

РЕАЛИЗАЦИЯ МЕТОДА НЕЧЁТКОГО СОПОСТАВЛЕНИЯ РЕЧЕВЫХ ОБРАЗОВ В НЕЙРОСЕТЕВОМ БАЗИСЕ

Рассматривается применение метода нечёткого сопоставления образов к решению задачи автоматического распознавания изолированных слов и вопросы аппаратной реализации метода в нейросетевом базисе для ускорения процесса распознавания.

Речевой интерфейс, как более естественный для человека, приобретает всё большую востребованность в современных человеко-машинных системах. В основе его построения лежит задача распознавания речи, для решения которой, несмотря на множество предложенных способов, не найден приемлемый метод [1]. Одним из перспективных подходов к решению подобных задач распознавания и классификации является использование аппарата нечётких множеств [2]. Практическое применение этого метода в интеллектуальном интерфейсе требует его выполнения в реальном времени. Ускорение процесса распознавания может быть достигнуто путём аппаратной реализации применяемой методики. При этом необходимо выбрать рациональную структуру вычислительной системы и элементную базу для её реализации. Поэтому применить эффективный метод и оптимально реализовать его аппаратно является важнейшей задачей на пути создания речевых интерфейсов.

Поиск эффективного метода распознавания речи продолжается до настоящего времени ввиду нетривиальности задачи из-за нестабильности речевого сигнала по многим параметрам. Намечились основные направления, базирующиеся на вероятностном, метрическом и нейросетевом подходах [3]. Сравнительный анализ, проведенный в [3], показывает достоинства и недостатки этих подходов. Отмечается, что наименьший процент ошибок наблюдается при использовании искусственных нейронных сетей. Также перспективен для решения трудноформализуемых задач, к которым относится задача распознавания речи, подход на основе нечёткой логики [4]. В сборнике японских авторов [2] описан метод нечёткого сопоставления образов и приведена высокая оценка его эффективности в распознавании английских, немецких и японских слов. Однако, вопросы применения данного метода к распознаванию русскоязычных слов и аппаратной реализации вычислительного процесса не рассмотрены.

Данная статья посвящена анализу эффективности метода нечёткого сопоставления при распознавании изолированных русских слов и аппаратного ускорения процесса распознавания. Идея аппаратного ускорения базируется на высокой эффективности реализации в нейросетевом базисе матричных и векторных операций, составляющих основу вычислительного процесса.

Речевой сигнал представляется в виде двумерного спектрального временного образа (СВО), получаемого с помощью оконного преобразования Фурье (рис.1а). Такой образ отражает изменение по времени амплитуд заданных частотных составляющих речевого сигнала и хорошо выражает особенности речи, что даёт возможность его использовать для автоматического распознавания произносимых слов [2]. СВО позволяет выделить местоположение резонансных частот, т.е. локальных выбросов, что является определяющей особенностью речевого сигнала [2]. На этом основании СВО можно преобразовать к двоичному виду, не теряя указанных информативных признаков речи, с помощью следующей замены: 1 – на месте локального выброса, 0 – в других местах. Полученный образ называют двоичным спектральным временным образом (ДСВО) и используют его как отражение особенностей речевого сигнала (рис. 1б).

