

А Н М Базлур Рашид, Тонмой Чоудури

## **Аналіз відбору характеристик високорозмірних фінансових даних в управлінні знаннями з точки зору спільної коєволюції та програмної моделі MapReduce**

Термін «великі дані» характеризує великі об'єми даних, створених завдяки передовим технологіям, у різних сферах з використанням 4Vs – volume, velocity, variety, veracity для визначення об'єму даних, які можуть бути оброблені лише шляхом обчислювально-трудомісткого аналізу, швидкості їх створення, різних типів даних, а також їх точності. Високорозмірні фінансові дані, наприклад, дані часових рядів і просторово-часові дані, містять велику кількість характеристик (показників) і малу кількість моделей, використовуваних для оцінки стану бізнесу в режимі реального часу у фінансових організаціях. Такі набори даних є зазвичай зашумленими. Між цими характеристиками можуть існувати сильні зв'язки. У багатьох сферах, включаючи фінансову, немає аналітичних інструментів для пошуку даних для видобутку знань через високорозмірність. Відбір характеристик є задачею оптимізації для відбору мінімального ряду важливих характеристик, що підвищує точність класифікації та скорочує обчислення. Традиційних статистичних методів відбору характеристик недостатньо для подолання «прокляття розмірності», пов'язаного з великими даними. Спільна коєволюція, мета-алгоритми та принцип «розділяй і володарюй» ділить високорозмірні задачі на менші під-задачі. Далі програмна модель MapReduce пропонує готову до використання поширену, адаптовану та відмовостійку інфраструктуру для паралелізування розробленого алгоритму. У статті представлено аналіз еволюційних методів відбору характеристик в управлінні знаннями з точки зору сучасної спільної коєволюції та програмної моделі MapReduce, а також напрямки майбутніх досліджень.

**Ключові слова:** великі дані, оптимізація, обчислювальні методи, мета-алгоритми, декомпозиція задач, паралельне програмування, видобуток знань

**Класифікація JEL:** M11, M15, C61, C63



Стаття знаходиться у відкритому доступі і може розповсюджуватися на умовах ліцензії Creative Commons Attribution-NonCommercial 4.0 International license, що дозволяє повторне використання, розповсюдження та відтворення, забороняє використання матеріалів у комерційних цілях та вимагає наявності відповідного посилання на оригінальну версію статті.

А Н М Базлур Рашид, Тонмой Чоудури

## **Анализ отбора характеристик высокоразмерных финансовых данных в управлении знаниями с точки зрения совместной коэволюции и программной модели MapReduce**

Термин «большие данные» характеризует большие объемы данных, созданных благодаря передовым технологиям, в различных сферах с использованием 4Vs – volume, velocity, variety, veracity для определения объема данных, которые могут быть обработаны только путем вычислительно-трудоемкого анализа, скорости их создания, различных типов данных, а также их точности. Высокорамерные финансовые данные, например, данные временных рядов и пространственно-временные данные, содержат большое количество характеристик (показателей) и малое количество моделей, используемых для оценки состояния бизнеса в режиме реального времени в финансовых организациях. Такие наборы данных являются обычно зашумленными. Между данными характеристиками могут существовать сильные связи. Во многих сферах, включая финансовую, нет аналитических инструментов для поиска данных для добычи знаний из-за высокоразмерности. Отбор характеристик является задачей оптимизации для отбора минимального ряда важных характеристик, что повышает точность классификации и сокращает вычисления. Традиционных статистических методов отбора характеристик недостаточно для преодоления «проклятия размерности», связанного с большими данными. Совместная коэволюция, мета-алгоритмы и принцип «разделяй и властвуй» делит высокоразмерные задачи на меньшие под-задачи. Далее программная модель MapReduce предлагает готовую к использованию распространенную, адаптированную и отказоустойчивую инфраструктуру для параллелизации разработанного алгоритма. В статье представлен анализ эволюционных методов отбора характеристик в управлении знаниями с точки зрения современной совместной коэволюции и программной модели MapReduce, а также направления будущих исследований.

**Ключевые слова:** большие данные, оптимизация, вычислительные методы, мета-алгоритмы, декомпозиция задач, параллельное программирование, добыча знаний

**Классификация JEL:** M11, M15, C61, C63



Статья находится в открытом доступе и может распространяться на условиях лицензии Creative Commons Attribution-NonCommercial 4.0 International license, что позволяет повторное использование, распространение и воспроизведение, запрещает использование материалов в коммерческих целях и требует наличия соответствующей ссылки на оригинальную версию статьи.