ISSN: 0974-5823 Vol. 7 No. 4 April, 2022

International Journal of Mechanical Engineering

A CONTEXT-AWARE AND SELF CONSISTENT SOLUTION FOR HIGH DIMENSIONAL FEATURE EXTRACTION PROBLEM FROM HOUSING PRICE PREDICTION DATASET

S Vishnudarshini, Saureesh Narayan Roy , Shubham Kumar , K. Rugveda

Department of Computer Science and Engineering

SRM Institute of Science and Technology

Ramapuram, Chennai - 600089

ABSTRACT

The main obstacle for the use of deep learning in medical and engineering sciences is its interpretability. The neural network models are strong tools for making predictions however, they often provide little information about the features that play significant roles in influencing the accuracy of prediction. To overcome this, many regularization procedures about learning the neural networks have been proposed for dropping non-significant features. The lack of theoretical results casts doubt on the applicability of such pipelines is unfortunate. We guarantee the use of the adaptive group lasso for selecting important features of neural networks. In many high dimensional classification or regression problems set in a biological context, the complete identification of the set of informative features is often as important as predictive accuracy, since this can provide mechanistic insight and conceptual understanding. Lasso and related algorithms have been widely used since their sparse solutions naturally identify a set of informative features. However, Lasso performs erratically when features are correlated. This limits the use of such algorithms in biological problems, where features such as genes often work together in pathways, leading to sets of highly correlated features. In this paper, we examine the performance of a Lasso derivative, the exclusive group Lasso, in this setting.

INTRODUCTION

There is an increase in demand for houses day by day as we are moving towards the aim of being a more developed civilization. Accurate forecasting of the house

prices have always impressed sellers and the buying person. The demand for the housing market is always increasing every year due to population growth due to relocation for their financial purpose. Long-term real estate price forecast which is especially important for the remaining people to stay for a long time but not permanently with such people they do not want to take risks during the construction of the house.In order to predict the price of a house one person usually tries it find the same structures in his place again based on the data collected that person will try to predict house price. In this project, house price forecasts for the house is made using different machine learning algorithms such as Random Forest Regression, Ridge Regression, LASSO Descent, Descent of Decision Tree, XGBoost Down and we use the Ada-Boost algorithm to upgrade weak students to strong students. This function works various strategies such as variance influence factor, size reduction strategies and data conversion strategies such as outliers and lost the importance of treatment and box-cox modification strategies. A few things that affect the price of a house include visual features, location and a few influential economic factors at the time. All this indicates that a real estate price forecast appears on a research site that requires information on machine learning.

EXISTING SYSTEM

We propose a neural-based (FS) network feature selection system that can control the rate of reuse of selected features by combining two penalities into a single function. Group Lasso Compensation aims to produce a minimum of features in a collected format. The redundancy-control penalty, which is denied based on a measure of dependence among features, is utilized to control the level of redundancy among the selected features. Both the penalty terms involve the L2,1-norm of weight matrix between the input and hidden layers. These penalty terms are nonsmooth in nature, and hence, one simple but efficient smoothing technique is used to overcome this issue. The monotonicity and convergence of the proposed Then, extensive experiments are conducted on both artificial and real data sets. Empirical results explicitly demonstrate the potential of the proposed FS scheme and its effectiveness in controlling redundancy. Empirical simulation is seen as consistent with theoretical effects.

In this paper, we have proposed an integrated FS scheme with control on the level of redundancy. The Group Lasso regularization is applied to the weights which connects the input and hidden layers of a neural network to produce group sparsity and select useful features. Most data sets, however, usually contain some redundant features. Keeping all discriminatory redundant features will increase the cost and complexity of designing the system. On the other hand, the removal of all redundant features is also not good. A system with some redundancy brings about an easier learning process and is more robust to measurement and other noise. Hence, another essential penalty based on correlation measure is designed to control the level of redundancy in the selected features. However, both penalties are non-differential at the origin. This not only leads to difficulties in the theoretical analysis but also generates oscillations in the experiments. A smoothing technique is used to overcome this drawback. On this basis, the theoretical analysis for monotonicity and convergence is presented. Although we have used the Pearson Correlation coefficient as a measure of dependence, in a more general setting, other measures of dependence, such as mutual information, can also be used without any changes in the theory and learning algorithm. Twenty-six data sets, which cover low-, medium-, and high- dimensional data sets, are used to test the proposed method.

