

МАТЕМАТИКА

*Сергей Михайлович ДИДЕНКО —
аспирант кафедры информационных систем*

УДК 004.93

ИССЛЕДОВАНИЕ МОДЕЛИ ДИНАМИКИ ПАРАМЕТРОВ ИНФОРМАЦИОННОГО ПОЧЕРКА ПОЛЬЗОВАТЕЛЯ

АННОТАЦИЯ. Приводятся результаты исследования динамики параметров манипулятора «мышь» и оценки распознаваемости по ним информационного почерка пользователя.

In the article published results of the research dynamics «mouse» manipulator parameters and authentication by mouse movements.

Введение. Информатизация общества является причиной все большей виртуализации образа жизни человека. Это объясняется ростом количества операций, производимых с использованием компьютерных сетей, сетевых ресурсов с целью проведения различных операций (передача информации, торговля, обучение, тестирование, анкетирование). Через 10-15 лет подавляющее большинство услуг высшего образования будет реализовываться дистанционно.

Перспектива развития информационных систем связана, в частности, с появлением аппаратно-программных средств поддержки деятельности конкретного пользователя с учетом его индивидуальных особенностей. В этом контексте одной из важных задач выступает мониторинг пользователя посредством отслеживания динамики его работы с устройствами ввода.

Достаточно новым способом биометрической верификации по поведенческим признакам является отслеживание характера взаимодействия человека с ЭВМ посредством различных манипуляторов, например световых перьев, манипуляторов типа «мышь», клавиатуры, джойстика и т. д. В связи с этим появилось понятие «информационный почерк пользователя» (ИПП), который отражает стиль работы пользователя с устройствами ввода [1]. Известны исследования «клавиатурного почерка» [2].

В настоящей работе рассматривается задача отслеживания характера работы пользователя с манипулятором «мышь» с точки зрения его использования как существенного компонента информационного почерка.

1. Об аналитическом описании динамики мыши

Математическая модель динамики D устройств ввода (клавиатура, мышь, джойстик и пр.) в общем виде может быть представлена как совокупность множеств реализаций случайных процессов на интервале наблюдения $[t_1, t_2]$:

$$D(t_1, t_2) = (\{\alpha_i(\tau)\}, \{\beta_i(\tau)\}, \{\gamma_i(\tau)\}, \{\eta_i(\tau)\}, \dots), \quad \tau[t_1, t_2], \quad (1)$$

где функции времени — реализация процессов, отражающих динамику конкретного устройства ввода (клавиатуры, мыши и т. д.);

$i = 1, \dots, I; j = 1, \dots, J; k = 1, \dots, K; l = 1, \dots, L;$ и т. д. I, J, K, L, \dots — в общем случае дискретные случайные величины в Z^+ .

Более конкретный и в то же время простой вид этой модели не представляется возможным. Дискретный процесс нажатия клавиш — случайный нестационарный поток событий, неоднородный как по количеству событий в реализации, так и по вероятностным свойствам интервала времени между событиями. Разнообразие траекторий указателя монитора (просмотрено около 50000 реализаций), отражающих движение мыши, настолько велико, что нет возможности предложить единую аналитическую структуру соответствующих кривых. Попытка аппроксимировать полиномом эти траектории привела к большому разбросу значений оцениваемых коэффициентов. Аналогична картина с другими средствами ввода информации в компьютер.

В связи с этим для распознавания ИПП нужно в идеале в качестве эталона брать всю совокупность реализаций процессов в начальный отрезок времени работы пользователя и в дальнейших измерениях сопоставлять в определенной метрике эталон с наблюдаемыми реализациями, уточняя компьютерный образ модели.

