Curriculum Vitae
Georgios Douzas

Business Email: gdouzas@icloud.com

Business Phone: 0030 6951310086

Pandoras 17, Varkiza

16672 Athens

Greece

Professional Experience

Machine Learning Researcher
NOVA IMS University of Lisbon, Portugal
09/13 - 09/14 & 09/18 - present

Designed, implemented and tested various new approaches to deal with the class imbalance problem. Research work focused on clustering based over-sampling methods that deal with the within-the-class imbalance problem. Additionally, Geometric SMOTE, an extension of the SMOTE algorithm, was proposed and implemented. The final publication presented results that show a significant improvement over SMOTE and its variations. Deep learning models, particularly Conditional Generative Adversarial Networks (CGANs), were also used as over-sampling methods with great success. The frameworks of the implementation were TensorFlow, Keras and PyTorch. Work is published in high impact machine learning journals. Implementation of the above algorithms was developed and made available as open source software. Work in progress includes comparative experiments between variations of CGANs as over-samplers as well as the investigation of novel algorithms in the context of Reinforcement Learning.

Machine Learning Engineer
Tripsta, Greece
10/17 - 08/18

Designed and implemented the main parts of the company's automated pricing system. These parts included machine learning estimators for the add-ons and the competitor's prices as well as the application of metaheuristic algorithms for the budget multi-objective optimization problem. The training data of the various estimators were at the order of TB while the prediction time of the automated pricing system was required to be less than 100 msec for the incoming 50K requests/sec. The languages of the implementation were Python, Java and Scala while Spark, Dask, Scikit-Learn and jMetal were used as distributed data processing, machine learning and optimization frameworks/libraries.

Senior Data Scientist
Quantum Retail Technology, Remote
12/16 - 09/17

Worked on demand forecasting and clustering for retail companies. Proposed and applied machine learning methods to improve the company's main forecasting solution that was based on exponential smoothing of the time series data as well as adjustments guided by a seasonality curve. Boosting trees were selected as the final machine learning

model. Applying feature extraction that integrated the business logic as well as applying extensive model hyperparameter tuning, the forecasting precision was improved by 30% compared to the original model.

Machine Learning Engineer
CERN, Remote
05/16 – 09/16

Developed the parallelization of various features for TMVA, the Toolkit for Multivariate Data Analysis with ROOT, as a part of a project funded by Google. ROOT is the main framework developed by CERN to deal with the big data processing, statistical analysis, visualization and storage of the massive amounts of data produced from the particle physics experiments. The legacy version was implemented in C++. The parallelized features included the application of brute-force and metaheuristic algorithms to the hyperparameter grid search of machine learning algorithms. The implementation was based on Python and Spark.

Scientific Software Engineer
IRI, Greece
10/14 – 05/16

Member of the IRI's "Solutions and Innovation Team" (R&D) working on the company's transition towards Open Source and Elastic Computing. Participated in an agile team working on the migration of IRI's main US "Price & Promo Analytics" Solution, generating more than $25M Annual Revenues, to Hadoop distributed storage and Spark cluster computing. Python was the core language of the implementation, but integration with R and Julia was performed to leverage special functionality. The legacy version was implemented in SAS. The main objectives of the project were the design of the parallelization schema, the enhancement of data manipulation with the use of distributed processing and the migration of the statistical modeling algorithms (regression mixed models). The final system was able to process 5 years of data for more than 300 categories containing 1 million products.

Co-owner / CTO
Sports Performance Training*, Greece*
09/09 – 09/13

Co-owner and CTO of Sports Performance Training (SPT) startup. SPT provided consulting services to individual professional athletes and athletic organizations in multiple sports. Using standard ergometric measurements and a custom process through biometric sensors, a variety of signals were recorded and preprocessed. A combination of descriptive statistics, predictive modeling and domain knowledge was applied to adjust the training load of the athlete or team in the pre-competition period and to maximize performance in the competition period. The deliverable to the client included also training guidelines on individual basis that aimed to avoid sports specific injuries.

Research Experience

Machine Learning Researcher, NOVA IMS University of Lisbon, Portugal.

Member of the MagIC research and development center. The previous section provides a detailed description.

Education

PhD in Theoretical Particle Physics, Department of Physics, National Technical University of Athens, Greece (Demokritos Institute Graduate Program Fellowship, awarded yearly after national exams).

MSc in Physics and Technological Applications, Department of Physics, National Technical University of Athens, Greece.

BSc in Physics, Department of Physics, National and Kapodistrian University of Athens, Greece.

Skills

Programing Languages.

Professional level: Python, C++, Java, Scala, R, JavaScript.
Basic level: C, Go.

Software Engineering.

Professional experience in front-end technologies and libraries (HTML, CSS, JavaScript and React), back-end web application frameworks (Flask and Django), software engineering best practices and DevOps tools.

Personal Projects

Open Source.

Geometric SMOTE.
Clustering based over-sampling.
Self-Organizing Map.
Research-Learn.
Generative oversampling.
Sports Betting.

Research

Publications.

Publications

Douzas, G., Bacao, F., Fonseca, J. and Khudinyan, M., Imbalanced Learning in Land Cover Classification: Improving Minority Classes' Prediction Accuracy Using the Geometric SMOTE Algorithm. Remote Sensing, 11, 24 (refereed).

Douzas, G. and Bacao, F. (2019), Geometric SMOTE a geometrically enhanced drop-in replacement for SMOTE. Information Sciences, 501, 118-135 (refereed).

Douzas, G., Bacao, F. and Last F. (2018), Improving imbalanced learning through a heuristic oversampling method based on k-means and SMOTE. Information Sciences, 465, 1-20 (refereed).

Douzas, G. and Bacao, F. (2018), Effective data generation for imbalanced learning using conditional generative adversarial networks. Expert Syst. Appl. 91, C (refereed).

Last, F., Douzas, G. and Bacao, F. (2017), Oversampling for Imbalanced Learning Based on K-Means and SMOTE.

Douzas, G. and Bacao, F. (2017), Geometric SMOTE: Effective oversampling for imbalanced learning through a geometric extension of SMOTE.

Douzas, G. and Bacao, F. (2017), Self-Organizing Map Oversampling (SOMO) for imbalanced data set learning. Expert Syst. Appl. 82, C (refereed).

Douzas, G., Grammatikopoulos, T. and Zoupanos, G. (2009), Coset space dimensional reduction and Wilson flux breaking of ten-dimensional N=1, E_8 gauge theory. Eur. Phys. J. C, 59: 917 (refereed).

Douzas, G., Grammatikopoulos, T., Madore, J. and Zoupanos, G. (2008), Coset space dimensional reduction and classification of semi-realistic particle physics models. Fortschr. Phys., 56: 424-429 (refereed).