Clustering analysis for muon tomography data elaboration in the Muon Portal project

Clustering analysis for muon tomography data elaboration in the Muon Portal project

M. Bandieramonte*[1,2], V. Antonuccio-Delogu[2], U. Becciani[2], A. Costa[2], P. La Rocca[1,3], P. Massimino[2], C. Petta[1,3], C. Pistagna[2], F. Riggi[1,3], S. Riggi[2], E. Sciacca[2], F. Vitello[2]

[1] Dept. of Physics and Astronomy, University of Catania, Italy
[2] INAF, Astrophysics Observatory of Catania, Italy
[3] INFN, Section of Catania, Italy

E-mail: *marilena.bandieramonte@oact.inaf.it

Abstract. Clustering analysis is one of multivariate data analysis techniques which allows to gather statistical data units into groups, in order to minimize the *logical distance* within each group and to maximize the one between different groups. In these proceedings, the authors present a novel approach to the muon tomography data analysis based on clustering algorithms. As a case study we present the *Muon Portal* project that aims to build and operate a dedicated particle detector for the inspection of harbor containers to hinder the smuggling of nuclear materials. Clustering techniques, working directly on scattering points, help to detect the presence of suspicious items inside the container, acting, as it will be shown, as a filter for a preliminary analysis of the data.

1. Introduction

Clustering or group analysis is a set of multivariate data analysis techniques that aims to select and to group homogeneous items in a data set. Since its appearance in the work of R. C. Tyron in 1939 [1], it was experienced with extended applications, starting from the '60s. Clustering techniques involve the division of the data into homogeneous subgroups. Through this multivariate data analysis technique, statistical units can be grouped together in order to minimize the *logical distance* within each group and to maximize the one between different groups. The *logical distance* is quantified by measures of similarity/dissimilarity between the statistical units. Informally, the goal of this partitioning is twofold. Data elements within a cluster must be similar to each other, while those within different clusters should be dissimilar, in such a way that the observations are the most possible homogeneous within classes and the most possible uneven between the different classes. The concept of homogeneity is specified in terms of distance and several criteria can be used, as it will be clarified later. Each of these classes is known as a *cluster*. It defines also a region where the objects density is locally higher than in other regions. Clustering techniques are successfully applied to problems as classification, pattern recognition and multivariate analysis. In these proceedings the authors present a density-based clustering algorithm and its application to the analysis of muon tomography data. Muon tomography is a technique that exploits the secondary particles of cosmic radiation and their interaction properties with matter. Recently it has been successfully applied to the problems of scanning and detection of radioactive material. The *Muon Portal* project, presented as a case

study in this report, is aimed at creating a large prototype of a scanner for containers control at the borders, in order to combat the illicit transport of fissile elements and face terrorist attacks. To achieve this goal, a large-scale scanner that is composed of four logical planes made of plastic scintillators strips is being realized. The idea of using clustering algorithms to process data in the *Muon Portal* project arises from the need to make the tracks reconstruction and the visualization of the container content, to be independent from the grid and the 3D-voxels[1]. The clustering algorithms, working with scattering points, could be useful to detect the presence of suspicious items inside the container, and could act as a filter for a preliminary analysis of the data. Several algorithms have been tested so far. In these proceedings, a modified Friends-of-Friends (FOF) algorithm is presented and results of simulated scenarios are also showed. The FOF algorithm is a domain-specific clustering algorithm; it is a percolation algorithm often used to identify dark matter halos in N-body simulations [2]. We implemented a multiphase unsupervised clustering version of the algorithm which performs the clustering in two different steps, using at the first step the scattering angle and the euclidean distance at the second step. The algorithm has been optimized by using space-partitioning data structures for organizing points in a k-dimensional space (kd-trees [3]) and further optimization strategies to achieve a $\mathcal{O}(N \log(N))$ complexity which makes it usable in real time analysis. The report is organized as follows: in Section 2 a literature review is presented. Section 3 is dedicated to the description of the FOF algorithm and its modification. Section 4 shows some simulation results applied in the context of the muon tomography imaging. Conclusions are drawn in Section 5.

