

ДОСЛІДНИЦЬКА ПРОПОЗИЦІЯ З ОБРАНОЇ СПЕЦІАЛЬНОСТІ

вступника до аспірантури Яременка Вадима Сергійовича

Тема: **Моделі машинного навчання для розв'язання
задач інтелектуальної обробки даних в
мультиагентних системах**

Спеціальність: (122) Комп'ютерні науки

1) Обґрунтування запропонованої теми дисертації

Надвеликим масивом даних вважається такий обсяг даних, який не може бути цілком розміщений в оперативній пам'яті комп'ютера без застосування спеціальних методів, що створює проблеми аналізу таких даних відомими методами інтелектуального аналізу даних.

Описана проблема ускладнюється тим, що найчастіше такі дані надходять з різних джерел даних в розподіленій комп'ютерній системі зі швидкістю, яка може перевищувати швидкість обробки та зберігання цих даних в пам'яті комп'ютера, що може стати причиною втрати деяких даних.

Сучасні технології обробки надвеликих масивів даних спираються на використанні розподілених файлових систем, які управляють процесами кластерних обчислень, та спеціальних технологій розподілу масивів даних на пов'язані між собою групи даних меншої розмірності і їх паралельного аналізу. До цих методів належить, наприклад, технологія MapReduce.

Головною тезою досліджень, які пропонується виконати в рамках даної теми, є те, що на підставі використання технологій, подібних до MapReduce, а також завдяки застосуванню методів машинного навчання можна налаштувати мультиагентну систему на розв'язання задач інтелектуального аналізу надвеликих масивів даних з високою продуктивністю, яка буде перевищувати продуктивність досі створених систем цього типу.

2) Виклад дослідних запитань; актуальність теми

Передбачається виконання теоретичних та експериментальних досліджень методів машинного навчання та розвиток цих методів з метою їх використання в мультиагентних системах, призначених для інтелектуального аналізу надвеликих масивів даних.

Серед дослідницьких запитань важливо зазначити наступні:

- створення методів для обробки надвеликих масивів даних для розв'язку задач інтелектуального аналізу;
- налаштування мультиагентної системи з використанням розроблених методів для ефективного її застосування при вирішенні задач інтелектуального аналізу даних.

Також слід зазначити, що ця тема пов'язана з науковою тематикою кафедри Системного проектування, бо є частиною наукової тематики з застосування сервіс-орієнтованих обчислень і архітектур в медицині.

3) Мета і завдання дослідження

Метою даної роботи є дослідження технологій MapReduce та подібних до неї (наприклад, Spark) для їх використання у вирішенні задач обробки великих масивів даних та створення моделей, що базуються на цих технологіях для розв'язку задач інтелектуального аналізу даних у мультиагентних системах. Результатом проведених досліджень передбачається налаштована для розв'язку задач інтелектуального аналізу мультиагентна система, продуктивність роботи якої є значно вищою за сучасні засоби обробки великих даних.

4) Аргументований виклад потенційної наукової новизни результатів дослідження

Під час попередніх досліджень були запропоновані моделі обробки великих масивів даних, в основі яких – технологія MapReduce. Одна з яких дозволяє вирішувати задачу кластеризації, а друга модель – задачу пошуку частих предметних наборів. В основі моделі для вирішення задачі кластеризації, знаходиться модифікований алгоритм CURE, ця модифікація

дозволяє обробляти дані паралельно. Також перевагою даної моделі є те, що використовується репрезентативна вибірка кластеру і не зберігаються усі його точки, що дозволяє зменшити завантаженість пам'яті комп'ютера. Іншою перевагою є те, що не потрібно завантажувати всю вхідну вибірку даних у оперативну пам'ять. В основі другої моделі, для вирішення задачі пошуку частих предметних наборів, знаходиться модифікований алгоритм SON. Аналогічно до попереднього пункту, алгоритм модифіковано для того, щоб була можливість обробляти дані паралельно. У цьому випадку за два проходи по даним знаходяться часті набори, що задовільняють заданим користувачем параметрам.

Для тестування запропонованих моделей було налаштовано фреймворк Hadoop MapReduce. На виборках даних обсягом від 10 МБ до 6 ГБ було проведено аналіз швидкодії створених моделей у порівнянні із моделями, в основі яких немає технології MapReduce. На основі цих результатів був зроблений висновок, що запропоновані моделі дають правильний результат, а також – що використання даних моделей дозволяє пришвидшити обробку даних. Особливо це помітно для даних об'ємом більше за 500 МБ.

В якості апробації результати досліджень були оприлюднені на 19-й Міжнародній науково-технічній конференції SAIT 2017, а також – у міжнародному науковому журналі «Інтернаука», випуск №6 2017 року.

Відповідно до вищезазначених аргументів можна прийти до висновку, що використання технологій обробки великих даних на прикладі MapReduce є ефективним при їх застосуванні у розв'язку задач інтелектуального аналізу, тому є доцільним продовження досліджень у даному напрямку. Передбачається виконання теоретичних та експериментальних досліджень методів машинного навчання та розвиток цих методів з метою їх використання в мультиагентних системах, призначених для інтелектуального аналізу надвеликих масивів даних. Передбачається створення експериментальної мультиагентної системи, в якій мають бути впроваджені розроблені методи та доведена їх ефективність.

5) Практичне значення очікуваних наукових результатів

За останні роки обсяг збережених даних став занадто великим для того, щоб була можлива їх обробка традиційними алгоритмами. Відповідно, кількість даних зростає і паралельно із цим розвиваються алгоритми для їх швидкої обробки з використанням сучасного технічного забезпечення.

Тому практичним значенням очікуваних наукових результатів є створення експериментальної мультиагентної системи для розв'язку задач інтелектуального аналізу з високою продуктивністю, яка буде перевищувати продуктивність досі створених систем цього типу, а також – створення рекомендацій по налаштуванню та впровадженню цієї системи у сучасні сфери людської діяльності.

Вступник

Яременко В. С.

Науковий керівник

проф. Рогоза В.С.

Завідувач кафедри

проф. Петренко А.І.
