

ДОДАТОК А

Графічний матеріал кваліфікаційної роботи

ХАРКІВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
РАДІОЕЛЕКТРОНІКИ

КВАЛІФІКАЦІЙНА РОБОТА

**Моделі, методи та засоби
побудови паралельних алгоритмів**

ст. гр. СПзм-22-1 Черниш А.В.

Керівник: доц. каф. ЕОМ Федченко В.М.

Мета та завдання роботи:

2

Метою кваліфікаційної роботи є розробка моделей і методів побудови паралельних алгоритмів аналізу даних з урахуванням умов їх виконання.

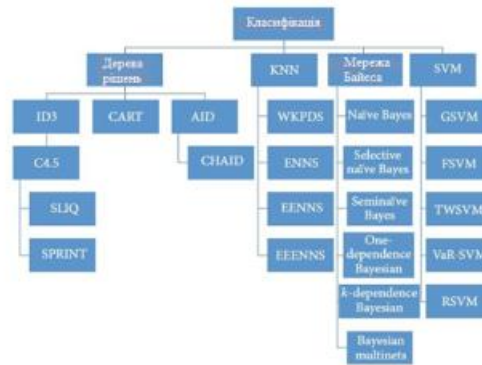
Об'єкт дослідження: паралельні алгоритми аналізу даних.

Завдання:

- аналіз існуючих моделей, методів і засобів побудови паралельних алгоритмів, в тому числі алгоритмів аналізу даних;
- створення моделі паралельного алгоритму аналізу даних як для розподіленої, так і для загальної пам'яті;
- програмна реалізація запропонованих моделей і методів побудови паралельних алгоритмів аналізу даних.

Алгоритми класифікації

3



Порівняння систем розподіленого аналізу даних

4

Характеристики	BC-PDM	Azure ML	Amazon ML	GoogleCloud ML	Watson Analytics
Хмарна модель	SaaS	SaaS	SaaS	SaaS	SaaS
Інтерфейс користувача	Web	Web	Web	API	Web
Користувальницький рівень	розробник	аналітик	аналітик	розробник	аналітик
API	Ні	REST	REST	REST	REST
Масштабованість обчислень	Так	Так	Так	Так	Так
Місце зберігання даних	Всередині хмари	Всередині хмари	Всередині хмари	Всередині хмари	Всередині хмари
Платформа розподілених обчислень	Apache Hadoop	Vowpal Webbit (Hadoop)	Apache Hadoop	Apache Hadoop	-
Повний цикл аналізу	Так	Так	Так	Так	Ні
Завдання аналізу	Classification, Clustering, Association	Classification, Anomaly detection, Regression, Clustering	Classification, Regression	Classification, Regression	Classification, Regression
Додавання нових алгоритмів	Ні	Its Machine Learning Marketplace	Ні	Ні	Ні
Використання	Ні	Платно	Платно	Платно	Платно

Порівняння систем розподіленого аналізу даних

4

Характеристики	BC-PDM	Azure ML	Amazon ML	GoogleCloud ML	Watson Analytics
Хмарна модель	SaaS	SaaS	SaaS	SaaS	SaaS
Інтерфейс користувача	Web	Web	Web	API	Web
Користувачський рівень	розробник	аналітик	аналітик	розробник	аналітик
API	Hi	REST	REST	REST	REST
Масштабованість обчислень	Так	Так	Так	Так	Так
Місце зберігання даних	Всередині хмари	Всередині хмари	Всередині хмари	Всередині хмари	Всередині хмари
Платформа розподілених обчислень	Apache Hadoop	Wowza Webbit (Hadoop)	Apache Hadoop	Apache Hadoop	-
Повний цикл аналізу	Так	Так	Так	Так	Hi
Завдання аналізу	Classification, Clustering, Association	Classification, Anomaly detection, Regression, Clustering	Classification, Regression	Classification, Regression	Classification, Regression
Додавання нових алгоритмів	Hi	Its Machine Learning Marketplace	Hi	Hi	Hi
Використання	Hi	Платно	Платно	Платно	Платно