2. Related Works

The clustering problem has been addressed in many contexts and by researchers in many disciplines; it has been effectively applied in a variety of engineering and scientific disciplines such as psychology, biology, medicine, computer vision, communications and remote sensing (e.g. see Refs. [4, 5, 6]). Cluster analysis organizes data (a set of patterns where each pattern could be represented by a vector) by abstracting the underlying structure, playing therefore a vital role in the field of data mining. The clustering method implemented should be very fast, efficient, and robust. In literature, there are several clustering algorithms described which can be roughly grouped into four categories: partitioning methods, hierarchical methods, grid-based and density-based clustering (see e.g. the review in Ref. [7]). They differ not only in their algorithm principles (which determine the behavior at runtime and the scalability of the algorithm itself), but also in many of their most basic properties such as the type of data processed, the assumptions on the cluster shape, the final form of the partitioning or the parameters that must be provided in the input. Partitioning clustering algorithms generate a single partition with a specified or estimated number of non-overlapping clusters of the data in an attempt to recover natural groups present in the data. Depending on the kind of prototypes, it is possible to distinguish k-means [8], k-modes [9] and k-medoid [10] algorithms. In the k-means algorithm, the similarity between clusters is measured with respect to the mean value of the objects belonging to a cluster. The k-modes extends the k-means paradigm to categorical domains. The k-medoid algorithms use a prototype, called the medoid, that is one of the objects located near the center of a cluster, the so-called centroid. The algorithm *Clarans* introduced in Ref. [11] is an improved k-medoid type algorithm restricting the large search space by using two additional user-supplied parameters. It is significantly more efficient than the well-known k-medoid algorithms *Pam* and *Clara* presented in Ref. [10], nonetheless producing a result of nearly the same quality. Hierarchical clustering algorithms construct a hierarchy of partitions,

[1] A voxel (volumetric pixels, or more precisely volumetric picture element) is a volume element, representing a value of signal intensity or color in a three dimensional space, similar to the pixel that represents a given two-dimensional image. The voxels are often used as a basis for the visualization and analysis of medical and scientific data.

represented as a dendrogram in which each partition is nested within the partition at the next level in the hierarchy [12]. A common way to find regions of high-density in the data space is based on grid cell densities [13]. A histogram is constructed by partitioning the data space into a number of non-overlapping regions or cells. Cells containing a relatively large number of objects are potential cluster centers and the boundaries between clusters fall in the "valleys" of the histogram. The success of this method depends on the size of the cells which must be specified by the user. The limit of this algorithm is that cells of small volume give a very "noisy" estimate of the density, whereas large cells tend to excessively smooth the density estimate. Density-based approaches apply a local cluster criterion and are very popular for the purpose of database mining. Clusters are regarded as regions in the data space in which the objects are dense, and which are separated by regions of low object density (noise). These regions may have an arbitrary shape and the points inside a region may be arbitrarily distributed.

In Ref. [14], a density-based clustering algorithm not grid-based, called *DbScan*, is presented. The basic idea of the *Dbscan* algorithm is that for each point of a cluster the neighborhood of a given radius, ϵ, has to contain at least a minimum number of points, N_{min}, where ϵ and N_{min} are input parameters. In Ref. [15] the density-based algorithm, named *DenClue*, is proposed. This algorithm uses a grid but is very efficient because it only keeps information about grid cells that do actually contain data points and manages these cells in a tree-based access structure. This algorithm generalizes some other clustering approaches which, however, result in a large number of input parameters. Another recent approach to clustering is the *Birch* method [16]. It cannot entirely be classified as a hierarchical or partitioning method since it constructs a CF-tree [17]. CF-tree is a hierarchical data structure designed for a multiphase clustering method. First, the database is scanned to build an initial in-memory CF-tree which can be seen as a multi-level compression of the data that tries to preserve the inherent clustering structure of the data. Then, an arbitrary clustering algorithm can be used to cluster the leaf nodes of the CF-tree. Because *Birch* is reasonably fast, it can be used as a more intelligent alternative to data sampling in order to improve the scalability of clustering algorithms.

3. Multivariate Data Analysis Techniques: a multi-phase unsupervised FOF algorithm

Clustering analysis provides a very sensible tool to detect "objects" within a spatial domain. It is a customary tool adopted e.g. in Astrophysics to quantify the hierarchical evolution of gravitationally bound systems like clusters and superclusters of galaxies. Clustering analysis has a significant advantage over other tools, such as classification algorithms [18, 19], statistical learning algorithms [20, 21] and association analysis algorithms [22, 23]. In fact, because of its sensitivity it can detect objects (clusters) composed of a small number of units.

