

Базовые принципы общей теории нейронной сети с внутренней обратной связью.

Д. Песчанский

Progress Inc., West Bloomfield, MI, US, pro@p-progress.com

Аннотация

В данной работе предлагается и обосновывается новая модель формального нейрона, аналоговые механизмы обучения нейрона и новая модель биологической обратной связи. Положения работы подтверждаются нейробиологическими данными, опубликованными другими авторами, а также данными эксперимента, полученными путем компьютерного моделирования. Демонстрируются ключевые качественные и количественные отличия предлагаемой модели нейросети от принятой сегодня концепции. Новая концепция отражает механизмы формирования памяти. Описываемая модель предоставляет возможность протянуть «мостик» между микро-уровнем молекулярных процессов, протекающих в отдельном нейроне, и макро-уровнем хранения и обработки информации в мозге. Что более важно, появляется реальная возможность искусственно конструировать и моделировать процессы, протекающие в мозге, а также строить искусственные нейронные сети, обучающиеся в режиме реального времени и неограниченные по структуре, количеству и сложности связей. Предлагаемая модель легко реализуется, как в виде виртуальных эмуляций, так и в виде цифровых и аналоговых схем.

Ключевые слова: Нейронная сеть, Алгоритм обучения, Синапс, Память, Обратная связь, Аналоговое обучение

1 Введение

Несмотря на большое разнообразие разработанных на сегодняшний день моделей нейросетей и алгоритмов их обучения [1], в основе любой из них (с небольшими вариациями) лежит концепция одного и того же формального нейрона (формальный нейрон (ФН) Мак-Каллока и Питтса) (рис. 1) [2].

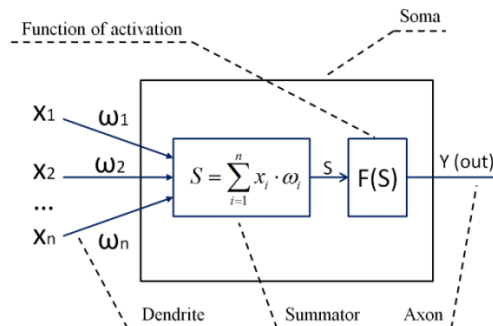


Рис. 1 Формальный нейрон

x_1, x_2, \dots, x_n – Входные сигналы нейрона

$\omega_1, \omega_2, \dots, \omega_n$ – Веса нейрона, взвешивающие входные сигналы

S – Сумма взвешенных сигналов

Y – Выходной сигнал

ФН является минимальным информационным уровнем для современной теории искусственных нейронных сетей (ИНС), которую не интересуют процессы, протекающие на более низком (например молекулярном) уровне.

В ФН предполагается движение сигналов лишь в прямом направлении: от дендритов, через сому и аксон на нижележащие синапсы. Конечно, в соответствии с рядом моделей (например сеть Хопфилда и сеть Хэмминга) [1], такой нейрон может обладать и рекуррентной связью, при которой аксон (выход) образует синапс с дендритом того же нейрона (вход) или дендритами вышележащих нейронов. Направление движения сигнала в таких сетях остается прежним. Хотя такие связи скорее исключение, чем правило для структуры мозга.

Моделирование процессов, происходящих в биологическом мозге, было первоначальной целью разработки теории ИНС. Однако модель ФН упрощена по сравнению с биологическим нейроном настолько, что её лишь условно можно называть нейроном. Подобное упрощение делает невозможным создание реалистичных моделей.

Одной из наименее правдоподобных особенностей известных нейросетевых моделей и алгоритмов обучения является их «цифровая» вычислительная природа. Совершенно неясно, кто же может в мозге выполнять все эти алгоритмические вычисления для обучения сети? Практически во всех известных на

сегодняшний день моделях, нейросеть, со всеми весами для каждого из нейронов, представляет собой пассивный объект обучения. Сами нейроны ничего не делают для своего обучения – это делает некий обучающий алгоритм, который осуществляется некой внешней, по отношению к самой нейросети, программой. В связи с недостатками нынешнего подхода необходимо искать альтернативные решения.

