Machine Learning Models and Methodologies

CMPE-258 Deep Learning - Short Story

Tripura Chandana Gayatri Gorla

Introduction

Sophisticated statistical models are believed to often bring improved accuracies and efficiencies. But, due to their non-interpretable nature of outputs, they are not very much used by organizations, institutions and governments. They are hence named Black-Boxes. Model interpretability is desired in practical world problems where decisions can have a huge impact (eg. criminal justice, estimating credit scores, health risks etc). Here novel methods that form the state-of-the-art for addressing this particular problem by trying to give a guide to practitioners for appropriate methods to their problems.

Local Vs Global Models

In case of very complex models, the scope of local model is restricted to only a particular neighborhood and the best case prediction is determined. In contrast, global models aim at understanding the whole model and hence these atm at understanding how the features affect the result rather than the interpretability.

Model Agnostic Methods

(i) Model flexibility

(ii) Explanation flexibility

(iii) Representation flexibility

Pros are (I) flexibility (II) Compatibility and cons being (I) Time Consuming, (II) Sampling variability

Model-Agnostic Method Approaches
(A) Perturbation Approach:

(i) Partial Dependence Plots (PDP)

(ii) Individual Conditional Expectation (ICE)

(iii) M-Plots

(iv) Accumulated Local effects (ALE)

(v) Shapley Values (SHAP)

(vi) LOCO

(vii) Decomposition of predictor

(viii) Feature Importance

(ix) Sensitive Analysis

(x) LIME

(B) Contrastive Approach :