The reason for this sensitivity is in its scaling with the number of composing units N_c. Therefore the algorithm has a computational complexity $\mathcal{O}(N_c^2)$. This arises because all clustering algorithms are based on pair counting and the number of pairs in a given set of N_c objects is $N_c(N_c-1)/2$. This scaling however is also computationally very expensive, in general. For this reason all the numerical implementations of the clustering algorithms have to be thoroughly tested to determine the minimum/optimal value(s) of N_c, as a result of a trade off between the statistical significance and the numerical complexity.

Common to all clustering algorithms is the comparison between the number of pairs in the sample with the number of cells in a random comparison sample, prepared under the controlled conditions. Two questions arise naturally:

- How much "random" should the control sample be?
- How large should N_c be to guarantee a reasonable accuracy?

There is no universal answer to these two questions. Instead what one can often do is to test

many samples carrying a given signal against the null hypothesis, and to use this statistics to determine the level of significance of a given problem. This is also what was done in the presented work.

Probably the most well known clustering algorithm is the *Friends-of-Friends* algorithm (FOF) [24, 2]. It depends on two parameters, the *linking length* r_{ll} and the minimum number of points N_{min}. Its principle is very simple. Starting from an arbitrary initial element in the dataset it looks for all the elements contained within a distance r_{ll}. If r_i and r_j are two independent elements and their relative distance is $< r_{ll}$, then they are considered as members of a new group. The algorithm defines uniquely groups that contain all the particles separated by a distance smaller than the given linking length r_{ll}. Once the length is defined, the algorithm identifies all pairs of particles which have a mutual distance smaller than the linking one. These pairs are designated friends and clusters are defined as sets of particles that are connected by one or more of the friendly relations, so that they are friends of friends. Another parameter in FOF algorithm is the minimum number of particles N_{min}, in the input list. This is particularly important to reject spurious clusters, i.e. transient objects which have a low statistical significance. Since it is much more likely that a spurious cluster (noise) involves a small number of events then viceversa, choosing N_{min} sufficiently large allows to eliminate these spurious clusters.

To summarize, the FOF requires: r_{ll} and N_{min} as input parameters and it is based on the following steps:

1. Choose an arbitrary unvisited data point as starting point.
2. Find the neighborhood of this point, e.g. all points within the radius r_{ll}. If more than N_{min} neighborhoods are found around this point, a cluster is started and the point is marked as visited, otherwise the point is labelled as noise.
3. If a point is found to be a part of the cluster, then its r_{ll} neighborhood is considered as a part of the cluster and the above procedure from step 2 is repeated for all r_{ll} neighborhood points. This is repeated until all points in the cluster are determined.
4. A new unvisited point is retrieved and processed, leading to the definition of a new cluster or noise.
5. This process continues until all points are marked as visited.

The complexity of this algorithm grows at most as $N_c(N_c-1) \sim \mathcal{O}(N_c^2)$. In practice, as elements are added to the list of clusters they are subtracted from the "active" list so that, at each step, the actual complexity grows as $N_{act}(N_{act}-1)$, where $N_{act} \leq N_c$.

The only freedom left lies in the choice of r_{ll}. A common choice is to take it as the average of the distances between elements for a representative subset of data. This proves to be a very effective choice, as it has also been demonstrated that both the final list of objects and their average properties are largely independent from the choice of r_{ll}, provided that the input data catalogue is an unbiased, statistically representative sample. FOF is thus stable to its only parameter.

