



Novelty detection in data streams

André C. P. L. F. de Carvalho*

University of São Paulo, São Carlos, Brazil – andre@icmc.usp.br

João Gama

University of Porto, Porto, Portugal - joao.jgama@gmail.com

Elaine Faria

Federal University of Uberlândia, Uberlândia, elaine@ufu.br

Isabel José da Costa Ribeiro Gonçalves

Polytechnic Institute of Viana do Castelo, Viana do Castelo, Portugal - isagoncalves@estg.ipv.pt

Business, government, industries and science organizations are increasingly moving towards decision-making processes that are based on information. In parallel, the amount of data representing the activities of organizations that is stored in databases is also growing. Therefore, the pressure to extract as much useful information as possible from this data is very strong. Conventional data mining tasks assume that the whole dataset is available. However, most practical problems continuously generate new data, in a data stream. Data stream mining is a research area that investigates the extraction of knowledge from large volumes of continuously generated data. One important application of data stream mining is novelty detection. Novelty detection can be defined as the ability to identify new or unknown situations. Novelty detection is usually a classification task. This paper will discuss some of the main approaches and show applications examples of the application of novelty detection in data stream mining.

Keywords: novelty detection, data mining, data stream.