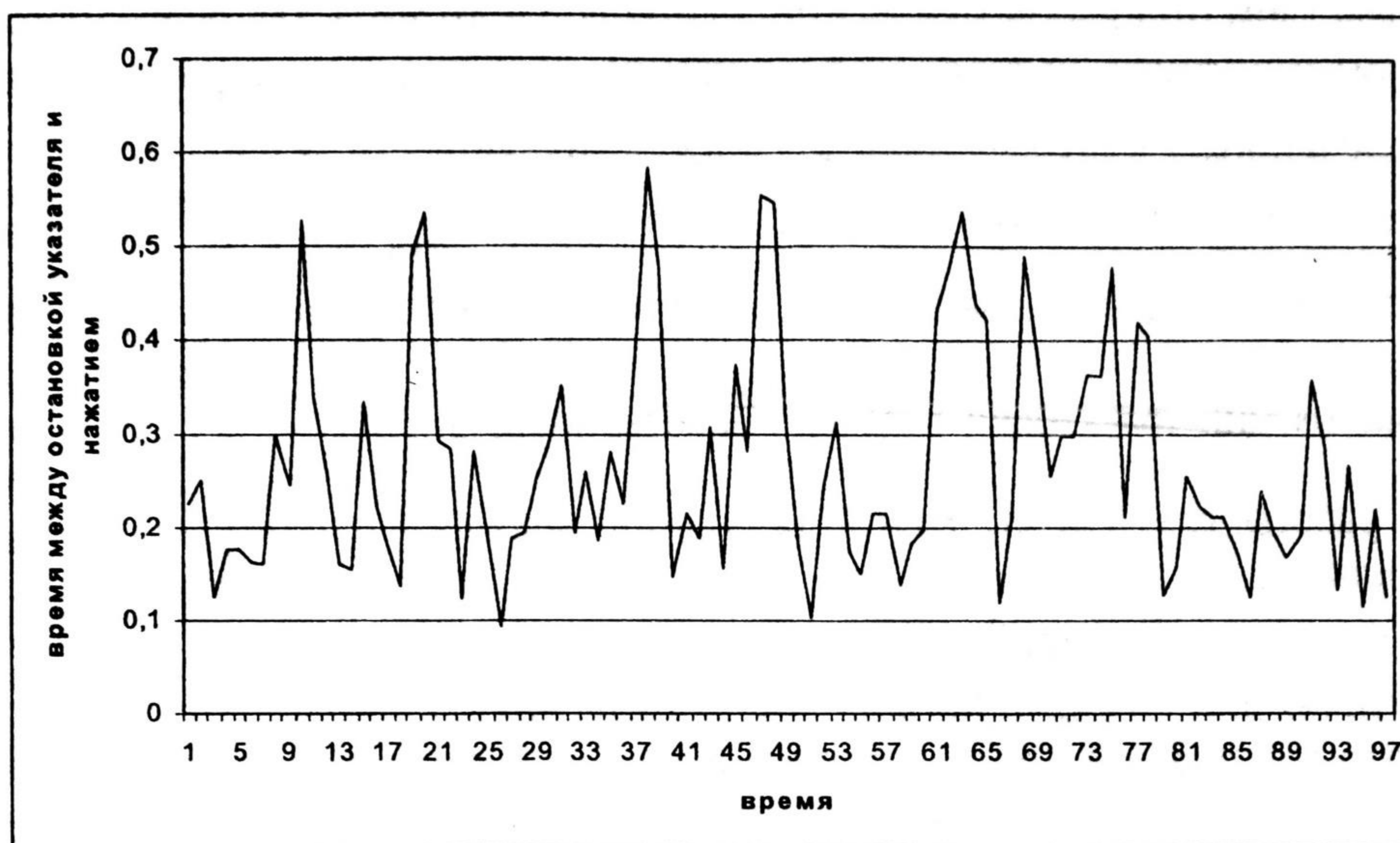
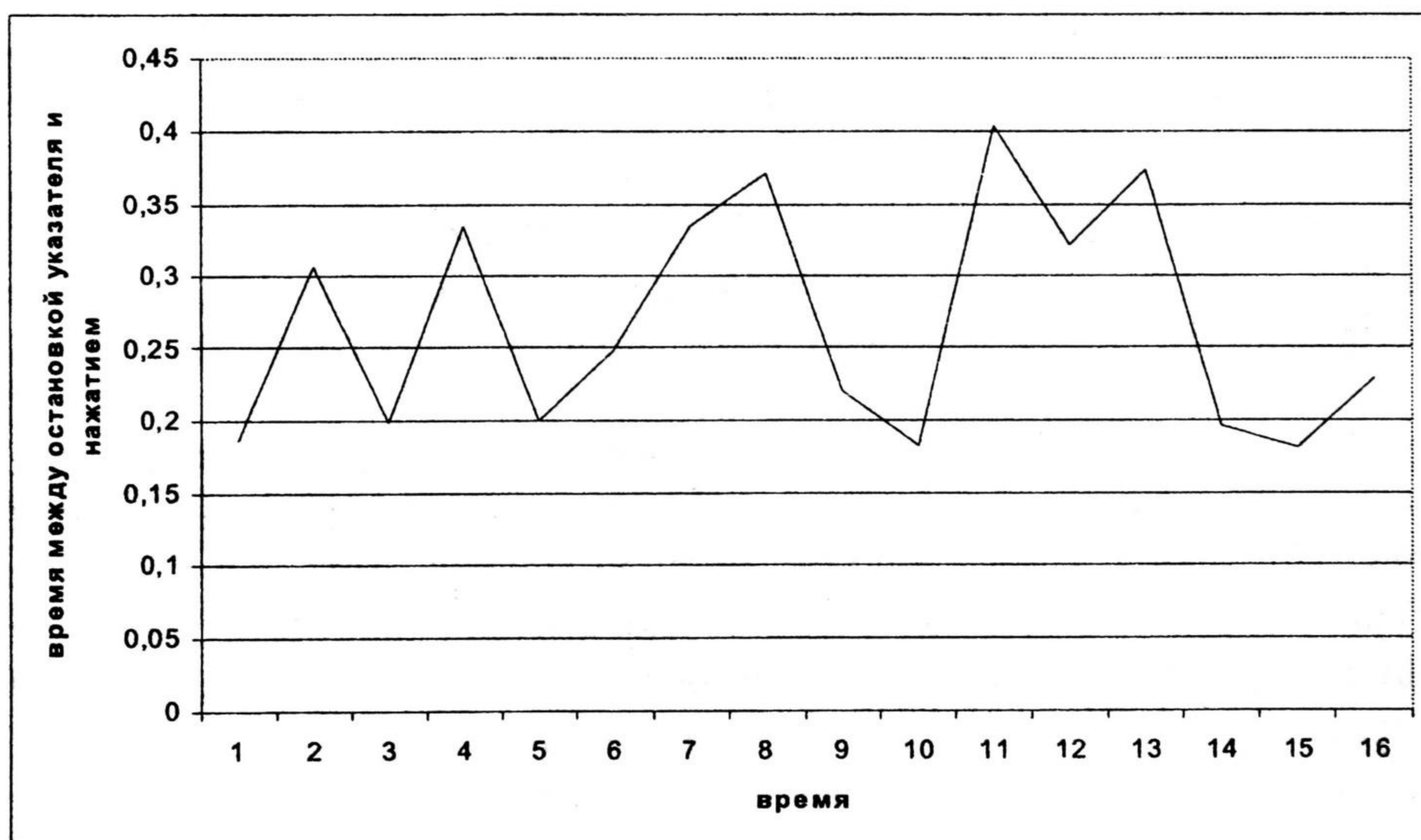
Более реально в развитие имеющихся работ по «клавиатурному почерку» отражать динамику мыши некоторой совокупностью оценок информативных параметров реализаций рассматриваемых случайных процессов.

