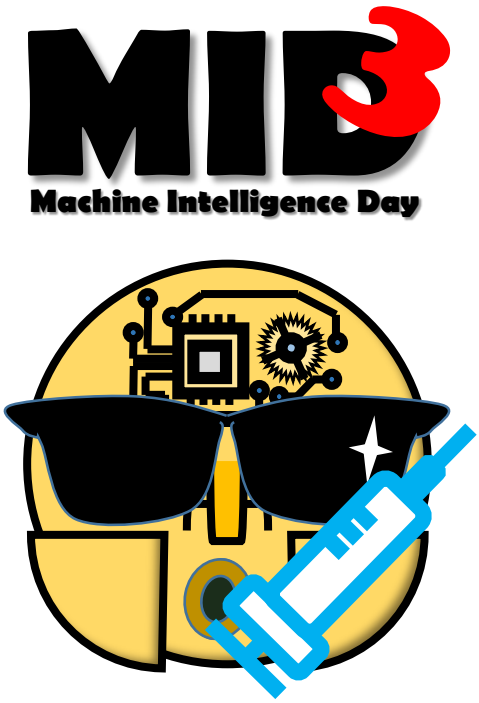
Proceedings of the third

Machine Intelligence Day

December 10 2021



Hosted by
Seidenberg School of Computer Science and Information Systems, Pace University
New York, New York

edited by Sung-Hyuk Cha
D. Paul Benjamin

Proceedings of the 3rd Machine Intelligence Day

Editors:

Sung-Hyuk Cha	Pace University
D. Paul Benjamin	Pace University

Program Committee:

Teryn Cha	Essex County College
Soon Ae Chun	CUNY Staten Island
James Geller	NJIT
Yegin Genc	Pace University
Anthony Joseph	Pace University
Sukun Li	Adelphi University
Damian M. Lyons	Fordham University
Francis Parisi	Pace University
Juan Shan	Pace University
Lixin Tao	Pace University

Hosted by:

The Seidenberg School of Computer Science and Information Systems

Published by Pace University Press

Available online
http://csis.pace.edu/∼scha/MID2021

Table of Contents

Preface

We are very pleased to have the opportunity to organize the third Machine Intelligence Day 2021. Machine Intelligence Day is an annual New York based conference hosted by Seidenberg School of Computer Science and Information Systems at Pace University. It occupies a unique place among conferences, presenting both new research and exceptional student papers, providing opportunities for both faculty and student participation. The purpose of Machine Intelligence Days is to provide a learning and sharing experience on recent developments in Artificial Intelligence, Computer Vision, Data Mining, Machine Learning, and Pattern Recognition. The conference is welcoming to a range of participants, open to both researchers in the field and students. While experts give talks, they are targeted at audiences in general computer science with an eye dedicated towards students. We have strived to publish well-written abstracts that present important original research results and/or open problems relevant to Machine Intelligence.

Two faculty members of Seidenberg school delivered the invited talks: Professors Yegin Genc and D. Paul Benjamin. Mr. Armen Pischdotchian, an academic tech mentor from IBM Watson delivered an invited talk and a workshop on "Buiding and deploying models using AutoAI." We are grateful to them. We also would like to express the gratitude to the IBM University Guest Lecture Program which made this guest lecture possible.

Machine Intelligence Day 2021 is held both in-person and online. There were both recorded video and poster presentation competitions. We received over 24 abstracts and 18 were selected in the Proceedings. 16 participated the recorded video presentation competition and 11 participated the poster presentation competition. We would like to express our gratitude to all the contributors and participants. Finally, we hope that you will benefit from this conference and its proceedings.

S.-H. Cha and D. P. Benjamin

Explainable AI and Human Cognitive Biases

Yegin Genc
Seidenberg School of Computer Science and Information Systems, Pace University NY
ygenc@pace.edu

When complex AI systems supplement decisions that are play an essential role in humans lives (as in e.g. medicine, law or defense), understanding how complex models furnish the AI predictions become even more critical. Therefore, there has been a push from the community to increase the focus on exploring the integration of explainable AI (XAI) implementations when human and AI agents work together. In this talk, I will survey the current trends in Explainable AI. I will also report on a recent study that explores the effects of AI explanations on human decisions in the context of a cognitive bias that gravitates human decisions towards presented AI recommendations.

