
УДК 004.051

І.В. Турченко

Тернопільський національний економічний університет, Тернопіль

НЕЙРОМЕРЕЖЕВИЙ ПРОГНОЗ ЦІНИ РЕСУРСІВ ХМАРНОГО СЕРВІСУ В СЕРЕДНЬОСТРОКОВІЙ ПЕРСПЕКТИВІ

Метод середньострокового прогнозування ціни обчислювальних ресурсів хмарного сервісу за допомогою багатошарового перцептрон розглянуто в даній статті. Проведено аналіз відомих рішень та представлено модель використаної нейронної мережі. Експериментальні дослідження проведено на даних хмарного сервісу Amazon EC2. Результати прогнозування показали високу точність запропонованого методу, який може бути застосований для виконання обчислювальних задач користувачів хмарних сервісів найбільш економічним чином.

Ключові слова: *хмарні сервіси, ціна ресурсів, прогнозування, багатошаровий перцептрон.*

Вступ

Постановка проблеми. Хмарні обчислення (cloud computing) розглядаються сьогодні як інформаційна технологія загального призначення, реалізо-

вана по схемі «за вимогою». Це означає, що доступ до обчислювальних ресурсів хмарних сервісів (cloud services) є завжди відкритий, і користувачі можуть використовувати ці сервіси влюбий, зручний для них, момент часу. При цьому основною «рушійною

силою» такої схеми «за вимогою» є економічна вигода користувачів, що показано в дослідженні Армбруста та інших [1]. На початковому етапі провайдери хмарних сервісів виставляли постійну ціну для користувачів за використання хмарних ресурсів [2 – 3]. З часом, коли хмарних сервісів ставало все більше, і вони почали розділятися на окремі конфігурації, використання фіксованої ціни стало неефективним через те, що сумарний об'єм вимог користувачів був значно меншим від загальної наявної кількості хмарних ресурсів. Тому провайдери хмарних ресурсів почали розробляти механізми стимулювання користувачів до використання їх хмарних сервісів, в тому числі і через відповідну цінову політику.

Проілюструвати взаємодію ціни користувача з наявністю хмарного ресурсу, що може бути використаний за цю ціну, доцільно на прикладі хмарного сервісу Amazon EC2 (Elastic Computing Cloud) [4]. В 2009 році компанія Amazon запропонувала набір хмарних ресурсів, що раніше постійно знаходилися в надлишку їх обчислювальних центрів, за гнучкою ціною. Це дало змогу користувачам «торгуватися» за ціну, за якою вони згодні використовувати той чи інший хмарний ресурс. Таким чином, цей механізм ціноутворення в Amazon EC2 став дуже подібним до стандартного аукціонного механізму формування ціни, при якому ціна залежить від попиту та пропозиції. В інтерпретації Amazon EC2 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 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 ресурси на вимогу, а також (iii) забезпечує максимізацію використання обчислювальних ресурсів провайдера хмарних сервісів.

Отже, поточна ціна обчислювальних ресурсів в хмарних сервісах періодично змінюється на основі поточних значень попиту та пропозиції, і користувачі, що вказали вищу ціну, перші отримують доступ до відповідних обчислювальних ресурсів. Користувачі, чії обчислювальні задачі можуть завтажуватися 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 використання обчислювальних ресурсів, що дозволить використовувати їх в момен-

ти часу коли ціна найменша, а отже найбільш економічним та ефективним чином. В загальному випадку коливання попиту і 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 багатошарового перцептронну, який є універсальним апроксиматором [5].

Метою статті є дослідження точності прогнозу ціни обчислювальних ресурсів хмарних сервісів за допомогою апарату штучних нейронних мереж.

