

Е.А. Высоцкая, НАУ, г. Киев

А.Н. Давиденко, ИПМЭ им. Г.Е.Пухова НАН Украины, г. Киев

## **АНАЛИЗ ТЕХНОЛОГИИ ПРЕДВАРИТЕЛЬНОЙ ОБРАБОТКИ ДАННЫХ ПРИ АУТЕНТИФИКАЦИИ ПОЛЬЗОВАТЕЛЕЙ КОМПЬЮТЕРНЫХ СИСТЕМ ПО КЛАВИАТУРНОМУ И РУКОПИСНОМУ ПОЧЕРКАМ**

Summary: In this article the stages of formation of a database of educational samples at the authentication on keyboard handwriting and at the authentication on hand-written handwriting are submitted. The stages of preprocessing of analyzable analyzed data at the specified two kinds of authentication are stated. The necessity of such processing is argued.

### **Введение**

С достижением достаточно высокого уровня развития вычислительной техники получили свое широкое применение биометрические динамические методы аутентификации. В ранее опубликованных работах [1-4] уже исследовались технологии биометрической аутентификации и давались рекомендации по разработке таких технологий. В данной работе внимание уделено вопросам предварительной обработки данных о почерке, которые впоследствии анализируются для выполнения аутентификации. Этот вопрос рассматривается в контексте двух разновидностей биометрической аутентификации: по клавиатурному почерку и по рукописному почерку. При этом подразумевается, что в качестве механизма распознавания будет использоваться вероятностная нейронная сеть [5].

### **Постановка задачи**

Целью исследований в данной работе было сравнить этапы предварительной обработки анализируемых данных при аутентификации по клавиатурному почерку и при аутентификации по рукописному почерку.

Для решения поставленной задачи было сделано следующее:

1. Проанализированы перечни анализируемых характеристик при аутентификации по клавиатурному почерку и при аутентификации по рукописному почерку.
2. Проанализирована специфика применения вероятностной нейронной сети для решения задачи распознавания образов.
3. Созданы на языке Borland Builder C++ программа на базе вероятностной нейронной сети для аутентификации пользователей компьютерной системы по клавиатурному почерку и на языке Borland Delphi аналогичная программа, анализирующая для аутентификации рукописный почерк.
4. С помощью созданных программ была накоплена информация о кла-

виатурном почерке и о рукописном почерке некоторого количества людей.

5. На основе накопленных данных, с помощью созданных программ было проведено ряд экспериментов для определения того, какая именно предварительная обработка анализируемых данных целесообразна при аутентификации пользователей по клавиатурному почерку и по рукописному почерку.

### **Решение поставленной задачи**

Биометрическая динамическая аутентификация обеспечивает достаточно хорошую вероятность правильного распознавания, однако динамическим характеристикам человека свойственна некоторая нестабильность. Кроме того, некоторая нестабильность характеристик почерка связана с особенностями используемого оборудования. С небольшой нестабильностью анализируемых данных вероятностная нейронная сеть, которая используется в данной работе в качестве механизма распознавания, хорошо справляется, но более значительные искажения, для обеспечения качественного распознавания, необходимо исправлять или исключать из процесса распознавания. Почерк (и клавиатурный, и рукописный) – это динамическая характеристика человека, которая может незначительно изменяться по нескольким причинам. Человек может спешить куда-то, может отвлечься на какие-то внешние обстоятельства, может заболеть – при этом он теряет внимательность и это сказывается на его почерке, т.е. происходит отклонение почерка от его нормы для данного человека. Такие отклонения полезны при выполнении мониторинга за работой сотрудников предприятия или при приеме на работу, поэтому эту информацию нельзя ни исправлять, ни исключать. Отклонения же возникшие по иным причинам необходимо нейтрализовать до начала распознавания. Кроме того, очень важно правильно выбрать то слово, которое используется в качестве пароля, т.е. слово (или фразу), характеристики набора на клавиатуре или характеристики написания на планшете, которого анализируется при распознавании. Предпочтительнее, чтобы это слово человек хорошо знал и часто его печатал/писал, т.е. слово на ввод/написание, которого у человека выработался свой почерк. Обычно это слово связано с областью работы данного человека или с его увлечениями. Если это новый человек в организации или он только учится работать за компьютером, тогда через некоторое время следует обновить базу данных его учебных образцов. Учитывая важность правильного выбора пароля и тех характеристик его ввода/написания, которые будут анализироваться при распознавании, до того как система сможет работать в режиме аутентификации необходимо выполнить следующее:

◆ Накопить пробную базу данных учебных образцов. Для этого выбрать предположительный пароль и те характеристики ввода/написания, которые предположительно должны анализироваться при распознавании. После чего накопить базу данных учебных образцов по выбранным характеристикам выбранного пароля, небольшого объема.

◆ Оценить накопленные учебные образцы. То есть проверить обеспечивают ли выбранные характеристики необходимую вероятность правильного

распознавания. Если требуемое качество распознавания не достигается, тогда выбрать другие анализируемые характеристики для этого же пароля и заново выполнить оценку накопленных образцов. Если повторная оценка не дала требуемого результата, тогда выбрать другой пароль и вернуться к этапу накопления пробной базы данных.

◆ Накопить полную базу данных учебных образцов, т. е. базу данных учебных образцов в объеме, требуемом для правильного распознавания.

Кроме указанных действий, как уже было сказано, для правильного распознавания необходимо выполнить предварительную обработку анализируемых данных.

Сначала рассмотрим, какую предварительную обработку анализируемых данных необходимо выполнить при аутентификации по клавиатурному почерку. При таком виде аутентификации, в данной работе, предлагается анализировать временные промежутки между нажатием на клавиатуре соседних символов пароля. Кроме правильного выбора пароля и анализируемых характеристик его ввода, на правильное распознавание оказывает влияние следующее. По причине различия в значениях параметров используемых клавиатур, могут искажаться характеристики ввода пароля. Поэтому, во-первых, в одной организации (если в ней используется единая база данных учебных образцов) лучше использовать одну модель клавиатур. А во-вторых, для всех клавиатур необходимо установить одинаковые настройки (задержку перед началом повтора и скорость повтора). При этом не будет иметь значение, на какой именно клавиатуре человек набирал пароль. Кроме того, как уже было сказано, человек может отвлечься и сделать какую-то ошибку (нажать не ту клавишу). При этом ввод слова лучше повторить заново (так как временной интервал перед вводом правильного символа будет искаженным), но признак сделанной ошибки лучше сохранить и впоследствии использовать при аутентификации, так как это достаточно индивидуальная характеристика. Если же человек по какой-либо причине сделал слишком большую паузу (значение должно устанавливаться в системе) перед вводом очередного символа, тогда ввод необходимо выполнить заново. Но, если временной промежуток не больше, чем допустимое значение, установленное в системе для всех пользователей, но значительно отличается от значения соответствующего параметра из других образцов для этого же пользователя, тогда этот образец ввода является не характерным для данного пользователя и его лучше исключить из базы данных учебных образцов (если это учебный образец). В работе был проанализирован существующий алгоритм [6] исключения учебных образцов с грубыми ошибками с помощью сортировки и исключения крайних значений (алгоритм ИУОГОСИКЗ) и предложен свой алгоритм – алгоритм исключения учебных образцов с грубыми ошибками с помощью сравнения признака с его средним значением. Суть предложенного алгоритма заключается в следующем: сначала для каждого пользователя находятся значения среднего арифметического каждого анализируемого признака; а затем каждый признак очередного об-

разца сравнивается с соответствующим средним значением и если отклонение больше допустимого, тогда этот образец необходимо исключить из базы данных учебных образцов, а если это неизвестный образец, тогда признать этот образец ложным и отказать аутентифицируемому в доступе к системе. В результате анализа написанных по этим алгоритмам программ и результатов проведенных экспериментов можно сказать, что предложенный алгоритм исключения учебных образцов с грубыми ошибками с помощью сравнения признака с его средним значением:

1. Проще.
2. Затрачивает меньше временных и вычислительных ресурсов.
3. Является, в большинстве случаев, более эффективным.