An Overview of Work in the Robotics Lab

D. Paul Benjamin
Department of Computer Science, Pace University New York, NY
dbenjamin@pace.edu

This short talk will give an overview of research underway in the Pace Robotics Lab. We will describe our new robots and their capabilities and demonstrate how they are used in simulation and the real world. This will be followed by a brief overview of some of our research projects, together with the difficulties they face.

Mobile Manipulator

Jackal

Figure 1: MID logo.

References

[1] http://csis.pace.edu/robotlab

Cross-roads of AI and Data Science

Armen Pischdotchian
IBM Watson, 550 King Street, Littleton MA 01730, USA
apischdo@us.ibm.com

Perhaps a better way to think of AI is that of prediction machines. To gain insights from what it is that you don't know that you don't know. And so the modern Data Scientist initiates that journey, where data is not just stored in databases or CSV files, rather, it is unstructured in nature and resides on our mobile devices; all that has to do with vision and language. Let's explore the cross roads of machine learning and data science.

The laborabory workshop on "Building and deploying models using AutoAI" will be given followed by the invited lecture. In this lab you build models that predict the risk factor of individuals not being able to pay back a bank loan and further discussions regarding the fairness of the prediction with regard to bias mitigation potentially stemming from the algorithm or the data set.

References

[1] https://www.ibm.com/blogs/ibm-training/author/apischdous-ibm-com/

Boston House Prices - Regression Predictive Modeling Machine Learning Problem from End-to-end Python

Pratikkumar Chaudhari
Computer Science Department, Pace University, New York, NY, USA
pratikkumar.b.chaudhari@pace.edu

In this paper, I have taken the dataset of Boston having the various factors like crime rate in town, size of the house etc. The aim of this paper is to predict the housing prices in Boston considering the above factors. I use linear regression, random forest regression and decision tree regression which is part of machine learning algorithm to predict housing prices in Boston by performing data visualization using scatter plot considering the crime rate in the area nearby, number of rooms in the house against the housing prices. I have performed some of the calculations like Mean Squared Error (MSE), Mean Absolute Error (MAE) and R^2 score to predict the housing prices. From the calculations, we can see that the value of R^2 for Linear Regression is 0.7640047, for Decision Tree Regression is 0.79829676 and that for Random Forest Regression is 0.8806006 which is higher among all three. Based on the calculation, maximum the value of R^2, better is the regression technique to predict the housing price. So, here I have used some of the Python and Data Science libraries like NumPy for numerical calculations and Pandas for data processing.

So, according to the data visualization, more the crime rate in the town, lesser the price of the house and more number of rooms per dwelling, more is the price of the house. And more the value of R^2, more is the accuracy of the predictive analysis. So, Random Forest Regression is the best suitable algorithm for the predictive analysis of the housing prices.

References

[1] Karim, R., Alam, M. K., and Hossain, M. R., Stock Market Analysis Using Linear Regression and Decision Tree Regression, in Proc. of *the 1st International Conference on Emerging Smart Technologies and Applications*, pp 1-6, 2021

[2] Vovk, V., Nouretdinov, I., and Gammerman A., On-Line Predictive Linear Regression, *the Annals of Statistics*, Vol. 37, No. 3, pp. 1566-1590, 2009

[3] Kurniawati, N., Putri, D. N., and Ningsih, Y.. Random Forest Regression for Predicting Metamaterial Antenna Parameters. in Proc. of *the 2nd International Conference on Industrial Electrical and Electronics*, pp 174-178, 2020

Recommendation System for Book Readers

Gaurav A. Haral
Computer Science Department, Pace University, New York, NY, USA
gauravharalgh@gmail.com

Several machine learning approaches are used in a recommendation system. Finding commonalities between different Book recommendations is one of the most important functions of a recommendation system today for the book readers. This recommendation model seeks out books that are like or better than those youve read or liked in the past. The dataset for this work was obtained from Kaggle, and the purpose is to create models capable of offering correct suggestions to online Book shoppers. During the development of this system, techniques such as Collaborative (Cosine Similarity rule), Nearest Neighbors, and Content Based methods were employed.

