

ПРИМЕНЕНИЕ СЕМАНТИЧЕСКОЙ НЕЙРОННОЙ СЕТИ В ЭКСПЕРТНОЙ СИСТЕМЕ, ПРЕОБРАЗУЮЩЕЙ СМЫСЛ ТЕКСТА НА ЕСТЕСТВЕННОМ ЯЗЫКЕ

ШУКЛИН Д. Е.

Рассматривается структура семантической нейронной сети, выполняющей функции экспертной системы, преобразующей смыслы текста на естественном языке. Предлагается реализация параллельных вычислений в семантической нейронной сети на базе последовательной вычислительной системы. Рассматривается оптимизация времени принятия решений и управления вниманием в экспертной системе ELEX-4, выполненной на основе семантической нейронной сети.

Каждый нейрон в семантической нейронной сети представляет собой некоторое элементарное понятие обрабатываемого смысла [1]. Смысл, извлеченный из текста на естественном языке, формализовано представляется в семантической нейронной сети как мгновенное состояние множества нейронов – эффекторов, находящихся в слое извлечения смысла из входной символьной последовательности [2]. Градиентное значение на выходе нейрона представляет собой нечеткий фактор уверенности (certainty factor) – степень уверенности в том, что данное элементарное понятие содержится в обрабатываемом тексте. Логическому значению “истина” градиентного значения на выходе аксона соответствует полная уверенность в том, что данное понятие присутствует в тексте. Градиентному значению “ложь” соответствует полная уверенность в отсутствии данного понятия в обрабатываемом тексте. Промежуточные значения соответствуют предположительному присутствию или отсутствию понятия в тексте со степенью уверенности, равному градиентному значению.

Тексты на естественном языке, содержащие различные смыслы, при обработке в нейронной сети будут вызывать различные мгновенные состояния эффекторов слоя извлечения смысла. Для преобразования одного смысла текста в другой достаточно преобразовать состояние одной группы нейронов в состояние другой группы. Обработка смысла текста будет заключаться в обработке состояний нейронов, характеризующихся набором их выходных градиентных значений. Для преобразования смысла в виде набора градиентных данных могут применяться экспертные системы.

Обычно экспертные системы помогают человеку принимать решения и решать разнообразные задачи. Экспертные системы применяются в областях человеческой деятельности, где требуется решать задачи, проводить классификацию, принимать решения в условиях недостаточности информации о поставленной задаче, а также при решении задач, для которых отсутствуют алгоритмы решения.

Так как в реальном мире редко можно утверждать что-либо с полной уверенностью, то в процессе рассуждений экспертные системы используют нечеткую логику. Для выяснения степени определенности утверждения в нечеткой логике используется фактор уверенности – число, обычно находящееся в интервале от 0 до 1 (похоже на теорию вероятностей). Но в отличие от теории вероятностей фактор уверенности выражает не вероятность возникновения события, а субъективную уверенность в этом. Отказ от применения классической теории вероятностей обусловлен многими причинами. Вот некоторые из них:

– Человек практически никогда не оперирует с точными значениями вероятностей и ему трудно, а иногда и невозможно точно определить вероятность возникновения события.

– Экспертные системы практически никогда не получают оптимального решения, а довольствуются лишь минимально допустимым уровнем решения. Поэтому отпадает необходимость проведения точных расчетов. Становится возможным только приблизительно определять фактор уверенности. В результате разработчики экспертных систем нередко используют свои собственные алгоритмы обработки факторов уверенности.

– При необходимости вычислять вероятности в сложных логических выводах расчеты, выполненные по классическим алгоритмам, усложняются. Эффективность работы экспертных систем резко падает.

– Для хоть сколько-нибудь точного расчета вероятности в сложных логических выводах необходимо иметь точную информацию о степени корреляции входных данных. Часто эта информация для экспертной системы недоступна.

Экспертные системы обычно состоят из двух основных частей: базы знаний и механизма логического вывода. В процессе проектирования экспертной системы группа экспертов вначале тщательно анализирует ситуацию, для которой строится экспертная система. Затем эти люди разрабатывают правила, которыми экспертная система должна руководствоваться при решении возникшей проблемы. Полученные правила помещаются в базу знаний экспертной системы. Одновременно идет проектирование механизма логического вывода, который определяет, какие правила и в каком порядке будут применяться экспертной системой для решения поставленной задачи. Он также решает вопрос о необходимости запроса у пользователя дополнительной информации. Порядок применения правил и порядок генерации запросов на получение дополнительных данных можно отнести к задаче управления вниманием. Механизм логического вывода также обеспечивает объяснение полученных результатов.

