

УДК 004.7

Интеллектуальный подход к моделированию работы GRID-системы в условиях неопределенности

С.И. Лавренюк

Институт кибернетики им. В.М.Глушкова НАН Украины

lsi@bigmir.net

Abstract

Lavreniuk S., Intelligent Approach to the Modeling of the GRID-Based Systems under Uncertainty. In this paper an approach to Grid system parameters estimation is described. This approach is based on using the mixing density neural networks. Proposed architecture has an advantage over traditional models as far as it provides estimation for whole distribution density function of the estimated value.

Введение

Можно смело констатировать тот факт, что на сегодняшний день имеет место широкое распространение Grid-систем [1], а также всестороннее использование их для решения разнообразных научных задач [2, 3]. Сложность задач, которые решаются такими системами, возрастает. Сегодня Grid-системы становятся более доступными возрастающему числу пользователей. При функционировании Grid-системы возникает вопрос об повышении эффективности использования таких систем и об гарантированном уровне качестве сервиса (SLA) и качества обслуживания (QoS) [4, 5] предоставляемого такой системой для пользователей в целом и ее узлами, компонентах и сервисами в отдельности.

При выполнении задач в Grid-системе для пользователя важно, чтобы его задача гарантированно выполнялась в положенные сроки и при этом в системе были доступны запрошенные ресурсы и сервисы.

Выполнение данных требований неразрывно связано с оптимизацией системы, оценкой определенных ее параметров и своевременным реагированием на отклонения рабочих параметров от заданных.

Постановка задачи

Важнейшей задачей повышения эффективности использования Grid-систем является планирование распределения задач между ресурсами и оценивание доступности информации и сервисов системы и ее компонентов для пользователей, которое осуществляется на основе оценки текущего состояния распределенной системы и загруженности ее ресурсов. Под доступностью информации и системы понимается возможность получить необходимую информацию из системы и работу сервиса в необходимое время и в необходимом объеме [6]. При постановке задач

оптимизации Grid-систему можно рассматривать как объект управления, функционирующий в условиях неопределенности, к которому применимы методы идентификации и оценивания. С помощью обратной связи можно обеспечить оптимальное управление этим объектом, например, оптимальную нагрузку на функциональные элементы системы, оценивать состояние системы на несколько шагов вперед и прогнозировать доступность системы.

Однако до сих пор теория управления широко не использовалась для управления Grid-системами, поскольку такие системы по своей природе являются иерархическими и трудно поддаются строгому аналитическому описанию. В последние годы только начали появляться работы, связанные с применением теории идентификации и управления [7] для моделирования нагрузки и планирования выполнения заданий в Grid-системе.

Одной из немногих работ, в которых задача оптимизации производительности Grid-системы рассматривается с точки зрения теории управления и для оценивания состояний модели применяется фильтр Калмана, является [8]. Однако в ней используется очень упрощенная линейная модель системы, которая не адекватна сложности исследуемого объекта.

На основе результатов структурно-функционального анализа (СФА) [9] предлагается несколько моделей таких систем, отражающих функционирование разных компонентов Grid-инфраструктуры. Полученные модели исследуются с позиций теории управления и идентификации. Нагрузка на отдельный узел Grid-системы описывается линейной моделью объекта управления с неизвестными возмущениями. Для представления неопределенности применяется теоретико-множественный подход [10, 11, 12]. Из-за отсутствия достоверной априорной информации о границах изменения неизвестных величин, оценивание неизвестных переменных выполняется в классе нечетких эллипсоидальных множеств [13].

Для описания функционирования сегмента или Grid-системы, включающей несколько узлов, и более полного учета показателей функционирования предлагается информационный или кибернетический подход к моделированию, когда система рассматривается как черный ящик и с учетом ее нелинейности представляется нейросетевой моделью. Для преодоления недостатков традиционных алгоритмов настройки нейросетевых моделей, в частности с целью обеспечения адаптивности и робастности процесса идентификации матриц весовых коэффициентов, используется метод рекуррентного оценивания параметров этой модели.

