BUSINESS INFORMATION SYSTEMS 2019

Grant number:
788186

Hyperparameter Optimisation for Improving Classification under Class Imbalance

Jiawen Kong

Universiteit Leiden
The Netherlands

BUSINESS INFORMATION SYSTEMS 2019

Grant number: 788186

Contents:

BUSINESS INFORMATION SYSTEMS 2019

EU MSCA-ITN project

Project introduction

- University: Leiden University, the Netherlands
- Topic: Class imbalance classification through semi-supervised and active learning
- Funding: EU Marie Skłodowska-Curie Actions (MSCA) ITN project ECOLE (grant NO. 788186)
- Cooperation: University of Birmingham, Honda Research Institute (HRI) Europe, NEC Europe

ECOLE
Experience-based Computation:
Learning to Optimise

Universiteit Leiden
The Netherlands

UNIVERSITY OF BIRMINGHAM

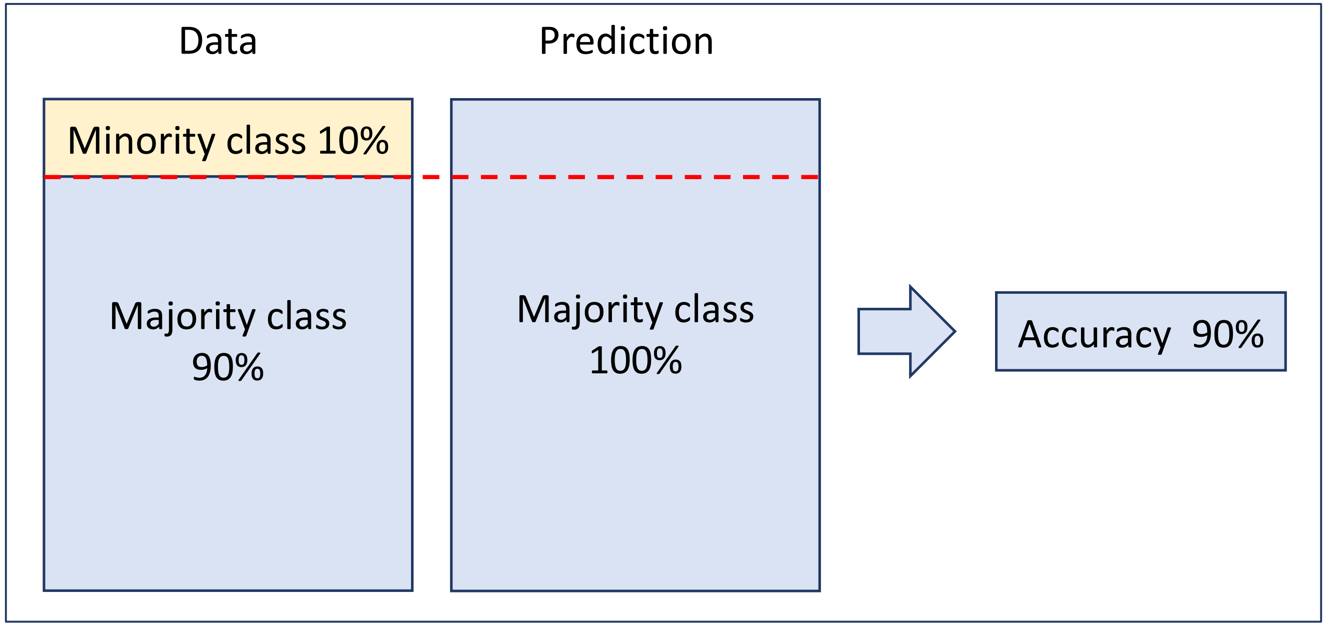
HRI
Honda Research Institute EU

NEC

Research background

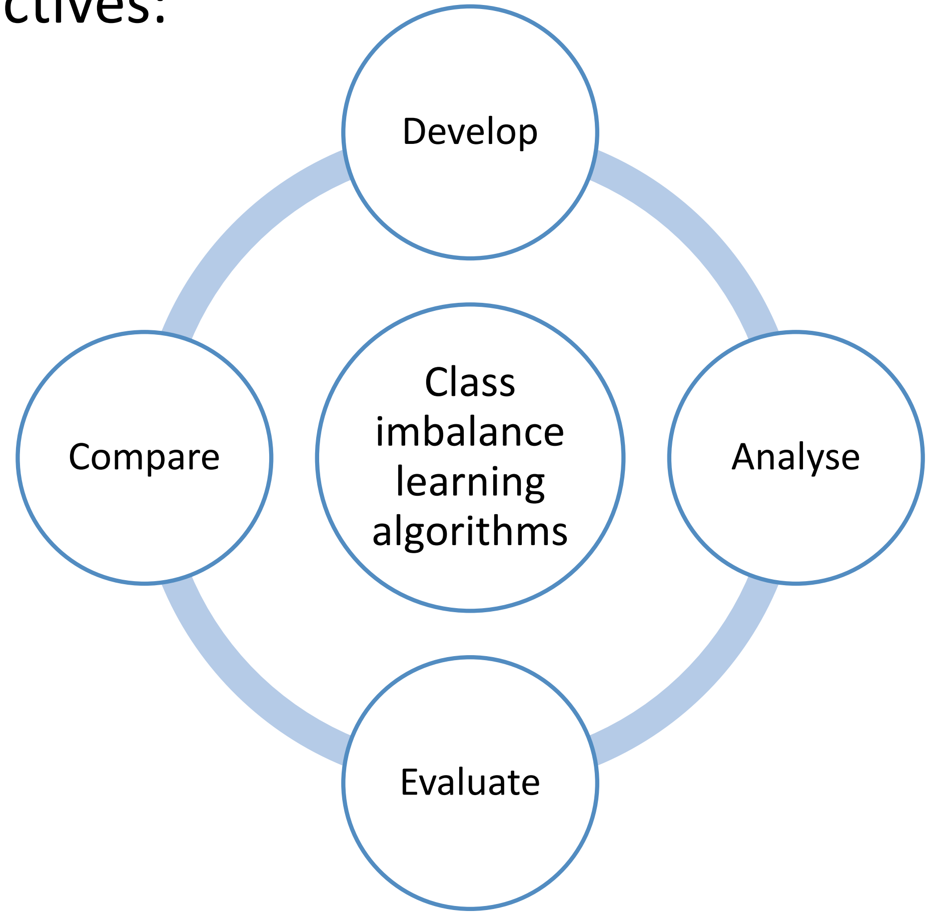
Hyperparameter Optimisation for Improving Classification under Class Imbalance