2. Информативные параметры динамики работы пользователя с манипулятором «мышь»

Для отбора информативных параметров траекторий использованы экспериментальные данные, полученные в результате наблюдения за группой пользователей из 20 человек (50000 траекторий в одинаковых условиях работы). Критерием информативности взято отсутствие значимой статистической связи с другими параметрами, отображаемое низкими значениями коэффициентов корреляции Пирсона и Спирмена (ввиду «негауссовости» распределений). Динамика манипулятора «мышь» адекватно описывается шестью параметрами: T время реализации траектории указателя на экране монитора, L длина этой траектории, V_{cp} средняя скорость движения указателя, $t_{уд}$ время удержания кнопки, δ время между остановкой манипулятора и нажатием клавиши, α угол направления движения указателя [3].

Кроме того, ощутимую корреляцию показали две пары информативных параметров: 1) L_{cp} и V_{cp} ($R = 0,6$), 2) t_n и δ ($R = 0,6$).

Вычислительные эксперименты по наблюдению значений параметров траекторий у опытных пользователей выявили их вариабельность в широких пределах от траектории к траектории (рис. 1.). В связи с этим были проанализированы средние их значения за 30-минутные интервалы (пример — на рис. 2). Эти флуктуации уже прогнозируемы. Вопрос только в том, насколько хороша будет распознаваемость компьютерного образа ИПП, основанного на средних значениях параметров динамики мыши (траекторий).

Рис. 1. Динамика значений параметра δ Рис. 2. Динамика средних за 30 мин. значений параметра δ

3. Компьютерный образ ИПП

Компьютерный образ ИПП — это весовые коэффициенты синапсов нейронов искусственной радиальной нейросети [4], обучаемой по результатам мониторинга динамики мыши в течение нескольких дней (10 в нашем эксперименте). Эксперимент с пятью пользователями показал уровень распознаваемости такого образа — 0,83 (см. п. 4). Клавиатурный почерк распознается с надежностью 0,99 путем его наблюдения в момент ввода парольной фразы [5]. При этом он используется только для повышения надежности аутентификации при входе в открытую систему. Данные о распознавании пользователей по клавиатурному почерку в процессе работы в системе отсутствуют. Нас же интересует непрерывный мониторинг ИПП по динамике мыши. Существующие исследования по манипулятору «мышь» при работе в системе показывают надежность распознавания 0,8-0,9 [6].

Для повышения распознаваемости ИПП по динамике мыши в нашем исследовании была предпринята попытка учесть статистическую связь между информативными параметрами \bar{t}_n и $\bar{\delta}$. Это сделано следующим образом.

Посредством отдельной нейросети сигмоидального типа для очередного ожидаемого 30-минутного интервала формируется прогнозируемое значение параметра \bar{t}_n по полученному в предыдущем интервале значению $\bar{\delta}$ и измеряемым в этот момент значениям 4-х дополнительных параметров: $t_{\text{мек}}$ — время суток (в мин.), $t_{\text{раб}}$ — время работы пользователя (с начала рабочего дня, в мин.), $N_{\text{тр}}$ — количество реализаций траекторий с начала работы пользователя, $I = \frac{N_{\text{тр}}}{\tau}$ — интенсивность работы, где $\tau = 30$ мин. В результате получаем встроенную в нейросеть зависимость:

$$\bar{t}_n = \omega(\bar{\delta}, N, I, t_{\text{мек}}, t_{\text{раб}}),$$

специфичную для каждого 30-минутного интервала.

Нейронная сеть представляет собой 3-слойный персептрон. Обучающая выборка состоит из данных наблюдения, полученных за 10 рабочих дней. Входной вектор сети определен параметрами: $\bar{\delta}, N, I, t_{\text{мек}}, t_{\text{раб}}$. На выход сети подается прогнозируемое значение параметра \bar{t}_n .

Далее в текущем интервале времени посредством нейросети, выполняющей функцию распознавания, на вход которой подаются как средние измеренных значений 6 параметров, так и спрогнозированное значение параметра \bar{t}_n . Это позволило повысить распознавание ИПП по динамике мыши до 0,9-0,93. Причиной улучшения ситуации, таким образом, стало повышение устойчивости текущего образа ИПП за счет привязки его к предшествующему интервалу измерений.

