Multi-Objective Semi-Supervised Explanation System

Het Patel College of Engineering
North Carolina State University
Raleigh, NC 27695
Email: hpatel28@ncsu.edu
Shaunak Patel College of Engineering
North Carolina State University
Raleigh, NC 27695
Email: shpate25@ncsu.edu

I. INTRODUCTION

Multi-Objective semi-Supervised Explanation System is a system that provides results, explanations, and justification based on the input from the model using a different combination of the objects that are provided and learning or predicting them over the data that is being used for the system. So, we can break this into two parts one is multi-objective which aims to make different combinations and try to get the best result out of it (which is accurate and fair), and another is Semi-supervised which aims to create different combinations of the label, and unlabeled data from the data-set provided (which is best and rest data). So, the overall main goal of having a Multi-Objective semi-Supervised Explanation System is to provide a helpful and meaningful explanation of the output by training on multiple objectives and different combinations of a label and unlabeled data.

With this goal, Multi-Objective A semi-Supervised Explanation System provides us with a solution or more of a specific point from the model which is the best, but most people out in today's world don't need a specific point, they want a two-spaced solution. The solution which they can work around, they can get the best result according to their needs. For example, in Credit Scoring, let's say a person wants to buy a house by taking a loan, but he is unsure about his credibility. So, we need to determine whether he can get a loan or not. To measure his credibility we can use his credit score and his annual income as parameters. Like if his credit score is high and his income is high then his credibility or acceptance of his loan is higher whereas if the credit score is low and his income is low then his credibility or acceptance of his loan is lower. These are the rules we try to create which should work for the whole data. Another example can be a medical issue or symptoms. So, through Multi-Objective A semi-Supervised Explanation System we can provide a two-spaced solution.

From the above example, we understand that a multi-objective semi-Supervised Explanation System needs to provide a two-spaced solution. So, for this project, we were provided a baseline code that has an implementation of sway, which deals in a way that it will find the best cluster from the tree by running down through the tree and creating some rules according to its finding. This rule might work best for the cluster of data it found but need not work for the whole data. As they were created just keeping those best data clusters in mind, whereas the whole data has varied points and might have some outliers. In, order to make it work over the whole data we need to create better rules which work best for the whole data. To overcome these issues we can use many different approaches like Random Search, linear search, Jump search, Exponential search, and many more. Through these search algorithms, we can get the best data cluster and make the best rule which can improve our overall data result.

A. Structure of this paper

There are definitely some important questions to be figured out from this paper such as How do different sampling methods compare in terms of classification performance? How does the performance of the proposed method change as the sampling budget increases? How does the proposed method compare to prior state-of-the-art methods, random selection, and human-level sampling processes?

The suggested approach combines uncertainty and diversity sampling, choosing samples based on their level of uncertainty first, and then using diversity sampling to make sure that the chosen samples are diverse. With a step size of 15, several y-sampling budgets, ranging from 10 to 500, are used to evaluate the sampling techniques. To lessen the impact of unpredictability, each experiment is run 20 times with different random number seeds.

II. RELATED WORK

There are many kinds of research conducted on the Multi-Objective semi-Supervised Explanation System, and there were many approaches that were used some gave good results some did not. This paper, discussing on some of the methods that were used previously like random projection, Semi-supervised learning, and why heuristics work.

The first method is Random projection is a technique that deals with reducing the dimensionality of points that lie in Euclidean space. So, how does this random projection work, before that we need to understand what reducing the dimensionality means. Reducing dimensionality means removing different variables like duplicate values, manipulative data entry, and more, by using different methods and machine learning algorithms, like principal component analysis (PCA), and more. Through this technique of minimizing the dimensionality, it makes the data set smaller which helps the model to run on a much smaller data set and reduces the distance between two points of the data set, which overall makes the processing time faster.

The Second method is heuristic, in the paper why heuristic Work, author Gred Gigerenzer (a German psychologist) compares three central ideas that humans used to learn or train for something. They are Logic, Probability, and heuristic. In his work, he compares why humans tend to overcome a pattern or learn a pattern using logic and probability as their prime resources and use heuristic as number two. He lists some of the misconceptions with heuristic vs logic and probability, some of the misconceptions he found were "Heuristics produce second-best results optimization is always better" or "People with higher cognitive capacities employ complex weighting and integration of information; those with lesser capacities use simple heuristics". He also provides the clarification of these misconceptions first, there can be a situation where optimization is impossible and less accurate as there are too many errors. Overall the main objective of the paper is that even though logic and probability is being developed over the years but no one can beat cognitive heuristics.

