

Daugiamačių duomenų vizualizavimas

Visualization of multidimensional data

Šiuolaikiškomis technologijomis galima gauti ir saugoti didelius duomenų kiekius ir srautus. Tačiau tebelieka esminė problema – kaip tuos duomenis suvokti ir interpretuoti, o taip pat priimti geriausius sprendimus žinių, gautų iš šių duomenų, pagrindu. Suvokti duomenis yra nelengvas uždavinys, ypač kai jie yra daugiamačiai, t. y. kai nurodo sudėtingą objektą ar reiškini, apibūdinamą daugeliu parametrų (požymių, savybių, rodiklių, ypatybių), kurie gali būti ne tik skaitiniai, bet loginiai, tekstiniai ir kt. Dažnai iškyla būtinybė nustatyti ir giliau pažinti tokių duomenų struktūrą: susidariusias grupes (klasterius), itin išsiskiriančius objektus, objektų tarpusavio panašumą ar skirtingumą ir pan. Vizualizavimas – tai galimybė duomenis pateikti žmogui suprantama forma, padedančia geriau juos suvokti. Vizualią informaciją žmogus pajėgus suvokti daug greičiau negu skaitinę, ji palengvina naujų žinių atradimą. Daugiamačių duomenų vizualizavimas ir dimensiškumo mažinimas yra neatskiriama duomenų tyrybos ir mašininio mokymosi dalis. Yra sukurta daug vizualizavimo metodų, savaip pateikiančių žinias apie duomenų rinkinį. Jie skiriasi ne tik vizualaus pateikimo forma, bet ir prielaidomis apie duomenų rinkinio struktūrą. Vizualizavimo uždavinys dažnai turi savyje optimizavimo uždavinį, kurio tikslo funkcija yra daugiaekstremė ir kurio kintamųjų skaičius yra labai didelis. Kyla iššūkiai greičiau ir tiksliau išspręsti tą uždavinį, ypač kai vizualizuojamų duomenų kiekiai yra labai dideli. Čia teks pasitelkti ir specialius optimizavimo metodus, ir našiuosius ir lygiagrečiuosius skaičiavimus. Nauji vizualizavimo metodai taip pat bus kuriami.

Modern technologies can access and store large amounts and streams of data. However, the fundamental problem remains: how to understand and interpret this data, and how to make the best decisions on the basis of the knowledge derived from it. Making sense of data is a challenging task, especially when it is multidimensional, i.e. when it refers to a complex object or phenomenon characterised by a large number of parameters (features, attributes, properties, indicators, characteristics), which may be not only numerical, but also logical, textual etc. There is often a need to identify and gain a deeper understanding of the structure of such data: the groups (clusters) that have been formed, the objects that are particularly distinctive, the similarity or dissimilarity of the objects, etc. Visualisation is the ability to present data in a human-readable form that helps to understand it better. Visual information can be grasped much more quickly than textual information and facilitates the discovery of new knowledge. Visualisation and dimensionality reduction of multidimensional data is an integral part of data mining and machine learning. Many visualisation techniques have been developed to represent knowledge about a dataset in different ways. They differ not only in the form of the visual representation but also in the assumptions made about the structure of the dataset. A visualisation problem often contains an optimisation problem with a multiextremal function and a very large number of variables. There are challenges in solving that problem faster and more accurately, especially when the amount of data to be visualised is very large. This will require the use of both special optimisation techniques and high-performance and parallel computing. New visualisation techniques will also be developed.