использованными при проектировании и строительстве программами для решения эксплуатационных задач.

Сегодня наиболее широко распространены трехмерные пространственные модели объектов 3D-BIM. Оно касается инженерно-геологических работ, материально-технического обеспечения строительства и безопасности. Однако, можем предположить, в ближайшее время ожидается рост интереса к 4D (временное планирование), 5D (финансовые затраты), 6D (потребление энергоресурсов и экологические параметры) и даже 7D-моделям (эксплуатационно-техническое обслуживание).

Возможности для повышения экологичности, заложенные в BIM, позволяют минимизировать количество отходов при строительстве, организовать их утилизацию на протяжении всего жизненного цикла здания, рассчитать объемы вредных выбросов и разработать программу по борьбе с ними.

Препятствием развития является устаревшая и с трудом развивающаяся нормативная база, отсутствие в Российской Федерации государственных BIM-стандартов и национальных классификаторов. Частично в приоритете использование BIM для оценки стоимости проекта, точной оценки строительных объемов, мониторинга реализации проекта, в том числе план-фактного анализа.

Весьма близка к BIM сформулированная компанией Dassault Systemes в 1998 году концепция Product Lifecycle Management (далее PLM) – управление жизненным циклом изделия, которой сегодня активно пользуется практически вся индустрия машиностроительного САПР. BIM является отражением и уточнением концепции PLM в специализированной области человеческой деятельности – архитектурно-строительном проектировании. Вполне логично, что по аналогии с PLM даже начал появляться термин BLM (Building Lifecycle Management) – управление жизненным циклом здания.

1. Талапов В.В. Основы BIM: введение в информационное моделирование зданий. Издательство ДМК-Пресс. 2015 г.
2. IFC – Industry Foundation Classes [Electronic resource] // IFC Wiki. – URL: http://www.ifcwiki.org/index.php?title=IFC_Wiki/.
3. ISO 19650-1:2018 Организация и оцифровка информации о зданиях и строительных работах, включая информационное моделирование зданий (BIM). Информационный менеджмент с использованием технологии информационного моделирования зданий.
4. Приказ Минстроя России №1230/пр от 18.09.17г. Об утверждении свода правил «Информационное моделирование в строительстве. Правила обмена между информационными моделями объектов и моделями, используемых в программных комплексах».

Новоженков В.А., Ковтушенко А.П.

Методы аутентификации пользователя по работе с клавиатурой

*МГТУ им. Н. Э. Баумана
(Россия, Москва)*

doi: 10.18411/lj-04-2020-28

idsp: ljjournal-04-2020-28

Аннотация

В представленной статье приводится обзор, краткое описание и анализ методов аутентификации пользователя по работе с клавиатурой (так называемым «клавиатурным почерком»), а также вариант распределенного и параллельного обучения таких алгоритмов. Были рассмотрены архитектура системы аутентификации и математические основы моделей, применяемых при распознавании пользователя, а также метод распределенного обучения алгоритма.

Ключевые слова: аутентификация, методы машинного обучения, нейронные сети, клавиатурный почерк, высокопроизводительные вычисления.

Abstract

This article provides an overview, a brief description and analysis of user authentication methods for working with the keyboard (the so-called "keyboard handwriting"), as well as a variant of distributed and parallel training of such algorithms. We examined the architecture of the authentication system and the mathematical foundations of the models used in user recognition, as well as the method of distributed learning of the algorithm.

Keywords: authentication, machine learning methods, neural networks, keyboard handwriting, high-performance computing.

Введение

Сложность и эффективность атак на корпоративные системы (например, банковские) возрастает ежегодно. Также, вместе с ними ежегодно совершенствуются и методы защиты информации. Помимо этого, очень важно разграничивать доступ пользователей к информации с разным уровнем доступа.

Одними из перспективных систем защиты информации являются системы класса «User and Entity Behavior Analytics» (UaEBA), которые производят непрерывную аналитику поведения пользователей на основе их работы с системой. Эти системы могут отследить несанкционированное взаимодействие с системой, даже если пользователь успешно аутентифицировался (или, например, получил доступ к системе с пройденной аутентификацией и открытым доступом).

