АНАЛИЗ РАЗМЕЩАЕМЫХ В СЕТИ ОТКРЫТЫХ ДАННЫХ В ЦЕЛЯХ ПОЛУЧЕНИЯ ИНФОРМАЦИИ О КРИМИНОГЕННОЙ ОБСТАНОВКЕ

Жарова А. К.[1], Елин В. М.[2], Атласов И. В.[3]

DOI: 10.21681/2311-3456-2025-4-152-159

Цель статьи: *предложить методику формирования цифрового профиля человека, который может быть использован для анализа и прогнозирования криминогенной обстановки.*

Метод исследования: *использованы логико-математические методы, такие как типологическая модель, детерминистская модель и имитационное моделирование. Кроме того, используется метод анализа математических моделей, включая стохастическую модель, что позволяет получить более точную и детализированную картину. Входными данными для анализа являлись данные, оставляемые человеком в процессе своей деятельности в интернете*

Результат: *системы анализа данных могут быть применены для извлечения, анализа, преобразования и представления информации, имеющей существенное значение при проведении оперативно-розыскных и следственных мероприятий. Авторы, ссылаясь на имеющуюся судебную практику, раскрывают значение коммуникационных данных для получения цифрового профиля человека, формально не относящихся к персональным данным как категории информации ограниченного доступа. Экспериментальная часть статьи представляет собой математическое моделирование криминогенной обстановки на основании анализа независимых цифровых данных, оставленных пользователем социальной Сети. Таким образом, в результате проведенного исследования выявлены закономерности, позволяющие в дальнейшем предсказывать поведение групп людей, осуществляющих передачу вредоносной информации в сети, либо размещение информации указанной категории.*

Практическая ценность: *на основании проведенного теоретического эксперимента сделан вывод о возможности применения математических методов в криминологическом анализе преступности*

Ключевые слова: *информационные технологии, коммуникационные данные, анализ цифровых теней и цифровых следов, персональные данные, математическое моделирование.*

Введение

Активность пользователей Интернета фиксируется и отражается в тех или иных интернет-данных, которые в дальнейшем могут быть проанализированы в целях получения цифрового профиля человека. Каждый фрагмент интернет-контента окружён множеством элементов коммуникационных данных. Даже в случае, когда контент зашифрован и системы анализа данных не могут получить информацию о содержании персональных данных (какую-либо личную информацию об отправителе или получателе), существующие системы анализа коммуникационных данных (в том числе, связанных с зашифрованным контентом), могут содержать обширный массив личной информации, включая личность, его друзей, географические координаты отправителя и получателя, IP устройства передачи сообщения, включая его полные технические характеристики. Анализ связанных коммуникационных данных усиливает возможности получения полной информации о человеке[4].

Европейский суд по правам человека (ЕСПЧ) обратил внимание на проблему, связанную с тем, что системы анализа данных могут формировать достоверный цифровой профиль человека на основе остаточных цифровых следов человека, оставленных им в интернете. В своём постановлении ЕСПЧ пришёл к выводу, что в интернете содержится гораздо больше коммуникационных данных, чем самого контента[5].

В России, согласно Федеральному закону «О персональных данных»[6], коммуникационные данные сами по себе не являются персональными. Однако, когда они собираются вместе и обрабатываются с помощью специальных систем анализа данных, они могут быть преобразованы в персональные

1 Жарова Анна Константиновна, доктор юридических наук, профессор Финансового университета при Правительстве Российской Федерации, Москва. E-mail: anna_jarova@mail.ru

2 Елин Владимир Михайлович, кандидат педагогических наук, доцент кафедры информационной безопасности Московского университета МВД России имени В.Я.Кикиотя, доцент кафедры информационной безопасности Финансового университета при Правительстве Российской Федерации, Москва. E-mail: elin_vm@mail.ru

3 Атласов Игорь Викторович, доктор физико-математических наук, профессор Московского университета МВД России имени В.Я.Кикиотя, Москва. E-mail: atlasov.igor.777@gmail.com

4 Big Brother Watch and Others v. the United Kingdom, Application Nos. 58170/13, 62322/14 and 24960/15 (ECtHR May 25, 2021).

5 Big Brother Watch and Others v. the United Kingdom.

6 Федеральный закон от 27.07.2006 N 152-ФЗ «О персональных данных»//СЗ РФ. 2006. № 31 (1 ч.) Ст. 3451.

данные [1, 2]. Более того, в нашей стране уже сложилась судебная практика по делам, в которых оспаривается правомерность применения подобных систем анализа данных[7,8].

Системы анализа больших данных способны как обрабатывать коммуникационные данные без необходимости идентификации пользователя сети [3, 4], так и идентифицировать конкретного человека. Например, в целях выявления лица, совершившего правонарушение с использованием информационно-коммуникационных технологий (ИКТ).

Можно ли скрыть данные для систем анализа данных?

