Важливим елементом (характеристикою) управлінської звітності є можливість її візуалізувати. Візуалізація управлінської звітності – подання обліково-аналітичної інформації у вигляді, який сприяє її найкращому розумінню користувачами (управлінським персоналом). Слід відмовитися від стереотипів сприйняття управлінської звітності як таблиці з даними. Управлінський звіт може бути поданий у вигляді діаграми, графіків, таблиць, рисунків тощо, які сприймаються користувачами інформації краще, ніж таблиці. Правильно сформований звіт дає змогу не лише скоротити час на сприйняття інформації, а й прискорити прийняття управлінських рішень, тому при розробці системи управлінської звітності компанії фактор візуального сприйняття інформації має бути врахований [2]. Від цього залежатиме оперативність, обґрунтованість та ефективність прийняття рішень управлінським персоналом. Тобто управлінська звітність повинна бути не просто набором текстової та/або цифрової інформації, а певним чином упорядкована, систематизована та представлена; слід уникати дублювання показників та додавання непотрібних, бо це може призвести до її викривлення.

Список використаних джерел

1. Король С. Я. Управлінська звітність: сутність і алгоритм формування // Бізнес Інформ. 2014. № 7. С. 325–331.

2. Зима Ю. П. Розробка моделі управлінської звітності з використанням інформаційних технологій // Економічний аналіз. 2014. Т. 158, № 2. С. 148–153.

ОСОБЛИВОСТІ, ПРОБЛЕМИ ТА СУЧАСНІ ІНСТРУМЕНТИ ВІЗУАЛІЗАЦІЇ BIG DATA

Лазебник Юлія Олександрівна,
доктор економічних наук, доцент,
професор кафедри статистики, обліку та аудиту,
Харківський національний університет імені В.Н. Каразіна

Візуалізація даних представляє дані в деякій систематизованій формі, включаючи атрибути та змінні для одиниці інформації. Методи виявлення даних на основі візуалізації дозволяють діловим користувачам збільшити різні джерела даних для формування спеціальних аналітичних поглядів. Розширена аналітика може бути інтегрована в методи підтримки створення інтерактивної та анімованої графіки на настільних комп’ютерах, ноутбуках або мобільних пристроях, таких як планшети та смартфони [5].

Великі дані – це великий об’єм, висока швидкість та/або набір даних з високим розмаїттям, які потребують нових форм обробки, щоб забезпечити розширену оптимізацію процесів виявлення, розуміння та прийняття рішень. Проблеми великих даних полягають у збиранні, зберіганні, аналізі, обміні, пошуку та візуалізації даних [1].

Візуалізація використовується для створення таблиць, діаграм, зображень та інших інтуїтивно зрозумілих способів відображення даних. Візуалізація великих даних не така проста, як традиційні невеликі набори даних. Розширення традиційних підходів до візуалізації вже здійснюється, але далеко не достатньо. Під час широкомасштабної візуалізації даних багато дослідників використовують витяг функцій та геометричне моделювання, щоб значно зменшити розмір даних до фактичного надання даних. Вибір правильного подання даних також дуже важливий при візуалізації великих даних [1].

Метою цього дослідження є представлення нових методів та досягнень візуалізації великих даних за допомогою впровадження звичайних методів візуалізації та розширення деяких з них на обробку великих даних, обговорення проблем візуалізації великих даних та аналіз прогресу технологій у візуалізації великих даних.

Часто застосовується багато традиційних методів візуалізації даних. Це: таблиця, гістограма, графік розсіювання, лінійна діаграма, смугова діаграма, кругова діаграма, діаграма з областями, блок-схема, множинні серії даних або комбінація діаграм, часова лінія, діаграма Венна, схема потоку даних та діаграма взаємозв'язків. Додатковими методами є: паралельні координати (parallel coordinates), карта дерева (treemap), конусне дерево (cone tree) та семантична мережа (semantic network) тощо [4].

