

## Прогнозування розвитку конфліктів гібридного типу на основі Big Data

**Резюме.** Розглядається проблема прогнозування розвитку конфліктів гібридного типу (*Hybrid War*), які включають одночасне протистояння в різних сферах та значну кількість об'єктів і суб'єктів конфлікту. Розглядається можливість застосування великих даних (*Big Data*) та методів їх обробки для прогнозування їх розвитку. Пропонується послідовність обробки великих обсягів різнорідних даних, які отримуються як з оцінок воєнно-політичної обстановки, так і з району конфлікту.

**Ключові слова:** конфлікти гібридного типу; великі дані; технології обробки великих даних; аналіз великих даних; прогнозування.

**Постановка проблеми.** Глобалізація і інформаційно-технологічна революція стали інтеграторами класичних і нових форм, способів і технологій сучасних конфліктів. Набули поширення конфлікти гібридного типу, які направлені на досягнення політичних цілей і характеризуються збільшенням кількості об'єктів, які задіяні в конфлікті, їх різними комбінаціями, одночасним проведенням декількох фаз конфлікту, і застосуванням військових та невійськових засобів [1]. Такі конфлікти характеризуються петабайтами ( $10^{15}$  байта) даних, які генеруються в процесі конфлікту. Аналіз поточного стану таких конфліктів потребує врахування значної кількості різнорідних параметрів (політичних, військових, економічних, інформаційних тощо), а прогнозування наступних фаз – встановлення кореляцій між об'єктами і суб'єктами конфлікту, які, як правило, ретельно приховуються [2]. На відміну від конфліктів минулого століття, прогнози щодо розвитку конфліктів гібридного типу на основі причинно-наслідкових зв'язків забезпечують рівень вірогідності їх реалізації не вище 0,8 [1]. Актуальним є пошук якісно нових підходів до прогнозування розвитку конфліктів гібридного типу.

**Аналіз останніх досліджень та публікацій.** Різним аспектам конфліктів гібридного типу війни присвячена значна кількість досліджень як провідних аналітичних центрів світу, так і низки вітчизняних науковців [1–3]. Незважаючи на те, що цей термін почав широко використовуватися в наукових публікаціях з 2014 року, проблема прогнозування фаз розвитку таких конфліктів не достатньо

теоретично обґрунтована і потребує подальшого вивчення.

Технології великих даних, як і конфлікти гібридного характеру, є принципово новою технологією. За результатами бібліографічного пошуку в базі даних Pro Quest Military Database знайдено 571 наукову статтю, які містять словосполучення "*Big Data*". Понад дві третини авторів цих статей дають своє визначення великих даних, тобто в ключових словах міститься словосполучення "definition". Швидше за все, технології великих даних, як і конфлікти гібридного типу, вміщують значно більше ознак, ніж ті, що можливо визначити відповідними дефініціями. У межах дослідження визначимо великі дані (*Big Data*) як сукупність обсягів інформації одного контексту, що безперервно збільшуються, але різних форматів подання, а також методів і засобів для ефективної та швидкої обробки.

Методи аналізу великих даних та питання їх практичного застосування для вирішення широкого кола завдань досліджені в монографії Кейта [4]. Монографія [5] присвячена проблемним питанням застосування великих даних під час планування операцій коаліційних сил в Іраку. Авторами встановлено прямий зв'язок між обсягом даних, який використовувався під час оперативного планування та кількістю жертв під час бойових дій. У роботі [6], яка виконана групою дослідників університету Сил спеціальних операцій США, проведено аналіз використання великих даних під час планування операції в Афганістані. Як один із прикладів, розглянуто вплив погодних умов, дані за якими за тривалий період спостереження були у розпорядженні

командування, на кількість збройних нападів на військові підрозділи. Розглядаються окремі питання прогнозування розвитку конфліктів на основі технологій великих даних. У [7] розглядаються питання застосування великих даних на тактичному рівні. У статті [8] надаються рекомендації щодо внесення змін у стратегії національної безпеки та воєнні доктрини з питань застосування великих даних під час оборонного та оперативного планування, підготовки військових аналітиків зважаючи на аналіз великих даних та впровадження технологій великих даних і методів їх аналізу для забезпечення національної безпеки і оборони.