Возможность обеспечения конфиденциальности информации ограниченного доступа зависит от используемых методов обработки информации. Для технологий анализа интернет-данных основным является факт их размещения в интернете. Сочетание технологий анализа больших данных и искусственного интеллекта (ИИ) позволяет улучшить прогнозирование, принятие решений, оптимизировать процессы и автоматизировать решение задач, например, провести анализ любой информации, в том числе, данных, которые так или иначе можно отнести к персональным данным [5, 6.]. Например, для решения задачи анализа информации, размещенной на странице социальной сети, в отношении которой ее обладатель поставил ограничения по доступу к ней только определенной группы людей, например, друзей.

Но и даже в том случае, если информация не размещена в Сети, технологии анализа данных могут получить ее самостоятельно на основе анализа цифровых теней и цифровых следов [7]. Например, швейцарские учёные провели эксперимент, в целях подтверждения гипотезы – могут ли большие языковые модели (LLM) собирать и раскрывать личную информацию пользователей. В качестве примера ученые взяли 1,5 тысячи случайных профилей с площадки Reddit и проанализировали их активность с помощью LLM. LLM смогли точно определить место рождения и жительства, а также уровень дохода людей по вторичной информации, оставленной пользователями Сети. GPT-4 идентифицировал с точностью 85 %, а LlaMA-2-7b с точностью 51 %[9].

Пока наиболее действенным методом противодействия получению доступа к содержанию информации является метод криптографического шифрования данных. Но, со временем, как только будет создан квантовый компьютер, все зашифрованные данные станут доступны для изучения [8].

В связи с этим проблема защиты права человека на неприкосновенность частной жизни в связи с массовым использованием различных технологий анализа данных [9, 10], со временем не только не теряет своей остроты, но и становится всё более значимой.

Для охвата всех возможных цифровых данных, оставленных пользователями в Сети, например, в 2022–2024 годах в странах Евросоюза (Германия[10] и Англия[11]) внесены изменения в законодательство, разрешающие массовый перехват сообщений. Тем самым в европейских странах требования о массовом перехвате данных стали законным инструментом для обработки больших данных, применяемым в целях получения разведывательной информации и выявления новых угроз, которые могут исходить как от известных, так и от неизвестных источников.

Рассмотрим возможность получения представления о причастности лица к тем или иным аспектам криминогенной деятельности с помощью анализа открытых данных из сетей с применением математических и логико-математических методов. Также оценим интернет-активность людей. Это позволит нам получить количественные характеристики данных процессов, выявить их закономерности, оптимизировать и прогнозировать поведение человека.

Использование в криминологической деятельности математических методов и моделей

В работе правоохранительных органов широко применяются специальные математические методы. В правовой сфере на смену приближенным качественным оценкам все чаще приходят точные количественные оценки. В правоохранительной деятельности актуально понимание возможностей математического моделирования, анализа, поддержки принятия решений, причинно-следственного анализа и вывода. Так, в целях математической оценки криминогенной обстановки[12] может проводиться оценка процессов и параметров математическими методами. На основании применения этих методов, строятся различные модели оценки криминогенной обстановки. Анализируемыми процессами в этом случае выступают действия пользователей в социальных сетях, а параметры меняются в зависимости от применяемой модели.

7 Определение Конституционного Суда РФ от 02.10.2003 N 345-О «Об отказе в принятии к рассмотрению запроса Советского районного суда города Липецка о проверке конституционности части четвертой статьи 32 Федерального закона от 16 февраля 1995 года «О связи» // Вестник Конституционного Суда РФ», N 1, 2004.

8 Определение Верховного Суда РФ от 29 января 2018 г. N 305-КГ17-21291 // СПС «КонсультантПлюс».

9 Нейросети раскрыли личные данные пользователей соцсетей // https://lenta.ru/news/2023/11/01/llm/

10 Gesetz zur Beschränkung des Brief-, Post- und Fernmeldegeheimnisses (Artikel 10-Gesetz – G 10) // https://www.gesetze-im-internet.de/g10_2001/BJNR125410001.html

11 Interception of communications code of practice 2022 // https://www.gov.uk/government/publications/interception-of-communications-code-of-practice-2022/interception-of-communications-code-of-practice-2022-accessible

12 Криминогенная обстановка определяется как совокупность процессов и их параметров, влияющих на состояние и динамику преступности.

Например, ***стохастическая модель динамики преступности*** базируется на том, что преступность — это случайный процесс, который зависит от множества факторов. Поэтому основным математическим аппаратом, применяемым в этой модели, является теория вероятностей и случайных процессов. Наличие множества возможных состояний, а также связей между ними и с окружающей средой позволяет считать структуру моделируемой системы заданной [11].

Если в исследуемых негативных процессах можно определить два состояния k и l, то можно построить граф, который будет отображать эти состояния и возможные переходы между ними в течение небольшого периода времени Δt. Этот граф представляет собой развитие модели Марковского процесса, где узлы отражают состояния моделируемого объекта, а дуги — вероятность перехода из одного состояния в другое.

Вероятно, модель динамики преступности можно отнести к непрерывной цепи Маркова, где система может менять свое состояние в любой момент времени (q-схема).

Граф изменения состояний в стохастической модели преступности представлен на рис. 1.