Основным преимуществом нейронных сетей по сравнению с классическими методами регрессионного анализа является упрощение процедуры идентификации модели. Идентификация нейронно-сетевой модели состоит в выборе типа и архитектуры нейронной сети, количества скрытых нейронов, а также передаточных функций. Существенного сужения исследуемого класса функций при этом не происходит, поскольку класс зависимостей, выражаемый нейронными сетями заданной архитектуры, может быть весьма широким [7].

4. Алгоритм оценки распознаваемости ИПП по информативным параметрам динамики манипулятора «мышь»

В качестве инструмента для оценки распознаваемости ИПП по параметрам траекторий, формируемых манипулятором «мышь», использована нейронная сеть с радиальными базисными элементами, отличающаяся высокой скоростью обучения и способностью моделировать любую нелинейную функцию с помощью лишь одного скрытого слоя [4].

Радиальная сеть имеет фиксированную структуру с одним скрытым слоем и линейным выходным нейроном (рис. 3). При этом выполнены условия: центр c_i нейрона (функции φ_i — (2)) сохраняется в пространстве весов от входного слоя к слою шаблонов; в качестве расстояния входного вектора (оценок параметров траекторий) до центра c_i используется евклидово расстояние; функцией, определяющей выходной сигнал нейрона, является кривая Гаусса:

$$\varphi(x) = \varphi(\|x - c_i\|) = \exp\left(-\frac{\|x - c_i\|^2}{2\sigma_i^2}\right). \quad (2)$$

