

Старший викладач Копичко С.М., магістрант Гордієнко Р.О.

**Національний технічний університет України
«Київський політехнічний інститут»**

ПРОГНОЗУВАННЯ СТАНУ ВУЗЛІВ У GRID-СИСТЕМІ

Abstract

**Kopychko S., student Gordiienko R.
*Prediction of state of the Grid nodes***

The paper describes needs of estimation of parameters which identify nodes in Grid middleware schedulers' for right distribution of load in system. Sampling of main parameters are made. Proposed method for estimate a reliability of node on the base of probability. Described result of forecasting sampled parameters of nodes in UAG and EGEE with the neural networks. The prospects for further research and modification are proposed.

Вступ

Наукова робота сьогодення пов'язана з величезною кількістю обчислювань. Обчислювальної потужності персональних комп'ютерів не вистачає для поточних потреб, тому для великих наукових спільнот були створені розподілені обчислювальні системи, які дістали назву наукових GRID-систем. Як приклад, GRID-технології застосовуються для моделювання та обробки даних в експериментах на Великому андронному колайдері. Розвиток GRID-технологій в Україні розпочався з державної програми Української академічної GRID-системи.

Постановка задачі

Виділення ресурсів та планування задач є основною проблемою у GRID-обчисленнях, а це, без сумнівів, найважливіший фактор, що впливає на продуктивність, оскільки це допомагає уникнути простоїв у GRID-системі та скорочує час виконання задач користувача [1].

Прогнозування для підвищення продуктивності

Будь-яка GRID-система оснащена системою моніторингу [2], на яку орієнтується планувальник під час розподілу задач. Проте, він не може узагальнити принцип динамічної зміни ресурсів. Для виконання цієї задачі необхідно мати прогноз або оцінку продуктивності кожного вузла у майбутньому. Далеко не кожний GRID має у своєму арсеналі компоненти, які виконують ці функції. Саме для цих систем і проводяться дослідження та розробки у цій області[3].

Головний ресурс GRID-системи – обчислювальні кластери. Їх потужності можливо умовно поділити на обчислювальні елементи та елементи для зберігання даних. В цій роботі увага приділена обчислювальним елементам вузлів. Їх стан можливо оцінити такими параметрами, як кількість вільний процесорів, максимальна кількість процесорів, яка може бути виділена під одну задачу, можливість локального розпаралелювання задачі на декілька процесорів, типи процесорів, довжина черги задач на вузлі в даний момент часу і т.д. Для повноцінної оцінки вузлів необхідно врахувати якомога більше параметрів, хоча деякими з них можливо знехтувати (наприклад, такими, як тип програмного забезпечення на вузлі) для спрощення моделі вузла.

Стратегії прогнозування

Припустимо, що GRID складається з j вузлів, кожен вузол має k ресурсів, таких як кількість процесорів, кількість задач у черзі, кількість доступної пам'яті тощо. $rs \in \mathbb{R}$, а $j, k \in \mathbb{N}$, $\mathbb{N}=\{1,2 \dots n\}$, і матриця RS^{GRID} GRID ресурсів визначається наступним чином:

$$RS_{j \times k}^{GRID} = \begin{bmatrix} rs_{1,1} & \dots & rs_{1,k} \\ \vdots & \ddots & \vdots \\ rs_{j,1} & \dots & rs_{j,k} \end{bmatrix}$$

Кожен ресурс може бути представлений функцією $rs(t)$, оскільки його стан динамічно змінюється. Т.ч. спрогнозувати стан вузлів в GRID-системі означає спрогнозувати rs . Системи моніторингу збирають інформацію про ресурси та зберігають її у вигляді так званих слідів роботи системи.

Таким чином, для кожного ресурсу ми маємо часовий ряд, заданий системою моніторингу, котрий необхідно оцінити для представлення інформації планувальнику про стан системи в конкретний момент часу в майбутньому.

Методи прогнозування

Існують різні методи прогнозування: від лінійної регресії до прогнозування за допомогою нечіткої логіки. Детерміновані процеси характеризуються достатньою інформацією для визначення функціональної залежності $y=f(t)$. В нашому випадку процес протікає в умовах невизначеності, тому традиційний підхід - використовувати стохастичні моделі часових рядів, розділяючи часовий ряд на систематичну та випадкову компоненту. Задача моделювання часових рядів в загальному вигляді може бути сформульована наступним чином. Нехай задані значення часового ряду $Y = \{y(1), y(2), \dots, y(N)\}$, де $y(t)$ – значення показника досліджуваного процесу, що був зареєстрований на t -м такті часу ($t = 1, 2, \dots, N$). Потрібно побудувати оцінки майбутніх значень ряду $\hat{Y} = \{\hat{y}(N+1), \hat{y}(N+2), \dots, \hat{y}(N+\tau)\}$, $1 \leq \tau \leq N$, де τ – горизонт прогнозування [4].

