

23. Sarah K. McCord, Linda Frederiksen, Nicole Campbell, An accessibility assessment of selected Web-based health information resources. *Library Hi Tech*, 2002, vol. 20, iss. 2, pp.188–198. Available at: <http://www.emeraldinsight.com/journals.htm?articleid=861320&show=abstract/> (accessed 23.07.2013).

24. Yogesan K., Kumar, S., Goldschmidt L., Cuadros J. (eds.) *Teleophthalmology*. Springer, 2006. 306 p.

УДК 004.8, 004.94

**АНАЛИЗ ПРОИЗВОДИТЕЛЬНОСТИ
РАСПРЕДЕЛЕННОЙ АСИНХРОННОЙ МОДЕЛИ
МНОГОСЛОЙНОГО ПЕРСЕПТРОНА,
ПОСТРОЕННОЙ НА БАЗЕ ПЛАТФОРМЫ ERLANG/OTP**

Статья поступила в редакцию 03.10.2013, в окончательном варианте 13.10.2013.

Плотникова Наталья Павловна, младший научный сотрудник, Мордовский государственный университет, 430005, Российская Федерация, г. Саранск, ул. Б. Хмельницкого, 39, e-mail: linsierra@yandex.ru

Федосин Сергей Алексеевич, кандидат технических наук, профессор, Мордовский государственный университет им. Н. П. Огарева, 430005, Российская Федерация, г. Саранск, ул. Б. Хмельницкого, 39, e-mail: fedosinsa@mrsu.ru

Казakov Виталий Гаясович, научный сотрудник, Мордовский государственный университет, 430005, Российская Федерация, г. Саранск, ул. Б. Хмельницкого, 39, e-mail: vgkazakov@gmail.com

В статье приводятся результаты исследования двух подходов к асинхронному распределенному моделированию многослойного перцептрона: модели с диспетчеризацией блоков нейронов в пределах слоя и вычислительного узла; модели без диспетчеризации, использующей подход «один нейрон – один процесс». Для проведения вычислительных экспериментов оба подхода были реализованы с использованием сущностей и механизмов платформы Erlang/OTP. В качестве основного алгоритма обучения сети был применен алгоритм RPROP. Для указанных подходов проведено имитационное моделирование процесса обучения сети, замерено среднее время обучения на эпоху. По результатам экспериментов проведен сравнительный анализ указанных подходов, выделены достоинства и недостатки каждого из них. Итогом работы является демонстрация эффективности модели без диспетчеризации на кластерных распределенных вычислительных системах.

Ключевые слова: многослойный перцептрон, обучение нейронной сети, асинхронная распределенная модель, модель с диспетчеризацией, модель без диспетчеризации

**PERFORMANCE ANALYSIS OF DISTRIBUTED
ASYNCHRONOUS MULTILAYER PERCEPTRON MODEL
FOR ERLANG/OTP PLATFORM**

Plotnikova Natalya P., Junior Researcher, Ogarev Mordovia State University, 39 B. Khmelniyskiy St., Saransk, 430005, Russian Federation, e-mail: linsierra@yandex.ru

Fedosin Sergey A., Ph.D. (Engineering), Professor, Ogarev Mordovia State University, 39 B. Khmelniyskiy St., Saransk, 430005, Russian Federation, e-mail: fedosinsa@mrsu.ru

Kazakov Vitaliy G., Researcher, Ogarev Mordovia State University, 39 B. Khmelniyskiy St., Saransk, 430005, Russian Federation, e-mail: vgkazakov@gmail.com

This article presents research results for two multilayer perceptron asynchronous distributed modeling approaches: model with neurons groups dispatching and model based on “one neuron – one process”

principle without dispatching. Implementation of these approaches uses entities and mechanisms of Erlang/OTP platform. Multilayer network models in this research use RPROP training algorithm. Research work includes training process modeling, average training speed measuring and comparison of discussed approaches for different neural networks. The article analyzes advantages and disadvantages of each model type for distributed cluster experimental systems with high speed interconnection bus. Main result of this research is proof of systems without dispatching efficiency.