(i) Counterfactuals Naturally Observed

(ii) Prototype and Criticism

(iii) Justified Counterfactual Explanations

Model specific fields :
(A) Machine vision models

(i) Masks

(ii) Real Time Saliency Maps

(iii) Smooth Grad

(iv) Layer wise Relevant Propagation

(v) Heat Maps

(B) General Neural Networks

(i) Differentiable Models

(ii) DeepLIFT

(iii) Taylor decomposition

(iv) Integrated Gradients

(v) I-GOS

(vi) Grad-cam

(C) Decision Tree Methods

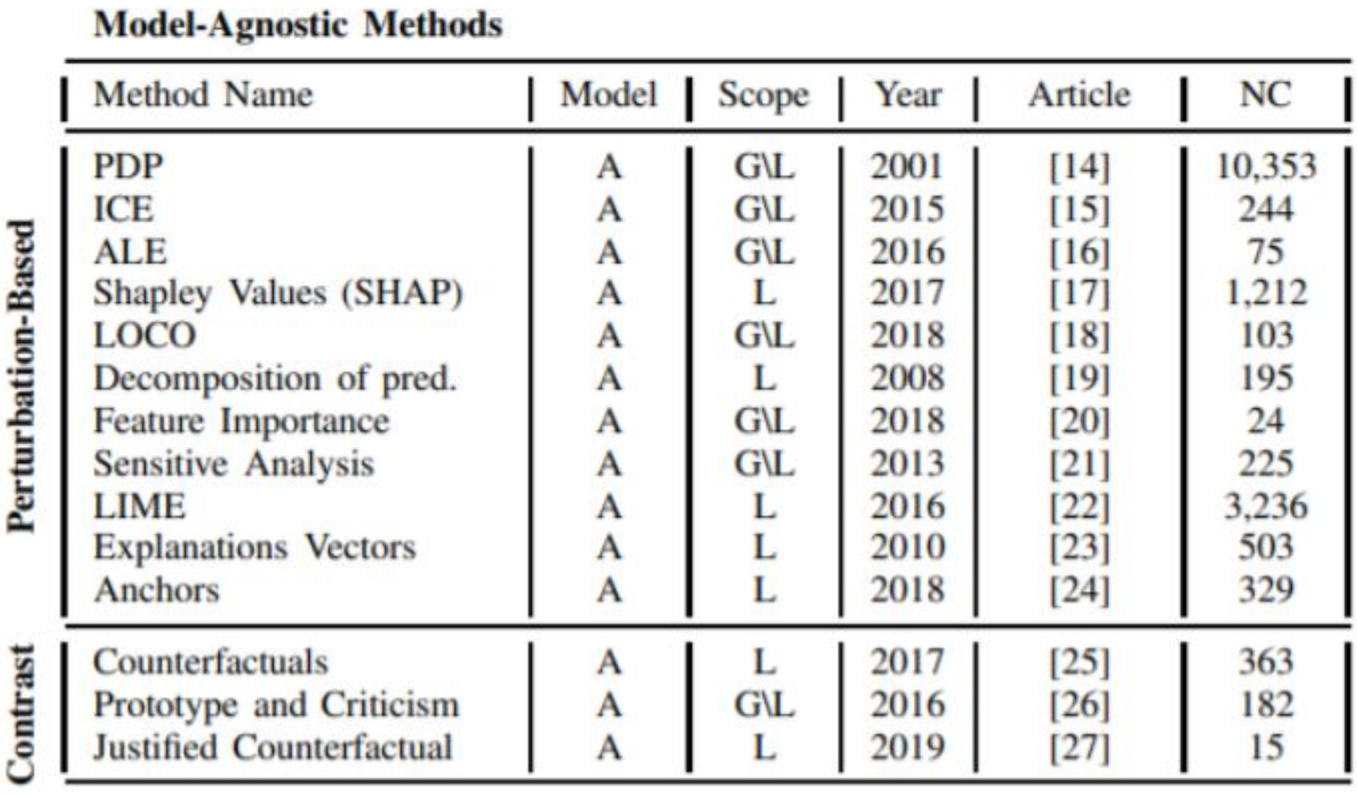
Tree Explainer —

1. Reporting the decision path
2. Assigning the contribution of individual feature
3. Applying model agnostic approach

Limitations:

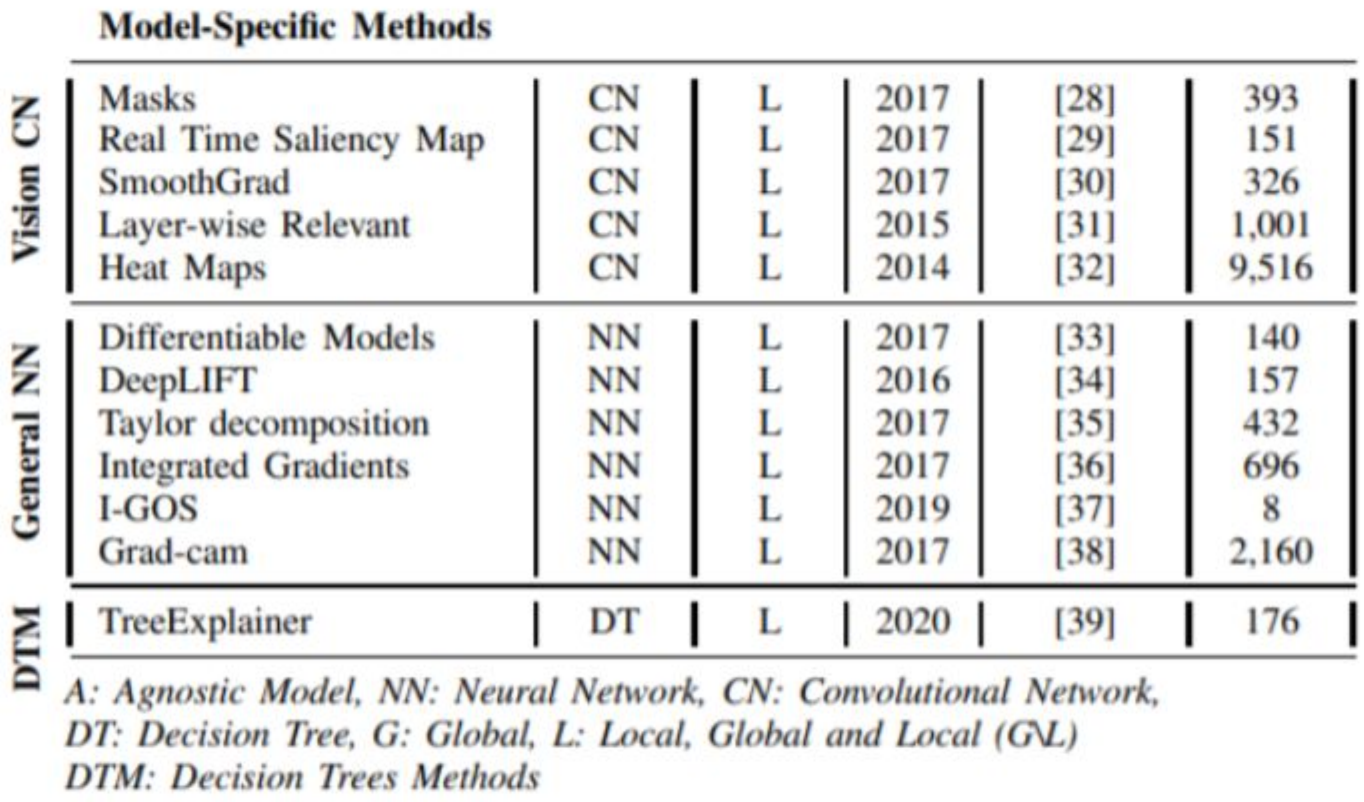
1. Not useful when the model utilizes multiple trees for final prediction
2. Explanation might be biased
3. Might be slow and suffer sampling variability