Сформулируем задачи идентификации и оценивания, которые возникают в процессе функционирования Grid-систем.

1. Оценивание загруженности ресурсов узла Grid-системы на основе линейной модели. Рассматривая задачу оценивания нагрузки на вычислительные и информационные ресурсы одного узла Grid-системы, можно ограничиться линейной моделью объекта управления с неизвестными возмущениями. Линейные модели оценивания параметров и состояний Grid-систем используются, в частности, в работах [8, 14]. Однако в этих работах рассматриваются слишком простые модели, не адекватные ложности задачи, в которых оценивается скалярная величина нагрузки на ресурс. В [15] предлагается использовать более адекватную модель объекта управления в векторном пространстве состояний, оставаясь при этом в классе линейных моделей. Для оценивания состояния такой модели в условиях нестатистически заданной неопределенности целесообразно использовать теоретико-множественный подход [12], в частности метод эллипсоидального оценивания [10, 11]. В условиях отсутствия достоверной априорной информации об оцениваемых величинах предлагается строить нечеткие множественные оценки неизвестных векторов [16].

2. Структурно-параметрическая идентификация модели в переменных «вход-выход». В работах [17, 18] для оценивания и прогнозирования нагрузки на ресурс Grid-системы используются регрессионные модели, преимущественно первого порядка. Такой выбор простейшей модели объясняется недостатком априорной информации. При наличии данных о структуре и порядке модели, систему можно представлять в виде объекта управления в переменных «вход-выход», решая для него задачу параметрической идентификации одним из классических методов теории управления и идентификации [19]. Однако при оценивании состояния конкретных Grid-систем порядок

модели точно неизвестен, поэтому приходится решать задачу не только параметрической, но и структурной идентификации. Для ее решения предлагается использовать интеллектуальные методы поиска, в частности генетические алгоритмы [20].

3. Идентификация нейросетевой модели Grid-системы. При отсутствии уточняющей информации о структуре объекта управления его целесообразно моделировать на основе информационного или кибернетического подхода, представляя модель в виде черного ящика с измеряемыми входами и выходами. Для математического описания подобных нелинейных моделей удобно использовать формализм нейронных сетей, обеспечивающих адекватное представление как моделей в терминах «вход-выход» (модель NARX).

Задача 1 решена в [15], задача 2 решена и детально описана в [20].

Рассмотрим задачу 3 более подробно.

Построение модели

Перед рассмотрением разработанного метода, рассмотрим детальнее применение искусственных нейронных сетей, а именно многослойных персептронов к оцениванию состояния Grid-систем. На основе данных мониторинга работы Grid-сети [22] сформируем обучающую выборку из распределения $p(x, d)$:

$$\{(d_i, x_i), i = \overline{1, n}\}, \quad (1)$$

где d_i — вектор измерения — входной вектор нейросети, x_i — вектор входных параметров модели — выходной вектор нейросети, n — размер выборки. В большинстве случаев в качестве распределения шума $p(\varepsilon)$ выбирается гауссовское с нулевым средним [21].

Задача обучения нейросети состоит в нахождении оптимальной зависимости между входными и выходными элементами выборки в виде:

$$x_i = f(d_i, w) + \xi_i, \quad (2)$$

где $f(d, w)$ — функция, определяемая нейронной сетью, которая зависит от вектора входа d и вектора весовых коэффициентов w , ξ_i — вектора невязки. Оптимальность понимается в смысле минимизации критерия:

$$J(w) = \sum_i \|\xi_i\|^2 = \sum_i \|x_i - f(d_i, w)\|^2. \quad (3)$$

Данные мониторинга Grid-систем [22] собираются в течении года в УАГ, Nordugrid. Поэтому в данной задаче многочисленные проблемы, связанные с ограниченностью

выборки не так актуальны.