Figure 1: Working Of Recommendation System

References

[1] Venkatesan, R. and Sabari, A., Issues in various recommender system in e-commerce *Journal of Critical Reviews*, vol. 7 no. 7, pp. 604-608, 2020

[2] Alamdari, P. M., Navimipour, N. J., Hosseinzadeh, M., Safaei, A. A. and Darwesh, A., A systematic study on the recommender systems in the E-commerce. IEEE Access, 8, pp.115694-115716, 2020

Setting prices using Van Westendorp Price Sensitivity Meter with Trial and Revenue Curves

Yelim Na
Peter J. Tobin College of Business, St. John's University, New York, NY
yelim.na21@my.stjohns.edu

Market research is essential when entering an existing market. Data-driven pricing gives marketers the flexibility of evaluating expected performance at different price points, making specific product launch plans tailored to market conditions. In addition, it is possible to reduce the company's loss due to failure by predicting the success of the market entry before launching the product. Moreover, reasonable pricing can lead to profits but also customer satisfaction.

The preference and price perception of 439 consumers for a fictional new pizza were collected by survey. The survey results were made into an SPSS file and analyzed for consumers' price perceptions using the Van Westendorp Price Sensitivity Meter(VWPSM) with Trial and Revenue Curves using SPSS and Excel.

Through VWPSM, a determination could be made regarding the possible acceptable price points by determining the following:

- Range of Acceptable Prices (RAP)
- Point of Marginal Cheapness (PMC)
- Point of Marginal Expensiveness (PME)
- Indifference Price Point (IDP)
- Optimum Price Point (OPP)

Using additional Trial and Revenue Curves, it is possible to identify the price that maximizes market share, and the price that maximizes revenue upon entering the market; these price points typically differ.

The significance of this project is that it demonstrates how to identify the range of prices that consumers can accept using the VWPSM. In addition, also incorporating trial and revenue modeling allows managers to determine the optimal price point and its impact on consumer demand.

References

[1] Risen, E., and Larry R.. The use of intent scale translations to predict purchase. *Biotrak* White Paper, 2008

[2] Weiner, J. L., and Barry Z., Pricing New-to-Market Technologies: An Evaluation of Applied Pricing Research Techniques, *Ipsos-Insight* White Paper, 2004

[3] Whitlark, D. B., Geurts M. D., and Swenson, M. J. New product forecasting with a purchase intention survey. *Journal of Business Forecasting Methods & Systems*, vol. 12, no. 3, pp 18-21, 1993

Sentiment Analysis for Reviews

Yashraj V. Bhandare
Computer Science Department, Pace University, New York, NY, USA
yashraj.bhandare@pace.edu

The purpose of the social media platform is to create more opportunities for people to express their beliefs in public, as long as they are hired to present an idea and become a major problem. Sentiment Analysis is a matter of natural language processing that may determine people's attitudes about any particular product through analysis. Sentiment Analysis is the process of automatically releasing features in the form of others' opinions about a particular product, service or information.

The available data was separated in positive reviews and negative reviews, some additional data was kept unprocessed in order to use unsupervised learning methods. Multiple approaches were used on the data such as changes in data pre-processing, different word embeddings such as BoW, tf-idf, Word2Vec were used. Additionally, to investigate accuracy removal of stop words was also done. Classification models such as SVM, Nave Bayes used.

The best results were acquired with Naïve bayes and linear regression. Further exploration with polarity of words for that different clusters are made.

References

[1] Hsu, C.-W., Chang, C.-C., and Lin, C.-J., A Practical Guide to Support Vector Classification, Technical report, Department of Computer Science, National Taiwan University, 2003

[2] Shimodaira, H., Text classification using naive bayes, Learning and Data Note 7 , pp. 1-9, 2014

[3] Ramkumar, A. S. and Nethravathy, R., Text document clustering using k-means algorithm, Int. Res. J. Eng. Technol, 6, pp.1164-1168, 2019

Modeling Climate Change Through DNN and LSTM

Ray Jennings, Marwan Shouery, and Kaleemunnisa
Computer Science Department, Pace University, New York, NY, USA
{rj07609p, ms96172p,klnu}@pace.edu

Reputable data sources have shown that multiple factors are increasing both the air and the ocean temperatures which are contributing to significant climate change. Many of these factors are attributed to population- specifically overall population size, urban population size, educational level, life expectancy, poverty rate and population density. Additional attributes for which there is data include land usage types such as agriculture, farming, and forestry as well as energy consumption of both renewable and non-renewable sources. Furthermore, greenhouse gas emissions, which may contain CO2, methane, and nitrous oxide are also contributing factors. These factors are all believed to play a significant role in climate change. Through this analysis, we demonstrate how all these attributes have direct correlations to the increase in the global temperature which is a primary contributing factor to climate change. The data sources are Our World In Data [1] and The World Bank Group [2]. The data sets are labeled by year, attribute, and country. Deep Neural Network (DNN) and Long Short-Term Memory Network (LSTM) models were built using Keras to model the dataset features. Furthermore, prediction visualization was used to model climate change based on the various data attributes within the primary countries of concern.

