JYOTI NIVAS COLLEGE AUTONOMOUS
POST GRADUATE CENTRE

DEPARTMENT OF MCA
TECH-ON-TAP

ISSUE 3
September
2019

CONTENTS

INTERIOR DESIGNING

AFNAAN K (172MCA34)

Problem Statement:

Completely automated intelligent system that produces realistic and aesthetically appealing interior designs for homes. The particularly striking feature of our system is that it generates multiple plausible options for an empty room. The relationships between different elements of a room and items placed in the room are represented as Bayesian networks. The causal relationships defining the network structure are derived from standard thumbrules of interior designing. The parameters for every node in the network are learnt from information extracted semi-automatically from the top view images of furnished living rooms and conversation areas. New layouts based on user inputs are generated upon inferencing from this learnt network.

Learning System:

An almost completely automated intelligent system that produces realistic and aesthetically appealing interior designs for homes.Mimics the human interior designer as it generates multiple plausible options for the end-user to choose from.The relationships between different elements of a room and items placed in the room are represented as Bayesian networks. The locations, orientations and stylistic attributes of basic entities of a room form the nodes of the network.

Training Phase:

Picture of the room is Clicked First ,then Object extraction along with their positions and orientations semi-Automatically (supported by little manual annotation) from 2d-layout diagrams.A modeling framework entirely in terms of Bayesian networks thatcreates multiple interior designs for a user.Parameter learning in a hybrid Bayesian network with both continuous and discrete nodes. Automatic selection of furniture items. A generic algorithm that can be easily extended to find shape, size, color , texture etc. of the objects and also to generate designs for other rooms.

Suitable Attributes:

- Object and relationship Extraction
- Ceiling
- Type of Flooring
- Depth Estimation
- Size of the Room
- Theme of the Room
- Texture size and color of objects
- Light fixtures
- Colour Scheme

Algorithms used for various attributes:

- SIFT (Scale-invariant Feature Transform),
- SURF (Speeded Up Robust Features)
- Histogram of oriented gradients (HOG) ,
- Local binary patterns (LBP)

- Deep Convulution neural network , K- means alorithm, K-nearest Neighbour Algorithm,
    - Forming the Hierarchical Bayesian Networks. discrete nodes use a Dirichlet distribution, continuous nodes use a multivariate Gaussian distribution.

Performance Metric:

The app would analyse chosen interiors or can also suggest interiors as per considering existing attributes**.** Confusion Matrix, it is the easiest way to measure the performance of a classification problem where the output can be of two or more type of classes. A confusion matrix is nothing but a table with two dimensions viz. "Actual" and "Predicted" and furthermore, both the dimensions have "True Positives (TP)", "True Negatives (TN)", "False Positives (FP)", "False Negatives (FN)". In this case true match and false match of interiors.

References:

- https://uk.blastingnews.com/tech/2018/05/what-does-ai-artificial-intelligence-mean-for-interior-design-002539097.html
- https://www.collaborativepractice.com/system/files/2%20-%20Elements%20of%20Interior%20Design.pdf
- http://matrixae.com/services/
- https://www.builderonline.com/products/home-technology/algorithim-based-interior-design-not-far-off_o

PREDICTING CASE OUTCOMES

AISHWARYA M (17MCA01)

Problem Statement:

Can Lawyers predict the outcome of their cases? How accurately do lawyers predict their case outcomes? These forecasts play a pivotal role in practical legal decision-making, and affect many stakeholders: the lawyer; the client; and the justice environment as a whole. Prediction errors can cost the client and their lawyer. A lawsuit can be very challenging especially for those who are unfamiliar with the legal process which includes the following major acts: meeting with the potential lawyers, hiring a right lawyer, deciding in which court to file the lawsuit

Learning Task:

The need for case assessment comes as early as the initial investigation into whether there are sufficient grounds on which to file (or defend) a lawsuit and can continue through trial, where settlement or a plea bargain is an option up until the time a verdict has been reached.

At every stage along the way, the lawyers on each side of a case have to consider the possible outcomes (e.g., summary judgment, settlement, plea bargain, jury trial, bench trial) and estimate the probability of a successful outcome for their client. A client's decision about whether to proceed with litigation can be affected in large part by her lawyer's predictions about case outcomes. For example, a lawyer in a civil lawsuit who predicts an unfavorable outcome at trial can strongly recommend and pursue pretrial settlement. An attorney who predicts a favorable trial outcome can recommend that a client accept or reject settlement offers in light of the expected damage award at trial.