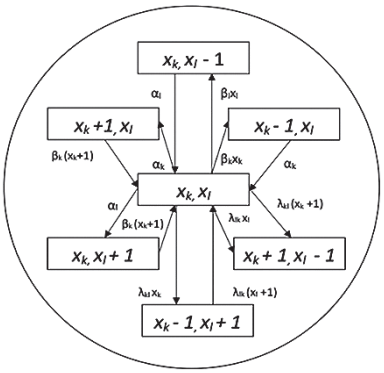

Рис. 1. Граф изменения состояний в стохастической модели динамики преступности

Детерминистская модель анализа динамики преступности позволяет оценить негативные процессы, происходящие в интернете. Она помогает анализировать особенности и закономерности преступной деятельности в Сети. Хотя эта математическая модель не может точно воспроизвести уникальность каждого отдельного преступления, она позволяет выявить факторы, влияющие на динамику процессов.

Поскольку в отношении каждого фактора достаточно четко определены количественные и качественные характеристики, параметры оценки могут быть положены в основу моделируемой системы.

Благодаря этому можно представить общую криминологическую картину и объяснить многие эмпирические наблюдения в этой области. Кроме того, создание математической модели динамики преступности может служить методологической основой для разработки системы криминологических гипотез.

При создании этих моделей индивид выступает в роли элемента системы, а группа индивидов формирует определённое множество процессов. Между ними существует взаимосвязь, которая определяется цепочкой переходов.

Модель устанавливает статистическую зависимость между следующими параметрами:

- Количество индивидов, находящихся в конкретном состоянии в определённый момент времени.
- Параметры, описывающие переходы между состояниями.
- Потоки индивидов, входящих в систему и покидающих её (рис. 2), где блоки – это люди, а дуги графа – это отношения между ними.

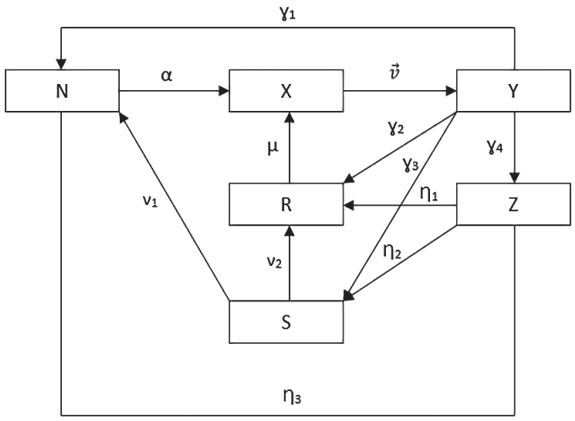

Рис. 2. Граф детерминистской модели динамики преступности.

Имитационная модель анализа динамики преступности является следующим этапом в развитии моделей, применяемых для анализа процессов и представляет собой универсальный программный комплекс, который позволяет воссоздать функционирование сложного процесса. Суть имитационного моделирования заключается в проведении эксперимента, позволяющего определить во времени варианты поведения людей при изменении внешних воздействующих факторов.

В настоящее время в криминологической практике имитационные модели стали эффективным инструментом для анализа, прогнозирования и управления в сфере уголовной юстиции. Они дают возможность комплексно исследовать преступность, учитывая влияние изменяющихся факторов. Так, например, российскими исследователями описана имитационная модель получения вероятностно-временных оценок длительности действий органов внутренних дел

при возникновении чрезвычайных обстоятельств криминального характера на примере массовых беспорядков [12], а также произведено использование модели на примере одного из видов массовых беспорядков с учетом фактора латентности [13].

Математическое моделирование криминогенной обстановки на основании независимых цифровых данных, оставленных пользователем социальной Сети

В данном разделе представлено решение задачи получения достаточно достоверных сведений о физическом лице на основании обработки данных, оставленных человеком в Сети, которые в дальнейшем будут служить входными данными для моделирования криминогенной обстановки, криминогенных наклонностях или интереса некоторой социальной группы к противоправной деятельности, вредоносному материалу, информации вредоносной направленности.

Для этого выберем такие натуральные числа n, s, l, при которых будет справедливо равенство $n = sl$. Здесь считаем $n = 1000$ и $l = 100$.

Далее рассмотрим n человек в социальных сетях. Выделяем тех людей, которые достаточно часто (например, 10 раз) размещают информацию, связанную со своими интересами, в том числе информацию криминальной направленности, которую можно определить как вредоносную информацию.

Для каждого i:$1 \le i \le n$ человека построим случайную величину ξ_i, принимающую два значения 1 и 0 в зависимости от того, было ли 10 упоминаний в социальных сетях об определенной вредоносной категории информации, лежащих в некотором промежутке, необходимой нам для исследования.

$$\xi_i = \begin{pmatrix} 1, & \text{были упоминания с вероятностью} \quad p = P(\xi_i = 1) \\ 0, & \text{не было упоминаний с вероятностью} \quad q = P(\xi_i = 0) \end{pmatrix}, \quad (1)$$

где $p + q = 1$ и значения этих величин неизвестны. Предположим, что случайные величины $\{\xi_i\}_{i=1}^n$ независимы в совокупности, то есть справедливо равенство

$$P(\xi_1 < t_1, \ldots, \xi_n < t_n) = P(\xi_i < t_1) \ldots P(\xi_n < t_n),$$

справедливое для всех действительных чисел $\{t_i\}_{i=1}^n$. Это условие говорит о том, что люди в социальных сетях, размещающие информацию о своих интересах, практически не знают друг друга. Они могут размещать значимую информацию, в том числе одобрять или не одобрять какую-либо информацию.

