring out a result for each, then mixes the results of the rules. At last, the output stage turn over the amalgamated result back into a certain controlled output value

IF variable IS property THEN action

Example • IF temperature IS very cold THEN stop fan • IF temperature IS cold THEN turn down fan

• IF temperature IS normal THEN maintain level

• IF temperature IS hot THEN speed up fan

C. Rules for Diagnosis

Case-1: IF the score is low in categories of essay, reading, comprehension, spelling, perception, solve, word problems, mental sums, time, money; THEN itis Dyslexia, Dysgraphia and Dyscalculia

Case-2: IF the score is low in fields of reading, comprehension, perception, word problems

THEN Dyslexia

Case-3: IF score low in fields of spelling, comprehension, essay;

THEN it is Dysgraphia

Case-4: IF the score is low in fields of solve, mental sums, word problems, sums related to time, calendar, money; THEN it is Dyscalculia

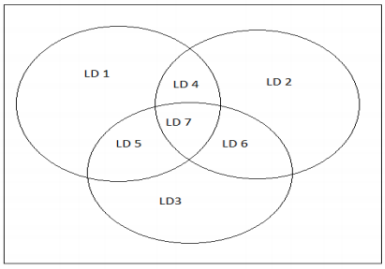
Case-5: IF the score is low in fields of reading, essay comprehension, perception word problem, spelling,

THEN it is Dyslexia and Dysgraphia

Case-6: IF the score is low in categoriesof reading, comprehension, perception, solve, mental sums, word problems, sums related to time, calendar and money; THEN it is Dyslexia and Dyscalculia

Case-7: IF the score is low in categories of spelling, comprehension essay, solve, mental sums, word problems, sums related to time, calendar, money; THEN it is Dysgraphia and Dyscalculia

Most of times, one type of Disability may leadto other two types as well; as such, most of the children are detected with all three types of LDs. Less common cases have one type of LD diagnosed or sometimes.



Distribution of classification of LD; represented by a Venn diagram showing the several overlapping cases of LD. Here LD 1: Dyslexia; LD 2: Dysgraphia; LD 3: Dyscalculia;LD 4: Dyslexia, Dysgraphia;LD 5: Dyslexia,Dysgraphia; LD 6: Dyslexia,Dyscalculia;LD 7: Dyslexia,Dysgraphia,Dyscalculia

4.TESTING

Introduction: After finishing the development of any computer-based system the next complicated time consuming process is system testing.During the time of testing only the development company can know that, how far the user requirements have been met out, and so on. Following are the some of the testing methods applied to this effective project.

Source code Testing: This examines the logic of the system. If we are getting the output that is required by the user, then we may say that the logic is perfect. **Specification Testing:** We can set with, what sort of program should be done and how it should be performed under various condition. This testing is a comparative study of evolution of system performance and system requirements.

Module Level Testing: In this the error will be found at each individual module, it encourages the programmer to find and rectify the error without affecting the other modules.

Unit Testing: In unit testing, the individual classes are tested. It is often seen if the class attributes are implemented in accordance to the design and if the methods and the interfaces are error-free. This testing is often perceived as the responsibility of the application engineer who sorts to implements the structure.

Integration Testing: Integration testing also regarded as integration and testing,) is the stage in software testing where an independent software modules are grouped and tested . It comes after unit testing and before validation testing. Integration testing takes as its 52 input modules that are to be unit tested, grouped in larger aggregates, applicable tests are defined in an integration test plan to those of the aggregates, and delivers as its output the integrated system ready for system testing.

Validation Testing: The process of assessing software at the time of development process or at its end to decideif at all it satisfies specified business requirements. It makes sure that the product in real meets the client's requirement.

Recovery Testing: Recovery testing is a kind of non-conventional testing technique carried in order to determine how quick the system will be able to recover after it had gone through system crash or hardware failure. Recovery testing is the forced failure of the software to verify if the recovery is successful or not.

