Percentage of Manipulative News Published by News Websites in India

Aditi Dave and Kushal Shah
Department of Electrical Engineering and Computer Science, IISER Bhopal.
Email : aditid19@iiserb.ac.in, kushals@iiserb.ac.in

Abstract

Media is the fourth pillar of our democracy and a significant source of getting correct and precise information about various events around the world. In the present digital age, online news websites have become an important source of news and information for an ordinary person. However, nowadays, it becomes difficult for us to get precise information, due to the menace of fake news. In order to analyse the extent of this problem, using tools of Machine Learning, we have performed an analysis of the content provided by these news websites.

There can be two types of fake news, one which is factually incorrect, and the other which is factual, but presented in a manipulative manner and thus, representing fictional style of writing. In our analysis, we are dealing with the latter, and an algorithm for detecting it was developed earlier by a research group at IISER Bhopal [1]. This algorithm takes an English text (in our case, a news article) as input and, using Logistic Regression, outputs the probability of the text being written in a manipulative (fiction) or straight forward (non-fiction) manner. From this earlier analysis, it was found that fiction texts tend to have a higher value for the ratio of number adverbs to adjectives and non-fiction articles tend to have a higher value for the ratio of the number of adjectives to pronouns.

For our analysis, we have collected the articles from a few news websites on a daily basis, and then calculated the percentage of these manipulative/fiction articles published by each of these news websites. We found that there is a large variation between various news websites in this regard.

References :

1. M. R. Qureshi, S. Ranjan, Rajakrishnan P. R., K. Shah, *A Simple Approach to Classify Fictional and Non-Fictional Genre*, Proceedings of the Second Storytelling Workshop, Association for Computational Linguistics, pages 81–89 Florence, Italy, August 1, 2019.

ANALYSIS OF SEARCH TECHNIQUES AND STRATEGIES FOR SOLVING 8 PUZZLE PROBLEM:

ABSTRACT:

Search is characteristics of almost all AI problems. Search strategies can be compared by their time and space complexities using big O notation. Here we discuss the search techniques and control strategies in solving 8 puzzle problem.

*N-puzzle problem has been one of the basic problems since the beginning of artificial intelligence. The most popular version of N-puzzle problem is 8 puzzle problem. It is simple and yet obeys a combinatorically large problem space of 9! /2 states. The N*N extension of the 8 puzzle is NP-hard. It is a sliding puzzle that consists of a frame of 8 numbered square tiles in random order with one tile missing. We are given an INITIAL state and we have to reach GOAL state which is also specified. In the first part of the paper we present complete statistical data based on an exhaustive evaluation of all possible tile configurations. Our results include data on the expected solution, the 'easiest' and 'worst' configuration and the density and distribution of solution nodes in search tree. In a search tree we make use of an appropriate heuristic which guides us to take a decision from a current state to next state in an intelligent manner.*

In our second set of experiments when heuristic search methods fail to describe how the shortest distance to a goal should be estimated, we use the 8-puzzle problem as a workbench model to evaluate the node ordering schemes in Iterative Deeping A(IDA*). More importantly, the search parameters that enables the highest gains are quite consistent and reduce the state space significantly resulting in memory and time optimization.*

Author: Sudipta Mukhopadhyay

KEYWORDS: Time and space complexity, NP-hard, search space and search tree, heuristic search, IDA*.

APPLYING TRANSFER LEARNING AND ATTENTION TO NER

Nirnay Roy

IISER-Bhopal

Email: nirnay16@iiserb.ac.in

Named Entity Recognition for social media data is tough because of its inherent noisiness. In addition to unfitting grammatical structures, it contains spelling discrepancies and many informal abbreviations. An attentional mechanism has lately been used to improve neural machine translation (NMT) by selectively focusing on parts of the source sentence during translation. Attention was also found to be useful in other NLP tasks like dependency parsing and other sequence labelling tasks. Inductive transfer learning has greatly impacted computer vision, but existing approaches in NLP still require task-specific modifications and training from scratch. This project attempts to apply these novel techniques to Named Entity Recognition on the best submission in the 3rd Workshop on Noisy User-generated Text (WNUT-2017) and look for improvements in the F1 Score metric.

Automated Pollen Source Analysis using Deep Convolutional Neural Networks

Anvit Mangal[1], Jainendra Shukla[2], Swapna R. Purandare[3]

Indraprastha Institute of Information Technology Delhi, New Delhi, India 110020
[1] anvit16135@iiitd.ac.in
[2] jainendra@iiitd.ac.in
[3] swapna@iiitd.ac.in

Palynology is the study of pollen grains and spores produced by plants. It has applications in a plethora of disciplines like geology, botany, paleontology, archeology, pedology, physical geography, and ecology. Recently, pollen source analysis is being actively used in the field of forensics, biometrics, and pharmacy. This variety of applications stems due to the fact that pollen does not easily decay and hence it can be used to determine its source of origin. Pollen can be identified to the genus or even to the species level by observing the size, shape, and surface texturing of the pollen grain. Traditionally, pollen identification and source analysis were manually done by botanical experts using microscopy. More recently, automatic, accurate, and faster techniques for pollen identification and source analysis are being developed using pollen image data. Our goal is to improve the accuracy as well as the speed of the image-based automatic pollen identification using Convolutional Neural Networks (CNN). We are in the process of building a large dataset of pollen grain images (over 10,000 images) belonging to 900 plant genera from all around the world. We plan to apply our methods for automatic classification of these images with high speed and accuracy.

Psychometric Profile Capture by Tracking Involuntary Physiological Response to Video Stimulus

Afzal Rao[1], Utkarsh Gupta[2], Anshul Rai[3], and Nitish Deshpande[4]

[1]Email: afzalrao@iitk.ac.in, Electrical Engineering, IIT Kanpur
[2]Email: utkarshg@iitk.ac.in, Electrical Engineering, IIT Kanpur
[3]Email: anshulra@iitk.ac.in, Electrical Engineering, IIT Kanpur
[4]Email: nitishvd@iitk.ac.in, Electrical Engineering, IIT Kanpur

ABSTRACT

INTRODUCTION:

Psychometric profiling identifies the extent to which a candidate's personality and cognitive abilities match those required to perform a role. Psychometric Profiling is achieved through Psychometric Tests.

Even though there are various kinds of questionnaires that are used for Psychometric Profiling, there does not exist a well established easily scalable automated process for the same.

OBJECTIVE:

The primary objective of this study was to design an automated procedure that was not privacy invasive, did not involve asking any direct questions & could do Psychometric Profiling effectively.

METHOD:

A set of classifiers was trained according to the ASCERTAIN [1]data set that was publicly available for use.

The user was exposed to stimuli in the form of movie clips. These movie clips were grouped according to their Arousal and Valence values. The user's Galvanic Skin Response and Electrocardiogram was measured simultaneously. The duration of the clips ranged from 60 to 88 seconds. The duration of the clips was set so as to ensure that the user was sufficiently emotionally affected and did not bear any fatigue.

From the physiological responses of the user, a set of features were extracted, which were then fed to a pre-trained Classifier. The classifier predicted the values of the traits according to the Big-5 Model.[2]

RESULT:

Accuracy of the classifier on various traits ranged from 70% to 80%.

CONCLUSION:

These high accuracy rates implied that the usage of an automated procedure for Psychometric profiling was actually feasible. The hardware for the same was also cheap, and could be easily mass produced for large scale usage.

Keywords: Psychometric, ASCERTAIN, Galvanic Skin Response, Electrocardiogram, Big-5 Model, Pupil Size, Blink Rate

[1]ASCERTAIN– A multimodal databa**AS**e for impli**C**it p**ER**sonali**T**y and **A**ffect recognit**I**o**N** using commercial physiological sensors.
[2]A taxonomy for personality traits. The five traits are **extraversion**, **agreeableness**, **openness**, **conscientiousness**, and **neuroticism**.

