Rule Learning from Time-Dependent Data Applied to Fraud Detection

Marine Collery

Rule Learning from Time-Dependent Data Applied to Fraud Detection

Marine Collery[1,2]

[1] IBM France Lab
[2] Inria Saclay Ile-de-France

Abstract. In financial environment, fraud detection is a challenging problem with tremendous financial impacts where data is highly unbalanced, sequential and timestamped. An additional constraint comes from the fact that common machine learning methods cannot be used alone for fraud detection, as every decision made in order to label a transaction as fraudulent needs to be explainable and the complete model understandable.The use of a symbolic language, such as understandable classification rules, is therefore preferred or even required.

Keywords: Rule Learning · Fraud Detection · Time-Dependant Data · Business Rules and Interpretability.

1 Introduction

For few decades now, rule systems have been widely adopted in different industrial fields. Business Rule Management Systems (BRMS) offer an intuitive, human readable and comprehensible way to define business rules and hides the computational aspect for the business user.

With the growth of machine learning in the past years due to the newly available computational power combined with a growing number of accessible datasets, improving quality of a learned predictive model was an important research interest. Today, impressive models are learned but can lack transparency, interpretability and understandability characteristics that are required and essential for numerous application fields. Those models, and especially the ones based on neural networks, are commonly referred to as "black boxes". Focus is progressively shifting towards providing an explanation for decisions a learned model took as well as building interpretable, understandable and transparent models from scratch.

Combining comprehensibility of business rules and machine learning power to tackle the problem, is the approach we are focusing on for this research project.

This strategy is considered in the context of fraud detection that comes with a complex learning problem as well as a full transparency requirement.

2 Related work

Interpretability, interpretation, and explainability With the growth of high performance non interpretable black-box models, an important question is raised: to what extent a model can be considered trustworthy, especially for high-stakes decision making ? Different terms are commonly used when referring to this problem, we clarify their meaning here for further use. *Model interpretability* is the ability (of the model) to explain or to present in understandable terms to a human [15, 8]. *Model rationale* is how the model takes decisions. *Interpretation and explanation (methods)* will be considered equivalent in this paper (subtle differences are not considered). They both refer to methods that explain or translate the model rationale.

The context of fraud detection There are multiple types of financial frauds from credit card fraud to insurance fraud which come with different detection solutions as described by J. West et al. [23]. Credit card fraud detections were for example studied with sequential and non-sequential learning methods by J. Jurgovsky et al. [13] and lead to different types of frauds detected with both approaches. A spatio-temporal attention-based neural network for fraud detection on credit card was recently introduced by D. Cheng et al. [5] and brought promising results for detecting 'suspicious transactions and mining fraud patterns'. However as pointed by J. Guo et al. [11], allowing for more long-range dependencies than common machine learning models can help identify repeated or cyclical appearances of fraudulent events which seems to be the harder to catch. Very recently, tensor networks were used for anomaly detection [22] where the model outperformed deep and classical algorithms on tabular datasets and achieved competitive results on image datasets.

Rule learning Another approach to detect anomalies in runtime process logs took by K. Böhmer et al., is rule mining [3]. It comes with some specific benefits, especially explainability. In opposition to machine learning models, rules are symbolic and key to bring understandable artificial intelligence.

Interpretable models should even be preferred to explaining uninterpretable model a posteriori for any high stakes decisions according to C. Rudin [20]. However, in some context rule based models are not considered as fully-interpretable. Indeed, as presented by Z. C. Lipton in [16], given the limited capacity of human cognition, when we reach a sufficient high dimension, we could consider the model to be less interpretable than a simple compact neural network.

Combining logic rules and deep neural networks is proposed by Z . Hu et al. [12] to enhance the neural network capabilities. This approach could actually also be used for rule learning. We can also mention recent work from I. Kraiem who applied rule learning for multiple anomaly detection [14] and G. Bert who presented an association rule learning approach for temporal noisy data [10].

More global approaches are proposed in [21] to induce if-then-else rules to explain predictions of supervised learning models, or in [18] to learn composi-

tional rules with very little data. As explained in [9], there are two main base families of methods to induce ruleset from training data: extracting rules from a decision tree (examples: CART [4] and C4.5 [19]) or sequential covering that is learning rules directly from data (examples: CN2 [6] and RIPPER [7]).

We can also refer to Inductive Logic Programming (ILP) introduced by S. Muggleton in 1991 [17] where an ILP system is a program that combines positive and negative examples with background knowledge and outputs a correct logical hypothesis. ILP systems result of two main steps: searching for hypothesis and then selecting the best one.