Inaccurate predictions, both in terms of overconfidence and under-confidence, can be costly. A lawyer who overestimates the possible outcome of a case (e.g., winning at trial and/or receiving a large damage award) may advise a client to reject reasonable settlement offers. A lawyer who underestimates the possible outcome may advise a client to accept a settlement offer significantly lower than what could have been achieved through a favorable trial verdict.

Training Phase:

The system is trained based on the grading system.The grading reflects factors like how strong the brief's arguments are, how persuasive the relied upon cases are, and the extent to which the brief cites precedent that supports the desired outcome. The higher a brief's grade, the more likely it is that the brief will win.

Lawyer grades a brief's:

Arguments: measuring the odds of a brief winning or losing based on brief-specific features that cover the strength of the legal arguments, the persuasiveness of the cases relied upon, and the balance of the argumentation.

Context: measuring the odds of a brief winning or losing based on contextual (non-brief specific) information like the Causes of Action, the Procedural Posture, and the Trial Court Judge. The mathematical difference between these two grades (Difference = Arguments - Context) quantifies how strongly or weakly the brief is performing, with briefs falling into one of five categories:

Strongly Overperforming: the brief's Arguments grade is 10 or more points higher than the brief's Context grade, indicating that the Arguments score suggests the brief is far more likely to win than the Context score suggests.

Weakly Overperforming: the brief's Arguments grade is fewer than 10 points higher than the brief's Context grade, indicating that the Arguments score suggests the brief is a little more likely to win than the Context score suggests.

Performing as Expected: the brief's Arguments grade equals the brief's Context grade, indicating that the Arguments and Context agree on the odds of the brief winning.

Weakly Underperforming: the brief's Arguments grade is fewer than 10 points lower than the brief's Context grade, indicating that the Arguments score suggests the brief is a little less likely to win than the Context score suggests.

Strongly Underperforming: the brief's Arguments grade is 10 or more points lower than the brief's Context grade, indicating that the Arguments score suggests the brief is far less likely to win than the Context score suggests.

Features / Attributes with at least 10 sample data:

A wide range of data is collected for this system from the various legal Indian judicial sites (ref). This collected data is then prepared according to the need of the algorithms for the training and testing. This step of feature extraction is performed by using the textual information of the cases and removes the stopping words or words irrelevant to describe the feature of the case, thus we use vectorization technique to define the weightage of a word in the description of the case. Apart from giving this information, vectorization says how important that word is to that document with respect to the corpus. Vectorization technique is also used to convert our feature extracted into a numerical value which then can be used for the evaluation of the models and can help us to plot the point graphically.

Dataset :

Dataset used here is the publicly available data which is published by the ECtHR. The European Court of Human Rights is an international court that was established in 1959. A judicial decision of the ECtHR contains the following main parts: • Introduction, consisting of the title (e.g., Lawless vs. Ireland), date, Chamber, Section of the Court and its constitution (i.e. judges, president, registrar); • Procedure, containing the procedure that took place from lodging and application until the judgement by the Court; • Facts, consisting of 2 parts: • Circumstances, containing a relevant background information on the applicant and events and circumstances that led them to seek justice due to alleged violation of their rights in accordance to ECHR; • Relevant Law, containing relevant provisions from legal documents other than the ECHR (these are typically domestic laws, as well as European and international treaties); • Law, containing legal arguments of the Court with each alleged violation discussed separately; • Judgement, containing the decision of the Court per alleged violation; • Dissenting/Concurring Opinions, containing the additional opinions of judges, explaining why they voted with the majority (concurring opinion) or why they did not agree with the majority (dissenting opinion).

CONVERTING TEXT TO SONG

AMALA SEELAN GV(17MCA02)

Problem statement:

If he/she is a slow learner and if the person faces difficulty in understanding the subject related concepts and also if he/she is interested in music kind of applications can utilize this system.

This system converts the text oriented concepts into song which can be easily understandable without any modification (converted song contains all the text based concepts). A song category is user friendly. They can pick up any type from the different categories like pop, Jazz, classic, rock, melody and so on.

Learning Task:

1. The system should identify the concepts according to the user needs.
2. The system should covert the text based concepts into a song according to their song categories.
3. The system should able to match the sentence to corresponding lyrics without any modification.
4. The system should accept the algorithm to make use of any audio formats like mp3, wav, AIF.
5. The system prepares itself for the outcome.

Attributes:

- Pitch
- Rhythm
- Dynamics
- Texture
- Timbre

Algorithm:

To implement this system, we use an algorithm called, Restricted Boltzmann Machine(RBM). RBM is a neural network with two layers, the visible layer and the hidden layer. Each visible node is connected to each hidden node (and vice versa), but there are no visible-visible or hidden-hidden connections (nodes are simply where calculations take place). This is the restriction. Each visible node takes one chord. Each chord is multiplied by a weight and then the node's output at the hidden layer. Unlike most of the neural networks, RBMs are generative models that directly model the probability distribution of data.