The critical rain gauge node classification in the Ganga Basin using the concept of complex networks

Shubham Tiwari[1], Sanjeev Kumar Jha[1*], Ankit Singh[1]

[1]Indian Institute of Science Education and Research Bhopal, Madhya Pradesh, India

*Corresponding Author

Email: sanjeevj@iiserb.ac.in

Abstract

Rainfall is a phenomenon marked by high variability both in space and time which makes its measurement difficult. Thus, reliable rain gauge networks are essential to provide a strong basis for the interpretation of the spatial-temporal characteristics of rainfall. High density of rain gauges is always desirable in a river basin, however rarely found. The number of rain gauges and its location in a river basin largely depends on accessibility of the area, available budget, and the research or operational objectives for which rain gauge network is being established. However, having large number of rain gauge in a river basin doesn't guarantee good database to capture spatio-temporal characteristics of rainfall. The rainfall measured from rain-gauges suffer from systematic and random errors, contains gaps (missing data) from few days to several years. A rainfall data user always asks one of the key questions: what is the optimum number of rain gauge one needs to have in a basin and where they should be located to record the spatio-temporal variability of the rainfall within the selected basin? From an application point of view, rainfall data is required at different temporal scales (3 hourly for flood forecasting, daily for streamflow prediction, monthly for reservoir operation). Hence the additional query for a data user is how the optimum number and location of rain gauges will change with the variation in temporal scale of the rainfall data.

Considering rain gauges as nodes in a complex network, quantification of node importance becomes significant. In recent years, the theory of complex networks has been used to study the spatial and temporal evolution of a broad spectrum of complex systems and associated phenomenon from diverse fields such as social networks, transportation networks, communication networks, and networks from computer science and mathematics.

In this study, we use the complex network-based Clustering Coefficient (CC) and the entropy theory-based Mutual Information (MI) to find the critical nodes in a rain gauge network. The CC is the standard metric for quantifying the extent to which the edges of a network cluster. The MI quantifies the amount of information of one random variable that is stored in another random variable. The Indian Meteorological Department (IMD) monitored-network of 692 rain gauges in the Ganga basin is considered for the analysis. Multiple rain gauge network scenario with varying Network Size (i.e. NS = 150, 300, 450 and 600) and Temporal Resolution (i.e. TR = 3 hours, 1 day and 1 month) are introduced to study the effect of network density, gauge location and temporal resolution on the network's critical nodes classification. As the amount of rainfall directly influence the streamflow output, proxy

validation of the methodology was done using a hydrological model. Our results indicate that the network density and temporal resolution strongly influence a node's importance in a rain gauge network. The proposed methodology can be useful in studies where parameter variability at a point (in comparison to its neighbourhood) quantifies its importance. The incorporation of hydrological model with node importance quantification will be highly beneficial for a hydrologist and water resource planners.

Title: Precipitation downscaling using Stacked Super-Resolution Methods

Rajib Chattopadhyay[1], Bipin Kumar[1], Manmeet Singh[1, 3], Karthik Kodari[2], Somanchi V S S N V G K Murthy[2]

[1]Indian Institute of Tropical Meteorology, Homi Bhaha Road, Pashan, Pune, 411008
[2]Defence Institute of Advance Technology, Girinagar, Pune, 411025
[3]IDP in Climate Studies, Indian Institute of Technology, Bombay

Abstract:

Indian Summer Monsoon Rainfall (ISMR) is critical to the survival of around 2 billion inhabitants living in the Indian subcontinent. Prediction of this system is thus crucial for the planning of various water requirements over the region. Traditionally, global climate models (GCMs), used for generating forecasts, are run at coarse spatial resolution underpinning the need to downscale these outputs at high resolution, to be made practically useful for the local communities. Such downscaling techniques are useful to generate district level or even city level forecast. .

Several statistical and dynamical downscaling methods are available in the literature. Machine learning-based downscaling techniques show promise in such downscaling problems. Recent developments in Single Image Super-Resolution (SR) using various Deep Learning convolutional techniques have offered promising hope in downscaling climate model outputs to high resolution. In this work, we show the use of state of the art SR techniques stacked in multiple layers applied on precipitation data over the Indian region. Results present a gateway to a new method that can potentially aid in the downscaling of ISMR and to understand extreme events at high resolutions.

Data driven discovery of nonlinear dynamics of anode-glow oscillations in a plasma device

Bhumika Thakur[1] , Neeraj Chaubey[2] and Abhijit Sen[1]

[1] *Institute for Plasma Research, HBNI, Bhat, Gandhinagar - 382 428, Gujarat, India*

[2]*Dept. of Physics & Astronomy, The University of Iowa, Iowa City, Iowa 52242, USA*

E-mail : bhumika@ipr.res.in

Abstract

Data driven techniques such as machine learning are increasingly finding applications in physical sciences and plasma physics is no exception. With a lot of data being generated from various plasma devices all over the world, some of the areas where machine learning has found applications are predicting plasma disruptions, advanced diagnostic analysis, and model extraction and reduction. The present work focuses on the third problem of data driven discovery of system dynamics. Many plasma processes are highly complex and nonlinear and often the exact form of the equations governing their dynamics is not known. If we can construct these equations from the experimental data, then we can further our understanding of these processes and use techniques such as model reduction to isolate dominant physical mechanisms. Here we take the experimental time series measurements of anode glow oscillations from a glow discharge plasma device [1] and attempt such a model construction. In the past, these oscillations were found to display a behavior that could be qualitatively modeled by a van der Pol-like equation, but the exact form of the equations governing their dynamics remained unknown. Our goal is to determine model equations that best represent the experimental data and thereby obtain some insight into the dynamics of anode glow oscillations. We investigate two approaches for the determination of such equations – symbolic regression via genetic programming [2] and sparse regression [3]. Genetic programming is a type of evolutionary algorithm that uses the principles of biological evolution to find the solution to a complex problem. Sparse regression uses the sparsity-promoting techniques and machine learning and is based on the observation that most dynamical equations are sparse in the space of possible functions. We test these methods using the numerical data from the dynamical systems whose exact governing equations are known such as Lorenz system and van der Pol equation. The performance of the two methods is compared in terms of convergence, execution speed and computational resources required. Next, we use these methods to obtain the model equations for glow discharge oscillations using the actual experimental data [1] and discuss the efficacy of the model for prediction purposes.

References:

[1] Neeraj Chaubey, S. Mukherjee, A. N. Sekar Iyengar, and A. Sen, “Synchronization between two coupled direct current glow discharge plasma sources”, Phys. Plasmas 22, 022312 (2015).

[2] John R. Koza. Genetic Programming: On the Programming of Computers by Natural Selection. MIT press, Cambridge, MA, (1992).

[3] Steven L. Brunton, Joshua L. Proctor, and J. Nathan Kutz. "Discovering governing equations from data by sparse identification of nonlinear dynamical systems." PNAS 113(15), 3932-3937 (2016).

Title: Metabolic, Functional and Evolutionary insights into the Genomes and Microbiomes using Bioinformatics and Machine Learning

Shubham Kumar Jaiswal[1], Vineet K. Sharma[1]

[1]Metagenomics and Systems Biology Laboratory, Department of Biological Sciences, Indian Institute of Science Education and Research Bhopal.

Abstract: The poster elaborates on two different works; the first part elaborates on the functional and evolutionary analysis of the Indian peacock genome, and the second part elaborates on a novel approach of predicting the metabolism of xenobiotic/drug molecules by human gut microbiota.

Part-1: The unique ornamental features and extreme sexual traits of Indian peacock (*Pavo cristatus*) have always intrigued scientists and naturalists for centuries. However, the genomic basis of these phenotypes are yet unknown. In this work, the first genome sequence (Genome Size: 1.13 GB, Protein-coding genes: 15,970) of peacock and comparative analysis of peacock with the high-quality genomes of chicken, turkey, duck, flycatcher, and zebra finch has been reported. Genes involved in early developmental pathways, including TGF-β, BMP, and Wnt signaling, which have been shown to be involved in feather patterning, bone morphogenesis, and skeletal muscle development, revealed signs of adaptive evolution and provided useful clues on the phenotypes of a peacock. Innate and adaptive immune genes involved in the complement system and T-cell response also showed signs of adaptive evolution in peacock, suggesting their possible role in building a robust immune system, which is consistent with the predictions of the Hamilton–Zuk hypothesis. This study provides novel genomic and evolutionary insights into the molecular understanding toward the phenotypic evolution of Indian peacock.

Part-2: The human gut microbiota is constituted of a diverse group of microbial species harbouring an enormous metabolic potential, which can alter the metabolism of orally administered drugs leading to individual/population-specific differences in drug responses. Considering the large heterogeneous pool of human gut bacteria and their metabolic enzymes, investigation of species-specific contribution to xenobiotic/drug metabolism by experimental studies is a challenging task. Therefore, in this work, a novel computational approach has been developed to predict the metabolic enzymes and gut bacterial species, which can potentially carry out the biotransformation of a xenobiotic/drug molecule. A substrate database was constructed for metabolic enzymes from 491 available human gut bacteria. The structural properties (fingerprints) from these substrates were extracted and used for the development of random forest models, which displayed average accuracies of up to 98.61% and 93.25% on cross-validation and blind set, respectively. After the prediction of the EC subclass, the specific metabolic enzyme (EC) is identified using a molecular similarity search. The performance was further evaluated on an independent set of FDA-approved drugs and other clinically important molecules. The approach and prediction models have been implemented as the 'DrugBug' tool for the prediction of xenobiotic/drug metabolism by metabolic enzymes of human gut microbiota.

