

ПРОГНОЗУВАННЯ ЦІНИ ХМАРНИХ РЕСУРСІВ AMAZON ЗА ДОПОМОГОЮ РЕКУРЕНТНОЇ НЕЙРОННОЇ МЕРЕЖІ

Метод середньострокового прогнозування ціни обчислювальних ресурсів хмарних сервісів за допомогою рекурентної нейронної мережі представлено в даній статті. Проведено аналіз відомих рішень та представлено модель рекурентної нейронної мережі. Експериментальні дослідження проведено на даних хмарного сервісу Amazon EC2. Результати прогнозування показали високу точність запропонованого підходу, що дозволяє використати його для виконання обчислювальних задач користувачів хмарних сервісів найбільш економічним чином.

Ключові слова: хмарні сервіси, ціна ресурсів, прогнозування, рекурентна нейронна мережа.

V.O. SHULTS

Ternopil National Economic University

SPOT PRICE PREDICTION OF AMAZON CLOUD RESOURCES USING RECURRENT NEURAL NETWORK

A method for the middle-term prediction of spot prices for cloud services using recurrent neural network is considered in this paper. The analysis of the existing solutions is done and the model of a recurrent neural network is presented. Simulation modeling results on the Amazon Elastic Cloud Computing spot instances showed a high accuracy of the proposed approach. It allows to use the proposed method for running computational tasks on a cloud in a most economic way.

Keywords: Spot Market; Cloud Computing; Prediction; Recurrent Neural Network.

ПОСТАНОВКА ЗАДАЧІ

Хмарні обчислення є новою інформаційною технологією «за вимогою», тому доступ до обчислювальних ресурсів хмарних сервісів є завжди відкритий [1]. Спочатку провайдери хмарних сервісів виставляли постійну ціну за використання хмарних ресурсів [2-3], однак з часом, коли хмарних сервісів ставало все більше, ресурси почали розділятися на окремі конфігурації. Використання фіксованої ціни стало неефективним через те, що сумарний об'єм запитів користувачів був значно меншим від загальної наявної кількості хмарних ресурсів. Тому провайдери почали розробляти механізми стимулювання користувачів до використання їх хмарних сервісів, в тому числі і через відповідну цінову політику. В 2009 році компанія Amazon [4] запропонувала гнучку ціну на набір хмарних ресурсів, що перебували у надлишку компанії. Це дало змогу користувачам «торгуватися» за ціну, за якою вони згодні використовувати той чи інший хмарний ресурс. Такий механізм ціноутворення став дуже подібним до стандартного аукціонного механізму формування ціни, при якому ціна залежить від попиту та пропозиції. Таким чином, ціна хмарного ресурсу залежить від максимальної ціни за годину, яку користувач згідний заплатити за цей ресурс, бажаного географічного розміщення ресурсу та кількості ресурсів, що бажає використати користувач. Якщо максимальна ціна пропозиції перевищує поточну ціну обчислювального ресурсу, то обчислювальні задачі, що виконуються на цьому ресурсі, завершуються користувачем, і ціна ресурсу збільшується до максимальної ціни пропозиції. Ціни обчислювальних ресурсів при цьому встановлюються на початку кожної години. Таким чином, аукціонний механізм формування ціни хмарних ресурсів (i) робить цю ціну динамічною, (ii) забезпечує найкращим чином вимоги користувача, формуючи мінімальну ціну на обчислювальні ресурси на вимогу, а також (iii) забезпечує максимізацію використання обчислювальних ресурсів провайдера хмарних сервісів [1]. Отже, користувачі, що вказали вищу ціну, перші отримують доступ до відповідних обчислювальних ресурсів. Користувачі, чії обчислювальні задачі можуть завантажуватися на виконання не обов'язково негайно, а у будь-який інший момент часу, можуть суттєво знизити свої фінансові витрати на оплату хмарних сервісів. Це може бути досягнуто шляхом прогнозування поточної ціни обчислювальних ресурсів, що дозволить їх використовувати в моменти часу, коли ціна найменша, а отже найбільш економічним та ефективним чином. В загальному випадку коливання попиту та пропозиції на ринку важко описати детермінованою математичною моделлю через випадкову природу цього процесу. Тому для прогнозування таких процесів доцільно використовувати універсальні моделі, що можуть відтворювати перетворення вхід-вихід певної системи з достатньою точністю. Однією з таких моделей є апарат штучних нейронних мереж, зокрема модель рекурентної нейронної мережі [5].