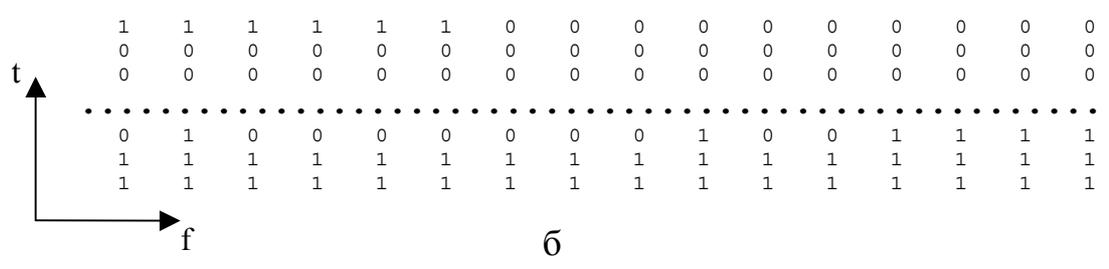
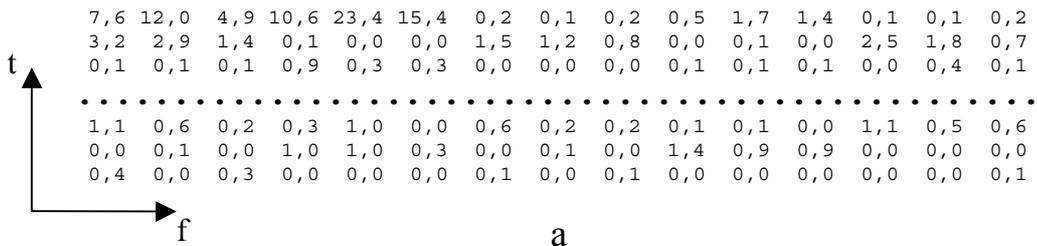


Рисунок 1 – Пример спектрально-временного представления слова «пять»: а – СВО; б – ДСВО

В качестве единиц речи рассматриваются слова, набор которых определяет словарный состав речевого общения. Для распознавания изолированных слов, произносимых в реальном времени, применялся метод нечёткого сопоставления с эталоном [2]. Эталонные образы для каждого слова словаря формировались как среднее арифметическое ДСВО различных вариантов произношения данного слова и затем подвергались нормализации. В результате формируется бинарное нечёткое отношение между множеством F (номеров частот f) и множеством T (номеров временных интервалов t) в виде:

$$f \in F, t \in T: F R T,$$

где R – нечёткое отношение, которое ставит каждой паре элементов $(f, t) \in F * T$ величину функции принадлежности $\mu_R(x, y) \in [0, 1]$ [4].

Обозначим число записанных слов через n, множество слов через $I = \{i_1, i_2, \dots, i_n\}$ и множество нечётких отношений, характерных для каждого слова, через $R = \{r_1, r_2, \dots, r_n\}$. Входной неизвестный образ y рассматривается как обычное (чёткое) отношение между множеством частот и множеством временных интервалов. Для него вычисляются степени сходства S_j с каждым нечётким отношением r_j . Результатом распознавания является слово j, такое, что

$$j = \max_{j \in I} \{S_j\}. \tag{1}$$

Степень подобия вычисляется по следующей формуле:

$$S_j = \frac{D_j}{\overline{D_j}}, \tag{2}$$

где $D_j = \int r(f, t) \wedge y(f, t) df dt,$
 $\overline{D_j} = \int \overline{r(f, t)} \wedge y(f, t) df dt.$

В дискретном случае имеет место

$$D_j = \sum_t \sum_f r(f, t) \wedge y(f, t) = \sum_t \sum_f a(f, t), \quad (3)$$

$$\overline{D_j} = \sum_t \sum_f \overline{r(f, t)} \wedge y(f, t) = \sum_t \sum_f b(f, t). \quad (4)$$

Система распознавания, основанная на методе нечёткого сопоставления (без акустического модуля, реализующего чтение речевого сигнала из файла и преобразования его в ДСВО), сначала тестировалась на множестве условных ДСВО, представляющих собой изображения цифр от одного до пяти с добавлением шума (рис.2).

