

УДК 004.032.26

Д.Ю. ВОРОНИН

Севастопольский национальный технический университет, Украина

## ОПЕРАТИВНАЯ ДИСПЕТЧЕРИЗАЦИЯ В РАСПРЕДЕЛЕННЫХ ВЫЧИСЛИТЕЛЬНЫХ СИСТЕМАХ НА ОСНОВЕ НЕЙРОСЕТЕВЫХ МОДЕЛЕЙ

*Рассмотрена задача построения оперативных сервисов по диспетчеризации заданий в распределенных вычислительных системах, ориентированных на обработку медицинской информации. Использование нейросетевых технологий позволило справиться с NP-полнотой рассматриваемой задачи и тем самым повысить гарантоспособность всей распределенной вычислительной системы в целом.*

**Ключевые слова:** гарантоспособность, нейронные сети, объекты критического применения, распределенная вычислительная система, диспетчеризация, нейросетевая оптимизация, NP-полнота.

### Введение

Широкое использование распределенных вычислений [1] в качестве опорного средства большинства информационных технологий (в здравоохранении, энергетике, банковских системах, бизнес процессах и т.д.), а также стремительно растущие требования к обеспечению гарантоспособности вычислительных сервисов [2], резко обострили проблему диспетчеризации в распределенных вычислительных системах [3] (РВС). Согласно [4], задача диспетчеризации (ЗД) является NP-полной и не может быть эффективно решена аналитически даже в случае наличия полной информации о пакете обрабатываемых в РВС заданий. Таким образом, в большинстве случаев приходится довольствоваться субоптимальными решениями, полученными при помощи различных эвристик, методов искусственного интеллекта, агентного подхода, имитационного моделирования [5], нейронных сетей [6] и т.п. В данной статье рассматривается подход, использующий аппарат нейронных сетей для повышения точности, а главное – реактивности диспетчерских решений в РВС, обрабатывающих медицинскую информацию.

### Постановка задачи

Схему организации вычислительного процесса в РВС представим при помощи кибернетической модели «черный ящик», то есть на одном из самых верхних уровней абстракции. В качестве входных данных выступает информация о структуре РВС и пакете решаемых в ней заданий по обработке медицинской информации. Под внутренними параметрами рассматриваемой системы понимается множество способов решения задачи диспетчеризации. Выходные данные формируются при помощи диспетчерских сервисов в РВС, учитывающих входные данные и внутренние управ-

ляющие параметры системы. Существует множество различных вариантов формализации функционирования гарантоспособных сервисов по диспетчеризации заданий в РВС. Предлагаемый в рамках данной статьи подход ориентирован на распределенную обработку медицинской информации в РВС.

Под ЗД будем понимать нахождение эффективного (с точки зрения определенных критериев) распределения заданий по узлам РВС. Сформулируем ЗД в терминах теории графов. Граф  $G_i = (N, \Gamma_N)$  задает особенности обеспечения заданий ресурсами РВС, здесь  $N$  – множество вершин графа (соответствуют узлам РВС), а  $\Gamma_N$  – множество дуг, которые поставлены в соответствие каналам перераспределения ресурсов в РВС.

Каждый элемент множества  $n_i \in N$  представляет собой вектор вида  $\langle n_i^{nbr}, n_i^{type}, n_i^r, n_i^q \rangle, i = \overline{1, |N|}$ , где  $n_i^{nbr}$  – номер  $i$ -ого узла РВС;  $n_i^{type}$  – тип  $i$ -ого узла (диверсный или нет);  $n_i^r$  – максимальный объем воспринимаемых ресурсов  $i$ -ым узлом;  $n_i^q$  – календарь загрузки  $i$ -ого узла. Граф  $G_j = (Z, \Gamma_Z)$  задает отношение связности между различными заданиями, здесь  $Z$  – множество вершин графа, которым поставлены в соответствие многоэтапные задания, представляющие собой иерархически упорядоченную совокупность информационно-вычислительных работ (ИВР), имеющих между собой функционально-логические связи. Множество дуг  $\Gamma_Z$  задает функционально-логические связи между заданиями. Каждый элемент множества  $z_j \in Z$  представляет собой вектор вида  $\langle z_j^{nbr}, z_j^{type}, z_j^{Tr}, z_j^{DATA}, z_j^q \rangle$ , где  $z_j^{nbr}$  – номер зада-

