ITEA automatic report

ITEA_summarizer

Sunday 13th June, 2021, 18:28

Automatic report created by *ITEA_summarizer* package. This report makes usage of several methods to automatically inspect and explain the final expression found in the evolutionary process performed by the ITEA algorithm.

Descriptive statistics of the data

Showing descriptive statistics for 5 (from a total of 8) features contained on the given data. The features were selected based on the absolute final importance.

	AveBedrms	MedInc	Latitude	Longitude	AveOccup
count	13828.000000	13828.000000	13828.000000	13828.000000	13828.000000
mean	1.097533	3.876745	35.651238	-119.585098	3.128660
std	0.445688	1.903102	2.134064	2.005127	12.646130
min	0.333333	0.499900	32.550000	-124.350000	0.692308
25%	1.006623	2.568575	33.940000	-121.810000	2.432189
50%	1.049552	3.538750	34.270000	-118.510000	2.819702
75%	1.100283	4.756600	37.720000	-118.010000	3.282093
max	25.636364	15.000100	41.950000	-114.310000	1243.333333

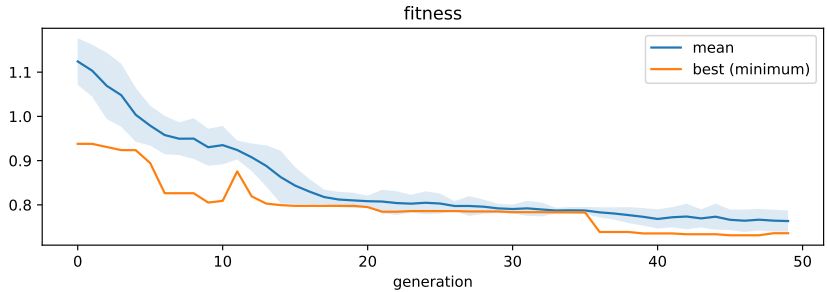
Algorithm Hyper-parameters

The following hyperparameters were used to execute the algorithm. If the random_state parameter was set to a integer value, then it is possible to repeat the exact execution by using the parameters listed below.

```
expolim : (0, 2)
gens : 50
max_terms : 5
popsize : 50
random_state : 42
simplify_method : simplify_by_var
verbose : 10
tfuncs : [log, sqrt.abs, id, sin, cos, exp]
```

Evolution convergence

The algorithm took 94.871 seconds to completely run. Below are the plots for the average fitness of the population and the best individual fitness for each generation.



Best expression

The best expression corresponds to the expression with best fitness on the last generation before the evolution ends. Not necessarily it will be the simplier or the global optimum of the evoution. The final expression is a regressor with fitness of 0.73588, and the number of IT terms is 5. Below is an representation of the expression:

$$
\begin{aligned}
ITExpr = & \underbrace{\beta_0 \cdot log(AveRooms^2 \cdot Population^2 \cdot AveOccup^2 \cdot Latitude^2 \cdot Longitude^2)}_{\text{term 0}} \\
& + \underbrace{\beta_1 \cdot log(MedInc \cdot HouseAge^2 \cdot AveRooms^2 \cdot AveBedrms \cdot Population^2 \cdot AveOccup^2 \cdot Latitude \cdot Long}_{\text{term 1}} \\
& + \underbrace{\beta_2 \cdot log(MedInc^2 \cdot AveRooms^2 \cdot AveBedrms^2 \cdot AveOccup^2 \cdot Latitude \cdot Longitude^2)}_{\text{term 2}} \\
& + \underbrace{\beta_3 \cdot log(MedInc^2 \cdot HouseAge \cdot AveRooms^2 \cdot AveBedrms^2 \cdot Population^2 \cdot AveOccup^2 \cdot Latitude^2 \cdot Lo}_{\text{term 3}} \\
& + \underbrace{\beta_4 \cdot log(MedInc^2 \cdot HouseAge^2 \cdot AveRooms^2 \cdot AveBedrms^2 \cdot Population^2 \cdot AveOccup^2 \cdot Latitude^2)}_{\text{term 4}} \\
& + I_0
\end{aligned}
\tag{1}
$$

Best expression metrics

In the next page is reported a table containing the coefficients for the previous expression, as well as some metrics calculated for each term individually:

- **coef:** coefficient of each term (or coefficients, if the itexpr is an instance of IT-Expr_classifier);
- **coef stderr:** the standard error of the coefficients;
- **disentang.:** mean pairwise disentanglement between each term when compared with the others;
- **M.I.:** mean continuous mutual information between each term when compared with the others;
- **pred. var.:** variance of the predicted outcomes for each term when predicting the training data.

