

*Орлов С. С., студентка I курсу
СО «Магістр» спеціальності
113 «Прикладна математика»
Потапова Н. А., к.е.н, доцент,
доцент кафедри інформаційних
технологій*

ОСОБЛИВОСТІ ВИКОРИСТАННЯ BIG DATA В ІНФОРМАЦІЙНОМУ АНАЛІЗІ

Донецький національний університет імені Василя Стуса, м. Вінниця

Big Data (великі дані) – це дані, які є настільки комплексними, великими та швидкими, що їх обробка традиційними методами стає неможливою [1]. Концепція великих даних була сформована приблизно на початку 2000-х років. Сформулював концепцію великих даних галузевий аналітик Даг Лейні, який визначив правило трьох “V”:

1. Volume (об’єм) – корпорації зберігають значні обсяги інформації з соціальних мереж, транзакцій, розумних пристроїв, тощо. Раніше зберігати таку кількість даних було надзвичайно дорого, або й взагалі неможливо. Але ситуація змінилась через здешевлення сховищ, хмарних технологій.

2. Velocity (швидкість) – з кожним роком кількість пристроїв, що беруть участь в обміні інформацією зростає великими темпами. Наразі це не лише комп’ютери та смартфони, а й розумні помічники, датчики систем безпеки розумних домів та ін. Це змушує компанії обробляти дані, які надходять, в реальному часі.

3. Variety (різноманітність) – дані надходять не лише у звичайних числових форматах, які можливо швидко обробляти. Це можуть бути текстові документи, фото, відео, інші медійні формати, дані з фінансових бірж тощо.

Виникає питання про необхідність використання big data. За словами Уейна Томпсона, голови відділу Data Science у компанії SAS Institute: “Глибоке навчання потребує великих даних, тому що великі дані необхідні для виділення прихованих закономірностей і пошуку відповідей без переобладнання даних. Завдяки глибокому навчанню, чим більше якісних даних у вас є, тим кращі результати”. [2]

З великої кількості якісних великих даних та із використанням машинного глибокого навчання, замість традиційних методів обробки інформації, дослідники можуть знаходити такі приховані закономірності, які неможливо було б побачити, якщо б такі об’єми даних намагалися досягнути традиційними методами обробки даних. Таким чином, обґрунтованим є зв’язок між великими даними та їх використанням в засобах глибокого навчання.

Важливим є не великі дані самі по собі, а методи їх обробки, тобто їх цінність полягає в тому, як їх використовувати. Ефективність використання

даних обумовлена першоджерелом їх виникнення та направлена на вирішення ряду задач, зокрема:

- полегшення управління ресурсами;
- підвищення ефективності операцій;
- оптимізацію виготовлення продукції;
- створення нових можливостей для збільшення прибутку та зростання бізнесу;
- використання в системах прийняття рішень.

Поєднавши великі дані та високоефективну аналітику можна спростити виконання таких бізнес-завдань, як:

- визначення причин збоїв або появи дефектів в реальному часі;
- розпізнавання різноманітних аномалій;
- покращення розуміння стану та діагнозу пацієнтів за рахунок пришвидшення аналізу їх даних;
- перерахунок ризиків для усього портфеля за лічені секунди;
- загострення точності моделей глибокого навчання та покращення швидкості та коректності їх реагування на зміни;
- виявлення шахрайської поведінки до того, як вона встигне вплинути на організацію.

Як доказ реальної корисності великих даних можна навести наступні приклади застосування їх в наукових дослідженнях:

1. У 2000-х роках розпочалася дослідницька кампанія Sloan Digital Sky Survey (SDSS), що збирала дані зі швидкістю 200 Гб за ніч. З такою швидкістю за перші кілька тижнів даних було зібрано більше, ніж за всю попередню історію астрономії, всього було зібрано більш ніж 140 Тб даних, але за прогнозами телескоп-наступник буде збирати таку кількість даних за 5 днів досліджень [2].

2. Центр симуляції клімату NASA (NCCS) зберігає 32 Пб кліматичних спостережень і моделювання на суперкомп'ютерному кластері Discover [3].

3. Раніше розшифровка геному людини потребувала більше 10 років, а тепер достатньо й одного дня.

4. Компанія Google створила базу даних DNABank, яка збирає та впорядковує дані ДНК з усього світу для виявлення генетичних аномалій та захворювань. Завдяки цьому значно підвищується швидкість аналізу генетичних даних, а також зменшується ризик “людського фактору”, що впливає на кількість потенційних помилок та неправильних висновків [4].

5. Також під час пандемії COVID-19 великі дані допомогли мінімізувати поширення вірусу, завдяки тому що уряди багатьох країн збирали та аналізували дані захворювань, переміщень та інші. Найбільших досягнень на цьому полі досягли Південна Корея та Ізраїль [5].

6. Також великі дані успішно використовуються для вирішення бізнес задач, аналізу політичного становища, для таргетингу реклами, прогнозування погоди та результатів спортивних матчів, тощо.

Отже, великі дані та машинне глибоке навчання – дуже важливі інструменти у всіх сферах життя, тому їх необхідно освоювати та використовувати для досягнення вищих цілей, ніж тих, що можна було

досягнути раніше.

Список літературних джерел.

1. Мартін Гілберт, Прісцилла Лопез. "The World's Technological Capacity to Store, Communicate, and Compute Information", 2011.
2. "Data, data everywhere". The Economist. 25 лютого 2010.
3. Філ Вебстер. "Supercomputing the Climate: NASA's Big Data Mission". CSC World. Computer Sciences Corporation. 4 січня 2013.
4. "DNAstack tackles massive, complex DNA datasets with Google Genomics". Google Cloud Platform. 1 жовтня 2016.
5. Аміт Рой Клаудхарі. "Gov in the Time of Corona". Gov Insider. 27 березня 2020.

УДК 004.6:004.82

*Сірков О. О., студент курсу спеціальності
122 «Комп'ютерні науки»*

*Антонов Ю.С. кан. фіз.-мат. наук, доцент,
доцент кафедри інформаційних технологій*

ДЕТЕКТУВАННЯ ЗБРОЇ НА ФОТОГРАФІЯХ АБО У ВІДЕОПОТОЦІ ЗА ДОПОМОГОЮ БІБЛІОТЕК КОМП'ЮТЕРНОГО ЗОРУ

Донецький національний університет імені Василя Стуса, м. Вінниця

Системи комп'ютерного зору в багатьох випадках вирішують завдання по доповненню або заміні людини в областях діяльності, пов'язаних зі збором і аналізом зорової інформації [1-2]. Розпізнавання зброї на фотографіях або кадрах відеопотоку може дозволити попереджати кримінальні злочини або знаходити заборонений контент у мережі Інтернет. Тому дослідження у даній галузі є актуальними. Серед існуючих аналогів систем розпізнавання зброї можна виділити програмне забезпечення від компаній “СпецЛаб” та “PatriotOne”, а також систему “ZeroEyes”, яку було інтегровано в школах в Америці [3].

Для задач розпізнавання в умовах використання великих об'ємів даних необхідно, щоб модель мала високу здатність до навчання та великий відсоток правильних припущень щодо ознак зображення. У порівнянні з традиційними нейронними мережами прямого поширення зі схожою кількістю шарів, згорткові нейронні мережі мають вищу здатність до навчання, оскільки містять набагато менше параметрів та зв'язків [4].

Нейронна мережа Mask R-CNN концептуально проста: нейронна мережа Faster R-CNN має два виходи для кожного об'єкта-кандидата, мітку класу та зміщення обмежувальної рамки; до цього додається третя гілка, що виводить маску об'єкта [5].