Novel feature selection methods and BDNN for Deceit Detection

Craig Li

Research School of Finance, Actuarial Studies and Statistics,
The Australian National University, Canberra, Australia
u7239585@anu.edu.au

Abstract. This paper investigates some novel feature selection methods such as Stochastic Gates, LassoNet and Fuzzy Roguh Sets to fetch appropriate features for deceit detection, designing a bidirectional neural network with Gaussian noise to check any prediction accuracy improvement.

Keywords: BDNN · Stochastic Gates · LassoNet · Fuzzy Rough Sets.

1 Introduction

With growing popularity of deep learning and neuropsychology, people's interest to understand more about ourselves and how we interact with each other and surrounding environment, such as affective computing [1], is gaining momentum in academia and education industry. Nowadays we are surrounded by fake news,manipulated information both on social media and in real business world. How to detect deceit and owning trust is playing a paramount role in facilitating our daily lives as well as maintaining business collaboration and social bonds. However, according to a research conducted by C. F. Bond Jr et al, people detect deception consciously at only around chance levels[4]. In order to increase deceit detection accuracy, researchers are analysing physiological signals such as Blood Volume Pulse (BVP), Galvanic Skin Response (GSR),Skin Temperature (ST) and Pupillary Dilation (PD), hoping to uncover "God's secrets". In this experiment, I will carefully examine data sampled from these signals and go through feature selection and data representation learning to classify veracity of subjective belief based on research conducted by X. Zhu et al[2].

1.1 Data Exploration

The dataset contains sampling data from four main categories of physiological signals, in the form of statistic summaries from each genre of physiological signals. Using statistical techniques for data preparation to increase the generalisation and the reliability of neural networks have been suggested by many researchers [3]. In total, there are 119 features across the four physiological signals: 34 (BVP) + 23 (GSR) + 23 (ST) + 39 (PD).[2]. Statistics summaries of physiological signal still convey information such as typical range, gradient, and variation of the signals[5]. The data was normalized and smoothed through lowpass Butterworth filter, consisting of 368 samples with 119 features and the target subjective belief category(0 or 1). During the data exploration, I found that some obvious discrepancies in BVP data whose values are very large, which largely affect feature selection, but due to the medical signal's inherent noise, uncertainty and incompleteness, data management in this domain is worth careful future investigation.

1.2 Problem Definition and Investigation Procedures

In X. Zhu's research [2], the generalized NN approach with full feature set gained the highest accuracy with 63%. However, promising feature selection methods such as Genetic Algorithm(GA) combined with neural network did not contribute too much in the accuracy performance compared with the combination of full features with neural network.Normally speaking, the upper bound of machine learning model prediction accuracy is determined by data and feature, proper use of model and algorithms is trying to approximates to the upper accuracy bound. So in this experiment, we will try several feature selection methods except for GA and other feature selection methods, such as SFS (Sequential Forward Selection),Statistical Dependency (SD) and etc., which had been evaluated with GA in processing signals emitted from medical device such as [7]. Besides, the combination of selected features with SVM gained better performance compared with SVM model-based prediction without feature selection.

1.3 Feature Selection

Increasingly demanding prediction accuracy in both academia and industry, exponentially growing data volume brought by Internet, bioinformatics and IoT, followed by critics regarding model interpretability and explainability, cost-effectiveness, all leads to a resurgence of feature engineering even in the deep learning era. Feature selection methods can be classified into three categories based on feature selection process: Filter , wrapper and embedded [8], gaining insights in the learning process, providing interpretability and reducing computational overhead and over-fitting whereas increasing prediction accuracy. Nowadays hybrid approaches are developed by researchers, combining different soft computing techniques like artificial neural network, fuzzy inference system, approximate reasoning and optimization methods such as evolutionary computation, swarm optimisation, rough sets etc.[9]. Based on fuzzy set and rough sets theory, fuzzy rough sets [10] based algorithms for feature selection is a good way to tackle uncertain and incomplete information, which proves to be a very effective tool for feature selection [11]. With popularity of deep learning both in academia and industry, an increasing number of feature selection experiments are conducted on deep neural networks to gain estimated prediction accuracy with existing data, because the main advantage of deep learning, originated from artificial neural network, over traditional machine learning, is its strong data representation learning capabilities over large-scale datasets without manual feature extraction but with good prediction accuracy. According to the Hierarchy Principle, deep learning combines low-level features to form more abstract high-level features, and discovers distributed representations of data [12]. Such kind of experiments are conducted especially in user scenarios of high dimensional feature spaces such as bioinformatics, environmental and atmospheric sciences, as a novel approach, really worth investigation, methods discussed in [3] [13][14][15][16] can be leveraged to combine machine learning techniques such as support vector machine, etc to overcome over-fitting problem and provide interpretability in predictive machine learning.