Quantitative Assessment of Vulnerability of Agricultural sector to Climate Change in Gaya District of Bihar

Akshay Singhal, Sanjeev Kumar Jha

Department of Earth and Environmental Sciences

Indian Institute of Science Education and Research Bhopal

Abstract

A major chunk of India's economic share relies on climate sensitive sectors like agriculture which is also the primary source of livelihood for most of the population. Sporadic climatic behavior coupled with anthropogenic infringes have made agricultural dependent communities vulnerable. Assessment of vulnerability to climate change and extremes is crucial to classify regions as well as communities in order to formulate appropriate adaptation strategies and efficiently allocate resources to mitigate vulnerability. Since vulnerability comprises multidimensional components relative to local geo-climatic characteristics, its characteristics vary significantly as the size of the area of study increases. In other words, better vulnerability assessment can be done if the area of study is smaller. This study focuses on the assessment of vulnerability of 24 blocks of Gaya district in the state of Bihar by constructing a Composite Vulnerability Index (CVI). The CVI index is composed of several indictors representing domains of climatic variability and extremes, agro-ecological sensitivity, demographic resistivity and socio-economic capacity facilitating wide ranging, cross domain, relevance of the study. These indicators form the basis for the quantification of vulnerability for each block. The values of each of these indicators are normalized depending upon the kind of relationship followed by the allocation of weights. The index is categorized into three individual indices namely exposure, sensitivity and adaptive capacity whose aggregation gives shape to the CVI. All the blocks are classified into various vulnerability levels depending upon their scores. Smaller scores denote lesser vulnerability while larger scores denote higher vulnerability. Such quantification of vulnerability of blocks enable in the identification of grass-root problems which helps in planning and investment. This not only prioritizes adaptive management actions and strategies

among the vulnerable blocks, but also identifies those which could slip into higher vulnerability levels in the future.

Machine Learning based Interaction Potentials for Ionic Fluids

Anuj Kumar Sirohi*

*School of Computational and Integrative Sciences, Jawaharlal Nehru University, New Delhi 110067

anujsirohi.amu@gmail.com

In this study, we have proposed a machine learning technique based interaction potential namely Neural Network Potential (NNP) for ionic fluids in biological systems. These ionic fluids are interact through Coulombic charge-charge long-range interactions which can be described by Restricted Primitive Model (RPM). In this model Ewald Sum method has been used for calculating long-range interactions in Monte Carlo simulations. Then, the neural network is implemented using Tensor-Flow, a powerful machine learning library from Google, on system configurations consisting charge and positions of the particles in the system. The data generated through Monte Carlo simulation has been used for training, testing and validation.

Our results for NNP compare nicely with the results obtained through Ewald Sum with good accuracy(99.0%). This encourages us for further use of NNP where the atomic configurations are handled using quantum mechanics, for example by density functional theory. This technique demonstrate an entirely new way of running computer simulations, with unprecedented computational efficiency and great accuracy.

References

[1] M.P. Allen, M.P. Allen, D.J. Tildesley, and D.J. Tildesley. *Computer Simulation of Liquids*. Oxford Science Publ. Clarendon Press, 1989.

[2] J¨org Behler and Michele Parrinello. *Generalized neural-network representa- tion of high-dimensional potential-energy surfaces*. Phys. Rev. Lett., 98:146401, 2007.

Keywords: Neural Network Potential, Restricted Primitive Model, Ewald Sum.

Audio to Score Alignment: Score Following

Nitish Vikas Deshpande [1], Prof Vipul Arora[2]

[1]) Junior Undergraduate Student, Department of Electrical Engineering, IIT Kanpur, nitishvd@iitk.ac.in

[2]) Assistant Professor, Department of Electrical Engineering, IIT Kanpur

The main objective of this work is to study various techniques for the score following problem and make a comparison between the conventional methods and state of the art methods for the problem. The problem consists of 2 components: Feature extraction from audio and alignment algorithms. We have chosen the state of the art algorithms for both of the components and merged them to implement an accurate score following system which has been evaluated on Bach10 as well as Altered Bach10 dataset. Due to advancements in deep learning, autoencoder networks have been used to extract features from audio rather than taking a simple transform like STFT and CQT. The incorporation of a deep learning method over the standard feature extraction procedure highly improves the accuracy of the alignment. Moreover variants of DTW like ShapeDTW which take into account the local temporal features by comparing local subsequences instead of point wise comparison and FastDTW which reduces the time complexity have been incorporated in the final implementation. This accurate alignment system forms a base for applications like music tutoring system.

Keywords: Score Following, DTW, Deep learning, autoencoders, audio features, Bach10

BowlerShip: Analyzing bowling partnerships

Praharsh A. Nanavati

Department of Engineering Sciences, Indian Institute of Science Education & Research, Bhopal

praharsh19@iiserb.ac.in

Cricket is a sport widely played and followed throughout our country. It has millions of followers worldwide. There are various formats of the game varying from five-day long matches to 20-over matches with each format requiring a different skillset and mindset. Apart from the scores, it has a lot of interesting statistics associated with it – be it runs, batting average, strike-rate, bowling average, bowling economy, batting partnerships. Websites such as Cricbuzz, Cricinfo, CricketNext collect detailed data making it possible to analyze this data and hence extract interesting patterns.

One of the popular statistics analyzed is batting partnerships: "Which was the most successful opening pair in World Cup 2019?", "Which is the biggest partnership in Test Cricket?", etc. In their paper (1), the authors investigate the importance of partnerships in Test and one-day cricket with respect to improved performance. In cricket, batsmen bat in pairs, and the pair is known as a partnership. A partnership continues batting until one of the batsmen is dismissed. It has long been believed that synergies exist between effective partnerships such that the whole is greater than the sum of its parts. However, based on their statistical analyses, the authors conclude that synergies in opening partnerships may be considered a sporting myth.

In this paper, we investigate bowling partnerships: Is a pair of bowlers more effective when bowling alternative overs with each other than with other bowlers? A pair of bowlers constitute a *bowling pair* **if they bowl consecutive overs (alternately). To measure a bowler's performance, Croucher (2) defines the bowling index as:**

$$\textit{Bowling Index} = \textit{Bowling average} \times \textit{Bowling Strikerate}$$

where, bowling average is the the number of runs conceded by a bowler per wicket taken and bowling strikerate is the average number of balls bowled for every wicket taken. Since we are analysing one-day performances, a bowler may be considered successful if he takes wickets and/or gives away few runs. The economy of a bowler, i.e., the runs conceded per over bowled is therefore an important metric of performance in this format of the game. So, we define our bowling index as:

$$\textit{Bowling Index} = \textit{Bowling economy} \times \textit{Bowling hitrate}$$

where bowling hitrate is the number of wickets taken per over and the economy is the number of runs conceded per over.

We downloaded ball-by-ball data for all World Cup 2019 matches from (3). The data is provided in YAML format (4). The meta-data includes the details of the venue, date, toss and teams of the match. For each ball bowled, the data includes the batsman, non-striker, runs scored, extras and wickets (if any). We are using Python for the analyses.

The logic of the data analysis involves a comparison of the performance of bowlers with their partners to investigate if their performance with some *preferred* **partners (the ones with whom synergies allegedly exist) is superior to their performance with the rest. Should synergies exist, we would expect to see superior performance when bowling with preferred partners.**

Bibliography

1. Juan Valero and Tim B Swartz. An investigation of synergy between batsmen in opening partnerships. *Sri Lankan Journal of Applied Statistics*, 13:87–98, 2012.
2. John S Croucher. Player ratings in one-day cricket. In *Proceedings of the fifth Australian conference on mathematics and computers in sport*, pages 95–106. Sydney University of Technology Sydney, NSW, 2000.
3. cricsheet.org. *One-Day-Internationals*, 2009 (accessed August 4, 2019).
4. YAML. *YAML: YAML Ain't Markup Language*, 2001 (accessed August 4, 2019).

Classification of Similar Foodborne Pathogens using Deep Convolutional Neural Network for Raman Spectroscopic data.