3.1. A modified multivariate unsupervised FOF algorithm

An important property of many real datasets is that their intrinsic cluster structure can not be characterized by a single global density parameter. Very different local densities may be needed to reveal clusters in different regions of the data space. Furthermore, one of the main weaknesses of clustering algorithms is that most of them are supervised and require the inclusion of a set of input parameters, and the result is dependent on the input values. In order to perform an optimized exploratory data analysis of the tomographic data from the *Muon Portal* detector, we implemented a modified version of the FOF an unsupervised and multi-phase FOF clustering version. To make the algorithm independent from the settings of input parameters and in

order to identify the density-based clustering structure, we implemented an augmented sorting technique of the dataset, based on the method called *OPTICS* [25]. In this method, the points in the dataset are ordered linearly according to the criteria of spatial distances. The points that are spatially close, are placed next to each other in the ordered database. An object in the dataset is defined a *core object* if it has at least N_{min} points in its ϵ neighborhood. The algorithm define two metrics. The Core Distance is the minimum value of ϵ that would allow an object to be a core object (ϵ'). The Reachability Distance between two objects, p and q, is the maximum between the Euclidean distance and the Core Distance $\max(\, d(p,q),\, \epsilon')$. For every object which has not been processed, the algorithm finds the points close to the ϵ neighborhood of the object, calculates its core distance ϵ' and save the object in the new database. During the data sorting operation, this distance is used as a criterion for two points to be accepted as a part of the same cluster. The algorithm proceeds ordering the points in ascending order to the nearest core object from which they can be reached. This process can be visualized by a dendrogram (see Fig. 1). It is a versatile base both for the automatic and interactive analysis of the cluster, useful to extract the intrinsic structure of clusters.

The grouping operation is performed in two different steps. First, the scattering angle θ_{scatt} is used as a criterion for grouping, performing an initial screening of the dataset and filtering out the noise from the data of interest. In this way, the scattering points are grouped according to the scattering angle. Then, the Euclidean distance is used in the second step, in order to distribute the clusters in the spatial domain and to distinguish different items inside the scanned volume.

The FOF requires the computation of the nearest neighbour of each point in the volume, so it has, in its original implementation, a computational complexity of $\mathcal{O}(N_c^2)$. The modified version of the algorithm has been optimized by using kd-trees to achieve a $\mathcal{O}(N_c \log(N_c))$ complexity which makes the algorithm usable for real time analysis.

4. Simulation results in the context of muon tomography imaging: the *Muon Portal* project

As mentioned in the *Introduction*, the *Muon Portal* project, used as a case study in these proceedings, has as its target the construction of a prototype of a large area detector for cargo containers inspection in ports in order to counteract the illicit carrying of fissile material. The originality and strength of the project lie in the use of the muon tomography technique to identify the presence of suspicious radioactive materials and to reconstruct the container content, discriminating the type of materials without the need to open it for inspection. The logical principle of the technique is the following: two detection planes are placed above the container that has to be inspected and two below it. The planes are made up of modules segmented in plastic scintillators strips, sensitive to muons, which emit light as they cross the strip. This light is then collected by Wavelength Shifting fibers, which re-emit the light in the green wavelength and which convey it towards Silicon Photomultiplier devices, which are able to read the light signal, then processed by the front-end electronics which provides as output the coordinates of the particle impact point on the plane. More details concerning the detector geometry, the electronic readout and the mechanical structure simulation study can be found in Refs. [26, 27, 28]. Using the coordinates, it is then possible to derive the trajectories of the incident ray and of the outgoing one, to determine the position and the amplitude of the scattering. This process is accomplished through the application of tracks reconstruction algorithms and 3D visualization techniques which allow to realize the tomography of the container. The track reconstruction is a challenging task and consists in the elaboration of data from the planes in order to obtain information on the deflection occurred by the muons within the scanned volume. Different statistical algorithms have been applied within the project for tomographic image reconstruction, as for example the simple POCA (*Point of Closest Approach*), the EM-

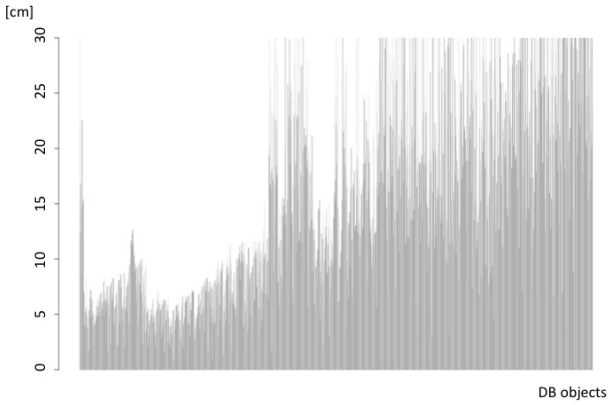


Figure 1. Reachability-plot for a simulation dataset with two blocks of high-Z material.