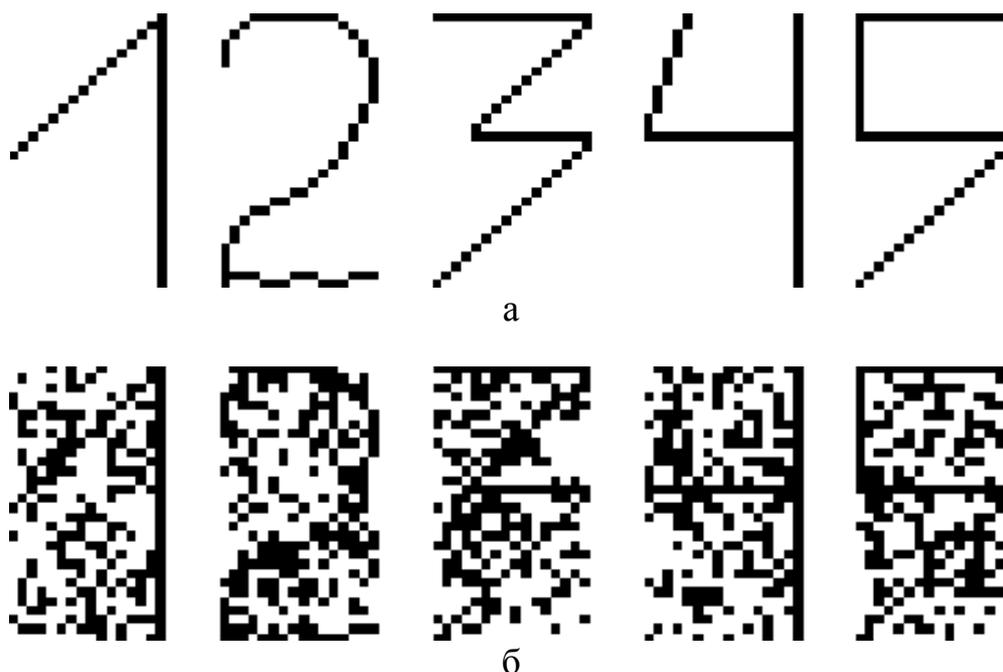


Рисунок 2 – Примеры графических изображений матриц ДСВО, использованных для тестирования метода нечёткого сопоставления: а – незашумлённые образы, б – зашумлённые образы

Система распознавания правильно распознала все образы тестового множества. Результаты представлены в табл. 1.

Таблица 1 – Степени подобия образа «один» словарным образам

Название образа	один	два	три	четыре	пять
Степени подобия незашумлённого образа	∞	0,0624	0,0512	1,6802	0,0241
Степени подобия зашумлённого образа	3,3289	0,4259	0,5887	1,0396	0,3653

После подключения к системе акустического модуля был поставлен эксперимент по распознаванию речевых образов числительных русского языка от одного до пяти. В обучающее множество входило 75 вариантов произношения этих слов (по 15 на каждое слово), в тестовое множество – 25 вариантов (по 5 на каждое слово). В результате система правильно распознала 19 слов, т.е. качество распознавания составило 76%.

В данной работе предлагается реализовать вышеприведённые вычисления в нейросетевом базисе, что позволит значительно ускорить процесс распознавания. Общая схема нейронной сети, решающей эту задачу, приведена на рисунке 3.