Первые два достоинства этого алгоритма объясняются тем, что:

1. В этом алгоритме выполняется отбор сразу по всем анализируемым признакам, в отличие от алгоритма ИУОГОСИКЗ, в котором многие действия выполняются по несколько раз.

2. В отличие от предложенного в данной работе алгоритма, в алгоритме ИУОГОСИКЗ для каждого признака выполняется сортировка, которая занимает достаточно много времени и оперативной памяти, особенно если сортируемый массив большой.

Более высокая эффективность, в большинстве случаев применения предложенного алгоритма, выражается в следующем:

1. Анализ качества анализируемых признаков после выполнения исключения из базы данных учебных образцов с грубыми ошибками по предложенному алгоритму показал, что все параметры имеют примерно одинаковый достаточно неплохой уровень качества, а после исключения по алгоритму ИУОГОСИКЗ уровень качества для различных анализируемых признаков сильно различается между собой.

2. При относительно большом количестве анализируемых признаков предложенный алгоритм обеспечивает лучшее качество распознавания, а при малом количестве (недостаточном для хорошего распознавания) анализируемых признаков алгоритм ИУОГОСИКЗ оказывается эффективнее.

После выполнения указанной предварительной обработки анализируемых данных вероятность правильного распознавания повышается.

При аутентификации по рукописному почерку необходима большая предварительная обработка данных, чем при аутентификации по клавиатурному почерку. Это вызвано следующими причинами. При распознавании по клавиатурному почерку анализируются, во-первых, параметры одного типа (временные промежутки), а, во-вторых, их немного (обычно, до 10). При распознавании по рукописному почерку аутентификация состоит из двух этапов: во-первых, анализируется написанный пароль, а во-вторых, стиль написания этого пароля. На первом этапе анализируются координаты  $X$  и  $Y$  точки и ее тип. На втором этапе анализируются такие характеристики: тип точки; давление, с которым человек давит ручкой на планшет при написании данной точки; угол изменения направления письма; время, прошедшее от начала напи-

сания символа до создания данной точки; скорость перемещения ручки из предыдущей точки в текущую, площадь изображения символа; угол наклона символа; количество контрольных точек в изображении символа; частота зафиксированных точек (соотношение пути, пройденного ручкой при написании символа к количеству зафиксированных точек); количество повторов точек (подряд идущих точек с одинаковыми обеими координатами) в изображении символа. Таким образом, можно сделать вывод, что при аутентификации по рукописному почерку, во-первых, анализируются характеристики разных типов, а, во-вторых, этих характеристик намного больше, чем при аутентификации по клавиатурному почерку. Этим и объясняется тот факт, что при аутентификации по рукописному почерку необходима большая предварительная обработка данных, чем при аутентификации по клавиатурному почерку.

Кроме правильного выбора (описанного ранее) пароля и анализируемых параметров написания пароля необходима следующая обработка данных.

В данной работе предлагается выполнять распознавание не всего пароля целиком, а посимвольно. Выполнение данного предложения не является обязательным, но является целесообразным, для значительной экономии ресурсов. Это объясняется двумя причинами.

Во-первых, при проверке правильности написанного пароля предпочтительней распознавать пароль не целиком, а посимвольно. Это вызвано следующими причинами:

1. Легче накопить в базе данных образцы написания  $N$  символов (где  $N$  – количество символов в используемом алфавите, с учетом того, что могут использоваться не только строчные буквы, но и заглавные буквы, цифры, различные знаки пунктуации), чем образцы написания всех возможных паролей, которыми могут являться любые комбинации из  $N$  символов. Количество возможных паролей (комбинаций символов), в данном случае будет равняться

$\sum_{Ks=1}^{Mks} Ks^N$  (где  $Ks$  – длина пароля,  $Mks$  – максимально возможная длина пароля)

или же будет равняться  $Ks^N$ , если  $Ks$  известно. Накопление такого количества данных значительно труднее, чем в случае посимвольного распознавания.

2. Классифицируемый объект легче проверить на принадлежность к одному из  $N$  классов, чем на принадлежность к одному из  $\sum_{Ks=1}^{Mks} Ks^N$  классов. То

есть, процесс распознавания легче при посимвольном анализе.