Dipak Bhikaji Kumbhar[1], Dhanya Reghu[1], Divya Shrungar[1], Sanchita Sil[2], Siva Umapathy[1,3*]

[1] Department of Inorganic and Physical Chemistry, Indian Institute of Science, Bangalore 560012, India

[2] Defense Bioengineering and Electromedical Laboratory (DEBEL), Bangalore 560093, India

[3] Indian Institute of Science Education and Research, Bhopal,462066, India

Abstract:

Raman Spectroscopy is a vibrational technique which is widely used in identification and classification of materials and biological samples etc. The accuracy of the classification model increases when the analysis methods includes the concepts of Artificial Intelligence (AI). Most commonly used classification algorithms in Raman spectroscopy are Linear Discriminant Analysis (LDA), Principal component analysis (PCA), Support Vector Machines (SVM) etc. However, all these algorithms need feature extraction to be done manually and are limited by the accuracy, speed and performance. Use of Convolutional Neural Network (CNN) in Raman spectroscopic data analysis automates the feature extraction and improves the classification accuracy. CNN is a bottom- up hierarchical multilayer perceptron that performs pre-processing, feature extraction and classification. Because of this characteristic, CNN has already found its place in Raman spectroscopy for identification and regression and is rapidly dominating other state-of-the-art machine learning methods. We have used deep learning technique to discriminate spectral data of similar type bacteria and it showed very promising result.

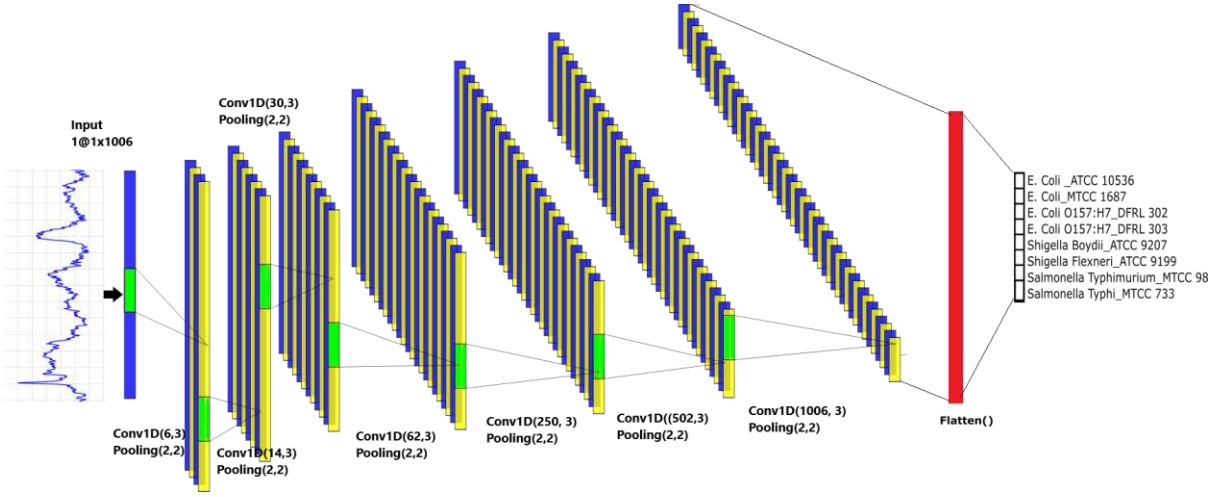

Figure: Schematic of Deep CNN with various convolution layers for feature extraction from spectral data of similar foodborne pathogens.

References:

[1] R. Gautam, S. Vanga, F. Ariese and S. Umapathy. EPJ Techniques and Instrumentation (2015) 2:8.
[2] J. Liu, M. Osadchy, L. Ashton, M. Foster, C. J. Solomon, S. J. Gibson, Analyst (2017), 142, 4067–4074.
[3] Acta Geodyn. Geomater., Vol. 8, No. 1 (161), 47–58, (2011).
[4] Yessi Jusman, Siew Cheok Ng, and Noor Azuan Abu Osman ScientificWorldJournal. (2014); 2014: 810368.
[5] A. Krizhevsky, I. Sutskever and G. E. Hinton, Advances in neural information processing systems, (2012), pp. 1097–1105.
[6] C. M. Bishop, Pattern Recognition and Machine Learning (Information Science and Statistics), (2006).

DETECTION OF TEMPORAL CHANGES IN DROUGHT. A CASE STUDY FOR MIRZAPUR (UTTAR PRADESH)

by Shivani Gond[1], Nitesh Gupta[1], S.K. Gupta[2]

[1]*PhD Research Scholar, Dept. of Civil Engineering, Indian Institute of Technology (BHU), Varanasi, 221005, Uttar Pradesh, India,*

[2]*Associate Professor, Dept. of Civil Engineering, Indian Institute of Technology (BHU), Varanasi, 221005, Uttar Pradesh, India,*

E-mail: shivanig.rs.civ16@itbhu.ac.in

ABSTRACT

The objective of the proposed study is to assess the temporal variation and forecasting drought for Mirzapur district of Uttar Pradesh. Drought indices are continuous function of rainfall, temperature and other hydro-meteorological variable commonly used to quantify the definition of drought. Drought Indices used to characterised the region were calculated by Standardized Precipitation Evapotranspiration Index (SPEI) for different time scale 1, 3, 6 ,9 and 12 monthly over period of 1951 to 2013. SPEI uses water balance as input variable which includes effect of both the Precipitation and Potential evapotranspiration. Thornthwaite method was applied for the calculation of Potential evapotranspiration (PET). Long term variability and monotonic trend of SPEI time series and the climatic variable (Precipitation, Temperature) is investigated by Non- parametric test by Mann-Kendall (MK) and Spearman's rho (SR) and the magnitude of the trend have been calculated by Sen's slope estimator. The monthly precipitation showed significant decreasing trend whereas temperature displayed increasing trend. For the assessment of vulnerability of drought due to change in climatic condition for future in Mirzapur district, a linear stochastic model i.e., Autoregressive Integrated Moving Average (ARIMA) model was developed to fit and quantitative forecasting of SPEI. The best model was selected based on Akaike Information Criteria (AIC) and Bayesian Information Criteria (BIC). The forecast results illustrate that the ARIMA model are more significantly used for short-term forecasting. Further study was made on the correlation coefficient between the actual SPEIs and the predicted ones for the forecasting. It is concluded that the ARIMA models can be used to forecast 1- 3 month leading values of all SPEI series and 6-month leading values for SPEI with time scales of 9, and 12 months. Our study testified that the ARIMA models developed for Mirzapur district can be effectively used in drought forecasting, thereby facilitating water resource managers and engineer's future to prepare contingency plans for water resource management.

***Keywords*:** Climate Change; Drought; SPEI; Trend; Mann-Kendall; ARIMA.

s

TITLE: Using Gaia DR2 Astrometry and Photomety Data to determine Distance Constraint to the North Polar Spur

AUTHORS: Kaustav Kashyap Das[1], IIT KANPUR, Alyssa Goodman, HARVARD UNIVERSITY, Catherine Zucker, HARVARD UNIVERSITY

ABSTRACT: The North Polar Spur (NPS) is one of the largest structures observed in the Milky Way in both the radio and soft x-rays. While several predictions have been made regarding the origin of the NPS, modeling the structure is difficult without precise distance constraints. Building on the work of Zucker et al. 2019[2], which determined accurate distances to local molecular clouds, we fit for distance and extinction to stars for sightlines along the southern terminus of the NPS and towards high latitudes ranging up to 53°, using optical and near-infrared photometry and Gaia DR2 astrometry. We model these per-star distance-extinction estimates as being caused by dust screens at unknown distances, which we fit for using a nested sampling algorithm. We then compare the extinction to the Spur derived from our 3D dust model with integrated independent measures from XMM-Newton X-ray absorption and HI column density measures. We find that we can account for nearly 100% of the total column density to the NPS as lying within 200 pc for latitudes > 26° and within 750 pc for latitudes < 11°. Based on the results, we conclude that the NPS is not associated with the Galactic Center or the Fermi bubbles as is commonly believed, rather it is associated with, especially at higher latitudes, with the Sco-Cen association. The robust distance constraints to the NPS obtained will be of paramount importance for modeling the evolution of the Milky Way. This is also a significant step towards coming up with a complete 3D model of the Milky Way.

