

Vizualizace dat příprava na zkoušku

Richard Zvonek

6. ledna 2021

Obsah

1	Základní členění grafických forem užitých při vizualizaci dat a jejich stručný historický přehled	3
1.1	Explorační grafika	3
1.2	Prezentační grafika	3
1.3	Historie	3
2	Definice pojmu dataset, spojitost, charakteristika dimenzí; vzorkovaná data a jejich zpětná rekonstrukce, formální zápis, příklady bazových funkcí a způsobů členění domény	4
2.1	Dataset	4
2.1.1	Spojítost	4
2.1.2	Charakteristika dimenzí	4
2.2	Vzorkovaná data a jejich zpětná rekonstrukce	4
2.2.1	Formální zápis	4
2.2.2	Příklady bazových funkcí a způsobů členění domény	4
3	Interpolace (skalárních) dat přímo z mračna bodů bez použití mřížky, modely barev a barevné přechody užitá ve vizualizaci dat.	5
3.1	Interpolace bez mřížky	5
3.2	Barevné přechody	5
3.2.1	Černobílá škála	5
3.2.2	Škála dvou odstínů	5
3.2.3	Tepelný škála - Heatmap	5
3.2.4	Rozlišující škála - Diverging	6
4	Vizualizace vektorových polí, divergence, rotace, vektorové čáry(proudnice, charakteristické směry).	7
4.1	Vektorová pole	7
4.2	Divergence	7
4.3	Rotace	7
4.4	Vektorové čáry	7
5	Rekonstrukce izoploch ve 3D, algoritmus Marching cubes	8
5.1	Marching Cubes	8
6	Tenzorová data a možnosti jejich vizualizace, příklady tenzorů druhého řádu, základní operace s tenzory	9
6.1	Vizualizace	9

7	Metody vizualizace volumetrických dat, užité optické modely, způsoby klasifikace a kompozice vzorků, možnosti interpolace	10
7.1	Maximum intensity projection	10
7.2	Direct volume rendering	10
7.3	Optické modely	10
7.4	Klasifikace vzorků	10
7.5	Interpolace	10
8	Vizualizace abstraktních dat a jejich charakteristika, možnosti vizualizace grafových struktur a vícerozměrných dat, redukce dimenze	11
8.1	Vícerozměrná data	11

1 Základní členění grafických forem užitých při vizualizaci dat a jejich stručný historický přehled

Při vizualizaci dat rozdělujeme grafiku na dva základní typy - prezentační a explorační.

1.1 Explorační grafika

Explorační grafika se používá primárně pro potřeby výzkumu a dalšího zpracování. Měla by být precizní, obsahovat definice, správně uvedené jednotky, osy, popisky, legendy...

1.2 Prezentační grafika

Prezentační grafika cílí na masové využití, měla by být zjednodušená a přehledná. Můžou být vynechány některé nadbytečné prvky potřebné v explorační grafice.

1.3 Historie

Historicky, jedny z nejstarších zobrazovaných dat byly data pro navigaci (mapy, pozice hvězd). Egypťané už ve 2. století př.n.l. užívali základní myšlenku souřadného systému, která byla v nezměněné formě používána až do 14. století.

Na začátku 17. století se začaly rozmáhat vizualizace měření matematicko-fyzikálních dat, zároveň se začaly objevovat i statistická měření (Vzdálenost Toledo-Rím, Sluneční skvrny).

V 18. století se postupně zlepšovaly mapy a přidávaly se nové informace, izoliny (vrstevnice). Lambert přišel s interpolací dat získaných empirickým měřením. Playfair přišel se spojnicovým, sloupcovým a koláčovým grafem.

V 19. století se díky modernějším možnostem zpracování grafiky se vizualizace dat velmi posunula. Vznikla většina do dnes používaných typů grafů pro statistiku: Sloupcové a koláčové grafy, histogramy, bodové grafy... V kartografii se více objevovaly komplexní atlasy map, rozšířila se i symbolika. Grafy fyzikálních úkazů se začaly častěji objevovat ve vědeckých článcích. Objevilo se stínování pro vyjádření poměru (negramotnost ve Francii. Moderní mapy se začaly používat pro ekonomické plánování.