Системы класса UaEBA являются самообучающимися системами, которые формируют свою базу знаний о пользователях и сущностях (например, бизнес-приложениях, базах данных, серверах и т. д.) на основе данных об их работе в прошлом (однако обычно таких данных очень большой объем, что требует значительных вычислительных ресурсов для успешного функционирования системы). Таким образом, в процессе работы пользователя или некоторой сущности формируется шаблон действий, характеризующий его, который необходимо периодически дообучать для поддержания его актуальности. Также этот шаблон периодически или непрерывно сопоставляется с текущим поведением системы или пользователя и в процессе этого сопоставления происходит аутентификация пользователя. Если аутентификация не успешна – возможны следующие варианты:

- 1) система или же аккаунт пользователя, взаимодействующего с ней, блокируются и вызываются дополнительные механизмы и процедуры проверки подлинности (например, для проверки личности пользователя запрашивается пароль от его учетной записи).
- 2) или же пользователю становятся недоступны критически важные ресурсы без явной блокировки системы

В рамках данной статьи рассматривается аутентификация пользователя по работе с клавиатурой, а также методы, которые могут применяться для осуществления аутентификации. Одним из главных преимуществ данного вида аутентификации является то, что она происходит в фоновом режиме и скрыта от глаз пользователя. Это означает, что пользователь может вовсе не знать о её существовании и работе, а следовательно, её сложнее обмануть и обойти злоумышленнику.

Концептуальная схема взаимодействия пользователя и системы приведена на рисунке 1.

Рисунок 1. Концептуальная схема взаимодействия пользователя и системы

Система представляет из себя комплекс программного обеспечения. На каждый компьютер компании устанавливается драйвер-фильтр верхнего уровня, перехватывающего нажатия на клавиши клавиатуры и записывающего в какой-либо файл следующие метрики: время удержания каждой клавиши клавиатуры, время между нажатиями разных клавиш, последовательность нажатия-отпускания каких-либо двух клавиш. Затем все собранные значения отправляются на обработку на сервер, где на их основе производится аутентификация пользователя. В случае успешного прохождения аутентификации алгоритм обновляется и дообучается (в случае использовании базы данных - база данных пополняется новым значением характеристики, а в случае использования нейронной сети или классических алгоритмов – происходит дообучение алгоритма на новых данных).

Для постоянного обучения на новых данных нужно определить задачу «дообучения», которую можно будет вызывать в нужный момент времени для нужного пользователя. Вызов такой задачи должен содержать текущее состояние и данные, на которых следует дообучать модель. Задачи обучения нейронных сетей и классических алгоритмов машинного обучения могут занимать довольно большое количество времени, поэтому имеет смысл эти задачи либо ускорить, либо исполнять параллельно. В данной статье затрагивается техническое решение, благодаря которому, вычисление задачи обучения нейронной сети ускоряется и производится параллельно и распределено с использованием высокопроизводительного кластера. Для этого необходимо внешнее API, позволяющее отправлять задачи на кластер с целью проведения дообучения алгоритмов. В таком случае компонент системы должен иметь список отправленных задач и отслеживать их состояние.

Нейронные сети, способные производить аутентификацию пользователей могут быть обучены параллельно и распределенно с использованием open-source библиотеки horovod, которая базируется на основных понятиях и концептах MPI, таких как size, rank, local rank, allreduce, allgather и broadcast.

Предположим, имеется 4 сервера с 4 GPU-ускорителями на каждом. Тогда запуск обучения нейронной сети с помощью Python-скрипта на каждом узле и каждом GPU-ускорителе будет вызывать следующие MPI-методы:

- Size – сообщает о количестве процессов (в данном случае 16);
- Rank – присваивает уникальный ID процессу (от 0 до 15);

- Local Rank – Уникальный ID процесса на одном узле кластера (от 0 до 3);
- Allreduce – операция, объединяющая данные с нескольких процессов и передающая результат обратно. Используется для усреднения тензоров в нейронной сети;
- Allgather – операция, собирающая данные со всех подпроцессов. Используется для сбора значений разреженных тензоров;
- Broadcast – операция, которая транслирует данные из одного процесса, идентифицируемого корневым рангом, на каждый другой процесс. [13]