[1] kaustavn@iitk.ac.in, +917318019134
[2] The Astrophysical Journal, 10.3847/1538-4357/ab2388 (2019)

Understanding drought dynamics over India

Md Saquib Saharwardi, Pankaj Kumar & Disha Sachan
Department of Earth & Environmental Sciences
Indian Institute of Science Education and Research Bhopal

Drought frequency and severity play a critical role to determine the hazard, vulnerability and associated dynamics. The frequency and intensity of drought events has increased over India and expected to increase in future as well. Hence understanding and classifying different levels of severity and drivers is very crucial for better planning and preparedness. In the present study, different drought statistics are explored using hydro-climatic parameters and drought indices. Drought variability is largely controlled by both regional and global tele-connection mechanisms that have been correlated from interannual to interdecadal scale. However, the prediction of drought based on these predictors remain a challenge using pre-established methods. For understanding the performance of climate models, several GCMs/RCMs ensembles using arithmetic mean are used to assess their efficacy in the present climate. These processes of classification, prediction and ensemble may perform better using advanced machine learning techniques like Artificial Neural Network, Support Vector Machine etc. We explored the prospects of utilizing these methods for advancement of current techniques and knowledge.

Screening of Abdominal Aortic Aneurysms Based on Machine Learning Approach

*Deepak Bajhaiya, Anju R Babu**

Department of Biotechnology and Medical Engineering, National Institute of Technology Rourkela, India-769008, Phone: 0661-2462298, e-mail: babua@nitrkl.ac.in, anjurababu@gmail.com

Abstract

An abdominal aortic aneurysm (AAA) is identified as a localized expansion of the abdominal aorta with a 50% increase in the aortic diameter [1]. This cardiovascular condition is usually asymptomatic and rupture can occur at any time without warning, making it difficult for patients to seek help and treatment. As a consequence, a robust early AAA predictor is highly relevant for minimizing the mortality rate due to the aortic wall rupture. This work aims to identify potential attributes for AAA prediction and to find a machine learning model that can be used in a clinical setting for prediction of AAA. In this study, N=424 subject's data (age and maximum aortic diameter) were collected from the works of literature (includes healthy and AAA population).70% and 30% of N=424 subject's data were used for training and testing of the model respectively. Five models named K-Nearest Neighbor (KNN), Decision Tree, Random Forest, and Logistic Regression, and Naïve Bayes were employed with cross-validation for better evaluation. For the prediction of AAA, among five models, Random Forest achieved the best classification accuracy of 99.22 %, precision of 0.98, recall of 0.98 and f1-score of 0.99 with features (age and maximum aortic diameter) incorporation. KNN achieved the 2nd best classification accuracy of 98.22 %, precision of 0.99, recall of 0.98 and f1-score of 0.98 with same features incorporation. Future work will incorporate additional attributes that will likely improve the accuracy of the models and help clinicians in their decision-making.

References:

[1] F. A. Lederle *et al.*, "Immediate repair compared with surveillance of small abdominal aortic aneurysms," *N. Engl. J. Med.*, 2002.

Role of Formal and Informal Rural Credit on Crop Yield Distributions and Downside Risk

Author(s): Sumedha Shukla, Gaurav Arora[1]

Abstract:

Farming is inherently risky given weather fluctuations during the growing season and an inconsistent revenue stream. Credit in agriculture plays an important role in risk mitigation through consumption smoothing and stabilized seasonal input use. Though agricultural risk can be categorized into production risk (i.e., impacts yields) and price risk (i.e., impacts farm incomes), we will focus on the role of credit in mitigating production risk. The three major coping instruments for production risk (excluding input subsidies) are credit, insurance and direct benefit transfers (Swain, 2014). While credit is an instrument of ex-ante income-smoothing or ex-post consumption-smoothing with respect to farm-level decisions, insurance provides ex-post consumption smoothening (Morduch, 1995).

Credit markets in developing countries are characterized by the formal sources (e.g. commercial banks) and informal sources (e.g. moneylenders), differentiated mainly on the basis of transaction costs and interest rates. Unique to the Indian scenario is the compulsory linkage between formal credit and crop insurance. This is expected to have an impact on yield risks, with the interaction between credit and insurance playing an important role. In order to assess its risk mitigation properties, we systematically estimate the role of credit in determining farm yields, particularly the yield risk, i.e., (left tail of the yield distribution).The lessons from this exercise could be useful for understanding the implications of this linkage on the effectiveness of the crop insurance in India.

We posit credit as a treatment to model agricultural productivity [1] and utilize plot-level ICRISAT-VDSA dataset on sorghum yields for 512 agricultural households, who may cultivate on one or more farms, across ten semi-arid villages in five Indian states during 2001-2014. An econometric issue arises when estimating the impact of credit on crop yield. In particular, crop yield and credit might be endogenous because on one hand credit might cause increased yields through investment in higher quality inputs and land management, while on the other hand credit access to a farmer might depend on an expectation of higher output on her plot. We employ a quasi-experimental framework to control for the endogeneity problem and estimate the impact of the first instance of agricultural productive credit on crop yields [2,3]. Specifically, we first conduct propensity score matching on the likelihood of obtaining credit conditional on household demographics (e.g., caste), land ownership status, soil quality and wealth (value of existing livestock & non-farm employment). We estimate the treatment effect of credit by comparing the average conditional yields on the *matched* households, while also controlling for inputs (e.g., fertilizer), weather, education, and distance between house and plot. Finally, in order to comment upon the effect of credit on downside yield risk (i.e., conditional estimates of left tail of the yield distribution) we estimate a quantile yield regression on the 25^{th} and 50^{th} quantiles. To our best knowledge, ours is the first attempt at a formal estimation of the impact of institutional and informal credit on crop yield risk.

[1] Sumedha Shukla is a research associate in the Econometrics Lab, IIIT-Delhi; and Gaurav Arora is an assistant professor of economics at IIIT-Delhi.

Our analysis reveals that the marginal impact of credit on sorghum yields is positive, on-average about 1/3rd of mean yields in our sample, and statistically significant. However, the difference of yield impact due to formal and informal credit is quite small. Furthermore, the positive impact of credit is largely concentrated on higher yield values and is statistically insignificant in the left-tail of the distribution for Sorghum yields. These results posit a critical question on the social utility of the expansion of formal institutions for rural credit in the past decades and its mandatory linkage with crop insurance.

Auxiliary regression for likelihood of attaining credit reveals that the formal credit access is twice as large for forward-caste households relative to the backward-castes. On the other hand, informal credit access too is skewed away from backward-castes, however the magnitude is slightly less at 1.5 times higher for the forward castes. Land ownership creates better odds of attaining informal credit than formal credit, though the difference is small. Education is found to be associated with higher credit access. Last but not the least, better soils are not found to have a statistically significant impact on credit access to rural households.

References

Feder, G., Lau, L. J., Lin, J. Y., & Luo, X. (1991). *Credit's Effect on Productivity in Chinese Agriculture: A Microeconomic Model of Disequilibrium.* Agriculture and Rural Development Department, World Bank.

ICRISAT-VDSA. (2018, November 21st). *Village Dynamics in South Asia, Micro Level Dataset*. Retrieved from http://vdsa.icrisat.ac.in/vdsa-database.aspx

Morduch, J. (1995). Income Smoothing and Consumption Smoothing. *Journal of Economic Perspectives, Volume 9, Summer*, 103-14.

Planning Comission, Government of India. (2007). *Report of the Working Group on Risk Management in Agriculture for XI Five Year Plan (2007–2012)* . New Delhi: Planning Comission, Government of India.

Swain, M. (2014). *Crop Insurance for Adaptation to Climate Change in India Working Papers (61)* . London, UK: Asia Research Centre(ARC), The London School of Economics and Political Science.

Notes:

[1] Note that credit data is recorded only once in a year, at the start of Kharif season.

[2] We confine our analysis to the *first instance* of agricultural credit, i.e. even if a household takes credit in more than one year, we only include it in our analysis for the *first year* in which it takes credit. An alternative strategy to calculate the treatment effect would be to compare the average yields for households which have taken credit to households which have never taken credit. But this strategy is not feasible given our sample where only 36 out of the 512 Sorghum-growing households have not accessed credit during the study period.

[3] A loan is categorized as *agriculturally productive* if it is spent on any agricultural input. By restricting our analysis to this category of loans, we effectively control for the utility of credit and its potential dilution of treatment effect estimates.

Automatic Segmentation of Peripheral and Transition Zone of Prostate Gland on MR Images.

