

УДК 004.02

## К ВОПРОСУ ПОСТРОЕНИЯ ИНФОРМАЦИОННОЙ СИСТЕМЫ ПОДБОРА ОПТИМАЛЬНЫХ ХАРАКТЕРИСТИК ИСКУССТВЕННЫХ НЕЙРОННЫХ СЕТЕЙ ДЛЯ ЗАДАЧ ПРОГНОЗИРОВАНИЯ В ЭЛЕКТРОЭНЕРГЕТИКЕ

Дьяченко Р.А.

*ФГБОУ ВПО Кубанский государственный технологический университет, Краснодар, Россия (350072, г. Краснодар, ул. Московская, 2А), [emessage@rambler.ru](mailto:emessage@rambler.ru)*

---

В статье рассмотрены вопросы, связанные с созданием информационного обеспечения для подбора оптимальных характеристик искусственных нейронных сетей. Указанная задача решается в рамках решения задачи создания системы прогнозирования параметров объектов электроэнергетики. Автором проанализировано существующее в настоящее время программное обеспечение в области прогнозирования в электроэнергетике. В выделенных базовых моделях функционирования информационных систем прогнозирования в электроэнергетике, описанных на основе UML-диаграмм активности, выявлены достоинства и недостатки, тенденции развития. Далее автором задача подбора оптимальных характеристик искусственных нейронных сетей была формализована в виде задачи минимизации по критерию времени обучения на фиксированной выборке, решение которой для больших обучающих выборок занимает длительное время. Для решения проблемы длительного обучения автором предложена методика, основанная на концепции распределенных вычислений MapReduce.

---

Ключевые слова: прогнозирование, искусственные нейронные сети, оптимизация, функции высшего порядка.

## ON THE CONSTRUCTION OF INFORMATION SYSTEM OF SELECTION OF THE OPTIMAL CHARACTERISTICS OF ARTIFICIAL NEURAL NETWORKS FOR FORECASTING PROBLEMS IN THE POWER SECTOR

Djachenko R.A.

*Kuban State Technological University, Krasnodar, Russia (350072, Krasnodar, Moskovskaya street, 2A) [emessage@rambler.ru](mailto:emessage@rambler.ru)*

---

The questions related to the creation of information support for the selection of the optimum characteristics of artificial neural networks. This object is achieved in the solution of a system of projections of energy facilities. The author has analyzed the extant software for forecasting in power. In the selected base models of information systems in forecasting power, described in the UML-activity diagrams, identified strengths and weaknesses, trends. The author further the task of selecting the optimum characteristics of artificial neural networks has been formalized in the form of a problem by minimizing training time at a fixed sample solution is for large training samples takes a long time. To solve the problem of long training by the author proposed a method based on the concept of distributed computing MapReduce.

---

Key words: prediction, artificial neural networks, optimization, higher-order functions.

В настоящее время одним из самых эффективных и точных методов прогнозирования в электроэнергетике является инструментарий искусственных нейронных сетей (ИНС). Указанный метод может применяться для прогнозирования значений различных параметров объектов электроэнергетики, в том числе и потребления электрической энергии. Существует достаточно большое количество компаний, которые предлагают решения в области прогнозирования в электроэнергетической отрасли.

Компания ABB Group разработала систему долгосрочного прогнозирования электрической нагрузки Long-Range Spatial Load Forecasting System. Компания «АББ»

работает в области технологий для автоматизации, позволяющих промышленным предприятиям и энергетическим компаниям повышать свою производительность, снижая воздействие на окружающую среду. «АББ» является компанией в энергетике, способной выполнять прогнозирование нагрузки, используя методы моделирования на небольших площадях.

ЗАО НЦП «Физтех» - система прогнозирования AR-Control. Удобная и эффективная система построения прогнозов временного ряда в экономике, науке, технике, бизнес-планировании, прогнозирование котировок акций курсов валют.

