ESANN 2021 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Online event, 6-8 October 2021, i6doc.com publ., ISBN 978287587082-7.
Available from http://www.i6doc.com/en/.

Context-specific sampling method for contextual explanations

Manik Madhikermi[1,2], Avleen Malhi[2,3], Kary Främling[1,2]

1- Umeå University, Department of computing science
Umeå, Sweden

2- Aalto University, Computer science Department
Espoo , Finland

3- Bournemouth University, Department of computing and informatics
Bournemouth, United Kingdom

Abstract. Explaining the result of machine learning models is an active research topic in Artificial Intelligence (AI) domain with an objective to provide mechanisms to understand and interpret the results of the underlying black-box model in a human-understandable form. With this objective, several eXplainable Artificial Intelligence (XAI) methods have been designed and developed based on varied fundamental principles. Some methods such as Local interpretable model agnostic explanations (LIME), SHAP (SHapley Additive exPlanations) are based on the surrogate model while others such as Contextual Importance and Utility (CIU) do not create or rely on the surrogate model to generate its explanation. Despite the difference in underlying principles, these methods use different sampling techniques such as uniform sampling, weighted sampling for generating explanations. CIU, which emphasizes a context-aware decision explanation, employs a uniform sampling method for the generation of representative samples. In this research, we target uniform sampling methods which generate representative samples that do not guarantee to be representative in the presence of strong non-linearities or exceptional input feature value combinations. The objective of this research is to develop a sampling method that addresses these concerns. To address this need, a new adaptive weighted sampling method has been proposed. In order to verify its efficacy in generating explanations, the proposed method has been integrated with CIU, and tested by deploying the special test case.

Keywords: CIU, weighted adaptive sampling, black-box explanations, XAI

1 Introduction

Recently, explainability is growing interest among researchers specifically when transparency and trust is foremost requirement in AI based applications. Despite several methods that have been developed to interpret the black-box model, the area still lacks maturity. Currently, there are several methods that have been developed for providing explanations to the decisions of an otherwise black- box machine learning model such as SHAP [7], Locally Interpretable Model-Agnostic Explanation (LIME) [8] and CIU [1]. These methods differ from each other in the way they try to explain the black-box model. XAI methods can be classified into categories model explanation, outcome explanation and model inspection

ESANN 2021 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Online event, 6-8 October 2021, i6doc.com publ., ISBN 978287587082-7.
Available from http://www.i6doc.com/en/.

according to [5]. Model explanation signifies providing a global explanation of the black-box model through an interpretable and transparent model called surrogate model. Rule extraction methods and estimation of global feature importance are model explanation methods. Outcome explanation provides an explanation of the outcome of the black-box for a specific instance (or context) and can therefore be considered local. Model inspection consists in providing a representation (visual or textual for instance) for understanding the black-box model or its outcome, i.e. how explanations are produced and presented based on model or outcome explanation methods. Most (or all) current outcome explanation methods are so-called post-hoc methods, i.e. they require creating an intermediate interpretable model to provide explanations. A major challenge of all methods that use an intermediate interpretable model (the âexplanation modelâ in [7]) is to what extent the interpretable model actually corresponds to the black-box model. CIU differs radically from the existing state-of-the-art in XAI because CIU does not create or use an intermediate interpretable model and provide better explanation in several use cases [6].

2 Contextual Importance and Utility (CIU)

CIU does not create or use an intermediate surrogate model or make linearity assumptions [2]. Here, Contextual Importance (CI) and Contextual Utility (CU) are used for generating the explanations and interpretation based on the contributing features of the dataset. The capabilities of these explanations are *contextual* since one feature might be important for taking a decision in one circumstance but can be irrelevant in another circumstance. The mathematical definition (detailed in [4]) of CI and CU is given in equation 1 and equation 2 respectively.

$$CI_j(\vec{C},\{i\}) = \frac{cmax_j(\vec{C},\{i\}) - cmin_j(\vec{C},\{i\})}{absmax_j - absmin_j} \tag{1}$$

$$CU_j(\vec{C},\{i\}) = \frac{out_j(\vec{C}) - cmin_j(\vec{C},\{i\})}{cmax_j(\vec{C},\{i\}) - cmin_j(\vec{C},\{i\})} \tag{2}$$