В экспертных системах правила обычно представляются в следующем виде [3]:

ЕСЛИ А ТО В,

ЕСЛИ не А ТО В,

ЕСЛИ А и В и В ТО Г и не Д,

и. т. д., где А, В, Г, Д – логические высказывания.

Каждое правило делится на две части. В первой части указываются условия, которые должны быть выполнены для срабатывания этого правила, а во второй — следствия из этих условий, например:

ЕСЛИ производится опасная разбраковка деталей, ТО распределение погрешности параметра детали антимодальное.

Правила помещаются в базу знаний. Туда же, в виде фактов, помещаются входные данные и результаты срабатывания правил. Если ситуация, анализируемая экспертной системой, удовлетворяет условиям срабатывания правила, то его следствия считаются верными для данной ситуации. Условия правила содержат операции конъюнкции, дизъюнкции и инверсии. Если условия у двух или более правил выполняются одновременно, то необходимо выбрать следствия для дальнейшей обработки. Выбор может быть осуществлен различными способами: 1- выбирается следствие у первого попавшегося правила; 2- из множества допустимых правил одно следствие выбирается случайно; 3- выбирается следствие у правила с наиболее жесткими условиями; 4- все следствия у допустимых правил выбираются одновременно. Для обеспечения обработки многозначных лексем текста будем считать истинными все следствия у всех правил, у которых выполняются условия. При этом одинаковые понятия, являющиеся следствиями разных правил, объединяются операцией дизъюнкции.

Таким образом, экспертная система представляет собой сеть правил. Все правила выполняются одновременно. Отдельные правила представляют собой отдельные узлы, а понятия, входящие в условия и следствия правил, представляют собой связи между узлами этой сети. Отдельные операции над обрабатываемым смыслом текста представляются в этой сети как отдельные правила в базе знаний экспертной системы.

Фрагмент семантической нейронной сети, как экспертная система, представляет собой три слоя нейронов. Слой рецепторов собирает информацию для обработки в нейронной системе. Градиентное значение нейрона - рецептора представляет собой нечеткий фактор уверенности элементарного понятия, которое символизирует нейрон. Слой эффекторов выдает результаты работы экспертной системы в виде совокупности градиентных данных нейронов-эффекторов. Как и в слое рецепторов, градиентное значение нейрона - эффектора представляет собой нечеткий фактор уверенности элементарного понятия, которое символизирует нейрон. Третий слой - слой обработки расположен между слоем рецепторов и слоем эффекторов, в нем расположены нейроны, реализующие базу знаний экспертной системы. В случае работы такой экспертной системы в высокораспараллеленном режиме каждый нейрон обрабатывает градиентные данные со своего дендрита и передает результат выполненной операции следующим нейронам по аксону. Время принятия решения такой экспертной системы зависит не от объема базы знаний, а от ее сложности. Максимальная задержка принятия решения после установки на слое рецепторов всех

входных данных будет определяться следующей формулой:

$$T_{\max} = T_1 N_{\max}, \quad (1)$$

где T_{\max} — максимальная задержка принятия решения; T_1 — время обработки градиентных данных в одном нейроне; N_{\max} — максимальное количество последовательно включенных нейронов.

Применив разложение правил вывода по совершенной дизъюнктивной форме или по совершенной конъюнктивной форме [4], можно реализовать слой обработки в два слоя: слоя нейронов, выполняющих операции конъюнкции, и слоя нейронов, выполняющих операции дизъюнкции. Это позволит экспертной системе принимать любое решение за два такта (времени обработки данных в одном нейроне) работы.

Однако на сегодняшний день параллельные вычислительные системы, реализованные аппаратно, труднодоступны. Поэтому возникает задача реализации модели параллельного аналогового вычислителя средствами последовательных цифровых вычислительных систем. Как описано в [1], для реализации параллельного аналогового нейронного вычислителя необходима имитация несинхронизированных нейронов синхронизированными с многопроходным алгоритмом их обработки.

Для реализации промышленной экспертной системы с помощью семантических нейронных сетей не нужно, а возможно и вредно, вводить в систему такие понятия, как усталость нейрона, время его реакции на поступившие воздействия, время нарастания сигнала, переходные характеристики и подобные им. Это приведет к снижению скорости работы экспертной системы и к влиянию на ее решения внутренней реализации нейрона, не относящейся к контексту решаемой задачи. В связи с этим ограничимся моделированием передачи между нейронами градиентных данных и их обработкой с использованием функций нечеткой логики, таких как конъюнкция, дизъюнкция и инверсия [1].