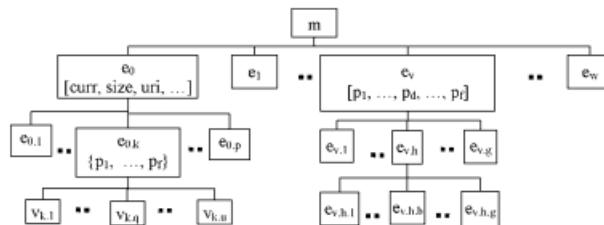
6

Модель представлення знань

$$m = [e_0, e_1, e_2, \dots, e_v, \dots, e_w]$$

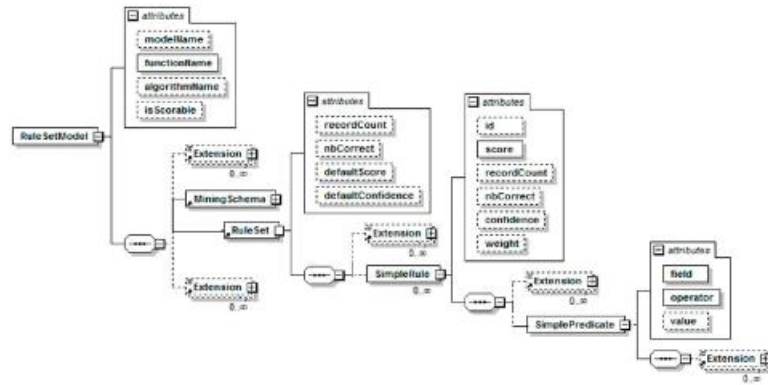
$$e_0 = \langle [\text{curr, size, uri, login, password, \dots}], [e_{0.1}, e_{0.2}, \dots, e_{0.k}, \dots, e_{0.p}] \rangle,$$

$$e_{0.k} = \langle [\text{pos, size, name, role \dots}], [vk.1, \dots, vk.q, \dots, vk.u] \rangle$$



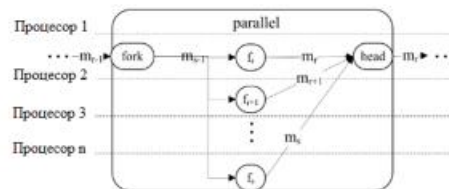
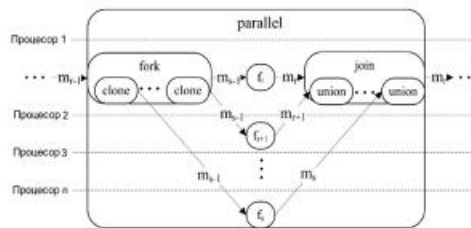
PMML модель знань RuleSetModel

