

УДК 681.3.06

Александр Андреевич Щербина

Кандидат технических наук, доцент, доцент кафедры информационных технологий

Людмила Алексеевна Терейковская

Ассистент кафедры информационных технологий

*Київський національний університет будівництва і архітектури, Київ***МЕТОД ОПРЕДЕЛЕНИЯ НЕЙРОСЕТЕВОЙ АРХИТЕКТУРЫ В ЗАДАЧАХ ГОЛОСОВОГО ВЗАИМОДЕЙСТВИЯ ДИСТАНЦИОННОГО ОБУЧЕНИЯ**

Вперше розроблено метод визначення архітектури нейронної мережі, призначеної для розв'язання типових задач голосової взаємодії в інформаційній системі дистанційного навчання. Метод, що базується на концепції відповідності вимог задачі і можливостей апробованих архітектур, дозволяє визначити оптимальну нейромережеву архітектуру. Наведено приклад встановлення оптимальної архітектури нейронної мережі, що використовується для розпізнавання фонем в процесі голосової авторизації.

Ключові слова: дистанційне навчання, голосова взаємодія, нейронна мережа, багатошаровий перцептрон, фонема

Впервые разработан метод определения архитектуры нейронной сети, предназначенной для решения типовых задач голосового взаимодействия в информационной системе дистанционного обучения. Метод, базируясь на концепции соответствия требований задачи и возможностей апробированных архитектур, позволяет определить оптимальную нейросетевую архитектуру. Приведен пример установления оптимальной архитектуры нейронной сети, которая используется для распознавания фонем в процессе голосовой авторизации.

Ключевые слова: дистанционное обучение, голосовое взаимодействие, нейронная сеть, многослойный перцептрон, фонема

First developed a method for determining the optimal architecture of a neural network for solving typical problems of vocal interaction in the information system of distance learning. In determining the optimal architecture used multi-criteria evaluation approach to ensure the possibility of the main requirements of these tasks in the tested neural network architectures. Established a list of optimization criteria: learning on noisy data, the use of correlated training examples need to be displayed in a training set of all aspects of the process, proportionality studies, and recognized classes, the ability to self-learning, the quality of teaching, the ability to automate learning adaptability to up learning, durations required for training volume of computing resources, memory, generalization error, calculate the duration required to detect the amount of computing resources. The importance of the criteria for the task proposed to estimate using weights that are based on expert data. Proved that to proven network architectures include: multilayer perceptron, radial basis function network, a network of Grossberg, Kohonen map, deep neural networks. For all these architectures are obtained when evaluating the feasibility of each of these criteria. An example of determining the optimal architecture of a neural network for phoneme recognition in the voice user authentication at login.

Keywords: distance learning, voice interaction, neural network, multilayer perceptron, the phoneme

Постановка проблемы

Анализ тенденций развития современных информационных систем дистанционного обучения (СДО) указывает на то, что одним из основных направлений ее усовершенствования является

расширение функциональных возможностей за счет внедрения интерактивных средств обучения, которые базируются на методах голосового взаимодействия. Соответственно, указанные методы в первую очередь должны быть направлены на

обеспечение автоматического распознавания голосовой информации в процессе ее обмена между тремя основными компонентами СДО преподавателем, учебными материалами и студентом. При этом система распознавания должна, как минимум, одновременно решать две типовые задачи:

1. В процессе компьютерного тестирования определить правильность голосового ответа.

2. Провести аутентификацию и идентификацию пользователя как при входе в систему, так и в процессе взаимодействия.

В более совершенном варианте методы голосового взаимодействия должны обеспечить интеллектуальный анализ полученной от пользователя голосовой информации. Отметим, что в распространенных СДО ("Moodle", "Прометей", "SharePointLMS", "Микротест", "ИнтраЗнание" и "Батисфера") инструментальные средства голосового взаимодействия не могут решить определенные типовые задачи. Вместе с тем, в таких информационных системах как Google+, Microsoft Office и VoiceNavigator указанные задачи решаются достаточно успешно. Соответственно [6; 9], такой успех достигается за счет использования нейросетевых методов обработки и распознавания голосовой информации.

Указанные предпосылки определяют основную проблему данной статьи – повышение эффективности СДО за счет использования нейросетевых методов обработки и распознавания голосовой информации.