В данной работе в качестве обыкновенной нейронной сети используются многослойные перцептроны с одним скрытым слоем. При этом нейроны скрытого слоя имеют сигмоидальную активационную функцию, а нейроны выходного слоя — линейную. В этом случае функция $f(d, w)$ имеет следующий вид:

$$f(d, w) = b_2 + \sum_k \left[u_k \cdot S(d^T v_k + b_1^{(k)}) \right], \quad (4)$$

где векторы v_k и u_k формируют матрицы весовых коэффициентов нейронов скрытого и выходного слоев, а векторы b_1 и b_2 — весовые коэффициенты связей bias-нейронов с нейронами скрытого слоя и выходного слоя соответственно, $b_1^{(k)}$ — k -ая компонента вектора b_1 . В совокупности векторы b_1, b_2, v_k и u_k формируют вектор параметров нейронной сети w . В данной работе в качестве сигмоидальной функции S используется гиперболический тангенс. Схематически работа нейронной сети изображена на рис. 1.

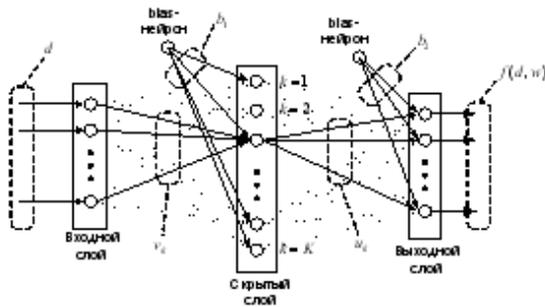


Рисунок 1 – Схема работы многослойного перцептрона

Основная проблема в применении обычных нейронных сетей для решения обратных задач заключается в неадекватном выборе функционала $J(w)$ [24], т.е. квадрате векторов невязки. В этом случае нейросеть будет приближать значение условного матожидания $M(x | d = d_i) = \int x p(x | d_i) dd_i$ вектора выхода при условии вектора входа [23].

Поэтому минимум $M \|x - f(d, w)\|^2$ будет достигаться для функции $f(d, w) = M(x | d = d)$. А это в частности означает, что ошибка аппроксимации нейронной сети в задаче (2) будет не меньше, чем $M \|x - M(x | d = d_i)\|^2 - \delta$, где значение

положительного числа δ определяется скоростью сходимости

$$\frac{1}{n} \sum_i \|x_i - f(d_i, w)\|^2 \xrightarrow[n \rightarrow \infty]{P} M \|x - f(d, w)\|^2$$

[23]. Тогда при условии больших обучающих выборок ошибка нейросети определяется дисперсией условного распределения $p(x | d)$, а выход — его матожиданием.

В случае нелинейных моделей распределение $p(x | d)$ может существенно отличаться от гауссовского, например, иметь бимодальную или мультимодальную форму. Тогда среднее значение, выдаваемое нейронной сетью, может лежать в области нулевой плотности и, соответственно быть нефизическим [24].

Для преодоления указанных недостатков нейронных сетей с квадратичным критерием оптимальности в работах [24, 25] были предложены специальные искусственные нейронные сети плотностей смесей (Mixture Density Networks, MDN).

Рассмотрим применение сетей плотностей смесей для решения задачи мониторинга работы Grid-сети, оценивания ее состояния. Для этого кратко рассмотрим отличия сетей плотностей смесей от обыкновенных нейронных сетей.

В данной вероятностной постановке обучение обыкновенной нейронной сети можно рассматривать как задачу нахождения оценки максимального правдоподобия

$$ML = -\ln \left(\prod_i p(x_i | d_i) \right) = -\sum_i \ln(p(x_i | d_i)) \quad (5)$$

при условии, что распределение шума ξ — нормальное с нулевым средним и диагональной матрицей ковариации $\sigma^2 I$, не зависящей от величины входа d . Тогда

$$ML(w) = -\sum_i \ln \left[\frac{1}{(\sqrt{2\pi}\sigma)^{n_x}} * \exp \left(-\frac{\|x_i - f(d_i, w)\|^2}{2\sigma^2} \right) \right] = n \cdot n_x \ln(\sqrt{2\pi}\sigma) + \frac{1}{2\sigma^2} \sum_i \|x_i - f(d_i, w)\|^2 = A + B \cdot J(w) \quad (6)$$