Паралельні координати використовуються для побудови окремих елементів даних у багатьох вимірах. Паралельна координата дуже корисна при відображенні багатовимірних даних. Карта дерева – це ефективний метод візуалізації ієрархій. Розмір кожного прямокутника являє собою одну міру, в той час як колір часто використовується для представлення іншого вимірювання даних. Конусне дерево – це ще один метод відображення ієрархічних даних, таких як організаційний орган у трьох вимірах. Гілки ростуть у вигляді конуса. Семантична мережа – це графічне зображення логічних зв'язків між різними поняттями. Він створює спрямований графік, комбінацію вузлів або вершин, країв або дуг і мітку над кожним краєм [4].

Візуалізації не лише статичні; вони можуть бути інтерактивними. Інтерактивну візуалізацію можна здійснити за допомогою таких підходів, як збільшення (збільшення та зменшення масштабу), огляд та деталізація, масштабування та панорама, фокус та контекст [4].

Нові технології бази даних та перспективні підходи до візуалізації на базі Інтернету можуть бути життєво важливими для зниження витрат на створення візуалізації та надання їм допомоги у вдосконаленні наукового процесу. Через технології на основі веб-зв'язку візуалізація змінюється у міру зміни даних, що значно скорочує зусилля для того, щоб візуалізації були своєчасними та актуальними. Ці «низькі» візуалізації часто використовуються в бізнес-аналітиці та відкритих урядових системах даних, але вони, як правило, не використовуються в науковому процесі.

Масштабованість та динаміка – дві основні проблеми візуальної аналітики. Методи аналітики, що базуються на візуалізації, приймають

виклики, представлені «чотирма VS» великих даних, і перетворюють їх на наступні можливості [4]:

– обсяг (*Volume)*: методи розроблені для роботи з величезною кількістю наборів даних і дозволяють отримати значення з великого обсягу даних;

– різноманітність (*Variety)*: методи розроблені для об'єднання стількох джерел даних, скільки потрібно;

– швидкість (*Velocity)*: за допомогою методів підприємства можуть замінити пакетну обробку потоковою обробкою в режимі реального часу;

– цінність (*Value)*: методи не лише дозволяють користувачам створювати привабливі інфографіки та теплові карти, але й створюють ділову цінність, отримуючи уявлення про великі дані.

Візуалізація великих даних з їх різноманітністю та неоднорідністю (структуровані, напівструктуровані та неструктуровані) є великою проблемою. Швидкість – бажаний фактор для аналізу великих даних. Розробити новий інструмент візуалізації з ефективним індексуванням у великих даних непросто. Хмарні обчислення та сучасний графічний інтерфейс користувача можуть бути об'єднані з великими даними для кращого управління масштабністю великих даних [6].

Системи візуалізації повинні взаємодіяти з неструктурованими формами даних, такими як графіки, таблиці, текст, дерева та інші метадані. Великі дані часто мають неструктуровані формати. Через обмеження пропускної здатності та вимоги до потужності візуалізація повинна наближатися до даних, щоб ефективно витягувати змістовну інформацію. Через великий розмір даних необхідність масової паралелізації є проблемою у візуалізації. Завдання паралельних алгоритмів візуалізації – це розкладання проблеми на самостійні завдання, які можна виконувати одночасно [6].

Ефективна візуалізація даних є ключовою частиною процесу відкриття в епоху великих даних. Для викликів високої складності та великої розмірності великих даних існують різні методи зменшення розмірності. Однак вони не завжди можуть бути застосовані. Чим більше розмірності візуалізуються ефективно, тим вищі шанси розпізнати потенційно цікаві зразки, кореляції чи інші структури.

Існують також такі проблеми щодо візуалізації великих даних [4]:

– візуальний шум: більшість об'єктів у наборі даних занадто відносно один одного. Користувачі не можуть розділити їх на окремі об'єкти на екрані;

– втрата інформації: зменшення наборів видимих даних можна використовувати, але це призводить до втрати інформації;

– велике сприйняття зображення: методи візуалізації даних обмежуються не лише співвідношенням сторін та роздільною здатністю пристрою, але й обмеженнями фізичного сприйняття;

– висока швидкість зміни зображення: користувачі спостерігають за даними та не можуть реагувати на кількість змін даних та їх інтенсивність на дисплеї;

– високі вимоги до продуктивності: навряд чи можна помітити статичну візуалізацію через менші вимоги до швидкості візуалізації – висока вимога до продуктивності.