DRAWBACKS OF THE EXISTING SYSTEM

- Difficult to be used in large-scale parallel computing.
- Can't learn the relation between factors.
- Time-consuming approach.
- Computational cost in training the model is high.

PROPOSED SYSTEM

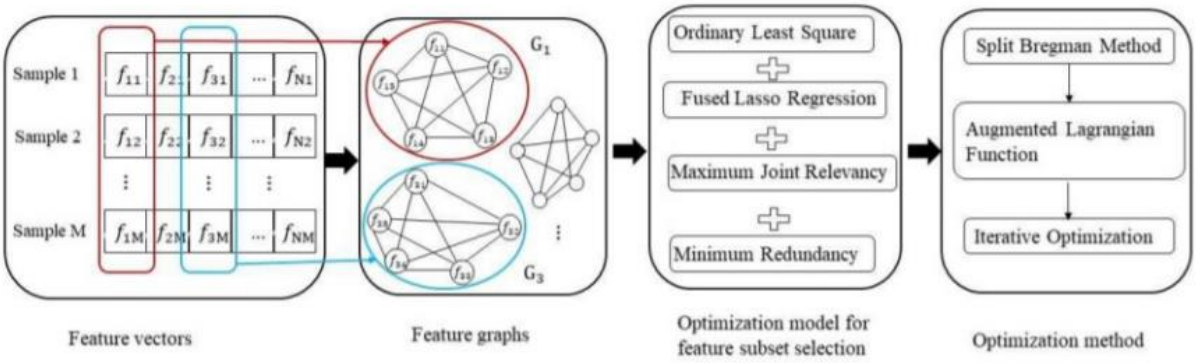
We propose fast algorithms to solve the exclusive group Lasso, and introduce a solution to the case when the underlying group structure is unknown. The solution involves choosing the stability with random group allocation and introduction of artificial features.

Informative features are more likely to be discovered when no other informative features coexist in the same group, while the probability of selecting an irrelevant feature is lower if a group contains at least one informative feature. The ideal group allocation is therefore to allocate each informative (and correlated) feature into a separate group, and set the number of groups to the number of informative features. We refer to this ideal group allocation as fixed groups. In most real-world cases informative features are unknown. We define and capture a fine-grained location profile powered by a diverse range of location data sources. We observe that the location of houses play a critical role in house price prediction. Therefore, we focus on enriching the location-driven house features and grouping them into four profiles for further fine-grained, namely house, education, transportation and facility, respectively.

ADVANTAGES OF THE PROPOSED SYSTEM

- Improved traceability.
- Provides a tangible improvement of final classification performance.
- Quick and Efficient to use Excellent empirical performance.
- Relatively simple and computationally inexpensive method.
- Simple, fast and less complex.
- It is a fast and easy procedure to perform.

OVERALL SYSTEM ARCHITECTURE



CONCLUSION

In this paper, we have developed a new fused lasso feature selection with structural information. The proposed method incorporates knowledge of the structural correlation between pairwise samples into the feature selection process, and has the potential to maximize joint relevance of pairwise feature combinations in relation to the target and minimize redundancy of selected features. In addition, the proposed method can promote sparsity in the features and their successive neighbors. An effective iterative algorithm is proposed to solve the proposed feature subset selection problem based on the split method.

FUTURE WORKS

In future, we plan to implement optimized feature extraction for image processing.