Водночас великі дані є другим за значимістю трендом розвитку сучасного суспільства (після віртуальності). Такі країни, як США, Великобританія, Японія, Австралія вже на законодавчому рівні визначають великі дані як найважливіший ресурс стратегічного значення [9].

Тим часом питання аналізу та прогнозування розвитку конфлікту гібридного типу між Україною і Росією [1], з великою кількістю задіяних об'єктів і суб'єктів, на основі значного обсягу накопичених за сім років однорідних та неоднорідних даних у науковій літературі досліджено не повністю.

**Мета статті** – обґрунтування можливості застосування технологій великих даних для аналізу та прогнозування розвитку конфлікту гібридного типу та визначення послідовності їх обробки й аналізу.

**Виклад основного матеріалу.** Існує множина визначень конфліктів гібридного

типу і ці визначення продовжують еволюціонувати [1]. У сукупності ці терміни увібрали складний характер війни XXI ст., до якої залучена множина суб'єктів, розмиті відмінності між традиційними видами збройних конфліктів і навіть між поняттями війни і миру. У традиційній війні конвенційні і нерегулярні операції мають тенденцію проводитися одна за одною, операції нерегулярних сил, як правило, вторинні порівняно з військовими кампаніями з використанням конвенціональних військових сил. Конфлікти гібридного типу включають застосування конвенціональних і нерегулярних засобів одночасно, за єдиним задумом, гібридна агресія за своєю природою асиметрична. Асиметрія природним чином включає використання воєнних, невоєнних і невійськових інструментів у “сірій зоні” між війною і миром [2].

Аналіз останніх досліджень та публікацій дав змогу встановити чіткий зв'язок між революційними змінами в сучасних інформаційних технологіях та застосуванням нових підходів до ведення сучасних конфліктів і війн, який наведений у Табл. 1. Квінтесенцією наступного етапу розвитку технологій є набуття можливостей збору, накопичення та аналізу великих обсягів даних стало розроблення та практичне застосування стратегій ведення конфліктів гібридного типу. Будь-яка протидія гібридній агресії стає неможливою без наявності та застосування низки сучасних технологій, які стають елементами систем великих даних.

Таблиця 1

**Зв'язок між розвитком інформаційних технологій і сучасними воєнними стратегіями**

Інформаційна технологія	Воєнні стратегії і операції
Мережеві технології	Мережецентричні бойові дії і операції
Хмарні технології	Єдина інформаційне поле збройних сил – Об'єднане багатодоменне командування і управління [10] – багатодоменні операції [11]
Технології великих даних ( <i>Big Data</i> )	Конфлікти і операції гібридного типу

Інформаційно-технологічні аспекти конфліктів гібридного типу полягають у здатності синхронізувати декілька засобів сили одночасно та використати інформаційно-аналітичні підходи, багатозначність, нелінійний характер та когнітивні елементи війни. Конфлікти гібридного типу ведуться, як правило, нижчою інтенсивністю, ніж пороги виявлення та реагування на класичні загрози воєнного характеру, і на пряму залежать від швидкості, обсягу та засобів збору та поширення даних. Кількість таких конфліктів

та обсяги даних, що їх супроводжують, зростатиме за геометричною прогресією [1].

Однією з визначальних рис опису сучасного конфлікту гібридного типу є використання різних типів структурованих та неструктурованих даних. Як правило, аналіз поточної фази конфлікту здійснюється описом даних DIME/PMESII/ASCOPE:

DIME – дипломатія, інформація, збройні сили та економіка;

PMESII – політичні, безпекові, економічні, соціальні, інформаційні та інфраструктурні параметри;

ASCOPE — райони, структури, спроможності, організація, люди та події.

Починаючи з певних числових значень обсягу даних і далі за мірою їх зростання, відбуваються якісні зміни, які дають змогу з великим рівнем імовірності визначити нові взаємозв'язки (кореляції) об'єктів і суб'єктів конфлікту, інформація про які зосереджена в одному масиві даних. Характеристики або параметри об'єктів і суб'єктів (тобто фактори), значення яких складають аналізований масив даних і через які проявляються ці взаємозв'язки, можуть бути різними. Кореляції дають змогу оцінити з якою ймовірністю зміна одного чинника конфлікту призведе до зміни інших чинників цього конфлікту. Виявлення таких кореляцій є особливо корисним на етапі оперативного планування операцій. Як правило, ці фактори тісно взаємопов'язані, на один із факторів можна ефективно впливати з метою досягнення позитивних змін у всьому середовищі конфлікту гібридного типу. Однак слід ураховувати, що встановити ці кореляції у разі малих обсягів масиву даних неможливо.

