

RESUME DARI ARTIKEL YANG BERJUDUL MAKING DATA VISUALIZATION MORE EFFICIENT AND EFFECTIVE: A SURVEY

Rafa Aqilla Jungjunan

2024-08-28

Visualisasi data dapat mengubah suatu data yang abstrak menjadi sebuah data yang dapat dilihat, contohnya seperti panjang, posisi, bentuk, warna dan lain-lain. Sekarang banyak organisasi yang memiliki data untuk dianalisa lebih lanjut dalam membantu menyusun strategi dan keputusan operasional, dan juga memudahkan data scientist untuk menginterpretasikan hasil dari data yang sudah di analisa.

Berkembangnya visualisasi data telah banyak berkontribusi di beberapa komunitas, seperti komunitas grafik komputer telah menggunakan teknologi yang lebih maju dan mudah ditafsirkan visualisasinya. Komunitas visualisasi juga telah memudahkan pengguna dan berinteraksi dengan visualisasi. Komunitas database juga secara drastis meningkatkan eksperien dari pengguna dalam melihat dan berinteraksi dengan visualisasi data secara langsung meskipun itu big data. Visualisasi data juga telah banyak digunakan di berbagai aplikasi yang berelasi dengan banyak database seperti Excel, Google Sheets, Oracle Data Visualization Dekstop, IBM DB2, Amazon Quicksight, Microsoft Power BI dan banyak lainnya.

Alur dalam visualisasi data merupakan tipe visualisasi data yang berulang. Berdasarkan alur, 3 hal yang membuat visualisasi data lebih efisien dan efektif serta nyambung ke database peneliti, yaitu:

1. Spesifikasi Visualisasi memberikan banyak cara supaya pengguna dapat secara spesifik membuat apa yang mereka mau. Hal tersebut sangat penting untuk bagaimana menggenerasikan visualisasi data dan secara spesifik mengetahui perbedaan informasi peta ke elemen visual, namun hal tersebut mirip dengan manipulasi data seperti menyatukan kelompok dan memberi perintah operasi secara spesifik dapat memicu masalah dalam memilih desain antara bahasa database
2. Pendekatan yang efisien untuk visualisasi data harus efisien dan terukur, khususnya pada manipulasi data dan pemetaan. Banyak peneliti yang telah mencoba antarmuka mesin proses data yang kuat, dan mendesain ulang sistemnya. Ada juga solusi aproksimasi dan Solusi yang progressif untuk digunakan pada big data, hal tersebut untuk memberikan respon secara langsung.
3. Rekomendasi visualisasi data secara spesifik sulit bahkan untuk para ahli dikarenakan data mana untuk divisualisasikan, bagaimana menceritakannya, dan bagaimana memvisualisaikan merupakan latihan percobaan.

Relasi survey Beberapa survei tentang visualisasi berfokus pada topik tertentu. Pada spesifikasi visualisasi, memberikan survei tentang klasifikasi, sumber data, media klasifikasi dan lain-lain. Bahasa visualisasi dari perspektif stack dan menekan supaya Bahasa tersebut dapat digunakan dari perspektif praktis. Pendekatan yang efisien pada visualisasi data, dapat dipertimbangkan bagaimana mengintegrasikan database, visualisasi data dan analisis data supaya pengguna dapat bekerja pada satu sistem tanpa berdiskusi supaya lebih efisien.

Spesifikasi Visualisasi

1. Spesifikasi pada visualisasi data Secara general, bahasa visualisasi data terdiri dari 3 bagian, yaitu:

- data : *records, transformation*
- marks : *type, size, legend, miscellaneous*
- mapping

operasi visual GUI-based merupakan tipe yang mengartikan bahasa visualisasi data.

2. Kategori pada bahasa visualisasi data

Bahasa tingkat tinggi merangkum beberapa detail tingkat rendah dengan bawaan yang masuk akal dan banyak batasan.

Bahasa tingkat rendah dimana pengguna butuh untuk menentukam semua elemen dalam pemetaan.

Bahasa tingkat tinggi merangkum rincian konstruksi visualisasi seperti fungsi pemetaan serta beberapa properti untuk menandai seperti ukuran kanvas, legenda dan properti lainnya.

3. Visual berbasis GUI operasi

Untuk menentukan visual, banyak pengguna megggunakan cara lain untuk memberikan spesifikasi adalah dengan mengikuti “Prinsip Manipulasi Langsung” dan merupakan konsep yang telah banyak digunakan dalam aspek interaksi antara manusia dengan komputer. Tujuan visual berbasis GUI operasi adalah untuk menunjukkan berbagai cara yang dapat ditentukan pengguna visualisasi.

Data interaktif visual adalah proses eksplorasi, dimana pengguna perlu menjaga dan menyempurnakan spesifikasi seperti tambah, hapus, ubah atribut, ganti tipe chart. Eksplorasi aspek yang dapat membantu pengguna dengan mudah navigasikan visualisasi yaitu :

- penyempurnaan kueri bertahap
- navigasi faceted

4. Spesifikasi yang kurang ditentukan

Dikarenakan banyak pengguna kurang paham semua aspek dari data, tugas visualisasi adalah untuk menafsirkan input yang kurang ditentukan dari petunjuk yang diberikan pengguna. Jenis-jenis petunjuknya dapat berupa berbasis referensi, berbasis kata kunci, dan berbasis bahasa alami.

Pendekatan yang efisien untuk visualisasi data

1. Visualisasi data yang tepat

Banyak sistem visualisasi data membaca data dari database dan data tersebut dapat dimanipulasi data dari SQL menggunakan alat visualisasi untuk merender visualisasi. Dapat dilakukan terjemahan kueri pada bahasa, kemudian dapat mengintegrasikan sistem visualisasi dengan DBMS, melakukan *column stores, indexes, parallel computation*, prediksi dan prefetching, studi kasus menggunakan Kyrix dan Tableau.