Обучение нейронных сетей плотностей смесей также состоит в нахождении оценки максимального правдоподобия, однако требование нормальности распределения шума

ξ (и, соответственно, плотности распределения $p(x|d)$) существенно ослабляется. Вместо этого плотность распределения $p(x|d)$ представляется в виде свертки гауссовских плотностей:

$$p(x|d) = \sum_{l=1}^L \alpha_l(d, w) \cdot \phi(x; m_l(d, w), \sigma_l(d, w)), \quad (7)$$

где

$$\phi(x; m, \sigma) = \frac{1}{(\sqrt{2\pi}\sigma)^{n_x}} \exp\left(-\frac{\|x-m\|^2}{2\sigma^2}\right) -$$

гауссовская плотность со средним m и диагональной матрицей ковариации $\sigma^2 I$, α_l — коэффициенты свертки ($\sum_l \alpha_l = 1$), L — количество элементов в свертке.

Для построения функций $\alpha_l(d, w)$, $m_l(d, w)$ и $\sigma_l(d, w)$ используется многослойный персептрон вида (4) с модифицированным выходным слоем. Количество нейронов в выходном слое равно $L(n_x + 2)$, поскольку для представления одной гауссовской функции плотности необходимо n_x переменных для задания вектора среднего и по одной для задания стандартного отклонения и коэффициента свертки. Поскольку на значения стандартного отклонения σ налагаются ограничения положительности, а на коэффициенты свертки α — нормированности на 1, то для задания $\sigma_l(d, w)$ используются экспоненциальные активационные функции, а для задания $\alpha_l(d, w)$ — функция softmax:

$$m_l(d, w) = b_{l,2}^{m_l} + \sum_k u_{l,k}^{m_l} \cdot S(d^T v_k + b_1^{(k)})$$

$$\sigma_l(d, w) = \exp\left\{ b_{l,2}^{\sigma_l} + \sum_k u_{l,k}^{\sigma_l} \cdot S(d^T v_k + b_1^{(k)}) \right\}, \quad (8)$$

$$\alpha_l(d, w) = \text{softmax} \left\{ b_{l,2}^{\alpha_l} + \sum_k u_{l,k}^{\alpha_l} \cdot S(d^T v_k + b_1^{(k)}) \right\}$$

где значение функции softmax определяется как:

$$\text{softmax}_i(x) = \frac{\exp(x_i)}{\sum_i \exp(x_i)}. \quad (9)$$

Схематически работа нейронной сети плотностей смесей изображена на рис. 2.

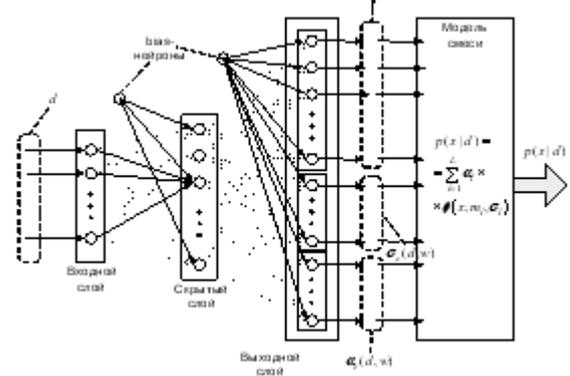


Рисунок 2 – Схема работы нейронной сети плотностей смесей

Рассмотрим работу сетей плотностей смесей на двух тестовых примерах, для которых распределение $p(x|d)$ можно вычислить аналитически. Обозначим конфигурацию многослойного персептрона в виде вектора $(n_{in}, n_{hid}, n_{out})$, где n_{in} — количество нейронов во входном слое, n_{hid} — в скрытом слое, а n_{out} — в выходном. Конфигурацию сети плотностей смесей будем обозначать с помощью вектора $(n_{in}, n_{hid}, n_{mix}, dim_{out})$, где n_{in} — количество нейронов во входном слое, n_{hid} — в скрытом слое, n_{mix} — количество ядер в смеси гауссовских функций, а dim_{out} — размерность выходного пространства сети.