	coef	func	coef stderr	disentang.	M.I.	pred. var.
term 0	-8.684	log	0.147	0.531	0.482	217.871
term 1	14.821	log	0.304	0.657	0.796	767.421
term 2	-0.402	log	0.011	0.300	0.327	0.352
term 3	17.139	log	0.296	0.746	0.866	1176.852
term 4	-23.274	log	0.453	0.710	0.917	2301.542
term 5	-163.347	intercept	3.271	0.000	0.000	0.000

Partial derivatives

$$\frac{\partial}{\partial MedInc} ITExpr \tag{2}$$
$$= \underbrace{1\beta_1 \cdot log'(MedInc \cdot HouseAge^2 \cdot AveRooms^2 \cdot AveBedrms \cdot Population^2 \cdot AveOccup^2 \cdot Latitude \cdot Longitude^2)(H}_{\text{term 1}}$$
$$+ \underbrace{2\beta_2 \cdot log'(MedInc^2 \cdot AveRooms^2 \cdot AveBedrms^2 \cdot AveOccup^2 \cdot Latitude \cdot Longitude^2)(MedInc \cdot AveRooms^2 \cdot}_{\text{term 2}}$$
$$+ \underbrace{2\beta_3 \cdot log'(MedInc^2 \cdot HouseAge \cdot AveRooms^2 \cdot AveBedrms^2 \cdot Population^2 \cdot AveOccup^2 \cdot Latitude^2 \cdot Longitud}_{\text{term 3}}$$
$$+ \underbrace{2\beta_4 \cdot log'(MedInc^2 \cdot HouseAge^2 \cdot AveRooms^2 \cdot AveBedrms^2 \cdot Population^2 \cdot AveOccup^2 \cdot Latitude^2)(MedInc}_{\text{term 4}}$$

$$\frac{\partial}{\partial HouseAge} ITExpr \tag{3}$$
$$= \underbrace{2\beta_1 \cdot log'(MedInc \cdot HouseAge^2 \cdot AveRooms^2 \cdot AveBedrms \cdot Population^2 \cdot AveOccup^2 \cdot Latitude \cdot Longitude^2)(M}_{\text{term 1}}$$
$$+ \underbrace{1\beta_3 \cdot log'(MedInc^2 \cdot HouseAge \cdot AveRooms^2 \cdot AveBedrms^2 \cdot Population^2 \cdot AveOccup^2 \cdot Latitude^2 \cdot Longitud}_{\text{term 3}}$$
$$+ \underbrace{2\beta_4 \cdot log'(MedInc^2 \cdot HouseAge^2 \cdot AveRooms^2 \cdot AveBedrms^2 \cdot Population^2 \cdot AveOccup^2 \cdot Latitude^2)(MedInc}_{\text{term 4}}$$

$$\frac{\partial}{\partial AveRooms} ITExpr \tag{4}$$

$$= \underbrace{2\beta_0 \cdot log'(AveRooms^2 \cdot Population^2 \cdot AveOccup^2 \cdot Latitude^2 \cdot Longitude^2)(AveRooms \cdot Population^2 \cdot AveOccup^2}_{\text{term 0}}$$

$$+ \underbrace{2\beta_1 \cdot log'(MedInc \cdot HouseAge^2 \cdot AveRooms^2 \cdot AveBedrms \cdot Population^2 \cdot AveOccup^2 \cdot Latitude \cdot Longitude^2)}_{\text{term 1}}$$

$$+ \underbrace{2\beta_2 \cdot log'(MedInc^2 \cdot AveRooms^2 \cdot AveBedrms^2 \cdot AveOccup^2 \cdot Latitude \cdot Longitude^2)(MedInc^2 \cdot AveRooms \cdot}_{\text{term 2}}$$

$$+ \underbrace{2\beta_3 \cdot log'(MedInc^2 \cdot HouseAge \cdot AveRooms^2 \cdot AveBedrms^2 \cdot Population^2 \cdot AveOccup^2 \cdot Latitude^2 \cdot Longitud}_{\text{term 3}}$$

$$+ \underbrace{2\beta_4 \cdot log'(MedInc^2 \cdot HouseAge^2 \cdot AveRooms^2 \cdot AveBedrms^2 \cdot Population^2 \cdot AveOccup^2 \cdot Latitude^2)(MedInc}_{\text{term 4}}$$