Чтобы обеспечить имитацию одновременной параллельной работы множества нейронов на последовательной вычислительной машине, необходимо последовательно обрабатывать состояния всех нейронов в сети для каждого кванта времени.

Обработка нейронов в экспертной системе производится тактами. Такты выполняются до момента, когда в сети не останется ни одного нейрона, требующего обработки. После завершения всех тактов на эффекторах выставляется текущее решение, принятое экспертной системой. При изменении информации на рецепторах запускаются новые такты, и по их окончании на эффекторах выставляется уточненное решение. Работа экспертной системы проходит либо в один такт, по выставлению всех входных градиентных данных на слое рецепторов, либо итерационно, по мере накопления этих данных.

Каждый такт состоит из нескольких проходов. Проход представляет собой последовательный обход всех нейронов, зарегистрированных для прохода. При этом для каждого нейрона выполняется специальная функция, соответствующая данному проходу.

Для обеспечения оптимизированной обработки проходов каждый нейрон будет иметь следующие флаги состояния: *активен/пассивен; изменен/не изменен; зарегистрирован/не зарегистрирован*. Так же каждый нейрон будет иметь свое градиентное значение, являющееся результатом работы функции, выполняемой нейроном, над входными градиентными данными. Флаги состояния нейрона и его градиентное значение назовем словом состояния нейрона. Каждый нейрон будет иметь два таких слова состояния: актуальное и текущее. Оба слова состояния нейрона помещаются в соответствующую секцию [2] нейрона. Актуальное слово состояния — это слово состояния нейрона, которое считается актуальным для данного кванта времени для этого нейрона. Именно актуальные градиентные данные получает нейрон по своему дендриту для обработки, и именно они передаются остальным нейронам для дальнейшей обработки. Однако результат обработки входных градиентных данных, поступающих из дендрита, помещается в текущее слово состояния нейрона. В текущем слове состояния накапливаются результаты обработки входных градиентных данных, поступивших из других нейронов. *Активным* считается нейрон, требующий обработки входных градиентных данных. *Пассивным* считается нейрон, не требующий обработки входных градиентных данных. *Измененным* считается нейрон, у которого градиентное значение текущего слова состояния отличается от градиентного значения актуального слова состояния. *Не измененным* считается нейрон, у которого после обработки входных градиентных данных текущее значение не изменилось по сравнению с актуальным градиентным значением. *Зарегистрированным* считается нейрон, зарегистрированный для обработки в последующем проходе. *Не зарегистрированным* считается нейрон, не требующий обработки в последующем проходе. Понятие *зарегистрированный/не зарегистрированный* нейрон фактически совпадает с понятием *активный/пассивный* нейрон и введено для оптимизации обработки больших (более 1000 нейронов) баз знаний.

Для каждого нейрона определены четыре функции проходов: две основные (1), (2) и две вспомогательные (3), (4). Основные функции проходов последовательно вызываются в такте обработки. Вспомогательные вызываются основными функциями для передачи или получения данных от других нейронов:

Проход-1: расчет градиентного значения:

1- если нейрон актуально-активен, то по дендриту рассылается сообщение *Проход-3*: запрос об актуальных входных градиентных данных, при этом полученные актуальные градиентные данные обрабатываются, и результат функции, выполняемой нейроном, помещается в текущее градиентное значение;

2 — сравниваются текущее и актуальное градиентные значения, в случае их неравенства устанавливается текущий флаг *изменен*;

3 — если текущий флаг *изменен* установлен, то по аксону передается сообщение *Проход-4*, вынуждающее все нейроны-приемники перейти в активное зарегистрированное состояние.

Проход-2: сброс, установка и регистрация:

1 — если нейрон *текуще не активен* и *текуще не зарегистрирован*, то он deregистрируется, текущий флаг *зарегистрирован* сбрасывается;

2 — текущее слово состояния нейрона присваивается актуальному слову состояния нейрона, при этом градиентное значение присваивается актуальному значению, флаги текущего слова состояния присваиваются флагам актуального слова состояния;

3 — после этого флаги состояния *активен* и *изменен* текущего слова состояния сбрасываются в неактивное состояние.

Проход-3: запрос об актуальных входных градиентных данных:

1 — нейрон возвращает свое актуальное слово состояния, включая актуальное градиентное значение.