Аналіз останніх досліджень та публікацій. Автори праці [6] дослідили характеристики ціноутворення ресурсів хмарного сервісу Amazon EC2, використовуючи річні дані про ціну обчислювальних ресурсів 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. Вони також запропонували статистичну модель, що відтворювала аналізовані дані про ціну ресурсів. Модель базувалася на Гаусівському розподілі випадкової величини. Автори констатували, що точність відтворення аналізованих даних склала 95%, хоча не вказали, на якій вибірці даних ця точність була досягнута, та не навели власне самих результатів дослідження точності прогнозування в графічній чи табличній формах. Автори праці [7] запропонували механізм встановлення «контрольних точок» для зменшення ціни обчислювальних ресурсів. Вони використали реальні дані з хмарного сервісу Amazon EC2 і показали, як адаптивна технологія встановлення «контрольних точок» може зменшити ціну ресурсів та зменшити час завершення обчислювальних задач. Автори роботи [8] запропонували модель прийняття рішень для оптимізації продуктивності, коштів та надійності на основі так званої «згоди про рівень надання послуги» (Service Level Agreement (SLA) використовуючи дані про ціну ресурсів Amazon EC2. Вони також використали реальні дані з цього хмарного сервісу та показали, як ця модель може використовуватись для здійснення оптимальних «торгів» з метою досягнення різних цілей з бажаним рівнем точності. Автори роботи [9] запропонували метод використання ціни обчислювальних ресурсів для прискорення виконання MapReduce задач. В статті використано Марковські послідовності для прогнозування тривалості використання ресурсів, однак не наведено дані про точність цього підходу.

В останні роки багато різних методологій було запропоновано для забезпечення надійного прогнозу-

вання. Ці методології включали в себе широкий спектр різноманітних методів, від статистичного аналізу [10] до використання штучних нейронних мереж. При цьому нейромережеві технології забезпечують дуже добру альтернативу математичним методам, завдяки їх здатності до самоадаптації та самонавчання, що дозволяє налаштуватись на динамічні, нелінійні та нестаціонарні властивості вхідних даних. Так як ціна ресурсів формується ринковими категоріями попиту та пропозиції, то природа цих даних відповідає природі багатьох фінансових показників. Аналіз праць [11 – 13] показав, що відома модель багатошарового персептрону успішно була застосовна для прогнозування різноманітних фінансових показників. Тому доцільно розглянути застосування даної моделі для прогнозу ціни ресурсів хмарного сервісу.

Матеріали та результати дослідження

Багатошаровий персептрон (БШП) з нелінійними функціями активації нейронів є простою, добре вивченою та широко використовуваною моделлю нейронної мережі (НМ) [14 – 15]. Вихідне значення тришарового персептрону

$$y = F_3 \left(\sum_{i=1}^N w_{i3} h_i - T \right),$$

де N – кількість нейронів схованого шару (рис. 1), w_{i3} – ваговий коефіцієнт від i -го нейрону схованого шару до вихідного нейрону, h_i – вихідне значення i -го нейрону схованого шару, T – поріг вихідного нейрону, F_3 – функція активації вихідного нейрону [14].

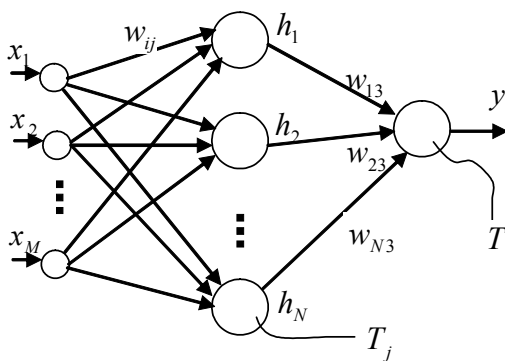


Рис. 1. Структура тришарового персептрону

Вихідне значення j -го нейрону схованого шару

$$h_j = F_2 \left(\sum_{i=1}^M w_{ij} x_i - T_j \right),$$

де w_{ij} – вагові коефіцієнти від вхідних нейронів до j -го нейрону схованого шару, x_i – вхідні значення, T_j – поріг j -го нейрона.