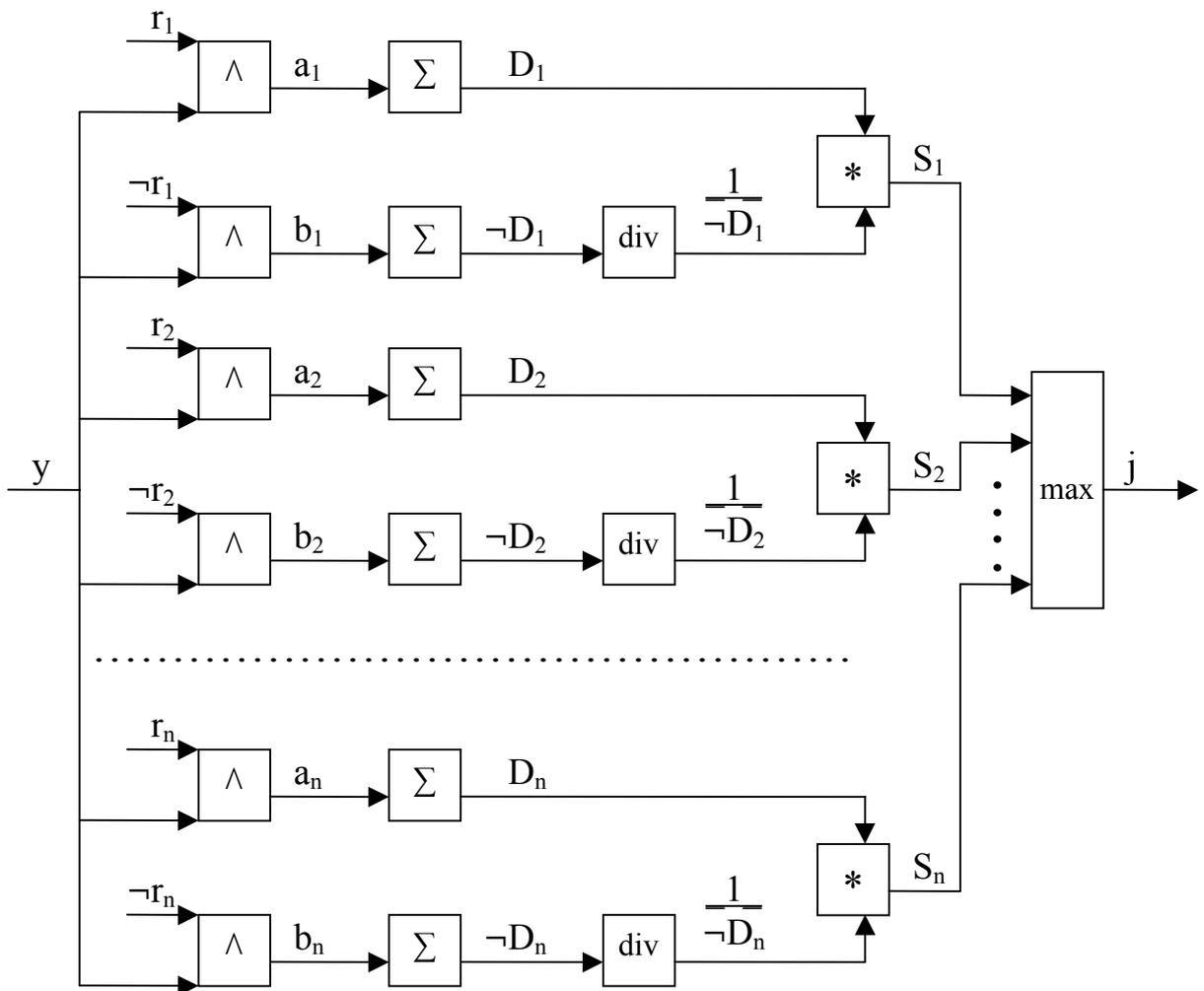


Рисунок 3 – Схема реализации метода нечёткого сопоставления образов с помощью нейросетевых блоков

Блоки Λ составляют слой из нечётких двухвходовых нейронов «И» [4]. Входные сигналы и веса этих нейронов объединяются с помощью t-конормы, в данном случае – операции \max , а выход образуется с помощью t-нормы, в данном случае – операции \min . Каждый нейрон этого слоя реализует операцию \min для соответствующих элементов входного ДСВО y и матрицы нечёткого отношения r_j , поэтому весовые коэффициенты этих нейронов имеют нулевые значения. Количество нейронов в каждом блоке слоя равно количеству элементов в соответствующей матрице нечёткого отношения, а количество нейронов во всём слое – количеству элементов во всех матрицах нечётких отношений.