Нам понадобится еще одно определение.

Определение 1. *Пусть задана последовательность функций* $\{f_n(x)\}_{n=1}^{\infty}$ *и функция* $f(x)$, *со значениями в множестве действительных чисел, заданных для всех* $x \in D$, *где* D – *замыкание некоторого открытого множества на прямой*.

Скажем, что последовательность функций $\{f_n(x)\}_{n=1}^{\infty}$ равномерно сходится к функции $f(x)$, если для некоторого $\varepsilon > 1$ существует такое натуральное n_0, что для всех натуральных $n > n_0$ справедливо равенство

$$f_n(x) - f(x) \vee \varepsilon$$

для всех $x \in D$. Обозначим этот факт через $f_n(x) \longrightarrow f(x)$.

Заметим, что случайная величина ξ_i обладает математическим ожиданием $M(\xi_i) = 1 \cdot p + 0 \cdot q = p$. Квадрат случайной величины ξ_i^2 также обладает математическим ожиданием $M(\xi_i^2) = 1^2 \cdot p + 0^2 \cdot q = p$. Поэтому случайная величина ξ_i обладает дисперсией $D(\xi_i) = M(\xi_i^2) - M^2(\xi_i) = p - p^2 = pq$.

Согласно центральной предельной теореме, для одинаково распределенных случайных величин, имеющих дисперсию [14], справедливо утверждение

$$P\left(\frac{1}{\sqrt{lD(\xi_i)}}\sum_{k=1}^{l}\left(\xi_{l\cdot(j-1)+k} - M(\xi_i)\right) < t\right) =$$

$$= P\left(\frac{1}{\sqrt{lpq}}\sum_{k=1}^{l}\left(\xi_{l\cdot(j-1)+k} - p\right) < t\right) \longrightarrow \frac{1}{2\pi}\int_{-\infty}^{t} e^{\frac{-x^2}{2}}dx,$$

которое означает, что для некоторого $\varepsilon > 0$ существует натуральное n_0, такое, что для всех натуральных $l > n_0$ справедливо равенство

$$\left|P\left(\frac{1}{\sqrt{lpq}}\sum_{k=1}^{l}\left(\xi_{l\cdot(j-1)+k} - p\right) < t\right) \longrightarrow \frac{1}{2\pi}\int_{-\infty}^{t} e^{\frac{-x^2}{2}}dx\right| < \varepsilon,$$

$$\left|P\left(\frac{1}{l}\sum_{k=1}^{l}\xi_{l\cdot(j-1)+k} < t\sqrt{\frac{pq}{l}} + p\right) - \frac{1}{2\pi}\int_{-\infty}^{t} e^{\frac{-x^2}{2}}dx\right| < \varepsilon.$$

Обозначим $y = t\sqrt{\frac{pq}{l}} + p$, $t = (y - p)\sqrt{\frac{l}{pq}}$.

В этом случае имеем

$$\left|P\left(\frac{1}{l}\sum_{k=1}^{l}\xi_{l\cdot(j-1)+k} < y\right) - \frac{1}{2\pi}\int_{-\infty}^{(y-p)\sqrt{\frac{l}{pq}}} e^{\frac{-x^2}{2}}dx\right| < \varepsilon.$$

Упростим интеграл

$$\frac{1}{2\pi}\int_{-\infty}^{(y-p)\sqrt{\frac{l}{pq}}} e^{\frac{-x^2}{2}}dx = \left|\begin{array}{l} x = (z-p)\sqrt{\frac{l}{pq}}\ dx = \sqrt{\frac{l}{pq}}\ dz \\ x = (z-p)\sqrt{\frac{l}{pq}}\ z = y \\ x = -\infty,\ z = -\infty \end{array}\right| =$$

$$= \frac{1}{2\pi}\sqrt{\frac{l}{pq}}\int_{-\infty}^{y} e^{\frac{-(z-p)^2}{l}}dz.$$

Окончательно имеем

$$\left|P\left(\frac{1}{l}\sum_{k=1}^{l}\varepsilon_{l(j-1)+k} < y\right) - \frac{1}{2\pi}\sqrt{\frac{l}{pq}}\int_{-\infty}^{y} e^{\frac{-(z-p)^2}{2\left(\sqrt{\frac{pq}{l}}\right)^2}}dz\right| < \varepsilon$$

Таким образом мы доказали, что для всех возможных, $1 \le j \le \frac{n}{l}$ функция распределения случайной величины $\xi_i = \frac{1}{l}\sum_{k=1}^{l}\varepsilon_{l(j-1)+k}$ равномерно сходится к нормальной случайной величине с математическим ожиданием $m = p$ и дисперсией $\sigma^2 = \frac{pq}{l}$. То есть можно считать, что случайная величина $\xi_i = \frac{1}{l}\sum_{k=1}^{l}\varepsilon_{l(j-1)+k}$ является нормальной случайной величиной с математическим ожиданием p и дисперсией $\sigma^2 = \frac{pq}{l}$.