В первом тестовом примере случайные вектора x и d являются одномерными, равномерно распределенными на отрезке $[0;1]$ и независимыми, т.е. распределение d полностью определяется распределением шума e , а модель $h()$ тождественно равна 0. В этом случае распределение $p(x|d)$ — равномерное на отрезке $[0;1]$. Для оценки работы многослойного персептрона и сети плотности смеси в качестве обучающей выборки используется выборка размером 100000. Для оценки использовался многослойный персептрон в конфигурации (1,5,1) и сеть плотности смесей в конфигурации (1,5,3,1).

Продолжение работ

В дальнейшем планируется провести заверку предложенных методов и моделей на данных, собранных из реально работающих Grid-узлов УАГ, NorduGrid, а также некоторых узлов EGEE. И выработать комплексные методы оценивания параметров Grid-узлов их доступности и качества работы в контексте выполнения. Сейчас ведутся работы по определению наиболее информативных параметров системы.

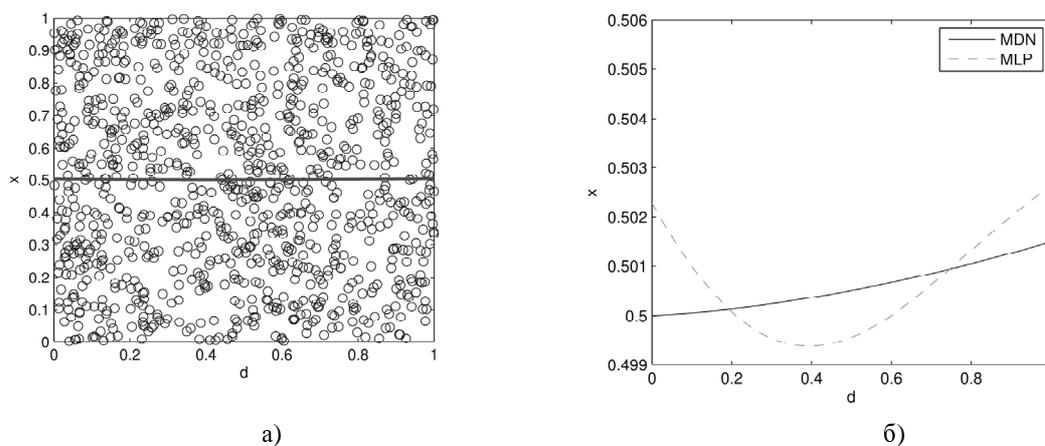


Рисунок 3 – Результаты работы сетей MLP и MDN: а) точки обучающего множества и результат работы MLP в зависимости от входных данных d (d в $[0,1]$); б) выход сети MLP и MDN, теоретическое значение выхода — 0,5