References

[1] https://ourworldindata.org

[2] https://www.worldbank.org

Music Genre Classification Using Capsule Network

Austin Puthenpurackal Blaise[1] and Vinodh Ewards [2]
[1]Computer Science Department, Pace University, New York, NY, USA
austin.puthen@pace.edu
[2]Department of Computer Science Technology, Karunya University
ewards@karunya.edu

Aimed to solve the music genre classification problem using the raw audio input and modified Convolutional Neural network. We have used multiple dataset including the GTZAN and Free Music Archive Dataset. The music features are generated using LibROSA and used by state-of-the art algorithms to predict the genre. The model proposed by Geoffery E. Hinton [**?**], the Capsule Network is primarily concentrated in this paper and its accuracy to classify the genre. The neural network model of Capsule Network is implement to produce a state of the art classification model using the new approach of capsules rather than neurons. The paper also compare Hintons model with other state of the art Neural Models including LSTM, CNN, CNNLSTM and other Machine Learning models including Random Forest, SVC, KNN classifiers.

References

[1] Sabour S., Frosst, Ni. and Hinton, G. E., Dynamic Routing Between Capsules, in Proc. of the 31st Conference on Neural Information Processing Systems, Long Beach, CA, 2017

On Dichotomized ANNs for Multi-Class Problem

Akankshya Biswal
Computer Science Department, Pace University, New York, NY, USA
ab94615n@pace.edu

Physicians used Cardiotocography (CTG) to knowing of fetal well-being and potential complications from pregnant women. They used a continuous electronic record of the babys heart rate took from the mothers abdomen. They visualized the unhealthiness that will give an opportunity for early intervention. CTG class status is classified in this paper with machine learning methods by using attributes of data obtained from the uterine contraction (UC) and fetal heart rate (FHR) signals and visualized the acquired information. This classification and visualization will help the doctor while treatment the patient. Experimental results have shown good accuracy scores and low error rate.

References

[1] Chen, X., Ye, Y., Xu, X., and Huang, J. Z., A feature group weighting method for subspace clustering of high-dimensional data, *Pattern Recognition.* vol 45, Issue 1, pp 434-446, 2012

[2] Cömert, Z. and Kocamaz, A. F., A Study of Artificial Neural Network Training Algorithms for Classification of Cardiotocography Signals, *Bitlis Eren University Journal of Science and Technology.* vol 7, n. 2, pp 93-103, 2017

[3] Nidhal, S., Ali, M. M., and Najah, H. A novel cardiotocography fetal heart rate baseline estimation algorithm, *Scientific Research and Essays.* vol 5, pp4002-4010, 2011

CysPred: A Proposed Machine Learning Strategy to Predict Cystoid Macular Edema

Rachel Vardi[1] and Rebecca Y. D'Agostino[2]
[1]Edgemont Junior Senior High School, New York, NY, USA
[2]Computer Science Department, Pace University, New York, NY, USA
rvardi05@gmail.com, rdagostino2@pace.edu

Cystoid Macular Edema (CME) is a disease affecting the central retina of the eye. A gathering of cyst-like (cystoid) spots develops in the macula, causing swelling in the retina. This swelling can lead to visual impairment, or even complete blindness. Using optical coherence tomography (OCT), cross sectional images of the retina can be captured and used for study. We hypothesize that an aggregate of prognostic imaging biomarkers learned by an algorithm will yield a better predictive visual acuity outcome than any factor in isolation.