Even though there is much research being done on Multi-Objective semi-Supervised Explanation Systems, we still don't completely rely on this work. It is not because they are not good but there are always some or other better ways to do the problem. In the above papers, one of the methods used is Random projection which used different techniques to minimize the data set and reduce the processing time one of the methods we can use is PCA. Even today we use PCA as one of our main methods to make a prediction for a model. But there are some problems that cannot be done using PCA and in the Why heuristic work we discussed the importance of heuristic over logic and probability. But we cannot use this technique every time. Sometimes we need to create a two-spaced solution and this technique does not provide us with it a multi-objective solution it just provides us with a specific solution.

One of the techniques we can use from these papers in our project or in further works is removing manipulative or duplicate data sets which can reduce the size of the model and then we can reduce the processing time. As we discussed above what a Multi-Objective semi-Supervised Explanation System is and how it uses different combinations and try to get the best result out of it (which is accurate and fair), and different combinations of the label, and unlabeled data from the data set provided (which is best and rest data). So, in this, we can factor in a new technique or we can optimize the system by minimizing the data set or getting the best values from the data set only to provide the best result. We can use the Random projection technique which can help in reducing the data set and also will reduce the processing time.

After reading this paper and researching the Multi-Objective semi-Supervised Explanation Systems, we found a way to optimize the code and also reduce the data set. we reduced the dimensionality data set and then we used Random search on the generated data set from reducing the dimensionality and applying random search on the model which will help us get the data cluster which is the best and then make a set of rules which are best for the whole data set.

III. Method

In these project, we have implemented different methods like Multi-objective, Semi-supervised, Explanation, sway, explanation tax, and sampling tax. Each method is being used either to improve the accuracy or to understand the data and the result. The first method we used is :

1) **Multi-objective:** multi-objective which aims to make different combinations and try to get the best result out of it (which is accurate and fair)
2) **Semi-supervised:** Semi-supervised aims to create different combinations of the label, and unlabeled data from the data set provided (which is best and rest data)
3) **Explanation:** In Explanation we try to predict the best data cluster from the given data set. Let's say we have a graph of plus and minus and we need to get the best data set from it we cannot just get the value from the data on the graph. we need to train and understand the data and go through the data and find the best data cluster which can be used. This method we use in Multi-Objective semi-Supervised Explanation System to set the rules we can create a tree and we go the tree and analyze the tree to find the best data cluster this cluster can be used and create a set of rules these rules are then used in the entire data set to get the result.
4) **sway:** In Sway we try to get the best data cluster from the data set. This best cluster is used to create a set of rules for the entire data set.
5) **explanation tax:** In the explanation tax method we find the difference between the xplan and sway. This difference occurs because of using the rules created from the best data cluster from the data and improvising it on the whole data set. this makes the xplan go lower accuracy than sway.
6) **sampling tax:** In the sampling tax we find the difference between the sway and all. This difference occurs because of using the best data cluster from the data and improvising it on the whole data set which is how sway works. And in All, we used all data to get the accuracy which is more than the sway as it uses all data whereas sway uses only the specified amount of data.

We are also using the top as one of our methods, which has the highest accuracy. So, in this project, we were given a baseline code that had an implementation of sway and xplan which had lower accuracy than the top and we had to create a new method for increasing the accuracy but it will be lower than the top method.

The data set of the Multi-Objective semi-Supervised Explanation System is a system project of different kinds, and sizes and also had a different objective. Each file had some objective that either needed to be improved or predict the future. here is the data set we used:

1) **auto2:** Auto2 is a data set of size 93 which consists of information regarding Auto car. The main objective of this data set is to provide an accuracy of having City MPG as positive which means it should increase, Highway MPG as positive which means it should increase Weight as negative which means it should be low, Class as negative which means it should be low.
2) **auto93:** Auto93 is a data set of size 398 which consists of information regarding Auto car. The main objective of this data set is to provide an accuracy of having City MPG as positive which means it should increase, Highway MPG as positive which means it should increase Weight as negative which means it should be low, Class as negative which means it should be low.
3) **china:** China is a data set of size 499 which consists of information regarding software projects and estimation, In this data set, we try to minimize the NEffort and not use EffortX for data training.
4) **coc1000:** coc1000 is a data set of size 1000 which consists of information regarding software projects and estimation, In this data set, we try to minimize the Analyst's Experience, Risk, and Effort and increase the LOC (lines of code).
5) **coc10000:** coc1000 is a data set of size 10000 which consists of information regarding software projects and estimation, In this data set, we try to minimize the Risk, and Effort and increase the LOC (lines of code), this data set is different from coc1000 as we are not trying to minimize the Analyst's Experience.
6) **healthCloseIsses12mths0001-hard:** healthCloseIsses12mths0001-hard is a data set size of 10000 which contains information about issues and close. This data set is used to predict the future of the project that is how many issues will be created and closed for the next 12 months or more. here we are trying to decrease the error of random forest, increase accuracy and decrease or minimize the (PRED).
7) **healthCloseIsses12mths0011-easy:** healthCloseIsses12mths0001-easy is a data set size of 10000 which contains information about issues and close. This data set is used to predict the future of the project that is how many issues will be created and closed for the next 12 months or more. here we are trying to decrease the error of random forest, increase accuracy and decrease or minimize the (PRED). This dataset is named easy because it is easy to optimize than hard.
8) **nasa93dem:** nasa93dem is a data set size of 93 which contains information about software effort+detects estimation In this data set, we try to minimize the months, efforts, and defects while increasing the Kloc.
9) **pom:** pom is a data set size of 10000 which contains information about agile project management In this data set, we try to minimize the cost and idle whereas increasing the completion.
10) **SSM:** SSM is a data set size of 239360 which contains information about computational physics In this data set, we try to minimize the NUMBERITERATIONS and TIMETOSOLUTION.

11) **SSN:** SSN is a data set size of 53662 which contains information about computational physics In this data set, we try to minimize the PSNR and Energy.

We have used Decision Tree Classifier for sway2 and expln2 analysis. Random search have also been implmented. Using the scikit-learn library, this code constructs a decision tree algorithm for categorization. An input dataset is used by the method, which divides it into two groups based on the values of a number of attributes. Then, until a stopping requirement, such as a maximum depth or a minimum number of samples per leaf, is satisfied, the process is repeated for each subgroup.

To implement the decision tree algorithm, the DtreeOptimizer class extends the BaseOptimizer class and overrides its _run function. The data is first preprocessed by using a LabelEncoder from the scikit-learn module to encode categorical characteristics. The best subset, which contains the rows with the greatest target value, and the rest subset, which contains the other rows, are then used to train a decision tree classifier. The rows that are classified as being a part of the best or remainder subset are returned after the trained classifier has been applied to the input data.

When the target variable is binary and the input features are a combination of categorical and numerical variables, this code is helpful for classification jobs. Both types of variables and non-linear relationships between them can be handled by the decision tree method, which is a straightforward and understandable approach. By employing suitable splitting criteria, it can also deal with missing numbers and outliers. A reliable implementation of the technique with numerous hyperparameters that can be adjusted to enhance performance is offered by the scikit-learn library.

Random Search examines a predetermined amount of hyperparameter sets randomly, rather than looking for all of them in the search space. The user chooses this quantity. In comparison to grid search, the method uses less computation and run time because it executes fewer trials for hyperparameter tuning. Unfortunately, the random search runs the danger of omitting the ideal collection of hyperparameters and skipping the peak model performance because it tries hyperparameter sets at random.

For case of sway2, We will rank the data using the Zitzler predicate and normalize the ranks to a scale of 1 to 100 in order to assess the performance of the optimizer. In order to obtain a more precise approximation of the ranking, we will carry out this process 20 times.

The top k elements of the data will then be predicted while working within a strict budget. These top k estimates will all be evaluated, giving us a total of k evaluations. This procedure will be repeated several times, and the distribution of the final ratings will be collected.

We will utilize statistical techniques like effect size and significance testing to assess the effectiveness of various optimizers, and then we will summarize the results. To better comprehend the final assessments' distribution and spot any trends or outliers, we'll also employ cutting-edge visualizations like box plots and scatter plots.

IV. Result

The below results tables are of "auto93.csv" dataset. From the table below we can analyse that the sway2 outperforms sway1. Consider for "Lbs" which should be decreased, sway2 has lower value than sway1, similarly for "Acc" and "Mpg", both the value should be increased and sway2 values are increased as compared to sway1. We also see that the values for sway2 and xplan2 has optimized than pervious sway1 and xplan1 but they can still be improved as top is the upper bound.