3 Problem, goals and method

Problem statement Modeling data in an interpretable and understandable way is very challenging when working with large-scale and real-world datasets. Interpretable models are commonly simple and have difficulties learning complex patterns. Rule-based approaches typically tend to overfit complex patterns because of the inappropriate simplicity of the rule language available (operators, aggregates...). Dimensionality of overfitted models make human understanding of the model much harder. In the context of fraud detection, with imbalanced datasets, evolving patterns and time dependency, those limitations are highlighted.

Problem How can we learn accurate, understandable and time-dependent rules for decision making and in particular for fraud detection problems?

Hypothesis The hypothesis on which the project helds are:

- rule based-models are fully-interpretable or at least more interpretable than other models;
- machine learning models bring relevant statistical information to learn rules from;
- sequential models (Hidden Markov Models, Matrix Product State based model, ...) can bring interesting statistical information to learn rules from;
- fraud detection is a relevant application domain to illustrate the problem.
- an ideal trade off between bias and variance can be found to generate rules out of different fraud patterns (the more complex patterns are, the harder it is to learn rules and generalize).

Purpose The purpose of this project is to induce sets of accurate and understandable rules with or from machine learning models in time dependent data. It will help achieving fraud detection and prediction in the challenging context of finance and banking environments where full interpretability is required. A longer term objective is to be able to integrate the induction solutions found in IBM products (Operational Decision Manager (ODM) and Automation Decision Service (ADS)).

Goals The goal of the project is to build, tune, test and validate one or multiple solid models and rule learning solutions to detect fraudulent patterns and events resulting in a fraudulent event. This main project goal can be divided in multiple goals:

- Acquiring expertise in fraud detection, rule induction and machine learning models.
- Building one or more models and rule learning solutions as well as an evaluation process to answer the stated problem.
- Experimenting and validating proposed solutions with synthetic and real data.
- Sharing results.

Tasks The following tasks will take part of this project:

- Write a state of the art analysis of the fraud detection models and solutions as well as an inventory of known fraud patterns.
- Write a state of the art analysis of rule learning algorithms as well as existing solutions to optimize parameters values.
- Propose a mathematical model of the problem by specifying inputs and outputs.
- Analyze available open source datasets applicable to the stated problem.
- Experiment with different supervised and unsupervised models found in state-of-the-art papers (reproduce when possible).
- Define an evaluation and test protocol.
- Work deeply on different approaches of the problem to improve results.
- Experiment on external synthetic data before experimenting in vivo on real data.
- Present and make available proof-of-concepts.
- Write papers for conferences, workshops, journals (attend when possible).
- Write final thesis.

The project will use empirical methods [2]. The work will be based on experimenting with specific datasets, performance metrics will be defined in order to evaluate and draw conclusions.

4 Preliminary findings

4.1 Fraud detection data

This research project benefits from the fact that an IBM partner in the financial area comes with a perfect use case for the project: detection of fraudulent events in bank transfers and credit card transactions. Experiments with real data will be feasible but with no access to the dataset, only resulting metrics will be shared. It provides a good final testing experimental environment but is not satisfactory at the research level.

Due to difficulties to generate or collect data for fraud detection for obvious confidentiality reasons, we have not found an existing reference dataset that combines all the following conditions:

– Data should be composed of events which are financial transactions (ideally not just credit card payment transactions).
– Profile of users should be extractable : we need to have the historical data of a client in order to predict fraudulent behavior.
– As a consequence, data should include a notion of time.

However, we could still use existing datasets that are not verifying the following conditions. For example, we can mention Kaggle Dataset : Synthetic Financial Datasets For Fraud Detection [1]. We learned the importance of features preprocessing with the use of this dataset as shown later in section 4.3.

We are currently searching for appropriate datasets to work on. An alternative we selected if we are not able to found fraud detection viable data, is to start with anomaly detection data which comes with comparable characteristics: temporal, unbalanced and evolving patterns (not known when appearing).

4.2 Rule language

What rule learning state-of-the-art analysis highlighted, is that there is an important rule conditions limitation when it comes to existing learning algorithms. Algorithms like RIPPER [7] or CN2 [6] for example, are not scalable for others than basic conditions operators. This comes from the fact that they have a weak internal data representation that is only based on original attributes. With this conclusions in mind, we list below increasingly complex rule structures. They reflect rules we want to be able to learn, in order to describe complex model like fraud detection. In the following rules, x are data attributes, $\{a, b, c, d, e, f\}$ are fixed values (numerical or categorical valid according to the operator in use in condition) and y_{pred} is the target class.