ния,  $z_j^{type}$  – тип  $j$ -го задания (связано с информационным обслуживанием объекта критического применения, или нет),  $z_j^{Tr}$  – трудоемкость  $j$ -ого задания,  $z_j^{DATA}$  – матрица исходных данных к заданию;  $z_j^q$  – календарь решения  $j$ -ого задания, учитывающий:  $d_j$  – момент поступления задания;  $t_{jr}$  – длительность обработки  $j$ -го задания на  $r$ -ом узле, а также директивный срок  $D_j$ , к которому желательно обработать данное задание;  $d_j, t_{jr}, D_j \in Z^+, j = \overline{1, |Z|}, r = \overline{1, |N|}, Z^+$  – множество целых положительных значений. В качестве упрощения [7] будем считать, что РВС является гомогенной, а длительность обработки заданий в узлах РВС не зависит от  $z_j^{nbr}$  и  $n_r^{nbr}$ , т.е.  $t_{jr} = 1, \forall j = \overline{1, |Z|}, \forall r = \overline{1, |N|}$ .

Отношения предшествования между заданиями, определенные при помощи дуг графа  $G_j = (Z, \Gamma_Z)$ , описываются матрицей смежности  $[\gamma_{i\mu}]$ , где  $i, \mu \in \overline{1, |Z|}$ .

В качестве примера отношения предшествования между заданиями по обработке медицинской информации рассмотрим базовую информационную структуру пакета (рис. 1), содержащего основные медицинские исследования для тяжелого пациента в нефрологическом отделении городской больницы №1 г. Севастополя, полученную на основании опроса группы квалифицированных экспертов, врачей-нефрологов. Расписание работы РВС определяется следующим образом:

$$s(\tau) = \{s_1(\tau), s_2(\tau), \dots, s_r(\tau)\} = z_j^q \Leftrightarrow n_i^q, \\ \forall r = \overline{1, |N|}, \forall j = \overline{1, |Z|} \forall i = \overline{1, |N|}, 0 \leq \tau < \infty$$