Для нейронів схованого та вихідного шару використана нелінійна сигмоїдна функція активації

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

Алгоритм зворотного поширення помилки [14 – 15] використано для навчання багатошарового персептрону. Він базується на методі градієнтного спуску та забезпечує ітераційну процедуру модифікації вагових коефіцієнтів та порогів НМ для кожного навчального вектора p з вибірки навчання:

$$\Delta w_{ij}(t) = -\alpha \frac{\partial E^P(t)}{\partial w_{ij}(t)}, \quad \Delta T_j(t) = -\alpha \frac{\partial E^P(t)}{\partial T_j(t)},$$

де α – крок навчання, $\frac{\partial E^P(t)}{\partial w_{ij}(t)}$ і $\frac{\partial E^P(t)}{\partial T_j(t)}$ – градієнти функції помилки на кожній ітерації t для навчального вектора p , $p \in \{1, \dots, P\}$, де P – розмір вибірки навчання.

Середньоквадратична помилка навчання на кожній ітерації навчання t обчислюється як

$$E^P(t) = \frac{1}{2} (y^P(t) - d^P(t))^2,$$

де $y^P(t)$ – вихідне значення НМ на ітерації навчання t , $d^P(t)$ – бажане значення НМ на ітерації навчання t для кожного навчального вектора p . Загальна помилка навчання обчислюється згідно з виразом

$$E(t) = \sum_{p=1}^P E^P(t).$$

Для проведення експериментальних досліджень використано історичні дані про ціну обчислювальних ресурсів хмарного сервісу Amazon EC2, 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 3247 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. Для навчання БШП використано 88 даних з початку кожного місяця, середньостроковий прогноз здійснювався для одного, двох, трьох, чотирьох та п'яти днів, наступних за вибіркою даних для навчання в рамках кожного місяця.

Тришаровий персептрон 5-10-1 (5 вхідних нейронів, 10 нейронів схованого шару, 1 нейрон вихідного шару) обрано для здійснення прогнозу з формуванням вибірки навчання згідно з методом Бокса-Дженкінса [10]. Для прогнозування використано дві моделі БШП з адаптивним і постійним кроком навчання, який був встановлений в значення 0.5 та 0.5 для схованого та вихідного шару відповідно. Програмна реалізація персептрону виконана на мові програмування С. БШП навчався до середньоквадратичної помилки 10^{-5} при 10^6 кількості навчальних епох. Час виконання одного експерименту з прогнозування склав близько 30 секунд на комп'ютері з

процесором AMD Xeon 2x4 965 з тактовою частотою 3.4 GHz та 4 GB оперативної пам'яті.

Результати середньострокового прогнозування при використанні 88 даних кожного місяця для навчання БШП наведено в табл. 1. Індекс нижнього регі-

стру біля кожного результату означає: 1 – використано БШП з адаптивним кроком навчання, 2 – використано БШП з постійним кроком навчання. Графічна інтерпретація результатів середньострокового прогнозування для кожного місяця наведена на рис. 2 – 4.

Таблиця 1

Результати середньострокового прогнозування при використанні 88 даних для навчання БШП