Table I: Survey's discussed methods



Model-Agnostic Methods

	Method Name	Model	Scope	Year	Article	NC
Perturbation-Based	PDP	A	G\L	2001	[14]	10,353
	ICE	A	G\L	2015	[15]	244
	ALE	A	G\L	2016	[16]	75
	Shapley Values (SHAP)	A	L	2017	[17]	1,212
	LOCO	A	G\L	2018	[18]	103
	Decomposition of pred.	A	L	2008	[19]	195
	Feature Importance	A	G\L	2018	[20]	24
	Sensitive Analysis	A	G\L	2013	[21]	225
	LIME	A	L	2016	[22]	3,236
	Explanations Vectors	A	L	2010	[23]	503
	Anchors	A	L	2018	[24]	329
Contrast	Counterfactuals	A	L	2017	[25]	363
	Prototype and Criticism	A	G\L	2016	[26]	182
	Justified Counterfactual	A	L	2019	[27]	15



Model-Specific Methods

Vision CN	Masks	CN	L	2017	[28]	393
	Real Time Saliency Map	CN	L	2017	[29]	151
	SmoothGrad	CN	L	2017	[30]	326
	Layer-wise Relevant	CN	L	2015	[31]	1,001
	Heat Maps	CN	L	2014	[32]	9,516
General NN	Differentiable Models	NN	L	2017	[33]	140
	DeepLIFT	NN	L	2016	[34]	157
	Taylor decomposition	NN	L	2017	[35]	432
	Integrated Gradients	NN	L	2017	[36]	696
	I-GOS	NN	L	2019	[37]	8
	Grad-cam	NN	L	2017	[38]	2,160
DTM	TreeExplainer	DT	L	2020	[39]	176

A: Agnostic Model, NN: Neural Network, CN: Convolutional Network,
DT: Decision Tree, G: Global, L: Local, Global and Local (G\L)
DTM: Decision Trees Methods

Conclusion

Relevant and Novel approaches were reviewed in this survey which gives light to the problem of explaining individual instances in Machine learning. Explaining the model prediction has become increasingly desirable as the trend of using the highly complex models for the explanation of algorithms has spread. Some of the interpretation models use natural language while others use visualization of models or learned representations. The methods are divided based on Model specific approach and Model agnostic approach. Model Agnostic approach can be used on any type of Machine Learning model. While, the Model Specific approach can be applied to only a particular group of models. Model Agnostic approach was sub-classified by taxonomy into SHAP and LIME. Model Specific approach was sub-classified into Computational Neural Networks, General Neural Networks and Decision Trees. Recently this family of Tree approach has out-performed the Neural networks.