A tour through learning from imbalanced data

Friday December 2 | 12-1pm | West 105

Speaker: Dr. Bartosz Krawczyk
Assistant Professor
Computer Science, VCU
Richmond VA

Abstract: Despite more than two decades of continuous development learning from imbalanced data is still a focus of intense research. Starting as a problem of skewed distributions of binary tasks, this topic evolved way beyond this conception. With the expansion of machine learning and data mining, combined with the arrival of big data era, we have gained a deeper insight into the nature of imbalanced learning, while at the same time facing new emerging challenges. This seminar will discuss basics and advanced concepts in the field of learning from imbalanced data, mainly in the context of classification problems. The nature of class imbalance will be analyzed, focusing on analyzing not only the disproportion between classes, but also other difficulties embedded in the nature of data. Pre-processing, algorithm-level and ensemble solutions will be presented, including both a selection of most effective approaches taken from the literature and a number of original solutions developed by the author.

Bio: Bartosz Krawczyk is an assistant professor in the Department of Computer Science, Virginia Commonwealth University, USA. He received an MSc degree in 2012 and PhD degree in 2015, both with distinctions from Wroclaw University of Science and Technology, Poland. His research is focused on machine learning, ensemble learning, data streams, class imbalance, one-class classifiers, and interdisciplinary applications of these methods. He has authored 32 international journal papers and over 80 contributions to conferences. Dr Krawczyk was awarded with numerous prestigious awards for his scientific achievements like IEEE Outstanding Leadership among others. He served as a Guest Editor in four journal special issues and as a chair of ten special session and workshops. He is a member of Program Committee for over 40 international conferences and a reviewer for 30 journals.