| Місяць | Відносна похибка прогнозу, % | | | | | | | | | |
|-------------------------|------------------------------|------------------|------------------|-------------------|------------------|-------------------|------------------|-------------------|------------------|-------------------|
| | 1 день | | 2 дні | | 3 дні | | 4 дні | | 5 днів | |
| | Серед. | Макс. | Серед. | Макс. | Серед. | Макс. | Серед. | Макс. | Серед. | Макс. |
| Грудень 2009 | 4.3 ₂ | 8.4 ₂ | 4.0 ₂ | 11.4 ₂ | 4.5 ₂ | 11.5 ₂ | 4.1 ₂ | 11.5 ₂ | 4.3 ₂ | 14.7 ₂ |
| Січень 2010 | 1.7 ₂ | 4.3 ₂ | 1.6 ₂ | 4.3 ₂ | 1.5 ₂ | 4.3 ₂ | 1.7 ₂ | 5.4 ₂ | 1.7 ₂ | 5.4 ₂ |
| Лютий 2010 | 2.1 ₂ | 4.3 ₂ | 2.4 ₂ | 4.9 ₂ | 2.5 ₂ | 4.9 ₂ | 2.5 ₂ | 4.9 ₂ | 2.4 ₁ | 4.9 ₁ |
| Березень 2010 | 2.2 ₂ | 4.6 ₂ | 2.2 ₂ | 4.7 ₂ | 2.3 ₂ | 4.8 ₂ | 2.4 ₂ | 4.9 ₂ | 2.4 ₂ | 5.0 ₂ |
| Квітень 2010 | 1.2 ₁ | 2.2 ₁ | 1.2 ₁ | 2.2 ₁ | 1.3 ₂ | 2.8 ₂ | 1.3 ₂ | 2.9 ₂ | 1.4 ₂ | 3.4 ₂ |
| Травень 2010 | 1.4 ₁ | 4.2 ₁ | 1.4 ₁ | 4.1 ₁ | 1.5 ₂ | 4.2 ₂ | 1.6 ₂ | 4.2 ₂ | 2.0 ₂ | 4.3 ₂ |
| Середня похибка: | 2.2 | 4.7 | 2.2 | 5.2 | 2.3 | 5.4 | 2.3 | 5.6 | 2.4 | 6.3 |

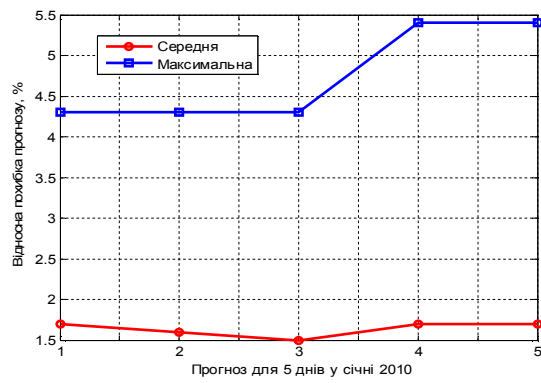
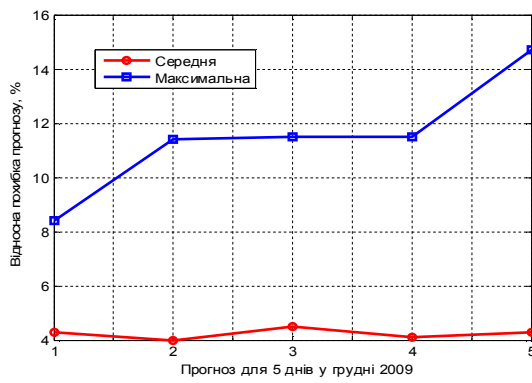


Рис. 2. Результати прогнозу ціни обчислювальних ресурсів для ml.linux за грудень 2009 та січень 2010

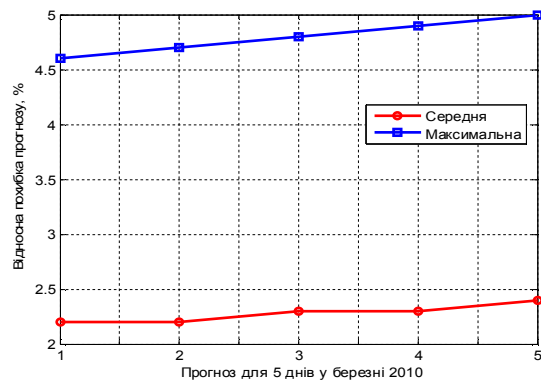
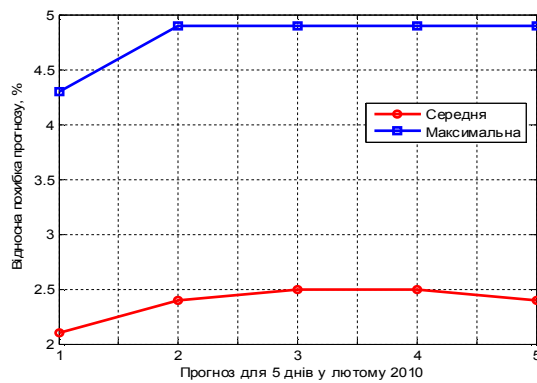