В дальнейшем необходимо найти некоторое среднее значение (математического ожидания) для группы случайных величин, рассмотренной выше, с определенной степенью достоверности γ, под которой нам следует понимать уровень уверенности в результатах исследования и выводах. (Так, при высоком уровне достоверности, если мы примем $\gamma = 0.95$, можно сделать вывод о том, что при повторении эксперимента независимой группой исследователей мы получим тот же результат в 95 случаев из 100).

Итак, если нам удастся оценить число m, то мы сможем сказать, что с достоверностью γ (в примере: 95 случаев из 100) каждые ml человек из l совершают действия в описанных нами промежутках, или sml из n осуществляют деятельность (интересуются вредоносной информацией) в описанных нами промежутках.

Далее наша задача заключается в оценке математического ожидания нормальной случайной величины τ_i. Для решения указанной задачи авторами предлагается использовать распределение Стьюдента, отвечающее признакам определения 2.

Определение 2. *Распределение τ с плотностью вероятности*

$$f_\tau(t) = \frac{1}{\sqrt{\pi n}} \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{t^2}{n}\right)^{\frac{-n+1}{2}},$$

называется распределением Стьюдента с n степенями свободы. Символом $\Gamma(x)$ обозначена гамма-функция

$$\Gamma(k) = \int_0^\infty t^{k-1}\, exp(-t)dt,$$

определенная для всех $k > 0$.

Далее, вычислим величины

$$\bar{\zeta} = \frac{1}{l} \Sigma_{k=1}^{l} \xi_k 2, \tag{2}$$

$$\bar{\theta} = \frac{1}{l} \Sigma_{k=1}^{l} \xi_k 2 - (\bar{\zeta})^2. \tag{3}$$

Случайная величина

$$\tau = \frac{\sqrt{l}\,(\bar{\zeta} - m)}{\sqrt{\frac{n\bar{\theta}^2}{l-1}}} = \sqrt{l-1}\,\frac{\bar{\zeta} - m}{\bar{\theta}} \tag{4}$$

имеет распределение Стьюдента с $l - 1$ степенями свободы, или

$$P\left(t_1 < \sqrt{l-1}\,\frac{\bar{\zeta} - m}{\bar{\theta}} < t_2\right) = \int_{t_1}^{t_2} f_\tau(x)dx.$$

Или то же самое утверждение – для плотности вероятности f_τ случайной величины τ с $l - 1$ степенями свободы справедливо равенство

$$P\left(\bar{\zeta} - \frac{\bar{\theta} t_2}{\sqrt{n-1}} < m < \bar{\zeta} - \frac{\bar{\theta} t_1}{\sqrt{n-1}}\right) = \int_{t_1}^{t_2} f_\tau(x)dx =$$

$$= \int_0^{t_2} f_\tau(x)dx - \int_0^{t_1} f_\tau(x)dx.$$

То есть за счет выбора n для любых значений t_1 и t_2 можно сделать величину $\gamma = \int_{t_1}^{t_2} f_\tau(x)dx$ сколь угодно близкой к единице. Далее за счет выбора t_1 и t_2 сделать интервал

$$\left[\bar{\zeta} - \frac{\bar{\theta} t_2}{\sqrt{n-1}} < m < \bar{\zeta} - \frac{\bar{\theta} t_1}{\sqrt{n-1}}\right]$$

сколь угодно малой длины и окончательно сделаем вывод – с вероятностью γ величина m лежит на отрезке

$$\left[\bar{\zeta} - \frac{\bar{\theta} t_2}{\sqrt{n-1}} < m < \bar{\zeta} - \frac{\bar{\theta} t_1}{\sqrt{n-1}}\right].$$

Таким образом, поставленная перед нами задача определения достаточно достоверных необходимых сведений для идентификации физического лица в результате обработки данных, формирующих косвенные признаки на основании определения группы людей, проявляющих интерес к вредоносной информации, имеет следующее решение:

Пусть изучаются 1000 человек на используемую ими вредоносную информацию. В результате, согласно формуле (1), получим 1000 независимых дискретных случайных величин ξ_i.

Далее, разобьем все испытания на непересекающиеся классы по 100 элементов в каждом и посмотрим сколько раз в каждом классе было более 10 упоминаний об информации, отнесенной к вредоносной. Далее, разобьем все испытания на непересекающиеся классы по 100 элементов в каждом и посмотрим сколько раз в каждом классе было более 10 обращений к вредоносной информации, что отобразим в таблице 1.

Будет получена таблица.

Таблица 1.

Количество обращений к вредоносной информации для каждого из классов.

классы	**1**	**2**	**3**	**4**	**5**
обращения	53	48	74	26	35
классы	**6**	**7**	**8**	**9**	**10**
обращения	66	79	81	95	18

Согласно таблице 1, в первом классе 53 раза было более 10 обращений к вредоносной информации, в втором классе 48 раз было более 10 обращений, и так далее.

То есть в соответствии с (1) случайные величины ε_i, i = 1, ..., 100 приняли значение 1 ровно 53 раза (т.е. 53 раза было упоминание указанной категории информации и значение 0 ровно 47 раз (не было упоминаний информации (100 – 53 = 47)).

