# A Hierarchical Conceptual Clustering Based on the Quantile Method for Mixed Feature-type Data

Manabu Ichino*
Tokyo Denki University, Saitama, Japan ichino@mail.dendai.ac.jp

Paula Brito
FEP & LIAAD INESC TEC, Universidade do Porto, Porto, Portugal mpbrito@fep.up.pt

Let each of $N$ objects be described by $d$ features composed of interval features, histogram features, multi-nominal features, and others. The quantile method transforms each of $N$ complex objects to $(m+1)$ $d$-dimensional numerical vectors, called quantile vectors, where $m$ is a preselected integer number and quantiles are obtained from the underlying distribution assumed for each observation of each feature. A $d$-dimensional rectangle spanned by a pair of quantile vectors represents a sub-object (sub-concept) of the given object in the $d$-dimensional representation space. Therefore, the selection of the integer number $m$ controls the granularity of the sub-concepts constructing the given object in the representation space. We define the concept size of a rectangular region in the representation space as the average of $d$ normalized side lengths of the rectangle. Then, we construct the concept hierarchy to minimize the concept size, i.e. the compactness of the concept is achieved, at each step of the clustering started from $N\times(m+1)$ $d$-dimensional quantile vectors (unit concepts) to a single largest concept. This hierarchical clustering satisfies dual monotone properties: the extensions of sub-objects satisfy the monotone property at the step by step merging process, and the corresponding concept sizes satisfy the same monotone property. We can see the mutual relationships of $N$ objects in the concept hierarchy (dendrogram) based on the level of unit parts (quantile vectors) representing the objects. We present two experimental results in order to show the usefulness of the proposed quantile method.

**Key Words:** Quantification by quantile vectors, Granularity of concept, Compactness, Dual monotone properties