2022年10月26日 11:05

Update since Sep 29th

- 1. Considering that there is always **unbalance in junctions with accidents and junctions without accidents**, two analysis are further carried out to reduce the quantity of junctions without accidents.
 - a. Reset the searching radius of each junction for road accidents from 20m to 30m, to have more junctions with accidents. I think this is acceptable, because in the real driving scenario, drivers have to make response to junctions earlier before they arrive junctions. Some accidents can also happen in this progress.
 - b. Delete dead ends from the current junction set,

 These dead ends are ignored in previous analysis. In fact there are few accidents happen in dead ends and they don't intersect/ connect with other roads. Now the total count of junctions are 1877(2303 for previous version).
- 2. Add a new **junction dimension**(and a bunch of related variables) in classifiers. Details can be found in the graph below.

Maxspeed of road, crossing, and traffic light data are collected from osm; Road level data are collected from Ordnance Survey road centre line

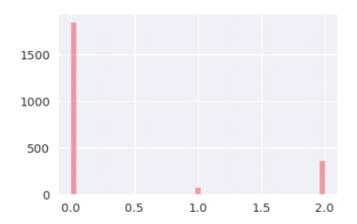
Just think a junction dimension can be necessary for a study on junction classification?

			Macro
		Meso	Ţ
		Micro	
Built Environment	Activities	Road Function	Road System
Street View	POI	Junction Info	Space syntax
'Building',	 	'Coordinates',	
'Sky',	'culture', 'food_drink',	'Minimum distance to other nearest junctions',	'NACHr800m_', 'NACHr1600m',
'Tree',	'hotel',	'count of road connected',	'NACHr2400m', 'NACHr3200m',
'Road',	'school',	'constitution of level of	'NACHr4000m',
'Sidewalk',	'shop_mall,	roads connected',	'NAINr800m
'Fence',		'Maximum maxspeed',	'NAINr1600m',
'TrafficLight',	'pub_bar',	'with_crossing',	'NAINr2400m', 'NAINr3200m',
		'with_traffic_light'	'NAINr4000m',

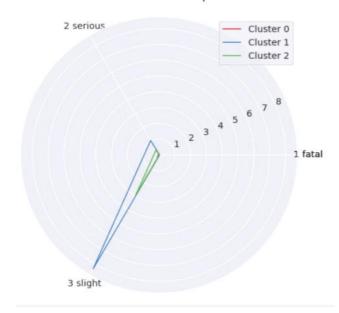
3. Test the k-means clustering method

Have tested the k-means clustering method and the count of slight/serious/fatal accidents are used as three dimensions.

- a. The main difference between clusters is the total count of accidents
- b. Significant difference in sample count included in each clusters
- c. Randomforest classifier receive unbalanced result. Poor performance on severe/fatal accident featured clusters.



Cluster centroid comparison



Without SMOTE

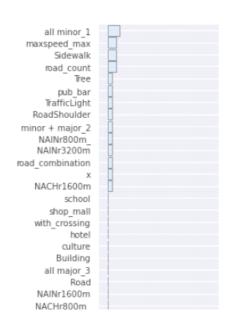
	precision	recall	f1-score	support
0	0.87	0.94	0.90	140
1	0.25	0.09	0.13	11
2	0.58	0.51	0.54	37
accuracy			0.80	188
macro avg	0.56	0.51	0.53	188
weighted avg	0.77	0.80	0.79	188

SMOTE + Without tuning

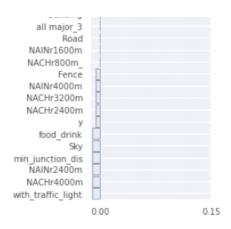
	precision	recall	f1-score	support
0	0.91	0.91	0.91	140
1	0.45	0.45	0.45	11
2	0.56	0.54	0.55	37
accuracy			0.81	188
macro avg	0.64	0.64	0.64	188
weighted avg	0.81	0.81	0.81	188

SMOTE + Tuning

SMOTE + Tuning Feature Importance



0					
SMOTE + Tuning					
SIVIOTE : Turning					
	precision	nacall	f1-score	support	
	precision	recuii	11-30010	Suppor c	
	0.01	0.00	0.00	4.40	
0	0.91	0.89	0.90	140	
1	0.45	0.45	0.45	11	
2	0.54	0.57	0.55	37	
accuracy			0.80	188	
macro avg	0.63	0.64	0.64	188	
weighted avg	0.81	0.80	0.81	188	



4. Classification method

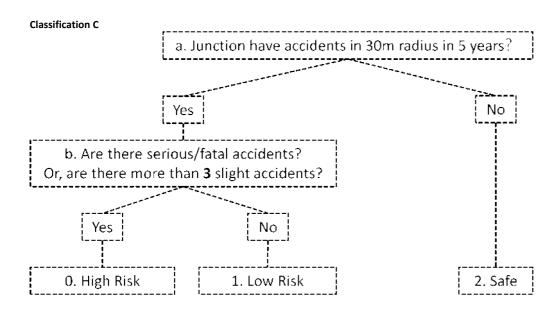
Have tested the classification C with new variables and with more balanced dataset

- a. The classification precision and recall for high risk junctions have increased compared to previous version.
- b. For safe and low risk junctions, seems there is a lack of supporting evidence.
- c. Poor performance in classifying low-risk junctions

Have raised another classification method E.