У програмах Big Data важко провести візуалізацію даних через величезні розміри великих даних. Більшість сучасних інструментів візуалізації Big Data мають низькі показники масштабування, функціональності та часу реагування. Невизначеність може спричинити великий виклик для ефективної візуалізації, що усвідомлює невизначеність, і виникати в процесі візуальної аналітики [1].

Було представлено потенційне рішення деяких викликів чи проблем щодо візуалізації великих даних [4]:

1. Задоволення потреби в швидкості: одне можливе рішення – апаратне забезпечення. Можна використовувати більшу пам'ять та потужну паралельну обробку. Іншим методом є введення даних у пам'ять, але з використанням підходу до обчислення сітки, де використовується багато машин.

2. Розуміння даних: одне рішення – мати належну експертизу домену.

3. Адресація якості даних: Необхідно забезпечити чистоту даних через процес управління даними або управління інформацією.

4. Відображення значущих результатів: Один із способів – це кластеризація даних у вигляд вищого рівня, де видно менші групи даних і дані можуть бути ефективно візуалізовані.

5. Робота з чужими людьми: Можливим рішенням є видалення залишків із даних або створення окремої діаграми для людей, що переживають люди.

Щодо того, як має бути спроектована візуалізація в епоху великих даних, підходи до візуалізації спочатку повинні забезпечити огляд, потім дозволити масштабування та фільтрацію та надати глибокі деталі за вимогою. Візуалізація може зіграти важливу роль у використанні великих даних для повного перегляду клієнтів. Стосунки є важливим аспектом багатьох сценаріїв великих даних. Соціальні мережі, мабуть, найвизначніший приклад і їх дуже важко зрозуміти в текстовому або табличному форматі; однак, візуалізація може зробити очевидними тенденції та закономірності мережі. Для візуалізації невідповідних стосунків користувачів у соціальній мережі запропоновано хмарний метод візуалізації. Метод може інтуїтивно представити соціальні відносини користувачів на основі кореляційної матриці, щоб представити ієрархічну залежність вузлів користувачів соціальної мережі. Крім того, метод використовує Hadoop на основі хмари для розподіленої паралельної обробки візуалізації, що допомагає прискорити великі дані соціальної мережі [3].

На платформі Hadoop працює багато інструментів візуалізації великих даних. Найпоширенішими модулями в Hadoop є: Hadoop Common, Файлова система розподіленої Hadoop (HDFS), Hadoop YARN та Hadoop MapReduce. Вони ефективно аналізують великі дані, але не мають достатньої візуалізації. Розроблено певне програмне забезпечення з функціями візуалізації та взаємодії для візуалізації великих даних [6]:

Pentaho: підтримує спектр функцій BI, таких як аналіз, приладова панель, звітність корпоративного класу та обмін даними.

Flare: це бібліотека ActionScript для створення візуалізації даних, що працює в Adobe Flash Player.

JasperReports: має новий програмний рівень для генерування звітів із великих сховищ даних.

Dygraphs: це швидка та еластична колекція графіків з відкритим кодом JavaScript, яка допомагає виявити та зрозуміти непрозорі набори даних.

Platfora: перетворює необроблені великі дані в Hadoop в інтерактивну систему обробки даних. Він має модульну функціональність двигуна даних пам'яті.

ManyEyes: це інструмент візуалізації, запущений IBM. Це публічний веб-сайт, на якому користувачі можуть завантажувати дані та створювати інтерактивну візуалізацію.

Tableau: це програмний інструмент для бізнес-розвідки (BI), який підтримує інтерактивний та візуальний аналіз даних. Він має двигун даних в пам'яті для прискорення візуалізації. Tableau має три основні продукти для обробки масштабних наборів даних, включаючи Tableau Desktop, Tableau Sever та Tableau Public.