Here, $CI_j(\vec{C},\{i\})$ is the contextual importance of a given set of inputs $\{i\}$ for a specific output j in the context $\vec{C}$. $absmax_j$ is the maximal possible value for output j and $absmin_j$ is the minimal possible value for output j. $cmax_j(\vec{C},\{i\})$ is the maximal value of output j observed when modifying the values of inputs $\{i\}$ and keeping the values of the other inputs at those specified by $\vec{C}$. Correspondingly, $cmin_j(\vec{C},\{i\})$ is the minimal value of output j observe. Similarly, for contextual utility $CU_j(\vec{C},\{i\})$, $out_j(\vec{C})$ is the value of the output j for the context $\vec{C}$.

2.1 Sampling in CIU

CIU does not create surrogate model but use sampled data for the computation of $cmax_j(\vec{C},\{i\})$ and $cmin_j(\vec{C},\{i\})$ for explanation. The estimation of

ESANN 2021 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Online event, 6-8 October 2021, i6doc.com publ., ISBN 978287587082-7.
Available from http://www.i6doc.com/en/.

$cmax_j(\vec{C}, \{i\})$ and $cmin_j(\vec{C}, \{i\})$ is done for defined value ranges of inputs $\{i\}$ which depends on the task parameters or the input values present in the training set. The current implementation [3] for estimating $cmax_j(\vec{C}, \{i\})$ and $cmin_j(\vec{C}, \{i\})$ uses Monte-Carlo estimation with uniformly distributed, randomly generated values within the provided value ranges of inputs $\{i\}$. This approach of computing $cmax_j(\vec{C}, \{i\})$ and $cmin_j(\vec{C}, \{i\})$ is suitable for many real life use cases however, presence of strong non-linearities in the model to explain that might be missed with lower sampling number. In order to address aforementioned concern, in this paper, we proposed Adaptive Weighted Random Sampling strategy that produces accurate $cmax_j(\vec{C}, \{i\})$ and $cmin_j(\vec{C}, \{i\})$ estimates also in the presence of strong non-linearities or exceptional input feature value combinations in the model, which may be missed by a uniform Monte-Carlo sampling. The proposed sampling strategy, detailed in section 3, is more adapted than the default in the following cases: (a) maintaining the input distribution is important (b) rare events sampling need to be performed (c)less sensitive to the outlier data point (d) co-related input features in dataset.

3 Adaptive Weighted Random Sampling Method for CIU

A new adaptive weighed random sampling method has been proposed and it might worth pointing that this sampling method is mainly focusing on numerical features and not for categorical features.

Algorithm 1: Set of representative input vectors using weighted random sampling

Input: $Data_{train} \leftarrow$ Training data set
$N \leftarrow$ Number of Sample (Default = 100)
$S_c \leftarrow$ Stratification constant (Default = 10)
$W \leftarrow$ Stratification weight (Optional)
Result: S is a $N \times M$ matrix
begin
  $Data_{jpdf} \leftarrow$ ***Get_jointPdf***$(Data_{train})$
  $Data_{cpdf} \leftarrow$ ***Get_cumulativePdf***$(Data_{jpdf})$
  $N_{strata} \leftarrow floor(\frac{N}{S_c})$
  $Data_{strata} \leftarrow$ ***generate_stratifiedtable***$(Data_{cpdf}, N_{strata})$
  if *W is null* **then**
    foreach *strata in* $Data_{strata}$ **do**
      $W.append \leftarrow \frac{Number\ of\ data\ point\ in\ strata}{Total\ number\ of\ data\ point\ in\ Data_{train}}$
    end
  end
  $S_{strata} \leftarrow$ W * N
  foreach *strata in* $Data_{strata}$ **do**
    $S.append \leftarrow$ ***get_sample***$(strata, S_{strata})$
  end
end

Algorithm 1 presents how the *Set of representative input vectors* is selected from the training dataset. The proposed implementation is based on stratification of the cumulative joint distribution function. The number of stratification depends on the number of samples to be selected. The number of strata is N/S_C