1. Base rule structure. CN2 and RIPPER -like rules.

   `if` x_1 `<` a `and` x_2 `>` b `and` x_3 `=` c
   `then` y_{pred} `=` d

2. Simple features comparisons.

   `if` x_1 `<` a `and` x_2 `>` x_1 `and` x_3 `=` c
   `then` y_{pred} `=` d

3. Linear combinations.

   `if` x_1 `<` a `and` $b_1 * x_2$ `>` b `and` x_3/c_1 `=` c_2
   `then` y_{pred} `=` d

4. Adding aggregates. For example sum, count, min, max, average ... that are applied to a set of data. This is particularly useful when working with time dependent data. We define α, a set of aggregation functions that can have parameters.

   `if` α_1 `<` a `and` $b_1 * x_2$ `>` $x_1 + b_2$ `and` $\alpha_2(c_1)$ `=` c_2
   `then` y_{pred} `=` d

5. Complex structures with aggregates.

   `if sum(`$a * e.x_2 - e.x_1$`) >` d
   `for` $e \in events$ `where` $e.x_1 > b$ `over timewindow(`c`)`
   `then` y_{pred} `=` d

6. Complex temporal expression between events e_1 and e_2.

   `if` $\exists e_1 : e_1.x_1 > 10$ `and` $\exists e_2 : e_2.x_1 = e_1.x_2$
   `where` $e_1.time \in [e_2.time, \text{now}]$
   `then` y_{pred} `=` d

7. Program induction extension. That is increasing complexity of the right part of the rule, by adding chaining or symbolic regression for example. A new variable *var* is defined.
   - Chaining

     `if` x_1 `<` a `and` x_2 `>` b `and` x_3 `=` c
     `then` var `=` $x_2 + d$
     `if` var `=` e
     `then` y_{pred} `=` f

   - Symbolic regression

     `if` x_1 `<` a `and` x_2 `>` b `and` x_3 `=` c
     `then` y_{pred} `+=` $x_2 + d$

4.3 First approach

The first approach took to learn rules with linear combinations (step 3), is to use a data-driven preprocessing approach. As pointed out by Li et al. [15], data preprocessing such as augmentation or regularization can impact interpretability considerably. Very few preprocessing techniques can be used without loss of interpretability, therefore a simple linear approach is chosen. It consists in adding new features to the data provided for the learning step. Those new features are actually linear combination of original features. This approach was chosen following first experiments done with *Synthetic Financial Datasets For Fraud Detection* dataset [1], that showed the difficulty of RIPPER and CN2 algorithms to model data that are not ruled by original features individually. With the manual introduction of a new feature, results improved considerably as shown in Table 1. An automated feature generation process is created with sum and difference operations. Interpretability is maintained thanks to a dimensional consistency filter. However this approach is not scalable for more complex operations and can have impacts on some learning algorithms (for example on RIPPER stopping criteria that depends on data dimensions).

4.4 Future work and ideas

An approach that we would like to develop is the use of intermediary models. Rather than working on the dataset directly, we want to try modeling the data

Table 1. Experiments with Synthetic Financial Datasets For Fraud Detection [1] dataset with and without manual preprocessing with CN2 [6] and RIPPER [7] algorithms

Dataset	Model	Metrics				
		acc	**bal_acc**	**f1**	**precision**	**recall**
raw	cn2	0.999	0.798	0.691	0.822	0.596
processed	cn2	1	0.994	0.993	0.997	0.988
raw	ripper	1	0.873	0.839	0.956	0.748
processed	ripper	1	0.998	0.996	0.997	0.996

first into an intermediary model (tensor networks, bayesian model etc.), before learning rules for that new representation of the data. Additionally further work on how to approach the temporal aspect of the data needs to be completed. With a fraud detection dataset, it would be interesting to apply anomaly detection strategy (supervised and unsupervised) as both domains share data characteristics (unbalanced, temporal).

5 Conclusion

In this paper, we presented the doctoral research project. There is a growing need for understandable AI models. A rule based approach is one potential solution, but they no longer have the same research interest as black boxes models do. We believe that this approach is a solution for many different kind of applications especially financial applications. Modeling with rules, a time-dependent dataset requires a rule language complexity that is not currently possible to learn with available methods. This research project aims at going in that direction.

Acknowledgements This thesis project is supported by PSPC AIDA 2019-PSPC-09. It is supervised by Philippe Bonnard at IBM France Lab and François Fages at Inria Saclay.