Konec 19. století se považuje za zlatou éru statistické grafiky, objevily se první zdařilé 3D grafy, mapy se stávaly ještě podrobnějšími, do statistických atlasů se zobrazovalo více a více dat...

Od začátku až do poloviny 20. století se vizualizace dat moc neposouvala, nebyl moc velký technický pokrok který by vizualizaci posunul. Vizualizace byly spíše formální a věcné.

V 50. letech se vznikem fortranu vznikl nový způsob zpracování dat. S rozšířením počítačů, programovacích jazyků a nových nástrojů se zjednodušila tvorba grafiky. Vznikly nové způsoby vstupu i výstupu informací. Vznikly první geografické informační systémy a první 2D a 3D interaktivní systémy.

Poslední čtvrtina 20. století se nesla ve znamení dat s vyšší dimenzí, a ve zlepšování interaktivity.

2 Definice pojmu dataset, spojitost, charakteristika dimenzí; vzorkovaná data a jejich zpětná rekonstrukce, formální zápis, příklady bazových funkcí a způsobů členění domény

2.1 Dataset

Dataset je soubor informací složený z jednotlivých prvků, se kterými lze manipulovat jako s jedním celkem. Např. položka datasetu měření teploty může být složena z času, naměřené teploty a místa měření.

2.1.1 Spojitost

Spojité dataset je tvořen pomocí spojitě funkce s daným definičním oborem a oborem hodnot. Z definice spojitě funkce vyplývá, že hodnoty spojitě datasetu se mění plynule. Typicky není jednoduché měřit reálná data spojitě a je potřeba je vzorkovat, viz 2.2 Vzorkovaná data a jejich zpětná rekonstrukce.

2.1.2 Charakteristika dimenzí

Dimenzí datasetu se rozumí počet jednotlivých prvků v rámci jednoho měření, pro tabulková data se jedná o počet sloupců. Např. dimenze datasetu měření teploty s informací o pozici, čase a měřené teplotě je 3.

Topologickou dimenzí se rozumí, kolika dimenzionálním prostor je potřeba pro zobrazení dat. Pro vizualizaci se v zásadě volí maximálně 3 dimenze. Rozdělujeme podle počtu dimenzí:

1	Křivka
2	Plocha
3	Volumetrická data

2.2 Vzorkovaná data a jejich zpětná rekonstrukce

Spojité data není často možné přesně naměřit spojitě. Proto je potřeba data vzorkovat. Zpětná rekonstrukce je poté možná formou aproximace, není možné získat absolutně přesná data. Rekonstrukce bere v potaz navzorkované hodnoty a z nich se snaží dopočítat aproximaci původní hodnoty.

Vzorkování je možné provádět rovnoměrně, nebo nerovnoměrně. Nerovnoměrné vzorkování je možné provést např. s ohledem na distribuci významnosti hodnot.

2.2.1 Formální zápis

Rekonstruovanou hodnotu lze vyjádřit jako sumu funkčních hodnot násobenou bazovou funkcí.

2.2.2 Příklady bazových funkcí a způsobů členění domény

Konstantní bazové funkce Jedná se o interpolaci nejbližším sousedem, výhodou je výpočetní nenáročnost a bezproblémová funkcionalita na libovolném počtu dimenzí. Nevýhodou je velmi špatná, nepřesná aproximace. Aproximace není příliš spojitá, je skoková.

Lineární bazové funkce Pro aproximaci se používají koeficienty pro určení vlivu funkčních hodnot původních bodů na funkční hodnotu interpolovaného bodu. Pro 2 body se změna hodnoty počítá lineárně, pro 3 body pomocí barycentrických souřadnic.

3 Interpolace (skalárních) dat přímo z mračna bodů bez použití mřížky, modely barev a barevné přechody užité ve vizualizaci dat.