References:

- https://towardsdatascience.com/deep-learning-with-tensorflow-part-3-music-and-text-generation-8a3fbfdc5e9b
- https://en.wikipedia.org/wiki/Restricted_Boltzmann_machine

ANALYSIS OF INFANT ACTIVITIES

ANNAPURNA PAILA(17MCA04)

Problem Statement:

As we are very well familiar with the hurdles faced by Parents to nurture their infant and especially in case if both the Parents are working Thus, we need to develop something unique that can help Parents to have a continuous surveillance on the Infant and according to the infant cry ask caretaker to take proper care. Parents can monitor baby's activity such as baby's fall detection and baby's cry by their movement of aged 1 to 2 years.

- **CRY DETECTION**

Crying is an infant behaviour and one of the way of communications and a positive sign of healthy life for the infant
1. Hungry: If the infant cries because of hunger, then the caretaker have to feed them to stop crying.
2. Stomach pain: If the infant cries due to stomach pain, then the caretaker have to take some remedies to stop their baby's cry.
3. Unhappy: If the infant cries due to unhappy, the caretaker should understand that they are feeling very uncomfortable with wet diapers etc.
4. Sad: If the infant cries out of sadness, then the caretaker should demand more attention towards their infant and has to make them feel safer.

Attributes:
Pitch Information, voice frequency of infant cry

Training: Cry signal of the infant is directly acquired and stored. By using various feature extracting methodology. One picked signal from the test folder will be compared with all the database signals one by one and when the corresponding feature is matched, then the reason for which the infant cry will be displaced.

Acquire cry signal | | Feature Extraction | | Automatic Classification | | Display result

Algorithm:
As the infants cry has some normal frequency range, the fundamental frequency (f0) is considered as a classifier. Pitch and frequency components are detected by using the two different combinations of different algorithms. i) Mel Frequency Cepstral Coefficient and Cross - correlation Method ii) ii) Linear Prediction Coefficient and auto - correlation Method.

- **FALL DETECTION:**

Missing to identify falls can have serious health and safety implications on an individual. Asinfants are at the highest risk for falling during sleeping or playing, it is sometimes cause of death. Fall can be - Fall from sleeping (bed), Fall from sitti,Fall from walking or standing attributts calculate the changes of body position of ankles knees, wrists elbows and shoulders

Algorithm:-
Head Motion - The principle used in this approach is that during a human fall vertical motion is faster than the horizontal motion.

HMM - HMM uses multiple features extracted from silhouette: height of bounding box, magnitude of motion vector, determinant of covariance matrix and ratio of width to height of bounding box of person

Training:

The main perception with this approach is that the shape of a person will change from standing to lying if a fall occurs. When baby will move the changes position will be calculated and will match with database and signal will be updated.

Reference:

https://www.netguru.com/codestories/how-to-use-machine-learning-models-to-detect-if-baby-iscrying.case-study
https://publisher.uthm.edu.my/ojs/index.php/ijie/article/view/1332/911
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5017327/

FACE OFF WITH MY DANCERS TEAM

-R BHAVYA SHREE(17MCA21)

Problem Statement:

An app which consists of dancers of various forms. The main objective of this app is to group dancers of their specialized form and perform in competitions.

Learning Task:

Dance has always been consider as an expression of rhythmic movement that work as an intensified sense of life, which basically arise from the inner perception whose function is to stimulates both mind and the body. It has been witnessed that people all around the world expressed and shared their feelings through the dance, since the origin of the human life on this planet.

Competition is a place where pure talent can be discovered and then launched into the next generation of professional dancers. The undertaking of a new type of competition is large, but the benefits are crucial. Reshaping the competitive dance industry will take time, but it starts with a brave individual to stand up and be the change for what the dancers of tomorrow need. The main purpose is the findings and ideas to preserve the integrity and artistry of the competitive dance culture.

Training Phase:

Predictive modeling, from Machine Learning Mastery, refers to *the mathematical problem of approximating a mapping function (f) from input variables (X) to output variables (y). This is called the problem of function approximation.* In other words, we are using historical data to make predictions about new data. Generally speaking, we could divide most function approximation problems into two major classes: regression and classification. A classification problem *consists of taking input vectors and deciding which of N classes they belong to, based on training from exemplars of each class.* Classification problems are discrete, meaning that each input vector belongs to one class, and the set of classes covers the entire output space.

In this paper we are going to classify dancers based of their specified features and their dance style.

