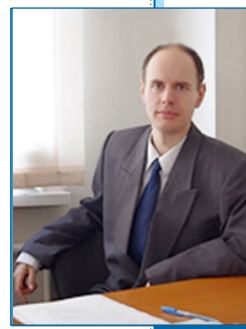


РАЗРАБОТКА НЕЙРОНЕЧЁТКОГО МЕТОДА РАСПОЗНАВАНИЯ ГЛАСНЫХ ЗВУКОВ

Фёдоров Е.Е.



Для создания интеллектуальной компьютерной системы в статье предложен метод распознавания гласных звуков на основе мер контрастностей непрерывного вейвлет-преобразования, искусственной нейронной сети, нечёткой логики и генетического алгоритма. Автором разработаны архитектура нечёткой искусственной нейронной сети и эффективные варианты операторов генетического алгоритма (репродукции, кроссинговера, мутации и редукции), которые позволяют учитывать этапы генетического алгоритма. Разработанный метод был исследован на стандартной базе TIMIT.

● *мера контрастности* ● *вейвлет-преобразование* ● *нейронная сеть*
● *нечёткая логика* ● *генетический алгоритм* ● *распознавание гласных звуков*

Vowels recognition method to create intelligent computer system is described in the article. It is based on measures of contrasting the continuous wavelet transform, artificial neural networks, fuzzy logic and genetic algorithm. The architecture of fuzzy artificial neural network and efficient operators of genetic algorithm (reproduction, crossover, mutation and reduction) have been proposed also. These operators allow for the steps of the genetic algorithm. The developed method has been tested for TIMIT database.

● *measures of contrasting* ● *wavelet transform* ● *artificial neural networks*
● *fuzzy logic* ● *genetic algorithm* ● *vowels recognition*

Общая постановка проблемы

На сегодняшний день актуальной является разработка программных компонент, предназначенных для распознавания речи человека, которые могут использоваться в интеллектуальных компьютерных системах.



Анализ исследований

Существующие методы и модели распознавания речевых образов обычно основаны на скрытых марковских моделях (СММ), алгоритме динамического программирования DTW и нейросетях и обладают одним или несколькими из перечисленных ниже недостатков [1–7]:

- длительность обучения;
- хранение большого количества эталонов звуков или слов, а также весовых коэффициентов;
- длительность распознавания;
- неудовлетворительная вероятность распознавания;
- необходимость большого количества обучающих данных;
- необходимость идентификации не только параметров, но и структуры модели (идентификация «в большом»).

С другой стороны, для решения различных задач в настоящее время активно развиваются гибридные интеллектуальные системы [8–14], использующие различные модели и методы искусственного интеллекта.

Постановка задач исследования

Для повышения вероятности распознавания речи людей, увеличения скорости обучения и автоматизации процесса определения структуры модели необходимо разработать метод, базирующийся на подходах искусственного интеллекта.

Решение задач и результаты исследований

Нейронечёткий метод распознавания гласных звуков предусматривает:

- формализацию признаков речи;
- создание структуры нейросетевой нечёткой системы распознавания гласных звуков;
- формализацию этапов создания нейросетевой нечёткой системы распознавания гласных звуков;
- построение генетического алгоритма.

1. Формализация признаков речи

После подавления шумов в речевом сигнале [15] на основе речевых характеристик формируется вектор, состоящий из мер контрастностей, впервые применённых Е.Е. Фёдоровым и Т.В. Ермоленко для преобразования и распознавания речи [16–18] и выделяемых посредством непрерывного вейвлет-преобразования с вейвлетом Морле [19].

2. Структура нейросетевой нечёткой системы распознавания гласных звуков

Нейросетевая нечёткая система распознавания гласных звуков, предложенная автором, обладает следующими достоинствами: возможность использования априорной информации (знаний экспертов); представление зна-

ний в виде правил, легко доступных для понимания человеком; возможность быстрого обучения и адаптации; параллельная обработка информации, которая повышает вычислительную мощность; отсутствуют трудности с определением структуры сети.

В качестве структуры нейросетевой нечёткой системы берётся нерекуррентная пятислойная нечёткая нейронная сеть (рис. 1) [20], ранее предложенная Е.Е. Фёдоровым и Э. Слесорайтите для распознавания зрительных образов.