Незалежно від методу, що використовується, ми припускаємо, що закономірність змін, яка виявлена для визначеного періоду часового ряду в минулому, збережеться на обмеженому відрізку часу в майбутньому.

В [5] та [6] були проведені експерименти з використання статистичного підходу до моделювання часових рядів, де модель часового ряду представлялась як $y_t = f(x_t, a) + \varepsilon_t$, де $f(x_t, a)$ – систематична компонента з параметром a , ε_t – випадкова компонента.

Дана робота має за мету поліпшити отримані результати та оцінити витрати на збільшення точності прогнозу/оцінки. Для цього пропонується використати інший популярний підхід до прогнозування – побудова та навчання нейронних мереж.

В нейромережевому підході задача прогнозування часових рядів формулюється як задача розпізнавання образів, для розв'язання якої формується навчальна послідовність даних часового ряду, і нейронна мережа навчається розпізнавати відповідні образи.

Опис даних та експерименту

Стан вузла буде характеризуватися такими параметрами, як ймовірність доступності вузла, кількість вільних процесорів на вузлі, середній час простою в черзі, кількість завдань у черзі, максимальний процесорний час для одного завдання.

Для навчання нейронних мереж використовувалися дані про стан вузлів, зібрані протягом тижня в двох GRID-системах: УАГ(український Академічний GRID [7]) и EGEE II (Enabling Grid for E-sciensE [8]). Були використані рекурентні нейронні мережі, нейронні мережі, що використовують радіально-базисні функції як активаційні функції, гібридні нейронні мережі. У перших двох видах присутній один прихований шар, що складається з п'яти нейронів, кількість нейронів на вхідному шарі дорівнює семи, кількість нейронів у вихідному шарі дорівнює одиниці. Гібридна нейронна мережа має два прихованих шари по п'ять нейронів в кожному. Кількість ітерацій для навчання - 672.

Результати експерименту

Як було сказано в [5] вузли GRID-системи можна умовно поділити на ті, в яких стабільне завантаження та ті, в яких нестабільне завантаження. За результатами проведених експериментів було з'ясовано що при оцінці вузлів зі стабільним завантаженням відносна похибка при використанні будь-якої мережі не перевищує 3%, що можна вважати дуже добрим результатом. При оцінці вузлів з нестабільним завантаженням похибка в залежності від типу мережі варіювалася від 14% в гібридній нейронній мережі і до 27% в рекурентній нейронній мережі при горизонті прогнозування рівному 49.

Висновки

Використання нейронних мереж дало можливість зменшити похибку оцінювання стану вузлів порівняно зі статистичними підходом до оцінювання. Найкраще себе зарекомендувала гібридна нейронна мережа, показавши похибку в 2% на вузлі зі стабільним завантаженням, і 14% на вузлі з нестабільним завантаженням. Недоліком даного підходу є «генетичний» недолік нейронних мереж - їх довгий період навчання і можливість надлишкового навчання.

У перспективі доцільно розглянути прогнозування стану вузлів за допомогою нечіткої логіки.

Література

1. *Гордієнко Р.О., Копичко С. М., Лавренюк А. М.* Методи оптимізації розподілення ресурсів та задач в GRID-мережі // Міжнародна конференція «Системний аналіз та інформаційні технології», 2011
2. *Форстер Я., Кессельман К.*, Физиология Грид // 2004-11-29
3. *Liang Hu, Xiaochun Cheng, Xilong Che*, Survey of Grid Resource Monitoring and Prediction Strategies // http://www.humanpub.org/ijiiip/ppl/08_IJIIIP1-026024JE.pdf
4. *Н.Г Ярушкіна, Т.В. Афанасьєва, И.Г. Перфильєва*, Интеллектуальный анализ временных рядов, Ульяновск, УлГТУ, 2010. – 320с.
5. *Лавренюк С.И., Копычко С.Н., Гордиенко Р.А.* Оценка параметров загрузки узлов Грид-системы для оптимизации ее производительности // Вісник НТУУ «КПІ». Інформатика, управління та обчислювальна техніка: Зб. наук. пр. – К.: Век+, – 2011. – № 53. – 211 с.
6. *Ю.Ю. Жебель, С.Н. Нечаусов, А.В. Шевченко*, Прогнозирование нагрузки потока задач на GRID-ресурсы, 2010
7. Ukrainian Academic Grid web-site, <http://com.bitp.kiev.ua>
8. European GRID infrastructure web-site, <http://www.egi.eu>