	Lbs-	Acc+	Mpg+	Avg evals
all	2800	15.5	20	0
sway1	2088.4	16.3	34.5	6
xpln1	2205.7	15.8	29	6
sway2	2063.9	16.9	35	22.2
xpln2	2111.7	16.6	35	6
top	1987	18.8	40	398

	Best	Beat Sway?	Beat Xpln?
Lbs-	top	True	True
Acc+	top	False	True
Mpg+	top	False	True

	Lbs-	Acc+	Mpg+
all to all	=	=	=
all to sway1	≠	≠	≠
all to sway2	≠	≠	≠
sway1 to sway2	≠	≠	≠
sway1 to xpln1	≠	≠	≠
sway1 to xpln2	≠	≠	≠
xpln1 to xpln2	≠	≠	≠
sway1 to top	≠	≠	≠

Now the above table was genereated by taking rest value as 10, for the rest value of 25, the table below is generated. For the table we can see that the sway2 values are changing in opposite direction than it should change.

	Lbs-	Acc+	Mpg+	Avg evals
all	2800	15.5	20	0
sway1	2101.4	16.4	32.7	6
xpln1	2189.7	15.9	29.3	6
sway2	2105.9	16.5	31.3	25.6
xpln2	2120.5	16.4	32	6
top	1987.7	18.8	40	398

The below tables are generated for "SSN.csv" dataset. From the below table we can see that both the "PSNR" and the "Energy" values should be decreased, and sway2 value are lesser than sway1, which makes the efficiency of sway2 better than sway1. We also see that the values for sway2 and xplan2 has optimized than pervious sway1 and xplan1 but they can still be improved as top is the upper bound.

	PSNR-	Energy-	Avg evals
all	45.7	1237.7	0
sway1	44.5	1222.2	9
xpln1	44.4	1230.9	9
sway2	42.7	989.2	44.5
xpln2	45	1226	9
top	26.7	465.9	53662

	Best	Beat Sway?	Beat Xpln?
PSNR-	top	True	False
Energy-	top	True	True

	PSNR-	Energy-
all to all	=	=
all to sway1	≠	≠
all to sway2	≠	≠
sway1 to sway2	≠	≠
sway1 to xpln1	≠	≠
sway1 to xpln2	≠	≠
xpln1 to xpln2	≠	≠
sway1 to top	≠	≠

The below tables are generated for the ”healthCloseIsses12mths0001-hard.csv” data set. From the below table, we can see that the ”MRE” value should actually decrease and sway2 has a lesser value than sway1, for the case of ”ACC”, the values have to be increased but it actually remains the same for the case of sway2, which means the sway2 is at least as better as sway1. Als for the ”PRED40” sway2 value is higher than the sway1 value, which gives us the correct classification. We also see that the values for sway2 and xplan2 has optimized than pervious sway1 and xplan1 but they can still be improved as top is the upper bound.

	MRE-	ACC+	PRED40+	Avg evals
all	75.3	7.2	25	0
sway1	73.9	7.5	23.8	8
xpln1	73.7	7.6	25	8
sway2	73.7	7.5	25	19.5
xpln2	74.1	7.4	25	8
top	64.9	11.3	25	10000

	Best	Beat Sway?	Beat Xpln?
MRE-	top	True	False
ACC+	top	False	False
PRED40+	all	True	False

	MRE-	ACC+	PRED40+
all to all	=	=	=
all to sway1	≠	≠	≠
all to sway2	≠	≠	≠
sway1 to sway2	≠	≠	≠
sway1 to xpln1	≠	≠	≠
sway1 to xpln2	≠	≠	≠
xpln1 to xpln2	≠	≠	≠
sway1 to top	≠	≠	≠

V. Discussion

In a multi-Objective semi-Supervised Explanation System project we need to improve the code accuracy, but there are threats to improving what if the sway works poorly than the xplan, or the new sway2 and xplan2 work worst than the previous sways? This can be caused if we have implemented the code poorly or if the data set we get to create rules is not proper. This can increase the coded threat. There can also be another problem where we get a data set not equal for each method we implemented. One of the toughest parts of the project would have been the number of iterations as we might have been playing with the iterations if we were not told about the iters to be 20 to at least get a better result.