АНАЛІЗ ОСТАННІХ ДОСЛІДЖЕНЬ ТА ПУБЛІКАЦІЙ

В статті [6] описано процес ціноутворення ресурсів хмарного сервісу Amazon EC2 (Elastic Computing Cloud) з використанням річних даних про ціну з чотирьох обчислювальних центрів компанії Amazon. Дане дослідження здійснено шляхом аналізу ціни ресурсів та інтервалу часу між зміною ціни, що дозволило визначити часову динаміку ціни обчислювальних ресурсів по годинах за день і по днях тижня. В статті [7] запропоновано механізм встановлення «контрольних точок» для зменшення ціни обчислювальних ресурсів на основі реальних даних з хмарного сервісу Amazon EC2. В статті показано, що адаптивна технологія встановлення «контрольних точок» може зменшити ціну ресурсів та зменшити час завершення

обчислювальних задач. В статті [8] запропонована модель прийняття рішень для оптимізації продуктивності, коштів та надійності використання обчислювальних ресурсів. Показано, що запропонована модель може бути використана для здійснення оптимальних «торгів» з метою досягнення різних цілей з бажаним рівнем точності. В статті [9] запропоновано метод використання ціни обчислювальних ресурсів для прискорення виконання MapReduce задач. В статті використано Марковські послідовності для прогнозування тривалості використання обчислювального ресурсу, однак не наведено, яку точність забезпечує цей підхід. В статті [1] досліджено метод прогнозування ціни ресурсів хмарного сервісу Amazon EC2 за допомогою багатшарового персеPTRону, і показано, що середня та максимальна похибка прогнозу не перевищують 2.4% та 6.3% відповідно про прогнозуванні ціни обчислювального ресурсу на 1-5 днів наперед. В статті показано, що в задачах прогнозування нейромережеві технології забезпечують дуже добру альтернативу математичним методам, завдяки їх здатності до самоадаптації та самонавчання, що дозволяє налаштуватись на динамічні, нелінійні та нестационарні властивості вхідних даних. Тому в даній статті досліджено метод прогнозування ціни обчислювальних ресурсів хмарного сервісу за допомогою рекурентної нейронної мережі.

МЕТА

Метою даної статті є дослідження методу прогнозування ціни обчислювальних ресурсів хмарного сервісу за допомогою рекурентної нейронної мережі.

МАТЕРІАЛИ ТА РЕЗУЛЬТАТИ ДОСЛІДЖЕННЯ

Рекурентна нейронна мережа (РНМ) зі зворотними зв'язками та нелінійними функціями активації нейронів є простою, добре вивченою та широко використовуваною моделлю нейронної мережі (НМ) [5, 10].

Математична модель РНМ (рис. 1) описується виразами

$$y = F_3 \left(\sum_{j=1}^M w_{j3} h_j - T_0 \right),$$

$$h_j = F_2 \left(\sum_{i=1}^N w_{ij} x_i + \sum_{k=1}^M w_{kj} h_k(t-1) + w_{3j} y(t-1) - T_j \right),$$

де M - кількість нейронів схованого шару, N - кількість нейронів вхідного шару, w_{j3} - ваговий коефіцієнт від j -го нейрону схованого шару до вихідного нейрону, x_i - вхідні значення, w_{kj} - ваговий коефіцієнт від контекстного нейрону k схованого шару до нейрону j того ж шару, $h_k(t-1)$ - вихідне значення контекстного нейрону k схованого шару у попередній момент часу $t-1$, w_{3j} - ваговий коефіцієнт від контекстного вихідного нейрону до нейрону j прихованого шару, $y(t-1)$ - значення контекстного вихідного нейрону у попередній момент часу $t-1$, T_j - поріг нейронів схованого шару, T_0 - поріг вихідного нейрону [5, 10].

