

Současné směry ve vývoji metod shlukové analýzy

Hana Řezanková

Ačkoliv podle některých statistiků je shluková analýza problematickým nástrojem analýzy dat (např. z důvodu, že různými postupy mohou být získána různá přiřazení objektů do shluků), u jiných vědeckých pracovníků je stále oblíbená a využívána. A především jsou stále vyvíjeny nové postupy či modifikace stávajících.

I když určité speciální postupy byly řešeny již v 70. letech 20. století (např. problematiku souborů s proměnnými různých typů řešil Gower v roce 1971 a algoritmy pro shlukování velkých souborů dat publikoval Hartigan v roce 1975), o jejich rozvoj se zvyšuje zájem na přelomu 20. a 21. století. Souvisí to především s nástupem technik pro tzv. data mining, kdy se analýzou dat začínají ve větší míře kromě statistiků zabývat také informatiči.

V současné době jsou hlavními směry vývoje

- metody pro rozsáhlé soubory dat,
- metody pro soubory s kategoriálními proměnnými,
- metody pro fuzzy a odvozené typy shlukování.

V příspěvku budou uvedeny principy navržených postupů pro shlukování v případě výše uvedených speciálních typů souborů a principy speciálních přístupů, které jsou protikladem k pevnému shlukování.

Current Trends in Development of Cluster Analysis Methods

Hana Řezanková

The main tasks for cluster analysis of the last decades are clustering very large data files, clustering data files with categorical variables, fuzzy clustering and other techniques expressing uncertainty. Some relating problems are solving, as outlier detection, determining the number of clusters, etc.

Although the tasks mentioned above were solving in the beginning of cluster analysis development, at the end of the 20th century and at the beginning of the 21st century the interest in these methods are growing in connection with the development of data mining techniques. The new algorithms for cluster analysis are proposed not only by statistics, but also by computer science researchers.

In the contribution, the principles of proposed techniques for clustering in case of special types of files mentioned above, and the principles of techniques expressing uncertainty are discussed.