7



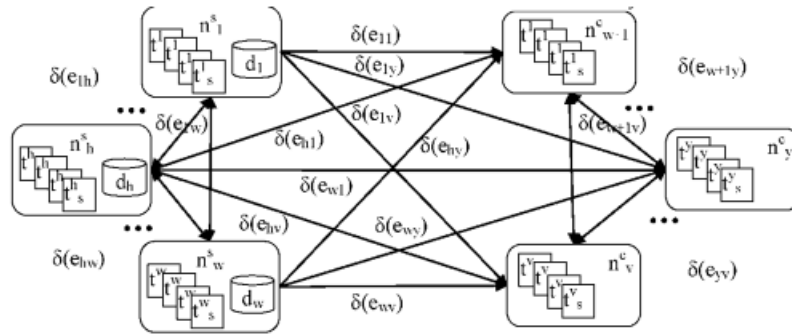
Паралельне виконання

8



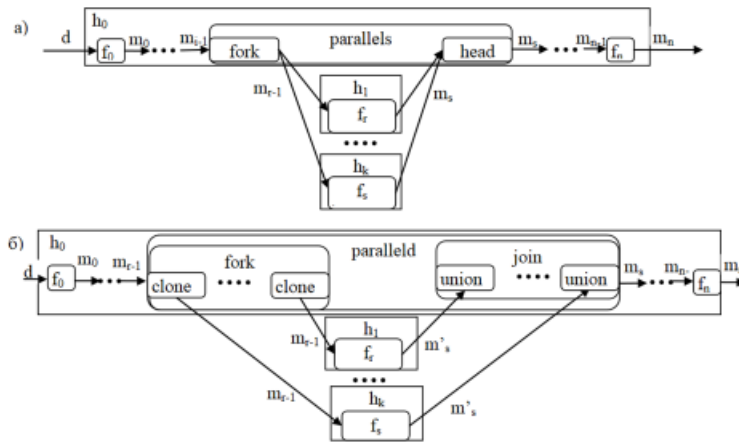
Фізичне середовище виконання алгоритмів аналізу даних

9



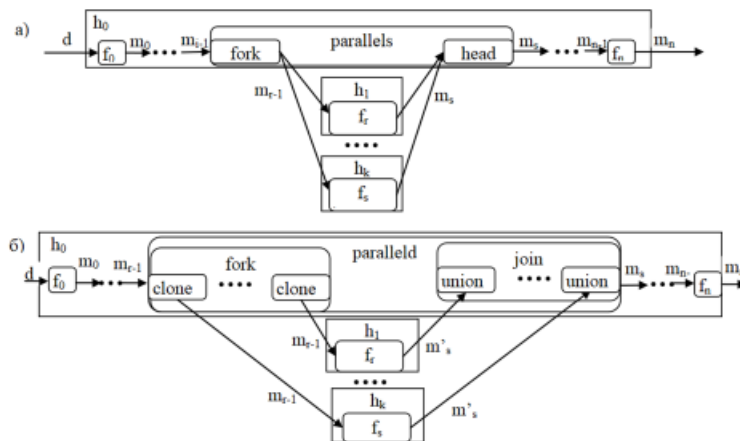
Паралельне виконання алгоритму аналізу даних

10



Паралельне виконання алгоритму аналізу даних

10



Модифікований метод

12

Розпаралелювання композиції алгоритму аналізу даних має виконуватися з урахуванням типу розподілу даних. У разі розподіленого зберігання даних ($|N^s| > 1$) безпосередньо перед розпаралелюванням повинна бути змінена композиція алгоритму аналізу даних відповідно до виду розподілу даних (горизонтальне або вертикальне). Для цього може бути використаний метод оптимізації структури алгоритму аналізу даних (`optimizeStructure(f, flag)`).

В результаті його виконання зовнішнім циклом буде або цикл по векторам `loopd` (в разі горизонтального розподілу) або цикл по атрибутам `loore [0]` (в разі вертикального розподілу).

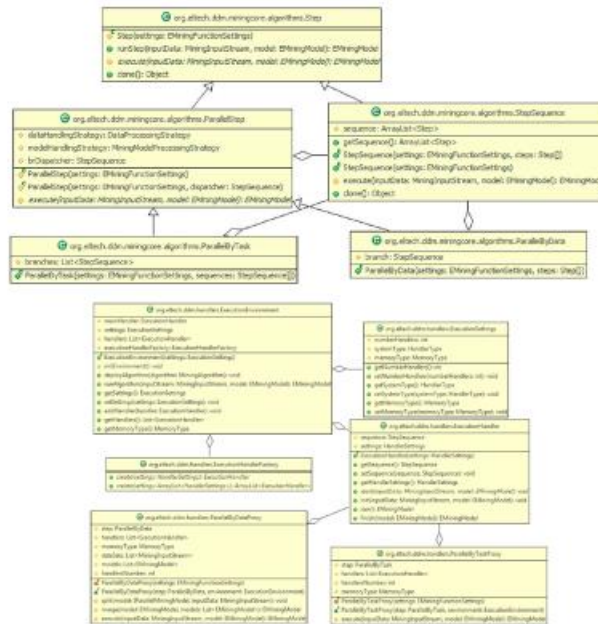
У разі розподіленого зберігання даних перевіряється можливість розпаралелювання відповідного циклу для систем з розподіленою пам'яттю перевірка здійснюється методом перевірки умов паралельного виконання пари ФОМ `verifyParallelize ((loopes1 ift), (loopes + 1 + 1 ift))`.