2 Положения новой концепции

2.1 Локальность процессов

Согласно теории Хэбба [3], предполагается только локальное взаимодействие между нейронами при отсутствии глобального учителя. Согласно Хэббу обучение является исключительно локальным явлением, охватывающим только два нейрона и соединяющий их синапс; не требуется глобальной системы обратной связи для развития нейронных образований. Иначе говоря, не может быть никаких механизмов пассивного обучения нейронов (нет никаких внешних, по отношению к нейрону, алгоритмов, использующих информацию о других нейронах). Только сам нейрон может обучать самого себя, используя для этого лишь ту информацию, которая поступает к нему через те связи с другими нейронами (синапсы), что он имеет. Все процессы нейрона локализованы. Кроме информации, поступающей по физическим связям, ни один нейрон не имеет больше никаких источников информации обо всей остальной сети. Основываясь на полученной информации, нейрон сам регулирует свою активность и таким образом обучается. Конечно, существуют механизмы массового регулирования нейронов, например, гормонами. Но эти механизмы не обеспечивают самого процесса обучения, а лишь регулируют массовую активность групп нейронов (глобальные переключатели).

Нейрон является полностью автономной самодостаточной минимальной адаптивной единицей сети. Из суммы обученных по отдельности нейронов складывается обучение всей сети. Это обеспечивает полную параллельность процесса обучения. Таким образом, оба основных нейросетевых процесса (обучение и распознавание) параллельны по своей природе.

2.2 Внутренняя обратная связь на основе ретроградного сигнала

В соответствии с одним из основных принципов кибернетики, любая адаптивная система, которой является и нейрон, обязана обладать обратной связью. Если исключить внешнюю рекуррентную связь, т.к. она отсутствует у большинства биологических нейронов, то остается предположить, что обратная связь реализуется

иным образом в самом нейроне (внутренняя обратная связь).

Прямая связь в нейроне реализуется, как известно, потенциалом действия (ПД) – разрядом, текущим в направлении от сомы к терминали аксона и далее через синапсы на дендриты нижележащих нейронов. Ряд фактов говорит о том, что внутренняя обратная связь иной природы в нейроне все же имеется. Например, было показано проявление пресинаптической долговременной потенциации (ДВП) в ответ на активацию постсинаптических рецепторов, что предполагает наличие ретроградного посредника в синапсе [4, Гл.16]. Роль внутринейронной (внутренней) обратной связи, доставляющей ретроградный сигнал от пресинаптической терминали аксона через аксон и сому к дендритам нейрона, могут выполнять связанные между собой микротрубочки цитоскелета. Внутри нейрона, в том числе внутри аксона и дендритов, микротрубочки, каждая из которых может достигать нескольких мм в длину, образуют сообщающиеся сети, охватывающие всю клетку, соединяясь между собой посредством белковых MAP-мостиков. Посредством особых белковых (актиновых) нитей через шипики микротрубочки связаны даже с синапсами (рис. 2).

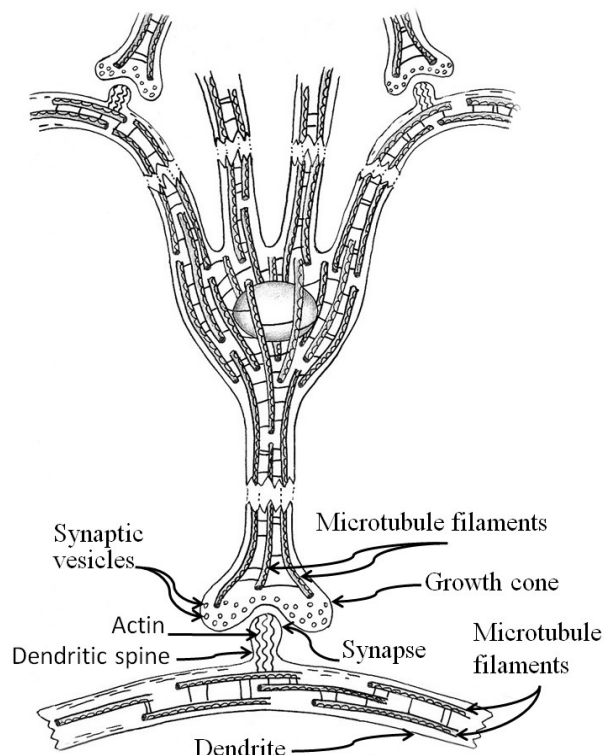


Рис. 2 Сеть микротрубочек в нейроне.

Каждая микротрубочка (рис. 3) представляет собой полую трубку, обычно состоящую из 13 рядов димеров тубулина. Каждая из молекул тубулина может существовать в двух конформациях. Было показано, что микротрубочки могут действовать как клеточные автоматы, передавая сложные сигналы в виде волн различных состояний электрической поляризации молекул тубулина, т.е. каждая микротрубочка способна

осуществлять на огромной скорости передачу сообщений [5]. Кроме того, полагают, что микротрубочки также ответственны за поддержание и изменение интенсивности синапсов.