Анализ последних исследований и публикаций, на которые опираются авторы

Проведенный в [7] анализ определения целесообразности применения нейросетевых методов для решения различных задач, указывает на то, что эффективность применения нейронной сети (НС) во многом зависит от того, насколько ее архитектура соответствует условиям поставленной задачи. При этом авторами не найдено методов определения оптимальной нейросетевой модели, предназначенной для решения типовых задач голосового взаимодействия в информационных системах СДО. Поэтому в качестве источника публикаций использованы работы [2-5; 8-10], которые посвящены применению НС для обработки, анализа и распознавания голосовой информации пользователей. Так, работа [5] посвящена общим вопросам построения технологий и средств распознавания и синтеза речи. В качестве модели НС, пригодной для распознавания речи и обучаемой без учителя, предлагается использовать

топографическую карту Кохонена (ТК). В работе [4] предложена возможная структура дикторозависимой системы распознавания голосовых команд, также основанная на НС типа ТК. Описаны основные этапы работы со звуковым сигналом, топология и процесс обучения предлагаемой НС. Указано, что рассматриваемая методика применения ТК может быть использована при проектировании систем голосового управления. При этом в работах [4; 5] отсутствует обоснование выбора нейросетевой модели типа ТК.

В работе [3] представлена методика распознавания ключевых команд с помощью набора НС типа многослойного персептрона (МСП), обучение которого реализовано с помощью генетических алгоритмов. Работа носит достаточно общий характер. Следует отметить недостаточно полное описание структуры МСП, в котором нет расчета количества скрытых нейронов, которое является одним из основных параметров данной нейросетевой модели. Также использование МСП предложено и в работе [8]. Отличием от [3] является обучение МСП методом обратного распространения ошибки. Предложенная структура МСП состоит из 25 входных, 1 выходного и 3 скрытых нейронов, что в соответствии с теоретическими результатами [7], не отвечает условиям применения НС. При этом показанные в [8] результаты экспериментов подтверждают перспективность использования МСП для распознавания отдельных слов в потоке слитной речи. Работа [2] посвящена разработке устойчивой к окружающим акустическим шумам дикторнезависимой системы автоматического поиска ключевых слов в потоке слитной речи. Основой предложенной системы является многослойная нейронная сеть, а для формирования первичных признаков речевого сигнала использовалось вейвлет-пакетное разложение. В работе применялись НС типа МСП, ТК и сеть Гроссберга (СГ). В работе [9] рассматривается новая концепция создания систем искусственного интеллекта на основе моделирования нейрофизиологических свойств мозга – безусловных и условных рефлексов с использованием нейроподобных растущих сетей. Предложено использовать указанные нейроподобные растущие сети для распознавания речевых команд. Работа имеет экспериментальный характер, а применение такого типа НС для распознавания речи требует дальнейших исследований. В работе [5] разработаны теоретические принципы обработки речевой информации на основе нейросетевой технологии с применением сетей радиальной базисной функции (РБФ). Показаны практические результаты, подтверждающие перспективность такой обработки.

Отметим, что в перечисленных работах [2-5; 8] нет обоснования выбора нейросетевой модели, тип которой был бы оптимизирован с точки зрения условий практического применения, а очевидные ошибки в определении параметров НС заставляют критически рассматривать положительные результаты экспериментальных исследований.

Особый интерес для анализа представила работа [10], в которой описаны нейросетевые методы распознавания речевой информации, применяемые в продуктах компании Google, являющийся на сегодня наиболее распространенными средствами распознавания речевых сигналов. В них применяются специализированные нейросетевые методы, в основе которых лежит технология машинного самообучения, нацеленная на повторение структурных и алгоритмических свойств коры головного мозга. Указанная технология получила название иерархической темпоральной памяти (Hierarchical Temporal Memory). Технология базируется на использовании так называемых глубоких нейронных сетей (ГНС). Особенностью данной технологии является иерархическая структура НС. Для обучения глубоких (скрытых) слоев применяется машина Больцмана. При этом НС используется практически на всех этапах анализа речевой информации – выделение фонем, распознавание отдельных слов и т.д. Основными достоинствами ГНС является самообучаемость, высокая точность распознавания и относительно короткое время обучения. К основным недостаткам относятся сложность реализации, относительно большое время распознавания и высокая ресурсоемкость, которая является следствием самообучения и необходимости запоминания большого количества речевых сигналов (слов, фраз). В свою очередь, самообучение также является следствием чрезвычайно большого объема обучающей выборки, которую практически невозможно представить примерами, характерными для обучения "с учителем". Отметим, что, по нашему мнению, самообучение НМ не является обязательным требованием к системам голосового взаимодействия СДО, в то же время ограниченность ресурсов сервера СДО в ряде случаев может быть критическим фактором применимости ГНС. Заметим, что реализация ГНС потребовала от Google создания самой большой в мире НС, функционирующей к тому же с использованием графических процессоров, что также вряд ли приемлемо для типовой СДО. Таким образом, использование ГНС в отечественных СДО возможно только после устранения указанных недостатков.