Některá data není možné jednoduše interpolovat pomocí mřížky (např. protože nejsou podle mřížky vzorkována). Takový případ může nastat např. v případě nějaké "divoké" distribuce měřených dat. Typickým případem můžou být 3D data získaná lidarem, tvořící mračno bodů. Taková data můžeme interpolovat pomocí vytvoření nějaké mřížky, nebo přímo, bez mřížky.

3.1 Interpolace bez mřížky

Interpolace bez mřížky je vhodná pro velké datasety. Není tolik náročná na paměť, ani na výpočetní výkon. Pro interpolaci se využijí RBF (radiální bazové funkce), které pro výpočet využívají vzdálenost od daného bodu. Na daný bod mají pak největší vliv nejbližší body. Jedny z nejčastějších radiálních bazových funkcí jsou Gaussova funkce, nebo inverzní vzdálenost. S použitím radiálních bazových funkcí se pak interpolace počítá podobně, jako interpolace mřížkou. Pro normalizaci hodnot se poté používá Shepardova interpolace, která používá radiální báze, ale výpočet je zatížen sumou vzdáleností interpolačních bodů.

3.2 Barevné přechody

Barevné přechody umožňují jednoznačně a jednoduše určit informace o vizualizovaných skalárních hodnotách jako je např. absolutní hodnota, pořadí a rozdíl hodnot, nebo rychlost změny hodnoty. Důležité je poté k datům vizualizovaných pomocí barevných přechodů uvádět i legendu. Důležité je taky správně pracovat s barevnou škálou. Např. teplé barvy oproti studeným barvám přitahují vyšší pozornost. Je tedy vhodné teplé barvy použít pro hodnoty, které chceme zdůraznit. Důležitý je také kontext. Cílový uživatel má zažitá očekávání a je vhodné je dodržet. Např. není vhodné pomocí teplých barev znázorňovat nízké teploty a pomocí studených barev vysoké teploty. Taková vizualizace není příliš přehledná. Vhodné taky je, aby barevná škála měla lineární průběh. Pokud by tomu tak nebylo, lineární průběh dat by mohl být interpretován nelineárně.

Důležité také je pro barevnou škálu vybrat celkový počet barev. Pokud vybereme vysoký počet barev, je možné, že data budou splývat. Nízký počet barev ale zase bude tvořit ve výsledné vizualizaci ostré hrany. Vysoký počet barev je vhodné zvolit pro zvýraznění linearitu hodnot, nízký počet je vhodný např. pro vrstevnice.

3.2.1 Černobílá škála

Černobílá škála je vhodná pro lineární data. Výhodou černobílé škály je přímý převod hodnoty na jas, nevzniká zmatení z různých odstínů různých barev. Výhoda může být i při tisku, kdy vizualizace není natolik závislá na přesných odstínech barev. Může být však problém s rozlišením dvou různých odstínů šedé barvy. Např. v případě nespojitého bodového grafu může vzniknout problém.

3.2.2 Škála dvou odstínů

Barevná škála je tvořena lineární interpolací dvou různých barev. Dalo by se říct, že černobílá škála je speciálním případem této škály. Volba dvou základních barev zásadně ovlivňuje výslednou vizualizaci. Vybrané barvy by měly být dostatečně rozdílné (ideálně komplementární barvy). Pokud jsou zvoleny dvě barvy, které nejsou příliš jasné, lze potom vizualizaci mít zobrazenou na bílém pozadí. Nevýhodou tohoto typu je pak nižší dynamický rozsah.

3.2.3 Tepelná škála - Heatmap

Tato škála intuitivně následuje teplotu materiálu. Nízké hodnoty jsou tmavé, střední hodnoty mají červený odstín a vysoké hodnoty pokračují žlutými odstíny až do bílé barvy. Tepelná škála obsahuje více odstínů

barev, lze tedy jednoduše rozlišit mezi více hodnotami. Jas na škále stoupá lineárně, což také napomáhá čitelnosti.

3.2.4 Rozlišující škála - Diverging

Tvorba této škály je podobná jako škála dvou odstínů. Okraje škály tvoří dvě barvy s podobným jasnem. Doprostřed škály se poté vloží třetí barva, ideálně s vyšším jasnem a výsledná škála vzniká lineární interpolací mezi těmito třemi barvami. Tato škála je vhodná pro zobrazení odchylky oproti střední hodnoty.