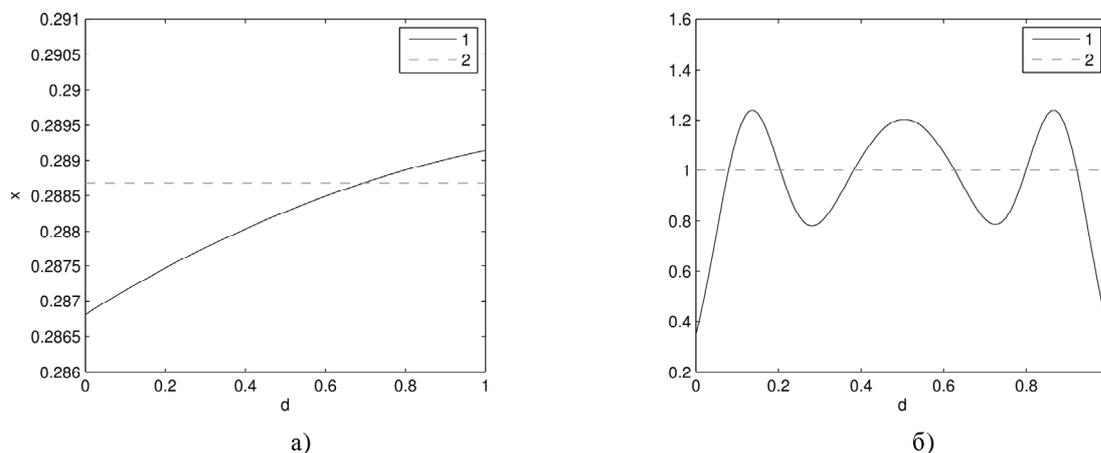


Рисунок 4 – Результаты работы сетей MLP и MDN: а) сравнение стандартного отклонения распределения $p(x|d)$, вычисленного в результате работы сети MDN (1) и теоретическое значение (2); б) графики плотности распределения $p(x|d)$ при $d=0,5$: вычисленного в результате работы сети MDN (1) и теоретическое значение (2)

Выводы

В зависимости от постановки задачи и наличия информации о структуре и порядке системы предложено использовать три класса моделей для оценивания состояния Grid-систем: линейная модель переменных состояния с неизвестными возмущениями, модель в переменных «вход-выход» и нейросетевая модель, применяемая для описания нелинейных объектов общего вида. Для оценки состояний и идентификации моделей во всех случаях применяются элементы интеллектуальных вычислений: нестатистическая неопределенность описывается в классе нечетких множеств, для структурно-параметрической идентификации модели в терминах «вход-выход» используется ГА, а для идентификации нейросетевой модели

предлагается рекуррентный робастный алгоритм множественного оценивания.

Временные характеристики работы систем, в частности время отклика или задержки при выполнении задачи, зависят от политики (динамики) обработки очередей для различных функциональных элементов системы: очереди выполнения задач процессором, очередей пакетов, передаваемых через сокеты, очередей на устройствах маршрутизации и т.п. В такой трактовке очередь выступает в роли элемента, который формирует разницу входных и выходных потоков сообщений в каждый момент времени, а значит, ее можно моделировать с помощью разностных (а в пределе — дифференциальных) уравнений с применением методов теории управления и идентификации.

На основе предложенного подхода

можно идентифицировать структуру и показатели узла Grid-системы, а затем параметры Grid-системы или ее отдельного узла, применить ее при реализации (оптимизации) той получить оценку обобщенного вектора или иной политики распределения задач.