LM (*Expectation-Maximization Likelihood*), and the modified multivariate unsupervised FOF algorithm. They have been developed in C++ using links to GEANT4 [29] and ROOT [30] frameworks for detector geometry building, navigation and mathematical routines. See Ref. [31] for further details. In the following we show the results of the application of the modified FOF algorithm to a simulated dataset.

The simulated dataset refers to a scenario in which there is a container at the center of which blocks of different materials have been placed. They create a MUON shape built of voxels of size 10 cm×10 cm×10 cm. Each letter is made of a different material: M = Uranium, U = Iron, Lead = O, N = Aluminum. The container load related to this scenario is $\simeq$ 480 kg. In the first step the algorithm OPTICS [25] is run. Figure 1 shows the reachability plot relative to our dataset, through a dendrogram. It shows on the x-axis data of the set reordered according to a criterion of "distance" defined initially as a spatial distance. Each line represents a single point. The y-axis shows the distance values (according to the metric set) that give rise to the presence of clusters. The presence of dips in the plot indicates the existence of a cluster. The analysis made using OPTICS shows that the value of the distance, which is useful for the identification of two clusters, is in the range between 5 and10 cm. It can be seen also graphically, indeed if we draw a horizontal line and move it along the dendrogram, the height of the line that corresponds to the identification of two valleys, it is found around that range.

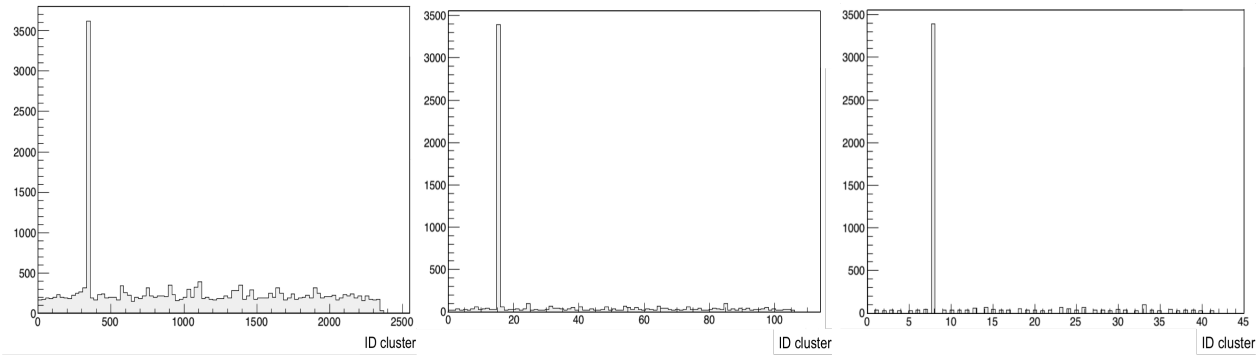
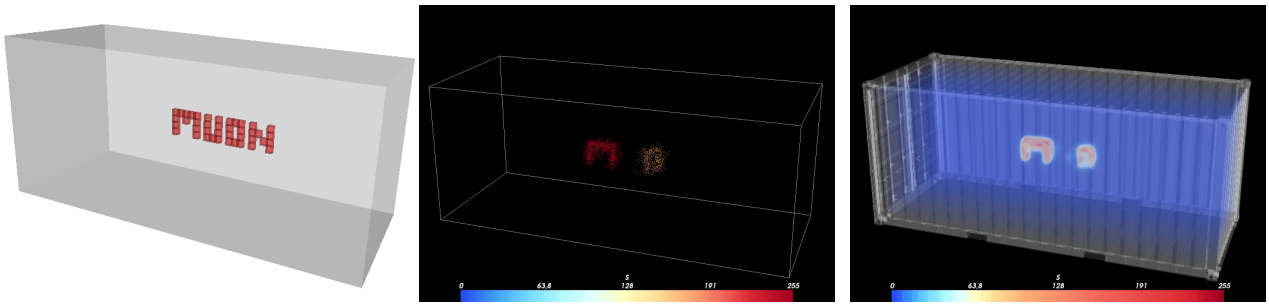


Figure 2. Results of the modified FOF algorithm applied on a dataset with high-Z material using a $r_{ll} = 5$ cm and N_{min} equals to 5, 20, 30 points respectively.