Підводячи підсумки проведеного дослідження, слід зазначити, що великі набори даних, що змінюються у часі, формують серйозний виклик процесу візуалізації даних. Візуалізація даних у режимі реального часу може дати можливість користувачам активно реагувати на виникаючі проблеми. Важко переоцінити достоїнства технік візуалізації даних. Це потужний і широко застосовуваний інструмент для аналізу та інтерпретації великих і складних за структурою даних, який повинен доносити складні ідеї чітко, точно та ефективно.

Список використаних джерел

1. Chen C.L. P., Zhang C.-Y. Data-intensive applications, challenges, techniques and technologies: A survey on Big Data. *Information Sciences*. 2014. Vol. 275 (10). P. 314-347.

2. Keahey T.A. Using visualization to understand big data. Technical Report. IBM Corporation. 2013. P. 1-16.

3. Kim Y., Ji Y.-K., Park S. Social Network Visualization Method using Inherence Relationship of User Based on Cloud. *International Journal of Multimedia and Ubiquitous Engineering*. 2014. Vol. 9(4). P. 13-20.

4. Lidong W., Guanghui W., Cheryl A.A. Big Data and Visualization: Methods, Challenges and Technology Progress. *Digital Technologies*. 2015. Vol. 1. No. 1. P. 33-38. URL: https://pdfs.semanticscholar.org/2975/4e4295a9ce4d51937c0712d6482634474628.pdf (дата звернення: 29.10.2019).

5. Matthew N.O. Sadiku, Adebowale E. Shadare, Sarhan M. Musa, Cajetan M. Akujuobi. Data Visualization. *International Journal of Engineering Research*

And Advanced Technology (IJERAT). 2016. Vol. 02. Issue. 12. URL: https://ijerat.com/uploads/2/3222_pdf.pdf (дата звернення: 29.10.2019).

6. Sucharitha V., Subash S.R., Prakash P. Visualization of Big Data: Its Tools and Challenges. *International Journal of Applied Engineering Research.* 2014. Vol. 9(18). P. 5277-5290.

ВИКОРИСТАННЯ БІБЛІТЕКИ GGPLOT2 ДЛЯ ВІЗУАЛІЗАЦІЇ ДАНИХ

Марець Оксана Романівна,
кандидат економічних наук, доцент кафедри статистики,
Львівський національний університет імені Івана Франка

Одним із популярних нині інструментів для візуалізації та аналізу статистичних даних є середовище та мова програмування R. Найбільш універсальна сфера їх застосування – це аналітика. R дає можливість проводити статистичні тести, перевіряти гіпотези, будувати графіки та робити прогнози. Перевагами використання мови програмування R для аналізу є такі:

– це програмне забезпечення з відкритим кодом;

– мова програмування з відносно інтуїтивним кодом, оскільки створена для аналізу даних;

– відтворюваність результатів аналізу, тобто будь-хто може подивитися на код і з'ясувати, як був зроблений аналіз;

– повторюваність результатів аналізу: якщо з'являються нові дані, можна легко запустити вже написаний код;

– дружня спільнота з блогами, статтями, форумами, де дадуть відповідь на питання різного рівня складності;

– наявність великої кількості бібліотек, призначених для виконання найрізноманітніших специфічних задач.

Одна з цих бібліотек – ggplot2 – представляє засоби для візуалізації даних і нині її вважають серед найкращих у цій сфері. Вона є частиною системи tidyverse – комплексу бібліотек, призначених для завантаження, обробки, очищення, візуалізації та аналізу даних. Основними бібліотеками tidyverse є ggplot2, dplyr, tidyr, readr, purrr, tibble, stringr та forcats, які надають інструменти для моделювання, перетворення та візуалізації даних.

Хочемо зазначити, що безпосередній візуалізації передує важливий і часто трудомісткий етап роботи з завантаження та підготовки даних. Саме для легкого виконання цих процесів призначені бібліотеки системи tidyverse. Ця система базується на принципах чистих даних (tidydata) та граматики графіки. *Чисті дані* (tidydata) – набори даних, де кожна змінна (показник) – це стовпець, а кожне спостереження (одиниця сукупності) – рядок. Термін *граматика графіки* запропонував Л.Уілкінсон у 1999 р. Вона охоплює два