Цель статьи

Разработка метода определения архитектуры нейронной сети, предназначенной для решения типовых задач голосового взаимодействия в информационной системе дистанционного образования.

Изложение основного материала исследования

По аналогии с работой [7], в основу проектирования нейросетевой архитектуры положена концепция – требования поставленной задачи должны быть оптимальным образом обеспечены в НС. При этом базовую постановку задачи определения оптимального типа нейросетевой модели, предназначенной для использования при голосовом взаимодействии в информационной системе СДО, запишем в виде:

$$\{R_k(x_i)\}_K \rightarrow \max, x_i \in X, i = 1, 2, \dots, I, \quad (1)$$

где R_k – k -й критерий оптимизации; x_i – i -я архитектура НС; X – множество допустимых архитектур; I – количество допустимых архитектур; K – количество критериев оптимизации.

В соответствии с работами [2-5; 8-10], множество допустимых архитектур состоит из следующих типов НС: МСП, РБФ, СГ, ТК и ГНС. Следовательно

$$A = \{МСП, РБФ, СГ, ТК, ГНС\}. \quad (2)$$

Таким образом, количество допустимых архитектур в выражении (1) $I = 5$.

Под k -м критерием оптимизации будем понимать некоторое k -е требование системы голосового взаимодействия СДО, которое должно выполняться нейросетевой моделью. Например, может ли НС определенного типа обучаться на зашумленных учебных примерах.

Отметим, что в соответствии с [7], допустимые архитектуры НС имеют различные возможности с точки зрения обеспечения обучаемости, вычислительной мощности и ресурсоемкости в процессе принятия решения. Детализируем указанные категории.

1. С точки зрения обучаемости тип НС характеризуется:

- возможностью обучения на зашумленных учебных примерах;
- возможностью обучения на коррелируемых учебных примерах;
- необходимостью отображения в обучающей выборке всех аспектов процесса голосового взаимодействия (например, необходимо ли в учебных примерах представить все возможные слова, которые может произнести слушатель СДО);

– необходимостью пропорционального представления в обучающей выборке примеров, соответствующих различным аспектам моделируемого процесса. Например, необходимо ли в учебной выборке пропорционально представить все различаемые фонемы;

– наличием в учебных примерах ожидаемого выходного сигнала НС (возможность самообучения). Этим определяется тип обучения НС – с учителем или без учителя;

– качеством обучения, которое обычно оценивают по величине максимальной и средней ошибки распознавания учебных данных. Следует отметить, что качество обучения для НС, обучаемых методом "с учителем", традиционно считается более высоким, относительно НМ, которые самообучаются;

– возможностью автоматизации процесса обучения, которая в основном зависит от количества эмпирически настраиваемых параметров;

– приспособленностью к дообучению на новых примерах;

– продолжительностью обучения, которую можно оценить так

$$T_{o,i} = \tau_{o,i} \times L_i \times W_i \times K_{o,i} = \tau_{o,i} \times L_i \times W_i \times f_{o,i}(N), \quad (3)$$

где $T_{o,i}$ – продолжительность обучения;

$\tau_{o,i}$ – продолжительность одной учебной итерации для одной синаптической связи; W_i – количество синаптических связей; L_i – количество нейронов для сети i -о типа; $K_{o,i}$ – количество учебных итераций; $f_{o,i}(N)$ – функциональная зависимость количества учебных итераций от количества учебных примеров для сети i -о типа; N – количество учебных примеров;

– объемом вычислительных ресурсов, необходимых для реализации процесса обучения, который, в соответствии с [7], оценивают так

$$V_{o,i} = \delta_{o,i} \times L_i \times W_i \times K_{o,i} = \delta_{o,i} \times L_i \times W_i \times f_{o,i}(N), \quad (4)$$

где $V_{o,i}$ – объем вычислительных ресурсов; $\delta_{o,i}$ – объем вычислительных ресурсов для реализации одной учебной итерации для одной синаптической связи сети i -о типа.

