International Journal of
Electrical and Computer System Design
ISSN: 2582-8134 www.ijecsd.com

Linear approximation Computational Complexity with Multivariate Correlation

Manjula R[1] , Logesh T[2]

[*1]Students, Department of computer science and application,Periyar maniammai institute of science and technology,Thanjavur ,India,manjuravi704@gmail.com

[2]Assistant Professor, Department of computer science and application.Periyar maniammai institute of science and technology,Thanjavur,India,logeshwr@gmail.com

Abstract: In recent years, the term "carbon footprint" has become common among meteorologists due to the pressing issue of climate change being a top priority for both corporations and politics. As a result, there is a high demand for calculating carbon footprints, with scholars proposing various approaches ranging from simple online calculations to more complex techniques such as input-output-based analysis. However, despite the widespread use of the term, there is no universally accepted academic definition of "carbon footprint." This lack of clarity is evident in the scientific literature, despite numerous studies on energy and ecological economics that should have defined the term precisely. Typically, carbon footprints are calculated by averaging the number of people living in a specific region with the total number of people in that region. However, we propose a system that will calculate the emissions from homes and industries separately, addressing the current system's drawbacks, which average out the footprints in a given region. Our system will use calculations that predict accurate results for each individual.

Keywords: Sparse regression, Sparse coding, Decision tree. Carbon emission, Multivariate correlation

1. Introduction

The calculation involved in evaluating emissions at each stage is complex, whereas calculating the carbon footprint of an individual or industrial area is relatively straightforward. The emissions occur during various stages such as the assembly plant, the creation of machinery, transportation of component parts, factories producing components, and mining. To estimate the greenhouse gases emitted to support one's lifestyle, calculators ask questions about household fuel use and travel patterns. Decision Trees are a type of supervised machine learning that segments training data based on a specific parameter. The algorithm used in Decision Trees is more efficient than existing systems of sparse regression. The output is efficient, time-saving, and accurate. By implementing this method, it is possible to calculate the carbon footprint of a particular region and identify emitters with a significant impact on the environment. Scholars have raised concerns about the carbon footprint, and the algorithm used for Decision Trees is a solution to this problem.

2. Categorization and Description of Works

The paper presents an algorithm for performing linear regression on high-dimensional, sparse data. The algorithm can identify the subset of relevant variables and compute an approximate solution to the regression problem. The paper presents theoretical results and experimental evidence of the algorithm's effectiveness. [1]. The paper proposes a new method for fitting additive models that are both sparse and accurate. The proposed method combines L1 regularization and shrinkage techniques to identify and eliminate irrelevant variables from the model. The authors provide experimental evidence to demonstrate the effectiveness of their method on various datasets. The paper was presented at the conference and its proceedings were published in the conference proceedings. [2]. The paper proposes a dimensionality reduction technique called Joint Ordinal-Cardinal Embedding (JOCE) for multiple ordinal regression problems. The JOCE framework integrates ordinal and cardinal information of data.

The paper proposes a feature selection method based on a local kernel regression score for high-dimensional data to address the curse of dimensionality problem. The method uses a kernel function to estimate the local regression score for each feature, and features with higher scores are considered more important and selected for further analysis. Experimental results show that the proposed method can achieve higher classification accuracy while using fewer features compared to other state-of-the-art methods. [4]. The article "Thirteen ways to look at the correlation coefficient" by J. L. Rodgers and W. A. Nice wander provides a detailed overview of the correlation coefficient and its properties, as well as offering thirteen different perspectives on how to interpret and use it in statistical analysis. The article covers topics such as definition and calculation, linear regression, outliers, hypothesis testing, limitations, and more, making it a valuable resource for researchers and practitioners in statistics. [5]. The article by S.M. Stigler explores the history of the invention of correlation, particularly the role played by Francis Galton. Stigler argues that Galton's account of the invention of correlation is not entirely accurate and that Galton's early approach to correlation had limitations that modern approaches do not have. The article sheds light on the contributions of early statisticians like Galton and Pearson and the ongoing development of statistical theory and methods. [6].

