

Алгоритм определения вероятностно-временных характеристик устройств динамического приоритета нейронных сетей

А. В. Греков

Пермский военный институт войск национальной
гвардии Российской Федерации
temagre@ieee.org

А. А. Южаков

Филиал "ОЦО" АО "ОХК "УРАЛХИМ" в г. Перми
alexander.yuzhakov@uralchem.com

С. В. Березняков

АО «ОДК-СТАР», г. Пермь
berser22@mail.ru

Д. М. Поляхов

Санкт-Петербургский государственный электротехнический университет
«ЛЭТИ» им. В.И. Ульянова (Ленина)

Аннотация. Одной из задач, решаемых в процессе проектирования устройств динамического приоритета нейронных сетей, является задача определения вероятностно-временных характеристик устройства. Высокие аппаратные затраты и низкая загрузка оборудования являются основными недостатками устройств динамического приоритета, реализованных на основе классических структур. Устройства динамического приоритета потоковой динамической архитектуры характеризуются изменением структуры под влиянием входных информационных потоков.

Ключевые слова: нейронная сеть; устройство динамического приоритета; отказоустойчивость; алгоритм

I. ВВЕДЕНИЕ

Изначально возникновение и развитие теории формальных нейронных сетей вдохновлялось биологией, т.е. рассматривая сетевые конфигурации и алгоритмы, исследователи мыслили их в терминах организации мозговой деятельности. Со времен работ Кахаля нейроанатомами было обнаружено большое количество различных нейронов, отличающихся друг от друга размерами, числом дендритов и аксонов и их распределением [1–4].

В настоящее время принято считать, что почти во всех биологических видах нервная система представляет собой сеть нейронов, каждый из которых состоит из тела с одним или более афферентных (входных) отростков, или дендритов, и одним или более эфферентных (выходных) отростков, или аксонов. Аксоны разветвляются на тонкие волокна; оставаясь отделенными от поверхности

мембраны клетки или дендрита, на которых они оканчиваются, эти волокна могут иметь с ними контакт.

II. СЕТЕВАЯ СТРУКТУРА УСТРОЙСТВА ДИНАМИЧЕСКОГО ПРИОРИТЕТА ПОТОКОВОЙ ДИНАМИЧЕСКОЙ АРХИТЕКТУРЫ

Специфика сетевой структуры устройства динамического приоритета (УДП) проявляется в аппаратно-программной реализации:

- специально организованной операционной среды (нейронной сети) на нейроподобных элементах (НЭ);
- параллельного выполнения основных функций устройства;
- модульной системе управления.

В рассматриваемой структуре устройства динамического приоритета потоковой динамической архитектуры (УДП ПДА) операционная среда (нейронная сеть) организована на основе коммутационной аппаратуры (КА) и нейронных элементов (НЭ_{*i*}), реализующих одно место в очереди. Количество входных сообщений i ($x_i, i = 1...n$), а число обслуживающих устройств y_j равно m . УДП ПДА функционирует следующим образом.

В начальный момент времени по каналам настройки (H_μ) приходит, например, информация, что для обслуживающего устройства y_2 необходимо организовать УДП₂ с количеством входных сигналов 3 ($3 \in n$), причем для x_2, x_3 и x_7 с дисциплиной обслуживания FIFO. УУ и МУУ с помощью $\{z_v\}$ и КА образует из НЭ_{*i*} для y_2 УДП₂ с заданными параметрами (НЭ_{*i*} объединяются через КА в

некоторую структуру, обеспечивающую формирование очереди и ее продвижение в соответствии с заданным алгоритмом обслуживания и требуемым числом параметров – каналов сообщений). После чего указанное устройство (УДП₂) переходит в рабочий режим функционирования под управлением созданного УДП (совокупность МУУ), а УУ имеет возможность обслуживать запросы на формирование других УДП_j ($j=1...m, j \neq 2$). Например, следующим приходит запрос на формирование УДП₁ через канал настройки H_μ для y_1 с количеством входных сигналов 8 для $x_1...x_8$ и дисциплиной – абсолютный приоритет. УУ с помощью $\{z_v\}$ и КА образует для y_1 УДП₁ с заданными параметрами. После чего УДП₁ переходит в рабочий режим функционирования, а УУ имеет возможность обслуживать запросы на формирование других УДП_j ($j=1...m, j \neq 1, j \neq 2$). Данный процесс будет продолжаться до тех пор пока не будут созданы УДП_j для всех обслуживающих устройств (y_j) или пока не будет исчерпан весь ресурс НЭ_k.

