Starshaped subgroup discovery with uniform Vapnik-Chervonenkis-guarantees

In this paper we introdcue a new type of subgroup discovery. While classical forms of subgroup discovery in its simplest form look at subgroup descriptions that are given by conjunctions of attribute specifications, here we look at a richer class of subgroups that may be called starshaped subgroups. As a description language of these subgroups we use here the general notion of betweenness that in a relational manner specifies if one data point lies between two other data points or not. It turns out that solving the exhaustive subgroup discovery problem on this richer family of starshaped subgroups is in a certain sense computationally simpler compared to classical subgroup discovery. On the other hand, because of the richer class of subgroups one has to deal with possible overfitting. For this we analyze the family of all starshaped subgroups within a concise Vapnik-Chervonenkis analysis. It turns out that with a notion of stylized betweenness we can in fact regularize the subgroup problem in a uniform way and with a uniform kind of Vapnik-Chervonenkis guarantee.