1.4 Bidirectional Neural Network Design

According to the design of bidirectional neural network done by A.F. Nejad and T.D. Gedeon [17], it suggests not using bidirectional neural networks for classes of problems which are inherently not invertible. For example, the output is ”yes” or ”no”, and in my case the input also has the same problem, with a large number of combinations of participants and videos watched with 119 features. But inspired by the XOR example proposed by A.F. Nejad and T.D. Gedeon in [17] and Generative Adversarial Networks(GANs) and Variational AutoEncoder, I plan to use Variational AutoEncoder neural network to mimic the bidirectional neural networks. The extra nodes solution in [17] will be implemented by the encode layer in AutoEncoder and their weights. This experiment uses six layers of neurons (the input layer is not included). The third layer serves as the encode layer, which learns the representation of input data distribution through neural network connections and weights, hoping to establish one-one mapping from each input neurons which ingest 119 features to the final subjective belief result (0 or 1). Underlying statistics concerns and one-one mapping function shall be carefully investigated to make the mimicked bidirectional neural network is fully invertible. In my implementation, the number of epochs to train the designed bidirectional neural network is set to 100, Adam optimiser is selected with both binary cross entropy and LK divergence as loss function, following by 0.001 learning rate. Due to time constraint, this part should be carefully investigated to make sure the encode layer to have appropriate encoding while still maintaining pivotal statistical properties and one-one mapping just like extra node solution in [17].

2 Method

2.1 Stochastic Gates Feature Selection

Feature extraction and selection is super paramount to model prediction accuracy, resolve over-fitting and learning interpretability. The correct path to improve prediction accuracy is to check data and feature selection. Due to the fact that the data has been preprocessed into statistics summary, such as mean, standard deviation and etc, hard to build models for the underlying noise of medical device signal. So I focus on feature selection, which plays a vital role in model prediction accuracy. With Stochastic Gates[18], we get a rough estimation of prediction accuracy based on existing data and proceed with feature selection simultaneously. Stochastic Gates is an embedded feature selection method for non-linear models, improves upon LASSO formulation, incorporating Bernoulli distribution into feature selection. The experiment with Stochastic Gates as feature selection and accuracy estimation is not satisfied. The overall prediction precision is around 56% as indicated in the confusion matrix in Figure 1.

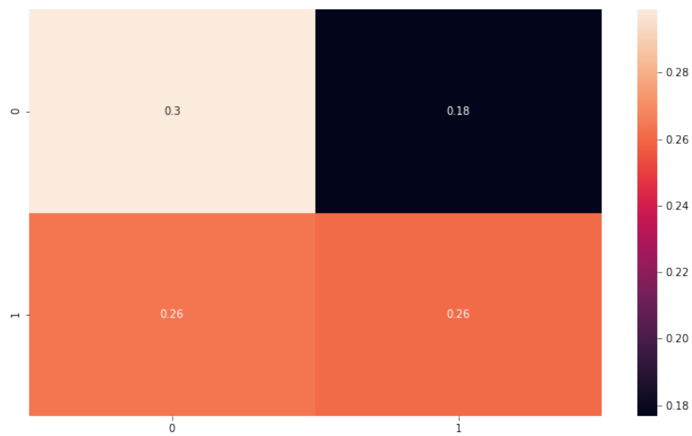

Fig. 1. Confusion Matrix of ST Gates for feature selection.

The feature selection result is not satisfied, but is still under analysis.