To start, we acquired data from a private database, including monthly OCT images of patients with CME from screening time to at least six months after, and specific measurements and details of each eye. Such measurements included the volume and thickness of different sectors of the retina, and the presence of different characteristics like cystoid spaces or fluid. Our goal is to determine what factors cause visual acuity to increase (and retinal thickness to decrease) over time. Decision Trees were used as a machine learning technique instead of typical linear regression to discover the interaction effects that linear regression often misses. An example of one said tree is shown below, correlating the center point thickness as a feature, and the presence of intraretinal fluid (cannot grade, absent, questionable, definite) as the class (Figure 1).

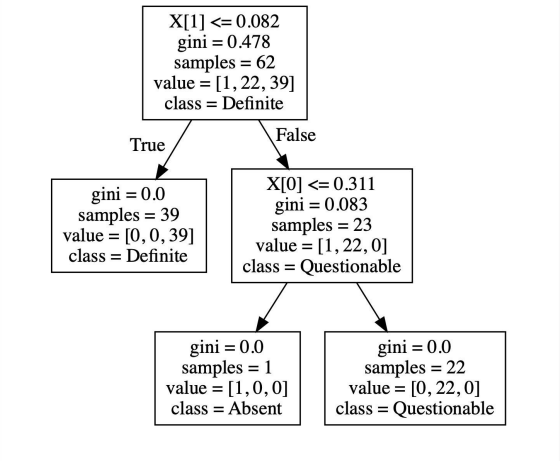


Figure 1: A decision tree splitting different patients center point thickness based upon whether intraretinal fluid is present in the retina

These results as of now are indicative of only one potential predicting characteristic, and more work will need to be done to find other correlations. This will heavily rely on making use of visual acuity scores measuring the progression of the patients sight accuracy throughout the months. The goal is to be able to have a fully predictive and accurate machine learning model that can pinpoint specific problematic areas in patients retina in early stages of CME, in order for doctors to prescribe the best treatment plan.

Senegalese Fashion Clothing Classification System by Deep Learning and Convolutional Neural Networks

Christelle Scharff[1], Adja Codou SECK[2], Kaleemunnisa[1], and Krishna Bathula[1]
[1]Department of Computer Science, Pace University, New York City, NY, USA
{cscharff, klnu, kbathula}@pace.edu
[2]Alioune Diop University of Bambey, Senegal, West Africa
seckadjacodou@gmail.com

Fashion is deeply clinched in and influenced by sociocultural and aesthetical practices and both local and international trends. Current Artificial Intelligence applications to fashion have mainly focused on Western fashion. Our study is inspired by the Fashion-MNIST dataset, used to benchmark machine learning algorithms that recognize 9 Western fashion apparels. It is an attempt to address the lack of data from the Global South and to propose a study case that relates to African culture, authenticity, living and heritage. African fashion is a 31 billion industry, the second largest economic sector after agriculture, that is now in the global spotlight. This paper presents a model capable of classifying various Senegalese fashion items by using transfer learning for image classification with MobileNetV2 convolutional model as the base model. Using this approach and building on the model used for Fashion-MNIST greatly reduces the enormous time required to train the AI model with increased accuracy. The model for Senegalese Fashion classification is built from an image dataset of two categories of dresses (grand boubou and taille mame) that we have gathered ourselves (with the consent of local Senegalese domain experts). Although the images acquired are not huge samples that a typical AI model expects, they are expanded with data augmentation techniques such as rotating, cropping, flipping, and resizing. More inclusive automated fashion classification processes are required to categorize, organize, identify, and advertise apparel for e-commerce purposes, and to avoid rampant piracy.

Generalized Optimal Proximity Measure

Lewis E. Westfall and Sung-Hyuk Cha
Computer Science Department, Pace University, New York, NY, USA
{lwestfall,scha}@pace.edu

One pervasive method for classifying an unknown input vector involves finding the most or top k similar instances in the reference set. This problem, so-called k-nearest neighbor or simply kNN problem has received a great attention in pattern recognition for its simplicity and practical value. While Euclidean L_2 or Manhattan L_1 distances are widely used in the instance based classifiers, they are rooted from the generalized Minkowski distance in eqn (1).

$$L_p(X,Y) = \left(\sum_{i=1}^{d} |x_i - y_i|^p \right)^{\frac{1}{p}} \tag{1}$$

Experimental results show that the optimal results are not conventional Euclidean or Manhattan distances and p value varies depending on datasets.