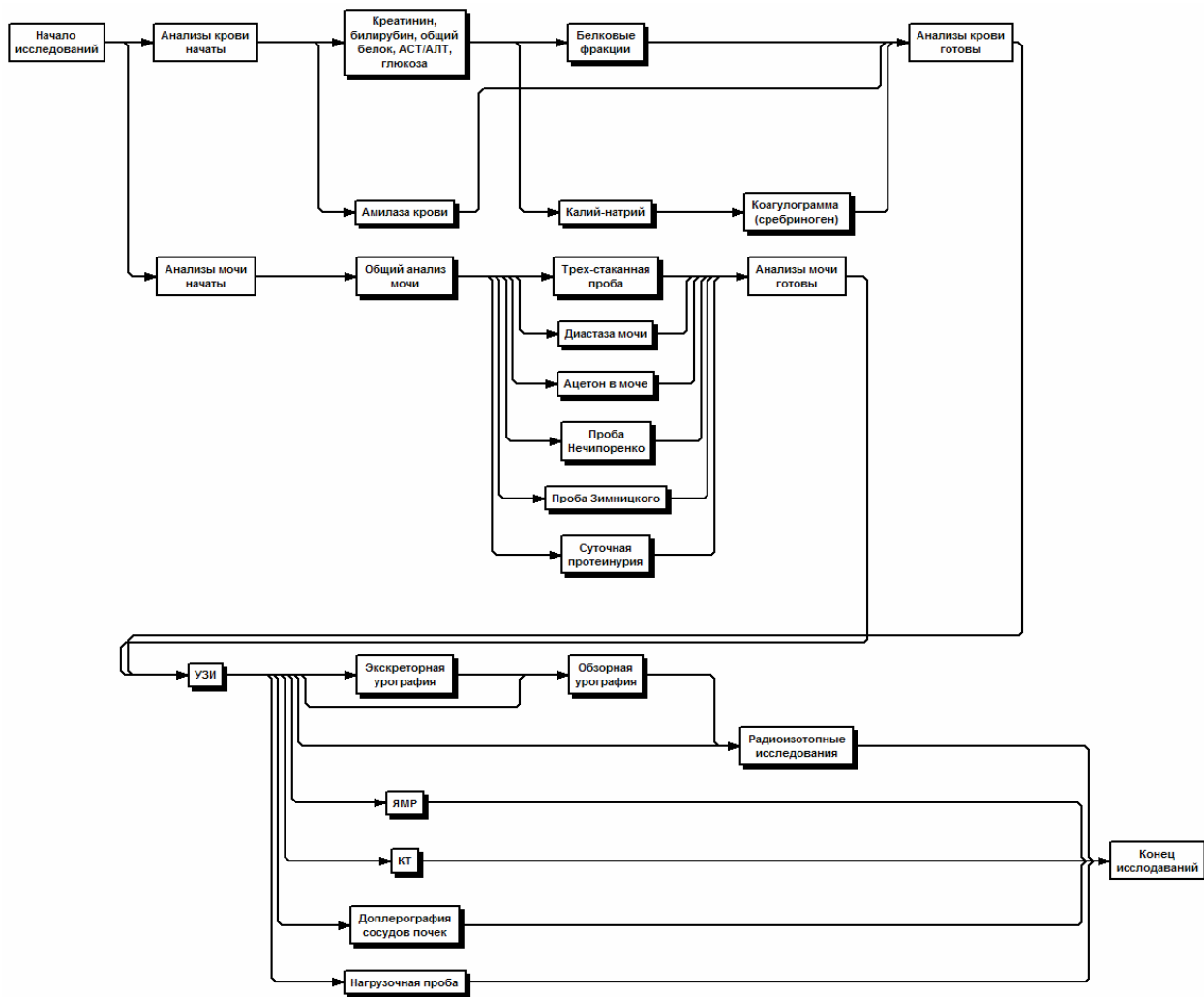


Рис. 1. Базовая информационная структура пакета задач пациента

Причем,  $s_q(\tau^*) = j$ , если  $j$ -ая задача в момент времени  $\tau^*$  обрабатывается в  $q$ -ом узле РВС.

Если  $s_q(\tau^*) = 0$ , то в момент времени  $\tau^*$   $q$ -ый узел РВС не задействован. В качестве решения ЗД бу-

дет выступать оптимальное расписание  $s(\tau)_{\text{опт}}$ , получаемое в ходе отображения графа  $G_j$  на  $G_1$ . Отображение  $G_j \xrightarrow{\varphi_k} G_1$  представляет собой  $k$ -ый вариант перераспределения ресурсов между узлами РВС в процессе обработки заданий; причем  $\varphi_k \in \Phi$ ,  $\Phi = \Phi_O \cup \Phi_{\text{Пр}}$ ,  $\Phi_O \cap \Phi_{\text{Пр}} = \emptyset$ . Где  $\Phi$  – множество всех допустимых вариантов отображений,  $\Phi_O$  – слабо эффективные отображения,  $\Phi_{\text{Пр}}$  – эффективные отображения. Таким образом,

$$s(\tau)_{\text{опт}} = \min_{s \in (\varphi_k \in V_{\text{Пр}})} \left( \sum_{j=1}^{|Z|} \alpha_j e_j(s) \right),$$

где  $e_j(s) = \max \{0, \overline{t}_j - D_j\}$  – запаздывание в обработке  $j$ -го задания;  $\overline{t}_j$  — момент завершения его обработки при расписании  $s$ ,  $\alpha_j$  – коэффициент важности  $j$ -го задания;  $k = \overline{1, |\Phi|}$ ,  $j = \overline{1, |Z|}$ .