Keywords: multilayer perceptron, neural network training, asynchronous distributed model, model with dispatching, model without dispatching

Для решения целого ряда задач проектируются искусственные нейронные сети (ИНС), которые могут достигать огромных размеров – количество нейронов в них измеряется тысячами, а то и миллионами [5]. К таким задачам относятся, например, долгосрочные прогнозы климатических изменений, геологических катаклизмов – землетрясений, извержений вулканов, столкновений тектонических плит, прогнозов цунами и разрушительных ураганов [3].

На современном этапе развития вычислительной техники одним из основных способов ускорения вычислений и достижения требуемой производительности является создание вычислительных систем, обеспечивающих одновременную реализацию нескольких вычислительных процессов, связанных с решением одной задачи, на разных процессорных элементах. Целью данной статьи является рассмотрение вопросов построения параллельной модели такого класса искусственных нейронных сетей, как многослойный перцептрон. Применительно к этому виду ИНС существует несколько подходов к распараллеливанию работы сети. Все они исходят из того, что наиболее ресурсоемким является не столько процесс функционирования сети, сколько процесс ее обучения.

Согласно наиболее популярному подходу элементарные вычислительные узлы сети (нейроны) организуются в группы. Шаги алгоритма обучения выполняются для каждой группы нейронов на отдельном вычислительном узле. Считается, что такой подход имеет преимущества перед распределением по вычислительным узлам отдельных нейронов, поскольку вычислительная нагрузка элементарного нейрона мала по сравнению с затратами на обмен сообщениями при синхронизации вычислений.

Однако асинхронная модель многослойной нейронной сети, построенная с использованием инструментов платформы Erlang/OTP в соответствии с принципом «один нейрон – один процесс», показала хорошие экспериментальные результаты на системах с общей памятью [2]. Поэтому для проведения полноценного сравнительного анализа в дополнение к разработанному ранее прототипу, работающему по указанному принципу, был создан прототип асинхронной распределенной модели многослойного перцептрона, логическая структура которого включает дополнительную сущность – диспетчер, контролирующей обмен сообщениями между группами нейронов.

Логические схемы взаимодействия отдельных сущностей систем с диспетчеризацией и без нее представлены на рис. 1 и 2 соответственно.