2. Вычислительная мощность НС, которая является косвенным показателем ее интеллектуальных способностей, характеризуется:

– отношением максимального количества примеров (объем памяти), которую может запомнить сеть, к количеству синаптических связей;

– ошибкой обобщения, которая показывает правильность классификации на примерах, которые не входят в обучающую выборку.

3. Требования к ресурсоемкости НС характеризуют:

– продолжительность принятия решения НС, которая оценивается так

$$T_{f,i} = \tau_{w,i} \times W_i \times \tau_{L,i} \times L_i, \quad (5)$$

где $T_{f,i}$ – продолжительность принятия решения; $\tau_{w,i}$ – продолжительность расчета прохождения сигнала по одной синаптической связи; $\tau_{L,i}$ – продолжительность расчета выхода одного нейрона для сети i -о типа;

– объемом вычислительных ресурсов для принятия решения НС, который оценивают так

$$V_{f,i} = \delta_{w,i} \times W_i \times \delta_{L,i} \times L_i, \quad (6)$$

где $V_{f,i}$ – требуемый объем; $\delta_{w,i}$ – требуемые ресурсы для реализации одной синаптической связи; $\delta_{L,i}$ – требуемые ресурсы для реализации одного нейрона для сети i -о типа.

Заметим, что при проведении оценочных сравнительных расчетов можно считать, что для допустимых нейросетевых архитектур величины $\tau_{o,i}$, $\delta_{o,i}$, $\tau_{w,i}$, $\tau_{L,i}$, $\delta_{w,i}$, $\delta_{L,i}$ одинаковы. Следовательно, при сравнении различных нейросетевых архитектур указанные параметры в выражениях (3) – (6) можно не учитывать. После соответствующих преобразований выражения (3) – (6) следует записать так:

$$T_{o,i} \sim L_i \times W_i \times f_{o,i}(N); \quad (7)$$

$$V_{o,i} \sim L_i \times W_i \times f_{o,i}(N); \quad (8)$$

$$T_{f,i} \sim W_i \times L_i; \quad (9)$$

$$V_{f,i} \sim W_i \times L_i. \quad (10)$$

Также следует отметить, что результаты [7] указывают на наличие функциональной зависимости между $(W_i \times L_i)$ и количеством примеров, которое может запомнить НС, количеством входных и выходных параметров. Следовательно:

$$(W_i \times L_i) \sim f_i(N, K_x, K_y), \quad (11)$$

где K_x, K_y – количество входных и выходных параметров НС.

После подстановки (11) в (7-10) получим:

$$T_{o,i} \sim f_i(N, K_x, K_y) \times f_{o,i}(N); \quad (12)$$

$$V_{o,i} \sim f_i(N, K_x, K_y) \times f_{o,i}(N); \quad (13)$$

$$T_{f,i} \sim f_i(N, K_x, K_y); \quad (14)$$

$$V_{f,i} \sim f_i(N, K_x, K_y). \quad (15)$$

Отметим, что вид функций $f_i(N, K_x, K_y)$ и $f_{o,i}(N)$ непосредственно влияет на продолжительность и ресурсоемкость обучения, а также на продолжительность распознавания и требуемый при распознавании объем вычислительных ресурсов НС. При этом оценить вид этих функций возможно на основании результатов [7].

Перечень критериев, которые отвечают описанным требованиям, показан в табл. 1.

В первом приближении для i -й допустимой архитектуры значение k -о критерия равно 1, если соответствующее k -е требование полностью в ней обеспечивается и равно 0, если не обеспечивается. Для определения этих значений возможно использовать результаты [7].