Также, применяя (1), получаем, что случайные величины ε_i, i = 101, ..., 200 приняли значение 1 ровно 48 раз и значение 0 ровно 52 раза и так далее.

Как доказано в работе, случайные величины $\zeta_j = \frac{1}{100}\sum_{k=1}^{100}\varepsilon_{100*(j-1)+k}$ можно считать нормальными случайными величинами. В итоге получим таблицу 2 нормальных случайных величин для каждого из классов.

Таблица 2.

Значения нормальных случайных величин

случайная величина	ζ_1	ζ_2	ζ_3	ζ_4	ζ_5
значение сл. вел.	0,53	0,48	0,74	0,26	0,35
случайная величина	ζ_6	ζ_7	ζ_8	ζ_9	ζ_{10}
значение сл. вел.	0,66	0,79	0,81	0,95	0,18

Подставляя значения случайных величин в формулы (2) и (3), получим:

$$\bar{\zeta} = \frac{1}{10}\sum_{k=1}^{10}\xi_k = \frac{1}{10}(0{,}53+0{,}48+0{,}74+0{,}26+0{,}35+ +0{,}66+0{,}79+0{,}81+0{,}95+0{,}18) = 0{,}575 \quad (5)$$

$$\bar{\theta}^2 = \frac{1}{10}\sum_{k=1}^{10}\xi^2 - (\bar{\zeta}) = = \frac{1}{10}(0{,}53^2+0{,}48^2+0{,}74^2+0{,}26^2+0{,}35^2+0{,}66^2+ +0{,}79^2+0{,}81^2+0{,}95^2+0{,}18^2) - 0{,}575^2 = 0{,}059345 \quad (6)$$

Подставляя полученные значения (5) и (6) в формулу (4), получаем, что значение случайной величины

$$\tau = \sqrt{l-1}\,\frac{\bar{\zeta}-m}{\bar{\theta}} = \sqrt{10-1}\,\frac{0{,}575-m}{\sqrt{0{,}059345}} = 12{,}31 \quad (7)$$

имеет распределение Стьюдента с 9 степенями свободы, или,

$$P(0{,}575 - 0{,}0812t_2 < m < 0{,}575 - 0{,}0812t_1) = = \int_{t_1}^{t_2} f_\tau(x)dx.$$

По таблицам для $\gamma = 0{,}95$ из неравенства $\int_{t_1}^{t_2} f_\tau(x)dx > \gamma$ выбираем $t_1 = 0{,}12$ и $t_2 = 0{,}83$.

$P(0{,}508 < m < 0{,}584) > 0{,}95$, примем усредненное $m = 0{,}54$.

В итоге получим, что в 95 случаях из 100 аналогичных случаях исследований, в среднем $mls = 0{,}54 * 100 * 10 = 540$ человек осуществляют обращение вредоносной информации.

Заключение

Таким образом, представленные в настоящей статье вычисления доказывают возможность применения математических методов в криминологическом анализе преступности и позволяют осуществлять цифровое профилирование человека на основании данных, размещаемых в социальных сетях и формально не относящихся к персональным данным, как категории информации ограниченного доступа.

Для достижения заявленных целей сформулирована задача получения с достаточной достоверностью криминологически значимой информации в результате обработки данных, оставленных в Сети группой пользователей, и на основе косвенных признаков, на основании анализа которых и полученных закономерностей можно сделать предположение о криминогенных наклонностях или интересе некоторой социальной группы к противоправной деятельности, вредоносному материалу, информации вредоносной направленности.

Таким образом, поставленная перед нами задача об определении достаточно достоверных необходимых сведений для идентификации физического лица в результате обработки данных, формирующих косвенные признаки на основании поведения группы незнакомых друг с другом людей, проявляющих интерес к вредоносной информации, либо же размещающих значимую информацию о своих интересах, решена.

С помощью проведенного исследования представлена модель оценки с высокой достоверностью γ (в примере: 95 случаев из 100) о том, что ml человек из l совершают действия в описанных нами промежутках, или sml из n осуществляют деятельность (интересуются вредоносной информацией) в описанных нами промежутках.

В результате проведенного исследования установлены закономерности, позволяющие в дальнейшем предсказывать поведение групп людей, осуществляющих обращение вредоносной информации в сети, либо предоставляющих информацию указанной категории.

Литература

1. Редкоус, В. М. О совершенствовании правовой основы деятельности органов внутренних дел по объявлению официальных предостережений / В. М. Редкоус // Закон и право. – 2020. – № 9. – С. 157–159. – DOI 10.24411/2073-3313-2020-10453.
2. Степанов, О. А. О правовых особенностях и рисках реализации цифрового профилирования / О. А. Степанов, Д. А. Басангов // Российская юстиция. – 2024. – № 1. – С. 59–69. – DOI 10.52433/01316761_2024_01_59.
3. «Цифровой поворот» в правовых исследованиях / И. Р. Бегишев, А. К. Жарова, Е. А. Громова [и др.] // Journal of Digital Technologies and Law. – 2024. – Т. 2, № 1. – С. 7–13. – DOI 10.21202/jdtl.2024.1.
4. Жарова, А. К. Система организационно-правового выявления лиц, разместивших информацию в интернете о намерении совершить преступление / А. К. Жарова // Пробелы в российском законодательстве. – 2024. – Т. 17, № 1. – С. 122–130. – DOI 10.33693/2072-3164-2024-17-1-122-130. – EDN OHAZYD.

