Overview of the evaluation methods for the interpretable business decision support model for cybernetics and organization of production

Dmitry A Zuev¹, Alexey P Kalistratov¹, Andrey G Zuev¹, Guennadiy I Afanasyev¹, Georgiy I Revunkov¹ and Alexey V Potudinskii²

¹Bauman Moscow State Technical University, 5-b1 2-yd Bauman str., Moscow, 105005, Russian Federation
²Military Training and Scientific Center of the Air Force "Air Force Academy named after Professor N.E. Zhukovsky and Yu.A. Gagarin"

E-mail: akalistrov@gmail.com

Abstract. Authors made an overview of approach towards creation business decision support model. Authors underline requirements and evaluation methods for proposed model to be used in cybernetics and organization of production. Authors described domain problems and proposed interpretable model as a solution. Authors explored levels of the evaluation and chosen one that suits their problem. Authors also discuss business model influencing factors as local interpretability, severity of incompleteness, time constraints and level of user expertise.

1. Introduction
Data mining is often achieved by machine learning: from adaptive advertisement to credit score calculation, machine learning systems are increasingly spreading. Models outperform humans on specific tasks like image recognition or text analyzing and often guide processes of human understanding and decisions. But such systems are usually treated as “black boxes”, hence, one can have small understanding, how given model is really working [1]. Malware detection, bike demand prediction, city traffic prediction - all compute their output without human help. There are two main reasons why explanation is not necessary: there are no significant consequences for unacceptable results and the problem is sufficiently well-studied and validated in real applications that we trust the system’s decision, even if the system is not perfect.

But, if model can be considered interpretable, there will be much more trust to it.. Over past two years a number of new methods for machine learning interpretability were published and the volume of research on interpretability is rapidly growing [2].

Authors try to underline factors influencing on the interpretable business decision support model and explore evaluation methods for the interpretable models in general.

2. Background
The machine learning model can be considered interpretable if its structure can be explained by a pre-selected expert. When interpreting the model, the expert explains how the individual characteristics and elements of the model affect the chosen variable and what their functional relationship is. Expert
should be also able to explain the influence of the model parameters on the accuracy and stability of the model’s approximation.

Model interpretability seems to be useful in any domain, but its importance is still unclear as on the first sight the same results can be achieved with “black-box” models. On the other hand, an incompleteness in the problem statement makes analysis and evaluation of a given model very difficult. There are some common cases when interpretation may be helpful:

- scientific understanding: explanation can be used to produce further knowledge;
- safety: complex systems can have unlimited number of test cases and it will be impossible to predict them all;
- ethics: “equal opportunity by design” [3];
- mismatched objectives: without understanding an algorithm we may find that we are optimizing an incomplete or completely wrong objective;
- multi-objective trade-offs: situation when two objectives are competing.

Applying machine learning in the commerce, it is hard to define a single goal. Almost all companies operate in a highly competitive environment, which leads to the process of expanding and maintaining the client base. This type of activity means work in two directions: attracting new customers and retaining of existing ones. These tasks are handled in parallel through marketing. In this work marketing tasks are reduced to the tasks of data mining, such as

1. Regression - finding a functional relationship between input parameters and a continuous output parameter, allowing to estimate the probability of occurrence of an event.
2. Classification - finding a functional relationship between the input parameters and the discrete output parameter, which allows the object to be assigned to a predetermined category.

From the business point of view, the main target is sales profit. So, maximizing sales becomes main objective. But the chosen algorithm may be optimizing an incomplete objective, e.g., the marketing manager may be interested not in the prediction of future sales but in the increasement of the customer loyalty and brand recognition [4,5].

In the presence of an incompleteness, explanation is one of the ways to ensure that effects of gaps in problem formalization are visible to us.

3. Level of the evaluation
In a classic machine learning approach, there are a number of well-known metrics, like F1-score for classification or separation index for regression [6]. From application point of view, evaluation of model performance has the role of efficiency metric: a go-playing agent might best a human player, a classifier may correctly identify cats against dogs.

Finale Doshi-Velez and Been Kim described three common evaluation approaches of interpretable ML models: application-grounded, human-grounded, and functionally-grounded [6]. These approaches can be ranked from task relevant to general. We also need to understand that while human evaluation is essential to assessing interpretability, human-subject evaluation is pretty tough.

First one is application-grounded but it is requires the biggest amount of time and effort. Application-grounded evaluation implies conducting forecast-based experiments within real cases. E.g., we have real case - business managers work on the sales income - the best way to show that the model works is to evaluate it with respect to the task: Does business gain sales? This method is precise and can grant us a lot of relevant data, but it has one main cons: it is risky and difficult to perform experiments on real business.