Для характеристики великих даних використовують характеристики [9]: обсяг (*volume*), швидкість накопичення нових даних та їх обробки (*velocity*), різноманіття типів даних, які можуть оброблятися (*variety*), достовірність (*veracity*), цінність (*value*), змінний характер (*variability*), актуальність (*volatility*) вразливість (*vulnerability*), обґрунтованість (*validity*), візуалізація (*visualization*).

Великі дані розділяють на дві категорії: структуровані та неструктуровані. *Структуровані дані* упорядковані певним чином і мають чітку структуру і, як правило, зберігаються у вигляді таблиць. Деякі дані можуть не вписуватися в жорстку структуру таблиці та можуть бути втрачені. Для виключення втрат даних використовуються реляційні бази даних – набір даних із зумовленими зв'язками між ними. Такі структуровані дані організовані як таблиці, у яких є лише частина даних, які зв'язані з даними інших таблиць, а спільний аналіз дає змогу отримувати інформацію з декількох таблиць одночасно.

*Неструктуровані дані* різноманітні (документи, зображення, відео), надходять із різних джерел (системи спостереження, дрони, супутникові дані тощо) та різних мереж. З такими даними складніше працювати, оскільки для отримання інформації з них необхідно підбирати

відповідні інструменти, методи зберігання та обробки.

Крім того, для аналізу поточних подій використовуються дані із соціальних мереж, пошукові запити користувачів Інтернету, корпоративні бази даних, фінансові системи, мобільні застосунки, пристрої аудіо- та відеореєстрацій, дані геолокації об'єктів конфлікту, отримані через мережі мобільного зв'язку, метеорологічна статистика та інші. Як приклад, такі дані були використані Bellingcat під час дослідження питання знищення літака рейсу MH-17 на сході України.

Різноманітність великих даних зумовлює специфічні методи їх аналізу. Для аналізу великих даних застосовується сукупність різних методів. Насамперед, це методи математики, статистики, методи інтелектуального аналізу та штучного інтелекту [3], розпізнавання образів, імітаційного моделювання, методи OSTIN [12] та інші. Саме вони дають змогу отримувати необхідну інформацію з великих даних, визначати наявність або відсутність зв'язків між суб'єктами і об'єктами конфлікту гібридного типу, визначати причини та наслідки різних подій, перевіряти версії тощо.

Пропонується послідовність з трьох етапів обробки великих даних для прогнозування конфліктів гібридного типу (рис. 1).

Етап *генерації даних* включає: збір з різних джерел, передавання, збереження, попередню обробку (інтеграція, очищення, оптимізація).

*Збір даних* здійснюється з різних джерел, включаючи дані візуальної розвідки, відеоспостереження, дронів, файлів журналів, датчиків, вебкамер, засобів радіотехнічної розвідки. Після того, як дані зібрані, вони *передаються* в інфраструктуру зберігання та обробки даних для подальшої обробки та аналізу. Це може бути виконано у два етапи: передача між динамічною мережею (DCN) та передача всередині DCN. Передача між DCN пов'язана з передачею даних від джерела даних до центру обробки даних (дата-центру), у той час як останній забезпечує передавання всередині мережі дата-центру. Дата-центр забезпечує *збереження, структурування та управління даними*. Розподілена система зберігання великих даних забезпечує надійний простір *зберігання і доступ* до великих даних і оцінюється такими параметрами, як узгодженість (*C*), доступність (*A*) та стійкість (*P*).

*Попередня обробка* включає *інтеграцію,*

очищення та оптимізацію даних. Дані, які збираються з різних джерел, можуть бути надмірними, зашумленими та непослідовними. Дані обробляються для підвищення їх якості до рівня, який потребує проведення аналізу. Ця процедура необхідна для підвищення точності аналізу та

скорочення витрат на зберігання.

Інтеграція призначена для об'єднання даних з різних джерел і забезпечення єдиного та одноманітного їх представлення. Для динамічної інтеграції дані можуть бути агреговані з різних джерел за допомогою віртуальної бази.