2.3 Частотно-импульсная модуляция выходного сигнала нейрона

Как было показано [4, Гл.1], у ПД есть четко определенный порог, после достижения которого амплитуда и длительность ПД не зависят от параметров стимуляции. Отдельно взятый ПД в принципе не способен тонко описывать промежуточные состояния в довольно широком диапазоне. Было показано [4, Гл.1], что интенсивность возбуждения нейронов кодируется не отдельным импульсом (ПД), а частотой ПД. Более эффективные стимулы вызывают большую деполяризацию и, как следствие, более высокую частоту генерации ПД в нейроне. Частотно-импульсная модуляция подтверждается еще и тем фактом, что и на выделение медиаторов в синаптическую щель частота ПД оказывает влияние [4, Гл.15].

Таким образом, если за основу выходного сигнала нейрона принять не отдельно взятый ПД, фактически бинарный по своей природе, а частоту генерации ПД в нейроне, то мы получаем совершенно иную функцию активации нейрона, нежели пороговую. Частота нейрона может плавно изменяться от нескольких герц до килогерц. Таким образом, функция активации нейрона является монотонной, непрерывной и возрастающей. Причём это не обязательно должна быть сигмоида, это может быть даже линейная функция.

2.4 Новая модель равновесного нейрона

2.4.1 Сомы, сумматор, функция активации

Очевидно, что сома нейрона так же, как это предполагалось и ранее, осуществляет аналоговое суммирование поступающих через синапсы сигналов – суммирование всех дендритных ПД. На основе полученной суммы сигналов нейрон формирует свой уровень возбуждения, который выражается определенной частотой единичных ПД. Таким образом, нейрон частотно модулирует монотонно возрастающую функцию активации. В принципе, это может быть и сигмоида, но вид этой функции не имеет принципиального значения – это может быть и прямая. Естественное сопротивление дендритов можно описать некоторым вектором неотрицательных констант:

$$C = \{c_1, c_2, \dots, c_n\},$$

где n – число дендритов.

Проходящие через дендриты сигналы можно описать вектором: $A = \{\alpha_1, \alpha_2, \dots, \alpha_n\}$

Тогда их взвешенная сумма:

$$S = \sum_{i=1}^n c_i \cdot \alpha_i \quad (1)$$

Функция активации нейрона – функция от этой суммы.

2.4.2 Синаптический механизм

В современной теории ИНС в качестве модели синапса принят принцип Дейла – «один нейрон – один медиатор» [6]. Однако сегодня известно, что в одном химическом синапсе могут освобождаться более одного типа медиатора, причём их набор постоянен для данной клетки. Известно несколько десятков разнообразных медиаторов. Очень вероятно, что еще большее число медиаторных веществ остается неисследованным [4, Гл.15].

Разнообразные медиаторы на одном синапсе действуют синергично. Для выброса низкомолекулярных медиаторов достаточно одиночных импульсов, тогда как для высвобождения нейропептидных котрансмиттеров часто требуются пачки импульсов. Выработка каждого медиатора на каждом синапсе может быть представлена функцией выработки $v_i(x_i)$ (рис. 3).

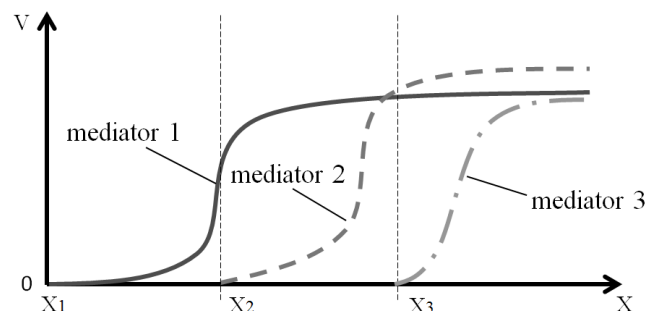


Рис. 3 Выработка медиаторов (V) в зависимости от сигнала (X)

x_1, x_2, x_3 – значения сигнала при которых начинается выработка нового медиатора.

Таким образом, формируется последовательность интервалов, на каждом из которых к общей сумме медиаторных сигналов подключается очередной тип медиатора. Иначе говоря, каждый интервал соответствует одному определенному только что подключившемуся медиатору или некоторой подгруппе только что подключившихся медиаторов.

