5th International Conference on

Big Data Analysis and Data Mining

June 20-21, 2018 | Rome, Italy

Ordinally equivalent patterns: an approach to improving visualization in parallel coordinates

Alexey Myachin and Boris Mirkin
National Research University Higher School of Economics—Moscow, Russia

Parallel coordinates is a well-known framework for data visualization. Rather than conventionally considering data objects as points in a multidimensional space, this approach visualizes the objects as polylines between parallel axes representing different features. Attempts at applying parallel coordinates to big data tables face issues. Of those, probably the most important is the so-called clutter at which no data structure can be reliably seen on the picture. There have been various attempts at overcoming the issue such as edge bundling or continuous parallel coordinates. We propose a different approach: extracting ordinally-equivalent, or co-monotone, patterns. Given an ordering $p=(x_1, x_2,\ldots, x_n)$ of data features, we refer to objects i_1 and i_2 as p–equivalent if the parallel coordinate polylines for these two objects are co-monotone, so that, for any k, $k=1, 2,\ldots, n-1$: (a) $x_k(i_1)>x_{k+1}(i_1)$ iff $x_k(i_2)>x_{k+1}(i_2)$, (b) $x_k(i_1)<x_{k+1}(i_1)$ iff $x_k(i_2) < x_{k+1}(i_2)$, and (c) $x_k(i_1) = x_{k+1}(i_1)$ iff $x_k(i_2) = x_{k+1}(i_2)$. We refer to the two objects as ordinally equivalent if they are p-equivalent at any order p of parallel coordinates. Obviously, the ordinal equivalence is an equivalence relation, thus, corresponding to a partition of the observations in classes of ordinally equivalent objects. Of course, no clutter may emerge at the parallel coordinate graphs of a class of the ordinal equivalence. Although, on the first glance, finding ordinal equivalence class is a complex, non-polynomial, combinatorial problem, we prove that, in fact, it is quadratic over n, and give several algorithms for finding the classes, depending on the user's preferences. The algorithms are scalable to big data sizes. We give examples of using ordinally equivalent patterns for extraction and visualization of patterns prevailing in data related to various application domains.

bmirkin@hse.ru

Computer Engineering & Information Technology
ISSN: 2324-9307

Data Mining 2018
June 20-21, 2018

Volume 7

Page 75