

ВЫБОР И ПРОГРАММНАЯ РЕАЛИЗАЦИЯ МЕТОДОВ ВИЗУАЛИЗАЦИИ ДАННЫХ КЛАВИАТУРНОГО ПОЧЕРКА

Н.Е. Юрьева

В данной работе проведён анализ методов визуализации данных клавиатурного почерка, описан ряд методов представляющих наибольший интерес с точки зрения прикладной области и обосновано их применение. Даётся описание разработанного веб-приложения предназначенного для сбора и предварительного анализа данных клавиатурного почерка.

This study presents an analysis of keyboard handwriting data visualization methods, the number of methods that are most suitable for application area are described, and their application is justified. The description of the developed web application intended for collection and preliminary analysis of the keyboard handwriting data is given.

КЛЮЧЕВЫЕ СЛОВА

Визуализация данных, клавиатурный почерк, веб-приложение.

1. ВВЕДЕНИЕ

Всё более важное место в жизни человека занимают информационные технологии, а использование компьютера для решения повседневных и профессиональных задач стало нормой. В подавляющем большинстве случаев, для реализации человеко-машинного взаимодействия, в случае компьютерных систем, используется клавиатура и мышь. Набор текстов на компьютере постепенно вытесняет классический рукописный способ выполнения текстов. Так же как и в «классической форме почерка» [1,2] неизбежно в процессе печати текста находят своё отражение многие особенности личности человека или свойства, присущие группам лиц. У каждого человека проявляется свой уникальный “почерк” – набор динамических характеристик печати. Индивидуальность пользователя проявляется в скорости набора символов, привычке использовать основную или дополнительную часть клавиатуры, характере «сдвоенных» и «строенных» нажатий клавиш, в излюбленных приемах управления компьютером и т. д. [4,3]

Методы оценки этих параметров по данным клавиатурного почерка вызывают всё больший интерес, по мере распространения электронных средств ввода.

Актуальность данной темы обусловлена двумя основными факторами – востребованностью исследований в области клавиатурного почерка и низким уровнем их программно-технического обеспечения.

В рамках данной работы, через создание тонкого клиента, решается одна из подзадач программно-технического обеспечения, а именно визуализация сырых экспериментальных данных. Стоит отметить, что даже незначительный набор методов визуализации, интегрированный в систему сбора данных о динамических характеристиках набора текста,

способствует интенсификации проведения исследований и повышает привлекательность инструмента для конечного пользователя (исследователя).

2. ВИЗУАЛИЗАЦИЯ И ПЕРВИЧНЫЙ АНАЛИЗ ДАННЫХ ИССЛЕДОВАНИЯ

В данном разделе будут подробно рассмотрены и описаны выбранные автором методы визуализации и обосновано их применение в задачах исследования клавиатурного почерка.

2.1 Тепловая карта

Тепловая карта — это графическое представление данных, где дополнительные переменные отображаются при помощи цвета. С помощью тепловой карты можно понять, какие зоны, например, сайта, используются чаще всего, или как в данном случае – какие клавиши чаще всего нажимал респондент во время исследования [5].

Преимущества подобного средства представления данных, трудно переоценить, так как тепловые карты просты для восприятия, но при этом часто позволяют делать достаточно точные выводы о динамике развития сложных процессов.

В данной работе тепловые карты используются для представления информации о среднем времени, затраченном пользователем на поиск символа при вводе текста. Можно описать несколько гипотез общего вида, предварительная проверка которых возможна с использованием подобной визуализации:

– Гипотеза о сходстве/различии в динамике набора текста группами испытуемых, имеющими различные уровни фактора, рассматриваемого в исследовании. При этом наличие различий не подчиняющихся линейному закону (группа А набирает служебные символы быстрее группы Б, при отсутствии различий в динамике набора печатных символов) будет легко дифференцироваться от случая линейных различий (группа А набирает текст, медленнее группы Б).

– Гипотеза о наличии/отсутствии влияния контролируемого внешнего фактора на динамику набора текста заданным испытуемым.

2.2 Столбчатая диаграмма

Столбчатая диаграмма позволяет легко определить не только наличие различий в выполненных по некоторому параметру измерениях, но и дать достаточно точную оценку отношения этих измерений друг к другу. Хотя, на первый взгляд, тепловая карта решает аналогичную задачу, на самом деле она обеспечивает более простое, в субъективном плане, сопоставление векторов данных, но даёт меньшее представление об их относительных различиях. В то время как тепловая карта для сопоставления значений параметра требует операций в терминах отношений между градациями цвета, столбчатая диаграмма позволяет перейти к отношениям высот. Такая задача имеет значительно более простое умозрительное решение. Среди примеров использования в контексте предметной области можно назвать:

– Уточнение и детализация интерпретаций выполненных с использованием тепловых карт.

– Поиск выбросов и артефактов в исходных данных.

– Оценка сходств различий в динамике ввода текста, для заданной пары (небольшой группы) символов.

2.3 Линейная диаграмма

Линейные диаграммы используются для характеристики вариации, динамики и взаимосвязи.

Данный вид диаграмм применяется для графического представления зависимости значений двух переменных. Одним из наиболее частых случаев применения линейной диаграммы можно считать отображение динамики измеряемого показателя во времени или сравнение динамики развития нескольких процессов, характеристики которых измерялись одновременно (в таком случае в одной области построения отображается несколько кривых/ломанных). Очевидно, что такая форма представления данных позволяет со значительной точностью установить наличие корреляции между парой значений, а так же дать общую характеристику отображаемому процессу.

