Data Mining in the Context of Monitoring Mt Etna, Italy

Marco Aliotta (1), Carmelo Cassisi (1), Marcello D'Agostino (1), Susanna Falsaperla (1), Ferruccio Ferrari (1), Horst Langer (1), Alfio Messina (2), Placido Montalto (1), Danilo Reitano (1), and Salvatore Spampinato (1)

(1) Istituto Nazionale di Geofisica e Vulcanologia, Osservatorio Etneo, Catania, Italy (horst.langer@ingv.it), (2) Istituto Nazionale di Geofisica e Vulcanologia, Sezione Roma 2, Roma, Italy, (alfio.messina@ingv.it)

The persistent volcanic activity of Mt Etna makes the continuous monitoring of multidisciplinary data a first-class issue. Indeed, the monitoring systems rapidly accumulate huge quantity of data, arising specific problems of andling and interpretation. In order to respond to these problems, the INGV staff has
developed a number of software tools for data mining. These tools have the scope of identifying structures in the data that can be related to volcanic activity, furnishing criteria for the identification of precursory scenarios. In particular, we use methods of clustering and classification in which data are divided into groups according to a-priori-defined measures of similarity or distance. Data groups may assume various shapes, such as convex clouds or complex concave bodies.The "KKAnalysis" software package is a basket of clustering methods. Currently, it is one of the key techniques of the tremor-based automatic alarm systems of INGV Osservatorio Etneo. It exploits both Self-Organizing Maps and Fuzzy Clustering. Beside seismic data, the software has been applied to the geo-chemical composition of eruptive products as well as a combined analysis of gas-emission (radon) and seismic data.
The "DBSCAN" package exploits a concept based on density-based clustering. This method allows discovering clusters with arbitrary shape. Clusters are defined as dense regions of objects in the data space separated by regions of low density. In DBSCAN a cluster grows as long as the density within a group of objects exceeds some threshold. In the context of volcano monitoring, the method is particularly promising in the recognition of ash particles as they have a rather irregular shape. The "MOTIF" software allows us to identify typical waveforms in time series, outperforming methods like cross-correlation that entail a high computational effort. MOTIF can recognize the non-imilarity of two patterns on a small number of data points without going through the whole length of data vectors.
All the developments aforementioned come along with modules for feature extraction and post-processing. Specific attention is devoted to the obustness of the feature extraction to avoid misinterpretations due to the presence of disturbances from environmental noise or other undesired signals originating from the source, which are not relevant for the purpose of volcano surveillance.