5. Шутова, А. А. Обеспечение цифровой безопасности системы здравоохранения уголовно-правовыми средствами / А. А. Шутова // Russian Journal of Economics and Law. – 2024. – Т. 18, № 4. – С. 936–953. – DOI 10.21202/2782-2923.2024.4.936-953.
6. Дейнеко, А. Г. Публичное право в киберпространстве: публично-правовое регулирование информационных отношений / А. Г. Дейнеко. – Москва: Общество с ограниченной ответственностью «Проспект», 2025. – 248 с.
7. Жарова, А. К. Парадигма цифрового профилирования деятельности человека: риски, угрозы, преступления / А. К. Жарова, В. М. Елин, А. В. Минбалеев. – Москва: Общество с ограниченной ответственностью «Русайнс», 2022. – 240 с.
8. Жарова, А. К. Обзор нормативных требований, обеспечивающих национальную безопасность США в сфере квантовых технологий / А. К. Жарова // Информационное общество. – 2023. – № 3. – С. 69–77. – DOI 10.52605/16059921_2023_03_69.
9. Залоило, М. В. Циклично-волновая модель интерпретации истории права на основе теории технологических укладов / М. В. Залоило // Историко-правовой ежегодник – 2023. – Москва: Infotropic Media, 2024. – С. 48–72.
10. Твердова, Т. В. § 3. Риски правового регулирования отношений, возникающих по поводу искусственного интеллекта / Т. В. Твердова // Теоретико-правовая парадигма существования кибернетической (информационной) цивилизации : монография. – Москва : Межрегиональная общественная организация «Межрегиональная ассоциация теоретиков государства и права», 2022. – С. 244–273. – EDN ZXOJCC.
11. Максимов, С. В. Стохастическая модель репрессивно-превентивного воздействия на преступность: от интуиции к расчетам / С. В. Максимов, Ю. Г. Васин, К. А. Утаров // Всероссийский криминологический журнал. – 2021. – Т. 15, № 6. – С. 665–680. – DOI 10.17150/2500-4255.2021.15(6).665-680.
12. Моделирование процессов принятия решения в правоохранительной деятельности / О. Ю. Данилова, А. В. Меньших, В. В. Меньших [и др.]. – Воронеж : Воронежский институт Министерства внутренних дел Российской Федерации, 2021. – 103 с. – ISBN 978-5-88591-856-5. – EDN FENSWM.
13. Минаев В. А. Моделирование динамики преступности с учетом фактора латентности// Криминологический журнал. Естественные науки. Компьютерные науки и информатика. 2022. № 2. С. 67–78.
14. Малахова, В. В. Анализ статистических данных с использованием математического аппарата искусственного интеллекта / В. В. Малахова, О. В. Малахов // Вестник Луганского государственного университета имени Владимира Даля. – 2023. – № 11. – С. 177–179. – EDN EADYTE.

ANALYSIS OF OPEN DATA POSTED ON THE NETWORK IN ORDER TO OBTAIN INFORMATION ABOUT THE CRIME SITUATION

Zharova A. K.[13], Elin V. M.[14], Atlasov I. V.[15]

Keywords: *information technology, communication data, analysis of digital shadows and digital footprints, personal data, mathematical modeling.*

The purpose of the article *is to propose a method for forming a digital profile of a person, which can be used to analyze and predict the criminogenic situation.*

Research method: *logical and mathematical methods, such as typological model, deterministic model and simulation modeling, are used. In addition, the method of analysis of mathematical models, including the stochastic model, is used, which allows to obtain a more accurate and detailed picture.*

Result: *data analysis systems can be used to extract, analyze, transform and present information that is essential for operational-search and investigative activities. The authors, referring to the existing judicial practice, reveal the importance of communication data for obtaining a digital profile of a person, which formally do not belong to personal data as a category of restricted information. is a mathematical modeling of the criminogenic situation based on the analysis of independent digital data left by the user of the social network. Thus, as a result of the study, patterns have been identified that make it possible to predict the behavior of groups of people who transmit harmful information on the network, or the placement of information of this category.*

Practical value: *on the basis of the theoretical experiment, a conclusion is made about the possibility of using mathematical methods in the criminological analysis of crime.*

References

1. Redkous, V. M. O sovershenstvovanii pravovoj osnovy deyatel'nosti organov vnutrennih del po ob"yavleniyu oficial'nyh predosterezhenij / V. M. Redkous // Zakon i pravo. – 2020. – № 9. – S. 157–159. – DOI 10.24411/2073-3313-2020-10453. – EDN EIBWAV.
2. Stepanov, O. A. O pravovyh osobennostyah i riskah realizacii cifrovogo profilirovaniya / O. A. Stepanov, D. A. Basangov // Rossijskaya yusticiya. – 2024. – № 1. – S. 59–69. – DOI 10.52433/01316761_2024_01_59. – EDN JJSSAU.