### Нейросетевая интерпретация 3Д

Будем рассматривать случай построения нейроподобной сети с непрерывными состояниями нейронов, динамика которых описывается дифференциальной моделью вида [7]:

$$\frac{\partial U_\mu}{\partial t} = \frac{U_\mu}{a_\mu} + \sum_{v=1}^q T_{\mu v} V_v - I_\mu; \quad V_\mu = g(U_\mu);$$

$V_\mu(t_0) = g(U_\mu(t_0)) = V_\mu^{(0)}$ , где  $t$  — время;  $U_\mu$  — внутренний потенциал  $\mu$ -го нейрона;  $a_\mu$  — постоянная времени  $\mu$ -го нейрона;  $V_\mu$  – его выходной сигнал;  $T_{\mu v}$  — коэффициент синаптической связи входа  $\mu$ -ого нейрона с выходом  $v$ -ого нейрона;  $\mu, v \in \overline{1, q}$ ;  $g$  — непрерывная функция активации, например вида  $g(z) = \frac{1}{1 + e^{-\lambda z}}$ ,  $\lambda > 0$ .

Описанная сеть будет переходить из некоторого начального состояния  $\{V_\mu^{(0)}\}$ ,  $\mu \in \overline{1, q}$ , в устойчивое конечное состояние [7, 8], соответствующее одному из минимумов энергетической функции сети, например

$$E = -\frac{1}{2} \sum_{\mu=1}^q \sum_{v=1}^q T_{\mu v} V_\mu V_v + \sum_{\mu=1}^q I_\mu V_\mu \quad (1)$$

при этом выходные сигналы нейронов будут принимать значения, близкие к 0 или 1.

Параметрам 3Д поставим в соответствие параметры состояния сети  $\{V_\mu\}$ ,  $\mu \in \overline{1, q}$  [7]. Сконструируем для рассматриваемой 3Д такую скалярную

выпуклую функцию параметров, чтобы ее минимизация обеспечивала  $s(\tau)_{\text{опт}}$ . Приравняв ее к энергетической функции в канонической форме (1), можно будет определить искомые параметры  $\{T_{\mu v}\}$  и  $\{I_\mu\}$  сети, решающей исходную задачу диспетчеризации. В качестве нейросетевой интерпретации наиболее общего случая диспетчеризации в РВС синтезируемая сеть будет представлять собой трехмерный массив нейронов размерности  $|Z| \times |N| \times p$  (отображен на рис. 2).

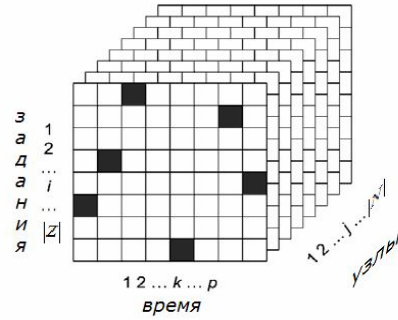


Рис. 2. Нейросетевая интерпретация 3Д

Состояние определяется множеством выходных сигналов  $V = \{V_{ijk} | i = \overline{1, |Z|}, j = \overline{1, |N|}, k = \overline{1, p}\}$ . Расписание  $s(x)$  отображается на трехмерный массив нейронов следующим образом:

$$(V_{ijk} = 1) \Leftrightarrow (S_j(\tau_k) = i), i = \overline{1, |Z|}, j = \overline{1, |N|}, k = \overline{1, p},$$

т.е. возведенное состояние  $ijk$ -го нейрона взаимно однозначно соответствует расписанию, согласно которому в  $k$ -ый момент времени  $i$ -е задание обрабатывается в  $j$ -ом узле. Значение параметра  $p$  можно определить, например пессимистической оценкой  $p = |Z|$ .