References

1. Synthetic Financial Datasets For Fraud Detection. https://kaggle.com/ntnu-testimon/paysim1
2. Bock, P.: Getting It Right: R&D Methods for Science and Engineering. Elsevier Science (Apr 2020)
3. Böhmer, K., Rinderle-Ma, S.: Mining association rules for anomaly detection in dynamic process runtime behavior and explaining the root cause to users. Information Systems **90**, 101438 (May 2020). https://doi.org/10.1016/j.is.2019.101438
4. Breiman, L., Friedman, J., Stone, C.J., Olshen, R.A.: Classification and Regression Trees. Taylor & Francis (Jan 1984)

5. Cheng, D., Xiang, S., Shang, C., Zhang, Y., Yang, F., Zhang, L.: Spatio-Temporal Attention-Based Neural Network for Credit Card Fraud Detection. Proceedings of the AAAI Conference on Artificial Intelligence **34**(01), 362–369 (Apr 2020). https://doi.org/10.1609/aaai.v34i01.5371
6. Clark, P., Niblett, T.: The CN2 Induction Algorithm. Machine Learning **3**(4), 261–283 (Mar 1989). https://doi.org/10.1023/A:1022641700528
7. Cohen, W.W.: Fast Effective Rule Induction. In: In Proceedings of the Twelfth International Conference on Machine Learning. pp. 115–123. Morgan Kaufmann (1995)
8. Doshi-Velez, F., Kim, B.: Towards A Rigorous Science of Interpretable Machine Learning. arXiv:1702.08608 [cs, stat] (Mar 2017)
9. Fürnkranz, J., Gamberger, D., Lavrač, N.: Foundations of Rule Learning. Springer Science & Business Media (Nov 2012)
10. Guillame-Bert, M.: Apprentissage de règles associatives temporelles pour les séquences temporelles de symboles p. 158
11. Guo, J., Liu, G., Zuo, Y., Wu, J.: Learning Sequential Behavior Representations for Fraud Detection. In: 2018 IEEE International Conference on Data Mining (ICDM). pp. 127–136 (Nov 2018). https://doi.org/10.1109/ICDM.2018.00028
12. Hu, Z., Ma, X., Liu, Z., Hovy, E., Xing, E.: Harnessing Deep Neural Networks with Logic Rules. In: Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). pp. 2410–2420. Association for Computational Linguistics, Berlin, Germany (Aug 2016). https://doi.org/10.18653/v1/P16-1228
13. Jurgovsky, J., Granitzer, M., Ziegler, K., Calabretto, S., Portier, P.E., He-Guelton, L., Caelen, O.: Sequence classification for credit-card fraud detection. Expert Systems with Applications **100**, 234–245 (Jun 2018). https://doi.org/10.1016/j.eswa.2018.01.037
14. Kraiem, I.B.: Détection d'Anomalies Multiples par Apprentissage Automatique de Règles dans les Séries Temporelles. Ph.D. thesis, Université de Toulouse-Jean Jaurès (Jan 2021)
15. Li, X., Xiong, H., Li, X., Wu, X., Zhang, X., Liu, J., Bian, J., Dou, D.: Interpretable Deep Learning: Interpretation, Interpretability, Trustworthiness, and Beyond. arXiv:2103.10689 [cs] (May 2021)
16. Lipton, Z.C.: The mythos of model interpretability: In machine learning, the concept of interpretability is both important and slippery. Queue **16**(3), 31–57 (Jun 2018). https://doi.org/10.1145/3236386.3241340
17. Muggleton, S.: Inductive logic programming. New Generation Computing **8**(4), 295–318 (Feb 1991). https://doi.org/10.1007/BF03037089
18. Nye, M.I., Solar-Lezama, A., Tenenbaum, J.B., Lake, B.M.: Learning Compositional Rules via Neural Program Synthesis. arXiv:2003.05562 [cs] (Mar 2020)
19. Quinlan, J.R.: C4.5: Programs for Machine Learning. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA (1993)
20. Rudin, C.: Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead. arXiv:1811.10154 [cs, stat] (Sep 2019)
21. Sushil, M., Šuster, S., Daelemans, W.: Rule induction for global explanation of trained models. arXiv:1808.09744 [cs, stat] (Aug 2018)
22. Wang, J., Roberts, C., Vidal, G., Leichenauer, S.: Anomaly Detection with Tensor Networks. arXiv:2006.02516 [quant-ph, stat] (Jun 2020)
23. West, J., Bhattacharya, M., Islam, R.: Intelligent Financial Fraud Detection Practices: An Investigation. arXiv:1510.07165 [cs] (Oct 2015)