ESANN 2021 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Online event, 6-8 October 2021, i6doc.com publ., ISBN 978287587082-7.
Available from http://www.i6doc.com/en/.

where N is number of samples and S_C is stratification constant. By default, N and S_C is 100 and 10 respectively. The choice of N and S_C depends on the use-case and tends to be a compromise between calculation speed and desired accuracy. In this proposed method, each strata has its own weight W. W determines how many samples will be selected from each strata. By default, the proportion of training data in each strata is used as weight for that strata and number of samples to be randomly selected from each strata is $S_{strata} = W * N$. The proposed method is adaptive in the sense that user can define their own preferable weight for each strata. However, the constraint is that the sum of weights for all strata must be equal to 1. The proposed sampling method is flexible and adaptive to select a set of representative samples based on their use-case.

4 Implementation and Result Evaluation

We have tested the explanations using our new proposed weighted sampling method and compared them with the ones generated with existing uniform sampling method using synthetically generated dataset.

4.1 Special case -Spiked Output

To test the special cases of the output distribution, we synthetically generated the dataset for our test cases. Then, the explanations are generated using CIU method and underlying sampling methods are tested based on the explanations. This is the special case where outputs are only spiked for the specific range of input features: "$Feature-1$" and "$Feature-2$", the output value is peaked when $Feature-1$ *between* 2990 & 3005 & $Feature-2 > 13$. In order show the impact of underlying sampling methods in CIU in aforementioned scenario, 600 synthetic data points with two features $Feature-1$ and $Feature-2$ has been generated using R. $Feature-1$ has 600 data points where 400 data points are normally distributed (generated using *rnorm* function of R) with mean of 3000, standard deviation of 800 and rest of 200 data points are normally distributed with mean of 3000, standard deviation of 0.01. Similarly, $Feature-2$ contains 400 uniformly distributed data points between 1 to 10(generated using $runif$ function of R) and rest of 200 are also uniformly distributed data points between 11 to 18. Here, we take a test instance with $Feature-1$ and $Feature-2$ of values as 3000, 32 to explain the output produce by the black-box. CIU generates its explanation CI and CU based on cmin and cmax which are calculated based on the samples generated by underline sampling method as explained in section 2.1. Figures 1 and 2 visually explain the CI, cmin and cmax for existing and proposed sampling methods. Here, the absolute min and max considered in these examples are 675.63 and 9500.02 respectively. The instance considered for the test is having input value as 3000 and output value as 9500. CI, cmin and cmax represent the output range, minimum and maximum output value selected on the chosen sampling method. It is clearly evident from the figures that the existing sampling method is not able to detect the cases where the output value

ESANN 2021 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Online event, 6-8 October 2021, i6doc.com publ., ISBN 978287587082-7.
Available from http://www.i6doc.com/en/.

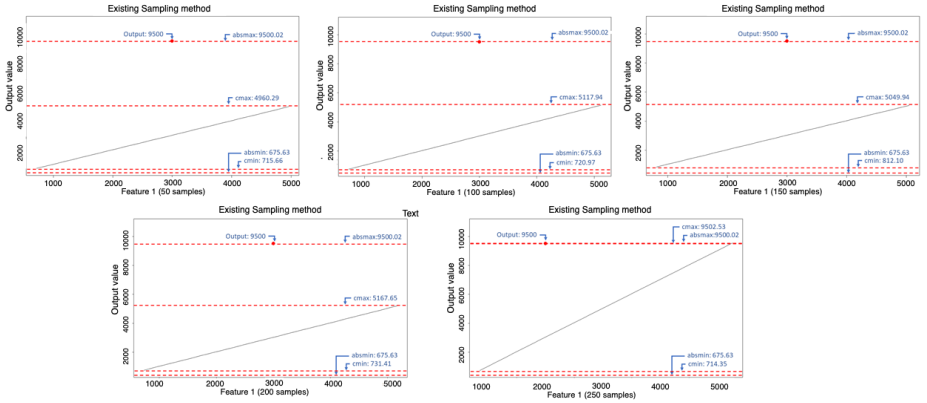

Fig. 1: Output as the function of Input with *Existing sampling method* for number of samples ranging from 50 - 250 for test case