The parameter r_{ll} can be chosen in the identified range and then it can be passed as input parameter to the modified multiphase FOF algorithm. Figure 2 shows several histograms representing the results obtained from FOF algorithm by setting the parameter $r_{ll} = 5$ cm, and assigning to N_{min} different values, 5, 20 and 30, respectively. On the y-axis we have the number of elements for each cluster, while in the x-axis we find the Id of the identified clusters. As it can be seen from the image, the parameter N_{min}, affects only the total number of clusters identified, but in all three cases, the presence of suspicious material is revealed and this can be seen from the high number of elements part of the same cluster. The algorithm calculates the mean value of the scattering angles for each cluster, giving consequently the alarm. In figure 3 is shown the word *MUON* simulated scenario with the results of the application of the clustering method. We can see a 3D visualization of the points filtered by the clustering and the volume rendering of the scenario. As it can be seen the method is successfully applied, and the two high-Z material letters (the M and the O) are identified.



(a) (b) (c)

Figure 3. A simulated scenario of a container with a M= Uranium, U= Iron, O= Lead, N= Aluminum writing inside (a). A 3D visualization of the points filtered by the clustering (b). Volume rendering of the scenario realized by the clustering technique (c).

5. Conclusions

In these proceedings a study of the application of density based clustering algorithms to a muon tomography dataset in the context of the *Muon Portal* project has been shown. A modified version of the FOF algorithm, multi-phase and unsupervised, has been implemented and the simulation results were shown. The clustering method proposed can be used both as a stand-alone tool to analyze and characterize a large dataset, for example for data mining applications and data processing, or it is suitable to be used as a filter, in pre-processing phases for other algorithms that operate, subsequently, on the identified clusters. The implemented method has many advantages. Firstly, there is a linear scaling behavior, which makes it interesting for the application on big datasets. Secondly, the nature of the algorithm makes it quite robust to the effects of outliers within the data. Also, the proposed clustering method has the ability to work with all types of datasets that can be described by a metric and does not impose any assumptions about the shape of the clusters within which it works, being free from the limits of three-dimensional grids. Finally, an important aspect of the algorithm is its ability to automatically determine the number of clusters without a supervision. Statistical analysis of different scenarios are currently in progress, in order to verify the validity of the proposed method even in realistic and highly dense scenarios.

References

[1] Tryon R 1939 *Cluster Analysis: Correlation Profile and Orthometric (factor) Analysis for the Isolation of Unities in Mind and Personality* (Edwards brother, Incorporated, lithoprinters and publishers)

