Deceit Detection Using Metaheuristics, Fuzzy Sets and Machine Learning

Craig Li

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

Abstract. This paper firstly uses logistic regression and Stochastic Gates to roughly estimate the deceit prediction accuracy and then investigates feature selection methods like meta-heuristic and fuzzy rough sets to obtain appropriate representation in predicting deceit, comparing prediction metrics of various machine learning models.

Keywords: genetic algorithm · feature selection · fuzzy rough.

1 Introduction

With growing popularity of AI, Cognitive Neuroscience and Psychology, people's interest to understand more about ourselves such as intelligence and how we interact with each other and our surrounding environment is soaring. Affective computing [1], a young interdisciplinary subject intersects at the aforementioned areas is gaining momentum in academia and education industry to help us better understand ourselves and seamlessly interact with artificial systems. Deceit detection, one of the research topics in affective computing is extremely worthwhile since 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[2]. 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". All these biomedical signals are full of noises and interferences, making it very hard to extract, construct and select meaningful features. In this paper, meta-heuristic and fuzzy rough sets are employed to do feature selection from constructed statistical features of biomedical signals, before being fed into popular machine learning models.

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 sig-

nals. Using statistical techniques for data preparation to increase the generalisation and reliability of neural networks have been suggested by many researchers [4]. In total, there are 119 features across the four physiological signals: 34 (BVP) + 23 (GSR) + 23 (ST) + 39 (PD)[3]. 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, we 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 really requires further cautious investigation in the hope that some unique features can be extracted from biomedical signal, which can significantly facilitate deceit prediction.

1.2 Problem Definition and Investigation Procedures

In X. Zhu's research [3], 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. Moreover, X. Zhu's research [3] in group and participant specific layer of neural network limited the model generalization to wider audiences without participant and group specific information although it achieved high prediction accuracy per group or participant. Normally speaking, the upper bound of machine learning model prediction accuracy is determined by data and feature, proper use of model and algorithms is just trying to approximate the upper accuracy bound. So in this paper, we will focus on feature selection methods such as fuzzy rough sets and metaheuristic method –genetic algorithm, hoping to improve prediction accuracy without participant or group level information added into the training and prediction cycle in that it also contradicted with the data preprocessing step to standardize or normalize data with min-max scaler to eliminate participant specific traits mentioned by X. Zhu.

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 [6], 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.[7]. According to fuzzy set and rough sets theory, fuzzy rough sets [8] 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 [9]. 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 [10]. 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 [4] [11][12][13][14] can be leveraged to combine with traditional learning techniques such as support vector machine, logistic regression, random forest and etc. to overcome over-fitting problem and provide interpretability in preaching machine learning.

1.4 Rough Estimate of Prediction Accuracy

Feature extraction and selection is super vital to model prediction accuracy, resolving over-fitting and learning interpretability. The correct path to improve prediction accuracy is to check data and features with domain experts first, then gain a rough estimate of prediction accuracy with popular and simple machine learning models and algorithms, ending with ensemble methods composed of some weak learners or deep learning to gain higher prediction accuracy based on data set size and quality. Due to the fact that data at hand has been preprocessed into statistics summaries, such as mean, standard deviation, etc., it is hard to extract extra features through some representation learning methods. So this paper focuses on feature selection, which plays a vital role in model prediction accuracy. With Stochastic Gates[15] and logistic regression, we gain a rough estimation of the prediction accuracy based on existing data and feature, proceeding with feature selection. Stochastic Gates is an embedded method for non-linear models, improves upon LASSO formulation, incorporating Bernoulli distribution into feature selection. The overall prediction precision is around 56% as indicated in the confusion matrix in Figure 1. With traditional machine learning prediction accuracy method –logistic regression, the classification accuracy can reach 62%, indicated in table 2, which is a good start to proceed with feature selection methods such as GA and fuzzy rough sets.

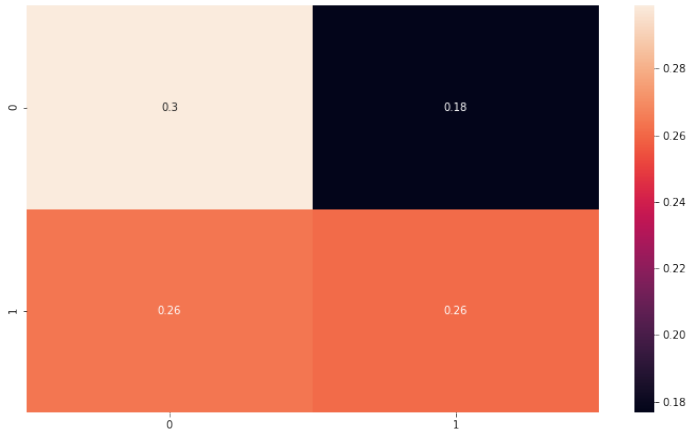

Fig. 1. Confusion Matrix of Stochastic Gates for feature selection.

2 Method

2.1 Fuzzy Rough Sets Based Feature Selection

Fuzzy Rough Sets based feature selection are good at tackling vague and incomplete information. The Gaussian kernel based fuzzy rough set approach proposed by Soumen Ghosh et al.[16] was used. The selected features and their prediction accuracy with traditional machine learning methods are in table 1.