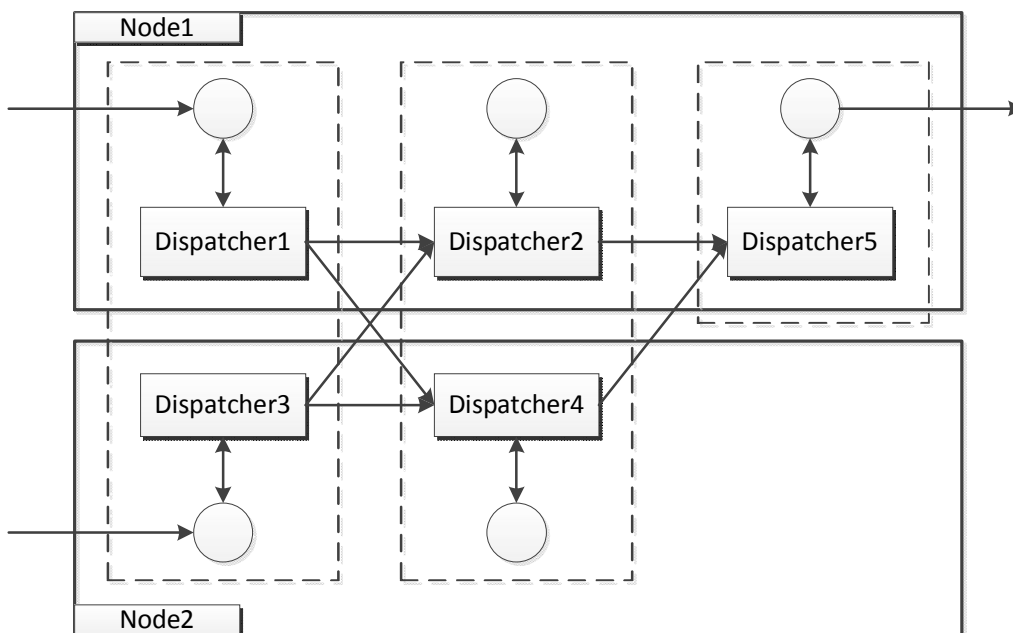


Рис. 1. Межнодное взаимодействие внутри нейронной сети посредством диспетчеров, запускаемых для групп нейронов слоя на каждой отдельной ноде (пунктирные линии обозначают слои нейронной сети, круги – нейроны)

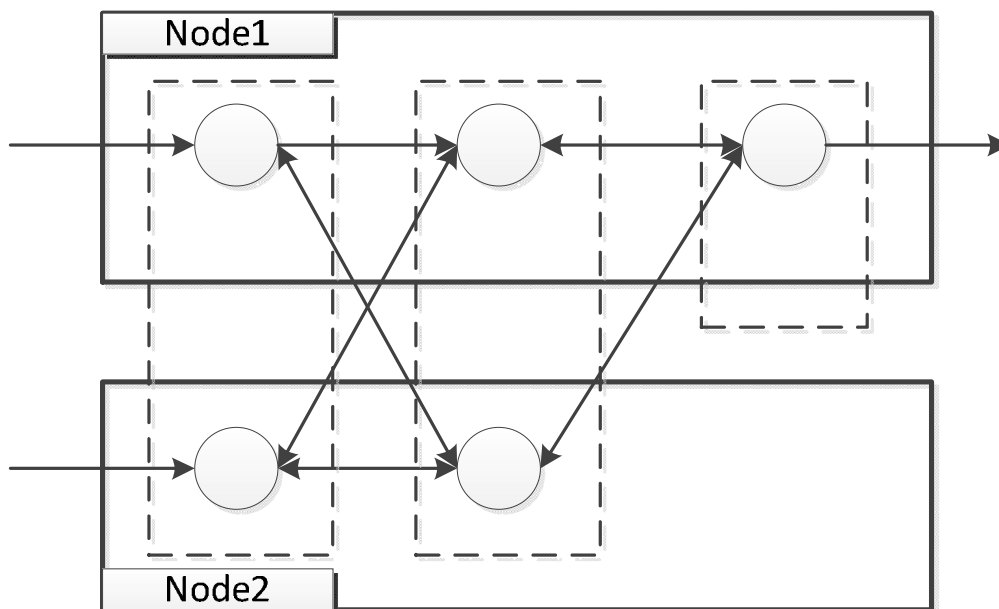


Рис. 2. Межнодное взаимодействие внутри нейронной сети без диспетчеризации (пунктирные линии обозначают слои нейронной сети, круги – нейроны)

Сущности «нейрон» и «диспетчер» представлены в прототипах в виде OTP behavior `gen_server`, реализующего функции обобщенного сервера, а именно, возможность отправки/обработки сообщений и изменения/сохранения текущего состояния процесса [4].

Node1 и Node2 на вышеприведенных рисунках – это стандартные ноды Erlang, внутри которых запускаются все процессы прототипов и осуществляется взаимодействие между ними [6].

Тестирование прототипов проводилось на базе вычислительного кластера IBM. Характеристики оборудования, а также характеристики систем, на которых проводились эксперименты, приведены в табл. 1 и 2.