When the instances are represented as binary vectors, there are various ways to generalize proximity measures [2]. One such proximity measure is called the generalized jaccard similarity coefficient.

$$\mathrm{GJS}(X,Y) = \frac{pa}{pa+b+c} = \frac{pXY^t}{pXY^t + X\bar{Y}^t + \bar{X}Y^t} = \frac{p\sum_{i=1}^{n} x_i y_i}{p\sum_{i=1}^{n} x_i y_i + \sum_{i=1}^{n} x_i(1-y_i) + \sum_{i=1}^{n}(1-x_i)y_i} \tag{2}$$

It generalizes famous conventional ones such as Jaccard, Dice, and Sokal & Sneath similarity coefficients when $p = 1$, 2, and 0.5, respectively. However, the optimal results are not conventional ones and p value varies depending on datasets. Moreover, interesting observations are found depending on the types of binary datasets.

Finally, weighted proximity measures are suggested using a genetic algorithm. While weighted Euclidean distance or weighted Hamming distance are pervasively used, improved weighted proximity measures are given.

References

[1] Dasarathy, B. V., Nearest Neighbor Pattern Classification Techniques, *IEEE Computer Society Press*, New York, 1991.

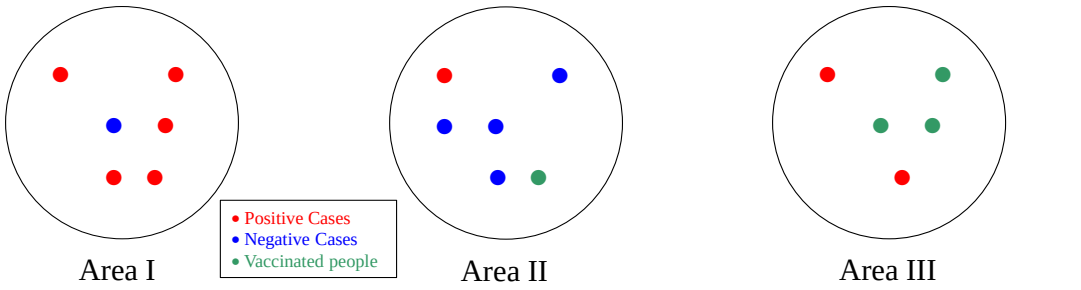
[2] Anomaly between Jaccard and Tanimoto Coefficients Cha, S.-H., Choi, v, and Tappert, C. C., Pace CSIS Research Day May 2009

Purity Measure for Cluster Analysis of the COVID-19 affected areas

Hetal Mistry
Computer Science Department, Pace University, New York, NY, USA
hm79869p@pace.edu

With the increase in Covid-19 cases, it is important to determine different strategies and the best possible solutions to provide support for any highly infected areas. Some studies present a novel analysis that results in clustering countries with respect to the active cases, active cases per population, and active cases per area based on Johns Hopkins epidemiological data [1]. This proposal analyzes the purity measure that could be useful to the medical, health, and government sector to plan any important preventive measures needed to be taken in that specific area. This analysis will also help us to determine the critical condition of the areas infected with COVID-19. The purity level per area will be measured by using the total number of positive cases, negative cases, and total vaccinated people with respect to that areas population.

$$\text{purity}(w_i) = \frac{1}{n_i} \max_{j \in C}(n_{ij}) \tag{1}$$



Area I: Purity = 1/6 (max(5, 1, 0))
= 5/6
= 0.83
The purity measure of the COVID-19 positive cases is high.

Area II: Purity = 1/6 (max(1, 4, 1))
= 4/6
= 0.67
The purity measure of the COVID-19 negative cases is high.

Area III: Purity = 1/5 (max(2, 0, 3))
= 3/5
= 0.60
The purity measure of the vaccinated people is high.