Для РВС ограничения одностадийности требуют, чтобы в каждой строке и в каждом столбце любого моментного  $k$ -сечения сети, представляющего собой матрицу  $[V_{ij(k)}]$ ,  $i = \overline{1, |Z|}, j = \overline{1, |N|}, k$  — фиксировано, было бы не более одного возбужденного нейрона. Кроме того, в каждом столбце любого узлового  $j$ -сечения нейронной сети, представляющего собой матрицу нейронов  $[V_{i(j)k}]$ ,  $i = \overline{1, |Z|}, k = \overline{1, p}$ ,  $j$  — фиксировано, также должно быть не более одного возбужденного нейрона. Наконец, в каждом столбце любого запросного  $i$ -сечения, представляющего собой матрицу нейронов  $[V_{(i)jk}]$ ,  $j = \overline{1, |N|}, k \in \overline{1, p}$ ,  $i$  — фиксировано, должно быть не более одного возбужденного нейрона [7].

Нейросетевую интерпретацию ограничения предшествования рассмотрим на примере пары требований, связанных отношением порядка  $i \rightarrow \mu$  (рис. 3).

Это означает, что матрица запросного  $i$ -сечения не должна содержать возбужденные нейроны с временной координатой  $k$ , большей временной координаты любого возбужденного нейрона в матрице запросного  $j$ -сечения. Ограничение предшествования для рассматриваемой пары заданий можно записать следующим образом [7]:

$$\max_{j \in \{1, \dots, N\}} \{k | V_{jik} = 1\} < \min_{j \in \{1, \dots, N\}} \{k | V_{\mu jk} = 1\}.$$

С учетом интерпретации временного запаздывания, описанной в [7], оптимальное расписание  $s(\tau)_{\text{opt}}$  будет вычисляться следующим образом:

$$s(\tau)_{\text{opt}} = \min_{s \in (\Phi_k \in V_{\Pi p})} \left( \sum_{i=1}^{|Z|} \alpha_i \left( \sum_{j=1}^{|N|} \Psi_0 \right) \right)$$

$$\Psi_0 = \left( \sum_{k=D_i}^p V_{ijk} + \sum_{k=D_i}^p \sum_{\chi=D_i}^k V_{ijk} (1 - V_{ij\chi}) \right)$$

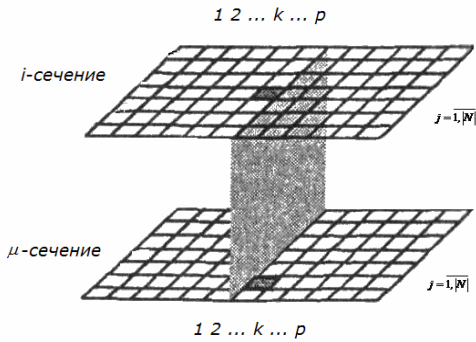


Рис. 3. Нейросетевая интерпретация ограничения предшествования

### Конструирование энергетической функции сети

Энергетическая функция сети равна сумме функции, принимающей минимальные значения на состояниях сети, удовлетворяющих рассмотренным ограничениям задачи и целевому функционалу

$$E^\circ(V) = E_O^\circ(V) + E_D^\circ(V) + E_{\Pi}^\circ(V) + E_C^\circ(V) + E_B^\circ(V),$$

где  $E_O^\circ(V)$  отвечает за выполнение ограничений одностадийности,  $E_D^\circ(V)$  отвечает за вычислительную устойчивость,  $E_{\Pi}^\circ(V)$  обеспечивает выполнение ограничений предшествования,  $E_C^\circ(V)$  соответствует минимизации функции стоимости,  $E_B^\circ(V)$  налагает ограничение бинарности.

$$E_O^\circ(V) = \frac{C_0}{2} \left[ \sum_{k=1}^p \Psi_1 + \Psi_2 \right]$$

$$\Psi_1 = \left[ \sum_{j=1}^{|N|} \sum_{i=1}^{|Z|} \sum_{\substack{\mu=1 \\ \mu \neq i}}^{|Z|} V_{ijk} V_{\mu jk} + \sum_{i=1}^{|Z|} \sum_{j=1}^{|N|} \sum_{\substack{v=1 \\ v \neq j}}^p V_{ijk} V_{ivk} \right]$$

$$\Psi_2 = \sum_{i=1}^{|Z|} \left[ \sum_{j=1}^{|N|} \sum_{k=1}^p V_{ijk} - 1 \right]^2,$$