Во-вторых, обработка данных сразу по всему паролю займет чрезмерно большие ресурсы. Поясним это подробнее. При выполнении аутентификации по рукописному почерку с помощью графического планшета во время написания одного символа система получает в среднем данные по ста точкам (иногда по нескольким сотням точек). Если же в пароле, например, 6 символов, то для проверки истинности введенного пароля необходимо обработать в среднем данные по 600 точкам. То есть вероятностной нейронной сети при

распознавании написанного пароля необходимо будет обработать почти 2000 признаков, а это займет очень много времени (а при распознавании стиля написания пароля признаков больше и, следовательно, времени необходимо еще больше). Поэтому при аутентификации можно выполнять анализ не сразу всего пароля, а посимвольно, и, если один из символов неверный, то остальные символы можно не анализировать. Однако при использовании такого жесткого условия необходимо обеспечить очень низкую вероятность ложного отказа при распознавании каждого символа, так ложное не распознавание одного символа приведет к ложному не распознаванию всего пароля, а, следовательно, отказ в доступе к защищаемой системе легальному пользователю.

В связи с вышеизложенным, в данной работе предлагается выполнять распознавание посимвольно, а, следовательно, необходимо разделить изображение всего слова на изображения отдельных символов.

Для правильного распознавания также необходимо исключить ошибочные данные следующих типов:

1. Последовательность точек с нулевым давлением (кроме первой подобной точки из каждой последовательности). Возникают из-за того, что пакеты данных о точке формируются, даже если человек не касается ручкой графического планшета, а передвигает его над рабочей областью планшета, на небольшом расстоянии. При этом в систему передаются данные о точках, которые не видны и не являются частью изображения пароля. Такая информация является ошибочной для аутентификации.

2. Случайные точки (небольшое количество). Возникают, если пользователь случайно дотронулся ручкой до планшета и, при этом, образовалась одна или несколько точек, невидимых глазом человека. В данном случае важно настроить систему таким образом, чтобы не принять часть символа за случайные точки, т.е. надо определить какое количество подряд идущих точек, составляющих отдельную линию, считать ошибкой, а какое частью символа.

3. Повторы, т.е. последовательность подряд идущих точек с неизменившимися координатами по осям  $X$  и  $Y$  (кроме случая, когда у одной из точек нулевое давление). Возникают потому, что пакет данных о точке передается в систему, если изменился хотя бы один параметр пакета, т.е. если изменилось, например, давление ручкой на планшет, а координаты ручки не изменились, то пакет данных все равно формируется и передается в систему. Такая информация может быть полезна при распознавании стиля написания пароля, но ошибочна при распознавании самого символа.

4. Случайные небольшие загибы в начале линий. В связи со спецификой работы с графическим планшетом, в начале линий иногда возникают небольшие загибы, как правило, под острым углом к линии символа. При фиксировании такой ошибки важно не спутать часть символа со случайным загибом.

5. Некачественный образец, отбрасываемый из-за невозможности разбиения изображения слова-пароля на заданное количество изображений символов (чаще возникает в результате отсутствия навыков работы с ручкой графического планшета). В данной работе, как уже было сказано, предлагается вы-

полнять посимвольное распознавание и в системе сделано ограничение, что пользователь должен писать символы пароля не слитно, а отдельно. Если же он написал какие-то символы слитно или же написал пароль из меньшего количества символов, тогда этот образец является ложным.

Кроме того, для правильного распознавания необходимо выполнить следующую корректировку данных:

1. Посимвольный поворот изображений символов для нормализации угла наклона их осей координат.

2. Посимвольный сдвиг изображений каждого символа в центр рабочей области выбранного размера.

3. Посимвольное пропорциональное масштабирование (растягивание/сжатие) изображений каждого символа на всю рабочую область выбранного размера.

Эта корректировка необходима по следующим причинам. Человек может написать пароль один раз слева планшета, другой раз, например, справа; может написать большими буквами, может маленькими; планшет может стоять ровно, по отношению к человеку, а может под углом. Кроме того, в разные дни могут использоваться планшеты разных размеров. При этом если не выполнить указанную коррекцию, т.е. если не привести все образцы к единому шаблону, тогда будет распознаваться не написанный символ, а его местоположение, что не является допустимым.