III. УСТРОЙСТВО ДИНАМИЧЕСКОГО ПРИОРИТЕТА ПОТОКОВОЙ ДИНАМИЧЕСКОЙ АРХИТЕКТУРЫ КАК СИСТЕМА МАССОВОГО ОБСЛУЖИВАНИЯ

Модель должна описывать процесс распределения общесетевого ресурса НЭ при поступлении заявок на создание УДП_i с учетом определенных требований обслуживания для источников (группы источников). С позиций теории СМО функционирование УДП может быть описано следующим образом.

Входящим потоком для УДП служит управляющая информация $\{H_\mu\}$, поступающая на вход устройства управления. Заявка H_μ направлена на создание УДП_i из общего ресурса нейронов с определенной дисциплиной обслуживания для требуемого числа входных каналов. Информация, содержащаяся во входном потоке заявок (управляющая информация) является неоднородной и характеризуется следующими параметрами:

- количество обслуживаемых каналов (номера $x_i, i=1...n$);
- тип дисциплины обслуживания;
- номер обслуживающего устройства y_j ($j=1 \dots m$).

В зависимости от ее содержания в нейронной сети (НС) будет создаваться требуемое количество УДП_i.

Выходящий поток представляет собой сигналы управления $\{z_v\}$, которые реализуют в нейронной сети требуемое количество устройств (внутренний поток управляющих сигналов, обеспечивающих реконфигурацию).

Обслуживающим устройством в рассматриваемой модели СМО является нейронная сеть, ресурсом сети являются нейронные элементы (НЭ). В зависимости от типа дисциплины обслуживания (ДО) и числа коммутируемых каналов устройство управления выделяет

из нейронной сети для обслуживания заявок данного типа определенное количество обслуживающих приборов (нейронных элементов). Если остаток общего ресурса нейронов не позволяет реализовать очередное УДП_i, то возникает отказ в обслуживании данной заявки. Для рассматриваемого примера ДО FIFO требует $G \times (G - 1)$ НЭ, для ДО абсолютный приоритет необходимо G нейронов, а для ДО относительный приоритет $2 G$ НЭ, где G – количество каналов в j -том создаваемом УДП [5].

Время обслуживания заявки состоит в затратах времени на формирование необходимого УДП из требуемого числа нейронов в соответствии с запрашиваемой дисциплиной обслуживания и числа входных каналов. Время обслуживания заявок ($T_{об}$), в общем случае имеет произвольный закон распределения. Причина этого – устройство может обрабатывать заявки от источников, которые могут иметь абсолютно разные алгоритмы своего функционирования. Нейронная сеть характеризуется возможностью одновременной реализации нескольких УДП_i ($i \in 1, G$). Причем, каждое из реализуемых в сети УДП_i может обладать своей ДО и обслуживать свое количество каналов (G_i).

УДП будем рассматривать как СМО с потерями заявок. С поступлением очередной заявки H_j в отсутствие необходимого числа НЭ для ее обслуживания, последние покидают систему не обслуженными, т.е. получают отказ в обслуживании (новый УДП_j не создается).

IV. ОПРЕДЕЛЕНИЕ ВЕРОЯТНОСТНО-ВРЕМЕННЫХ ХАРАКТЕРИСТИК УСТРОЙСТВ ДИНАМИЧЕСКОГО ПРИОРИТЕТА НЕЙРОННЫХ СЕТЕЙ

Высокие аппаратные затраты и низкая загрузка оборудования является основными недостатками устройств динамического приоритета (УДП), реализованных на основе классических структур. Внешние переменные (H_μ) являются случайными функциями, поэтому задачи реализации (реконфигурирования) конкретных УДП будут возникать в случайные моменты времени. В остальных интервалах времени нейронные УДП будут не загружены, т.е. часть оборудования будет простаивать. Поэтому возникает возможность передать это оборудование на обслуживание других каналов, т.е. перейти к многоканальному обслуживанию на перестраиваемых архитектурах, т.е. использовать устройства динамического приоритета потоковой динамической архитектуры (УДП ПДА).