Для нейронів схованого та вихідного шару використана нелінійна сигмоїдна функція активації

$$F(x) = \frac{1}{1 + e^{-x}}.$$

Для навчання РНМ використано алгоритм зворотного поширення помилки [5, 10]. Для проведення експериментальних досліджень використано історичні дані про ціну обчислювальних ресурсів середньої ланки хмарних сервісів, що працюють під управлінням операційної системи Лінукс. Ці дані позначаються логотипом m1.linux та знаходяться на веб-сайті компанії Amazon [4]. Для експериментів використано набір з близько 3000 даних за 6 місяців (185 днів) з грудня 2009 року по травень 2010 року включно. Це в середньому 17 записів про ціну обчислювальних ресурсів за один день. Подібно до статті [1], обрано режим середньострокового прогнозу, так як користувачі хмарних сервісів можуть планувати свої стратегії торгів на декілька днів наперед, при цьому період від одного до п'яти днів є найбільш важливим. Прогноз здійснювався одно-кроковим методом з перенавчанням РНМ на кожному кроці прогнозування [11]. Для навчання РНМ вибрано 176 даних з початку кожного місяця. РНМ мала структуру 5-10-1, використано адаптивний та постійний (0.1 для схованого та вихідного шарів) кроки навчання. РНМ навчалась до середньої квадратичної помилки 10^{-5} та 5×10^5 кількості навчальних епох. Час навчання склав в середньому 90 секунд при обробці 176 даних. Всі експерименти виконані на процесорі AMD Phenom II x4 965 із тактовою частотою 3.4 ГГц із 4-ма Гб ОЗУ. Загальний час обчислень всього експерименту для середньострокового прогнозу з перенавчанням на кожному кроці прогнозу становив понад 4 години. Середня та максимальна відносні помилки прогнозування на 1-5 днів показані в Табл. 1, їх графічне представлення для кожного із 6-ти місяців показані на Рис. 2-4. Результати експериментальних досліджень показують, що РНМ забезпечує достатньо точні результати середньострокового прогнозування для більшості випадків.