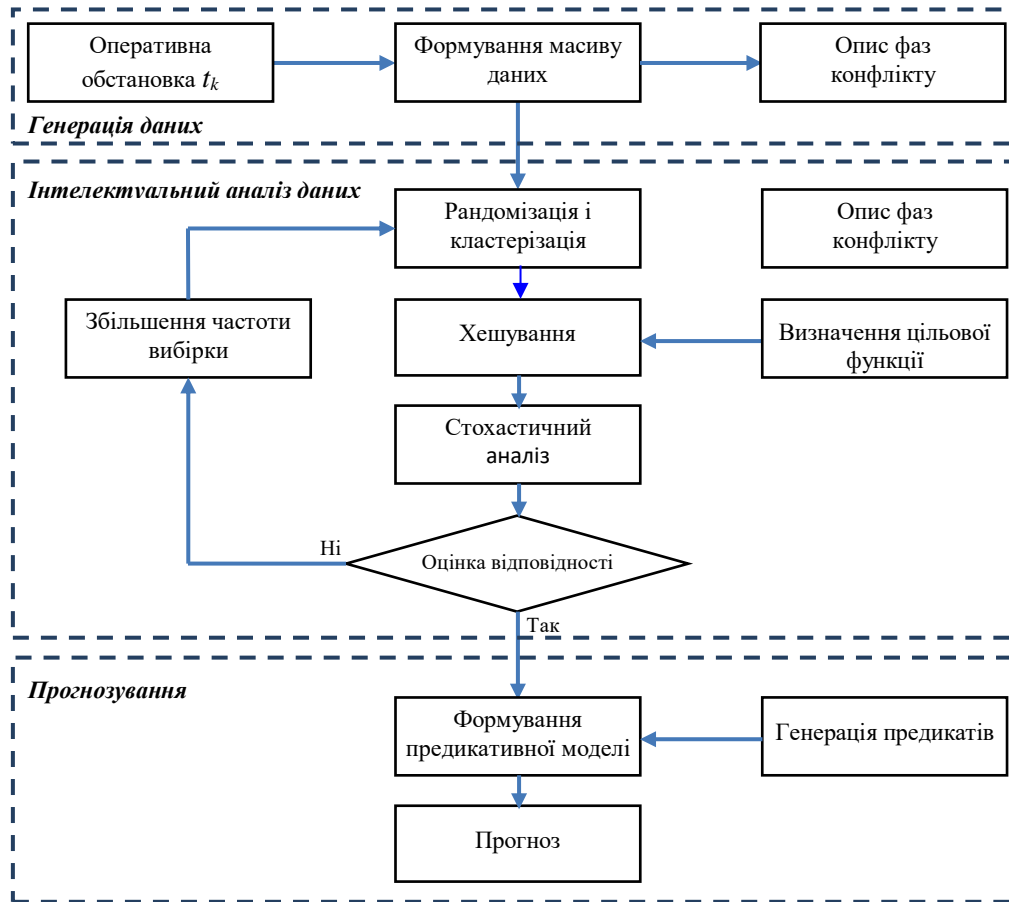


Рис. 1. Загальна послідовність процесу прогнозування конфліктів гібридного типу на основі великих даних

Процедура очищення даних включає перевірку на точність, повноту та послідовність. Під час цього процесу дані можуть бути видалені та змінені з метою їх покращення. Загальний процес очищення включає: визначення типів, ідентифікація, документація, виправлення помилок та зміна процедур окремих джерел і процедур генерації даних з метою уникнення помилок у майбутньому.

Процедура оптимізації направлена на виявлення наборів надлишкових даних або їх повторення. Це збільшує вартість зберігання, призводить до неузгодженості даних та впливає на якість даних. Для подолання цього, використовуються різні методи скорочення даних, такі як фільтрація та стиснення даних. Обмеження цих методів скорочення даних у тому, що вони збільшують обчислювальні витрати.

Інтелектуальний аналіз даних (Data

Mining) включає застосування класичних та сучасних процедур і алгоритмів для перетворення різнорідних неструктурованих даних на необхідну для прогнозування інформацію, тобто отримати нову, приховану та корисну інформацію з набору великих даних.

До таких процедур відносяться: рандомізація; хешування (згортання), включно з локально чутливим хешуванням для виявлення дублікатів даних; кластеризація багатовимірних масивів даних (BFR, k-means, CURE та інші); пошук за подібністю; стохастичний аналіз; аналіз конфліктного кошика; масштабований пошук асоціативних правил (A-Priori); оцінка відповідності.