Таблица 1

Характеристики кластера IBM

Количество лезвий	3
Процессор	Количество процессоров: 2 процессора на лезвие Количество ядер: 4 ядра на процессор
Оперативная память	16 Гб на лезвие
Скорость шины обмена данными между лезвиями	16 Гбит/с
Высокоуровневое программное обеспечение	VMware vCenter Server Version 5.0.0 Build 455964
Количество виртуальных машин, участвующих в эксперименте	15 (из расчета – 5 виртуальных машин на лезвие)

Таблица 2

Характеристики виртуальных машин

Оперативная память	2 Гб
Процессор	1 процессор, 1 ядро
Операционная система	CentOS 6.4 Linux 2.6.32 – 279.el6.x86_64
Версия программного обеспечения	Erlang R15B01 (erts – 5.9.1)
Командная строка запуска теста	erl -smp -rsh ssh -pa ebin/ -sname animalia -s start start mlp

Нейронная сеть, используемая в эксперименте, состояла из трех слоев. В первом (входном) и втором (скрытом) слоях содержится N нейронов. В выходном слое содержится 1 нейрон. Эксперимент проводился для различных значений N в диапазоне от 100 до 1000. В результирующих графиках приводится значение общего количества нейронов ($2*N + 1$).

Подобная архитектура обусловлена в первую очередь двумя положениями теории нейронных сетей:

- 1) трехслойный персептрон с достаточным количеством нейронов может решить любую задачу, которая в общем случае решается произвольным многослойным персептроном [1];
- 2) количество нейронов скрытого слоя не должно быть меньше количества входов сети.

На вход сети подаются бинарные данные. Обучающая выборка генерируется случайным образом. Размер обучающей выборки – 10 пар. Используемый алгоритм обучения – RPROP [7].

Эксперимент состоит в следующем. Прототипы последовательно запускаются на различном количестве Erlang-нод (от 1 до 15). При этом каждая нода запускается на отдельной виртуальной машине.

На рис. 3 приведены результаты распределенного тестирования для модели с диспетчеризацией. Как видно, данная модель демонстрирует стабильный рост производительности с увеличением количества задействованных нод. Особенно существенно уменьшение времени, затрачиваемого на обучение больших сетей с количеством нейронов от 1201 и более.

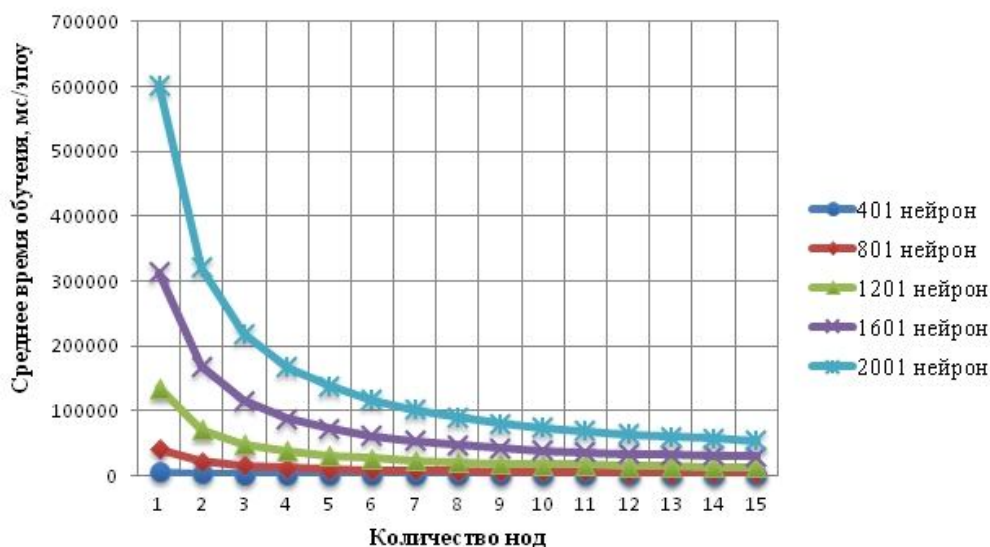


Рис. 3. Результаты распределенного тестирования модели многослойного персептрона с диспетчеризацией на виртуальных рабочих станциях кластера

На рис. 4 приведены результаты расчетов относительного ускорения, даваемого моделью с диспетчеризацией. Величина этого ускорения рассчитывается как отношение среднего времени обучения сети при работе на заданном количестве нод к среднему времени обучения сети при работе на одной нод. Таким образом, график на рис. 4 демонстрирует эффективность использования различного количества вычислительных узлов (нод).