ОАО «ВНИИЭ» - программный комплекс «Энергостат» («Энергостат-1.1») принят в опытную эксплуатацию в ОДУ Северного Кавказа и Северокавказском РДУ. Это программное обеспечение прогнозирования электропотребления с учётом изменения погодных условий. Сформированы и загружены в базу данных многолетние архивы фактических суточных графиков электропотребления, а также влияющих факторов. Реализована привязка комплекса к архивам «Суточной ведомости» для загрузки данных в темпе процесса. Осуществлена настройка математических моделей прогнозирования электропотребления по территориям диспетчерского управления ОДУ Северного Кавказа и Северокавказского РДУ с учётом температуры. Произведена установка программных модулей для обеспечения функционирования комплекса в ОДУ Северного Кавказа.

AleaSoft (Испания) - многоуровневая система прогнозирования потребления энергии для предприятий - современный западный программный продукт, специализирующийся на прогнозировании электропотребления, с модулями искусственного интеллекта.

Inron (США) - система краткосрочного прогнозирования электропотребления MetrixIDR - современный западный программный продукт, специализирующийся на прогнозировании электропотребления, с модулями искусственного интеллекта.

Деятельность предприятий энергетического комплекса связана с необходимостью прогнозировать потребление электроэнергии, производить расчет оптимальных тарифов и решать многие другие задачи анализа данных. При прогнозировании потребления электроэнергии необходимо учитывать погодные факторы, сезонные корректировки, корректировки на дни недели и т.д.

BIGroup Labs строит прогнозы на несколько дней вперед с ошибкой, не превышающей 3-5%.

Forecast Elektra - прогнозирование потребления электроэнергии для энергосбытовых компаний, предназначена для прогнозирования суточных графиков потребления

электроэнергии (ПЭ) энергосбытовой компании с упреждением до 7 суток и основана на нейросетевой модели.

BI ElektraProm - прогнозирование потребления электроэнергии для промышленных предприятий, предназначена для прогнозирования суточных графиков потребления электроэнергии (ПЭ) промышленных предприятий с упреждением до 7 суток и основана на нейросетевой модели.

StatSoft Russia, система «Энерго-Прогноз» - прогнозирование потребления электроэнергии промышленными предприятиями, отдельными объектами, регионами. Прогноз основывается на данных о потреблении электроэнергии в прошлом и строится с помощью оптимальных математических алгоритмов. Прогноз может быть скорректирован с помощью значений метеофакторов (включая температуру, скорость ветра, влажность), особенностей текущей нагрузки предприятия и т.д. Система прошла тестирование в реальных задачах, где был получен прогноз с точностью порядка 2-3%.

Система «Энерго-Прогноз» содержит самые современные технологии анализа данных и прогнозирования, визуализации и представления итоговых результатов, в том числе через Internet.

Ядром «Энерго-Прогноза» является система STATISTICA, что позволяет пользователю применять мощные аналитические методы в рамках одного пакета. Реализована возможность независимой оценки ошибки прогнозирования с помощью метода кросс-проверки. Модель, обеспечивающая наименьшую ошибку прогноза, может быть автоматически принята в качестве оперативной, и прогноз с её помощью представлен пользователю. Таким образом, возможна максимальная автоматизация прогнозирования.

Процесс получения прогнозных значений с использованием ИНС состоит из двух этапов:

- создание ИНС;
- эксплуатация ИНС с целью получения прогнозных значений.

UML-диаграмма активности [2], описывающая процесс создания ИНС [1; 3], представлена на рисунке 1.