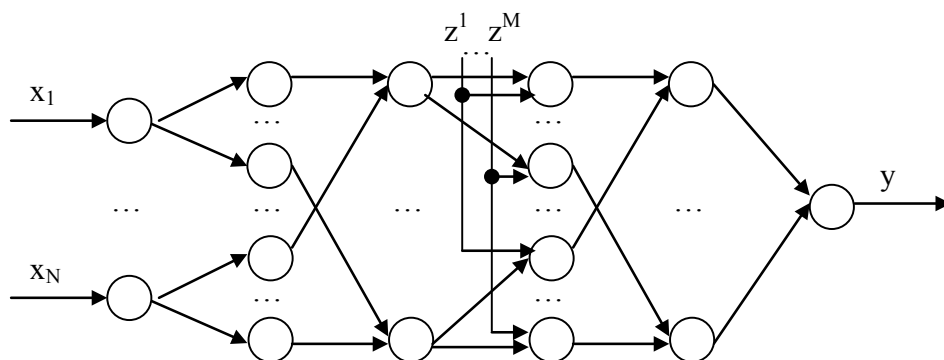


Рис. 1. Структура нейросетевой нечёткой системы

Входной (нулевой) слой содержит $N^{(0)}=N$ нейронов (соответствует количеству признаков). Первый скрытый слой реализует фаззификацию и содержит $N^{(1)}=MN$ нейронов (соответствует количеству значений лингвистических переменных). Второй скрытый слой реализует агрегирование подусловий и содержит $N^{(2)}=M$ нейронов (соответствует количеству правил M). Третий скрытый слой реализует активизацию заключений и содержит $N^{(3)}=M^2$ нейронов. Четвёртый скрытый слой реализует агрегирование заключений и содержит $N^{(4)}=M$ нейронов. Выходной слой реализует дефаззификацию и содержит $N^{(5)}=1$ нейрон.

Все весовые коэффициенты нейросети равны 1.

3. Этапы создания нейросетевой нечеткой системы распознавания гласных звуков

Создание нейросетевой нечёткой системы распознавания гласных звуков предусматривает выполнение следующих этапов:

- формирование базы нечётких правил;
- фаззификация;
- агрегирование подусловий;
- активизация заключений;
- агрегирование заключений;
- дефаззификация.

Представим j -е нечёткое правило в виде

R^j : ЕСЛИ \tilde{x}_1 есть $\tilde{\alpha}_1^j$ И ... И \tilde{x}_N есть $\tilde{\alpha}_N^j$ ТО \tilde{y} есть $\tilde{\beta}^j$, $j \in \overline{1, M}$, $i \in \overline{1, N}$,

где \tilde{y} – имя входной лингвистической переменной;



\tilde{y} – имя выходной лингвистической переменной;

$\tilde{\alpha}_i^j$ – нечёткая переменная (значение лингвистической переменной \tilde{x}_i);

$\tilde{\beta}^j$ – нечёткая переменная (значение лингвистической переменной \tilde{y});.

Нечёткое множество \tilde{A}_i^j является областью значений нечёткой переменной $\tilde{\alpha}_i^j$, нечёткое множество \tilde{B}^j является областью значений нечёткой переменной $\tilde{\beta}^j$.

На *этапе фаззификации* определим степень истинности i -го подусловия, то есть установим соответствие между входными переменными x_i j -го правила и значениями функции принадлежности $\mu_{\tilde{A}_i^j}(x_i)$. Поскольку ряд методов, связанных с распознаванием речевых образов, использует функцию Гаусса, то выберем эту функцию в качестве $\mu_{\tilde{A}_i^j}(x_i)$, то есть

$$\mu_{\tilde{A}_i^j}(x_i) = \exp\left[-\frac{1}{2}\left(\frac{x_i - m_i^j}{\sigma_i^j}\right)^2\right]$$

где m_i^j – математическое ожидание,

σ_i^j – среднеквадратичное отклонение.