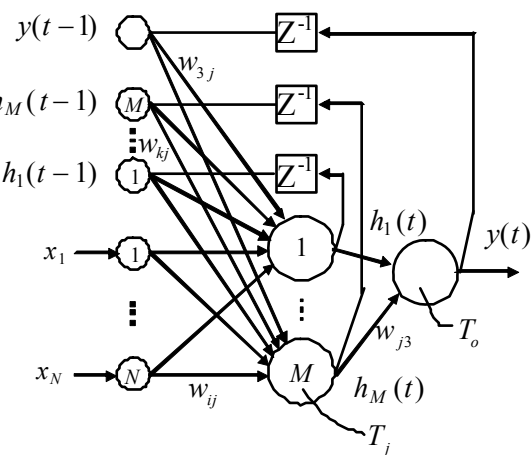


Рис. 1. Структура рекурентної НМ

Результати прогнозування

| Місяць | Відносні похибки прогнозування, % | | | | | | | | | |
|---------------------------|-----------------------------------|------------|----------|------------|------------|------------|------------|-------------|------------|-----------|
| | 1 день | | 2 дні | | 3 дні | | 4 дні | | 5 днів | |
| | сер | макс | сер | макс | сер | макс | сер | макс | сер | макс |
| Грудень 2009 | 4.6 | 10.9 | 3.4 | 7.4 | 5.6 | 31.3 | 11.9 | 27.5 | 14.4 | 27.5 |
| Січень 2010 | 1.7 | 6.2 | 2.0 | 7.8 | 1.8 | 7.8 | 1.7 | 7.8 | 1.8 | 7.8 |
| Лютий 2010 | 2.6 | 4.6 | 2.3 | 4.6 | 2.3 | 4.6 | 2.3 | 4.6 | 2.4 | 15.1 |
| Березень 2010 | 2.1 | 4.2 | 2.1 | 5.3 | 2.2 | 5.3 | 2.3 | 5.3 | 2.5 | 13.7 |
| Квітень 2010 | 1.1 | 2.7 | 1.2 | 2.9 | 1.5 | 3.5 | 2.0 | 16.3 | 2.1 | 16.3 |
| Травень 2010 | 1.4 | 2.5 | 1.3 | 3.5 | 1.4 | 3.8 | 1.4 | 3.8 | 1.5 | 4.4 |
| Усереднена похибка | 2.2 | 5.1 | 2 | 5.2 | 2.4 | 9.3 | 3.6 | 10.9 | 4.1 | 14 |

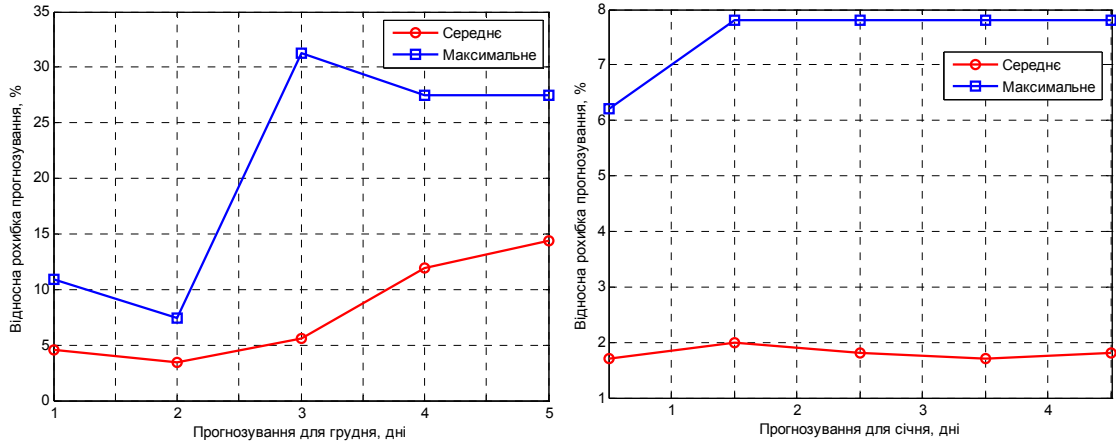


Рис. 2. Результати середньострокового прогнозування для m1.linux за грудень 2009-січень 2010

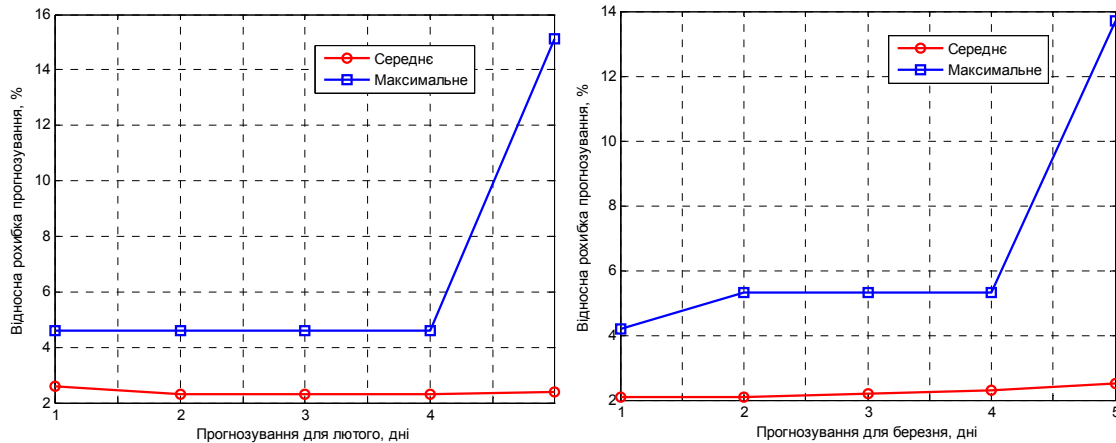


Рис. 3. Результати середньострокового прогнозування для m1.linux за лютий-березень 2010