Security Testing: Security testing is a process meant to reveal faults in the security mechanisms of an information system that is used to protect data and maintain functionality as intended. As a result of the logical limitations for security testing, passing this test could not become an indication of no flaws .**Performance Testing:** Performance testing, a non-conventional testing technique performed to determine the system parameters in terms of

responsiveness and stability under various workload.

Output Testing: After performing the validation testing, the next step is output testing of the proposed system since no system would be termed as useful until it does produce the required output in the specified format. Output format is considered in two ways, the screen format and the printer format.

User Acceptance Testing: User acceptance testing, a testing mechanism where the end users included in testing the product to validate the product against their requirements. It is performed at client location at developer's site

TEST CASES

SNO	Test case	Requirement specification	Expected output	Observed output	Status P=pass F=fail
1.	Register with correct details	RS1	The user should get registered	The user will be registered	Pass
2.	Login with correct details	RS2	The user should login	The user will login	Pass
3.	Login with wrong details	RS3	The user should not login	The user will not login	Fail
4.	Create dataset	RS4	The user should enter the values for parameters	The user will enter the values for parameters successfully	Pass
5.	View dataset	RS5	The user can view the dataset	The user will view the dataset	Pass
6.	Pre processing	RS6	The data given by user is preprocess	The will be preprocessed	pass
7	Classification	RS7	The dataset is trained and tested	The dataset will be classified into twoclasses	Pass
8	accuracy	RS8	The accuracy of the model is printed	The accuracy of the model is given	pass
8	prediction	RS8	The data enter is predicte to which class it belongs	The data will be predicted	pass
9	logout	RS9	Logout from the page	The user will exits from the page	pass

Experiment: The dataset consists of 107 record of which 58 are non-learning disability and 49 are learning disability The mean absolute error is one of a number of comparing forecasts with their eventual outcomes. The learning vector quantization has given a accuracy of 90.263% and decision stump has given accuracy of 69.4853% In this project the dataset is compared with Weka tool decision stump algorithm.

ALGORITHM	ACTUAL	FORECASTED	ERROR
DESCISION STUMP	107	75	32
LEARNING VECTOR QUANTIZATION	107	96	12

Confusion Matrix: • A confusion matrix is a table that is used to describe the performance of a classifier on a set of test data for which the true values are to be known. The confusion matrix itself is relatively simple to understand, but the related terminology can be confusing.

- **true positives (TP):** predicted yes and they do have the disability.
- **true negatives (TN):** predicted no, no disability
- **false positives (FP):** predicted yes, no disability (Also known as a "Type I error.")
- **false negatives (FN):** predicted no, theyactually do have the disability. (Also known as a "Type II error.")

N=107	Predicted: Non-LD	Predicted: LD	
Actual: Non-LD	TN=50	FP=8	58
Actual: LD	FN=4	TP=45	49
	54	53	

5.CONCLUSION

The project has shown us the skills needed to develop an application based on Learning Vector Quantization.

In the end we would like to conclude that our aim of developing a simple LVQ algorithm which when compared to other soft computing algorithms gives a competitive result. This method besides being simple, is also easy to replicate in large volumes but gives comparable results based on accepted benchmarks. It can also be seen that on increasing the number of data in the training set of the system, the overall accuracy shows a promising growth.

In future we are supposed to explore the possibility of parameter classification in order to distinguish irrelevant and superfluous variables which might lead to decrease in diagnosis process time and increase in accuracy. This could be beneficial to the special educators, doctors and teachers by providing suggestions that can lead to the debarringunnecessary tests and saving of time needed for diagnosing LD.

References:-

www.academia.edu/19803501/Learning_Disability_Diagnosis_and_Classification_A_Soft_Computing_Approach
• https://pythoninstitute.org/free-python-courses
• https://data.world/datasets/classification
• https://www.fullstackpython.com/flask.html
• https://www.datamini.co.in/
• Data Mining - Charu C. Aggarwal
• Learning Python — The Hitchhiker's Guide to Python