[2] Lacey C G and Cole S 1994 *Mon.Not.Roy.Astron.Soc.* **271** 676 (*Preprint* `astro-ph/9402069`)
[3] Bentley J L 1975 *Communications of the ACM* **18** 509–517
[4] Sciacca E, Spinella S, Ienco D and Giannini P 2011 Annotated stochastic context free grammars for analysis and synthesis of proteins. *Evolutionary Computation, Machine Learning and Data Mining in Bioinformatics* (*Lecture Notes in Computer Science* vol 6623) ed Pizzuti C, Ritchie M and Giacobini M (Springer Berlin Heidelberg) pp 77–88 ISBN 978-3-642-20388-6
[5] Frigui H and Krishnapuram R 1999 *Pattern Analysis and Machine Intelligence, IEEE Transactions on* **21** 450–465 ISSN 0162-8828
[6] Bandyopadhyay S, Maulik U and Mukhopadhyay A 2007 *Geoscience and Remote Sensing, IEEE Transactions on* **45** 1506–1511 ISSN 0196-2892
[7] Jain A K, Murty M N and Flynn P J 1999 *ACM Comput. Surv.* **31** 264–323 ISSN 0360-0300
[8] MacQueen J B 1967 Some methods for classification and analysis of multivariate observations. *Proc. of the fifth Berkeley Symposium on Mathematical Statistics and Probability* vol 1 ed Cam L M L and Neyman J (University of California Press) pp 281–297
[9] Huang Z 1997 A fast clustering algorithm to cluster very large categorical data sets in data mining. *In Research Issues on Data Mining and Knowledge Discovery* pp 1–8
[10] Kaufman L and Rousseeuw P J 2008 *An Introduction to Cluster Analysis* (John Wiley Sons, Inc.) pp 320–331 ISBN 9780470316801
[11] Ng R T and Han J 1994 Efficient and effective clustering methods for spatial data mining. *Proceedings of the 20th International Conference on Very Large Data Bases* VLDB '94 (San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.) pp 144–155 ISBN 1-55860-153-8
[12] Johnson S 1967 *Psychometrika* **32** 241–254 ISSN 0033-3123
[13] Jain A K and Dubes R C 1988 *Algorithms for Clustering Data* (Upper Saddle River, NJ, USA: Prentice-Hall, Inc.) ISBN 0-13-022278-X
[14] Ester M, Kriegel H P, Sander J and Xu X 1996 A density-based algorithm for discovering clusters in large spatial databases with noise. *Kdd* vol 96 (AAAI Press) pp 226–231
[15] Hinneburg A, Hinneburg E and Keim D A 1998 An efficient approach to clustering in large multimedia databases with noise. (AAAI Press) pp 58–65
[16] Zhang T, Ramakrishnan R and Livny M 1996 Birch: An efficient data clustering method for very large databases.
[17] Barbará D 2002 *SIGKDD Explor. Newsl.* **3** 23–27 ISSN 1931-0145
[18] Kukar M 2006 *Knowledge and information systems* **9** 364–384
[19] Li T, Zhu S and Ogihara M 2006 *Knowledge and information systems* **10** 453–472
[20] Fung G and Stoeckel J 2007 *Knowledge and Information Systems* **11** 243–258
[21] Tao D, Li X, Wu X, Hu W and Maybank S 2007 *Knowledge and Information Systems* **13** 1–42 ISSN 0219-1377
[22] Ahmed S, Coenen F and Leng P 2006 *Knowledge and information systems* **10** 315–331
[23] Bonchi F and Lucchese C 2006 *Knowledge and information systems* **9** 180–201
[24] Davis M, Efstathiou G, Frenk C S and White S D 1985 *Astrophys.J.* **292** 371–394
[25] Ankerst M, Breunig M M, Kriegel H P and Sander J 1999 Optics: Ordering points to identify the clustering structure. (ACM Press) pp 49–60
[26] Pugliatti C, Antonuccio V, Bandieramonte M, Becciani U, Belluomo F, Belluso M, Billotta S, Blancato A A, Bonanno D L, Bonanno G, Costa A, Fallica G, Garozzo S, Indelicato V, Rocca P L, Leonora E, Longhitano F, Longo S, Presti D L, Massimino P, Petta C, Pistagna C, Puglisi M, Randazzo N, Riggi F, Riggi S, Romeo G, Russo G V, Santagati G, Valvo G, Vitello F, Zaia A and Zappal G 2014 *Journal of Instrumentation* **9** C05029
[27] Antonuccio V, Bandieramonte M, Becciani U, Belluomo F, Belluso M, Billotta S, Blancato A, Bonanno D, Bonanno G, Carbone B *et al.* 2012 Design of a large area tomograph to search for high-z materials inside containers by cosmic muons. *Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC), 2012 IEEE* (IEEE) pp 5–8
[28] Rocca P L, Antonuccio V, Bandieramonte M, Becciani U, Belluomo F, Belluso M, Billotta S, Blancato A A, Bonanno D, Bonanno G, Costa A, Fallica G, Garozzo S, Indelicato V, Leonora E, Longhitano F, Longo S, Presti D L, Massimino P, Petta C, Pistagna C, Pugliatti C, Puglisi M, Randazzo N, Riggi F, Riggi S, Romeo G, Russo G V, Santagati G, Valvo G, Vitello F, Zaia A and Zappala' G 2014 *Journal of Instrumentation* **9** C01056
[29] Geant4 C 2003 *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **506** 250 – 303 ISSN 0168-9002
[30] Brun R and Rademakers F 1997 *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **389** 81 – 86 ISSN 0168-9002 new Computing Techniques in Physics Research V

[31] Riggi S, Antonuccio-Delogu V, Bandieramonte M, Becciani U, Costa A, La Rocca P, Massimino P, Petta C, Pistagna C, Riggi F *et al.* 2013 *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **728** 59–68