

ВІДГУК

офіційного опонента про дисертаційну роботу

Аксак Наталії Георгіївни

«МЕТОДИ ТА МОДЕЛІ РОЗПОДІЛЕНОЇ ІНТЕЛЕКТУАЛЬНОЇ ОБРОБКИ
ВЕЛИКИХ ДАНИХ У СПЕЦІАЛІЗОВАНИХ КОМП'ЮТЕРНИХ СИСТЕМАХ»,

подану на здобуття наукового ступеня доктора технічних наук
за спеціальністю 05.13.05 – комп'ютерні системи та компоненти

Актуальність теми виконаних досліджень

Швидкий розвиток вбудованих інтелектуальних пристроїв і комп'ютерних мереж породив багато мережевих додатків і послуг. У зв'язку з цим виникає питання як задовольнити всі вимоги, пов'язані зі швидким зростанням нових мережевих додатків і послуг, використовуючи централізовану парадигму обчислень. Це означає, що існують надзвичайно важливі завдання, пов'язані зі спілкуванням і прийняттям рішень, в яких при взаємодії людини з різними комп'ютерними системами очікується оперативна реакція на виникаючі перетворення в реальному часі. Ці завдання безпосередньо пов'язані з методами обробки постійно зростаючого обсягу даних і обчислювальними пристроями, на яких будуть вирішуватися такі задачі.

Проблема розподіленої інтелектуальної обробки великих даних в тому, що більша їх частина представлена в форматі, яка погано відповідає традиційному структурованому формату. Дані самих різних типів, як структуровані, так і неструктуровані зберігаються в безлічі різноманітних сховищ, іноді навіть за межами організації. Корпорації можуть мати доступ до величезного обсягу своїх даних і не мати необхідних інструментів, щоб встановити взаємозв'язок між цими даними і зробити на їх основі значущі висновки. Необхідно врахувати, що дані оновлюються все частіше, і виникає ситуація, в якій традиційні методи та моделі інтелектуальної обробки не



можуть наздогнати величезні обсяги постійно оновлюваних даних. Тому розробка методів та моделей розподіленої інтелектуальної обробки великих даних в спеціалізованих комп'ютерних системах (СКС) є актуальною. Також недоліком сучасних технологій обробки великих даних є відсутність єдиної «інформаційної інфраструктури» корпоративних центрів обробки даних як структурованої, так і неструктурованою інформації та відсутність єдиного підходу до програмно-апаратних рішень і супутніх сервісів, спрямованих на виявлення прихованих інформаційних закономірностей в великих даних.

Основні особливості створення СКС полягають у такому: по-перше, це випробування нових методів обробки інформації (розпаралелювання, децентралізація обчислень, використання функціональних вирішувачів, багатовимірних числових систем, тощо); по-друге, процес створення СКС з оптимальними функціональними якостями має характер багаторівневої процедури, яка комбінує формальні та змістовні методи. Існують два теоретичних напрямки в організації структур СКС. У першому випадку особливості виконуваних задач реалізуються на стандартних процесорах шляхом спеціалізації програмного забезпечення. У другому – використовуються процесори, що орієнтовані на виконання спеціалізованих функцій апаратними засобами.

Отже, авторка формулює відповідні цілі і завдання дослідження, що пов'язані з розробкою теоретичних і практичних питань відповідної інноваційної методології побудови СКС розподіленої інтелектуальної обробки великих даних.

Зважаючи на це, вважаю тему дисертаційної роботи Аксак Наталії Георгіївни, яка направлена на розвиток методології та розробку моделей й методів інтелектуальної обробки великих даних у СКС, актуальною.

Відзначаючи актуальність теми дисертації, треба також підкреслити її відповідність тематичним планам держбюджетних науково-дослідних робіт Харківського національного університету радіоелектроніки, в яких авторка брала участь.

Структура і склад дисертаційної роботи

Дисертація складається з анотації, вступу, шістьох розділів, висновків, списку використаних джерел та додатків.

У **вступі** наведені основні характеристики роботи, такі як актуальність, мета та задачі дослідження, визначені наукова новизна, відомості про апробацію і впровадження результатів та аналіз публікацій за темою роботи з особистим внеском автора.