При этом выражение (1) возможно переписать так

$$\sum_{k=1}^K \alpha_k \beta_k R_k(x_i) \rightarrow \max, x_i \in X, i = 1, \dots, 5, \quad (16)$$

где $\alpha_k = [0 \dots 1]$ – весовой коэффициент k -го критерия относительно поставленной задачи; $\beta_k = [0 \dots 1]$ – весовой коэффициент, который учитывает возможность обеспечения k -о требования в i -й допустимой архитектуре; $R_k(x_i)$ – полученное в первом приближении значение k -о критерия оптимизации для сети i -о типа.

Таблица 1

Критерии оптимизации

Критерий	Категория	Формулировка требования
R_1	Обучаемость	Обучаемость на зашумленных данных
R_2		Использование коррелируемых учебных примеров
R_3		Необходимостью отображения в обучающей выборке всех аспектов процесса
R_4		Пропорциональность учебных примеров и распознаваемых классов
R_5		Возможность самообучения
R_6		Качество обучения
R_7		Возможность автоматизации обучения
R_8		Приспособленность к дообучению
R_9		Продолжительность обучения
R_{10}		Требуемый для обучения объем вычислительных ресурсов
R_{11}	Вычислительная мощность	Объем памяти
R_{12}		Ошибка обобщения
R_{13}	Ресурсоемкость	Продолжительность расчета
R_{14}		Требуемый для расчета объем вычислительных ресурсов

Отметим, что значение весовых коэффициентов α_k возможно определить с помощью экспертных оценок на основании анализа поставленной задачи голосового взаимодействия СДО. Эти коэффициенты указывают на значимость k -ого требования для поставленной задачи.

Таким образом, метод определения оптимальной с точки зрения системы голосового взаимодействия СДО архитектуры НС состоит из следующих этапов:

1. С точки зрения использования НС охарактеризовать особенности поставленной задачи.

2. Для решения поставленной задачи голосового взаимодействия СДО методом экспертных оценок определить значения весовых коэффициентов α_k , которые учитывают значимость критериев, представленных в табл. 1.

3. С использованием результатов [7] для каждой из допустимых архитектур, определенных выражением (2), в первом приближении определить значения критериев R_k , представленных в табл. 1.

4. Используя результаты [7] и выражения (12-15) для каждой из допустимых архитектур, определенных выражением (2), рассчитать значения весовых коэффициентов β_k .

5. Используя выражение (7) определить оптимальную архитектуру НС.

Рассмотрим использование разработанного метода на конкретном примере определения оптимальной архитектуры НС, предназначенной для распознавания фонем в процессе голосовой авторизации пользователя СДО при входе в систему.

Этап 1. Под авторизацией пользователя будем понимать процесс проверки его имени и пароля, введенных в СДО по аудиоканалу при входе в систему.

Распознавание фонем является одним из типовых блоков распознавания речи, который предшествует блоку распознавания слов. При этом под понятием фонемы понимают основную единицу звукового строя языка, как предельный элемент, выделяемый линейным членением речи. Всего в русском языке насчитывается 43 фонемы (37 согласных и 6 гласных) [1; 2; 8]. Базируясь на результатах [2; 8], в качестве входных параметров НС, характеризующих фонему, использованы мел-кепстральные коэффициенты, являющиеся характеристиками формант речевого сигнала на стационарном участке.

Поскольку длительность одной фонемы $t_\phi = 60$ мс [1; 8], а длительность стационарного участка речевого сигнала, который является базисом для определения мел-кепстральных коэффициентов, составляет $t_{cm} = 12$ мс, то t_ϕ следует разделить на перекрывающиеся между собой стационарные участки. В соответствии с [8], принято половинное перекрытие участков. Поэтому количество

стационарных участков, соответствующих одной фонеме, равно $10 (2 \times t_{\phi} / t_{cm} = 2 \times 60 / 12 = 10)$.

Из-за того, что один стационарный участок характеризуется 24 мел-кепстральными коэффициентами, на вход НС будет подаваться всего $K_x = 240 (24 \times 10)$ таких коэффициентов.

Выходные параметры НС соответствуют распознаваемым фонемам. Поэтому количество выходных параметров K_y равно 43.

Этап 2. Определенные экспертным методом значения коэффициентов α_k показаны в табл. 2.