В разработанном модуле по оси абсцисс на подобном графике располагаются символы текста введенного пользователем, от первого до последнего, а по оси ординат абсолютное время, затраченное на его поиск на клавиатуре. Анализ графика позволяет дать предварительный ответ на следующие общие вопросы:

- Существует ли временный эффект научения (ускоряется ли набор по мере работы) и, при существовании такого эффекта, существуют ли индивидуальные/групповые различия в его проявлении.
- Проявляется ли эффект утомления (замедление набора) и если да, то для каких групп пользователей и в какой степени.
- Существуют ли символьные группы (биграмы, триграммы, целые слова), для которых значительно выражено изменение динамики набора по отношению к остальному тексту.

2.4 Базовые статистики

Помимо визуализации данных исследования, модуль осуществляет вычисление простейших статистик для каждого респондента и позволяет экспортировать их прямым копированием таблицы из окна браузера в большинство современных табличных редакторов. В значительном числе случаев (в частности, проверке гипотез о групповых сходствах/различиях), сырые данные не представляют практического интереса. Система предоставляет данные о мерах центральной тенденции (медиане и математическом ожидании), а так же стандартном отклонении для времени ввода символов и времени ввода слов. Дополнительно в данной таблице хранятся данные об абсолютном количестве операций редактирования (использования клавиши `backspace`) и количестве случаев прерывания ввода (превышения интервала между вводом символов). Эти данные после копирования могут быть произвольно обработаны в стандартных статистических пакетах, а также использованы для настройки обучаемых структур, например, в качестве входов нейронной сети, выходом которой будет метка группы («высокий уровень стресса»/«низкий уровень стресса» и т.п.)

В заключении данного раздела хотелось бы отметить, что спектр гипотез, предварительная проверка которых возможна, в частности, за счёт реализованных средств визуализации достаточно широк и перекрывает значительный процент конкретных задач. А проведение пилотажного исследования от сбора данных до интерпретации требует от пользователя минимального набора действий. С технической точки зрения, ему необходимо отправить ссылку на форму ввода (размещённую в сети интернет) и инструкцию группе респондентов, а затем ознакомиться с описанными графиками. С практической точки зрения, продумывание дизайна эксперимента, всё ещё остаётся ключевой задачей, не поддающейся автоматизации в общем виде.

3. ПРОГРАММНАЯ РЕАЛИЗАЦИЯ

Реализацию разработанного в рамках данной работы проекта можно разделить на несколько независимых подзадач: построение тепловых карт скорости набора клавиш, построение столбчатых диаграмм времени поиска символа, построение графика динамики набора текста, формирование таблицы базовых статистик. Для решения каждой из этих подзадач требуется разработка кода клиентской и серверной части [6].

Модуль построения первичных статистик на основании собранных экспериментальных данных, формирует сводную таблицу, содержащую набор статистических показателей для каждого респондента, принимавшего участие в исследовании клавиатурного почерка (рис.3.1)

Имя	Отчество	Фамилия	Возраст	Пол	Медиана ввода симв. (мс)	Ср. инт. ввода симв. (мс)	Ст. отк. ввода симв. (мс)	Медиана ввода слов (мс)	Ср. инт. ввода слов (мс)	Ст. отк. ввода слов (мс)	Прерыв.	Backsp.
Поликсения	Ильинична	Шварц	69	Мужской	180	232	173.63	360	568	447.28	0	0
Анна	Михайловна	Карлина	22	Женский	270	435	642.42	1552	1747	1471.79	2	3
Иван	Иванович	Иванов	56	Мужской	202	330	497.09	1619	1858	1434.24	0	0
Петр	Петрович	Петров	18	Мужской	180	306	283.76	1733	2010	1434.92	1	0
Евгения	Германовна	Кузьмина	50	Женский	216	301	316.01	1890	2014	1749.79	0	2

Рис.3.1 Пример результатов работы модуля построения таблицы первичных статистических данных

Модуль визуализации тепловой карты состоит из двух взаимосвязанных частей – клиентской и серверной (рис.3.2). Клиентская часть отвечает за взаимодействие с пользователем, обработку его запросов и визуальное отображение информации, а серверная за выполнение вычислительных функций и работу с базой данных.



Рис.3.2 Графическая модель клавиатуры с нанесённой тепловой картой

окончательного, статистически обоснованного решения об отвержении или принятии выдвинутой гипотезы.

ЛИТЕРАТУРА

1. Юрьева Н. Е. Поддержка принятия решений при построении психологического портрета личности на основе нейросетевого распознавания почерка: диссертация кандидата технических наук: 05.13.01 / Юрьева Наталия Евгеньевна; [Место защиты: Нижегород. гос. техн. ун-т им Р.Е. Алексеева]. - Москва, 2013. - 171 с.
2. Гунько Н.Е. Подход к решению задачи составления психологического портрета человека по почерку - Нейрокомпьютеры: разработка и применение, №2, 2012, с. 54-62.
3. Брюхомицкий Ю. Клавиатурная идентификация личности. - Издательство: «LAP LAMBERT Academic Publishing», 2012 г. -321с.
4. Биометрическая аутентификация пользователя по клавиатурному почерку. Анализ динамики нажатия клавиш. [Электронный ресурс] // Образовательный сайт fullref.ru. – 2009.
5. Желязны Д. Говори на языке диаграмм. Пособие по визуальным коммуникациям. – Москва: Манн, Иванов и Фербер, 2010. – 260 с.
6. Адам Т., Дэвид С. РНР. Рецепты программирования. – 3-е издание. – Санкт-Петербург : Питер, 2015. – 784 с.

Работа поступила 16.07.2017г.