Table 1. Features selected by Fuzzy Rough Sets

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

2.2 GA Based Feature Selection

Prediction metrics of GA and Fuzzy Rough sets based selection and its combination with traditional machine learning methods are in table 2. We can clearly see from the table 1 that feature selection does not always help in model prediction. Some models increased their classification accuracy with feature selection, while some others do not. The best model for this preprocess data is the logistic

regression, which scored 62% in accuracy and 64% in F1 score with full features and 64% in accuracy, 66% in F1 score with GA feature selection; Random forest with fuzzy rough sets feature selection achieved the second good accuracy at 61%.

Table 2. Classifier Metrics

Classfier	Accuracy	Precision	Recall	F1 score
Logistic Regression (with GA)	64%	65%	67%	66%
Logistic Regression (with Fuzzy)	45%	47%	46%	47%
Logistic Regression (with full features)	62%	64%	64%	64%
RF(with GA)	58%	61%	56%	59%
RF (with Fuzzy)	61%	63%	61%	62%
RF (with full features)	55%	58%	58%	58%
SVM(with GA)	51%	56%	38%	45%
SVM (with Fuzzy)	53%	53%	100%	69%
SVM(with full features)	54%	58%	59%	69%
Neural Network(with GA)	57%	58%	64%	61%
Neural Network(with Fuzzy)	47%	50%	59%	54%
Neural Network(with full features)	57%	58%	67%	62%

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 64% 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 around 53%. Feature selection from GA and fuzzy rough methods greatly boosted Random forest prediction accuracy. The overall accuracy 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 57% accuracy and 62% in F1 score due to over-fitting with small dataset.

4 Conclusion and Future Work

4.1 Conclusion

The main purpose this paper is to achieve higher classification accuracy compared with that achieved by X.Zhu[3], so Stochastic Gates and logistic regression were employed to estimate the rough accuracy with existing data and features since there is no perfect data and model, just suitable models with roughly

correct data to approximate the highest classification and regression accuracy. During the first trial with Stochastic Gates, the estimated prediction accuracy is around 63% but it is hard to reproduce, but definitely the prediction accuracy of logistic regression with full features is always 62%, a little bit lower than that achieved by selected features through genetic algorithm that scores 64% accuracy.

From experiments, we can confirm that the research done by X. Zhu[3] for the generalised neural network with full features is reproducible and the accuracy is achievable, besides our experiment result with full features or features selected by GA is slightly better. The prediction accuracy of hybrid group and participant layer-wise approach is hard to achieve with traditional machine learning methods even if they are armed with feature selection techniques. Although X. Zhu's group/participant layer-wise method has indeed increased accuracy, it has very limited generalisation and transferable capability to audiences without participant and group information, which shall fall back to generalized prediction. So data engineering and stacked machine learning methods including deep learning shall be continually employed to explore this topic.

4.2 Future Work

(a) Detailed experiments will be conducted on ensemble learning methods besides GA and Fuzzy Rough Sets based feature selection;
(b) Medical device signals such BVP, GSR, ST and PD will be carefully examined and check whether it is able to reprocess the data to discover any unique features that can facilitate deceit detection just like some specific features to music genre classification;
(c) Examine biomedical signal to check whether it is possible to learn representations from signal directly to extract some features and discuss with biomedical professionals to facilitate feature engineering;
(d) The subject belief can be categorized into several classes such as high, medium, low and etc. to reflect the vagueness and uncertainty of observers' belief, gauging the degree of veracity of presenters in future experiments.
(e) To achieve higher accuracy, more data and data details are needed with representation learning models.

References

1. Picard, Rosalind, Affective Computing. Cambridge, MA: MIT Press. p. 1. (1997).
2. Bond Jr, Charles F., and Bella M. DePaulo. "Accuracy of deception judgments." Personality and social psychology Review 10, no. 3, 214-234, (2006).
3. 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).
4. 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].

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. Girish Chandrashekar and Ferat Sahin, "A survey on feature selection methods," Computers & Electrical Engineering, vol. 40, no. 1, pp. 16–28, (2014).
7. 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, (2014).
8. T.K.SheejaaA. SunnyKuriakose, "A novel feature selection method using fuzzy rough sets," Computers in Industry, vol. 97, pp. 111–116, (2018).
9. 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).
10. J. Cai, J. Luo, S. Wang, and S. Yang. Feature selection in machine learning: A new perspective. Neurocomputing, 300:70–79, (2018).
11. Y. Yamada, O. Lindenbaum, S. Negahban, and Y. Kluger, "Feature Selection using Stochastic Gates.," ICML, (2020).
12. 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).
13. 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).
14. Lemhadri, Ismael and Ruan, Feng et al, LassoNet: Neural Networks with Feature Sparsity,arXiv preprint arXiv:1907.12207, (2019)
15. Y. Yamada, O. Lindenbaum, S. Negahban, and Y. Kluger, "Feature Selection using Stochastic Gates.," ICML, (2020).
16. 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). .