1-й розділ присвячено аналізу предметної області, розглянуто питання інтелектуальної обробки великих даних у СКС. Наведено огляд мультиагентних технологій, технологій паралельних та розподілених обчислень, аналіз існуючих методів нейромережевої обробки різнорідних даних, методів візуалізації багатовимірних даних та обробки зображень досліджуваного об'єкта, формалізовано процес розподіленої інтелектуальної обробки великих даних у СКС, сформульована мета та задачі дослідження.

У **2-му розділі** сформульовані основні принципи, на яких базуються методологічні основи побудови персоніфікованого сервіс-орієнтованого середовища, які надають бізнес послуги. До найбільш значущих з них належить послідовність взаємопов'язаних дій: аналіз поведінки користувача за його серфінгом в Інтернеті; виявлення причинно-наслідкових закономірностей, що зумовило можливість вироблення гіпотез про наявність чи відсутність певних властивостей користувача; побудова інформаційного портрету користувача; визначення важливості відвідуваних користувачем веб-сторінок; прискорена кластеризація користувачів сервіс-орієнтованого середовища, що зумовило адаптувати веб-ресурс залежно від індивідуальних переваг користувача.

3-й розділ пропонує загальну структуру мультиагентної системи (МАС) з використанням організаційної концепції «агент/група/роль», метод кооперації агентів сервіс-орієнтованої веб-системи, метод розподілу робіт агентів у МАС, метод узгодження дій агентів та метод взаємодії агентів для оперативного реагування на екстрені обставини. Також у розділі наведено

опис Cloud-Fog-Dew архітектури для персоналізованих сервіс-орієнтованих систем, який дозволяє використовувати результати віддаленого моніторингу при прийнятті рішень для своєчасного реагування при настанні невідкладної ситуації. Отримані у третьому розділі результати дозволяють своєчасно надавати спеціалізовані послуги

4-й розділ містить розроблений комплекс методів нейромережевої обробки великих даних в обчислювальному середовищі з різними топологіями. Обгрунтовано вибір обчислювального розподіленого середовища залежно від обсягу даних – від систем із загальної пам'яттю й технологією програмування OpenMP, якщо обсяг даних невеликий, до кластерних систем з індивідуальною пам'яттю й моделлю програмування MapReduce, якщо обсяг даних великомасштабний.

Отримані у даному розділі результати можуть бути використані при вирішенні неформалізованих завдань великого розміру нейронними мережами, що дозволяє скоротити обсяги переданої інформації між вузлами, збільшити швидкодію в ході вирішення завдань, а також ефективно балансувати навантаженням обчислювальних ресурсів з різними топологіями передачі даних.

5-й розділ розкриває етапи аналізу багатовимірних даних шляхом відображення у простір малого розміру (лінія/площина) з метою запобігання їхнього відображення в одну точку. Показано доцільність використання параметричних сплайнів і сплайн-поверхонь для побудови відповідно одновимірних і двовимірних кусково-гладких мап Кохонена (КГМК). Пропонується сукупність критеріїв для визначення якості візуалізації даних мапами Кохонена, що відображаються в одну точку на мапі.

6-й розділ присвячено розподіленій інтелектуальній обробці інформації про досліджуваний об'єкт на високопродуктивних обчислювальних комплексах, що дозволило значно скоротити час обробки даних.

На основі запропонованих моделей інтелектуальної обробки великих

даних розроблено програмні модулі (розпізнавання емоцій, жестів руки та голосової верифікації особистості), які використовуються в сервіс-орієнтованій комп'ютерній системі для зменшення часу відклику.

У додатках наведено допоміжні таблиці, акти про впровадження результатів дисертаційної роботи у держбюджетні роботи та навчальний процес, приклади застосування розроблених методів і моделей для вирішення практичних завдань, список публікацій здобувача.

Ступінь обґрунтованості та достовірності нових наукових положень, отриманих в дисертаційній роботі

Достовірність роботи забезпечується тим, що вона ґрунтується на практичних і теоретичних викладках, які в роботі набувають потрібного розвинення згідно зі специфікою мети та задач досліджень, а також узгоджені із експериментальними даними, отриманими автором. Аналітичні оцінки, які отримані автором, підкріплені відповідними викладками.

Результати роботи достатньо апробовано на міжнародних наукових та науково-практичних конференціях та мають підтверджені впровадження.