У процедурах аналізу даних можуть використовуватися розподілені файлові системи та Map-Reduce, як інструмент для створення паралельних алгоритмів, ефективних для обробки великих обсягів

даних; теорія локально-чутливих (LSH) функцій: локально-чутливі сімейства для відстаней по Жаккару, по Хеммінгу, для евклідової відстані та відстані по косинусу.

Для аналізу великих даних доступні різні інструменти, включаючи програмне забезпечення з відкритим вихідним кодом та комерційне програмне забезпечення.

*Етап прогнозування включає:*

визначення набору показників для характеристики конкретного прогнозу;

вибір архітектури системи прогнозування, з урахуванням своєчасності та об'єктивності отримання прогнозів та організації заходів протидії, можливо використання існуючої архітектури – обробка великих даних на основі пам'яті та системи паралельної обробки;

кластеризація вхідних даних – угрупування даних щодо об'єктів і суб'єктів конфлікту на основі визначених характеристик;

вибір методу аналізу великих даних, в окремих випадках можливе застосування традиційних методів, таких як кластерний та регресійний аналіз, інтелектуальні алгоритми;

візуалізація даних для перевірки деталей у різних масштабах, статистичний аналіз та визначення закономірностей змін даних з часом;

формування прогнозу на основі аналізу та візуалізованих результатів, встановлення зв'язків між об'єктами, прийняття обґрунтованого рішення та планування необхідних заходів протидії.

Водочас, прогнозування на основі великих даних, незважаючи на широке застосування технологій та алгоритмів їх аналізу в системах оборонного і оперативного планування низки країн, має низку проблемних питань, які вимагають вирішення:

1. Масштабованість та зберігання: швидкість збільшення обсягу даних набагато вища, ніж у існуючих систем обробки, системи зберігання не можуть зберігати ці дані, існує необхідність розроблення системи обробки даних, яка буде задовольняти не тільки сучасні, а і майбутні потреби.

2. Своєчасність аналізу: цінність даних з часом знижується, більшість програм для виявлення загроз низького рівня, супроводження просування диверсійно-розвідувальних груп, потребують проведення транзакцій у реальному або близькому до реального часу.

3. Оброблення неоднорідних даних: дані, отримані з різних джерел, неоднорідні за

своєю природою, зображення, відео та дані соціальних мереж не можливо зберігати та обробляти за допомогою традиційних інструментів, таких як SQL, необхідні нові методи та алгоритми для представлення зображень.

4. Аналітика даних: традиційні СУБД підходять лише для структурованих даних, нереляційні бази даних використовуються для обробки неструктурованих даних, необхідне вирішення проблеми продуктивності.

5. Підготовка фахівців з питань обробки даних: зі збільшенням кількості генерованих (структурованих та неструктурованих) даних виникає потреба у фахівцях з аналітичними навичками у сфері обробки та аналізу великих даних, до 2024 року може знадобитися від 250 000 до 500 000 додаткових фахівців у галузі великих даних [8].

6. Конфіденційність та безпека: як правило, аналітики великих даних мають технічну інфраструктуру для доступу до будь-якого джерела даних, включаючи мережі державних установ та стратегічних об'єктів, необхідні нові підходи до забезпечення безпеки даних, з врахуванням конфіденційності та результативності.

7. Скорочення даних – один із поширених способів вписатися у математичну модель, дані, що знаходяться поза контекстом, втрачають цінність, водночас аналіз надлишкових даних може показати додаткові кореляції.

8. Цифровий розрив: отримання доступу до великих даних – одне з найважливіших обмежень, супротивник, який проводить моніторинг соціальних мереж, має також доступ до великих соціальних даних країни, що веде оборону.

9. Помилки даних: великі дані схильні до помилок і втрат, для вирішення проблеми необхідні характеристики джерела даних, властивості та обмеження набору даних, щоб уникнути системних помилок під час інтерпретації даних.

### **Висновки**

1. З появою конфліктів гібридного типу обсяг даних, необхідних для прогнозування їх розвитку, різко виріс.

2. Жодна технологія не дозволяє отримати прогноз розвитку конфлікту гібридного типу з повною достовірністю, однак, враховуючи переваги та недоліки мережецентричних або хмарних технологій, результативним може виявитися застосування технологій обробки великих обсягів даних для отримання достовірних прогнозів.