Virendra Kumar[1*], Dharmesh Singh[2], Chandan J Das[3], Anup Singh[2,4], and Amit Mehndiratta[2,4]

[1]Department of NMR, All India Institute of Medical Sciences, Ansari Nagar, New Delhi
[2]Centre for Biomedical Engineering, Indian Institute of Technology, Hauz Khas, New Delhi.
[3]Department of Radiology, All India Institute of Medical Sciences Delhi, Ansari Nagar, New Delhi.
[4]Department of Biomedical Engineering, All India Institute of Medical Sciences, Ansari Nagar, New Delhi.
*Corresponding Author: vrndra@gmail.com, virendrakumar@aiims.edu

Magnetic resonance imaging (MRI) have shown potential role in investigation of prostate cancer (PCa). The zonal anatomy of prostate gland, peripheral and transition zone are best depicted on MR images, compared to other imaging modalities like transrectal ultrasound (TRUS). Normal peripheral zone (PZ) appears hyperintense and transition zone (TZ) appears hypointense on T2 weighted MRI while PCa lesion appears hypointense compared to PZ and TZ. The Prostate Imaging Reporting and Data System version 2 (PI-RADS v2) is a standardized scoring system for reporting of MRI of the prostate. Automatic segmentation of the prostate zonal anatomy and subsequently, prostate cancer lesion on MRI images will help objective and reproducible reporting. This study evaluates the two novel algorithms for prostate segmentation and compares their accuracy, 1) Semi-automatic active contour based segmentation approach. 2) Fully automated Simple Linear Iterative Clustering (SLICS) based superpixel algorithm.

The study was carried out on MRI data of 14 patients of biopsy proven PCa. All MR images were acquired at 3T MRI scanner. Axial T2W images were used for identification of PCa lesion and DWI (b value=0 s/mm^2) were used for segmentation. The proposed method has been improvised over the active contour model developed by Chan and Vese, which is a semi-automatic algorithm. An appropriate rectangular mask for the prostate gland based on the shape information was selected by the user. After that region based Chen Vese active contour model applied to the shape prior and refine the surface of the prostate with different iterations. The morphological opening operator to improve the prostate segmentation result were applied. 2) SLICS Superpixel: In a pre-processing step, contrast-enhancement, normalization, morphological erosion operation, and median-filtering was performed on each slice. Simple Linear Iterative Clustering (SLIC) method is adopted to segment each slice in order to obtain superpixels. Each segmented superpixel was replaced with their mean intensity. Five-level thresholding and binarization were done for the generation of the mask of prostate-gland and multiplied binary-mask with original images.

Segmentation of PZ and TZ was carried out using an atlas based approach similar to sequential registration-based segmentation. Manual segmentation was performed by a radiologist for comparison of automated segmentation results. Performance of the segmentation algorithm was evaluated using Dice-coefficient (DC) and Jaccard-index (JI). The DC and JI for active contour segmentation of prostate-gland were 87.56 $\pm$ 5.20% and 78.46$\pm$7.89% respectively, whereas for superpixel (SLICS) based approach were 87.71$\pm$6.41% and 77.32$\pm$8.50% respectively, indicating a high similarity between segmentation obtained by using proposed algorithms and the manual segmentation.

Automatic segmentation of prostate and its zonal anatomy on MR images will eventually help to develop automatic artificial intelligence based automatic detection and segmentation of prostate cancer lesion. This may help in automatic objective and reproducible reporting of prostate nodules. This may find application in targeted biopsy or therapy of the detected nodule.

References

1. Hoeks CM, Barentsz JO, Hambrock T, et al. Radiology 2011;261:46–66.
2. Weinreb JC, Barentsz JO, Choyke PL, et al. Eur Urol 2016;69(1):16-40.
3. Kim CK, Park BK, Kim B, et al. AJR Am J Roentgenol. 2010;194:1461–1469.
4. Chan TF and Vese LA. IEEE Trans Image Proc 2001; 10(2):266-277.
5. Khalvati F, Salmanpour A, Rahnamayan S, et al. J Dig Imaging, 2016; 29(2):254–263.
6. Zhang J, Baig S, Wong A, et al. J Comput Vis Imag Syst 2016;2(1).

Bias correction and spatial downscaling based post-processor for improving medium range forecasts of the Indian summer monsoon rainfall

Nibedita Samal, Sanjeev Kumar Jha

Department of Earth and Environmental Science

Indian Institute of Science Education and Research, Bhopal, India

Abstract

Flood causes devastation in many states of India every year. An accurate warning of the flood event in few days in advance may save thousands of lives and crores worth of properties. In India, agencies like Indian Institute of Tropical Meteorology (IITM) and National Center for Medium Range Weather Forecast (NCMRWF) have taken initiatives to provide weather forecast for 5 to 15 days lead time. The raw forecast obtained from a Numerical Weather Prediction (NWP) model are known to have bias due to approximate simulation of atmospheric processes and errors in defining initial conditions. In this study, our objective is to improve the NWP forecast by correcting the bias and downscaling it spatially to the scale of our interest. Two daily precipitation datasets are used for the summer-monsoon period over Indian state of Madhya Pradesh region, namely, High Resolution Ensemble System (HRES) and Indian meteorological data (IMD) at 50 km and 25 km spatial resolution respectively. The goal is to bias correct and spatially downscale HRES data using the IMD data. We use 12 years of summer monsoon data (June, July, August and September) from 2007 to 2018. For downscaling, two approaches, namely bias corrected spatial downscaling (BCSD) and statistical downscaling and bias correction (SDBC) are used to bring the HRES 50 km data to 25 km. In BCSD, the HRES data is bias corrected using the quantile-quantile mapping method, then downscaled by multiplying a transfer function to the data interpolated to finer resolution. In SDBC, the HRES data is interpolated to finer scale and multiplied with the transfer function to downscale the data. This downscaled data is then bias corrected using quantile-quantile mapping method. The analysis is carried out by using first 11years data and the forecasts are generated for the 12^{th} year, which is assessed using the 12^{th} year IMD observation data. The error estimation of BCSD and SDBC outputs with IMD data shows that SDBC approach gives better results. The next objective is to develop a machine learning approach to improve any given future forecasts with the help of transfer functions already developed using past data. A comparison between the traditional and

machine learning approaches will be observed to find out the better approach for downscaling precipitation.

Making of an unbiased taxonomist: A machine learning approach for exploring morphological variations to delimit species

Preeti Saryan[1], Shubham Gupta[2], Vinita Gowda[1]
[1] Dept. of Biological Sciences, Indian Institute of Science Education and Research Bhopal, India
[2] Dept. of Computer Science and Automation, Indian Institute of Science, Bengaluru, India

Classical taxonomy deals with assessing the morphological variations and assigning the species identity to an individual. However, due to high morphological similarities sometimes a group of closely related species may have ambiguous species boundaries and such a group is identified as a "species complex". Generally, taxonomists use classification based methods to delimit species where groups are predefined (supervised analysis) which adds human bias to the analysis. Taxonomists also use unsupervised methods like ordination methods (such as principal component analysis, and non-metric multidimensional scaling) either as visualization tools or as feature extraction tools for supervised learning algorithms like linear discriminant analysis. Hence, a robust method for unsupervised discovery of clusters is required which not only identifies all the possible clusters but also a) can test the quality of cluster, and b) can deduce the defining morphological features for each of those clusters. Here, we present two case studies where we use spectral clustering along with character analysis to delimit species complexes in the genus *Hedychium* (Zingiberaceae). We propose that spectral clustering algorithm can be used to find different clusters because unlike simple *k*-means clustering, spectral clustering algorithm has the capability of discovering arbitrarily shaped clusters even when standard kernels (like RBF kernel) are used.

The genus *Hedychium* is known to have several species complexes and thus has challenged taxonomists for decades. We measured 150 morphological characters, both vegetative and reproductive, for each taxa (n=5 to 20) from multiple populations belonging to two major complexes (the Spicatum-complex and the Coronarium-complex). We identified seven groups in the Spicatum-complex (5 known species, one variety, and one unidentified) and three groups in the Coronarium-complex (3 known species). Spearman correlation test was performed to identify and remove the highly correlated traits. Spectral clustering suggests 5, 9 and 12 clusters. Character analysis (mutual information between characters and clusters, correlation and t-test) identified at least four characters that define the clusters: lateral staminode ratio, labellum ratio, notch and notch length to labellum length ratio. We also found that for one of the species clusters, ecological characters like the number of flowers opening per day was important suggesting the importance of ecology in delimiting species. Together with our proposed character analysis methods, spectral clustering enabled the unsupervised discovery of species boundaries along with an explanation of their biological significance.