$$\frac{\partial}{\partial AveBedrms} ITExpr \tag{5}$$
$$= \underbrace{1\beta_1 \cdot log'(MedInc \cdot HouseAge^2 \cdot AveRooms^2 \cdot AveBedrms \cdot Population^2 \cdot AveOccup^2 \cdot Latitude \cdot Longitude^2)(M}_{\text{term 1}}$$
$$+ \underbrace{2\beta_2 \cdot log'(MedInc^2 \cdot AveRooms^2 \cdot AveBedrms^2 \cdot AveOccup^2 \cdot Latitude \cdot Longitude^2)(MedInc^2 \cdot AveRooms^2}_{\text{term 2}}$$
$$+ \underbrace{2\beta_3 \cdot log'(MedInc^2 \cdot HouseAge \cdot AveRooms^2 \cdot AveBedrms^2 \cdot Population^2 \cdot AveOccup^2 \cdot Latitude^2 \cdot Longitud}_{\text{term 3}}$$
$$+ \underbrace{2\beta_4 \cdot log'(MedInc^2 \cdot HouseAge^2 \cdot AveRooms^2 \cdot AveBedrms^2 \cdot Population^2 \cdot AveOccup^2 \cdot Latitude^2)(MedInc}_{\text{term 4}}$$

$$\frac{\partial}{\partial Population} ITExpr \tag{6}$$
$$= \underbrace{2\beta_0 \cdot log'(AveRooms^2 \cdot Population^2 \cdot AveOccup^2 \cdot Latitude^2 \cdot Longitude^2)(AveRooms^2 \cdot Population \cdot AveOccup}_{\text{term 0}}$$
$$+ \underbrace{2\beta_1 \cdot log'(MedInc \cdot HouseAge^2 \cdot AveRooms^2 \cdot AveBedrms \cdot Population^2 \cdot AveOccup^2 \cdot Latitude \cdot Longitude^2}_{\text{term 1}}$$
$$+ \underbrace{2\beta_3 \cdot log'(MedInc^2 \cdot HouseAge \cdot AveRooms^2 \cdot AveBedrms^2 \cdot Population^2 \cdot AveOccup^2 \cdot Latitude^2 \cdot Longitud}_{\text{term 3}}$$
$$+ \underbrace{2\beta_4 \cdot log'(MedInc^2 \cdot HouseAge^2 \cdot AveRooms^2 \cdot AveBedrms^2 \cdot Population^2 \cdot AveOccup^2 \cdot Latitude^2)(MedInc}_{\text{term 4}}$$

$$\frac{\partial}{\partial AveOccup} ITExpr \tag{7}$$

$$= \underbrace{2\beta_0 \cdot log'(AveRooms^2 \cdot Population^2 \cdot AveOccup^2 \cdot Latitude^2 \cdot Longitude^2)(AveRooms^2 \cdot Population^2 \cdot AveOccu}_{\text{term 0}}$$

$$+ \underbrace{2\beta_1 \cdot log'(MedInc \cdot HouseAge^2 \cdot AveRooms^2 \cdot AveBedrms \cdot Population^2 \cdot AveOccup^2 \cdot Latitude \cdot Longitude^2)}_{\text{term 1}}$$

$$+ \underbrace{2\beta_2 \cdot log'(MedInc^2 \cdot AveRooms^2 \cdot AveBedrms^2 \cdot AveOccup^2 \cdot Latitude \cdot Longitude^2)(MedInc^2 \cdot AveRooms^2}_{\text{term 2}}$$

$$+ \underbrace{2\beta_3 \cdot log'(MedInc^2 \cdot HouseAge \cdot AveRooms^2 \cdot AveBedrms^2 \cdot Population^2 \cdot AveOccup^2 \cdot Latitude^2 \cdot Longitud}_{\text{term 3}}$$

$$+ \underbrace{2\beta_4 \cdot log'(MedInc^2 \cdot HouseAge^2 \cdot AveRooms^2 \cdot AveBedrms^2 \cdot Population^2 \cdot AveOccup^2 \cdot Latitude^2)(MedInc}_{\text{term 4}}$$

$$\frac{\partial}{\partial Latitude} ITExpr \tag{8}$$
$$= \underbrace{2\beta_0 \cdot log'(AveRooms^2 \cdot Population^2 \cdot AveOccup^2 \cdot Latitude^2 \cdot Longitude^2)(AveRooms^2 \cdot Population^2 \cdot AveOccu}_{\text{term 0}}$$
$$+ \underbrace{1\beta_1 \cdot log'(MedInc \cdot HouseAge^2 \cdot AveRooms^2 \cdot AveBedrms \cdot Population^2 \cdot AveOccup^2 \cdot Latitude \cdot Longitude^2)}_{\text{term 1}}$$
$$+ \underbrace{1\beta_2 \cdot log'(MedInc^2 \cdot AveRooms^2 \cdot AveBedrms^2 \cdot AveOccup^2 \cdot Latitude \cdot Longitude^2)(MedInc^2 \cdot AveRooms^2}_{\text{term 2}}$$
$$+ \underbrace{2\beta_3 \cdot log'(MedInc^2 \cdot HouseAge \cdot AveRooms^2 \cdot AveBedrms^2 \cdot Population^2 \cdot AveOccup^2 \cdot Latitude^2 \cdot Longitud}_{\text{term 3}}$$
$$+ \underbrace{2\beta_4 \cdot log'(MedInc^2 \cdot HouseAge^2 \cdot AveRooms^2 \cdot AveBedrms^2 \cdot Population^2 \cdot AveOccup^2 \cdot Latitude^2)(MedInc}_{\text{term 4}}$$