Модель с диспетчеризацией позволяет получить практически линейный рост ускорения по мере наращивании числа нод. При использовании 15 вычислительных узлов мы получаем 11-кратный прирост производительности для сетей больших размеров (от 801 нейрона и больше). Для относительно маленьких сетей (401 нейрон на рис. 4) наблюдается падение производительности. Это связано, в первую очередь, с тем, что затраты на обмен сообщениями между нейронами и диспетчерами слишком велики по сравнению с общим объемом вычислений. В связи с этим функциональные узлы большую часть времени простаивают в ожидании получения данных для обработки.

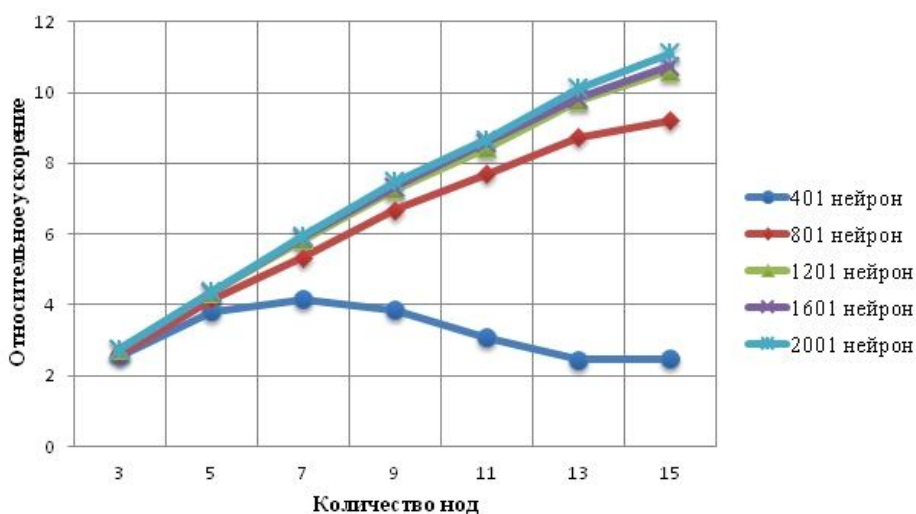


Рис. 4. Эффективность модели многослойного персептрона с диспетчеризацией при тестировании на виртуальных рабочих станциях кластера

На рис. 5 приведены результаты распределенного тестирования для модели без диспетчеризации. Для данной модели при использовании двух нод наблюдается увеличение времени обучения, а затем экспоненциальный спад, наиболее существенный для больших сетей (с количеством нейронов от 1601 и выше).