Рис. 3. Результати прогнозу ціни обчислювальних ресурсів для ml.linux за лютий 2010 та березень 2010

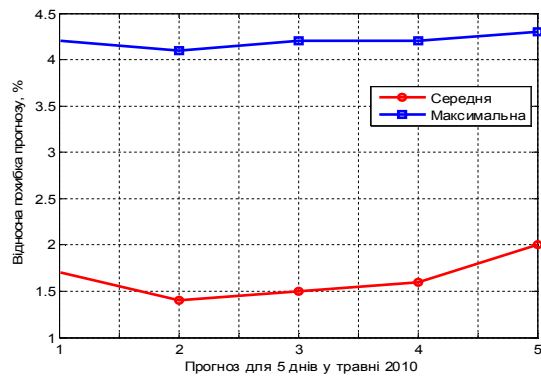
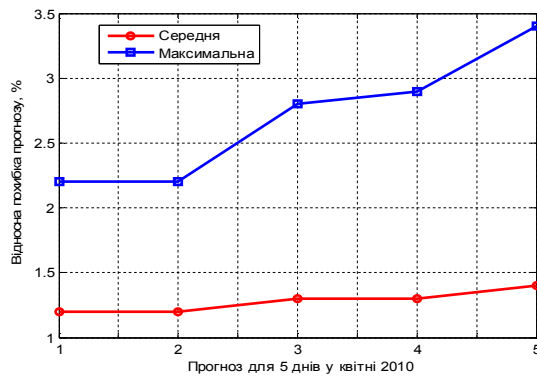


Рис. 4. Результати прогнозу ціни обчислювальних ресурсів для ml.linux за квітень 2010 та травень 2010

Результати експериментальних досліджень показали високу точність запропонованого підходу. При прогнозі ціни обчислювального ресурсу на 1-5 днів для 6-ти досліджуваних місяців середня відносна похибка прогнозування не перевищує 2.4%, а максимальна відносна похибка прогнозування не перевищує 6.3%.

Отримані результати експериментальних досліджень показали, що нейронні мережі добре підходять до виконання подібних задач прогнозування і можуть бути дуже корисними для користувачів хмарних сервісів, хто бажає виконувати свої обчислювальні задачі найбільш економним чином.

Висновки

В статті представлено метод середньострокового прогнозування ціни обчислювальних ресурсів хмарних сервісів за допомогою багатослового перцептрон. 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 2.4% та 6.3% 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 апарат штучних нейронних мереж добре підходить для виконання подібних задач прогнозування і може бути корисним для користувачів хмарних сервісів, хто бажає виконувати свої обчислювальні задачі найбільш економним чином.