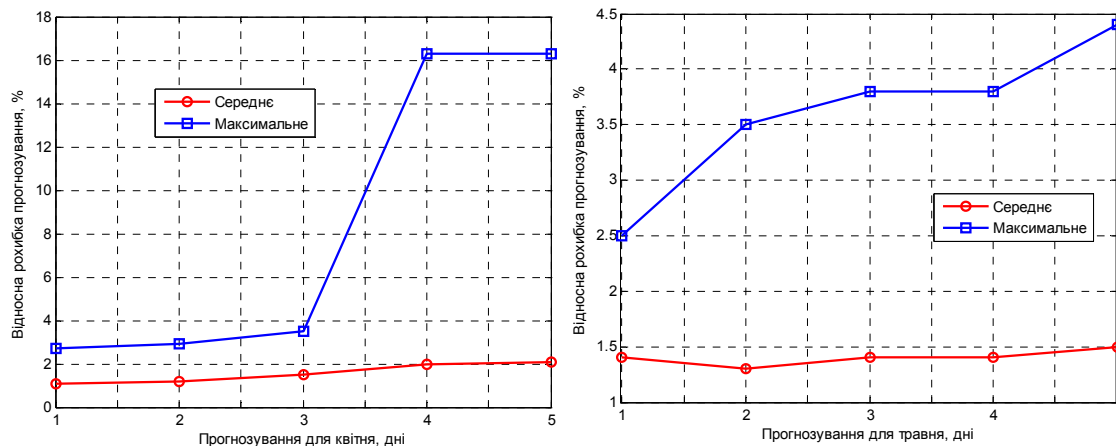


Рис. 4. Результати середньострокового прогнозування для m1.linux за квітень-травень 2010

ВІСНОВКИ

Метод середньострокового прогнозування ціни обчислювальних ресурсів хмарних сервісів за допомогою рекурентної нейронної мережі представлено в цій статті. Експериментальні дослідження проведені на даних хмарного сервісу Amazon EC2. Результати прогнозування показали високу точність запропонованого підходу, при цьому середня та максимальна відносна похибка прогнозу не перевищують 4.1% та 14% відповідно при прогнозуванні ціни обчислювального ресурсу на 1-5 днів наперед для 6-ти досліджуваних місяців 2009-2010 років. Отримані результати експериментальних досліджень показали, що рекурентна нейронна мережа добре підходить для виконання задач прогнозування і може бути корисна для користувачів хмарних сервісів для виконання їх обчислювальних задач найбільш економним чином.