Блоки Σ составляют слой из обычных нейронов с линейной функцией активации. Каждый блок представляет собой многовходовой нейрон, входами которого являются выходы нечётких нейронов соответствующего блока Λ . Каждый нечёткий нейрон из данного слоя реализует операцию суммирования выходов нечётких нейронов по формуле (3), а каждый чётный – по формуле (4). Для этого веса всех синапсов, соединяющих нейроны нечёткого слоя с нейронами суммирующего слоя, устанавливаются равными 1.

Каждый из блоков деления (div) представляет собой трёхслойную нейросеть для вычисления обратной функции $f(x) = 1/x$ [5], на вход которой подаётся результат

вычисления $-D_j$ из соответствующего суммирующего блока, а на выходе получается значение $1/-D_j$.

Блоки умножения (*) представляют собой одиночные одноходовые нейроны с линейной функцией активации. На вход нейрона подаётся значение D_j , в качестве весового коэффициента используется значение $1/-D_j$, а на выходе формируется произведение $D_j * 1/-D_j$ в соответствии с формулой (2).

Блок \max предназначен для поиска максимума и выполнен в виде трёхслойной нейросети [6]. Количество входов этой нейросети равно количеству слов словаря, а на выходе получается номер распознанного слова, вычисленный в соответствии с формулой (1).

Для этой схемы проведен анализ временной сложности и выполнено сравнение с временной сложностью последовательного алгоритма. Пусть размер всех матриц нечётких отношений равен $N * 16$, а количество слов в словаре и, соответственно, количество нечётких отношений составляет M . Количество входов нейросетевой структуры, изображённой на рис.3, равно $M * N$. Тогда временная сложность вычислительного процесса в такой нейросетевой структуре составляет 8 умножений, $32 * N + K + 2$ сложений (где K – число нейронов, аппроксимирующих обратную функцию) и 2 операции сравнения (нахождение максимального и минимального из двух значений). Временная сложность этого же вычислительного процесса, если он будет выполняться последовательно, составит $M * (32 * N + 1) - 1$ сравнений, $30 * M * (N - 1)$ сложений и M делений. Сравнение показывает, что нейросетевая реализация за счёт распараллеливания даёт выигрыш приблизительно в $2 * M$ раз, т.е. преимущество в быстродействии растёт линейно с увеличением объёма словаря.

В результате исследований установлено, что метод нечёткого сопоставления хорошо распознаёт даже сильно зашумлённые образы, если они не деформированы. Однако, при распознавании речевых образов русского языка, которые характеризуются сильной изменчивостью по структуре расположения локальных выбросов, метод дал худшие показатели. Очевидно, выделение информативных признаков речевого сигнала, используемое в данной работе, не адекватно отражает этот сигнал. Планируется улучшить методику предварительной обработки сигнала с целью выделения более значимых информативных признаков изолированных слов, что позволит с помощью параллельных нейросетевых вычислений создавать удовлетворительные системы речевого интерфейса.

Литература:

1. Колоколов А.С. Предварительная обработка сигнала для распознавания речи // Автоматика и телемеханика. 2002. №3. С. 160 – 168.
2. Прикладные нечёткие системы / Под ред. Т.Тэрано, К.Асаи, М.Сугено. – М.: Мир, 1993. – 287 с.
3. Жирков А.О., Корчагин Д.Н., Лукин А.С., Крылов А.С., Баяковский Ю.М. Графический метод представления и нейросетевое распознавание частотно-временных векторов речевой информации // Программирование. 2003. №4. С. 41 – 52.
4. Круглов В.В., Борисов В.В. Искусственные нейронные сети. Теория и практика. – М.: Горячая линия – Телеком, 2001. – 382 с.
5. Галушкин А.И., Судариков В.А., Шабанов Е.В. Нейроматематика: методы решения задач на нейрокомпьютерах // Математическое моделирование. 1991. Том 3. № 8. С.93 – 111.
6. Галушкин А.И. О решении задач сортировки с использованием нейронных сетей // Нейрокомпьютер. 1994. № 3, 4. С.35 – 40.