Таблица 2

Значения весовых коэффициентов α_k

Значение	Аргументация экспертов
$\alpha_1 = 0,1$	В СДО возможно проводить обучение НС в требуемых условиях, что сводит к минимуму зашумленность и коррелируемость учебных данных
$\alpha_2 = 0,1$	
$\alpha_3 = 0,1$	Ограниченность обучающей выборки позволяет пропорционально отобразить в ней все примеры парольных данных
$\alpha_4 = 0,1$	
$\alpha_5 = 0,1$	Поскольку парольные данные, которые предоставляются конкретному пользователю, известны СДО, то в учебных примерах НС ожидаемый выход также известен
$\alpha_6 = 1$	Качество обучения должно быть максимальным, так как это является необходимым требованием правильной классификации
$\alpha_7 = 0,9$	Возможность автоматизации обучения должна быть максимальной, так как для типовой СДО затруднительно выделить эксперта в области НС, который бы занимался настройкой процесса обучения
$\alpha_8 = 0,4$	Приспособленность к дообучению является не очень важным требованием, поскольку небольшой объем обучающей выборки позволяет достаточно быстро переобучить НС
$\alpha_9 = 0,8$	Продолжительность обучения и требуемый объем вычислительных ресурсов для обучения – это достаточно весомые требования по причине ограниченности технических средств СДО
$\alpha_{10} = 0,8$	
$\alpha_{11} = 1$	Объем памяти НС должен быть максимальным, а ошибка обобщения минимальна, так как это достаточные требования правильной классификации.
$\alpha_{12} = 1$	
$\alpha_{13} = 0,9$	Исходя из технических возможностей отечественных СДО, минимальная продолжительность расчета и минимальный требуемый объем вычислительных ресурсов являются важнейшими требованиями
$\alpha_{14} = 0,9$	

Этап 3. Определенные на основании [7] приближенные значения критериев R_k представлены в табл. 3.

Таблица 3

Значения критериев R_k

Критерий	Архитектура НС				
	МСП	РБФ	СГ	ТК	ГНС
R_1	1	0	0	1	1
R_2	1	0	0	1	1
R_3	1	0	0	1	1
R_4	0	1	1	1	0
R_5	0	0	0	1	1
R_6	1	1	1	1	1
R_7	1	1	1	1	1
R_8	0	1	1	1	1
R_9	0	1	1	1	0
R_{10}	1	1	1	1	1
R_{11}	1	1	1	1	1
R_{12}	1	1	1	1	1
R_{13}	1	1	1	1	1
R_{14}	1	1	1	1	1

Этап 4. Определенные на основании [7] с использованием (12) – (15) значения коэффициентов β_k представлены в табл. 4.

Таблица 4

Значения весовых коэффициентов β_k

Значение	Архитектура НС				
	МСП	РБФ	СГ	ТК	ГНС
β_1	1	0	0	0,8	1
β_2	1	0	0	0,8	1
β_3	1	0	0	0,8	1
β_4	0	0,9	0,8	0,8	0
β_5	0	0	0	1	1
β_6	1	1	1	0,6	0,9
β_7	1	0,7	0,7	0,8	1
β_8	0	1	1	0,8	0,8
β_9	0	1	1	1	0
β_{10}	1	0,5	0,3	0,6	0,9
β_{11}	1	0,5	0,3	0,7	0,9
β_{12}	1	0,4	0,2	0,6	0,9
β_{13}	0,8	0,9	0,9	0,8	0,7
β_{14}	0,8	0,9	0,9	0,8	0,7

Этап 5. На основании (7) рассчитана полезность каждой из допустимых архитектур НС. Результаты расчета показаны в табл. 5.

Таблица 5

Полезность допустимых архитектур

МСП	РБФ	СГ	ТК	ГНС
6,44	5,84	5,27	6,08	6,3

Анализ данных табл. 5 указывает на то, что оптимальной архитектурой является МСП.

Выводы

1. Впервые разработан метод определения архитектуры нейронной сети, предназначенной для решения типовых задач голосового взаимодействия в информационной системе дистанционного образования. Метод, базируясь на концепции соответствия требований задачи и возможностей апробированных архитектур, позволяет определить оптимальную нейросетевую архитектуру.

2. Перспективы дальнейших исследований связаны с проектированием нейросетевых архитектур, предназначенных для решения каждой из типовых задач голосового взаимодействия системы дистанционного образования.