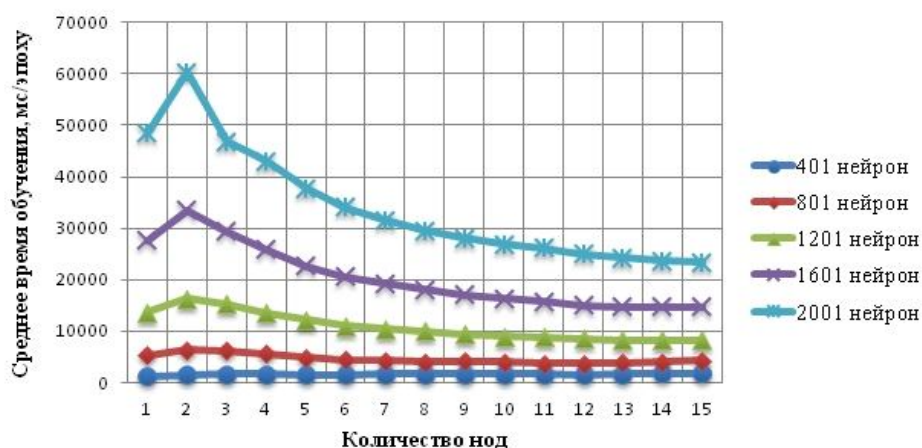


Рис. 5. Результаты распределенного тестирования модели многослойного перцептрона без диспетчеризации на виртуальных рабочих станциях кластера

На рис. 6 приведены графики зависимости относительного ускорения модели от количества задействованных нод для нейронных сетей различных размеров. Максимальное ускорение, которое демонстрирует модель без диспетчеризации, равно примерно 2.

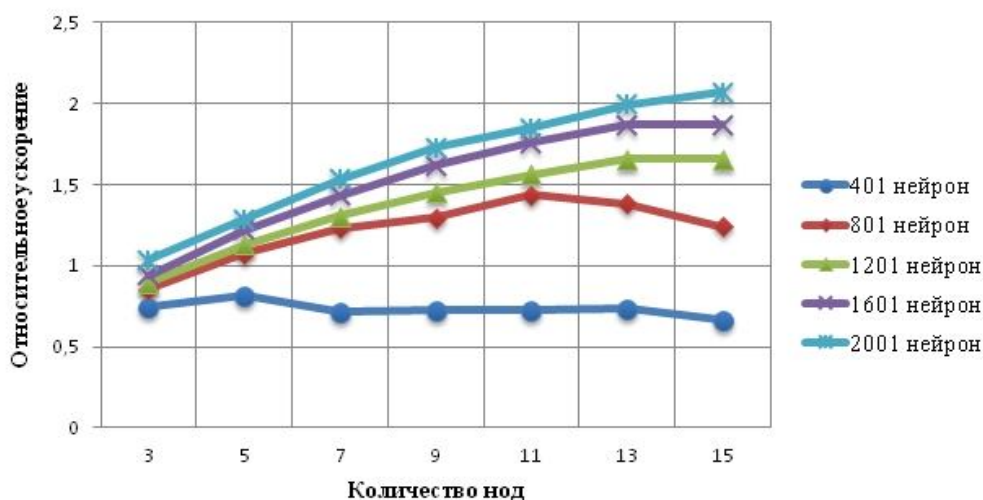


Рис. 6. Эффективность модели многослойного перцептрона без диспетчеризации при тестировании на виртуальных рабочих станциях кластера

На основе выполненных экспериментов был проведен сравнительный анализ рассматриваемых подходов. Результаты анализа приведены на рис. 7 и 8.