References

[1] Zarikas, V., Poulopoulos, S. G., Gareiou, Z., and Zervas, E., Clustering analysis of countries using the COVID-19 cases dataset, *Data in Brief*, vol 31, p 105787, 2020

[2] Yang, H. Y., Xu, J., Li, Y., Liang, X., Jin, Y. F., Chen, S. Y., Zhang, R. G., Zhang, W. D., and Duan, G. C. The preliminary analysis on the characteristics of the cluster for the COVID-19, *Zhonghua Liu Xing Bing Xue Za Zhi*, vol. 41 no. 5, pp. 623-62, 2020

[3] Afzal, A., Ansari, Z., Alshahrani, S., Raj, A. K., Kuruniyan, M. S., Saleel, C. A., and Nisar, K. S., Clustering of COVID-19 data for knowledge discovery using c-means and fuzzy c-means, *Results in Physics*, vol 29, 2021

[4] Vahabi, N., Salehi, M., Duarte, J. D., Mollalo, A., and Michailidis G., County-level longitudinal clustering of COVID-19 mortality to incidence ratio in the United States, *Sci Rep* 11, 3088, 2021

Regression for Categorical Feature Data

Vemulapalli Sai Pranav
Computer Science Department, Pace University, New York, NY, USA
saipranav.vemulapalli@pace.edu

Objective is to obtain the relationship between the dependent variable and the independent variable [1]. Linear regression was the approach to find the relationship. Twenty percent of the dataset was allocated as testing dataset and eighty percentage for the training dataset. The prediction which got after fitting the model was sixty five percent accurate (good score) which showed that as the age (x) increases, price (y) for the insurance increases accordingly.

Dataset had the outliers, which afterwards were removed using the Quantiles concept. Quantiles were used to set the maximum threshold and minimum threshold which are the lower and upper limits in a dataset. Outside these limits lie outliers. After the outlier removal normalization was performed on dataset which gave the improvement in the accuracy score for different random states (different random state mean different datapoints for the respective random state) used.

Linear Regression for the categorical data was also performed which gave bad score meaning Linear Regression cannot be done on the categorical data. Instead, it can be done with KNN Regression using Hamming distance. Future Objective is to perform KNN Regression on categorical data and achieve the better results.

References

[1] Maulud, D. H. and Abdulazeez, A. M., A Review on Linear Regression Comprehensive in Machine Learning, *Journal of Applied Science and Technology Trends*, Vol. 1 No. 4, 2020

[2] Kanyongo, G. Y., Certo, J., and Launcelot, B. I., Using regression analysis to establish the relationship between home environment and reading achievement: A case of Zimbabwe, *International Education Journal*, vol. 7 no. 5, pp. 632-641, 2006

Trimmed Weighted k-Nearest Neighbors

Jonathan I. Lee,
Department of Computer Science, Pace University, New York, NY, USA
jlee12@pace.edu

The nearest neighbor (NN), k-nearest neighbors (kNN) [4], and weighted k-nearest neighbors (wkNN) [3] algorithms comprise a family of pattern classifiers can be extended to serve more complex classifiers. This study proposes a modification to wkNN, called trimmed, weighted k-nearest neighbors (twkNN) to handle potential under-performance issues and improve its accuracy, especially where kNN outperforms it.

After defining a value t, the threshold of noise that the twkNN classifier will tolerate, the algorithm proceeds as normal under wkNN, with the exception that the furthest t values of each classification type are 'trimmed' away before the weighted vote calculation in wkNN. The intention is to maintain the strengths of wkNN's concern with proximity, while avoiding its sensitivity to noise, or outliers, close to the query point.

This study was run with several canonical classification datasets, to include irises, wine and heart disease. To further strain the classifiers, varying degrees of noise are applied to the datasets, along with varying selections for k.

Contrary to expectations, the twkNN modification fared best without the added noise, outperforming NN, kNN and wkNN, while its performance decreased as arbitrary noise was introduced to the datasets. The paper will further describe the difficulties encountered here, and future directions of study: variable t values, and ensemble methods in the NN family of classifiers.

References

[1] Bishop, Christopher M. *Pattern Recognition and Machine Learning*. New York: Springer, 2006.

[2] Duda, Richard O. and Peter E. Hart. "Pattern Classification and Scene Analysis," 1974.

[3] Dudani, S. A., "The Distance-Weighted k-Nearest-Neighbor Rule," in *IEEE Transactions on Systems, Man, and Cybernetics*, vol. SMC-6, no. 4, pp. 325-327, April 1976, doi: 10.1109/TSMC.1976.5408784.

[4] G. Gates, "The reduced nearest neighbor rule (Corresp.)," in IEEE Transactions on Information Theory, vol. 18, no. 3, pp. 431-433, May 1972, doi: 10.1109/TIT.1972.1054809.