Проход-4: перейти в активное зарегистрированное состояние:

1 — устанавливается флаг текущего *активного* состояния;

2 — если нейрон *текуще-активен* и *текуще-незарегистрирован*, то он регистрируется, флаг *текуще-зарегистрирован* устанавливается.

Регистрация нейронов проходит в соответствующей секции нейрона управления слоем обработки экспертной системой. Этим обеспечивается возможность нахождения в одной нейронной сети нескольких экспертных систем, при этом один и тот же нейрон может одновременно входить в разные зоны регистрации. На время выполнения одного такта нейрон управления замораживается, т.е. в памяти системы создается его копия, и проходы вызываются по замороженному состоянию секции регистрации нейронов базы знаний. Нейроны слоя знаний имеют доступ к замороженному нейрону. Это обеспечивает правильный процесс регистрации - deregистрации нейронов базы знаний в нейроне управления.

Работа нейрона управления заключается в рассылке сообщений зарегистрированным нейронам о выполнении проходов. Нейрон управления может работать в двух режимах — одноктактовом и многотактовом. В одноктактовом режиме нейрон управления один раз рассылает сообщение *Проход-1* и затем один раз сообщение *Проход-2*. Во многотактовом режиме нейрон управления последовательно рассылает сообщения *Проход-1* и *Проход-2*, пока существует хотя бы один зарегистрированный нейрон. В случае наличия в нейронной сети циклических ссылок многотактовый режим может привести к зависанию системы выполнения. Для реализации экспертной системы, основанной на правилах, средствами семантической нейронной сети необходимо преобразовать сеть правил в нейронную сеть. В процессе преобразования необходимо обеспечить решение следующих проблем: управление вниманием и объяснение полученных результатов. Управление вниманием разбивается на подзадачи определения неизвестного факта, наиболее важного в данный момент, и определение степени завершенности вывода по каждому возможному ответу.

При преобразовании правил экспертной системы в нейронную сеть первым этапом является разбиение каждого правила на элементарные операции дизъюнкции, конъюнкции и инверсии. Каждой такой операции назначается группа из нескольких нейронов. Каждая такая группа нейронов является элементарной операцией, выполняемой правилом экспертной системы. Группы нейронов связаны с другими группами посредством дендритов и аксонов. В каждой группе существует один главный нейрон, значение на его выходе является значением операции, выполняемой группой. Дендрит главного нейрона группы связан с главными нейронами групп, являющихся понятиями, от значения которых зависит значение текущей группы. Аксон главного нейрона связан с дендритами нейронов групп, значения которых зависят от значения текущей ассамблеи. Таким образом, текущее значение, передаваемое по аксону главного нейрона, является текущим значением элементарного понятия в экспертной системе. Остальные нейроны группы выполняют служебные функции, такие как слежение за процессом логического вывода для последующего объяснения результатов и управление вниманием.

Рассмотрим преобразование правил для определения типа ПЭВМ типа x86 в базу знаний в виде семантической нейронной сети. На рис. 1 приведен текст базы знаний для исследовательской экспертной системы ELEX-4.