Литература

1. Foster I., Kesselman C. The Grid 2: Blueprint for a New Computing Infrastructure. — Morgan Kaufmann, San Francisco, Calif, 2003. — 748 p.
2. Сергієнко І.В. Про основні напрямки створення інтелектуальних інформаційних технологій / І.В.Сергієнко // Системні дослідження та інформаційні технології. - 2002. - № 1. - С. 39–64.
3. Грід — нова інформаційно-обчислювальна технологія для науки / А.Г. Загородній, Г.М. Зінов'єв, С.С. Мартинов, С.Я. Свистунов, В.М. Шадура // Вісник НАН України. — 2005. — № 6. — С. 17-25.
4. Qin Liu , Xiaohua Jia , Chanle Wu, Optimal precomputation for mapping service level agreements in grid computing, Future Generation Computer Systems, v.24 n.8, p.849-859, October, 2008.
5. Al-Ali, R. von Laszewski, G. Amin, K. Hategan, M. Rana, O. Walker, D. Zaluzec, N. QoS Support for High-Performance Scientific Grid Applications, - Cluster Computing and the Grid, 2004. CCGrid 2004. IEEE International Symposium on .
6. Термінологія в галузі захисту інформації в комп'ютерних системах від несанкціонованого доступу: НД ТЗІ 1.1-003-99. — К.: ДСТСЗІ СБ України, 1999. — 26 с.
7. Кунцевич В.М. Управление в условиях неопределенности: гарантированные результаты в задачах управления и идентификации / В.М. Кунцевич. — К.: Наукова думка, 2006. — 206 с.
8. Tian Z., Liu L., Yang Y., Zhai Z. A stochastic control model for hierarchical grid service // NPC 2005, LNCS 3779. Ed. by H. Jin, D. Reed, W. Jiang. □ 2005. □ P. 72–79.
9. Куссуль Н. Grid-системы для задач исследования Земли. Архитектура, модели и технологии / Н.Куссуль, А.Шелестов. - К.: Наук. думка, 2008. - 452 с.
10. Бакан Г.М. Аналитический синтез алгоритмов гарантированного оценивания состояний динамических объектов / Г.М. Бакан // Проблемы управления и информатики. — 2003. — № 3. — С. 38–55.
11. Волосов В.В. Робастные алгоритмы эллипсоидального оценивания состояния многомерных нестационарных непрерывных динамических систем / В.В.Волосов // Проблемы управления и информатики. — 1999. — № 1. — С. 38–52.
12. Кунцевич В.М. Синтез оптимальных и адаптивных систем управления. Игровой подход / В.М.Кунцевич, М.М.Лычак. — К.: Наук. думка, 1985. — 286 с.
13. Куссуль Н.Н. Оценивание фазового состояния линейных многомерных динамических объектов с использованием размытых эллипсоидальных множеств / Н.Н.Куссуль, А.Ю.Шелестов // Проблемы управления и информатики. — 1995. — № 1. — С. 53–63..
14. Prem H., Srinivasa Raghavan N.R. A support vector machine based approach for forecasting of network weather services // J. of Grid Computing. — 2006. — 4. — P. 89 114.
15. Лавренюк С. И. Идентификация моделей и оценка состояния Grid-систем / С. И. Лавренюк, А.Ю.Шелестов // Кибернетика и системный анализ. — 2009. — № 6. — С. 42-50.
16. Куссуль Н.Н., Шелестов А.Ю. Нечеткий эллипсоидальный наблюдатель состояния линейных динамических объектов с неизвестными возмущениями / Н.Н.Куссуль, А.Ю.Шелестов // Праці міжнар. конф. з управління "Автоматика-2000". — Львів, 2000. — С. 149–154.
17. Jain R. The art of computer systems performance analysis. — New York: John Wiley and Sons. — 1991. — 685 p.
18. Martinich J. Production and operations management: an applied modern approach. — New York: John Wiley and Sons. — 1996. — 944 p.
19. Сейдж Э.П., Уайт Ч.С. Оптимальное управление системами / Э.П.Сейдж, Ч.С.Уайт. — М.: Радио и связь, 1982. — 392 с.
20. Шелестов А. Ю. Структурная и параметрическая идентификация моделей объектов управления на основе генетического алгоритма / А. Ю.Шелестов // Управляющие машины и системы. — 2007. — № 5. — С. 50–60.
21. Vasour, C., Baret, F., Béal, D., Weiss, M., & Pavageau, K. (2006). Neural network estimation of LAI, fAPAR, fCover and LAI × Cab, from top of canopy MERIS reflectance data: Principles and validation. Remote Sensing of Environment, 105, 313–325.
22. Лавренюк С.И. Построение базы данных мониторинга состояния узлов grid-инфраструктуры / Лавренюк С.И., А.Н.Лавренюк, Ю.А.Грипич // материалы тр. 6-й междунар. конф. Теоретические та прикладные аспекты построения программных систем (Киев, 09-10 декабря 2009 г.). – TAAPSD' 2009, - С. 124-131.
23. Кравченко А.Н. Нейросетевой метод решения обратных задач для радиационных моделей растительного покрова / А.Н. Кравченко // Кибернетика и системный анализ. — 2009. — № 3. — С. 159-172.
24. Bishop, C. M. Neural networks and their applications. Review of Scientific Instruments 65(6), 1803–1832.
25. Bishop, C. M. Bayesian methods for neural networks. Technical Report NCRG/95/009, Neural Computing Research Group, Aston University.

Поступила в редакцию 30.03.2010