Оформлення дисертаційної роботи

Структура та обсяг дисертаційної роботи відповідають вимогам щодо тексту дисертацій на здобуття наукового ступеня доктора технічних наук. Дисертацію написано грамотною мовою, з використанням сучасної бібліографії та наукової термінології. Зміст та результати досліджень викладено лаконічно та аргументовано. Задачі дисертації в логічній послідовності вирішуються на усіх етапах дослідження – від аналізу проблеми до побудови моделей та методів розв'язання задач, з наступним проведенням експериментів з конкретними чисельними даними та впровадженням результатів дослідження. Зміст автореферату відповідає змісту дисертаційної роботи.

Практичне та наукове значення результатів дисертаційної роботи

Практична значимість отриманих теоретичних результатів дисертаційної роботи Аксак Н.Г. підтверджена підвищенням продуктивності спеціалізованих комп'ютерних систем за рахунок зменшення часу виконання завдань та ефективного використання ресурсів.

На основі запропонованих методів та моделей розроблено експериментальну систему для розв'язання задачі діагностики та віддаленого моніторингу. Експериментальні дослідження доводять ефективність запропонованих рішень.

Результати, отримані автором, слід вважати важливими для подальшого розвитку проблемно-орієнтованих систем, що надають спеціалізовані послуги. Крім того, деякі з цих результатів можуть бути використані у навчальному процесі для підготовки фахівців в галузі створення СКС та використання інтелектуальної обробки великих даних.

Публікації за темою дисертації

Основні результати дисертації достатньо повно викладені в 41 статті, що опубліковані у наукових фахових виданнях. Зміст наведених публікацій не містить повтору. Вони адекватно відображають зміст та результати дисертації. Згідно з авторефератом кандидатської роботи здобувача, в роботі не використані наукові матеріали цієї роботи.

Наукова новизна

Вперше запропоновано методологічну основу побудови спеціалізованих комп'ютерних систем розподіленої інтелектуальної обробки великих даних, яка характеризується тим, що містить такі функціональні складові: метод збору інформації про користувачів веб-сервісів; метод адаптації предметно-орієнтованих веб-сервісів; методи прискореної обробки великих даних на основі високопродуктивних обчислень за допомогою нейронних мереж; спосіб своєчасного надання спеціалізованих послуг за

допомогою мультиагентних систем, що забезпечує високорівневий доступ предметних користувачів до проблемно-орієнтованих програмних комплексів та дозволяє погоджувати їхню взаємодію для адекватного і оперативного реагування на зміни, що відбуваються в реальному часі.

Вперше запропоновано принцип проектування проблемно-орієнтованих веб-сервісів, який характеризується налаштуванням такої сукупності моделей взаємопов'язаних компонентів СКС розподіленої інтелектуальної обробки великих даних: узагальненої моделі організації проблемно-орієнтованих обчислень; загальної структури мультиагентної системи для оперативного реагування в екстрених ситуаціях з використанням організаційних концепцій агент-група-роль; моделі категоризації користувачів веб-служб на основі методу автоматичного породження гіпотез; моделі процесу персоналізації веб-ресурсу, в якій поєднуються агентські і нейромережеві технології; об'єктної моделі документів проблемно-орієнтованого веб-сервісу у вигляді графа, що в цілому дозволяє вирішувати комплексну задачу адаптації та персоналізації веб-порталу спільно з наданням спеціалізованих послуг.

Вперше запропоновано принцип побудови адаптивної веб-служби та метод персоналізації Інтернет-сервісу, які характеризуються сукупністю взаємопов'язаних дій: автоматичне вироблення гіпотез, що дає можливість визначити наявність або відсутність цільових властивостей користувача; аналіз поведінки користувача за його серфінгом в Інтернеті, що дозволяє видавати більш релевантні результати; побудова інформаційного портрета для збору статистично значущої сукупності інформаційних характеристик з метою планування подальших дій; паралельна кластеризація користувачів з використанням мап Кохонена з метою прискорення обробки великих даних, що в цілому дозволяє налаштовувати контент та надавати актуальну інформацію з різних аспектів пропонованого сервісу.

Вперше запропоновано моделі MapReduce Hadoop для розподіленої нейромережевої обробки великих даних, які характеризуються адаптацією

топології передачі даних до відповідної архітектури нейронної мережі на обчислювальний кластер, що дозволяє скоротити обсяги переданої інформації між вузлами, підвищити швидкодію при вирішенні складних завдань, а також ефективно балансувати навантаженням обчислювальних ресурсів з різними топологіями передачі даних.