13 Anna K. Zharova, Dr.Sc. (of Law), Professor, Financial University under the Government of the Russian Federation, Moscow. E-mail: anna_jarova@mail.ru
14 Vladimir M. Elin, Ph.D., Associate Professor of the Department of Information Security of the Moscow University of the Ministry of Internal Affairs of Russia named after V. Y. Kikiot, Associate Professor of the Department of Information Security of the Financial University under the Government of the Russian Federation, Moscow. E-mail: elin_vm@mail.ru
15 Igor V. Atlasov, Dr.Sc. (of Physical and Mathematical), Professor of the Moscow University of the Ministry of Internal Affairs of Russia named after V. Y. Kikiot, Moscow. E-mail: atlasov.igor.777@gmail.com

3. Begishev I. R., Zharova A. K., Gromova E. A., Zaloilo M. V., Filipova I. A., Shutova A. A. «Cifrovoj povorot» v pravovyh issledovaniyah // Journal of Digital Technologies and Law. 2024. № 2(1). EDN: IWWUBP.
4. Zharova, A. K. Sistema organizacionno-pravovogo vyyavleniya lic, razmestivshih informaciyu v internete o namerenii sovershit' prestuplenie // Probely v rossijskom zakonodatel'stve. – 2024. – T. 17, № 1. – S. 122–130. – DOI 10.33693/2072-3164-2024-17-1-122-130. – EDN OHAZYD.
5. Shutova, A. A. Obespechenie cifrovoj bezopasnosti sistemy zdravoohraneniya ugolovno-pravovymi sredstvami / A. A. Shutova // Russian Journal of Economics and Law. – 2024. – T. 18, № 4 – S. 936–953. – DOI 10.21202/2782-2923.2024.4.936-953. – EDN SHZTFY.
6. Dejneko, A. G. Publichnoe pravo v kiberprostranstve: publichno-pravovoe regulirovanie informacionnyh otnoshenij / A. G. Dejneko. – Moskva : Obshchestvo s ogranichennoj otvetstvennost'yu \«Prospekt»\, 2025. – 248 s. – ISBN 978-5-392-42996-7. – EDN SBSONL.
7. Zharova, A. K. Paradigma cifrovogo profilirovaniya deyatel'nosti cheloveka: riski, ugrozy, prestupleniya / A. K. Zharova, V. M. Elin, A. V. Minbaleev. – Moskva : Obshchestvo s ogranichennoj otvetstvennost'yu \"Rusajns\", 2022. – 240 s. – ISBN 978-5-466-00766-4. – EDN DNKVPR.
8. Zharova, A. K. Obzor normativnyh trebovanij, obespechivayushchih nacional'nuyu bezopasnost' SShA v sfere kvantovyh tekhnologij / A. K. Zharova // Informacionnoe obshchestvo. – 2023. – № 3. – S. 69–77. – DOI 10.52605/16059921_2023_03_69. – EDN CCHNJY.
9. Zaloilo, M. V. Ciklichno-volnovaya model' interpretacii istorii prava na osnove teorii tekhnologicheskih ukladov / M. V. Zaloilo // Istoriko-pravovoj ezhegodnik – 2023. – Moskva : Infotropic Media, 2024. – S. 48–72. – EDN IETWNM.
10. Tverdova, T. V. § 3. Riski pravovogo regulirovaniya otnoshenij, voznikayushchih po povodu iskusstvennogo intellekta / T. V. Tverdova // Teoretiko-pravovaya paradigma sushchestvovaniya kiberneticheskoj (informacionnoj) civilizacii : monografiya. – Moskva : Mezhregional'naya obshchestvennaya organizaciya \«Mezhregional'naya associaciya teoretikov gosudarstva i prava»\, 2022. – S. 244-273. – EDN ZXOJCC.
11. Maksimov S. V. Stohasticheskaya model' repressivnopreventivnogo vozdejstviya na prestupnost': ot intuicii k raschetam / S. V. Maksimov, Yu. G. Vasin, K. A. Utarov. — DOI 10.17150/2500-4255.2021.15(6).665-680 // Vserossijskij kriminologicheskij zhurnal. – 2021. – T. 15, № 6. – S. 665–680.
12. Modelirovanie processov prinyatiya resheniya v pravoohranitel'noj deyatel'nosti / O. Yu. Danilova, A. V. Men'shih, V. V. Men'shih [i dr.]. – Voronezh : Voronezhskij institut Ministerstva vnutrennih del Rossijskoj Federacii, 2021. – 103 s. – ISBN 978-5-88591-856-5. – EDN FENSWM.
13. Minaev V. A. Modelirovanie dinamiki prestupnosti s uchetom faktora latentnosti// Kriminologicheskij zhurnal. Estestvennye nauki. Konp'yuternye nauki i informatika. 2022. № 2. S. 67–78.
14. Malahova, V. V. Analiz statisticheskih dannyh s ispol'zovaniem matematicheskogo apparata iskusstvennogo intellekta / V. V. Malahova, O. V. Malahov // Vestnik Luganskogo gosudarstvennogo universiteta imeni Vladimira Dalya. – 2023. – № 11. – S. 177–179. – EDN EADYTE.