4 Vizualizace vektorových polí, divergence, rotace, vektorové čáry(proudnice, charakteristické směry).

Vektorová data jsou v podstatě více dimenzionální skaláry. Vektor se skládá z n skalárů. Vektor může znázorňovat pozici, směr, rychlost... Většina vektorových dat je 2 nebo 3 dimenzionální.

4.1 Vektorová pole

Vektorová pole jsou více rozměrná vektorová data (vektory ve 2 nebo 3 dimenzích). Příkladem takových dat mohou být např. kapaliny, nebo proudění tlaku.

4.2 Divergence

Divergence je v podstatě součet parciálních derivací hodnot. Divergence symbolizuje přírůstek nebo úbytek v daném bodě v daném čase.

4.3 Rotace

Rotace - curl v bodě se počítá jako rozdíl derivací ve směrech. Tato hodnota znázorňuje rychlost a směr rotace v daném bodě.

4.4 Vektorové čáry

Vektorové čáry umožňují znázorňovat jak rychlost, tak i směr rotace ve vektorovém poli. Proudnice poté znázorňují přesný přesun masy z jednoho bodu do druhého v čase. Vizualizace může být pomocí spojitých čar, nebo pomocí glyfů. Glyfy mohou být ve formě čar, nebo např. pomocí 3D kuželů. Glyfy také mohou být rozmístěny uniformně, nebo s ohledem na hodnoty (v místech vyššího proudění vyšší koncentrace glyfů).

5 Rekonstrukce izoploch ve 3D, algoritmus Marching cubes

Rekonstrukce izoploch se používá pro tvorbu ploch z mračna bodů. Může být použita např. pro simulaci povrchu kapaliny, nebo např. při procedurálním generování terénu z 3D šumové funkce.

5.1 Marching Cubes

Algoritmus Marching cubes je rozšíření algoritmu Marching squares o třetí dimenzi. Oba algoritmy rekonstruují topologickou plochu objektů. Pro rekonstrukci je nejprve spočítat mřížku, která bude následně tvořit těleso. Čím větší je rozlišení mřížky, tím větší je potom přesnost výsledného tělesa. Následně se provede interpolace hodnoty v každém vrcholu mřížky pomocí všech okolních bodů. Následně se tvoří izoplocha. Pokud je hodnota menší, než prahovací hodnota, leží hodnota uvnitř tělesa. V opačném případě leží vně tělesa. Algoritmus poté prochází čtverce / krychle, počítá s jejich vrcholy a tvoří polygony plochy. Pro přesnější výsledek je vhodné vrcholy nových polygonů v rámci krychle lineárně interpolovat podle funkčních hodnot vrcholů krychle.

6 Tenzorová data a možnosti jejich vizualizace, příklady tenzorů druhého řádu, základní operace s tenzory

Tenzorová data jsou obecně data vyššího řádu. Tenzory nultého řádu jsou skalární hodnoty, tenzory prvního řádu jsou vektory, tenzory druhého řádu jsou hodnoty vyjádřené maticí (např. 3×3). Tenzorem 2. řádu lze popsat např. některé vlastnosti těles a ploch (zakřivení např.), nebo některé mechanické vlastnosti materiálů (zatížení působící na materiál, tah v jednotlivých osách...). V mozkové tkáni je rozptýl silnější ve směru neuronových vláken, lze to tedy popsat tenzorem 2. řádu.

Tenzor můžeme také interpretovat jako operátor, který transformuje jeden vektor v jiný vektor.

6.1 Vizualizace

Vizualizace tenzorů je v zásadě složitý problém. Pro každý bod ve 3D prostoru je potřeba vizualizovat 3×3 matici dat. Pokud bychom se jednoduše pokusili vyjádřit tenzor pouze skalární hodnotou, nevedla by taková vizualizace k pochopení dat.