2. Aproksimasi visualisasi data

Untuk mengaproksimasi visualisasi data, dapat menggunakan AQP-based, incremental sampling-based, human perception-based, range-based binning, dan konten based-binning.

3. Progressive data visualization

Dalam progressif visualisasi data, binning berbasis rentang menyediakan visualisasi dengan resolusi berbeda dengan mengubah ukuran tempat sampah. Pada rentang dan binning berbasis, menyediakan dua struktur pohon untuk eksplorasi hierarkis : HETree-R (HETree berbasis rentang) dan HETree-C (Konten berdasarkan HETree).

Rekomendasi Visualisasi

Pruning Meaningless Visualizations merupakan ada banyak sinyal atau kendala baik dari batasan yang ditentukan pengguna, dimana pengguna dapat menentukan minat visualisasi seperti kolom atau catatan data. Expertprovided constraints, merupakan beberapa kombinasi variabel, transformasi, dan pengkodean visual mungkin tidak menghasilkan visualisasi yang valid.

Setelah menghasilkan visualisasi kandidat(atau valid) dengan pemangkasan penelitian, sistem rekomendasi visualisasi akan mengenali makna visualisasi penuh berdasarkan metrik atau aturan yang telah ditentukan sebelumnya. Beberapa sistem juga dapat memberi peringkat visualisasi atau rekomendasi yang menarik kepada pengguna.

1. Specification-based recommendations

1.1 Spesifikasi yang tidak lengkap

Sistem rekomendasi visualisasi dengan spesifikasi kosong tidak memerlukan input pengguna, sedangkan sistem rekomendasi dengan spesifikasi parsial menerima input spesifikasi elemen visualisasi parsial yang diinginkan. Satu-satunya perbedaan antara spesifikasi kosong dan parsial adalah bahwa yang terakhir harus memangkas ruang pencarian dengan batasan yang ditentukan pengguna saat menghitung elemen visualisasi untuk menghasilkan kandidat visualisasi.

Beberapa rekomendasi untuk spesifikasi yang tidak lengkap yaitu :

- Rule-based visualization ranking
- Statistical Rules
- Perceptual Rules
- Machinelearning-based visualization ranking
- Learning with Soft Constraints
- Learning with Examples

1.2 Spesifikasi berbasis referensi

Beberapa sistem rekomendasi visualisasi merekomendasikan visualisasi berdasarkan data referensi atau referensi visualisasi. Biasanya sistem akan merekomendasikan visualisasi yang mirip atau berbeda dari yang diberikan referensi dalam aspek-aspek tertentu.

Aspek-aspek tersebut diantaranya adalah :

- Deviation-based
- Anomaly-based
- Similarity/Distance-based

2. Behavior-based recommendations

Sistem rekomendasi berbasis perilaku pengguna saat ini sebagai input, lalu menyimpulkan tugas yang dimaksudkan pengguna dan merekomendasikan visualisasi berdasarkan tugas mereka.

3. Personalized recommendations

Sistem personalized recommendations menangkap histori kebiasaan pengguna sebagai masukan dan merekomendasikan secara personal ketertarikan visual. Personalized recommendations dapat berupa :

- Linear Model
- Collaborative Filtering : collaborative Filtering, content-based Filtering, hybrid Filtering

4. A summary

Solusi berbasis aturan berdasarkan pemahaman intuitif seseorang tentang visualisasi, tetapi tidak membuat pemahaman lengkap tentang persepsi manusia, hanya berfokus pada beberapa metrik yang menarik. Solusi pembelajaran mesin perlu mengumpulkan data pelatihan, dan hasilnya sulit ditafsirkan, tetapi menangkap pengetahuan kognitif manusia tentang efektivitas manusia.

Other research directions

1. Data preparation for data visualization

Pada aslinya data tipe yang tidak bersih dan secara visual data sering menipu pengguna. Fenomena ini sering disebut sebagai visualisasi bias dari komunitas visualisasi data. Hasil dari data yang kotor adalah :

- What-if Analysis for Outliers
- Evaluating Visualizations with Missing Data
- Detecting biased visualizations
- Task-aware data cleaning

2. Data visualization benchmarks

Sangat penting untuk membuat tolak ukur pada kinerjandaan rekomendasi. Tolak ukur harus mengabdikan pada tugas analisis visual, memberikan jejak dan data yang dapat digunakan kembali, dan dalam hal rekomendasi, memiliki cakupan dan kualitas labelnya yang tinggi.

Hal yang harus banyak dilakukan adalah peluang penelitian, yaitu :

- Categorization of visualizations
- Training data

3. Data visualization for database-related applications

Visualisasi data juga berperan penting dalam aplikasi terkait database. Dengan perkembangan pesat teknik visualisasi, ada lebih banyak peluang tentang menggunakan visualisasi data untuk aplikasi terkait database. Peluang tersebut adalah :

- Data visualization for data discovery
- Data visualization for data debugging

Kesimpulan

Visualisasi data adalah bidang yang berkembang pesat dengan banyak hasil penelitian baru dan sistem baru dikembangkan baru-baru ini. Peneliti dan praktisi dari banyak bidang telah berkontribusi hingga keberhasilan visualisasi data yang luar biasa, didorong oleh sebagian besar domain dan aplikasi. Secara komprehensif menggambarkan karya-karya dalam visualisasi spesifikasi, metode yang efisien untuk visualisasi data, dan rekomendasi visualisasi.