2.2 Fuzzy Rough Sets Based Feature Selection

Fuzzy Rough Sets based feature selection are good at tackling vague and incomplete information. In this experiment, a Gaussian kernel based fuzzy rough set approach proposed by Soumen Ghosh et al.[20] was used. The selected features and their prediction accuracy with traditional machine learning methods are as follows:

Feature Selection Methods	Selected Features								
Gaussian Kernel Based Fuzzy Sets	10_bvp	26_bvp	17_bvp	24_eye	20_gsr	1_bvp	20_temp	29_bvp	12_eye

2.3 GA Based Feature Selection

GA based feature selection and its combination with traditional machine learning method as follows:

Classfier	Accuracy	Precision	Recall	F1 score
RF(with GA)	61%	63%	61%	62%
RF (with Fuzzy)	61%	63%	61%	62%
RF (with full features)	50%	53%	51%	51%
SVM(with GA)	53%	52%	100%	69%
SVM (with Fuzzy)	53%	53%	100%	69%
SVM(with full features)	55%	53%	97%	69%
Neural Network(with fill features)	58%	55%	51%	53%

2.4 LassoNet Based Feature Selection

LassoNet is a method for feature selection in neural networks, to enhance interpretability of the final network[16]. The experiment is still underway and close to get results.

3 Results and Discussion

Currently the feature selection does not improve much accuracy in general, including GA and fuzzy rough sets based methods. The logistic regression model with full features can achieve 63.5% accuracy, the combination of fuzzy rough sets and logistic regression lowered the accuracy to 45%, but both GA and fuzzy rough sets based selection do not affect support vector machine's prediction accuracy, staying at 53%. Feature selection from GA and fuzzy rough methods greatly boosted Random forest prediction accuracy. The overall accuracy of my experiment under GA and traditional machine learning methods is very close with that achieved by X. Zhu. The neural network does not change too much against feature selection, staying around 58% due to over-fitting and small dataset.

4 Conclusion and Future Work

4.1 Conclusion

The main purpose of my experiment is to achieve at least the same amount of classification accuracy or exceed that achieved by X.Zhu'[2], so Stochastic Gates and logistic regression were employed to estimate the rough accuracy with existing data since there is no perfect data and model, just suitable models with roughly correct data to approximate the best classification and regression accuracy. During the first trial with Stochastic Gates, the estimated prediction accuracy is around 63%, but forgot to set the random seed and hard to reproduce the result, which is extremely weird during the hyper parameters tuning, but definitely the prediction accuracy of logistic regression with full features is 63.5%, which is the same with that achieved by selected features through genetic algorithm. Due to time constraint and I got stuck in experiments by some tricky errors, fuzzy rough set based feature selection is fully finished with traditional machine learning methods as prediction models. LassoNet based feature selection experiment and other potential methods are not fully completed, from my intuition, LassoNet based feature selection is very promising.

So, from my experiments, I can confirm that the research done by X. Zhu[2] for the generalised neural network with full features is reproducible and the accuracy is achievable, besides my experiment result with full features or features selected by GA is only a little bit better. The prediction accuracy of hybrid group and participant layer-wise approach shall be compared with that of bidirectional neural network with selected features is not finished. Based on the prediction results from Stochastic Gates, the accuracy shall not be better than that of the hybrid group and participant layer-wise approach conducted by ZX.Zhu, since Stochastic Gates which incorporates Bernoulli distribution into feature selection, and my design of bidirectional neural network just learns the statistical distribution of input data, trying to establish an one-one mapping from input neurons to subjective belief result in the encode layer inspired by Variational AutoEncoder neural nets. Compared with X. Zhu's method, my way shall have higher generalisation capability based on statistical distribution of input data. X. Zhu's group/participant layer-wise method has increased accuracy, but has very limited generalisation and transferable capability.

4.2 Future Work

- Detailed experiments will be conducted to compare LassoNet, Stochastic Gates and Deep Forest [21] besides GA and Fuzzy Rough Sets based feature selection since feature engineering is paramount to the prediction accuracy and learning interpretability.
- Medical device signals such BVP, GSR, ST and PD will be carefully examined and check whether it is able to model the underlying noise that mixed into the signal.
- The subject belief can be categorized into several classes such as high, medium, low and etc. to reflect the vagueness and uncertainty of observers, gauging the degree of veracity of presenters in future experiments.
- To achieve higher accuracy, we still need more data and can have a try with more novel representation learning models.