где  $C_0 > 0$  – постоянный множитель.

$$E_D^\circ(V) = \frac{C_d}{2} \sum_{j=1}^{|Z|} \sum_{i=1}^{|N|} D_{ji} \left[ \sum_{j=1}^{|N|} \sum_{k=1}^p \sum_{v=1}^{|Z|} V_{ijk} V_{\mu vk} \right]$$

где  $C_d > 0$  – постоянный множитель;

$D_{ji} = z_j^{\text{type}} \Leftrightarrow n_i^{\text{type}} = w$ , где  $w \in Z^+$  и определяет приоритет назначения  $i$ -ого узла для обработки  $j$ -ого задания с точки зрения вычислительной устойчивости. Другими словами,  $D_{ji}$  учитывает, чтобы максимальное число диверсных узлов должно быть занято обработкой критических задач.

$$E_{\Pi}^\circ(V) = \frac{C_n}{2} \sum_{i=1}^{|Z|} \sum_{\mu=1}^{|Z|} \gamma_{i\mu} \left[ \sum_{j=1}^{|N|} \sum_{k=1}^p \sum_{v=1}^{|N|} \sum_{\chi=1}^k V_{ijk} V_{\mu v \chi} \right]$$

где  $C_n > 0$  – постоянный множитель;  $\gamma_{i\mu}$  – элемент матрицы смежности, задающей отношение порядка.

$$E_C^\circ(V) = C_\varphi \sum_{i=1}^{|Z|} \alpha_i \left( \sum_{j=1}^{|N|} \left( \sum_{k=D_i}^p V_{ijk} + \Psi_3 \right) \right)$$

$$\Psi_3 = \sum_{k=D_i}^p \sum_{\chi=D_i}^k V_{ijk} (1 - V_{ij\chi}),$$

где  $C_\varphi > 0$  – постоянный множитель.

$$E_B^\circ(V) = \frac{C_\beta}{2} \sum_{i=1}^{|Z|} \sum_{j=1}^{|N|} \sum_{k=1}^p \sigma_{ijk} V_{ijk} (1 - V_{ijk}),$$

где  $C_\beta > 0$  – постоянный множитель;  $\sigma_{ijk}$  – коэффициент, численно равный количеству ранее рассмотренных ограничений, накладываемых на  $ijk$ -нейрон. Энергетическая функция сети  $E(V, T, I)$ , заданная в канонической форме, описывается следующим образом:

$$E(V, T, I) = \Psi_4 - \frac{1}{2} \Psi_5, \quad \Psi_4 = \sum_{i=1}^{|Z|} \sum_{j=1}^{|N|} \sum_{k=1}^p V_{ijk} I_{ijk}$$

$$\Psi_5 = \sum_{i=1}^{|Z|} \sum_{j=1}^{|N|} \sum_{k=1}^p \sum_{\mu=1}^{|Z|} \sum_{v=1}^{|N|} \sum_{\chi=1}^p T_{ijk\mu v \chi} V_{ijk} V_{\mu v \chi}$$

Методика определения параметров искомой нейронной сети аналогична [7].

## Заключення

Разработана и представлена методика синтеза нейронных сетей для построения оперативных сервисов по диспетчеризации заданий в распределенных вычислительных системах. Область практического использования описанной методики ограничена пока достаточно узким кругом тех задач организации диспетчеризации, в которых время обработки задания в узле РВС инвариантно их номерам, а целевая функция представляет собой суммарное запаздывание.