The paper "Sparse regression via range counting" proposes a new algorithm called the Range Counting algorithm, which efficiently counts the number of points within a given range using a data structure called the Range Tree. The authors demonstrate how this algorithm can be used to solve the sparse linear regression problem and provide theoretical guarantees for its convergence. The paper also includes experiments on synthetic and real-world datasets, showing that their approach outperforms several state-of-the-art methods for sparse regression in terms of accuracy and speed. [7]. The paper proposes a new deep learning model called Stacked Broad Learning System (SBLS), which extends the Broad Learning System (BLS) framework by stacking multiple layers to form a deep architecture. The SBLS model retains the advantages of BLS such as fast training and good generalization performance while achieving competitive performance compared to state-of-the- art deep learning models. The paper also introduces a new incremental flattening method to convert the deep SBLS model into a flattened form for online learning and model compression. The article proposes a method for improving image classification algorithms by diversifying sparse representations using a regularization technique on augmented data. The method outperforms state-of-the-art algorithms in terms of accuracy and robustness to data perturbations. [9]. The paper proposes a Laplacian regularized nonnegative representation (LRNR) method for clustering and dimensionality reduction based on nonnegative matrix factorization (NMF) algorithm. The LRNR method includes a Laplacian regularization term to preserve local data structure and has shown promising results on various datasets. The paper was published in IEEE Transactions on Circuits and Systems for Video Technology in January 2021. [10]

3 Performance Analysis of the Proposed Methodology in terms of Existing and proposed approach

The fundamental difficulty with sparse regression in the current system is subset selection, which often has a high processing cost. Recently, a number of enhanced approximation methods for subset selection have been presented. However, less attention has been paid to the non-approximate method of subset selection, which is important for many data analysis issues. To improve the accuracy of anticipated CO2 emissions, nonlinear data can be used. Therefore, the algorithm currently in use, Sparse regression, predicts the desired data from previously trained data. This allows us to produce the desired result in the quickest and most effective way possible. Our recommendation of a sparse and trustworthy regression technique for predictive modelling is the algorithm's key feature. It finds the most significant anomalies. We demonstrate its performance on simulated data and a real-world application. In this paper, we introduce a regression analysis method dubbed "sparse shooting S" that considers these traits typical of large data sets. The most important predictors are chosen because of the sparseness of the resulting regression coefficients, many of which are zero. One of the method's distinctive features is its robustness against outliers in the data matrix's cells. The selection of resilient variables is shown through simulation study.

As a means of accelerating the process and forecasting precise outcomes in the assessment of carbon footprint for homes and businesses, our suggested algorithms are effective. Consequently, the system we have is designed to identify precise emissions for homes and businesses in a given area. Similar to how individuals make decisions, the Decision Tree supervised machine learning algorithm bases its decisions on a set of guiding principles. It is possible to imagine a machine learning classification algorithm as a decision- making tool. Typically, the model is claimed to predict the class of the novel, previously unobserved input, but in truth, the algorithm must decide which class to assign. When planning your upcoming trip, you adopt a rule-based strategy. You might opt for a different destination depending on how long you intend to stay on vacation, your spending limit, and whether or not your extended family will be visiting you. The decision is affected by the responses to these questions. Furthermore, this will be effective in giving us the precise information we needed. Whether it is a residential user or an industrial user, we have predicted the precise emission outers in the neighbor hood with this method.

4. Methodology and Results

The decision tree algorithm is a commonly used machine learning method for categorization and prediction assignments. It works by recursively partitioning the data into subsets based on the values of the features, and then creating a tree structure where each node represents a decision based on a feature value. The tree starts with a single node called the root, which represents the entire dataset. At each internal node of the tree, the algorithm selects a feature to split the data into two or more subsets. The goal of the split is to maximize the information gain, which measures the reduction in entropy (or increase in purity) of the resulting subsets. The entropy is a measure of the randomness or uncertainty in the data, and the goal of the algorithm is to reduce it as much as possible with each split. Once a split is made, the algorithm creates a new child node for each subset and repeats the process recursively until a stopping criterion is met. For example, the algorithm might stop when a certain maximum depth is reached or when the number of data points in a node falls below a certain threshold. To make a prediction for a new instance, the algorithm traverses the tree from the root to a leaf node, following the decision path based on the feature values of the instance. The prediction is then based on the majority class (in classification) or the mean value (in regression) of the training instances that belong to the same leaf node. Decision trees offer various benefits such as their capacity to interpret, handle categorical and continuous features, and scale up or down. However, they are also prone to over fitting and can be sensitive to small changes in the data. To mitigate these issues, various techniques such as pruning, ensemble methods, and random forests can be used.