Обнаружено множество методов регулирования передачи медиатором сигнала. Например, для каждого типа медиаторов на постсинаптической клетке существуют специфические к ним рецепторы. Более того, один и тот же медиатор может иметь разные типы постсинаптических рецепторов – как возбуждающие (положительная передача сигнала), так и тормозящие (отрицательная передача сигнала). Плотность таких

хемотрецепторов в зоне синапса определяет уровень передаваемого медиатором сигнала, как положительного, так и отрицательного [4, Гл.16].

Для простоты модели можно объединить все этапы регулирования передачи сигнала для каждого j -го медиатора в одну, общую для i -го синапса, функцию передачи сигнала $\mu_{ji}(v_i)$, где $v_i = v_i(x_i)$. Следовательно, в конечном счёте, функцию передачи медиатором сигнала можно представить функцией зависящей от сигнала X : $\mu_{ji}(x_i)$ (рис. 4а).

Подобно тому, как ПД от всех дендритов суммируются в соме, вклад каждого медиатора в формирование дендритного ПД суммируется в дендрите. Следовательно, передаваемый через синапс суммарный сигнал можно описать следующим образом:

$$\alpha_i = \alpha_i(x_i) = \sum_{j=1}^m \mu_{ji}(x_i) \quad (2)$$

где m – количество медиаторов, воспринимаемых рецепторами на синапсе i -го дендрита, x_i – выходной сигнал вышележащего нейрона, с которым i -ый дендрит создал синапс, а $\mu_{ji}(x_i)$ – функция передачи j -ым медиатором сигнала на i -ом синапсе (рис. 4б).

Фактически, передаваемый сигнал $\alpha_i(x_i)$ может быть описан вектором: $\mu_i = \{\mu_{1i}(x_i), \mu_{2i}(x_i), \dots, \mu_{mi}(x_i)\}$

Подставив (2) в (1) мы получаем, что сумма сигналов в нейроне:

$$S = \sum_{i=1}^n c_i \cdot \alpha_i(x_i) = \sum_{i=1}^n c_i \cdot \left(\sum_{j=1}^m \mu_{ji}(x_i) \right) \quad (3).$$

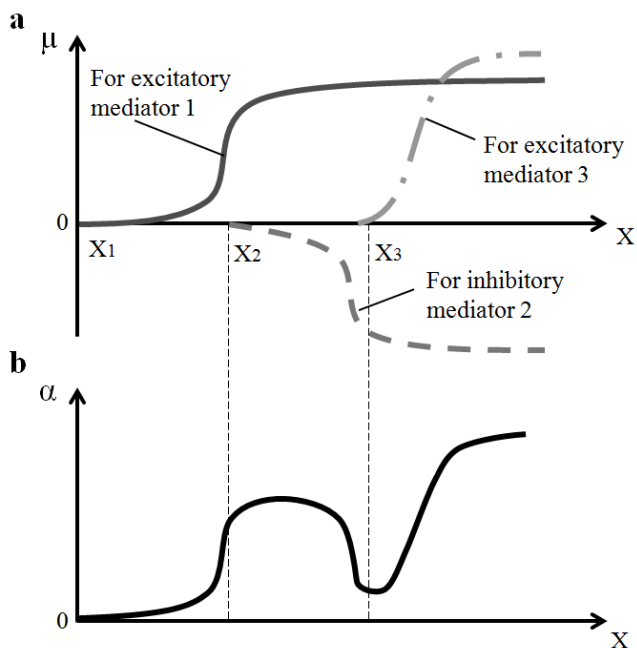


Рис. 4 Функция передачи медиатором сигнала (а) и синаптическая функция передачи сигнала (б)
 X – входной сигнал синапса

x_1, x_2, x_3 – значения сигнала при которых начинается выработка нового медиатора.

μ – уровень функции передачи медиатором сигнала
 α – уровень синаптической функции передачи сигнала

2.4.3 Обучение как равновесный процесс

Как было показано Хэббом [3], стимуляция пресинаптического волокна без поддержки стимулами с постсинаптической стороны ведёт к синаптической депрессии, т.е. снижению постсинаптических потенциалов в ответ на пресинаптические стимулы. И наоборот, пресинаптические стимулы, поддержанные постсинаптическими, ведут к синаптической потенциации, т.е. повышению потенциалов.