На *этапе агрегирования подусловий* функция принадлежности условия для j -го правила определяется в виде

$$\mu_{\tilde{A}^j}(\bar{x}) = \mu_{\tilde{A}_1^j}(x_1) \dots \mu_{\tilde{A}_n^j}(x_n), j \in \overline{1, M}$$

На *этапе активизации заключений* функция принадлежности заключения для j -го правила определяется в виде

$$\mu_{\tilde{C}^j}(y) = \mu_{\tilde{A}^j}(\bar{x}) \mu_{\tilde{B}^j}(y), j \in \overline{1, M}, j \in \overline{1, M}$$

$$\mu_{\tilde{B}^j}(y) = \begin{cases} 0, & y \leq j - 0.5 \\ \frac{y - (j - 0.5)}{0.5}, & j - 0.5 \leq y \leq j \\ \frac{(j + 0.5) - y}{0.5}, & j \leq y \leq j + 0.5 \\ 0, & y \geq j + 0.5 \end{cases}$$

где $\mu_{\tilde{B}^j}(y)$ – треугольная функция.

На *этапе агрегирования заключений* функция принадлежности итогового заключения определяется как

$$\mu_{\tilde{C}}(y) = \max(\mu_{\tilde{C}^1}(y), \dots, \mu_{\tilde{C}^M}(y))$$

На *этапе дефаззификации* для получения номера класса используется метод максимума функции принадлежности

$$y = \arg \max_{z^j} \mu_{\tilde{C}}(z^j), j \in \overline{1, M},$$

где z^j – центр нечёткого множества \tilde{C}^j .

Таким образом, математическую модель нейронечёткой системы распознавания гласных звуков (рис. 1) можно представить в виде

$$y = \arg \max_{z^k} \max_{j \in \overline{1, M}} \mu_{\tilde{B}^j}(z^k) \prod_{i=1}^N \mu_{\tilde{A}^i}(x_i), k \in \overline{1, M}.$$

4. Этапы построения генетического алгоритма

Построение генетического алгоритма предусматривает выполнение следующих этапов [21–25]:

- создание исходной популяции;
- определение фитнес-функции;
- задание оператора репродукции (селекции);
- задание оператора кроссинговера;
- задание оператора мутации;
- задание оператора редукции;
- определение условия останова.

На *этапе создания исходной популяции* выбраны вещественные гены в силу следующих причин:

- возможность поиска в больших пространствах, что трудно делать в случае двоичных генов, когда увеличение пространства поиска сокращает точность решения при неизменной длине хромосомы;
- способность к локальной настройке решений;
- отсутствие операций кодирования/декодирования, которые необходимы для двоичных генов, повышает скорость работы алгоритма;
- близость к постановке большинства прикладных задач (каждый вещественный ген отвечает за одну переменную или параметр).

В качестве хромосомы, которая представляет i -ю особь популяции $H = \{h_i\}$, выступает упорядоченный вектор параметров (математических ожиданий и среднеквадратичных отклонений)

$$h_i = (lx_1^1 + i * \Delta m_1^1, lx_1^2 + i * \Delta m_1^2, \dots, lx_n^1 + i * \Delta m_n^1, lx_n^2 + i * \Delta m_n^2, \\ lx_1^1 + i * \Delta \sigma_1^1, lx_1^2 + i * \Delta \sigma_1^2, \dots, lx_n^1 + i * \Delta \sigma_n^1, lx_n^2 + i * \Delta \sigma_n^2), i \in \overline{1, |H|}$$

$$\Delta m_k^j = \frac{rx_k^j - lx_k^j}{|H|}, \Delta \sigma_k^j = \frac{rx_k^j - lx_k^j}{|H|}, j \in \overline{1, M}$$

где $|H|$ – мощность популяции,

lx_k^j, rx_k^j – левая и правая границы значений k -го признака (гена), вычисленные экспериментально.



На *этапе определения фитнес-функции* предложена следующая фитнес-функция, соответствующая вероятности правильного распознавания:

$$F = \frac{1}{P} \sum_{p=1}^P I(y_p - d_p) \rightarrow \max_{\alpha, \sigma} I(\alpha) = \begin{cases} 1, & \alpha = 0 \\ 0, & \text{иначе} \end{cases}$$

где d_p – тестовый результат распознавания,

y_p – результат распознавания, полученный по нейросети,

P – количество тестовых речевых образцов.