Fig 1. User sign in form

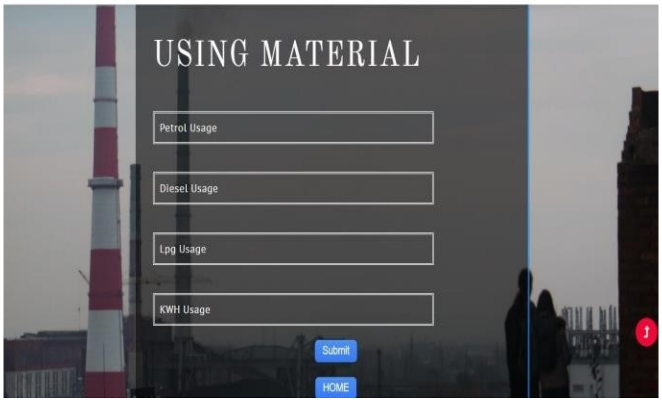

Fig 2. Output Results

5 Conclusion

In summary, the current system for measuring carbon footprint is inefficient in terms of time and accuracy. To address this issue, we have proposed a new system that incorporates an algorithm to address the core issue of sparse regression. Specifically, we have explored the Decision Tree algorithm, a supervised learning technique that can be used for both regression and classification tasks. Decision trees enable the division of datasets into trees based on a set of rules and criteria, with nodes such as root nodes, leaf nodes, sub-trees, splitting, and pruning. We have discussed the benefits and drawbacks of decision trees and highlighted that our proposed system can provide a more precise estimation of carbon emissions for homes and industries, avoiding the flawed approach of averaging and dividing equally. By identifying regions with high emissions, our system can better target the areas that require the most attention in reducing carbon emissions.

References

1. S. Har-Peled, P. Indyk, and S. Mahabadi, “Approximate sparse linear regression,” Sep. 2016. [Online]. Available: arXiv:1609.08739
2. G. Liu, H. Chen, and H. Huang, “Sparse shrunk additive models,” inProc. Int. Conf. Mach. Learn., 2020, pp. 6194–6204.
3. J. Zeng, Y. Liu, B. Leng, Z. Xiong, and Y.-M. Cheung, “Dimensionality reduction in multiple ordinal regression,” IEEE Trans. Neural Netw.Learn. Syst., vol. 29, no. 9, pp. 4088–4101, Sep. 2018.
4. Y.-M. Cheung and H. Zeng, “Local kernel regression score for selecting features of high-dimensional data,” IEEE Trans. Knowl. Data Eng., vol. 21, no. 12, pp. 1798–1802, Dec. 2009.

5. J.L.Rodgers and W. A. Nicewander, “Thirteen ways to look at thecorrelation coefficient,”Amer. Stat., vol. 42, no. 1, pp. 59–66, 1988.
6. S. M. Stigler, “Francis galton¨s account of the invention of correlation,”Stat. Sci., vol. 4, no. 2,pp. 73–79, 1989.
7. J. Cardinal and A. Ooms, “Sparse regression via range counting,”Aug. 2019.[Online].Available:arXiv:1908.00 31
8. Z. Liu, C. P. Chen, S. Feng, Q. Feng, and T. Zhang, “Stacked broadlearning system: From incremental flatted structure to deep model,”IEEE Trans. Syst., Man, Cybern., Syst., vol. 51, no. 1, pp. 209–222,Jan. 2021.
9. S. Zeng, B. Zhang, J. Gou, and Y. Xu, “Regularization on augmenteddata to diversify sparse representation for robust image classification,” IEEE Trans. Cybern., early access, Oct. 21, 2020,doi:10.1109/TCYB.2020.30257 57.
10. Y.-P. Zhao, L. Chen, and C. L. P. Chen, “Laplacian regularized non negative representation for clustering and dimensionality reduction,” IEEE Trans. Circuits Syst. Video Technol., vol. 31, no. 1, pp. 1–14, Jan. 2021.