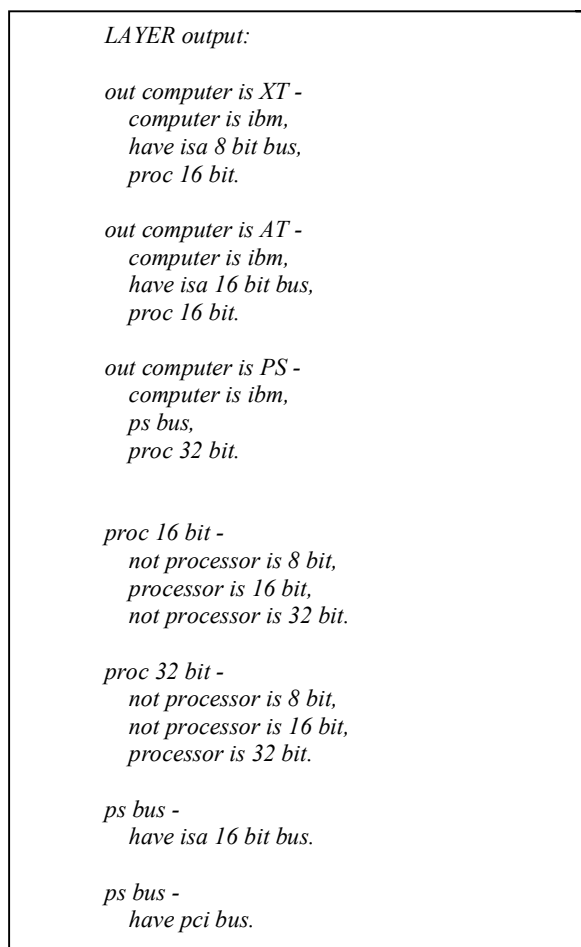


Рис. 1

Правило в экспертной системе ELEX-4 состоит из двух частей, в первой, перед символом “-” записывается следствие, во второй, после символа “-” через запятую записываются условия правила. Одинаковые следствия у разных правил объединяются группой нейронов, выполняющих операцию дизъюнкции. Условия, не имеющие соответствующих следствий, становятся рецепторами. Эффекторы описываются с использованием ключевого слова “out”, инверсия условия описывается с использованием ключевого слова “not”.

В экспертной системе ELEX-4 для обеспечения логического вывода и управления вниманием используются градиентные значения трех типов: значение операции V , завершенность вывода по операции C и важность операции I . Значение операции V и завершенность вывода по операции C являются значениями прямого распространения, т.е. они распространяются в направлении от фактов к выводам экспертной системы. Важность операции I является значением обратного распространения и распространяется в направлении от выводов к фактам. Таким образом, значения фактов V и завершенность выводов по операции C подаются на входы экспертной системы, а снимаются с ее выходов. Значение важности факта для вывода I подается на выходы экспертной системы, а снимается с ее входов.

После преобразования базы знания с рис. 1 в семантическую сеть экспертной системы ELEX-4 сеть главных нейронов будет иметь вид, приведенный на рис. 2.

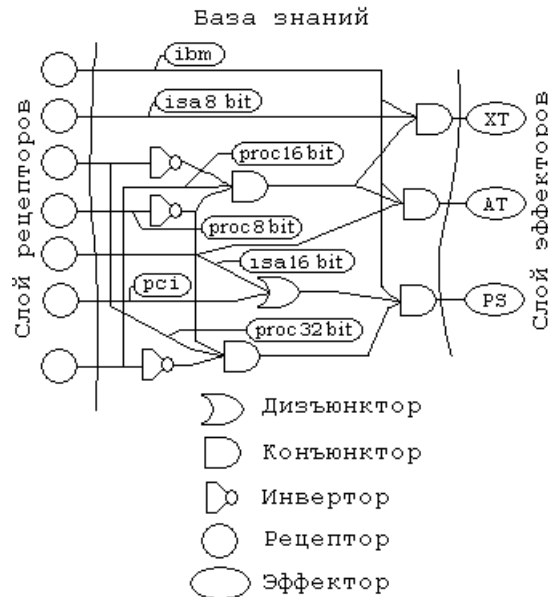


Рис. 2

Значение операции V рассчитывается главным нейроном группы и показывает степень возможности возникновения данного вывода экспертной системы. Завершенность вывода по операции C показывает степень достоверности полученного результата. Завершенность вывода по операции рассчитывается группой на основании типа операции, выполняемой группой, и степени завершенности входных данных, полученных группой. Важность

операции I определяется степенью завершенности вывода группы и важностью групп, использующих результат операции.

В экспертной системе ELEX-4 применяются следующие зависимости расчета градиентных значений, реализуемые семантической нейронной сетью:

Группа конъюнкции “ \wedge ”:

$$V = V_1 \wedge V_2 \wedge \dots \wedge V_N, \quad (2)$$

$$C = (C_1 \wedge C_2 \wedge \dots \wedge C_N) \vee (C_1 \wedge (\neg V_1)) \vee (C_2 \wedge (\neg V_2)) \vee \dots \vee (C_N \wedge (\neg V_N)), \quad (3)$$

$$I = (\neg C) \wedge (I_1 \vee I_2 \vee \dots \vee I_N), \quad (4)$$

где V – результат конъюнкции; V_1, V_2, \dots, V_N – входные значения; C – завершенность вывода конъюнкции; C_1, C_2, \dots, C_N – завершенность входных значений; I – важность операции для входных нейронов; I_1, I_2, \dots, I_N – важности операций от выходных нейронов.

Таким образом, в случае поступления на вход группы, реализующей операцию конъюнкции, хотя бы одного завершенного значения “ложь” эта группа примет завершенное значение “ложь” и минимальную важность.