Human-grounded approach is more applicable in real life, but still requires domain expert and a number of experiments on the relevant domain. Main difference from application-grounded approach is that we can work with models and simulation instead of conducting real-life experiments. Using this approach, we can estimate required time for one might understand a result, and understand overall system complexity that can be controlled. This approach’s main feature is dependency only on quality
of explanation, and quality of explanation can be evaluated without a specific end-goal. In our case, we have domain expert that can perform simulation process: analyze explanation and an input, and simulate the model’s output.

Functionally-grounded approach requires no human experiments. Functionally-grounded evaluations lay on models or regularizers that have already been validated. The most common generally interpretable models are linear regression model and decision trees. Work on model approximation focuses on deriving a simple, interpretable model that approximates a more complex, uninterpretable one. The disadvantage of this approach is that for even moderately complex models, a good global approximation cannot generally be found [7].

4. Task-related relevant factors
One of very important property of interpretable model is its scope of interpretability. Model can be considered interpretable if a given person can comprehend the whole model at once [8,9]. Model can be defined to global or local interpretability type. Global model interpretability helps to understand the distribution of your target variable based on the features. Global model interpretability is hardly archived in practice. On the other hand, local interpretability implies knowing the reasons for a specific decision. Expert can focus on a single model instance and examine aspects of prediction model makes for this input, and why model had made this decision. In business, local interpretability is more preferable than global as local interpretability is easier to be reached and hence more applicable.

The depth of the question’s incompleteness also affects explanation needs. Amount and characteristic of explanations needed depend on many factors: inputs, constraints, domains, internal model structure, costs, or even the need to understand the model’s training algorithm. We can list an example of depth question: one business manager can have plenty of questions how to gain sales. Another one wants to control and understand every single decision made. And the last one wants to check his list of hypothesis and their influence on result. So, in each case there will be difficult requirements towards the explanation even if the input factors (as real sales data) can be the same.

Time constraints depend mainly on domain. The least strict constraints suit scientific applications as scientists often can spare as much time for research as needed. An opposite example, aircraft collision avoidance systems must produce explanations that can be understood quickly. According to the [10], users of business decision support systems want to spend fair amount of time on working with such systems: they are not in hurry, but also don’t want to spend days understanding explanations.

The complexity of explanation should depend on the end user expertise in domain: domain experts may expect or prefer a somewhat larger and sophisticated model, which confirms facts they know, over a smaller, more “black-box” styled one [6]. In ecommerce, we can underline several group of end users: business managers, marketing managers and business analytics. All of them come with different background. So, explanation depth must be fine-tuned to suit everyone’s needs or at least can be tunable at all.

5. Conclusion and future work
In this work, we have laid the groundwork for future model design. This paper focus on the overview of the requirements and evaluation methods for the interpretable business decision support model.

In the problem formulation incompleteness section, we described our domain problems and proposed interpretable model as a solution of them. Then we explored level of the evaluation that suits our problem. Human-grounded approach was selected as most relevant one.

Then, in the task-related relevant factors section we discuss such factors as: local interpretability, severity of incompleteness, time constraints and level of user expertise. We can underline that in our case local explanation with medium complexity may be the best choice for end user.

In future work authors plan to do a research of supervised interpretable models and post-hoc approaches according to described premises.
References

[1] Krause J, Perer A & Ng K 2016 Interacting with predictions: Visual inspection of black-box machine learning models Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (ACM) 5686-97

[2] Rudin C 2014 Algorithms for interpretable machine learning Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining (ACM) 1519-28

[3] Big data: A report on algorithmic systems, opportunity, and civil rights 2016 https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/2016_0504_data_discrimination.pdf

[4] Haryanto J O, Moutinho L and Coelho A 2016 Is brand loyalty really present in the children's market? A comparative study from Indonesia, Portugal, and Brazil Journal of Business Research 69(10) 4020-32

[5] Sun Z L, Choi T M, Au K F and Yu Y 2008 Sales forecasting using extreme learning machine with applications in fashion retailing Decision Support Systems 46(1) 411-9

[6] Doshi-Velez F and Kim B 2017 Towards a rigorous science of interpretable machine learning. arXiv preprint arXiv:1702.08608

[7] Biran O and Cotton C 2017 Explanation and justification in machine learning: A survey IJCAI-17 Workshop on Explainable AI (Venue: RMIT University) p 8

[8] Molnar C, Casalicchio G and Bischl B 2018 Iml: An R package for Interpretable Machine Learning The Journal of Open Source Software 3 786

[9] Lipton Z C 2016 The mythos of model interpretability arXiv preprint arXiv:1606.03490

[10] Turban E, Sharda R and Delen D 2011 Decision Support and Business Intelligence Systems (Pearson) p 760