Список літератури

1. Armbrust M. *Above the clouds: A Berkeley view of cloud computing* // EECS Department, University of California, Berkeley, Tech. Rep. UCB/EECS-2009-28, Feb 2009 / M. Armbrust, A. Fox, R. Griffith et al. Available: <http://www.eecs.berkeley.edu/Pubs/TechRpts/2009/EECS-2009-28.html/>
2. Chun B.N. *Mirage: A micro-economic resource allocation system for sensornet testbeds* / B.N. Chun, P. Buonadonna, A. Auyoung, C. Ng, D.C. Parkes et al // Proc. of the 2nd IEEE Workshop on Embedded Networked Sensors, 2005. – P. 19-28.
3. Weng C. *An economic-based resource management framework in the grid context* / C. Weng, M. Li, X. Lu, Q. Deng // 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. Hornik K. *Universal approximation of an unknown mapping and its derivatives using multilayer feedforward networks* / K. Hornik, M. Stinchcombe, H. White // Neural Networks. – Vol. 3, Iss. 5. – 1990. – P. 551-560.
6. Javadi B. *Statistical modeling of spot instance prices in public cloud environments* / B. Javadi, R.K. Thulasiram, R. Buyya // UCC. IEEE Computer Society. – 2011. – P. 219-228.
7. Yi S. *Reducing costs of spot instances via checkpointing in the Amazon elastic Compute Cloud* / S. Yi, D. Kondo, A. Andrzejak // 3rd IEEE International Conference on Cloud Computing. – 2010. – P. 236-243.
8. Andrzejak A. *Decision model for cloud computing under SLA constraints* / A. Andrzejak, D. Kondo, S. Yi // 18th IEEE/ACM Int. Symp. on Modeling, Analysis and Simulation of Comp. and Telecom. Systems (MASCOTS). – 2010. – P. 257-266.
9. Chohan N. *See spot run: using spot instances for MapReduce workflows* / N. Chohan, C. Castillo, M. Spreitzer, M. Steinder, A. Tantawi, C. Krintz // 2nd USENIX conference on Hot topics in cloud computing. – 2010. – P. 7-17.
10. Box G. *Time Series Analysis: Forecasting and Control: 4th edition* / G. Box, G. Jenkins, G. Reinsel. – New York: Wiley, 2008. – 784 p.
11. Gurusen E. *Using artificial neural network model in stock market index prediction* / E. Gurusen, G. Kayakulu, T.U. Daim // Expert Systems with Applications. – Vol. 38, Issue 8. – 2011. – P. 10389-10397.
12. Lin X. *Short-term stock price prediction based on echo state networks* / X. Lin, Z. Yang, Y. Song // Expert Systems with Applications. – Vol. 36. – 2009. – P. 7313-7317.
13. Kim S.H. *Graded forecasting using an array of bipolar predictions: application of probabilistic neural networks to a stock market index* / S.H. Kim, S.H. Chun // Intern. Jour. of Forecasting. – Vol. 14. – 1998. – P. 323-337.
14. Haykin S. *Neural Networks: A Comprehensive Foundation: 2nd Edition* / S. Haykin. – New Jersey: Prentice Hall, 1998. – 842 p.
15. Головкин В.А. *Нейронные сети: обучение, модели и применение* / В.А. Головкин. – М.: Радиотехника, 2001. – 256 с.

Надійшла до редколегії 28.02.2014

Рецензент: д-р техн. наук, проф. І.В. Шостак, Національний аерокосмічний університет ім. М.Є. Жуковського «ХАІ», Харків.

НЕЙРОСЕТОВОЙ ПРОГНОЗ ЦЕНЫ РЕСУРСОВ ОБЛАЧНОГО СЕРВИСА В СРЕДНЕСРОЧНОЙ ПЕРСПЕКТИВЕ

І.В. Турченко

Метод среднесрочного прогнозирования цены вычислительных ресурсов облачного сервиса с помощью многослойного перцептрона рассмотрено в данной статье. Осуществлен анализ известных решений и представлена модель нейронной сети. Экспериментальные исследования осуществлены на данных облачного сервиса Amazon EC2. Результаты прогнозирования показали высокую точность предложенного метода, который может быть использован для выполнения вычислительных задач пользователей облачных сервисов наиболее экономным образом.

Ключевые слова: облачные сервисы, цена ресурсов, прогнозирование, многослойный перцептрон.

NEURAL-BASED SPOT PRICE PREDICTION OF CLOUD SERVICE IN MIDDLE-TERM PERSPECTIVE

I.V. Turchenko

A method for the middle-term prediction of spot prices for cloud service using multi-layer perceptron is considered in this paper. The analysis of the existing solutions is provided and the model of a standard multi-layer perceptron is presented. The experimental research is fulfilled on the Amazon Elastic Cloud computing spot instances. The obtained prediction results showed a high accuracy of the proposed approach. It allows the users of cloud services to apply the proposed method for the running computational tasks in a most economic way.

Keywords: cloud services, spot prices, prediction, multi-layer perceptron.