Группа дизъюнкции “ \vee ”:

$$V = V_1 \vee V_2 \vee \dots \vee V_N, \quad (5)$$

$$C = (C_1 \wedge C_2 \wedge \dots \wedge C_N) \vee (C_1 \wedge V_1) \vee (C_2 \wedge V_2) \vee \dots \vee (C_N \wedge V_N), \quad (6)$$

$$I = (\neg C) \wedge (I_1 \vee I_2 \vee \dots \vee I_N), \quad (7)$$

где V – результат дизъюнкции; V_1, V_2, \dots, V_N – входные значения; C – завершенность вывода дизъюнкции; C_1, C_2, \dots, C_N – завершенность входных значений; I – важность операции для входных нейронов; I_1, I_2, \dots, I_N – важности операций от выходных нейронов.

Таким образом, в случае поступления на вход группы, реализующей операцию дизъюнкции, хотя бы одного завершенного значения “правда” эта группа примет завершенное значение “правда” и минимальную важность.

Группа инверсии “ \neg ”:

$$V = \neg V_0, \quad (8)$$

$$C = C_0, \quad (9)$$

$$I = (\neg C) \wedge (I_1 \vee I_2 \vee \dots \vee I_N), \quad (10)$$

где V – результат инверсии; V_0 – входное значение; C – завершенность вывода инверсии; C_0 – завершенность входного значения; I – важность операции для входных нейронов; I_1, I_2, \dots, I_N – важности операций от выходных нейронов.

Группа инверсии имеет завершенность, равную завершенности входного значения, и важность, равную объединению важности групп – приемников результата инверсии.

Часть нейронов семантической нейронной сети, выполняющую функции базы знаний экспертной системы, будем условно обозначать, как показано на рис.3.

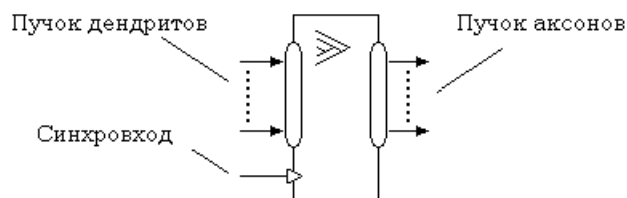


Рис. 3

В случае работы управляющего нейрона в одноканальном режиме по синхровходу будет поступать сигнал, инициирующий выполнение одного такта работы, в случае работы управляющего нейрона в многоканальном режиме по синхровходу будет поступать сигнал, инициирующий циклическое выполнение тактов, до тех пор, пока в соответствующей секции управляющего нейрона будут находиться зарегистрированные нейроны.

Участки семантической нейронной сети, выполняющие функции базы знаний экспертной системы, могут каскадом включаться либо друг за другом, либо вместе с участками сети, выполняющими другие функции. Рассмотренная реализация экспертной системы средствами семантической нейронной сети позволяет одновременно вести параллельную обработку множества правил в базе знаний. Таким образом, экспертная система на основе семантической нейронной сети в состоянии выполнять в режиме реального времени преобразование смысла текста как преобразование состояния участков нейронной сети. Правила экспертной системы представляются в семантической нейронной сети группами нейронов. Управление вниманием обеспечивают служебные нейроны группы. Объяснение полученных результатов выполняется путем анализа состояния градиентных значений нейронов, находящихся в слое обработки экспертной системы.

Литература: 1. Дударь З.В., Шуклин Д.Е. Семантическая нейронная сеть, как формальный язык описания и обработки смысла текстов на естественном языке // Радиоэлектроника и информатика. 2000. № 3. С. 72-76. 2. Дударь З.В., Шуклин Д.Е. Реализация нейронов в семантических нейронных сетях // Радиоэлектроника и информатика. 2000. № 4. С. 89-96. 3. Экспертные системы. Принципы работы и примеры: Пер. с англ./ А.Брукинг, П.Джонс, Ф.Кокс и др./ Под ред. Р. Форсайта. М.: Радио и связь, 1987. 224 с. 4. Шабанов-Кушнаренко Ю.П. Теория интеллекта. Математические средства. Х.: Вища школа, 1984. 143 с.

Поступила в редколлегия 27.10.2000

Рецензент: д-р техн. наук, проф. Шабанов-Кушнаренко Ю. П.

Шуклин Дмитрий Евгеньевич, аспирант кафедры программного обеспечения ЭВМ ХТУРЭ. Научные интересы: теория интеллекта, нейронные сети. Адрес: Украина, 61166, Харьков, пр. Ленина, 14, тел. 40-94-46.