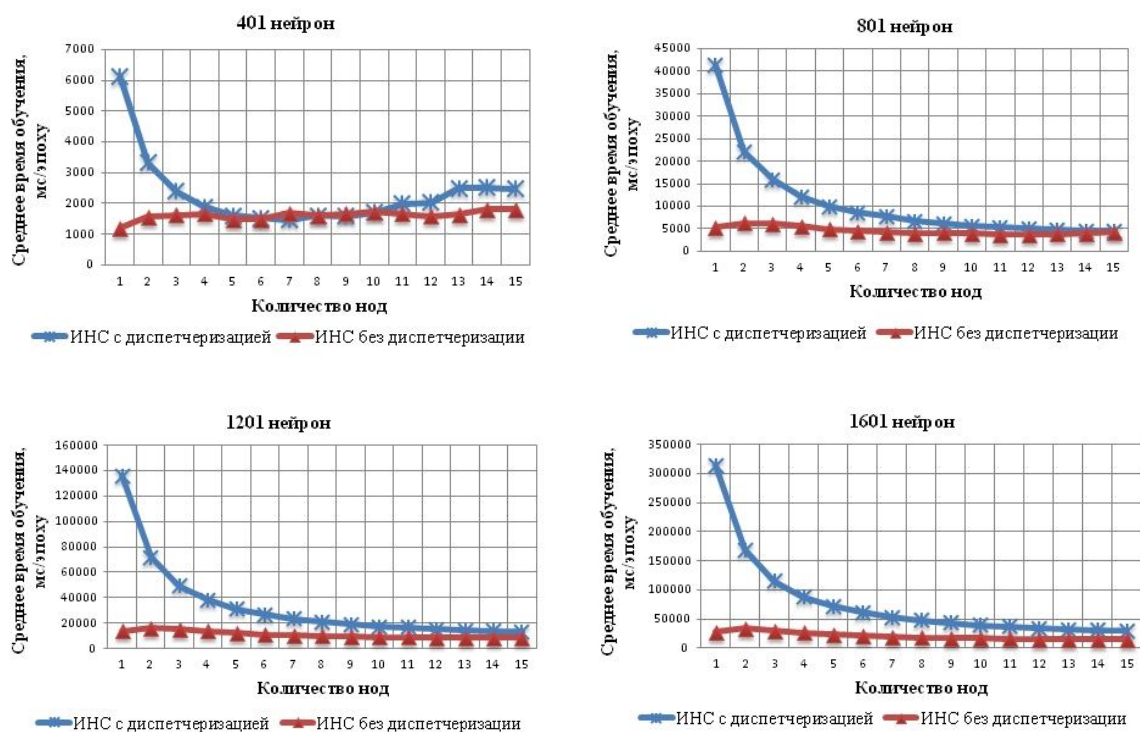


Рис. 7. Сравнение производительности моделей с диспетчеризацией и без для ИНС различных размеров (401, 801, 1201 и 1601 нейрон)

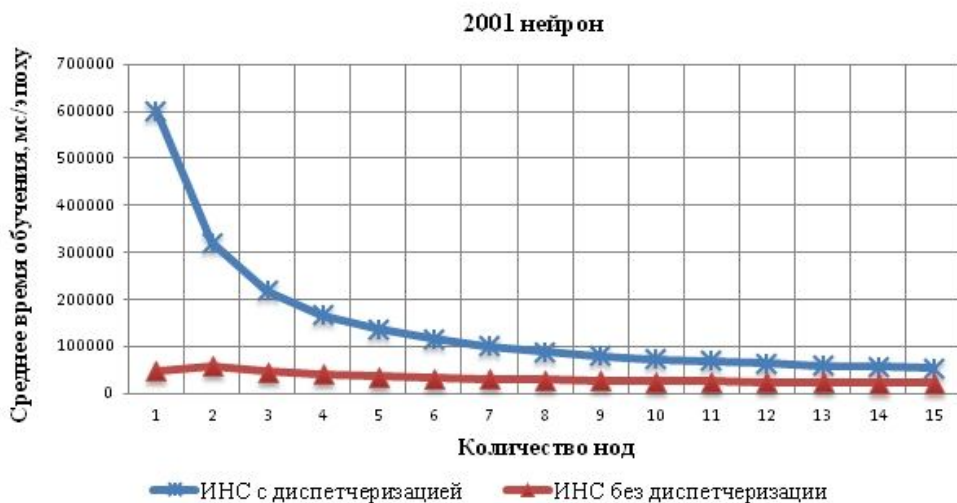


Рис. 8. Сравнение производительности моделей с диспетчеризацией и без для ИНС различных размеров (2001 нейрон)

Модель без диспетчеризации показывает лучшие по производительности результаты по сравнению с моделью с диспетчеризацией. Только в ряде случаев для небольших сетей (порядка 201 – 401 нейрона) модель без диспетчеризации работает медленней. На сетях средних размеров (801, 1001, 1201 нейрон) модель с диспетчеризацией на большом количестве нод (15) показывает примерно такие же результаты, как и модель без диспетчеризации.