Jednou z možných metod vizualizace tenzorů je snížení dimenze dat pomocí analýzy hlavních komponent. Další metodou může být také použití tenzorových glyfů. Pro vizualizaci pomocí glyfů je potřeba samplovat definiční obor, pro každý vzorek se poté vytvoří tenzorový glyf. Tenzorový glyf znázorňuje vlastní vektor a vlastní číslo tenzoru ve vzorkovaném bodě. Výsledným glyfem je elipsa ve 2D prostoru, nebo elipsoid ve 3D prostoru. Kromě elipsoidů lze použít i jiné tvary - kvádry, nebo válce.

Použití glyfů je analogické s glyfy a vektorovými poli. Kromě glyfů lze také tenzory zobrazit pomocí vláken. Zobrazení tenzorů pomocí vláken se často používá v anatomii, pro zobrazení např. neurony v mozku.

7 Metody vizualizace volumetrických dat, užité optické modely, způsoby klasifikace a kompozice vzorků, možnosti interpolace

Volumetrická data jsou v podstatě skalární data ve 3D prostoru. Jednou z možností vizualizace je pomocí redukce problému do 2D nařezáním na jednotlivé snímky. Další možností je sestavení 3D izoplochy pomocí algoritmu marching cubes. Třetí možností může být přímá vizualizace dat. Izoplochu je možné renderovat pomocí metody sledování paprsku.

7.1 Maximum intensity projection

Při metodě sledování paprsků je možné přímo zobrazit nejvyšší nalezenou hodnotu.

7.2 Direct volume rendering

Pomocí metody DVR je výsledná hodnota získána postupnou integrací hodnot při průběhu paprsku.

7.3 Optické modely

Je možné volit různé optické modely podle zobrazovaných dat. Pokud zobrazujeme data, která záření pouze pohlcují, můžeme vizualizovat pouze **absorpční složku**. Pokud zobrazujeme pouze průhledné materiály, které světlo propouští, můžeme zobrazovat pouze **emisní složku**. Tyto dva přístupy je také možné zkombinovat. Tato kombinace se často používá při DVR.

7.4 Klasifikace vzorků

Nejjednodušší klasifikace může být převod získaných dat do černobílého spektra. Lepší, komplexnější metodou může být klasifikace pomocí komplexnější funkce.

7.5 Interpolace

Při použití trilineární interpolace vznikají na tělese kruhové artefakty. Tento problém je možné řešit pomocí trikubické interpolace, kde se nepočítá interpolovaná hodnota z 6 vrcholů jedné krychle (trilineární), ale z vrcholů krychlí rozložených do $3 \times 3 \times 3$ matice. Vrcholů pro interpolaci tak vzniká celkem 64. (matice $4 \times 4 \times 4$ vzorků)

8 Vizualizace abstraktních dat a jejich charakteristika, možnosti vizualizace grafových struktur a vícerozměrných dat, redukce dimenze

Vizualizace abstraktních dat se primárně zabývá infografikou. Jsou to primárně spíše veličiny, které nelze jednoznačně měřit a neexistuje jeden jednoznačný způsob, jak je zobrazit. Infografika by v ideálním případě měla být reprezentativní, prezentovatelná a interaktivní. Oproti scivizu má infografika širší cílovou skupinu, měla by být tedy jednodušší.

Abstraktní data je možné vizualizovat např. pomocí tabulek (nejjednodušší). Pro vizualizaci vztahů mezi jednotlivými prvky je možné vizualizovat pomocí různých grafů, Vennových diagramů, stromů, nebo matic sousednosti.

Pro grafy je možné použít Force directed layout, který se snaží zobrazovat graf s co nejmenším překryvem cest.

8.1 Vícerozměrná data

Vícerozměrná data je možné vizualizovat např. pomocí paralelních grafů. Pro redukci náročnosti je možné redukovat dimenzi dat, např. pomocí FastMap algoritmu. Redukovat dimenze je možné v případě vynechání některých méně důležitých dat. Při vizualizaci je vhodné data rozdělit do kategorií a zachovat některé omezení. Např je vhodné zachovat vzdálenost mezi body, nebo sousednost bodů.