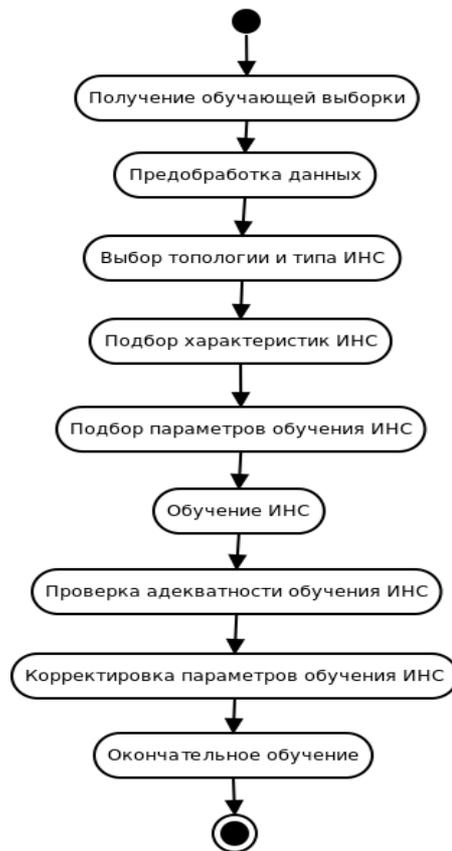


Рисунок 1 – UML-диаграмма активности процесса создания ИНС

Процесс эксплуатации ИНС с целью получения прогнозных значений может быть описан UML-диаграммой, представлен на рисунке 2.

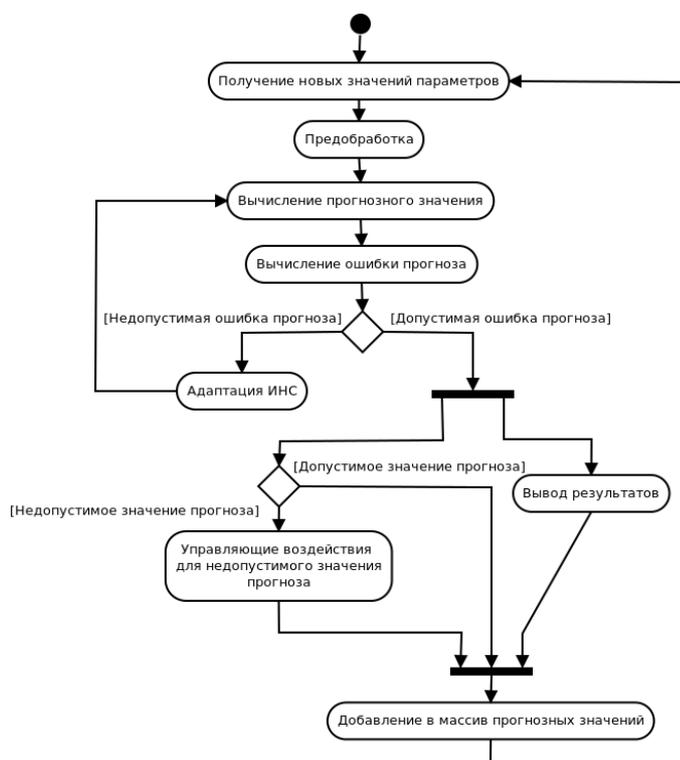


Рисунок 2 - UML-диаграмма активности процесса эксплуатации ИНС с возможной адаптацией

На рисунке 2 представлен вариант эксплуатации ИНС для получения прогнозных значений в режиме с контролем качества прогноза. В случае отсутствия выхода ошибки прогноза за допустимый предел системы эксплуатируется в обычном режиме. В противном случае проводится комплекс мероприятий по адаптации ИНС с последующим возвращением в нормальный режим функционирования. В контексте промышленного использования данный вариант является наиболее приемлемым, однако в зависимости от размера обучающих выборок процесс адаптации ИНС может занимать достаточно длительное время.

В этой связи проблема уменьшения времени адаптации сети является актуальной и востребованной.

Одним из самых важных (затратных по времени) этапов создания (или адаптации) ИНС является этап подбора характеристик ИНС.

**Целью** данного исследования является разработка методического (алгоритмического и программного) обеспечения оптимального подбора характеристик ИНС.