Prediction of Properties of Self-compacted Concrete Containing Fly Ash Aggregate using Machine Learning Tools

Naveena M.P[a],G.Narayana[b], Vijayalakshmi Akella[c]

[a]*Research Scholar, Department of Civil Engineering, SJCIT Chickballapur /Assistant Professor, Department of Civil Engineering, KSSEM, (Bengaluru-560109, India)*

[b]*Professor & H.O.D, Department of Civil Engineering, SJC Institute of Technology, (Chickballapur-562101, India)*

[c]*Professor & H.O.D, Department of Civil Engineering, K.S. School of Engineering and Management, (Bengaluru-560109, India)*

Abstract

The Artificial Fly Ash aggregates (AFA) are made from Fly ash and water and can be used as a replacement of natural aggregates in concrete production. On the other hand, the continuous use of natural aggregates causes environmental problems and it is also necessary to preserve the natural sources. The use of sustainable technologies such as the development of Self-compacting concrete (SCC) is an innovative type of conventional concrete, which can be placed and compacted under its self-weight with little or no vibration effort.

The Present study focused on the Prediction on the influence of Artificial Fly Ash aggregates properties on fresh and hardened and durability Property of self-compacting concrete as substitutes to natural Coarse aggregate. The Prediction of a Self-compacting concrete property with artificial aggregate and natural aggregate was investigated by using the MATLAB. Design of experiment was performed two factors with three levels. These factors were the Quantity of Cement and percentage replacement of Coarse aggregate in the concrete mix. The results indicate that there was an increase in the Flow ability and Durability characteristic with increasing Cement content and Increase in the strength with decreasing the size of Artificial Fly Ash aggregates (AFA).

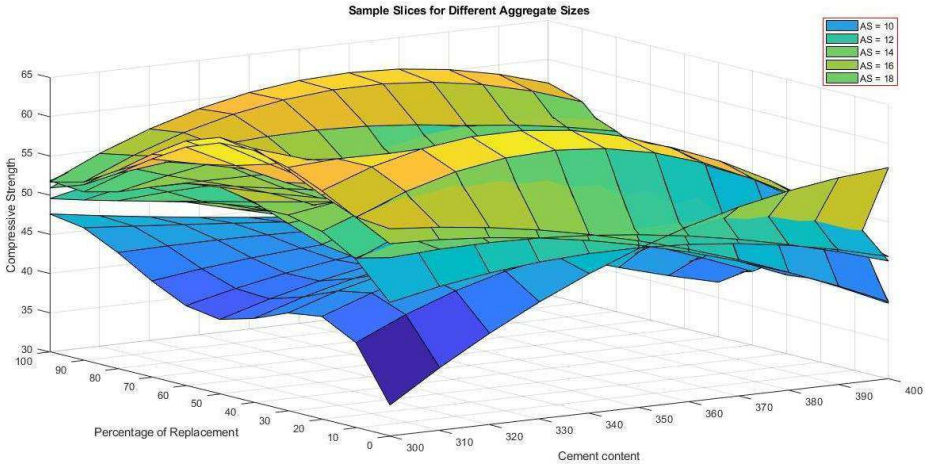

FIGURE 1. Surface plot

REFERENCES

1. Shivaraj et.al, "Prediction of Compressive, Flexural and Splitting Tensile Strengths of Concrete using Machine Learning Tools", International Journal of Engineering Research & Technology ISSN: 2278-0181 Vol. 4 Issue 05, May-2015.

2. Palika Chopra, et.al, "Prediction of Compressive Strength of Concrete Using Artificial Neural Network and Genetic Programming", Advances in Materials Science and Engineering, Volume 2016, Article ID 7648467, 10 pages.

3. Mahmoud Abu Yaman, et.al, "Predicting the ingredients of self-compacting concrete using artificial neural network", Alexandria Engineering Journal, Volume 56, Issue 4, December 2017, Pages 523-532

4. Dhanya Sathyan et.al,"Modeling the Fresh and Hardened Stage Properties of Self-Compacting Concrete using Random Kitchen Sink Algorithm" International Journal of Concrete Structures and Materials, December 2018

Climate Change Impacts on Irrigation Water Requirements of Wheat Using CROPWAT Model in Agro-Climatic Zones of India

Swatantra Kumar Dubey*, Pankaj Kumar and Md. Saquib Saharwardi

Department of Earth and Environmental Sciences
Indian Institute of Science Education and Research Bhopal, Bhopal, India

Corresponding Author*: swatantrakd@iiserb.ac.in

Abstract

Agriculture automation is the key concern and emerging area for India. Increasing population rate leading to stress on food demand briskly. Agricultural sector soil moisture monitoring is required for irrigation scheduling, water resource allocation and management. India is extremely diverse in its climate, soils structure, geological, physiography, cropping patterns and extensive agricultural planning. Planning Commission (Niti Aayog) of India divided India into 15 Agro-climatic regions. This study covers 14 Agro-climatic regions to determine the water requirement of wheat crop period (1979–2015). The Food and Agriculture Organization (FAO) CROPWAT 8.0 model has been used in these zones to find the crop water requirements (CWRs) and irrigation schedules of wheat crop. CROPWAT 8.0 model computed the past CWRs using Penman-Monteith method. The study result showed that ET_0 of the zones varied from 3.51 mm/day (Western Himalayan) to 5.63 mm/day (Western dry regions) and the effective rainfall varied from 336.8 mm (Western dry regions) to 1298.7 mm (Eastern Himalayan). The irrigation requirement of the wheat crop varies in all the Agro-climatic regions, and the maximum irrigation requirement shows in the Gujarat plains and hills followed by Western dry regions and minimum water requirement in Western Himalayan. This study demonstrated that the CROPWAT model could simulate the irrigation requirement needs in the changing climate for the wheat crop to the better management of the irrigation schedule. The Artificial Neural Network (ANN) may be used to predict the CWRs and adopted to estimate irrigation requirement, evapotranspiration (ET_0) and precipitation for the future.

Keywords: Agriculture, Crop water requirement, ET_0, Artificial Neural Network

Preliminary Detection of Pulmonary Diseases

Soumya Ranjan Dash[1], Jay Mundra[1], Utkarsh Gupta[1], Archit Bansal[1], Vinayak Agrawal[1], Jaya Srivastava[2], Kushangi Mittal[2], Netravat Pendsey[3], Prashant Raiger[4]

[1]) Electrical Engineering, IIT Kanpur
[2]) Mechanincal Engineering, IIT Kanpur
[3]) Department of Physics, IIT Kanpur
[4]) Aerospace Engineering, IIT Kanpur

Most of the fatalities caused by pulmonary disorders such as COPD, asthma, bronchiolitis etc. could be prevented by timely diagnosis of the patients. Combined with the extreme lack of medical practitioners and infrastructure in rural areas, the lack of screening programs, which would provide such timely diagnosis, has made pulmonary diseases the second most deadliest kind in India. To make such screening programs more practical and feasible, we propose an automated procedure for the screening of the rural populations, using an easy-to-use device that can be operated even by a community worker. As a proof of concept, we have developed a novel device to record auscultation sounds and the breathing rate of the patient. On this data, a machine learning model with Keras framework and Tensorflow backend is implemented to detect abnormalities. A 3 dense layer and a dropout layer convolutional neural network is used to identify Wheezes and Crackles in an audio file which is fed as Mel-Spectrograms as inputs. Data augmentation was employed in the form of audio stretching as well as Vocal Tract Length Perturbation, especially for the scarcer 'wheeze' and 'wheeze and crackles' classes. Overall validation accuracy currently stands at roughly 80 percent. On the user interface side, the mobile application displays an immediate report of the patient displaying the diseases which are suggested by the symptoms observed. Such automation using an inexpensive device removes the dependency on doctors and expensive equipment, which makes the screening programs impractical for large scale implementation today.

Keywords: Convolutional Neural Network, Tensorflow, Crackles, Wheezes, Mel-Spectrogram

Estimation of Cheeger Constant for Regular Graphs

Ambar Jain[1,*], Kashyap Rajeevsarathy[2], and Shivam Pal[1]

[1] *Department of Physics, Indian Institute of Science Education and Research, Bhopal*
[2] *Department of Mathematics, Indian Institute of Science Education and Research, Bhopal*
[*] *Corresponding author: ambarj@iiserb.ac.in*

Abstract

There are numerous known applications of graph theory in computer science, linguistics, social sciences, physics, chemistry etc. In computer science many problems such as networking, data mining, scheduling, optimization etc are solved using graphs. In statistical physics graphs are used in percolation theory. Graphs are also applied in quantum physics for perfect transfer of qubit states in the network of spin-1/2 particles.