The **method E** also consider accidents happened on road segments(beyond the junction) and provide more rules to define low risk and safe junctions.

- a. Good performance in classifying high-risk and low risk juncitons and relatively low performance in safe juncitons
- b. Relatively more balanced performance for the overall model.



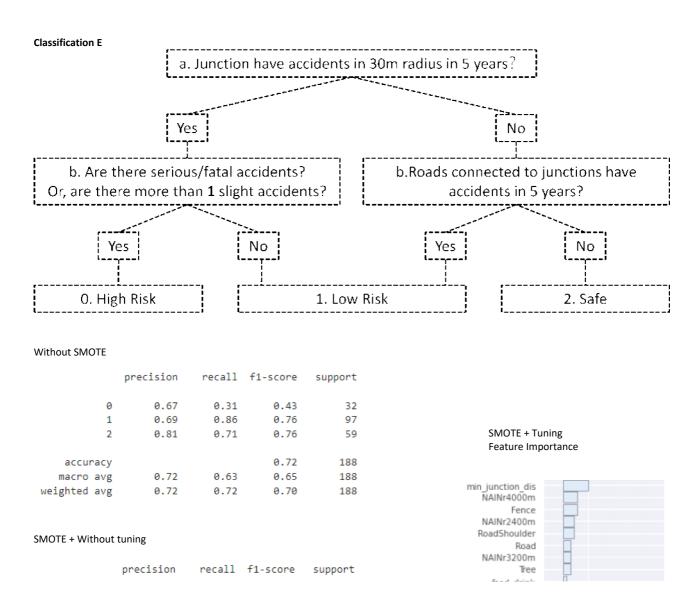
Without SMOTE

	precision	recall	f1-score	support
0	0.80	0.86	0.83	100
1	0.31	0.26	0.28	39
2	0.71	0.71	0.71	49
accuracy			0.70	188
macro avg	0.61	0.61	0.61	188
weighted avg	0.68	0.70	0.69	188

SMOTE + Tuning Feature Importance

-	
NACHr1600m	
all minor_1	
Tree	
NAINr3200m	
NAINr800m	

accumacy			0.70	188	all minor_1			
accuracy					Tree			
macro avg	0.61	0.61	0.61	188	NAINr3200m			
weighted avg	0.68	0.70	0.69	188	NAINr800m_			
					Sky			
					NAINr4000m			
SMOTE + Without	tuning				Fence			
	J				NAINr2400m			
	precision	recall	f1-score	support	with_crossing NACHr800m			
					min junction dis	H		
0	0.80	0.86	0.83	100	NACHr3200m			
1	0.31	0.26	0.28	39	with_traffic_light			
2	0.71	0.71	0.71	49	maxspeed_max			
					road_count NAINr1600m	H		
accuracy			0.70	188	×			
macro avg	0.61	0.61	0.61	188	minor + major 2			
weighted avg	0.68	0.70	0.69	188	NACHr4000m			
					school			
					pub_bar			
					у			
SMOTE + Tuning					hotel			
SIVIOTE + Turning					culture			
	precision	225211	f1-score	support	all major_3			
	precision	Lecall	11-Score	Support	TrafficLight			
					NACHr2400m			
0	0.80	0.87	0.83	100	road_combination	Н		
1	0.37	0.28	0.32	39	shop_mall	Н		
2	0.73	0.73	0.73	49	Road	Н		
					food_drink Sidewalk			
accuracy			0.71	188	RoadShoulder	П		
macro avg		0.63	0.63	188	Building			
weighted avg		0.71	0.70	188		0.00	0.04	0.55
MerBuren av8	0.09	0.71	0.76	100		0.00	0.04	0.15



SMOTE + Without tuning							
	precision	recall	f1-score	support			
0	0.54	0.59	0.57	32			
1	0.70	0.70	0.70	97			
2	0.77	0.73	0.75	59			
accuracy			0.69	188			
macro avg	0.67	0.67	0.67	188			
weighted avg	0.70	0.69	0.69	188			
SMOTE + Tuning							
	precision	recall	f1-score	support			
0	0.54	0.59	0.57	32			
1	0.70	0.71	0.70	97			
2	0.78	0.71	0.74	59			
accuracy			0.69	188			
macro avg	0.67	0.67	0.67	188			
weighted avg	0.70	0.69	0.69	188			

