

BIG DATA (ВЕЛИКІ ДАНІ) В ІНФОРМАЦІЙНО-ВИМІРЮВАЛЬНИХ СИСТЕМАХ

© М. Глушко, В.Кучерук, 2017

Вінницький національний технічний університет, Вінниця, Україна

У техніці використовується як термін Big Data, так і термін „великі дані”. Великі дані не мають строгого визначення. Не можна провести чітку межу - це 10 терабайт або 10 мегабайт? Сама назва дуже суб'єктивна. Слово „велике” - це як „один, два, багато” у первісних племен.

Однак є усталена думка, що великі дані - це сукупність технологій, які покликані здійснювати три операції. По-перше, обробляти великі в порівнянні з „стандартними” сценаріями обсяги даних. По-друге, вміти працювати з даними, що швидко поступають в дуже великих обсягах. Тобто даних не просто багато, а їх постійно стає все більше і більше. По-третє, вони повинні вміти працювати зі структурованими і погано структурованими даними паралельно в різних аспектах. Великі дані припускають, що на вхід алгоритми отримують потік не завжди структурованої інформації, і що з нього можна витягти більше ніж одну ідею.

Типовий приклад великих даних - це інформація, яка надходить з різних фізичних експериментальних установок - наприклад, з Великого адронного колайдера, сенсори якого постійно виробляють величезну кількість даних. Установка безперервно видає великі обсяги даних, а вчені з їх допомогою вирішують паралельно безліч завдань.

Багато що з того, що стосується великих даних, підходів, які використовуються для їх аналізу, насправді існує досить давно. Наприклад, обробка зображень з камер спостереження, коли ми говоримо не про одно зображення, а про потік даних. Або навігація роботів. Все це існує десятки років, просто зараз завдання з обробки даних торкнулися набагато більшої кількості людей та ідей. Багато розробників звикли працювати зі статичними об'єктами і мислити категоріями станів. У великих даних парадигма інша. Ми повинні вміти працювати з безперервним потоком даних, і це цікава задача. Вона зачіпає все більше і більше областей.

В даній роботі, пропонується застосовувати термінологію „великих даних” для роботи із інформаційно-вимірювальними системами. Уявімо, що нам потрібно виміряти температуру на виробництві. Якщо у нас тільки один сенсор температури у будівлі, необхідно переконатися, що він працює точно і безперервно. Якщо ж для кожного структурного елемента виробництва встановлений окремий сенсор, вірогідно, рано чи пізно якийсь із них почне надавати неправильні дані. Отримані дані можуть бути більш точними, ніж з одного сенсора. Будь-який із окремо взятих показників може бути помилковим, але в сукупності велика кількість показників дадуть правдоподібну картину. При цьому отримувану інформацію можна обробляти також із використанням методів детермінованого хаосу [1 - 3].

Візьмемо для прикладу безпроводні сенсори, впроваджені на виробництві. На всій території нафтопереробної фабрики BP Cherry Point в Блейні (Вашингтон) розставлені безпроводні сенсори, утворюючи невидиму мережу, яка виробляє величезні об'єми даних в режимі реального часу. Несприятливі навколишні умови можуть час від часу спотворювати покази, приводячи до похибки. Проте велика кількість інформації, яка надходить, компенсує це. Вимірюючи навантаження на труби безперервно, а не через певні проміжки часу, компанія BP виявила, що деякі види сирової нафти більш їдкі, ніж інші. Раніше це не вдавалось визначити, а значить, і запобігти.

Отже, отримуючи величезні масиви даних нового типу, в деяких випадках, можна знехтувати точністю, якщо вдається спрогнозувати загальні тенденції. Проте, при надходженні та обробленні великої кількості даних, наприклад декілька зеттабайт в інформаційну систему, ми можемо зменшувати похибку. Також, при цьому є можливість прогнозування роботи системи із надзвичайно високою точністю, оскільки чим більше даних - тим точніше результат. Для розуміння: Екзатабайт - мільярд гігабайт, зеттабайт - в тисячу разів більше екзатабайт.

Маючи таку велику кількість даних, можливо не тільки прогнозування результатів вимірювання або різних операцій системи, але і роботи самого обладнання на виробництві: коли і саме який сенсор із певною вірогідністю вийде з ладу, що дозволить суттєво скоротити проблеми на виробництві.

1. Виктор Майер-Шенбергер Кеннет Кукьер. *Большие данные (Big Data). Революция, которая изменит то, как мы живем, работаем и мыслим.* // Издательство „Манн, Иванов и Фербер”, Москва, 2014 г. ст. 48. 2. Rick Smolan, Jennifer Erwit. *The Human Face of Big Data First Edition Edition* //, USA 2014, p 234. 3. John W. Foreman. *Data Smart: Using Data Science to Transform Information into Insight* // 1st Edition, Kindle Edition, 2016, USA 2016 p. 132.