Here, we are interested in a particular graph invariant called the Cheeger constant (or Isoperimetric Constant), which is the best-known measure of the connectivity of a graph. In other words, networks modeled using graphs with higher Cheeger constants have better connectivity and reliability. However, the computation of Cheeger constant of an arbitrary graph involves the enumeration of the subsets of its vertex set. Consequently, as a graph increases in size, the number of subsets grows exponentially, which makes the calculation of Cheeger constant computationally complex. In fact, this is known to be an NP-hard problem. We will describe an efficient procedure of estimating the Cheeger constant of an arbitrary graph using its spectrum by applying machine learning techniques.

Machine Learning for Multi-Dimensional Integral Estimation

The problem of evaluating integrals of arbitrary functions over specified domains is common in high energy physics. A frequent use case is the calculation of integrals in Feynman diagrams, which in turn, motivates the development of algorithms for multidimensional numerical integration. Algorithms for multidimensional numerical integration are required to be efficient at the same time should yield high precision results. Existing algorithms are typically variations of Monte Carlo integration with importance sampling. We explore new Machine Learning based Monte Carlo techniques using Generative Deep Neural Networks for multi-dimensional integrals. We also explore alternate machine learning technique that approximates multi-dimensional functions as sum of products of one-dimensional functions, followed by standard integration techniques and study their scalability to multiple dimensions. In particular, we show that the machine learning based techniques can be used to exhibit substantial improvement in the computational efficiency of multi-dimensional integral estimation.

Predicting seasonal temperature over India using advanced data science techniques

Aditya Kumar Dubey, Aaquib Javed, Amita Kumari, and Pankaj Kumar

Department of Earth and Environmental Sciences
Indian Institute of Science Education and Research Bhopal

Warm temperature events such as heatwave, droughts, forest fires, etc. have witnessed an increase in their return period frequencies and intensities under the warming climate in recent decades. Heatwave during summer cause human discomfort, socio-economic damages, excess energy use, water scarcity, and often leading to loss of human life. A seasonal temperature prediction/forecast can be used for better preparedness and livelihood security by providing advance information with sufficient lead time for better decisions to minimize the adverse effect of temperature extremes. Predicting the seasonal temperatures over a small region is challenging due to a lack of understanding of the small scale physical process. Different approaches used in the seasonal forecast/prediction over India are dynamical and statistical downscaling. In the present study, prediction of summer (March-May) temperature carried out using a statistical approach canonical correlation analysis (CCA). Here, we elucidate the relationships between various meteorological parameters from the preceding winter (December-February) and the following summer (March-May). All data is detrended to show a purely predictive relationship between summer temperature and other winter variables. El Niño Southern Oscillation (ENSO), along with zonal circulation and pressure systems, play a significant role in determining summer temperature over the Indian region. The skill of all the results is quite significant compared to other studies. The other such techniques can be further explored for robust and reliable predictions.

Automating the health assessment of Honeybee colony with deep convolutional neural networks

Madhav Thakker[1], Saket Anand[1], Swapna R. Purandare[1], Trisha Datta[1]

[1)] IIIT-Delhi New-Delhi, India
{*madhav16159,anands,swapna,trisha18017*}*@iiitd.ac.in*

Honey bees play an important role in the production of food by pollinating crop plants. The western honey bee, *Apis mellifera* is a domesticated species of the honey bee, commonly maintained by beekeepers for their pollination services as well as for the production of honey. Since the past few years, a decline in *Apis mellifera* honey bees has been reported from many parts of the world. For example, in 2018, 60% of the *Apis mellifera* honey bee colonies in the United States have vanished. This decline in honey bees has an adverse impact on pollination and yield of many crops such as nuts, oilseeds, fruits, and vegetables. Therefore, it is important to assess and monitor the health of the honey bee colonies. The total number of adult worker bees in a beehive is considered a key indicator of beehive health. Our goal is to develop a generalized method for automatic counting of bees, which can be used for beehive images having different variations ranging from bee species (affecting texture and color), lighting conditions, and the size of the bees. The main challenges involved are generalizing the automatic counting and detection performance well across different domains (bee species and beehive imaging conditions), and significant occlusion caused by the density of bees in the beehive images. We shot videos of *Apis mellifera* bees on beehive frames and extracted 895 images from them. We then annotated 50 frames for testing our trained model. We used an annotated OIST dataset with 8,640 images containing a total of 375,698 labeled bees as a source domain and our own *Apis mellifera* dataset having 896 images without annotations as the target domain. We evaluated three different models on our dataset and analyzed the results obtained. Firstly, we trained a **Mask-RCNN** on the *Apis mellifera* Dataset and got an mAP of **0.7212**. Secondly, we also tried Domain Adaptation Techniques like **Domain-Adaptive Faster-RCNN** which resulted in an mAP of **0.124** on the target domain. Finally, we tried **Strong-Weak Distribution Alignment for Adaptive Object Detection** and got an mAP of **0.194**. To cater for the scale change between the source and the target domain images (a global transformation affecting all objects and backgrounds), we trained a weak-global alignment model by using focal-loss to focus on hard-to-classify examples and got an mAP of **0.309**. The results demonstrate that a weak-alignment model performs better than other techniques on our dataset with a big margin.

Distinguishing Quantum Many-Body Phases in Disordered Systems using Deep Neural Network

Nilanjan Roy* and Anshum Manchanda

Department of Physics, Indian Institute of Science Education and Research Bhopal, India

*Email: *nilanjan16@iiserb.ac.in*

Disorder and interactions are ubiquitous in nature. Certain isolated interacting many-body quantum systems undergo a crossover from the ergodic thermalized phase to the non-ergodic many-body localized (MBL) phase, as the disorder in potential is tuned from lower to higher values . The MBL phase has been realised in the experiments based on ultra-cold atoms, ions and materials with nitrogen-vacant (NV) centers. In our work, we carry out a systematic study of the localization properties of interacting fermions in the quasi-periodic potential in one dimension using a deep neural network based approach. We analyse the energy levelspacings and inverse participation ratio of the eigenstates by supervised learning techniques to establish a crossover from the thermalized phase at low disorder to the MBL phase at high disorder. Our neural network gives further insights into the crossover region, which cannot be extracted from the many-body systems using other numerical methods, e.g., the exact diagonalization techniques. In the end, we provide a comparative study between the phase diagram obtained from our neural-network based method and the same obtained via the exact diagonalization.

Measuring adaptation action to climate change using climate smart quotient (CSQ)

Climate change threatens the productions and productivity of crops. The changes in temperature and precipitation are expected to shift the agricultural seasons, pest and disease patterns and alter the production, prices, incomes and ultimately, livelihoods of rural poor. Such changes, in turn, are posing a challenge in preserving and enhancing food security, especially in developing countries, including India, which necessitates the farming system to be more resilient to climate change. Lot of policy action is done by the governments to build the capacity of the rural poor for climate change. However, in absence of data driven decision making evidence on effectiveness of interventions either remain unmeasured or result in inadequate policy directives. The concept of Climate Smart Village is being implemented under the National Adaptation Fund by Environment Planning and Coordination Organization Govt. of MP. (EPCO)

In view of evidence driven policy needs, IIFM Bhopal has partnered with EPCO to build Monitoring and evaluation framework for various intervention in 60 villages in 3 districts. A unified index termed as Climate Smart Quotient (CSQ) is designed to quantitatively assess the program effectiveness in the study villages. The index has process and outcome components. The set of indicators that comprises the components of CSQ are grouped in the following heads of Weather-smartness, Water-smartness, Carbon and Seed-smartness, Nitrogen-smartness, Knowledge-smartness, Energy-smartness, and Economic-smartness in the two cropping seasons of Kharif and Rabi. All the villages have reported positive Climate Smartness for the process indicators; while Quotient for outcome indicators vary across villages. This is due to effect of climate adversities. The impact of project effectiveness is measured using statistical double difference methodology.

The concept of unified Climate Smart Quotient enables a data driven decision support system for measuring the dynamics of climate smartness in view of program action. The data analysis for one year of intervention resulted in policy directives for next phase of action and strategy for information gathering on scaled up version of the project.

Transmission connectivity link status estimation

with partial measurement using Graphical learning method

Author: S.Naresh Ram *Supervisor* : Prof. Sukumar Mishra

Abstract

State estimation is a method in which states of a system are being determined by using Weighted Least Square method. In power system, the success of estimated measurements majorly depends on accuracy of network topology (breaker status). Due to inclement weather conditions, failure of Energy Management System (EMS) data link of major stations results in loss of complete data (Topology and line measurement data) at control centre. This results in non-convergence of available state estimation method.

The proposed method estimates the missing data and recreates the network topology with partial available measurements using Graph learning techniques. The given method has also been tested on IEEE-14 bus system data.