Вперше запропоновано Cloud-Fog-Dew архітектуру для персоналізованих сервіс-орієнтованих систем, яка характеризується поєднанням розподілених методів і засобів збору, зберігання, обробки та аналізу великих даних, що дозволяє використовувати результати віддаленого моніторингу при прийнятті рішень для своєчасного реагування з настанням екстреної ситуації, навіть при збоях віддаленого виклику або підключення до Інтернету.

Удосконалено методологію обробки великих даних нейронними мережами, яка відрізняється наявністю динамічного перерозподілу робіт між обчислювачами, що дозволяє рівномірно планувати навантаження обчислювального кластера з різними топологіями передачі даних.

Удосконалено методи відображення багатовимірних даних у простір малого розміру, які відрізняються запобіганням відображення багатовимірних даних в одну точку, що дозволяє усунути їх нерозрізненість у просторі малої вимірності.

Удосконалено методологію розподіленої інтелектуальної обробки зображень досліджуваного об'єкта, яка відрізняється можливістю online аналізу його поведінки, що забезпечує підвищення якості надання сервісів і зменшення часу відклику.

Набув подальшого розвитку метод узгодження дій агентів шляхом використання засобів спілкування взаємодіючих сторін та стану системи у вигляді стійкості та цілісності, що дозволяє успішно функціонувати агентно-орієнтованій системі для оперативного реагування на критичні зміни обставин.

Набули подальшого розвитку методи розподілу робіт і координації групових дій агентів в агентно-орієнтованій системі шляхом динамічного розподілу ролей між агентами, що дозволяє для кожного зареєстрованого користувача розподілити поставлені завдання між агентами для своєчасного реагування при виникненні інциденту та підвищити ефективність спільних дій всіх учасників.

Зауваження

1. В процесі аналізу предметної області не достатньо розглянуті сучасні методи створення та функціонування спеціалізованих комп'ютерних систем.

2. В рукопису дисертації було б докладніше відзначити зв'язок методів візуалізації багатовимірних даних (розділ 5) з послугами, які надає проблемно-орієнтована система..

3. При розробці методів нейромережевої обробки зображень (розділи 3, 6) передбачається використання бази даних для зберігання результатів обробленої інформації про користувача, але за текстом роботи не вказано, яка саме база даних була використана при практичній реалізації програмного забезпечення.

4. Не зрозуміло, як метод розподілу робіт (розділ 3, вираз (3.4)) між агентами агентно-орієнтованій системі впливає на скорочення часу відклику системи.

5. У процесі придбання знань (розділ 3) доцільним було б докладніше описати правила виведення.

6. Запропоноване рішення про надання спеціалізованих послуг (розділ 3) не враховує кількість зареєстрованих користувачів веб-сервісу.

7. При розробці інтелектуальних методів обробки даних достатньо велика увага приділялася нейронним мережам (розділи 4, 6). Доцільно було б використовувати також інші інтелектуальні методи.

Однак ці зауваження суттєво не впливають на загальну позитивну характеристику дисертації, що має визначені вище актуальність, наукову новизну і практичну значущість.

Висновок

Дисертаційна робота Аксак Наталії Георгіївни є завершеним науковим дослідженням, у ході якого вирішена актуальна науково-прикладна проблема створення спеціалізованих комп'ютерних систем для проблемно-орієнтованих обчислень. Результати отримані автором особисто, не перетинаються з результатами, отриманими у кандидатській роботі.

Вважаю, що за актуальністю вибраної теми, обсягом і рівнем виконаних теоретичних і експериментальних досліджень, достовірністю і обґрунтованістю висновків, новизною досліджень, значенням отриманих результатів для науки і практики дисертаційна робота задовольняє вимогам п. 9, 10, 12, 13,14 «Порядку присудження наукових ступенів», затвердженого Постановою КМУ № 567 від 24 липня 2013 р. Зміст роботи відповідає спеціальності 05.13.05 - комп'ютерні системи та компоненти, а її авторка заслуговує присвоєння наукового ступеня доктора технічних наук.

Професор кафедри автоматизації та комп'ютерно-інтегрованих технологій Харківського національного технічного університету сільського господарства ім. Петра Василенка доктор технічних наук, професор



I.O. Фурман

Підпис доктора технічних наук, професора Фурмана Іллі Олександровича. засвідчую

*керівник відділу
Я.Ловоросів ХНТУСГ*



«06» червень 2019 р.

Н.В. Валуйко