С помощью вызовов библиотеки `horovod`, можно достичь огромного прироста в скорости обучения нейронных сетей и некоторых классических алгоритмов кластеризации за счет параллельного и распределенного обучения алгоритмов на высокопроизводительных кластерах с использованием множества ядер графических ускорителей (GPU). В рамках поставленной бизнес-задачи скорость обучения алгоритмов является критически важным параметром, определяющим эффективность всей системы в целом.

Дообучение на новых данных позволяет алгоритму плавно адаптироваться к изменениям характеристик пользователя. В случае неуспешного прохождения аутентификации – система блокирует доступ пользователя к рабочему компьютеру, либо просит подтвердить свою личность с помощью ввода пароля или других дополнительных механизмов аутентификации.

Рассмотрим основные алгоритмы и методы аутентификации:

Метод №1. Проверка динамики набора конкретного текста (пароля)

Одним из базовых методов аутентификации пользователя является проверка динамики набора конкретного пароля на клавиатуре, а именно – средняя скорость набора конкретного текста, время между нажатием клавиш, необходимых для набора заданного текста, время удержания клавиш, а также среднее количество типичных ошибок, допускаемых при вводе конкретного текста. Данный метод основан на нечетких множествах. Аутентификация производится на основе стандартных алгоритмов кластеризации, позволяющих по ряду признаков определить достоверность личности пользователя, введившего определенный пароль. Однако, недостатком реализации является то, что он работает только для одного конкретного пароля, в момент его ввода, а не во время работы пользователя с системой в фоновом режиме. [1]

Метод №2. Кластеризация пользователей по «клавиатурному почерку» с использованием классической кластеризации

Другим возможным алгоритмом аутентификации является метод, основанный на кластеризации пользователей одним из классических алгоритмов машинного обучения (таких как, например, k-средних или k ближайших соседей). В данном методе предлагается использовать представление данных пользователя, основанные на методе потенциальных функций (частном случае метода ближайших соседей, позволяющего оценивать вес объектов обучающей выборки). Основной метрикой является расстояние между очередным нажатием и шаблоном пользователя.

Каждый пользователь характеризуется некоторым набором признаков, отражающих его «клавиатурный почерк». При периодической аутентификации, входной вектор или матрица определяется к соответствующему кластеру, определенному в системе. Если этот кластер соответствует пользователю, который в данный момент работает с системой, то прохождение аутентификации считается успешным.

Недостаток этого метода в том, что пользователь, которого нет в базе данных все-равно будет отнесен к какому-либо кластеру. Чтобы этого не произошло – следует ввести какую-либо метрику, при превышении которой пользователь не будет отнесен ни к какому кластеру (например, евклидово метрическое расстояние между объектами «пользователь в базе данных» и объектом-матрицей, характеризующей пользователя, проходящего аутентификацию в текущий момент времени). [2]

Метод №3. Аутентификация пользователя, основанная на кластеризации Partition Around Medoids

Данный метод основан на диграфах – структуры данных, характеризующей вектор времени между последовательными нажатиями на клавиши клавиатуры. Затем каждый объект, состоящий из набора диграфов, относится к какому-либо кластеру, который представляет из себя пользователя, считающегося эталоном в базе. Кластеризация осуществляется с помощью метода partition k-medoids или partitioning around medoids. [3]

Метод №4. Аутентификация на основе классификатора, представляющего из себя «дерево решений» или «случайный лес»

Данный метод представляет из себя развитие метода №2 с тем лишь различием, что используется не кластеризация, а классификаторы, такие как, например, дерево решений или случайный лес. Из недостатков этого метода можно выделить то, что для его корректной работы необходимо собрать достаточное количество статистических данных (для валидной классификации).