Основными **задачами** исследования являются:

- разработка алгоритма подбора оптимальных характеристик ИНС;
- адаптация модели распределенных вычислений MapReduce для решения задачи распределенного подбора характеристик ИНС с целью уменьшения затрачиваемого на расчет времени;

– UML-моделирование информационной системы, реализующей модель распределенных вычислений MapReduce.

После выбора общей структуры нужно экспериментально подобрать параметры сети.

В самом простом случае для сетей, подобных персептрон, подбираемыми характеристиками сети могут быть:

- число слоев сети  $P^{(Layers)}$ ;
- число нейронов в скрытых слоях  $P^{(Neurons)}$ ;

Дополнительными характеристиками также могут выступать:

- наличие или отсутствие обходных соединений;
- передаточные функции нейронов;
- др.

Исследования в области выбора оптимальных моделей ИНС по различным критериям рассмотрены в [3–5].

Далее будем рассматривать подбор оптимальных характеристик по величинам  $P^{(Layers)}$  и  $P^{(Neurons)}$ . К более сложным вариантам характеристик сети можно применить аналогичные рассуждения путем добавления дополнительных множеств и размерностей.

При выборе количества слоев и нейронов в них следует исходить из того, что способности сети к обобщению тем выше, чем больше суммарное число связей между нейронами. С другой стороны, число связей ограничено сверху количеством записей в обучающих данных.

Пусть даны натуральные числа  $P^{(minLayers)}$ ,  $P^{(maxLayers)}$ ,  $P^{(minNeurons)}$  и  $P^{(maxNeurons)}$ , такие что

$$P^{(minLayers)} \leq P^{(maxLayers)}$$

$$P^{(minNeurons)} \leq P^{(maxNeurons)}$$

$$D^{(Layers)} = \{d^{(layers)} \in N \mid P^{(minLayers)} \leq d^{(Layers)} \leq P^{(maxLayers)}\} \quad (1)$$

$$D^{(Neurons)} = \{d^{(Neurons)} \in N \mid P^{(minNeurons)} \leq d^{(Neurons)} \leq P^{(maxNeurons)}\} \quad (2)$$

Тогда Декартово произведение множеств  $D^{(Layers)}$  и  $D^{(Neurons)}$

$$D = D^{(Layers)} \times D^{(Neurons)} = \{(d^{(Layers)}, d^{(Neurons)})\} \quad (3)$$

назовем множеством характеристик ИНС с ограничениями  $P^{(minLayers)}$ ,  $P^{(maxLayers)}$ ,  $P^{(minNeurons)}$  и  $P^{(maxNeurons)}$ .

Элементом  $d$  множества  $D$  является одна характеристика (пара двух натуральных чисел), задающая для ИНС количество скрытых слоев и количество нейронов в слое.

Пусть также имеется некоторая обучающая выборка  $S=\{s_i\}$ . Тогда задачу оптимального подбора характеристик ИНС можно сформулировать в виде задачи минимизации функции

$$\min_{d \in D} E(d, S) \quad (4)$$

где  $E$  – функция, возвращающая ошибку обучения при конфигурации ИНС  $d$  на обучающей выборке  $S$ .

В общем случае задача 4 решается в три этапа.

1. Формирование множества  $D$  согласно (3).
2. Формирование множества (списка)

$$DE = \{(d_i, e_i) \mid d_i \in D, e_i = E(d_i, S)\}$$

(результат применения операции **map** (мэппинг, функция высшего порядка) над множеством (списком)  $D$ , на обучающей выборке  $S$  и функцией  $E$ .

3. Применение операции свертки **min** над списком  $DE$ , получившимся в результате выполнения предыдущего этапа. Результат выполнения операции пара  $(d_i, e_i)$  с наименьшим значением  $e_i$ .

Описанное решение может быть реализовано при помощи достаточно простых алгоритмов.