Метод застосовується до ФОМ f_t , спричиненої ітераційно у відповідному циклі на різних ітераціях $s_i + 1$. Незалежно від типу функції `parallel`, що повертається, цикл виконується паралельно функцією `paralleld`. Якщо розпаралелювання циклу не можливо (метод `verifyParallelize` повертає `null`), то для аналізу даних обраним алгоритмом необхідний їх попередній збір в єдиному сховищі.

Діаграма класів для реалізації ФОМ алгоритмів аналізу даних

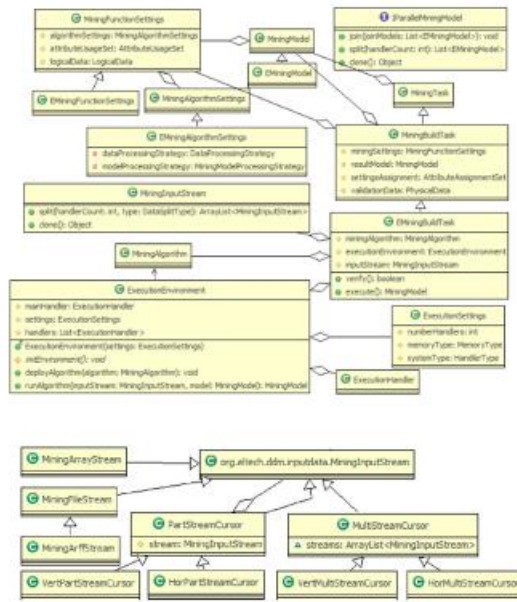


Діаграми класів



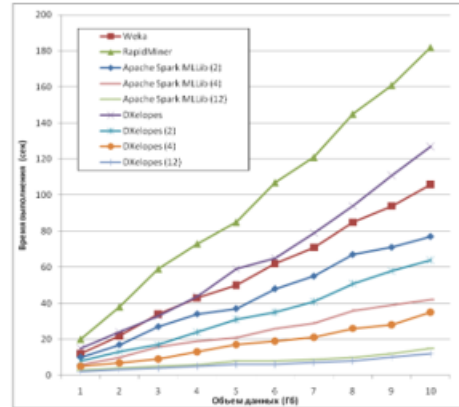
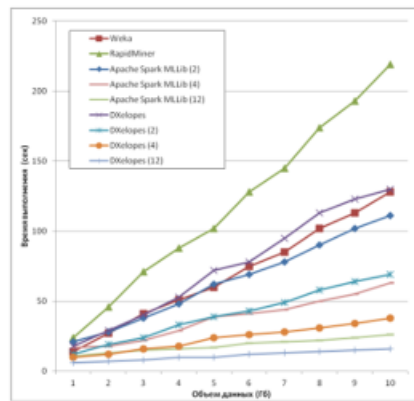
Базові класи конфігурації

15



Аналіз отриманих результатів

16



Висновки

16

Проведено аналіз існуючих моделей, методів і засобів побудови паралельних алгоритмів, який показав, що вони не дозволяють виконувати побудову паралельних алгоритмів аналізу даних з урахуванням умов їх виконання. Представлена модель знань, яка описує знайдені в даних закономірності у вигляді лісу дерев уніфікованих елементів і визначає функції, що забезпечують їх паралельне побудова в загальній і розподіленій пам'яті. Модифікований метод паралельного алгоритму аналізу даних, що дозволяє проводити розпаралелювання, як за даними, так і за завданнями, для виконання в загальній і розподіленій пам'яті. Розроблено бібліотека паралельних алгоритмів аналізу даних для виконання в заданому середовищі, що включає в себе програмні реалізації запропонованих моделей і методів. Проведено експериментальні дослідження побудови паралельних алгоритмів аналізу даних з урахуванням умов виконання, які підтвердили достовірність отриманих результатів.