Features:

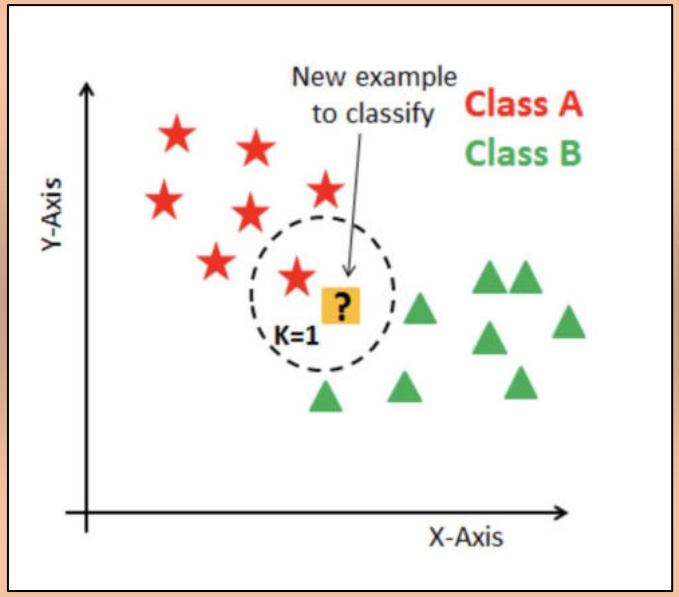
The dataset used in this paper consists of the following predictive attributes:

- Dance style
- Experience
- Place
- Competitions participated
- Current competition details
- Availability(Timing)

SUITABLE ALGORITHM TO SOLVE THE PROBLEM

K-Nearest Neighbors (KNN):

Suppose you're going to the Opera, but you've never been to the Opera before. You show up at the Opera not very sure about what to do. So you pick the person closest to you, observe, and imitate. Then you go on to the 5 or 10 people closest to you, observe, and imitate. This is the general idea behind KNN.



KNN stores the entire dataset and makes predictions for a new data point by searching for the *k* nearest instances, and summarizing the output variable for those instances. To determine the similarity between instances, it computes a measure of distance. This measure can be Euclidean distance, Hamming distance, etc.

The distance metric is again, a personal choice that depends on your specific problem and data; and so is the *k* number. Usually, you'd want to experiment with different measures of data and values of *k* together to see which combination results in the most accurate model.

REFERENCES:

https://www.researchomatic.com/historicalcultural-dance-analysis-164032.html
https://digitalcommons.pace.edu/cgi/viewcontent.cgi?article=1189&context=honorscollege_theses
https://towardsdatascience.com/to-dance-or-not-to-dance-the-machine-learning-approach-decc2a7782b5

PREDICTING HOUSING PRICES FOR REAL ESTATE COMPANIES

MAITHRY A (172MCA33)

Problem Statement:

Prices of real estate properties are sophisticatedly linked with our economy. Despite this, we do not have accurate measures of housing prices based on the vast amount of data available. Therefore, the goal of this project is to use machine learning to predict the selling prices of houses based on many economic factors.

Learning Task:

One heuristic dataset commonly used for regression analysis of housing prices. Former analyses have found that the prices of houses in that dataset are most strongly dependent on their size and the geographical location. Until recently, basic algorithms such as linear regression can achieve 0.113 prediction errors using both intrinsic features of the real estate properties (living area, number of rooms, etc.) and additional geographical features (socio demo-graphical features such as average income, population density, etc.).

Sample Data Set:

	area_type	availability	location	size	society	total_sqft	bath	balcony	price
0	Super built-up Area	Ready To Move	Brookefield	2 BHK	Roeekbl	1225	2.0	2.0	39.07
1	Plot Area	Ready To Move	Akshaya Nagar	9 Bedroom	NaN	2400	9.0	2.0	120.00
2	Plot Area	18-Apr	Hennur Road	4 Bedroom	Saandtt	1650	5.0	2.0	62.00
3	Super built-up Area	Ready To Move	Kodichikkanahalli	3 BHK	Winerri	1322	3.0	1.0	95.00
4	Super built-up Area	Ready To Move	Konanakunte	2 BHK	AmageSa	1161	2.0	1.0	51.00

Features:

- Area type – describes the area built
- Availability – when it can be possessed or when it is ready(categorical and time-series)
- Location – where it is located in Bengaluru
- Price – Value of the property in lakhs(INR)
- Size – in BHK or Bedroom (1-10 or more)
- Society – to which society it belongs
- Total sqft – size of the property in sq.ft.
- Bath – No. of bathrooms
- Balcony – No. of the balcony

Algorithm:

Regression is a machine learning tool that helps you make predictions by learning – from the existing statistical data – the relationships between your target parameter and a set of other parameters. According to this definition, a house's price depends on parameters such as the number of bedrooms, living area, location, etc. If we apply artificial learning to these

parameters we can calculate house valuations in a given geographical area.