Література

1. Турченко І. Нейромережевий прогноз ціни ресурсів хмарного сервісу в середньостроковій перспективі // Системи обробки інформації. – Х.: Харківський університет Повітряних Сил імені Івана Кожедуба. – 2014. – Вип. 2 (118). – С. 211-215.
2. Chun B. N., Buonadonna P., Auyoung A., Ng C., Parkes D. C. et al. Mirage: A microeconomic resource allocation system for sensor network testbeds // Proceedings of the 2nd IEEE Workshop on Embedded Networked Sensors, 2005.
3. Weng C., Li M., Lu X., Deng Q. An economic-based resource management framework in the grid context // IEEE International Symposium on Cluster Computing and Grid CCGrid. - Vol. 1. - 2005. - P. 542–549.
4. Amazon EC2 spot price history, 2012. Available: <http://aws.typepad.com/aws/2011/07/ec2-spot-pricing-now-specific-to-each-availability-zone.html>.
5. Golovko V.A. Neural Networks: learning, models and applications. – М.: Radiotekhnika, 2001. – 256 p (in Russian).
6. Javadi B., Thulasiram R. K., Buyya R. Statistical modeling of spot instance prices in public cloud environments // UCC. IEEE Computer Society. – 2011. - P. 219–228.
7. Yi S., Kondo D., Andrzejak A. Reducing costs of spot instances via checkpointing in the Amazon elastic Compute Cloud // 3rd IEEE International Conference on Cloud Computing. – 2010. – P. 236–243.
8. Andrzejak A., Kondo D., Yi S. Decision model for cloud computing under SLA constraints // 18th IEEE/ACM International Symposium on Modeling, Analysis and Simulation of Computer and Telecommunication Systems (MASCOTS). – 2010. – P. 257–266.
9. Chohan N., Castillo C., Spreitzer M., Steinder M., Tantawi A., Krintz C. See spot run: using spot instances for MapReduce workflows // 2nd USENIX conference on Hot topics in cloud computing. – 2010. – P. 7–17.
10. Haykin S. Neural Networks: A Comprehensive Foundation, 2nd Edition: Prentice Hall, 1998, 842 p.
11. Turchenko V., Beraldi P., De Simone F., Grandinetti L. Short-term Stock Price Prediction Using MLP in Moving Simulation Mode // Proceedings of the 6th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems (IDAACS2011). – Prague (Czech Republic). – 2011. - P. 666-671.

References

1. Turchenko I. Neural-based prediction of cloud service prices in middle-term perspective // Systemy obrobky informatsii. – H.: Harkivskiy universytet Povitrianyh syl imeni Ivana Kogheduba. – 2014. – Vyp. 2 (118). – P. 211-215.
2. Chun B. N., Buonadonna P., Auyoung A., Ng C., Parkes D. C. et al. Mirage: A microeconomic resource allocation system for sensor network testbeds // Proceedings of the 2nd IEEE Workshop on Embedded Networked Sensors, 2005.
3. Weng C., Li M., Lu X., Deng Q. An economic-based resource management framework in the grid context // IEEE International Symposium on Cluster Computing and Grid CCGrid. - Vol. 1. - 2005. - P. 542–549.
4. Amazon EC2 spot price history, 2012. Available: <http://aws.typepad.com/aws/2011/07/ec2-spot-pricing-now-specific-to-each-availability-zone.html>.
5. Golovko V.A. Neural Networks: learning, models and applications. – М.: Radiotekhnika, 2001. – 256 p (in Russian).
6. Javadi B., Thulasiram R. K., Buyya R. Statistical modeling of spot instance prices in public cloud environments // UCC. IEEE Computer Society. – 2011. - P. 219–228.
7. Yi S., Kondo D., Andrzejak A. Reducing costs of spot instances via checkpointing in the Amazon elastic Compute Cloud // 3rd IEEE International Conference on Cloud Computing. – 2010. – P. 236–243.
8. Andrzejak A., Kondo D., Yi S. Decision model for cloud computing under SLA constraints // 18th IEEE/ACM International Symposium on Modeling, Analysis and Simulation of Computer and Telecommunication Systems (MASCOTS). – 2010. – P. 257–266.
9. Chohan N., Castillo C., Spreitzer M., Steinder M., Tantawi A., Krintz C. See spot run: using spot instances for MapReduce workflows // 2nd USENIX conference on Hot topics in cloud computing. – 2010. – P. 7–17.
10. Haykin S. Neural Networks: A Comprehensive Foundation, 2nd Edition: Prentice Hall, 1998, 842 p.
11. Turchenko V., Beraldi P., De Simone F., Grandinetti L. Short-term Stock Price Prediction Using MLP in Moving Simulation Mode // Proceedings of the 6th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems (IDAACS2011). – Prague (Czech Republic). – 2011. - P. 666-671.

Рецензія/Peer review : 8.11.2014 р.

Надрукована/Printed : 1.1.2015 р.