Метод №5. Метод наказания-поощрения

Данный метод является модернизацией и расширением предыдущих, может быть построен на основе любого из них. Метод основан на динамическом изменении уровня недоверия к пользователю, что позволяет оперативно отслеживать подмену пользователя и доступ к каким-либо конфиденциальным ресурсам. Каждое регламентированное действие пользователя проверяется системой. В случае успешной проверки уровень доверия увеличивается. В случае неуспешной – уменьшается. При достижении критического уровня доверия – доступ к системе блокируется. [4]

Метод №6. Корреляционный метод аутентификации пользователя

Данный метод характеризуется тем, что из данных, собранных в процессе работы пользователя с клавиатурой, формируется матрица, содержащая длительность нажатия на каждую клавишу, время между нажатиями двух разных клавиш. Затем высчитывается корреляция между эталонными векторами в базе данных и векторами, которые характеризуют пользователя, проходящего аутентификацию в текущий момент времени. Из особенностей данного метода – он хорошо адаптируется под ситуации, когда пользователь, например, увеличил свою скорость печати.

Метод №7. Аутентификация на основе нейронных сетей

В данном случае сформированный батч данных, содержащий информацию о нажатых пользователем клавишах, отправляется на вход нейронной сети. Данный метод является одним из наиболее перспективных и современных. Нейронные сети представляют из себя универсальный аппроксиматор, позволяющий аппроксимировать практически любую математическую функцию, а значит, в теории и выявлять любые зависимости в данных.

Архитектура используемой нейронной сети может варьироваться, чтобы наиболее эффективно и достоверно определять пользователя. Это могут быть как рекуррентные нейронные сети (например, LSTM или GRU), способные «запоминать» предыдущие характеристики пользователя и менять в зависимости от них пороговое значение, либо сверточные нейронные сети, способные преобразовывать входные данные в

некоторые эмбединги, также это могут быть и обычные нейронные сети, представляющие из себя многослойный персептрон.

Таблица 1.

Сравнение методов аутентификации пользователей (вычисленный на основе имеющихся реализаций и экспертных оценок сотрудников кибер-безопасности)

№ метод а	Устойчивость к ошибкам 1-го рода (ложное срабатывание - неудачный доступ подлинного сотрудника)	Устойчивость к ошибкам 2-го рода (пропуск события - аутентификация злоумышленника)	Сложность реализации алгоритма или метода (0 - легкий в реализации, 5 - сложный в реализации)	Возможность дообучения алгоритма в процессе работы
1	++	+++	+	-
2	++	++	+	+
3	++	++	+	+
4	++	++	++	+
5	+++	+	+++	+
6	+++	+++	++++	+
7	От + до ++++ (в зависимости от архитектуры нейронной сети)	От + до ++++ (в зависимости от архитектуры нейронной сети)	+++++	+

Заключение

Рассмотренные методы являются наиболее популярными и позволяют осуществлять аутентификацию пользователя прямо во время работы с персональным компьютером и проверять достоверность его личности на протяжении всей работы с машиной. Также стоит отметить, что существующие методы продолжают совершенствоваться из года в год, но их точность часто определяется степенью эффективности выбора параметров, настройки и использования моделей. Все упомянутые в статье методы требуют долгого вычисления на базе данных даже в 100 пользователей (что эквивалентно компании малого размера), а с учетом необходимости дообучения – целесообразно производить обучения алгоритмов на высокопроизводительных кластерах с использованием графических ускорителей.