На больших сетях (1601 нейрон и более) прототип модели без диспетчеризации, как правило, работает быстрее.

Таким образом, исходя из всего вышесказанного, можно сделать следующий **вывод**.

Несмотря на то, что относительные показатели эффективности модели с диспетчеризацией выше, при непосредственном сравнении производительности двух рассматриваемых подходов больший абсолютный прирост производительности дает ИНС без диспетчеризации.

Список литературы

1. Горбань А. Н. Обобщенная аппроксимационная теорема и вычислительные возможности нейронных сетей / А. Н. Горбань // Сибирский журнал вычислительной математики. – 1998. – Т. 1, № 1. – С. 12–24.
2. Плотникова Н. П. Построение параллельной модели многослойного персептрона / Н. П. Плотникова, В. Г. Казаков, В. В. Тесля, С. А. Федосин // Информационно-коммуникационные технологии. – 2012. – Т. 10, № 3. – С. 8–15.
3. Савочкин А. Е. Применение нейросетевого подхода при проектировании информационно-измерительных систем для определения степени повреждения технически сложных объектов / А. Е. Савочкин // Прикаспийский журнал: управление и высокие технологии. – 2013. – № 2. – С. 151–159.
4. Armstrong J. Making reliable distributed systems in the presence of software errors. Thesis / J. Armstrong. – November, 2003. – 283 p.
5. Bose B. K. Neural network applications in power electronics and motor drives – An introduction and perspective / B. K. Bose // IEEE Trans. Ind. Electron. – 2007. – Vol. 54, no. 1. – P. 14–33.
6. Erlang programming language. – Available at: <http://erlang.org/> (accessed 03.09.2013).
7. Riedmiller M. A direct adaptive method for faster backpropagation learning: The RPROP algorithm / M. Riedmiller, H. Braun // Proc. Int. Conf. Neural Networks. – San Francisco, CA, 1993. – P. 586–591.

References

1. Gorban A. N. Obobshchennaya approksimatsionnaya teorema i vychislitelnye vozmozhnosti neyronnykh setey [General approximation theorem and neural networks computational abilities]. *Sibirskiy zhurnal vychislitel'noy matematiki* [Siberian Journal of Computational Mathematics], 1998, vol. 1, no. 1, pp. 12–24.
2. Plotnikova N. P., Kazakov V. G., Teslya V. V., Fedosin S. A. Postroenie parallelnoy modeli mnogosloynogo perseptrona [Construction of multilayer perceptron parallel model]. *Informatsionno-kommunikatsionnye tekhnologii* [Information and Communication Technologies], 2012, vol. 10, no. 3, pp. 8–15.
3. Savochkin A. Ye. Primenenie neyrosetevogo podkhoda pri proektirovanii informatsionno-izmeritelnykh sistem dlya opredeleniya stepeni povrezhdeniya tekhnicheskikh slozhnykh obektov [Using neural networks approach in the design of information measurement systems for determining the extent of damage of technically complex objects]. *Prikaspiyskiy zhurnal: upravlenie i vysokie tekhnologii* [Caspian Journal: Management and High Technologies], 2013, no. 2, pp. 151–159.
4. Armstrong J. Making reliable distributed systems in the presence of software errors. Thesis. November, 2003. 283 p.
5. Bose B. K. Neural network applications in power electronics and motor drives – An introduction and perspective. *IEEE Trans. Ind. Electron.*, 2007, vol. 54, no. 1, pp. 14–33.
6. Erlang programming language. Available at: <http://erlang.org/> (accessed 03.09.2013).
7. Riedmiller M., Braun H. A direct adaptive method for faster backpropagation learning: The RPROP algorithm. *Proc. Int. Conf. Neural Networks*. San Francisco, CA, 1993, pp. 586–591.