Основными проблемами при решении сформулированных выше задач являются:

- неэффективность распараллеливания большинства алгоритмов обучения;
- ограниченные вычислительные ресурсы отдельных информационных узлов;
- длительное время обучения ИНС.

В этой связи возможными решениями могут быть:

– разработка эффективных математических и алгоритмических методов обучения и самоорганизации ИНС;

– усовершенствование существующих алгоритмических методов оптимального подбора характеристик за счет увеличения вычислительных возможностей оборудования;

– разработка распределенных вычислительных систем (вычислительных кластеров), реализующих решения задач подбора оптимальных характеристик ИНС и задач обучения ИНС путем равномерного распределения вычислительных возможностей узлов (нодов) кластера.

Разработка эффективных математических и алгоритмических методов обучения и самоорганизации ИНС является наиболее перспективным направлением исследований, которые полностью не завершены.

Усовершенствование существующих алгоритмических методов оптимального подбора характеристик за счет увеличения вычислительных возможностей оборудования является достаточно эффективным решением в условиях отсутствия ограничений на стоимость создания и эксплуатации подобных систем. Однако современные реалии требуют рационального отношения ко всем финансовым издержкам, которые в конечном итоге оказывают непосредственное влияние на ценообразование в области оплаты за услуги энергетиков конечными потребителями. Поэтому решение подобных задач так называемым грубым методом наращивания вычислительных возможностей оборудования экономически необоснованно.

По упомянутым выше причинам разработка вычислительных кластеров, реализующих решения указанных выше задач путем равномерного распределения вычислительных возможностей узлов (нодов) кластера, в современных условиях, также является одним наиболее перспективных направлений.

В настоящее время одной из самых эффективных моделей распределенных вычислений, применяемых при создании распределенных информационных систем, является модель MapReduce [6].

Одной из причин успеха принципа MapReduce является то, что он разрабатывался как простая парадигма написания кода, пригодного для массового распараллеливания.

Преимущество MapReduce заключается в том, что он позволяет распределенно производить операции предварительной обработки и свертки. Операции предварительной обработки работают независимо друг от друга и могут производиться параллельно.

Типичная реализация алгоритма получает на вход 3 аргумента: исходную коллекцию, Map-функцию, Reduce-функцию и возвращает новую коллекцию данных после обработки. Алгоритм состоит из нескольких шагов.

Предлагаемое решение на основе концепции MapReduce при реализации распределенной информационной системы для подбора оптимальных характеристик искусственных нейронных сетей имеет DFD-диаграмму, представленную на рисунке 3.

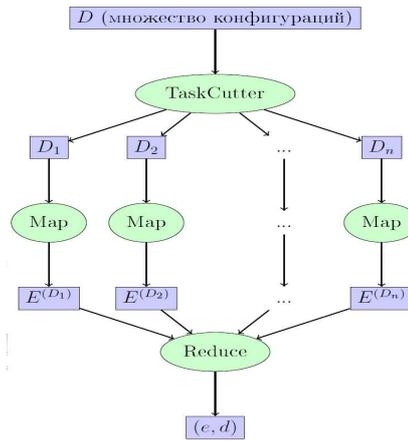


Рисунок 3 – DFD-диаграмма концептуальной модели распределенной информационной системы подбора оптимальных характеристик ИНС на основе модели MapReduce

Исходные данные и обозначения модели:

**D** - множество конфигураций ИНС;

**n** - количество частей, на которые будет делиться множество  $D$ ;

**TaskCutter** - процедура «нарезки» множества (списка)  $D$  на части;

$D_i$  - подмножество множества  $D$ ,  $i=1, \dots, n$ , получаемое методом *TaskCutter*;

**Map** - метод, вычисляющий функцию  $E$  над списком  $D_i$ ;

$E^{(D_i)}$  - множество пар вида

$$(e_k^i, d_k^i); d_k^i \in D_i; e_k^i = E(d_k^i, S); i = 1, 2, \dots, n; k = 1, 2, \dots, \dim(D_i);$$

**Reduce** - метод (свертка), вычисляющий функцию *min* над списками  $E^{(D_i)}$ ;

**(e,d)** - результат выполнения свертки, т.е. конфигурация  $d$ , на которой  $e = \min E(d, S)$ .

Диаграмма развертывания разработанной распределенной информационной системы, созданной на основе вышеописанного подхода, представлена на рисунке 4.