Возможно нахождение соответствующего значения вероятности отказа $P_{отк}$ при заданном значении числа нейронов сети. На практике, как правило, задаются значением $P_{отк}$ для которого и необходимо описать требуемое число нейронов сети [6–8]. Рассмотрим алгоритм указанной задачи.

1. Задаются исходные данные

$P_{отк} = P_{отк доп}$ (где $P_{отк доп}$ – допустимое значение вероятности отказа в обслуживании), λ, μ .

2. Задается произвольное значение числа нейронов в сети. Определяется тип моделируемой системы (разомкнутая или замкнутая). Для выбранного типа модели определяется $P_{отк}$.

3. Проверяется условие $P_{отк} \leq P_{отк\ доп}$. Если неравенство выполняется, то уменьшается число нейронов в сети и вычисляется новое значение $P_{отк}$. Если неравенство не выполняется, то увеличивается число нейронов сети и вычисляется новое значение $P_{отк}$.

4. Циклы итераций заканчиваются, когда уменьшение числа нейронов сети на «единицу» приводит к нарушению неравенства.

V. ВЫВОДЫ

Таким образом, в результате выполнения представленного алгоритма получаем минимальное значение объема оборудования нейронной сети. Сложность реализации предлагаемого алгоритма определяется выбранным методом определения $P_{отк}$. Выигрыш в объеме оборудования от использования УДП ПДА по сравнению с классическими системами УДП существенным образом зависит от загрузки оборудования. Чем ниже загрузка исходных УДП, тем больший эффект дает использование УДП ПДА, тем возможно большее сокращение объема оборудования нейронной сети УДП. Возможно применение алгоритма для построения отказоустойчивых цифровых устройств с применением функционально-полных толерантных элементов [9–10].

БЛАГОДАРНОСТИ

Авторы выражают глубокую признательность коллективу кафедры «Автоматика и телемеханика» Пермского национального исследовательского

политехнического университета и лично Заслуженному изобретателю РФ, профессору кафедры «Автоматика и телемеханика» доктору технических наук, профессору Сергею Феофентовичу Тюрину.

СПИСОК ЛИТЕРАТУРЫ

- [1] Масич Г. Применение нейрокомпьютерных систем [Электронный ресурс] / Российская академия наук / Уральское отделение. Институт механики сплошных сред. Пермь, 2004. Режим доступа: <http://neuralbench.ru>
- [2] Nazarenko M.A., Rumyantsev V.B. D-transformation and polynomial track recognition // Commun. JINR E11-96-337. Dubna, 1996. 337 p.
- [3] Уоссермен Ф. Нейрокомпьютерная техника / Ф. Уоссермен. М.: Мир, 1992. 184 с.
- [4] Короткий С. Нейронные сети: алгоритм обратного распространения [Электронный ресурс] / Лаборатория искусственного интеллекта. Электрон. дан. М., 2005. – Режим доступа: <http://lii.newmail.ru/NN/KOROTKY/N2/kog-nn2.htm>
- [5] Третьяков Н.В. Устройство динамического приоритета нейронной структуры / Н.В. Третьяков, Ал.Ал. Южаков, А.А. Южаков // Информационно-управляющие системы : сб. науч. тр. / Перм. гос. техн. ун-т. Пермь, 1999. С. 10–15.
- [6] Овчаров Л.А. Прикладные задачи массового обслуживания / Л.А. Овчаров. М.: Машиностроение, 1969. 324 с.
- [7] Назаров А.А. Анализ математической модели адаптивной терминальной измерительной системы / А.А. Назаров, А.А. Южаков, М.М. Нулик // Автоматика и телемеханика. 1993. № 11. С. 108–119.
- [8] Южаков А.А. Стохастические сети в проектировании технических систем / А.А. Южаков. Перм. гос. техн. ун-т. Пермь, 1999. 131 с.
- [9] Grekov A.V., Tyurin S.F. Fault Tolerant Logic Cell Fpga / Proceedings of the 2017 IEEE Russia Section Young Researchers in Electrical and Electronic Engineering Conference, ElConRus 2017 2017. Pp. 287-290.
- [10] Advanced FPGA Look Up Tables. Tyurin, S., Grekov, A., Vikhorev, R., Prokhorov, A. International Journal of Pure and Applied Mathematics. 2017. 117(22 Special Issue), Pp. 143-146.