Список литературы

1. Винцюк Т.К. Анализ, распознавание и интерпретация речевых сигналов / Т.К. Винцюк. – К: Наукова думка. – 1987. – 262 с.
2. Дмитриев В.Т. Дикторонезависимая система автоматического поиска ключевых слов в потоке слитной речи, устойчивая к акустическим шумам / В.Т. Дмитриев, И.В. Баландин // Вестник РГРТУ. – 2008. – № 2 (выпуск 24). – С. 17-22.
3. Ле Н.В. Распознавание речи на основе искусственных нейронных сетей / Н.В. Ле, Д.П. Панченко // Технические науки в России и за рубежом: материалы междунар. заоч. науч. конф. — М.: Ваш полиграф партнер, 2011. – С. 8-11.
4. Кладов С. А. Распознавание голосовых команд с помощью самоорганизующейся нейронной сети Кохонена. [Электронный ресурс]. Молодежный научно-технический вестник №05, май 2012. – Режим доступа: <http://sntbul.bmstu.ru/doc/458.html>
5. Кушнир Д.А. Исследование и разработка нейросетевых методов анализа и обработки речевого сигнала в задаче распознавания речи: дис. ... канд. техн. наук: 05.13.01 / Д.А. Кушнир – М., 2006. – 182 с.
6. Терейковська Л.О. Проблема голосової взаємодії в дистанційному навчанні вищого навчального закладу / Л.О. Терейковська, І.А.Терейковський // Управління розвитком складних систем: зб. наук. праць. – 2013. – Вип. 13. – С. 157–161.
7. Терейковський І. Нейронні мережі в засобах захисту комп'ютерної інформації / І. Терейковський. – К.: ПоліграфКонсалтинг, 2007. – 209 с.
8. Титов Ю. Н. Математическая модель органа слуха для автоматического распознавания речи / Ю. Н. Титов // Науч.-техн. вестн. СПбГУ ИТМО. – 2007. – № 37. – С. 307–310.
9. Яценко В.А. Вторичные автоматы в интеллектуальных системах / В.А.Яценко // Искусственный интеллект. – 2005. – № 3. – С.14–20.
10. Geoffrey Hinton Deep Belief Nets. [Электронный ресурс]. – Режим доступа: <http://www.cs.toronto.edu/~hinton/nipstutorial/nipstut3.pdf>.

References

1. Vintsyuk T.K. (1987). Analysis and interpretation of the speech recognition signals. Kiev, Naukova Dumka. 262.
2. Dmitriev, V.T., Balandin I.V. (2008). Speaker-independent automatic search for keywords in a stream of continuous speech, resistant to acoustic noise. RSREU, Gazette, № 2 (Issue 24), 17-22.
3. Le N.V., Panchenko D.P. (2011). Speech recognition based on artificial neural networks. Engineering science in Russia and abroad: Proceedings of International. absentia. scientific. conf. Moscow, Publishing Your partner, 8-11.
4. Hoards S.A. (2012) Recognition of voice commands using self-organizing Kohonen neural network. Youth Science and Technology Gazette, № 05, <http://sntbul.bmstu.ru/doc/458.html>
5. Kushnir D.A. (2006). Research and development of methods of analysis and neural network for speech signal processing speech recognition problem: dis. ... Cand. tehn. Sciences: 05.13.01. Moscow, 182.
6. Tereykovska L.A., Tereykovskyy I.A. (2013.) The problem of voice interaction in distance learning university. Management of Complex Systems: Coll. sciences. papers, Issue 13, 157-161.
7. Tereykovskyy I. (2007). Neural network means of information protection. Kiev, PolihrafKonsaltnh, 209.
8. Titov N. (2007). Mathematical model of the organ of hearing for automatic speech recognition. Scientific-technical. vestn. ITMO, № 37,307-310.
9. Yaschenko V.A. (2005). Secondary machines in intelligent systems. Artificial intelligence, № 3, 14-20.
10. Geoffrey Hinton.(2011) Deep Belief Nets. <http://www.cs.toronto.edu/~hinton/nipstutorial/nipstut3.pdf>

Статья поступила в редколлегию 27.01.2014

Рецензент: д-р техн. наук, проф. С.В. Цюцюра, Киевский национальный университет строительства и архитектуры, Киев.