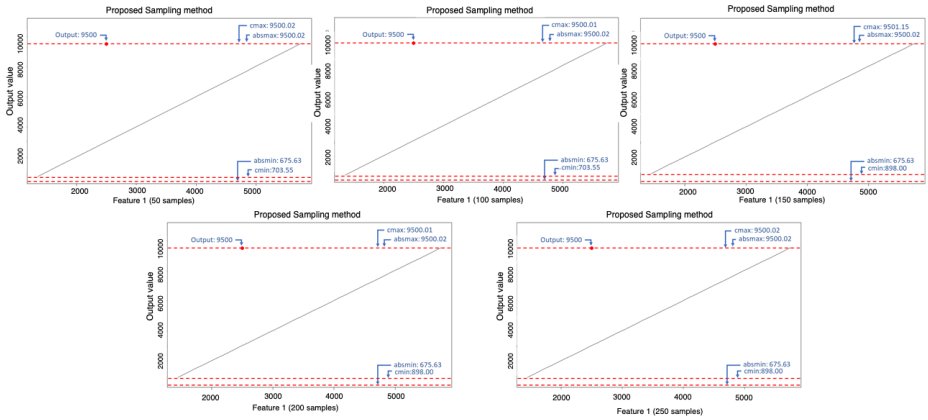

Fig. 2: Output as the function of Input with *Proposed sampling method* for number of samples ranging from 50 - 250 for test case

spikes until 250 samples are selected and the proposed method detects the output spikes even with 50 samples.

5 Discussion and Conclusion

Contextual Importance and Utility (CIU) provides an alternative to LIME and SHAP for generating model-agnostic outcome explanations in tasks comprising numerical features. Choosing an sampling strategy for generating the instances to fit the surrogate model has a major impact on the quality of the approxi-

ESANN 2021 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Online event, 6-8 October 2021, i6doc.com publ., ISBN 978287587082-7.
Available from http://www.i6doc.com/en/.

mation of the black-box decision boundary and thus on the effectiveness of the generated explanation. Due to the randomness of sampling, the resulting explanations may suffer from high discrepancies between repeated evaluations. This paper proposes a new sampling method called adaptive weighed random sampling method which helps in selecting the right samples from the input space. The proposed sampling method helps in generating samples which are representative in context and generates explanations which makes more sense compared to the the existing sampling method for numerical features. For non-surrogate XAI methods like CIU, the effect of sampling directly reflects on the output of CIU. Due to this reason, the proposed sampling method has been tested on CIU to evaluate its effectiveness. The proposed sampling method has been tested using test use-case with spiked output values to identify this special case but the existing sampling method is very sensitive to it.

Currently, the paper is limited to the single use-case. Our future research direction is to apply the proposed sampling technique in other use-case scenario and test with publicly available datasets as well.

References

[1] Främling, K.: Explaining results of neural networks by contextual importance and utility. In: Proceedings of the AISBâ96 conference. Citeseer (1996)

[2] Främling, K.: Modélisation et apprentissage des préférences par réseaux de neurones pour láide à la décision multicritère. (1996)

[3] Främling, K.: ciu: Contextual importance and utility. `https://cran.r-project.org/web/packages/ciu/index.html` (2020)

[4] Främling, K.: Explainable ai without interpretable model. arXiv preprint arXiv:2009.13996 (2020)

[5] Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., Pedreschi, D.: A survey of methods for explaining black box models. ACM computing surveys (CSUR) **51**(5), 1–42 (2018)

[6] Kary Främling, Marcus Westberg, M.J.M.M.A.M.: Comparative study of three outcome explanation methods: The 3rd international workshop on explainable and transparent ai and multi-agent systems, extraamas 2021, london, uk, may 3–7 2021, (In press)

[7] Lundberg, S.M., Lee, S.I.: A unified approach to interpreting model predictions. In: Advances in neural information processing systems. pp. 4765–4774 (2017)

[8] Ribeiro, M.T., Singh, S., Guestrin, C.: " why should i trust you?" explaining the predictions of any classifier. In: Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining. pp. 1135–1144 (2016)