На *этапе задания оператора репродукции (селекции)* для отбора векторов параметров для скрещивания и мутации в качестве оператора репродукции в статье предложена комбинация случайного и линейно упорядоченного отбора

$$P(h_i) = \frac{1}{|H|} \left(1 - \frac{t}{T}\right) + \frac{1}{|H|} \left(a - (2a - 2) \frac{i-1}{|H|-1}\right) \left(\frac{t}{T}\right),$$

где t – номер итерации,

T – максимальное количество итераций.

Таким образом, на ранних стадиях работы генетического алгоритма используется равновероятный отбор, обеспечивающий исследование всего пространства поиска (случайный выбор хромосом), а на заключительных стадиях используется линейно упорядоченный отбор, делающий поиск направленным (текущие лучшие хромосомы сохраняются). Эта комбинация не требует масштабирования и может использоваться при минимизации фитнес-функции.

Следует отметить, что автор не использует для вычисления вероятностей $P(h_i)$ имитацию отжига, поскольку она требует эмпирического определения значений дополнительных параметров.

На *этапе задания оператора кроссинговера* для комбинирования двух вариантов вектора параметров, отобранных оператором репродукции, в качестве оператора кроссинговера используется однородный (равновероятный) кроссинговер (гены обоих родителей выбираются с равной вероятностью). Для выбора родителей в статье предложена следующая комбинация – на ранних стадиях работы генетического алгоритма используется аутбридинг (первый родитель выбирается случайно, а второй – как максимально далёкий от первого либо как находящийся от первого на расстоянии большем, чем заданное ε , в качестве расстояния между родителями используется расстояние Манхэттена), обеспечивающий исследование всего пространства поиска, а на заключительных стадиях используется инбридинг (первый родитель выбирается случайно, а второй – как максимально близкий к первому либо как находящийся от первого на расстоянии меньшем, чем заданное ε , в качестве расстояния между родителями используется расстояние Манхэттена), делающий поиск направленным. Эта комбинация не требует масштабирования и может использоваться при минимизации фитнес-функции. После выбора родителей осуществляется скрещивание, и производятся два потомка.

Вероятность кроссинговера над выбранными родителями на основе аутбридинга определена в виде

$$P_c^{out} = P_{c0} \left(1 - \frac{t}{T} \right),$$

где t – номер итерации,

T – максимальное количество итераций,

P_{c0} – начальная вероятность кроссинговера.

Вероятность кроссинговера над выбранными родителями на основе инбридинга определена посредством имитации отжига в виде

$$P_c^{in} = P_{c0} \left(\frac{t}{T} \right).$$

Таким образом, при $P_c^{out} > P_c^{in}$ кроссинговер над выбранными родителями выполняется на основе аутбридинга, иначе на основе инбридинга.

Следует отметить, что автор не использует для вычисления вероятностей P_c^{out} , P_c^{in} имитацию отжига, поскольку она требует эмпирического определения значений дополнительных параметров.

Для глобального поиска оптимального вектора параметров необходимо повысить разнообразие вариантов.

На *этапе задания оператора мутации* для обеспечения разнообразия вариантов вектора параметров после кроссинговера используется неоднородная мутация.

Шаг мутации Δ определён в виде

$$r = rand(), \lambda = rand(), \Delta = \begin{cases} (Max_j - h_j) r \left(1 - \frac{t}{T} \right), & \lambda < 0.5 \\ -(h_j - Min_j) r \left(1 - \frac{t}{T} \right)^2, & \lambda \geq 0.5 \end{cases},$$

где $rand()$ – функция, возвращающая равномерно распределённое случайное число в диапазоне $[0,1]$,

Max_j , Min_j – максимальное и минимальное значение j -го гена,

t – номер итерации,

T – максимальное количество итераций.

Вероятность мутации определена в виде

$$P_m = P_{m0} \left(1 - \frac{t}{T} \right),$$

где P_{m0} – начальная вероятность мутации.

Таким образом, на ранних стадиях работы генетического алгоритма с высокой вероятностью происходит мутация с большим шагом, что обеспечивает исследование всего пространства поиска, а на заключительных стадиях вероятность мутации и её шаг стремятся к нулю, что делает поиск направленным. Для формирования новой популяции на основе предыдущей популяции и векторов параметров, полученных путём кроссинговера и мутации, используется оператор редукции.



Следует отметить, что автор не использует для вычисления вероятности $P_{т0}$ имитацию отжига, поскольку она требует эмпирического определения значений дополнительных параметров.