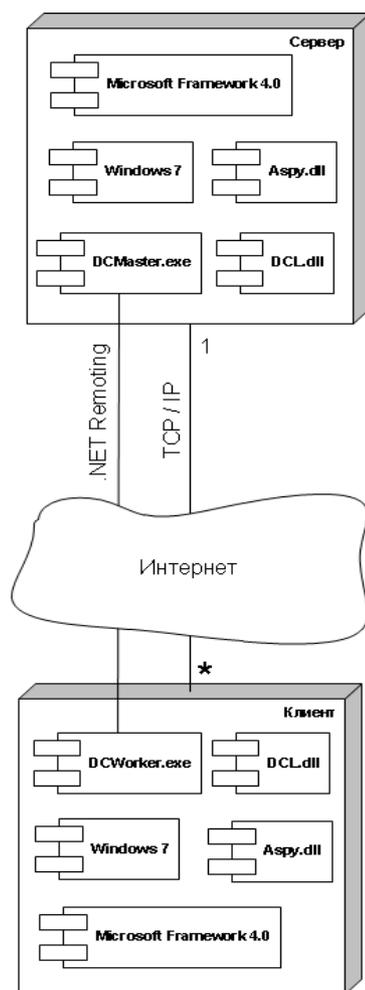


Рисунок 4 - Диаграмма развертывания

Согласно диаграмме имеются два типа физических устройств: сервер и клиент. Они соединены в сеть по протоколу TCP/IP.

На сервере установлены:

- операционная система Microsoft Windows;
- Microsoft Framework версии 4 или больше, необходим для работы программы AspyMaster;
- программа AspyMaster, является приложением, взаимодействует с AspyWorker через сеть Интернет (возможна работа в локальной сети), используя .Net Remoting.;
- в папке с программой находится библиотека Aspy.dll, в которой реализован функционал, необходимый для работы ИНС. Данная библиотека устанавливается вместе с программой AspyMaster;
- библиотека DCL.dll реализует поддержку технологии Microsoft .Net Remoting у программ AspyMaster и AspyWorker, то есть позволяет осуществлять взаимодействие данных

программ через сеть. Библиотека DCL.dll находится в папке с программой и устанавливается вместе с ней.

На клиенте установлены:

- операционная система Microsoft Windows;
- Microsoft Framework версии 4 или больше, необходим для работы программ AspyWorker;
- программа AspyWorker, является приложением, взаимодействует с AspyMaster через сеть Интернет (возможна работа в локальной сети), используя .Net Remoting;
- в папке с программой находится библиотека Aspy.dll, в которой реализован функционал, необходимый для работы ИНС. Данная библиотека устанавливается вместе с программой AspyWorker;
- библиотека DCL.dll реализует поддержку технологии Microsoft .Net Remoting у программ AspyMaster и AspyWorker, то есть позволяет осуществлять взаимодействие данных программ через сеть. Библиотека DCL.dll находится в папке с программой и устанавливается вместе с ней.

UML-диаграмма активности разработанной информационной системы представлена на рисунке 5.