The regression works on given data and finds the relationship between parameter in the input and the output. To show you how regression algorithm works we'll take into account only one parameter – a home's living area – to predict price. It's logical to suppose that there is a linear relationship between area and price. $y = k0 + k1*x$

In our case, y equals price and x equals area. Predicting the price of a home is as simple as solving the equation (where k0 and k1 are constant coefficients):

$$price = k0 + k1 * area$$

We can calculate these coefficients (k0 and k1) using regression. Regression takes every possible value for k0 and k1 and minimizes the total deviation; this is the idea of regression in a nutshell. But in real life, there are other challenges you need to deal with. House prices obviously depend on multiple parameters, and there is no clear linear relationship between all of these parameters.

References:

https://www.kaggle.com/erick5/predicting-house-prices-with-machine-learnining
https://yalantis.com/blog/predictive-algorithm-for-house-price/

PREDICTION OF CHEMICAL PROPERTIES OF MOLECULES.

RABIA FIRDOUS (172MCA34)

Problem statement:

Prediction of chemical properties using only image data of 2D drawings of molecules.

Learning task:

The system uses non-linear functions which allows it to learn representations and extract necessary features from input data to predict the property desired. Representation learning ability is of key importance in computer vision.

Training phase:

Utilizing "raw data" in the form of 2D drawings of molecules that requires the minimal amount (i.e. no higher than high-school level) of chemical knowledge to create. By training deep convolutional neural networks to predict chemical properties that spans a broad range of categories without the input of advanced chemistry knowledge, but instead allowing the network to develop its own representations and features from the images it is trained on. Requires minimal chemical knowledge beyond the generation of 2D chemical structures. SMILES strings are converted to their respective 2D molecular structures, which were then mapped onto an input array used to train the convolutional neural network in a supervised fashion. Apart from the generation of chemical images and the measured chemical properties used to train the model; no additional source of chemistry-inspired features, such as molecular descriptors or fingerprints were used.

After a SMILES to structure conversion, the 2D images are mapped onto an 80 x 80 image that serves as the input image data for training a deep neural network to predict toxicity, activity, and solvation properties.

Features/ attributes:

Chemicals in the databases are stored as smiles strings which are compact string structures that describes a molecule's structure. The resulting coordinates of the 2D structure of each molecule were then mapped onto an 80 x 80 grid, where each pixel had a resolution then resulted to be array the resulting 80 x 80 array 9 was then greyscale "colour coded" based on the presence of an atom or a bond. Specifically, atoms mapped onto the grid were assigned a number based on its atomic mass unit, and bonds mapped onto the grid were assigned the number 2, as it does not correspond to the identity of any element in the training set. The other parts of the grid were empty (i.e. vacuum) and were defaulted to the number 0. The resulting discretized image of the molecule was then parsed into a deep convolutional neural network for training.

Performance metric:

Achieves a level of performance that is on par with expert-developed QSAR/QSPR models based off molecular descriptors, fingerprints and other engineered features.

The evaluation metric is RMSE (root mean square error), which for a given dataset of n samples

is defined as:

$$\text{RMSE} = \sqrt{1\,n\sum(x\,i\,,\,p\,r\,e\,d\,i\,c\,t\,e\,d\;-x\,i\,,\,m\,e\,a\,s\,u\,r\,e\,d\,)\,2}$$

Suitable Algorithm:

Rectified linear activation function, RMS (Root mean square) propalgorithm, fine stage tuning with e stochastic gradient descent (SGD) algorithm.
Augmentation is performed on the input images.

References:

- Dahl, G. E.; Jaitly, N.; Salakhutdinov, R. Multi-task Neural Networks for QSAR Predictions arXiv:1406.1231 2014.
- He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. arXiv:1512.03385 2015.
- Krizhevsky, A.; Sutskever, I.; Hinton, G. E. ImageNet Classification with Deep Convolutional Neural Networks Advances in Neural Information Processing Systems 2012.