На *этапе задания оператора редукции* в качестве оператора редукции в статье используется схема $(\mu + \lambda)$ (селекционная схема) (особи предыдущей популяции численностью μ и потомки численностью λ объединяются и упорядочиваются по значению фитнес-функции, и в новую популяцию отбирается μ первых лучших особей). Эта схема не требует масштабирования и может использоваться при минимизации фитнес-функции.

На *этапе определения условия останова* предлагается условие $t \geq T$.

5. Численное исследование нейронечёткого метода распознавания гласных звуков

В табл. 1 приведено сравнение предложенного метода и существующих нейросетевых методов на основе базы данных ТИМТ. Распознаванию подлежали все гласные звуки. Ошибка распознавания представляет собой отношение количества правильно распознанных фреймов, содержащих звуки (длина каждого фрейма 512 при частоте дискретизации 22050 Гц), к их общему количеству в процентах, при этом фрейм, содержащий конец первого звука и начало второго звука, не учитывался. Ошибка распознавания гласных звуков, приведённая в табл.1, оказалась примерно сопоставима с ошибкой распознавания согласных звуков.

Таблица 1

Оценка нейросетевых методов распознавания

Приведённые в табл.1 стандартные нейронные сети были реализованы автором статьи посредством функций Neural Networks Toolbox и Bioinformatics Toolbox пакета Matlab. Исследование позволяет сделать вывод, что авторский метод обеспечивает высокую вероятность распознавания.

Нейронные сети	Ошибка распознавания (%)
Машина опорных векторов	15
Самоорганизующаяся карта признаков	12
Многослойный персептрон	10
Нейронная сеть на основе радиально-базисных функций	10
Вероятностная нейронная сеть	8
Авторская	4

Выводы

Новизна. В статье предложен метод распознавания гласных звуков, базирующийся на методах цифровой обработки сигнала, распознавании речевых обра-

зов, нейронных сетях, нечёткой логике и генетическом алгоритме. Этот метод был численно исследован на стандартной базе ТИМТ и сравнен с известными нейросетевыми методами. Разработанная нейронечёткая система распознавания гласных звуков обладает следующими достоинствами: возможность использования априорной информации (знаний экспертов); представление знаний в виде правил, легко доступных для понимания человеком; возможность быстрого обучения и адаптации; параллельная обработка информации, которая повышает вычислительную мощность; отсутствие трудностей с определением структуры системы; высокая вероятность распознавания. Предложены эффективные варианты операторов репродукции, кроссинговера и редукции, учитывающие номер итерации генетического алгоритма и повышающие эффективность обучения системы распознавания.

Практическое значение. Основные положения данной работы предназначены для реализации в интеллектуальных компьютерных системах распознавания речи человека.