[5] Zhang, Zhongheng. "Introduction to machine learning: k-nearest neighbors." *Annals of translational medicine*, vol. 4,11 (2016): 218. doi:10.21037/atm.2016.03.37

Hybrid Quantum Classical Neural Network

Avery Leider[1], Gio Abou Jaoude[1], Abigail Strobel[2] and Siddh Agarwal[3],
[1]Seidenberg School of Computer Science and Information Systems, Pace University NY
[2]Dyson School of Arts and Sciences, Pace University, New York, NY
[3]Hicksville High School, Hicksville, NY
{al43110n, ga97026n, as66693n}@pace.edu, sba17@protonmail.com

The Hybrid Quantum Classical Neural Network (HQCNN) is an ambitious combination of classical neural networks and quantum circuits[1]. The goal of the HQCNN is a machine learning model that utilizes the scalable nature of classical networks with the promising learning capabilities of variational quantum circuits. The model will consist of a quantum-classical pair for every class in the standard MNIST data set[2]. Convolution of the images is done with the use of quanvolutional kernels courtesy of PennyLane API[3]. Improvements are made to the weights and parameters that make up the model, through a combination of the parameter-shift rule for quantum parameters and standard Jacobian matrices for classical weights.

References

[1] G. Abou Jaoude *Demonstration on Jupyter Notebook on Google Colab*, https://colab.research.google.com/drive/1ioMnFGvfKeZDJHZzq431vlC5p5Twufc8?usp=sharing

[2] Y. LeCun, C. Cortes, and C. J.C. Burges, *The MNIST Dataset Of Handwritten Digits*, https://yann.lecun.com/exdb/mnist/

[3] *Pennylane API*, https://pennylane.ai/

Quantum Machine Learning Classifier

Avery Leider[1], Gio Abou Jaoude[1] and Abigail Strobel[2],
[1]Seidenberg School of Computer Science and Information Systems, Pace University NY
[2]Dyson School of Arts and Sciences, Pace University, New York, NY
{al43110n, ga97026n, as66693n}@pace.edu

This Quantum Machine Learning Classifier (QMLC), uses the mathematics of quantum computing in a deep neural network[1] to find and classify the specific flower type of the three different iris flower species: Versicolor, Setosa and Virginica, utilizing the SciKit-Learn dataset "Iris."[2] In that dataset, there are four characteristic features of each iris type: petal length, petal width, sepal length, and sepal width. This quantum computing machine learning classifier is an extension of earlier work by Cappelletti et al[3]. This quantum computing machine learning classifier outperformed the classical computing deep learning neural network method. Most significant is that this classifier trained in fewer epochs.

Figure 1: Versicolor, Setosa and Virginica[2]

References

[1] G. Abou Jaoude *Demonstration on Jupyter Notebook on Google Colab*, https://colab.research.google.com/drive/1YMPn7LTtfd73WpWc4XmGUsAxHruD0WBf?usp=sharing

[2] E. Anderson, *Iris Dataset*, https://scikit-learn.org/stable/auto_examples/datasets/plot_iris_dataset.html

[3] W. Cappelletti, R. Erbanni and J. Keller *Polyadic Quantum Classifier*, https://arxiv.org/pdf/2007.14044.pdf

The Effect of Two Qubit Errors on Quantum Computing Accuracy

Lewis E. Westfall
Computer Science Department, Pace University, New York, NY, USA
lwestfall@pace.edu

Quantum computers give hope that they will be able to solve computational problems that conventional computers cannot. They will do this by concepts not applicable to conventional computers such as entanglement, superposition, and not being limited to the states 0 and 1. Entanglement is the ability to connect two or more qubits in such a way that an operation on one qubit affects not only the operational qubit but also its entangled twin. Two proven uses of entanglement are superdense coding and quantum teleportation. This document describes a quantum computing architecture-based de-coherence model developed on the available IBM Q-Experience quantum computers. By using entangled qubits in superdense coding, it is possible to explore how the error rate between two qubits and the individual qubit error rate affects the accuracy of the outcome.

References

[1] International Business Machines: IBM Q-Experience. https://quantum-computing.ibm.com Accessed: 2021-06-29

[2] Westfall, L., Leider, A.: Superdense coding step by step. In: Future of Information and Communication Conference, pp. 357372. Springer, 2019

Index of Participants

A - Authors CC - Conference Chairs PC - Program Committee LA - Local Arrangement