PREDICTION OF FUTURE TV RATINGS

SANJANA K R (17MCA24)

TV ratings have been a mainstay of the media industry for over half a century. They're used to make programming decisions and have become part of our popular culture1, but they are also the basis for billions of rupees worth of advertising transactions every year between marketers and media companies. They help measure the success of TV shows, verify that their audience size and composition are delivering against strict media-buy targets, and provide a basis for make-goods if the numbers come up short. From that point of view, TV ratings are metrics that measure the past, or at best the present, of TV viewing.
But ratings are also used to predict the future. They set expectations and affect programming decisions from one season to the next, and they help set the cost of advertising (advertising rates) well in advance of when a program goes on the air

Features:

1. **Program characteristics**: known elements to access and categorise a show
2. **Program performance:** performance on measurable dimensions. Ex: historic ratings
3. **Promotional support:** investment in driving awareness among audience marketing spend on/cross-air promos
4. **Audience engagement:** audience interest and commitment to a show ex:television brand effect
5. **Social/on-line behavior:** social media information

Training Phase:

Algorithm randomly split the data into training and cross-validation testing sets. The model learned by making predictions based on the training set, testing those predictions on the cross-validation testing set, and repeating the process multiple times using different parameters. The final parameters were selected with consideration to the results of the cross-validation, helping limit the tendency to overfit the model to the training set

Also held out some data that was never used in the buildup process, but served as another layer to test the validity of our model and protect against overfitting. Holdout validation testing data provides an additional measure of quality control in the overall process. Models still tend to overfit even when using cross-validation. In order to choose the parameters most appropriate to apply to a new dataset, it is usually better to choose results that are slightly conservative, even for the testing dataset. The holdout validation testing set helped us achieve that balance
Once everything checked out and the final parameters were set, we retrained the model using the best parameters to leverage the most complete information available. Then we ran it on a new dataset and compared its performance to client projections, focusing on key demographic groups.

Methodology:

Many models and machine learning algorithms, including linear regression, penalized regression, multiple adaptive regression splines, random decision forests, support vector machines, neural networks and gradient boosting machine (GBM)3 can be used. While each method has its own advantages and disadvantages, in the end, the GBM method (specifically, the xgboost optimized library) is best algorithms that offer the best combination of accuracy and scalability for our project.

Performance Metrics:

WAPE (weighted mean absolute percentage error) is used to evaluate the accuracy of models. WAPE is a statistical measure that helped us ensure that the way our model fit new data was reasonably consistent with how it fit historical data.

Used WAPE to compare model's accuracy model at two different levels. The first was at the channel level, which placed little emphasis on the ability to distinguish between programs, but was focused on getting the high level trends right—such as overall TV viewership for each channel. We also compared WAPE at the hour-block or program level. The hour-block level looked at the model's ability to distinguish between shows, as well as its ability to understand the high-level effects that influence all shows.

COMPLETING A PARTIALLY DRAWN BASIC SHAPE

SHILPA S (172MCA35)

Problem Statement:-

The application aims at how to identify and suggest the shapes that user is about to draw on the screen Given a user hand drawing a shape on a touch screen and a database of known shapes, determine which shape the user was trying to draw. A model of this decision would allow a program to show the platonic version of that shape the user drew to make crisp diagrams.

Steps:

first the system should be capable of identifying the shape that the user is going to draw and then it should suggest the shapes similar to that in which user can select which he want to draw.Shape recognition and providing suggestions are the 2 important steps of this problem.

Implementation:-

We present an approach for recognizing multi-stroke hand-drawn symbols. The main feature of the approach is its capacity of recognizing partially drawn symbols. Furthermore, it is invariant with respect to scale, and supports symbol recognition independently from the number and order of strokes. The recognition technique is based on subgraph isomorphism and exploits a novel spatial descriptor, based on polar histograms, to represent relations between two stroke primitives. Using different symbol sets, both hand-drawn and artificially deformed, we evaluated the effectiveness of the approach in recognizing the symbols as a function of the number of primitives already drawn by the users. The results show that the approach gives a satisfactory recognition rate with partially drawn symbols, also with a very low level of drawing completion, and outperforms the existing approaches proposed in the literature. We also report the results of a user study aimed at evaluating whether the users can efficiently exploit symbol autocompletion.

Another set of algorithms that can be used for implementing the above problem statement are pattern recognition technique and learning paradigm. Traditional memory stores data at a unique address and can recall the data upon presentation of the complete unique address. An associative memory is a brain like distributed memory that learns by association. It takes one of the two form:

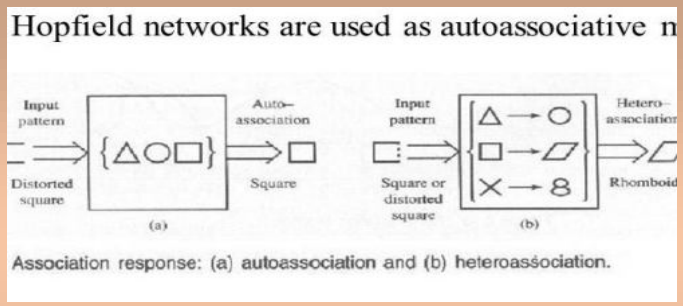
1) Auto association: required to store a set of patterns by repeatedly presenting them to the network. these memories are capable of retrieving a piece of data upon presentation of only partial information from that piece of data. The network is presented a partial description of an original pattern stored in it, and the task is to retrieve that particular pattern. It uses unsupervised learning

2) Hetero association: It differs from autoassociation in that an arbitary set of input patterns is paired with another arbitary set of output patterns. these memories, on the other hand, can recall an associated piece of datum from one category upon presentation of data from another category. It uses supervised learning

Training data:

four shapes dataset– this dataset consists of 16,000images of four basic shapes (star, circle, square, triangle). Each image is 200/200 pixels.

Example:



Association response: (a) autoassociation and (b) heteroassociation.

References :-

- https://www.kaggle.com/smeschke
- https://www.hindawi.com/journals/cin/2017/5705693
- https://patents.google.com/patent/US20090324076
- https://architosh.com/2015/06/think-with-your-fingers-instaviz-2-0-diagram-sketching-for-ios-launched

PET DOG HEALTH MONITORING SYSTEM

SWEETY LENKA (17MCA26)

Problem Statement:

Pets, especially dogs share a very close and deep relationship with people, and this bond has only grown over the years. Now, when the pet is unwell, people are more inclined to invest in healthcare and treatment, giving a significant boost to the animal healthcare and veterinary services industry. Usually a pet dog gives a sign to its owner by expressing unusual behavior or by the change of its body when its health is at risk or having a disease. However, without deep knowledge about the pet dog's disease, owners tend to neglect such signs but only depend on the regular check by veterinarians only to make the situation worse.

This could be resolved by developing a pet dog health pre-diagnosis/monitoring system with machine learning (an artificial intelligence technique) that can be easily checked by dog owners without deep knowledge of computer technology or dog diseases.

Learning Task:

The proposed system requires dog owner's only symptoms they found from their pets then the system answers some number of most probable diseases pets may have with computed confidence rate. Technologically, like many existing health diagnosis system for human beings the system requires a database for symptom-disease associations and the inference system from that database with given symptoms recognized by pet owners.

- **Disease-Symptom Data Collection**

  For the pet dogs' disease-symptom data, with respect to 13 body parts that the symptoms occur whole body, head, abdomen, leg, hip, eye, nose, mouth, ear, hair, skin, temper, and excretion. Those symptoms are associated with diseases and verified by veterinarians. Some representative symptoms associated with a certain disease also creates other queries to user for further considerations.

- **Database Construction**

  Obtained symptoms and corresponding diseases are associated with different strength that will be learned by our ART2 learning engine. The database consists of three main tables.

  • Disease (ID, Disease Code, Symptom Code, Description)

  • Symptom (Symptom ID, Body Part, Description)

  • Learned Result (Cluster ID, Input Neuron, Strength) Symptom Code of Disease table may contain multiple Symptom IDs of Symptom table. Since usually a disease has multiple symptoms. The learning result table stores the connection strength after learning between disease and symptom.

- **Overall Procedure**

  The system starts with asking representative symptoms currently a dog suffers from the user. This symptom input procedure is carefully designed to minimize misleading input.

ART2 Learning System:

The key of the diagnosis system is its learning engine based on ART2 that is a self organizing neural network presented as the solution for the plasticity-stability problem without suffering from the local minima problem.

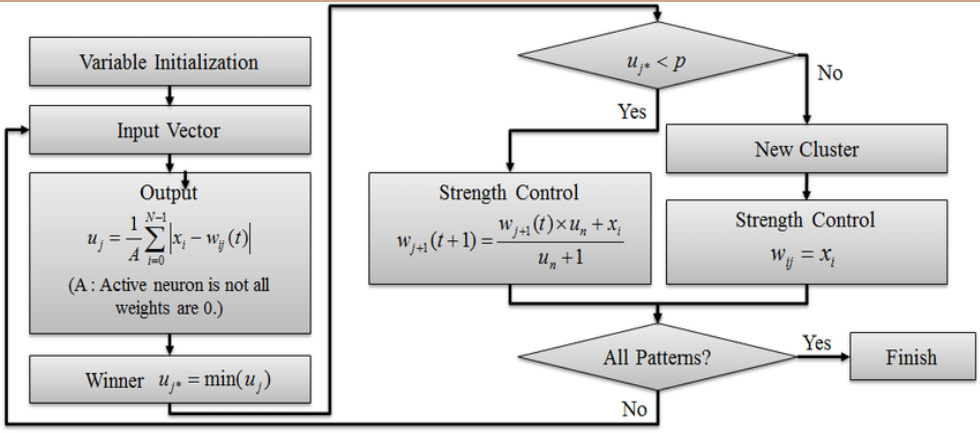
Its characteristic can be summarized as follows.