После выполнения указанной предварительной обработки анализируемых данных вероятность правильного распознавания повышается.

### **Выводы**

В данной работе представлены этапы формирования базы данных учебных образцов при аутентификации по клавиатурному почерку и при аутентификации по рукописному почерку; изложены, с аргументацией их необходимости, этапы предварительной обработки анализируемых данных при указанных двух видах аутентификации.

1. *Е.А. Высоцкая.* Выбор анализируемых параметров при аутентификации пользователей компьютерных систем по клавиатурному почерку при помощи вероятностной нейронной сети. – Киев: НАН України. Збірник наукових праць. Моделювання та інформаційні технології, 2005р., випуск 30, стор. 45-52.

2. *Е.А. Высоцкая.* Исключение учебных данных с грубыми ошибками, как один из способов повышения эффективности применения вероятностных нейронных сетей для аутентификации пользователей компьютерных систем по клавиатурному почерку. – Киев: НАН України. Збірник наукових праць. Моделювання та інформаційні технології, 2005р., випуск 29, стор. 52-59.

3. *Е.А. Высоцкая.* Влияние исключения учебных данных с грубыми ошибками на зависимость эффективности применения вероятностных нейронных сетей для аутентификации пользователей компьютерных систем по клавиатурному почерку от различных параметров. – Киев: НАН України. Збірник наукових праць. Інститут проблем моделювання в енергетиці, 2005р., випуск 28, стор. 3-10.

4. *Е.А. Высоцкая.* Задача распознавания написанного ключевого слова, как одна из задач, решаемых при выполнении аутентификации пользователей компьютерных

систем по рукописному почерку. – Київ: НАН України. Збірник наукових праць. Моделювання та інформаційні технології, 2006р., випуск 36, стор. 67-76.

5. Р. Каллан. Основные концепции нейронных сетей.: Пер. с англ. – М.: Издательский дом “Вильямс”, 2001. – 290с.

6. Расторгуев С.П. Программные методы защиты информации: Учебное пособие/ Пензенский государственный университет – Пенза: Издательство Пензенского государственного университета, 2000. – 95с.

Поступила 15.02.2010р.

УДК 621.396.677.4(045)

Л.Я. Ільницький, О.А. Щербина, І.І. Михальчук

### МАЛОГАБАРИТНА СПІРАЛЬНА АНТЕНА

In modern systems of mobile radiomonitoring the antennas of small dimensions are used. These antennas are ineffective in the ranges of high and very high frequency, as it is difficult to match them with the feeder due to low effective length and low input resistance.

**Вступ.** У більшості видів коротких антен ємнісне навантаження сприяє збільшенню діючої довжини антени і зменшенню ємнісного вхідного опору. За допомогою засобів антенної техніки можливо створити такі схемні конфігурації навантажень, які будуть вирівнювати амплітудний розподіл струму в лінійних частинах дровових антен і самі випромінюватимуть, що впливатиме на розподіл поля у просторі. До таких форм навантажень можна віднести дровові спіралі із стоячою хвилею струму.

**Постановка завдання** Завдання полягає в тому, щоб визначити вплив спіралі, яка обтікається стоячою хвилею струму, на розподіл поля в просторі.

**Теоретичні засади.** Розглянемо поле випромінювання спірального провода, якому надано форму спіралі.

Живлення підведемо до точки  $A$  (початок спіралі), а в точці  $B$  (кінець спіралі) будемо вважати, що струм має нульове значення. Рівняння спіралі в полярній системі координат має такий вигляд:

$$\rho = \rho_0 - a\varphi_s, \quad (1)$$

де  $\rho_0 = OA$  – початковий радіус спіралі;  $a$  – швидкість згортання спіралі;  $\varphi_s$  – полярний кут.

Радіус у кінці спіралі ( $\rho_k = OB$ ), очевидно повинен задовольняти нерівність:

$$\rho_k = \rho_0 - a\varphi_k > 0, \quad (2)$$