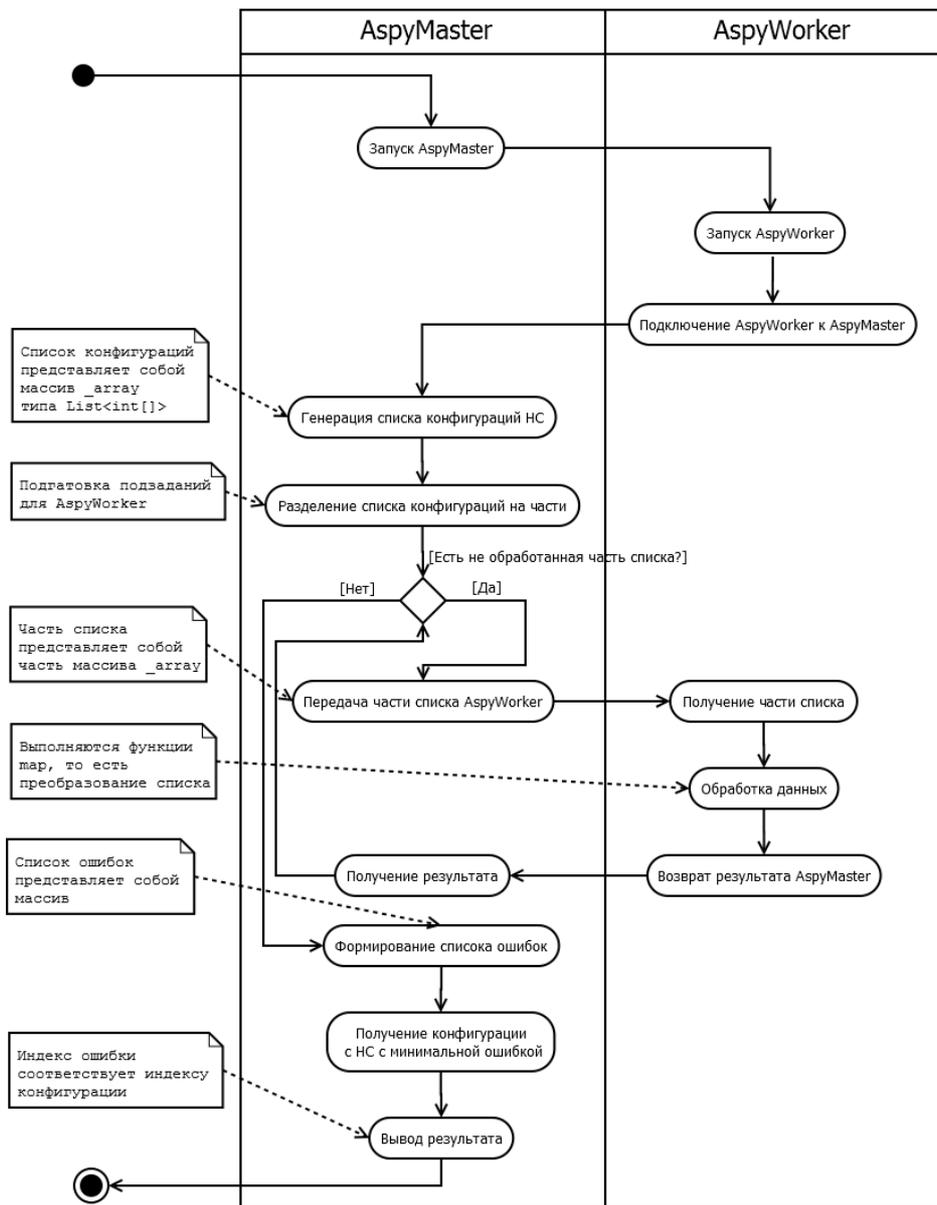


Рисунок 5 – UML-диаграмма активности

Диаграмма наглядно описывает динамику взаимодействия программ AspyMaster и AspyWorker.

После запуска программ AspyMaster на сервере и AspyWorker на клиентах клиентские программы AspyWorker пытаются найти в сети AspyMaster, после чего устанавливают соединение с ним. Далее AspyMaster генерирует список конфигураций HC, после этого делит список на части, то есть делит общее задание на подзадания. Следует отметить, что размер подзаданий задается в графическом интерфейсе программы AspyMaster.