Принимая в расчёт ретроградный сигнал (при стимуляции постсинаптической стороны), можно предположить, что прямой ПД ведёт к синаптической депрессии, а ретроградный сигнал ведёт к синаптической потенциации. Причём степень изменений, как в первом, так и во втором случае, прямо зависит от величины стимула.

Было показано, что ПД клетки возникает в точке на аксоне сразу за сомой и распространяется не только по аксону в прямом направлении, но и через сому на дендриты в обратном. Ретроградный кальциевый дендритный ПД, в свою очередь, может вызывать локальные изменения концентрации кальция и влиять на синаптическую передачу [4, Гл.8]. Возможно, именно этот механизм лежит в основе синаптической депрессии при прямом ПД.

Одна из основных идей данной работы заключается в следующем: механизм встречных сигналов нейрона, снижая и повышая синаптические веса, приводит их в состояние равновесия, что и лежит в основе обучения. Кроме того, при этом, как указывалось выше, для обучения нейрона используются лишь сигналы, поступающие из вышележащих синапсов при прямом сигнале и сигналы из нижележащих синапсов при ретроградном сигнале. Можно говорить о том, что ретроградный сигнал несёт информацию об ожидаемом уровне возбуждения нейрона при заданных входных сигналах. Природа возникновения ретроградных сигналов требует освещения в отдельной статье.

Таким образом, для отдельного синапса возможны следующие ситуации:

1. Входной образ приводит на выходе нейрона, вследствие особенностей его предыдущего обучения, к слабому выходному сигналу. Однако ожидался сильный выходной сигнал, что выражается в сильном ретроградном сигнале. Как указывалось выше, слабый прямой ПД ведёт к слабой депрессии синапсов. Если ретроградный

сигнал превышает прямой, то и уровень потенциации синапсов превысит уровень их депрессии. В следующий раз, при том же входном образе, уровень возбуждения нейрона будет выше.

2. Входной образ приводит на выходе нейрона к сильному выходному сигналу. Однако ожидался слабый выходной сигнал, что выражается в слабом ретроградном сигнале. Таким образом, сильный прямой ПД ведёт к сильной депрессии синапсов, а слабый ретроградный – к их слабой потенциации. Уровень депрессии синапсов в таком случае превысит уровень их потенциации. В следующий раз, при том же входном образе, уровень возбуждения нейрона будет ниже.
3. Сильный прямой ПД и сильный ретроградный сигнал приведут к сильной депрессии с одной стороны и сильной потенциации с другой, которые взаимно нейтрализуют друг друга и оставят уровень возбудимости на прежнем (высоком) уровне.
4. Слабый прямой ПД и слабый ретроградный сигнал подобны случаю 3, с той лишь разницей, что взаимно нейтрализуют друг друга слабая депрессия и слабая потенция, оставив нейрон на прежнем уровне слабой возбудимости.

Иными словами, обучение нейрона – это установление равновесия между депрессией и потенциацией синапсов при прямом и ретроградном сигналах.

Предложенная равновесная модель формирования памяти совершенно естественно соотносится с правилами Хэбба. Однако правила Хэбба не срабатывают в ситуации, когда обратный сигнал сильнее прямого, что приводит к усилению связи.

Описанный механизм сильно упрощён. Кажется очень вероятным, что приведённый равновесный процесс неравномерно распределён между всеми синапсами нейрона. Например, было показано, что обратное распространение ПД из сомы зависит от входного сопротивления различных ветвей (дендритов). Входное сопротивление, в свою очередь, зависит от степени активности возбуждающих и тормозных синапсов. Таким образом, обратное распространение возбуждения в дендриты зависит от синаптической активности [4, Гл.8].

2.5 Локализованный равновесный процесс

Описанный выше равновесный механизм обучения работает не на всём синапсе, а лишь на некоторых из его медиаторов.

Как указывалось выше, различным уровням пресинаптического сигнала соответствуют различные группы медиаторов. Это означает, что и описанный выше равновесный механизм компенсации ошибки

работает для определённого образа на определённой группе медиаторов. Для другого образа эта группа медиаторов будет уже иной, т.е. различные образы характеризуются наборами медиаторов лишь частично пересекающимися от образа к образу.