При построении энергетической функции нейронной сети были учтены ограничения предшествования, бинарности, одностадийности и гарантоспособности (вычислительной устойчивости). Необходимо отметить, что в статье не затронуты вопросы качества нейросетевого решения, а также связанные с этим проблемы достижения глобального минимума энергетической функции и вычисления постоянных множителей в конструируемой энергетической функции. Данные вопросы представляют самостоятельный интерес и требуют отдельного рассмотрения. Предложенная методика предназначена, в первую очередь, для определения структуры и параметров искомого нейронной сети, решающей ЗД в РВС.

## Литература

1. Топорков В.В. Модели распределенных вычислений. / В.В. Топорков. – М.: Физматлит, 2004. – 320 с.

2. Волковой А.В. Многоверсионные системы и технологии для критических приложений. Лекционный материал / А.В. Волковой, И.В. Лысенко ; Под ред. Харченко В.С. – Министерство образования и науки Украины, Национальный аэрокосмический университет им. Н.Е. Жуковского «ХАИ», 2008. – 224 с.

3. Воронин Д.Ю. Системотехнический анализ процессов диспетчеризации в распределенных вычислительных системах / Д.Ю. Воронин // Вестник СевГТУ. Вып. 93: Информатика, электроника, связь: Сб. науч. тр. – Севастополь: СевНТУ, 2008. – С. 38-42.

4. Коган Д.И. Задача диспетчеризации: анализ вычислительной сложности и полиномиально разрешимые подклассы / Д.И. Коган, Ю.С. Федосенко // Дискретная математика. – 1996. – Т.8, № 3. – С. 135-147.

5. Скاتков А. В. Обеспечение гарантоспособности распределенной вычислительной системы с использованием диверсной диспетчеризации / А.В. Скатков, Д.Ю. Воронин // Радиоэлектронные и компьютерные системы. – №7 (34). – Х.: Изд-во ХАИ, 2008. – С. 19-24

6. Хайкин С. Нейронные сети: полный курс, 2-е издание / С. Хайкин. – М.: Издательский дом «Вильямс», 2006. – 1104 с.

7. Ефимов В.В. Синтез нейроподобных сетей для задач составления расписаний / В.В. Ефимов // Изв. вузов. Сер. Приборостроение, 1996. – № 1. – С. 36-44.

8. Меламед И.И. Нейронные сети и комбинаторная оптимизация / И.И. Меламед // Автоматика и телемеханика, 1994. – № 4. – С. 3-40.

Поступила в редакцию 14.01.2010

**Рецензент:** д-р техн. наук, проф. А.В. Дрозд, Одесский национальный технический университет, Одесса, Украина.

## ОПЕРАТИВНА ДИСПЕТЧЕРИЗАЦІЯ В РОЗПОДІЛЕНИХ ОБЧИСЛЮВАЛЬНИХ СИСТЕМАХ НА ОСНОВІ НЕЙРОМЕРЕЖЕВИХ МОДЕЛЕЙ

*Д.Ю. Воронін*

Розглянуто задачу побудови оперативних сервісів по диспетчеризації завдань в розподілених обчислювальних системах, орієнтованих на обробку медичної інформації. Використання нейромережєвих технологій дозволило впоратися з NP-повнотою розглянутій задачі і тим самим підвищити гарантоздатність всієї розподіленої обчислювальної системи в цілому.

**Ключові слова:** гарантоздатність, нейронні мережі, об'єкти критичного застосування, диспетчеризація, нейромережева оптимізація, NP-повнота, розподілена обчислювальна система.

## OPERATIONAL SCHEDULING IN THE DISTRIBUTED CALCULATING SYSTEMS BASED ON NEURAL MODELS

*D. Y. Voronin*

The problem of building operational services for the tasks scheduling in distributed calculating systems is considered. The using of neural network technology coped with the NP-completeness of the problem and increased the dependability across the distributed calculating system as a whole.

**Key words:** dependability, neural networks, critical applications, distribution computer system, scheduling, neural network optimization, NP-completeness.

**Воронин Дмитрий Юрьевич** – аспирант кафедры кибернетики и вычислительной техники Севастопольского национального технического университета, Севастополь, Украина, e-mail: dima@voronins.com.