References

1. Picard, Rosalind, Affective Computing. Cambridge, MA: MIT Press. p. 1. (1997).
2. X. Zhu, T. Gedeon, S. C. 2. I. International, 2020, “Deceit Detection: Identification of Presenter’s Subjective Doubt Using Affective Observation Neural Network Analysis,” ieeexplore.ieee.org,(2020)
3. Tamás D. Gedeon, “Data mining of inputs-analysing magnitude and functional measures,” users.cecs.anu.edu.au, 1997. [Online]. Available: http://users.cecs.anu.edu.au/ Tom.Gedeon/pdfs/ContribDataMinv2.pdf. [Accessed: 24-Apr-2021].
4. Bond Jr, Charles F., and Bella M. DePaulo. ”Accuracy of deception judgments.” Personality and social psychology Review 10, no. 3 (2006): 214-234.
5. Gong, P., Ma, H.T., Wang, Y.: Emotion recognition based on the multiple physiological signals. In: International Conference on Real-Time Computing and Robotics, pp. 140–143. IEEE, Angkor Wat (2016)
6. Author, F.: Article title. Journal **2**(5), 99–110 (2016)
7. J. S. Rahman, T. Gedeon, S. Caldwell, R. Jones, M. Z. Hossain, and X. Zhu, “Melodious Micro-frissons: Detecting Music Genres from Skin Response,” in Proceedings of the International Joint Conference on Neural Networks, 2019, vol. 2019-July, pp. 1–8.
8. Girish Chandrashekar and Ferat Sahin, “A survey on feature selection methods,” Computers & Electrical Engineering, vol. 40, no. 1, pp. 16–28, Jan. 2014.
9. K. Jaganath, Mr. P. Sasikumar, Graph Clustering and Feature Selection for High Dimensional Data, International Conference On Global Innovations In Computing Technology, vol. 2, issue 1, pp. 3786-3791, March 2014.
10. T.K.SheejaaA. SunnyKuriakose, “A novel feature selection method using fuzzy rough sets,” Computers in Industry, vol. 97, pp. 111–116, (2018).
11. S. Vluymans, L. D’eer, Y. Saeys, C. Cornelis Applications of fuzzy rough set theory in machine learning: a survey Fundam. Inform, pp. 1-34, (2015)
12. J. Cai, J. Luo, S. Wang, and S. Yang. Feature selection in machine learning: A new perspective. Neurocomputing, 300:70–79, (2018).
13. Y. Yamada, O. Lindenbaum, S. Negahban, and Y. Kluger, “Feature Selection using Stochastic Gates.,” ICML, (2020).
14. Y. Yao, J. Plested, and T. Gedeon, “Deep Feature Learning and Visualization for EEG Recording Using Autoencoders,” in Neural Information Processing, Springer, Cham, pp. 554–566, (2018)
15. Yingkun Huang, Weidong Jin, Zhibin Yu, Bing Li, Supervised feature selection through Deep Neural Networks with pairwise connected structure, Knowledge-Based Systems, Volume 204, (2020)
16. Lemhadri, Ismael and Ruan, Feng et al, LassoNet: Neural Networks with Feature Sparsity,arXiv preprint arXiv:1907.12207, (2019)
17. A. F. Nejad and T. D. Gedeon, ”Bidirectional neural networks and class prototypes,” Proceedings of ICNN’95 - International Conference on Neural Networks, 1995, pp. 1322-1327 vol.3, doi: 10.1109/ICNN.1995.487348.
18. Y. Yamada, O. Lindenbaum, S. Negahban, and Y. Kluger, “Feature Selection using Stochastic Gates.,” ICML, (2020).
19. Lenz OU, Peralta D, Cornelis C,fuzzy-rough-learn 0.1: a Python library for machine learning with fuzzy rough sets. IJCRS 2020: Proceedings of the International Joint Conference on Rough Sets, pp 491–499. Lecture Notes in Artificial Intelligence, vol 12179, Springer.doi: 10.1007/978-3-030-52705-1_36, (2020).

20. S. Ghosh, P. S. V. S. S. Prasad, and C. R. Rao, “An Efficient Gaussian Kernel Based Fuzzy-Rough Set Approach for Feature Selection,” in Multi-disciplinary Trends in Artificial Intelligence, Springer, Cham, pp. 38–49, (2016).
21. Z. H. Zhou, J. F. N. S. Review, 2019, “Deep forest,” academic.oup.com .