1) ART2 is a self-organizing pattern clustering structure by competitive learning.

2) It is a stable and adaptable neural network with incremental learning ability, that is, new learning procedure does not affect already existing clusters.

3) No local minima problem.

4) It is learnable with binary input and analogue input. The change of connection strength is the average of all input patterns thus it is uniformly distributed to all clusters.



ART2 Learning Algorithm

Performance Metric:

In order to avoid misleading user input symptoms, the system has query process to refine the search. Thus, the user firstly gives a most obvious symptoms then the system reacts with showing related symptoms so that user can select observed symptoms more accurately and this procedure is done twice to minimize the risk of lack of available information. The main role of this system is not to lose the timing of treatment if the owner sees highly confident disease diagnosis through this system.

References:

https://pets.webmd.com/dogs/enriching-your-dogs-life#1https://www.dragonblogger.com/how-machine-learning-can-be-useful-in-animal-healthcare-and-veterinary-services/

N. Kugure, “Knowledge of Pet Dog Disease and Taming”, Yuwon Print Ltd., (2001).

Y. Dateuchi, “Pet Dog Diseases and Treatments”, Haseo Print Ltd., (2001).

PERSONALIZED MEDICINES

ZULFIN ARA (172MCA37)

Problem Statement:

Personalized medicine, or more effective treatment based on individual health data paired with predictive analytics, is also a hot research area and closely related to better disease assessment. The domain is presently ruled by supervised learning, which allows physicians to select from more limited sets of diagnoses, for example, or estimate patient risk based on symptoms and genetic information.

The Advantages of Personalized Medicine

- Better medication effectiveness, since treatments are tailored to patient characteristics, e.g., genetic profile.
- Reduction of adverse event risks through avoidance of therapies showing no clear positive effect on the disease, while at the same time exhibiting (partially unavoidable) negative side effects.
- Lower healthcare costs as a consequence of optimized and effective use of therapies.
- Early disease diagnosis and prevention by using molecular and non-molecular biomarkers.
- Improved disease management with the help of wearable sensors and mobile health applications.
- Smarter design of clinical trials due to selection of likely responders at baseline.

Learning Tasks:

Consume Medical Literature, Understand the data, Generate insight, Review treatment options

Training Phase:

1. Consume Medical Literature
2. Understand the data
3. Generate insight
4. Review treatment options

Consume Medical Literature:

It represents our knowledge about human diseases, including their subtypes.

For instance, the evolving definition of breast cancer molecular subtypes is an example for our incomplete knowledge of complex diseases such as cancer. However, the robustness of this new classification still remains to be demonstrated. Therefore, a disorder that is not known cannot be screened and investigated

Understand The Data:

Understand the longitudinal medical record and applying natural language processing and advanced cognitive algorithms to each unique patient case.

Generate Insight:

Generate a list of potential treatment options ranked by applicability—recommended, for consideration, and not recommended

Review Treatment Options:

Review treatment options and supporting evidence side by side to understand Watson's rationale and quickly access the relevant articles and clinical data.

Features / Attributes:

Age, gender, blood pressure, low-density lipoprotein (LDL) cholesterol, diabetes, obesity, inactivity, alcohol, smoking, demographic features, medical history, and auxiliary examinations (electroencephalogram [EEG] and magnetic resonance imaging [MRI])

Example:

Pt ID	Date	Diagnosis/Prescription/Procedure
207a3d56	2007.7	Lipitor
207a3d56	2010.8	Chest pain
207a3d56	2010.83	Angina pectoris
207a3d56	2011.2	Myocardial infarction

Suitable Algorithm/S to Solve the Problem

Five classical machine learning algorithms, i.e., Decision Tree, Random Forest, Support Vector Machine, XGBoost, and Logistic Regression, are selected and trained by our dataset to get classification models.

REFERENCES:

https://www.forbes.com/sites/insights-intelai/2019/02/11/how-machine-learning-is-crafting-precision-medicine/#3b5f70b05941
https://www.sciencedirect.com/science/article/abs/pii/S0005796719300981
file:///C:/Users/zulfi/Downloads/2438-Article%20Text-4072-1-10-20121206%20(2).pdf
file:///C:/Users/zulfi/Downloads/make-01-00009-v2.pdf