3. Крім визначення великих даних, необхідно розробляти методики застосування великих даних для прийняття рішень та прогнозування.

4. Аналітика великих даних має широке застосування у різних сферах гібридного конфлікту, включаючи військовий. Застосування технологій великих даних допомагає в аналізі та побудові сценаріїв розвитку конфлікту. Водночас, незважаючи на переваги, у аналітиці великих даних є свої обмеження та проблеми щодо застосування.

Таким чином, великі дані відкривають нові можливості для прогнозування сучасних конфліктів гібридного типу.

#### СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Semenenko V. M., Ivashchenko A. M. Military Aspects of Countering Hybrid Warfare: Experiences, Lessons, Best Practices. *NATO Science and Technologies Organization*. Paris, 2020. 188 p.
2. Семененко В. М., Іващенко А. М. Особливості воєнних аспектів сценаріїв протидії гібридній війні. *Збірник наукових праць Центру воєнно-стратегічних досліджень Національного університету оборони України імені Івана Черняхівського*. Київ, 2019. № 1 (65). С. 19–24.
3. Ivashchenko A., Chornodid I. The Business Assistant Service as One of the Promising Areas for the Adoption of AI Technologies in the Enterprise. *Business: Theory and Practice*. Vol. 21 Issue. 2. P. 588–597. (Scopus ID).
4. Кейт О. Зброя математичного знищення. Київ : Book Chef, 2020. 336 с.
5. Berman E., Felter J., Shapiro J. Small Wars, Big Data: The Information Revolution in Modern Conflict. Princeton NJ, Princeton University Press. 2018. 408 p.
6. Zorri D. M., Derezhovski M. Big Data Conflict Forecasting: Operationalizing the Data Science Team. Occasional paper. Joint Special Operating University, Department of Strategic Studies, June 2021. 32 p.
7. Tunnell H. D. Tactical Data Science. *Military Review*. Kansas, Fort Leavenworth, Jul/Aug 2020. Vol. 100, Iss. 4. P. 123–137.
8. Morabito D. National Security and the Third-Road Threat: Toward a Comprehensive Theory of Information Warfare. *Air&Space Power Journal*. 2021. Vol. 35, Iss. 3. P. 19–39.
9. The Biggest Data Management News Items During the First Half of 2021. Melbourne, New Bites, Jun 27, 2021. 127 p.
10. Метров О. Концепции применения вооруженных сил США в многосферных операциях. *Зарубежное военное обозрение*. 2019. № 10. С. 3–8.
11. Defence Primer: Army Multi-Domain Operations (MDO). *Congressional Research Service*. October 22, 2021. 3 p.
12. McHabb J.B.A Military History of the Modern Middle East. Santa Barbara, CA, Denver, COL, Praeger, 2019. 456 p.

Стаття надійшла до редакційної колегії 09.12.2021

### Predicting the development of hybrid conflicts based on Big Data

#### Annotation

Globalization and the information technology revolution have become integrators of classical and new forms, methods and technologies of modern conflicts. Hybrid conflicts, which achieve political goals and are characterized by an increase in the number of objects involving different combinations, the simultaneous conduct of different conflicts and the use of military and non-military means, have become widespread. Such conflicts are characterized by petabytes ( $10^{15}$  bytes) of data generated during the conflict.

Unlike the conflicts of the last century, forecasts of the development of hybrid conflicts based on causation provide a level of probability of their implementation not exceeding 0.8. The search for qualitatively new approaches to forecasting the development of hybrid conflicts is relevant.

*The purpose of the article* is to substantiate the possibility of using big data technologies for analysis and forecasting of hybrid conflict development and to determine the sequence of their processing and analysis.

We offer a sequence of three stages of big data processing for predicting hybrid conflicts:

data generation includes: data collection from various sources, transmission, storage, pre-processing (integration, cleaning, optimization) of data;

data mining (Data Mining) includes the use of classical and modern procedures and algorithms to convert heterogeneous unstructured data into information needed for forecasting;

formation of the forecast on the basis of the analysis and visualized results, establishment of communications between objects, acceptance of the reasonable decision and planning of necessary measures of counteraction.

**Keywords:** hybrid conflicts; big data; big data processing technologies, big data analysis; prognostication.