Литература

1. *Осовский С.* Нейронные сети для обработки информации / С. Осовский. — М.: Финансы и статистика, 2002. — 344 с.
2. *Хайкин С.* Нейронные сети: полный курс / С. Хайкин. — М.: Издательский дом «Вильямс», 2006. — 1104 с.
3. *Каллан Р.* Основные концепции нейронных сетей / Р. Каллан. — М.: Издательский дом «Вильямс», 2001. — 288 с.
4. *Rabiner L.R.* Fundamentals of speech recognition / L.R. Rabiner, B.H. Jang. — Englewood Cliffs, NJ: Prentice Hall PTR, 1993. — 507 p.
5. *Потапова Р.К.* Речь: коммуникация, информация, кибернетика / Р.К. Потапова. М.: Радио и Связь, 1997. — 528 с.
6. *Винцюк Т.К.* Анализ, распознавание и интерпретация речевых сигналов / Т.К. Винцюк. — К.: Наук. думка, 1987. — 261 с.
7. *Фёдоров Е.Е.* Методология создания мультиагентной системы речевого управления: монография / Е.Е. Фёдоров. — Донецк: изд-во «Ноулидж», 2011. — 356 с.
8. *Комарцова Л.Г.* Нейрокомпьютеры / Л.Г. Комарцова, А.В. Максимов. — М.: Изд-во МГТУ им. Н.Э. Батмана, 2002. — 320 с.
9. *Intelligent Hybrid Systems: Fuzzy Logic, Neural Networks, and Genetic Algorithm / Ed. Da Ruan.* — Boston: Kluwer Academic Publishers, 1997. — 258 p.
10. *Леоненков А.В.* Нечёткое моделирование в среде MATLAB и fuzzyTECH / А.В. Леоненков. — СПб.: БХВ-Петербург, 2005. — 736 с.
11. *Рутковская Д.* Нейронные сети, генетические алгоритмы и нечеткие системы / Д. Рутковская, М. Пилиньский, Л. Рутковский. — М.: Горячая линия — Телеком, 2006. — 452 с.
12. *Круглов В.В.* Искусственные нейронные сети. Теория и практика / В.В. Круглов, В.В. Борисов. — М.: Горячая линия — Телеком, 2002. — 382 с.
13. *Штовба С.Д.* Проектирование нечётких систем средствами MATLAB / С.Д. Штовба. — М.: Горячая линия — Телеком, 2007. — 288 с.
14. *Фёдоров Е.Е.* Методики интеллектуальной диагностики: монография / Е.Е. Фёдоров. — Донецк: изд-во «Ноулидж», 2010. — 303 с.
15. *Ермоленко Т.В.* Методика подавления цветных шумов в речевом сигнале / Т.В. Ермоленко, Е.Е. Фёдоров // Речевые технологии. — 2009. — № 3. — С. 3-13.
16. *Фёдоров Е.Е.* Модели и методы преобразования речевых сигналов: монография / Е.Е. Фёдоров. — Донецк: изд-во «Норд-Пресс», 2006. — 260 с.
17. *Ермоленко Т.В.* Исследование признаков, используемых для фонемного распознавания, с помощью нейросети // Искусственный интеллект. — 2007. — № 4. — С. 357–363.



18. Фёдоров Е.Е. Методика формирования акустических характеристик эталонов речи: монография / Е.Е. Фёдоров. — Донецк: изд-во «Вебер», 2008. — 282 с.
19. Малла С. Вэйвлеты в обработке сигналов / С. Малла. — М.: Мир, 2005. — 671 с.
20. Фёдоров Е.Е. Модели и методы распознавания зрительных образов / Е.Е. Фёдоров, Э. Слесорайтите. — Донецк: Изд-во «Ноулидж» (Донецкое отделение), 2013. — 422 с.
21. Evolutionary Computation 1. Basic Algorithms and Operators / Ed. by T. Back, D. B. Fogel, Z. Michalewicz. — Bristol: IOP Publishing Ltd., 2000. — 377 p.
22. Evolutionary Computation 2. Advanced Algorithms and Operators / Ed. by T. Back, D. B. Fogel, Z. Michalewicz. — Bristol: IOP Publishing Ltd., 2000. — 304 p.
23. Yu X. Introduction to evolutionary algorithms / X. Yu, M. Gen. — London: Springer-Verlag, 2010. — 433 p.
24. Скобцов Ю.А. Основы эволюционных вычислений / Ю.А. Скобцов. — Донецк: ДонНТУ, 2008. — 326 с.
25. Субботін С.О. Нейтеративні, еволюційні та мультиагентні методи синтезу: монографія / С.О. Субботін, А.О. Олійник, О.О. Олійник. — Запоріжжя: ЗНТУ, 2009. — 376 с.

Сведения об авторе:

Фёдоров Евгений Евгеньевич,

доктор технических наук, профессор кафедры автоматизированных систем управления Донецкого национального технического университета, доцент. В 2012 году защитил докторскую диссертацию в Национальном авиационном университете г. Киева. Автор свыше 110 научных публикаций, в том числе 10 монографий, посвящённых: моделям и методам преобразования и распознавания речевых образов; моделям и методам преобразования и распознавания зрительных образов; моделям и методам анализа и синтеза естественно-языковых объектов; моделям и методам вибрационной и акустической диагностики; интеллектуальным технологиям в логистике, метаэвристикам. Основная область интересов: идентификация и верификация диктора, распознавание и синтез речи, анализ и синтез естественно-языковых объектов, распознавание лица человека, интеллектуальная вибродиагностика и шумодиагностика, медицинская интеллектуальная диагностика (анализ биосигнала), интеллектуальные технологии в транспортной логистике (поиск оптимального маршрута и мультиагентное взаимодействие), оптимизация числовых функций и комбинаторная оптимизация на основе метаэвристик.