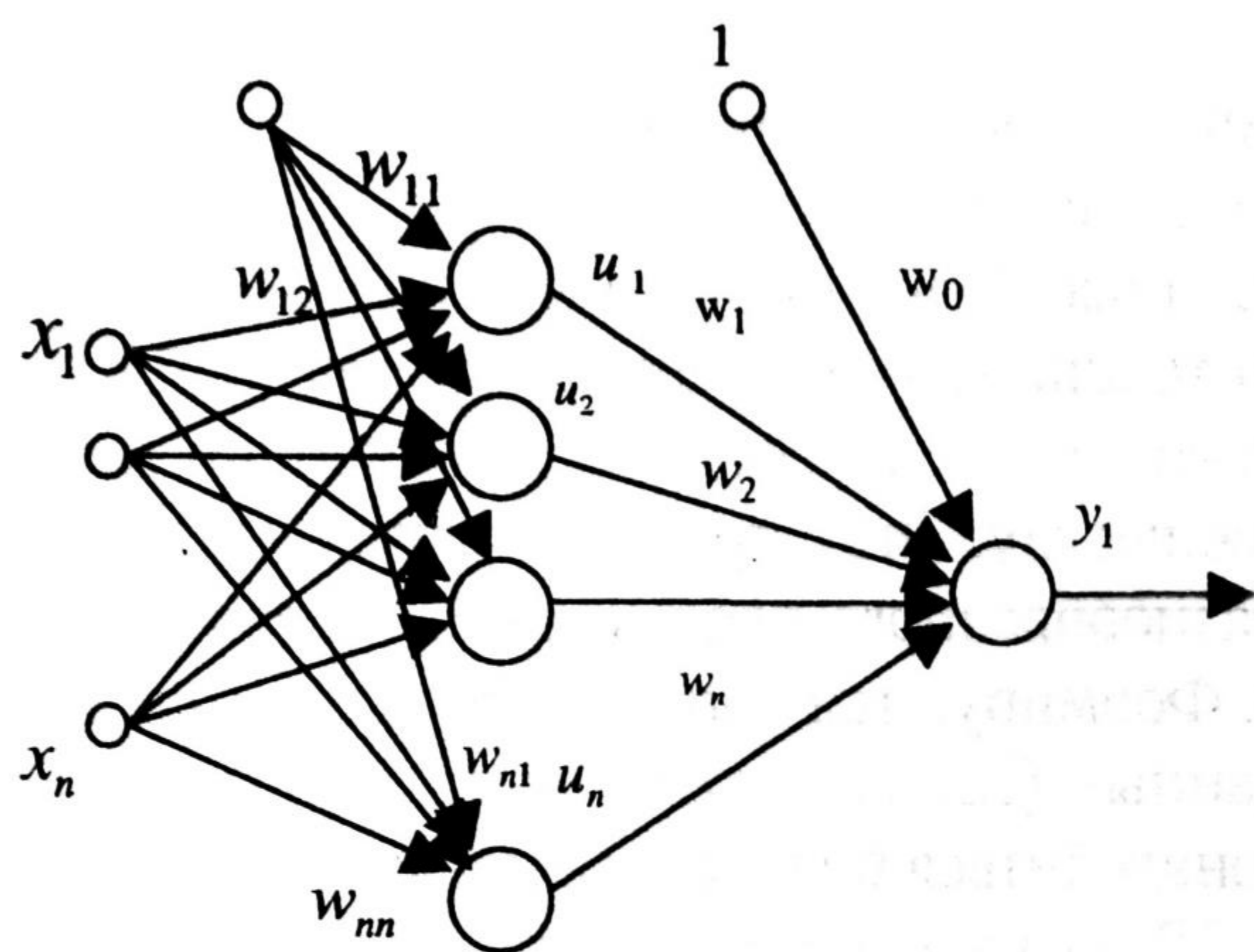


Рис. 3. Структура радиальной сети

Для обучения НС используется гибридный алгоритм [4]. Процесс обучения делится на два этапа: 1) подбор линейных параметров сети (весов выходного слоя); 2) адаптация нелинейных параметров радиальных функций (и характеристики ширины этих функций). В качестве обучающей выборки использовались данные наблюдений за пользователем на протяжении 10 дней, в течение всего рабочего дня.

Эффективность распознавания ИПП отображена процентом случаев правильного распознавания 5 пользователей в 1000-кратном эксперименте. Распознаваемость составила 90-93%. Увеличение количества зарегистрированных пользователей на одном рабочем месте (более 5) уменьшает вероятность распознавания в случае отсутствия пополнения обучающей выборки, т.е. без внесения дополнительной информации в компьютерные образы пользователей.

СПИСОК ЛИТЕРАТУРЫ

1. Власов А.Н. Способ представления координатной составляющей информационного почерка пользователя // Мягкие вычисления: Материалы междунар. науч. конф. Т. 1. СПб., 2003. С. 116-119.
2. Гузик В.Ф., Галуев Г.А., Десятерик М.Н. Биометрическая нейросетевая система идентификации пользователя по особенностям клавиатурного почерка // Нейрокомпьютеры. Разработка, применение. 2001. № 7-8. С. 104-118.
3. Диденко С.М., Шапцев В.А. Исследование динамики работы пользователя с манипулятором «мышь» // Математическое и информационное моделирование. Тюмень: Изд-во ТюмГУ, 2004, С. 295-304.
4. Оссовский С. Нейронные сети для обработки информации / Пер. с польского И. Д. Рудинского. М.: Финансы и статистика, 2002, 250 с.
5. Obaidat M.S., Sadoun B. Keystroke dynamics based authentication / Monmouth University W. Long Branch. NJ, 2000
6. Shivani Hashia, Authentication by Mouse Movements / CS 297 Report. 2004. May.
7. Секерин А.Б. Метод оценки устойчивости нейронно-сетевых моделей. 2005 (<http://zhurnal.ape.relarn.ru/articles/2005/031.pdf>).

*Денис Богданович КЕПЕЩУК —
аспирант кафедры информационных систем*

УДК 004.652.4

РАЗРАБОТКА МОДЕЛИ ОБМЕНА ДАННЫМИ В ГЕТЕРОГЕННЫХ РАСПРЕДЕЛЕННЫХ БАЗАХ ДАННЫХ

АННОТАЦИЯ. В статье приводятся основные результаты разработки модели обмена данными в распределенных базах данных с гетерогенными составляющими.

In the article are given the main results of the research of the model of data exchanging in distributed databases with heterogeneous constituents.