- Challenge
- Resampling techniques
- Hyperparameter optimisation

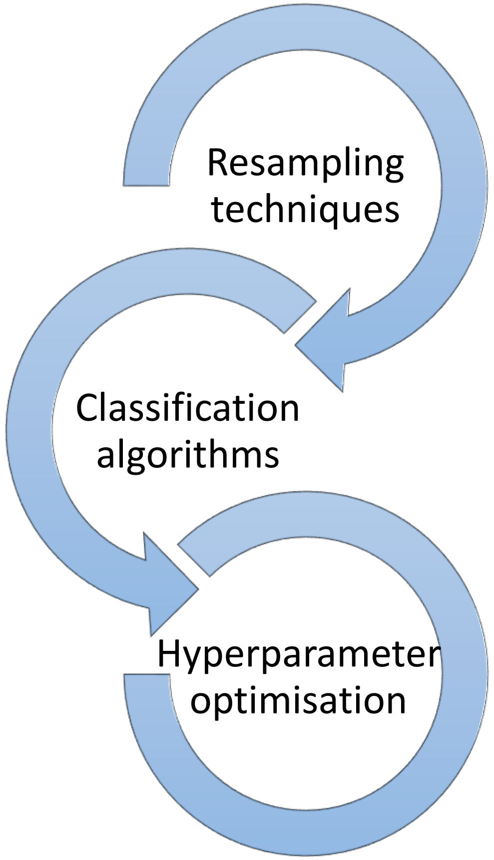


Thesis and research goals

Long-term objectives:



Short-term objectives:



Methodology

Hyperparameter optimisation for improving classification under class imbalance

- Resampling techniques (SMOTE, ADASYN and etc.)
- Hyperparameter optimisation (HyperOpt)
- Data complexity (Maximum Fisher's Discriminant Ratio (F1))

Scenario	Classification Algorithms	Resampling Approaches
(1) $A_{def} + R_{no}$	With default hyperparameters	No
(2) $A_{opt} + R_{no}$	With optimised hyperparameters	No
(3) $A_{def} + R_{def}$	With default hyperparameters	With default hyperparameters
(4) $A_{opt} + R_{def}$	With optimised hyperparameters	With default hyperparameters
(5) $A_{def} + R_{opt}$	With default hyperparameters	With optimised hyperparameters
(6) $A_{opt} + R_{opt}$	With optimised hyperparameters	With optimised hyperparameters

BUSINESS INFORMATION SYSTEMS 2019

Results and future plan

Results

Dataset	#Attributes	#Examples	#Classes	Imbalance Ratio	F1 value
glass1	9	214	2	1.82	0.92
glass6	9	214	2	6.38	0.53

Resampling Approach	NONE		SMOTE		ADASYN		SMOTEENN		SMOTETL	
	glass1	glass6	glass1	glass6	glass1	glass6	glass1	glass6	glass1	glass6
$A_{def} + R_{no}$	0.6754	0.9768	—	—	—	—	—	—	—	—
$A_{opt} + R_{no}$	0.8314	0.9848	—	—	—	—	—	—	—	—
$A_{def} + R_{def}$	—	—	0.7098	0.9758	0.7312	0.9727	0.7286	0.9751	0.7194	0.9744
$A_{opt} + R_{def}$	—	—	0.8364	0.9828	0.8390	0.9803	0.8218	0.9890	0.8459	0.9824
$A_{def} + R_{opt}$	—	—	0.7361	0.9768	0.7403	0.9744	0.7484	0.9822	0.7427	0.9795
$A_{opt} + R_{opt}$	—	—	0.8564	0.9852	0.8592	0.9845	0.8350	0.9877	0.8659	0.9857

Future plan

- More data complexity measures
- Data complexity & choice of resampling techniques
- Applications of class imbalance techniques in engineering

Summary

- A significant improvement can be achieved for the dataset with high F1 value .
- It is necessary to study the data complexity before optimising the hyperparameters.
- The trade-off between time cost and expected performance improvement should be considered.
- More data complexity measures are worth studying in future research.

BUSINESS INFORMATION SYSTEMS 2019

Grant number: 788186

Thank you !