2.6 Обучение нейрона

Обучение нейрона можно описать следующей простой схемой:

1. На синапсы нейрона подаётся входной образ.
2. Входные сигналы на синапсах активизируют строго определённые группы медиаторов.
3. Синаптические веса для выбранных медиаторов суммируются в соме формируя прямой ПД.
4. Величина прямого ПД определяет уровень депрессии выбранных медиаторов (снижения весов).
5. На выход нейрона подаётся (ожидаемый) ретроградный сигнал, вызывающий потенцию выбранных входным образом медиаторов (повышение весов). Причём уровень этой потенциации зависит от величины ретроградного сигнала. В результате прямого ПД и ретроградного сигнала нейрон приходит в состояние равновесия, таким образом обучаясь текущему образу.
6. На входы и выход нейрона подаётся следующий образ, и шаги 1 – 6 повторяются для нового образа.
7. После прохождения всех образов (одна эпоха), процесс для всех образов может быть повторён вновь – до достижения необходимого уровня точности обучения или до достижения минимальной скорости схождения.

В процессе обучения новому образу, уровень весов, формирующих предыдущую память, лишь частично модифицируется новым образом. При достаточно большом количестве синапсов и возможных уровней сигнала (медиаторов) влияние таких искажений оказывается чрезвычайно малым. Более того, при повторном обучении веса постепенно перераспределяются, находя положение с максимальным уровнем общего равновесия, т.е. такая система сама находит положение глобального минимума.

3 Выводы

Предложены новые модели формального нейрона, формирования памяти, обучения и распознавания, которые хорошо согласуются с экспериментальными данными. Описанная модель нейронной сети гораздо ближе к биологической нейронной сети, чем известные ИНС.

Предложенная новая модель формального нейрона и сети включает:

- Синапс с множеством медиаторов (весов) и с нелинейной функцией передачи сигнала.
- Внутреннюю обратную связь, реализующую проведение ретроградного сигнала в направлении от терминали аксона к дендритам и синапсам с вышележащими нейронами.

Модель обучения этого нейрона и сети включает:

- Выбор медиаторов в синапсе, участвующих в обучении и распознавании, в зависимости от входящего сигнала.
- Корректирование весов на синапсах путём депрессии и потенциации при прямом и ретроградном сигналах. Причём величина снижения синаптических весов прямо зависит от прямого ПД нейрона, а величина повышения этих весов прямо зависит от ретроградного сигнала.

Эксперименты [7] с виртуальными моделями предложенной сети показали, что сеть имеет:

- Высокую скорость обучения, линейно зависящую от размеров сети и объёмов данных, в отличие от других моделей ИНС, имеющих экспоненциальную зависимость. Предложенная сеть требует количество эпох обучения на несколько порядков меньше, чем любые из классических ИНС.
- Масштабируемость, позволяющая строить такие сети любого размера и сложности.
- Простоту в реализации в виде аналоговой или цифровой схемы, не требующей «внешнего учителя» в виде компьютера или чипа, реализующего длительные и сложные цифровые вычисления [8].

4 Благодарности

Я высоко ценю плодотворные дискуссии и поддержку А. Будишевской, В. Просяника, Б. Злотина, В. Песчанской, С. Вишнепольской и А. Зусман. Я также очень благодарен за поддержку, оказанную командой компании Progress Inc. и инвесторами.

5 Библиография

- [1] S.O. Haykin “Neural Networks and Learning Machines” (3rd Edition), Pearson Education, Inc., 2009
- [2] W.S. McCulloch, W.H. Pitts “A logical calculus of the ideas immanent in nervous activity”. Bulletin of Mathematical Biophysics, Vol. 5, p. 115-133, 1943
- [3] D.O. Hebb “The organization of behavior”. Wiley, New York, 1949
- [4] J.G. Nicholls, A.R. Martin, P.A. Fuchs, D.A. Brown, M.E. Diamond, D.A. Weisblat “From Neuron to Brain, Fifth Edition”, Publisher: Sinauer Associates, 2012

[5] T.J.A. Craddock, D. Friesen, J. Mane, S. Hameroff, J.A. Tuszynski. “The feasibility of coherent energy transfer in microtubules”. Journal of the Royal Society, Interface / the Royal Society, 11(100):20140677, 2014

[6] H.H. Dale “Pharmacology and nerve endings”. Proc Roy Soc Med 28:319-332, 1935

[7] D. Pescianschi, A. Boudichevskaia, B. Zlotin and V. Proseanic “Analog and digital modeling of a scalable neural network”, presented for the current conference.

[8] D. Pescianschi “Neural network and method of neural network training” PCT patent application US1519236, 2015