$$\frac{\partial}{\partial Longitude} ITExpr \tag{9}$$

$$= \underbrace{2\beta_0 \cdot log'(AveRooms^2 \cdot Population^2 \cdot AveOccup^2 \cdot Latitude^2 \cdot Longitude^2)(AveRooms^2 \cdot Population^2 \cdot AveOccu}_{\text{term 0}}$$

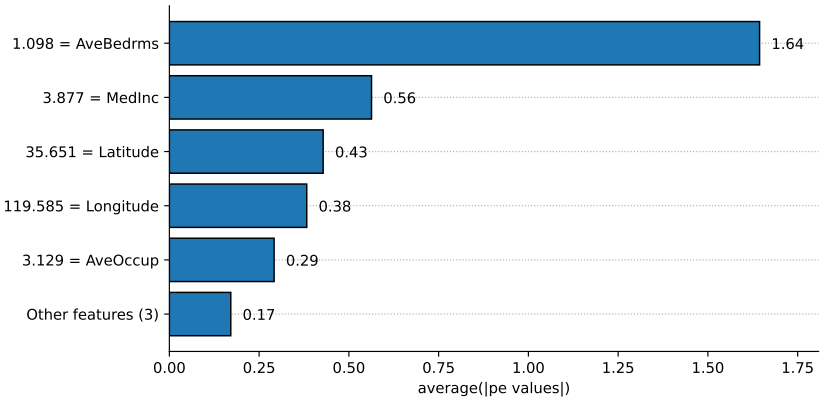
$$+ \underbrace{2\beta_1 \cdot log'(MedInc \cdot HouseAge^2 \cdot AveRooms^2 \cdot AveBedrms \cdot Population^2 \cdot AveOccup^2 \cdot Latitude \cdot Longitude^2)}_{\text{term 1}}$$

$$+ \underbrace{2\beta_2 \cdot log'(MedInc^2 \cdot AveRooms^2 \cdot AveBedrms^2 \cdot AveOccup^2 \cdot Latitude \cdot Longitude^2)(MedInc^2 \cdot AveRooms^2}_{\text{term 2}}$$

$$+ \underbrace{2\beta_3 \cdot log'(MedInc^2 \cdot HouseAge \cdot AveRooms^2 \cdot AveBedrms^2 \cdot Population^2 \cdot AveOccup^2 \cdot Latitude^2 \cdot Longitud}_{\text{term 3}}$$

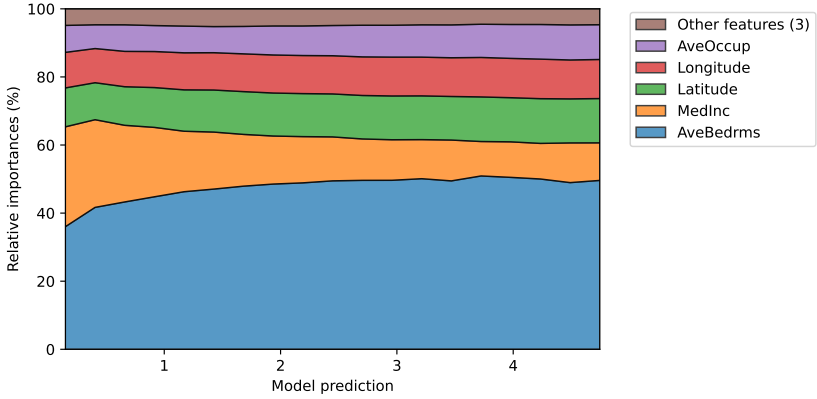
Average partial Effects

Feature importances with Average Partial Effects. This method attributes the importance to the i-th variable by calculating the average of the partial derivative w.r.t. i, evaluated for all data in the training set.



Normalized partial Effects

Feature importances with Normalized Partial Effects. To create this plot, first, the output interval is dicretized. Then, for each interval, the partial effect of all samples in the training set that results in an prediction within the interval are calculated. Finally, they are normalized in order to make the total contribution be 100%.



Partial Effects at the Means

Partial Effects plots created by fixing the co-variables at the means and evaluating the model's output when only one variable changes. For simplicity, at most 5 variables are selected to create the plot (the 5 most important variables considering their Average Partial Effects).