VI. Conclusion

In the Multi-Objective semi-Supervised Explanation System project we were given sway and xplan as our baseline code. Sway found the best cluster of data from the data set and then it made a set of rules which were used on the whole data set. Our job was to optimize this sway and xplan method and create our own method that can improve the accuracy. we used random search to find the data set and then we made the rules and then we used that whole data set for xplan. we got the result we wanted like when the parameter is need to be increased, it increased in sway2 as well xplan2 from the previous sway and xplan. we knew that our result is correct because we optimized it from the previous sway and xplan and we were lower than the top method. We also run the code with different iters and different rest values. We saw a change in value when iter is approx 10 to 12, whereas in the rest we see a change in value when the rest value is up to 30 -35.

VII. Future work

Multi-Objective semi-Supervised Explanation System is a very broad and vast project as there is a lot of room to improve and optimize the code. In this project we were provided with a baseline code of sway and xplan which worked in a way that a set of rules are created from the best data cluster from the data set which is an implementation of sway and then these rules are improvised and used on the whole data set which is the implementation of xplan, we assume that the xplan will work better but it does not as rules were created from the small data set. Then we implemented the sway2 and xplan2 which improved the accuracy but still there is room for improvement as the upper bound of the implementation was top.

For the future implementation we can improve the rules by getting the best data set, we used random search in sway2 but we can use Genetic Algorithm, what this will do is it will draw their inspiration from natural selection and evolution. By iteratively evolving the population through crossover (recombination) and mutation techniques, they mimic the process of genetic reproduction using a population of potential solutions. The ability of genetic algorithms is to solve complex and large-scale problems. Other algorithms that can be used are Simulated Annealing, Particle Swarm Optimization, Gradient-Based Methods, and many more.

Implementing this method will get the best data cluster from the data set and then we can create the best rules which can be used on the whole data set. Another thing we can improve is the number of iterations as the project requirement was about 20 iters but if the time would have been there we could have played with the iters to find a relationship and also find the best iters to find find the best accuracy.

Acknowledgment

The authors would like to thank you Professor Dr. Tim Menzies, TAs of the class Andre Lustosa , Sherry (Xueqi) Yang , and Rahul Yedida for giving us a wonderful opportunity to work on Multi-Objective semi-Supervised Explanation System project and the course.

References

[1] H. Kopka and P. W. Daly, *A Guide to LaTeX*, 3rd ed. Harlow, England: Addison-Wesley, 1999.

[2] Bro¨der, A., Schiffer, S. (2006). Adaptive flexibility and maladaptive routine in selecting fast and frugal decision strategies. Journal of Experimental Psychology: Learning, Memory, and Cognition, 32, 904–918.

[3] G. Gigerenzer, “Why Heuristics Work,” 2008. Available: http://library.mpib-berlin.mpg.de/ft/gg/GGWhy2008.pdf

[4] O. Chapelle, B. Schölkopf, and A. Zien, "Semi-Supervised Learning edited by." Available: https://www.molgen.mpg.de/3659531/MITPress–SemiSupervised-Learning.pdf

[5] "Random projection," Wikipedia, Feb. 08, 2021. https://en.wikipedia.org/wiki/Randomprojection

[6] Yee, M., Hauser, J., Orlin, J., Dahan, E. (2007). Greedoid-based non-compensatory two-stage consideration-then-choice infer- ence. Marketing Science, 26, 532–549.

[7] Martignon, L., Vitouch, O., Takezawa, M., Forster, M.R. (2003). Naive and yet enlightened: From natural frequencies to fast and frugal decision trees. In D. Hardman L. Macchi (Eds.), Thinking: Psychological perspectives on reasoning, judgment and decision making (pp. 189–211). Chichester, United Kingdom: Wiley.

[8] DeMiguel, V., Garlappi, L., Uppal, R. (2006, August). 1/N. Paper presented at the 2006 European Financial Association Meeting, Zurich, Switzerland. Retrieved November 7, 2007, from http://ssrn.com/abstract=911512

[9] Gigerenzer, G., Swijtink, Z., Porter, T., Daston, L., Beatty, J., Kru ger, L. (1989). The empire of chance: How probability changed science and everyday life. Cambridge, United Kingdom: Cambridge University Press.

[10] Cosmides, L., Tooby, J. (1992). Cognitive adaptations for social exchange. In J.H. Barkow, L. Cosmides, J. Tooby (Eds.), The adapted mind: Evolutionary psychology and the generation of culture (pp. 163–228). New York: Oxford University Press.