REFERENCES

- Y. Saeys, I. Inza, and P. Larranaga, "A review of feature selec- tion techniques in bioinformatics," Bioinformatics, vol. 23, no. 19, pp. 25072517, Oct. 2007.
- S. Li and D. Wei, "Extremely high-dimensional feature selection via feature generating samplings," IEEE Trans. Cybern., vol. 44, no. 6, pp. 737747, Jun. 2014.
- H. Yin, "ViSOMa novel method for multivariate data projection and structure visualization," IEEE Trans. Neural Netw., vol. 13, no. 1, pp. 237243, 2002.
- S. Solorio-Fernndez, J. F. Martnez-Trinidad, and J. A. Carrasco-Ochoa, "A new unsupervised spectral feature selection method for mixed data: A lter approach," Pattern Recognit., vol. 72, pp. 314326, Dec. 2017.
- K. Nag and N. R. Pal, "A multiobjective genetic programming-based ensemble for simultaneous feature selection and classication," IEEE Trans. Cybern., vol. 46, no. 2, pp. 499510, Feb. 2016.
- Z. Zhang, L. Bai, Y. Liang, and E. Hancock, "Joint hypergraph learning and sparse regression for feature selection," Pattern Recognit., vol. 63, pp. 291309, Mar. 2017.
- J. Wen, Z. Lai, W. K. Wong, J. Cui, and M. Wan, "Optimal feature selection for robust classication via l2,1-norms regularization," in Proc. 22nd Int. Conf. Pattern Recognit., Aug. 2014, pp. 517521.
- K. Sun, S.-H. Huang, D. S.-H. Wong, and S.-S. Jang, "Design and application of a variable selection method for multilayer perceptron neural network with LASSO," IEEE Trans. Neural Netw. Learn. Syst., vol. 28, no. 6, pp. 13861396, Jun. 2017. S. Scardapane, D. Comminiello, A. Hussain, and A. Uncini, "Group sparse regularization for deep neural networks," Neurocomputing, vol. 241, pp. 8189, Jun. 2017.
- L. Jiang, C. Li, S. Wang, and L. Zhang, "Deep feature weighting for naive bayes and its application to text classication," Eng. Appl. Artif. Intell., vol. 52, pp. 2639, Jun. 2016.
- G. Qu, S. Hariri, and M. Yousif, "A new dependency and correlation analysis for features," IEEE Trans. Knowl. Data Eng., vol. 17, no. 9, pp. 11991207, Sep. 2005.
- J. Xu, B. Tang, H. He, and H. Man, "Semisupervised feature selection based on relevance and redundancy criteria," IEEE Trans. Neural Netw. Learn. Syst., vol. 28, no. 9, pp. 19741984, Sep. 2017.
- S. Tabakhi and P. Moradi, "Relevanceredundancy feature selection based on ant colony optimization," Pattern Recognit., vol. 48, no. 9, pp. 27982811, Sep. 2015.

- F. Nie, S. Yang, R. Zhang, and X. Li, "A general framework for auto- weighted feature selection via global redundancy minimization," IEEE Trans. Image Process., vol. 28, no. 5, pp. 24282438, May 2019.
- I.-F. Chung, Y.-C. Chen, and N. R. Pal, "Feature selection with con- trolled redundancy in a fuzzy rule based framework," IEEE Trans. Fuzzy Syst., vol. 26, no. 2, pp. 734748, Apr. 2018.
- M. Banerjee and N. R. Pal, "Unsupervised feature selection with controlled redundancy (UFeSCoR)," IEEE Trans. Knowl. Data Eng., vol. 27, no. 12, pp. 33903403, Dec. 2015.
- Z. Xie and Y. Xu, "Sparse group LASSO based uncertain feature selection," Int. J. Mach. Learn. Cybern., vol. 5, no. 2, pp. 201210, Apr. 2014.
- M. Forti, P. Nistri, and M. Quincampoix, "Generalized neural network for nonsmooth nonlinear programming problems," IEEE Trans. Circuits Syst. I, Reg. Papers, vol. 51, no. 9, pp. 17411754, Sep. 2004.
- J. Wang, Y. Wen, Z. Ye, L. Jian, and H. Chen, "Convergence analysis of BP neural networks via sparse response regularization," Appl. Soft Comput., vol. 61, pp. 354363, Dec. 2017.
- C. Lazar, J. Taminau, S. Meganck, et al., "A survey on lter tech- niques for feature selection in gene expression microarray analysis", 2012.