После того как задание поделено на части, AspyMaster начинает отсылать каждому подключенному к нему AspyWorker подзадание и ждет от него решение в виде номера лучшей конфигурации и наименьшей ошибки. AspyWorker, получив подзадание (в виде части списка конфигураций HC), начинает обучать их данным. Если после того как

AspyMaster отослал задание, по какой-либо причине теряется связь с AspyWorker, то это же задание позже будет отослано другому AspyWorker.

Освободившиеся AspyWorker'ы получают новое подзадание от AspyMaster. Этот процесс продолжается, пока у AspyMaster'a есть не решенные подзадания. Получая результаты от AspyWorker'ов, AspyMaster формирует из них список ошибок. Позже, когда все подзадания решены, формируется окончательный список ошибок, из которого AspyMaster получит окончательный результат в виде конфигурации ИС с минимальной ошибкой.

Таким образом, используя мощности группы компьютеров, процесс поиска наилучшей конфигурации ИС происходит гораздо быстрее.

В результате проведения всех исследований было разработано методическое (алгоритмическое и программное) обеспечение оптимального подбора характеристик ИНС, которое включает:

- концептуальную модель распределенной информационной системы для подбора оптимальных характеристик искусственных нейронных сетей на основе модели распределенных вычислений MapReduce;

- UML-диаграмму вариантов использования распределенной информационной системы для подбора оптимальных характеристик ИНС на основе модели MapReduce;

- UML-диаграмму компонентов распределенной информационной системы для подбора оптимальных характеристик ИНС на основе модели MapReduce;

- UML-диаграмму развертывания распределенной информационной системы для подбора оптимальных характеристик ИНС на основе модели MapReduce;

- UML-диаграмму классов распределенной информационной системы для подбора оптимальных характеристик ИНС на основе модели MapReduce;

- UML-диаграмму активности распределенной информационной системы для подбора оптимальных характеристик ИНС на основе модели MapReduce.

Перечисленные UML-диаграммы в комплексе могут быть использованы при создании и развертывании информационных систем подбора оптимальных характеристик ИНС. Они также могут быть использованы для реинжиниринга уже существующих информационных систем прогнозирования в электроэнергетике, основанных на использовании ИНС.

## Список литературы

1. Бэин Д.В. Сравнительные модели прогнозирования электрической нагрузки / Д.В. Бэин, Е.Д. Фармер. — М. : Энергопромиздат, 1987. — С. 200.
2. Леоненков А.В. Самоучитель UML : учебное пособие. — СПб. : БХВ-Петербург, 2001. — С. 298.
3. Разработка алгоритма поиска оптимальной модели / Р.А. Дьяченко [и др.] // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. — 2012. — № 3 (77). — С. 378-387.
4. Солдатова О. Применение нейронных сетей для решения задач прогнозирования / О. Солдатова, В. Семенов // Исследовано в России : электр. науч. журнал. — 2006.
5. Фишер А.В. Организация хранения хронологических данных в базах данных систем мониторинга и прогнозирования / А.В. Фишер, Р.А. Дьяченко, И.С. Лоба // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. — 2012. - № 5 (79). - С. 485-495.
6. Mapreduce // [www.ru.wikipedia.org.wiki.mapreduce](http://www.ru.wikipedia.org.wiki.mapreduce) (дата обращения: 19.01.2013).

### Рецензенты:

Атрощенко Валерий Александрович, доктор технических наук, профессор кафедры информатики и вычислительной техники, декан факультета компьютерных технологий и автоматизированных систем, ФГБОУ «Кубанский государственный технологический университет», г. Краснодар.

Шевцов Юрий Дмитриевич, доктор технических наук, профессор кафедры информатики и вычислительной техники, ФГБОУ «Кубанский государственный технологический университет», г. Краснодар.