1. В.Ю. Каганов, А. К. Королев, М. Н. Крылов, И. В. Машечкин, М.И. Петровский (2015) - «Методы машинного обучения в задачах аутентификации с использованием информации о динамике набора пароля на клавиатуре»;
2. В.Ю. Каганов, А. К. Королев, М. Н. Крылов, И. В. Машечкин, М.И. Петровский (2013) – «Методы активной аутентификации на основе анализа динамики работы пользователей с клавиатурой»
3. Bertacchini M., Benitez C., Fierens P.I. User clustering based on keystroke dynamics // XVI Congreso Argentino de Ciencias de la Computaci, on. – Mor, on, 2010. P. 832– 841.
4. Monroe F., Rubin A.D. Keystroke dynamics as a biometric for authentication // Future Generation Computer Syst., 2000. Vol. 16. No.4. P. 351–359.
5. Савинова В.М., Бесхмельницкий А.А., Бибина Е.С., Осадчая А. Д.. Идентификация пользователей корпоративной системы с помощью поведенческого анализа с использованием модели искусственной нейронной сети, 2017
6. Saleh M. Abu-Soud (2013) – «A Hybrid Keystroke Authentication Dynamics»;
7. Marcus V.S. de Oliveira Paula, Eduardo A, Kinto, Emilio D.M. Hernandez, Tereza Cristina M. de B. Carvalho (2007) – «User Authentication based on Human Typing Pattern with Artificial Neural Networks and Support Vector Machine»
8. Kenneth Revett, Florin Gorunescu, Marina Gorunescu, Marius Ene, Sergio de Magalhaes, Henrique M. Dinis Santos (2008) – «A machine learning approach to keystroke dynamics based user authentication»
9. Edmond Lau, Xia Liu, Chen Xiao, Xiao Yu (2004) – «Enhanced User Authentication Through Keystroke Biometrics»
10. Nandini Chourasia, Prof. Dr. Prasanna Joeg (2014) – «Authenticate User by Keystroke Dynamics»

11. Файсханов И.Ф. — Аутентификация пользователей при помощи устойчивого клавиатурного почерка со свободной выборкой текста // Кибернетика и программирование. – 2018. – № 3. – С. 72 – 86
12. Савинов А. Н. Методы, модели и алгоритмы распознавания клавиатурного почерка в ключевых системах: автореферат диссертации на соискание ученой степени кандидата технических наук: 05.13.19. -Санкт-Петербург, 2013. - 19 с.
13. Электронный ресурс удаленного доступа: Horovod Concepts. Дата обращения 19.02.2020, <https://github.com/horovod/horovod/blob/master/docs/concepts.rst>

Прокофьев О.В.

Требования к системам поддержки принятия решений и их реализация

Пензенский государственный технологический университет

(Россия, Пенза)

doi: 10.18411/lj-04-2020-29

idsp: ljjournal-04-2020-29

Аннотация

Системы поддержки принятия решений (СППР), получившие широкую известность с начала 1980-х годов, сейчас являются одним из быстро развивающихся и востребованных направлений информационных систем. Статья посвящена причинам того, что ожидания и подразумеваемые требования к разрабатываемым системам не всегда соответствуют результату внедрения.

Ключевые слова: системы поддержки принятия решений, широкомасштабные СППР, корпоративные СППР, внедрение СППР

Abstract

Decision Support Systems (DSS), widely known since the early 1980s, are now one of the fastest growing and sought-after areas of information systems. The article is devoted to the reasons that the expectations and implied requirements for the developed systems do not always correspond to the implementation result.

Keywords: decision support systems, large-scale DSS, corporate DSS, implementation of DSS

Понятие систем поддержки принятия решений (DSS) включает в себя широкий диапазон различных систем, средств и технологий. Оно было и остаётся многофункциональным понятием для многих видов информационных систем, которые поддерживают процесс принятия решений [1]. Исполнительные системы для решений руководителей компаний (EIS), географические информационные системы (GIS), исполнительные системы поддержки принятия решений (ESS), аналитические процессы для принятия групповых решений онлайн (GDSS) могут быть включены в категорию систем под названием DSS.

Повышенное внимание вызывает внедрение широкомасштабных DSS с обширными информационными базами, которые, позволяют принимающим решение найти практически любую информацию о их компании в нужное время. Принимающие решение могут черпать, разделять, объединять, наглядно и графически изображать корпоративные и внешние данные. Широкомасштабная DSS может находиться в диапазоне, охватывающем как довольно простые системы, так и содержащие комплексную информацию, а также аналитически продвинутые EIS. На более продвинутом уровне мы можем найти системы для анализа информации [2-9], которые упрощают процесс использования информации при помощи таких компьютеризированных аналитических средств, как пакеты статистической обработки, поиск информации и т.д. Наиболее продвинутые широкомасштабные аналитические системы обеспечивают доступ к ряду ориентированных на принятие решений баз