САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
(СПбГУ)

УДК 004.8
ГРНТИ 27.47.23, 28.23.29, 04.15.41, 20.01.07
№ госрегистрации АААА-А19-119092090068-5

УТВЕРЖДАЮ
И. о. начальника
Управления научных исследований

_____ Е.В. Лебедева

___.09.2020

ОТЧЕТ

О НАУЧНО-ИССЛЕДОВАТЕЛЬСКОЙ РАБОТЕ

Машинное обучение и структурные особенности байесовской сети доверия со скрытыми переменными как модели социально-значимого поведения

по теме:

БАЙЕСОВСКИЕ СЕТИ ДОВЕРИЯ КАК МОДЕЛИ СОЦИАЛЬНО-ЗНАЧИМОГО ПОВЕДЕНИЯ, ВЫЧИСЛИТЕЛЬНЫЕ ЭКСПЕРИМЕНТЫ

Грант РФФИ № 19-37-90120

(промежуточный)

Руководитель темы,
д.ф.-м.н., проф.

_____ А.Л. Тулупьев

01.09.2020

Санкт-Петербург 2020 г.

СПИСОК ИСПОЛНИТЕЛЕЙ

Мл. науч. сотр. _____ 1.09.2020 Торопова Александра Витальевна (отчет)

РЕФЕРАТ

Отчет 24 с., 28 источников, 4 прил.

БАЙЕСОВСКИЕ СЕТИ ДОВЕРИЯ, НЕПОЛНАЯ ИНФОРМАЦИЯ, СОЦИАЛЬНОЕ ПОВЕДЕНИЕ, АНАЛИЗ СОЦИАЛЬНОГО ПОВЕДЕНИЯ, ИНТЕНСИВНОСТЬ ПОВЕДЕНИЯ

Объектом исследования являются алгебраические байесовские сети доверия как модели социально-значимого поведения в контексте получения оценок интенсивности поведения.

Проект поддержан грантом РФФИ № 19-37-90120, отчет представлен по первому году проекта.

Основной **целью** проекта является является создание правдоподобной модели поведения, позволяющей косвенно оценить параметры поведения респондента на основе неточной, неполной, нечисловой информации об отдельных эпизодах его поведения. Эти исследования тесно связаны с направлением развития алгоритмического обеспечения систем поддержки и принятия решений, используемых для выявления характеристик процессов в социуме при невозможности организовать классические формы длительного наблюдения и многофакторного измерения параметров процесса. При этом имеются сведения, полученные от экспертов, предположения о классах и семействах таких процессов, а также ограниченное число измеряемых особенностей такого процесса.

Методология проекта основывается на предложенном и апробированном в рамках предшествующих проектов использовании сочетания использования байесовских сетей доверия с различными методами (синтезирование, опросы, сбор данных) получения данных о последних эпизодах поведения.

Работа в исследовании велась по трем основным направлениям: 1) исследование моделей социально-значимого поведения на основе байесовских сетей доверия; 2) сбор данных об интенсивности поведения; 3) реализация программного комплекса, позволяющего работу с моделями.

Предложена модель социально-значимого поведения со «следующим эпизодом». Разработано программное обеспечение для сбора данных о постинге из социальных сетей Вконтакте и Instagram. Собраны данные об интенсивности поведения для работы с моделями. Проведены апробация модели и сравнение с другими моделями социально-значимого поведения.Реализована часть программного комплекса для работы с моделями. Состоялось участие в двух международных конференциях. Подготовлены 2 статьм для публикации в рецензируемых журналах. Опубликовано 2 работы (из них 2 РИНЦ).

СОДЕРЖАНИЕ

ВВЕДЕНИЕ

Настоящий отчет о НИР (промежуточный) содержит сведения о результатах первого года работы над проектом № 19-37-90120, поддержанным грантом РФФИ.

Актуальность. Задачи изучения процессов в социуме и моделирования поведения как индивидуального, так и на уровне популяции, часто встречаются во многих областях науки: в маркетинговых и экономических исследованиях анализируется поведение покупателей и пользователей каких-либо сервисов [1–4]; в социологии — поведение человека при взаимодействии с другими, а также в социальных сетях [5–7]; в эпидемиологии для оценки риска передачи вирусов [8, 9]; в информационной безопасности (социоинженерные атаки, т.е. атакующие действия через пользователя) [10–12] и др.

Интенсивность поведения является одним из основных параметров поведенческих процессов, однако точные данные об интенсивности во многих случаях невозможно из-за временных, финансовых или правовых ограничений. Таким образом актуальна задача оценки этого параметра с помощью использования косвенных методов.

На основе оценки интенсивности поведения появляется возможность делать выводы о значимых аспектах поведения как в настоящем, так и в будущем. Например, в компьютерной безопасности, используя данные об интенсивности взаимодействия между пользователями можно оценить вероятность распространения социоинженерных атак. В эпидемиологии по оценки интенсивности контактов между людьми можно сделать выводы о дальнейшем распространении болезни.

Одной из особенностей исследования процессов в социуме является то, что в случае опроса респондентов об особенностях их поведения данные поступают на естественном языке, т.е. являются в значительной степени нечеткими и неполными. В изучаемом нами поведении можно выделить четко определенные проявления – эпизоды поведения. В качестве примера можно привести рискованное поведение, например, поведение, связанное с риском передачи и получения такого неизлечимого инфекционного заболевания как

ВИЧ. Для такого поведения есть четко определенные эпизоды, во время которых может произойти передача заболевания: незащищенный половой акт, инъекции при потреблении наркотиков и т.д. Поведение рассматривается как последовательность эпизодов, и анализируются данные об эпизодах. В такой постановке исследуемая задача близка к задачам, возникающим при анализе временных рядов [1]. Однако, несмотря на удобство использования методологии моделирования и анализа временных рядов (особенно при работе с качественными данными в нечетких временных рядах), применение этих подходов к решению задачи, связанной с моделированием социально-значимого поведения, сталкивается с рядом ограничений. Например, одно из ограничений описанного подхода связано с рассмотрением рядов с равноотстоящими элементами, в то время как интервалы между эпизодами социально-значимого поведения не являются одинаковыми, как правило они представляют собой случайные величины.

Во многих случаях при изучении поведения кроме данных об эпизодах поведения становятся известными также дополнительные сведения — психологические, демографические, социальные характеристики, позволяющие лучше описать поведение. Кроме того, могут быть обоснованные предположения о характере поведения, о связях между параметрами. В частности, моделью рискованного сексуального поведения во многих исследованиях является пуассоновский процесс [14, 15]. Включение таких теоретических предположений позволяет получать более точный прогноз.

Ограниченное число и неточность, нечеткость естественно-языковых формулировок ответов, а также необходимость учета экспертных знаний о предметной области не позволяют напрямую использовать известные для оценки параметров поведения, поэтому возникает необходимость в предложении новых математических моделей.

Суворовой А.В. уже была предложены модели социально-значимого поведения, основанная на данных о последних эпизодах поведения респондентов [16], однако стоит вопрос о том, насколько этим данным можно доверять, ведь в зависимости от социального одобрения или неодобрения того или иного вида поведения респонденты могут дать не совсем верную информацию, а иногда и заведомо ложную. Кроме того, в этих моделях завершение исследуемого периода рассматривается также, как и эпизоды поведения.

Таким образом, требуется разработка подходов для решения этих проблем. Для решения первой проблемы В [17] предлагается расширить исходную модель скрытыми переменными, отвечающими за «реальные» последние эпизоды поведения респондентов, в ходе предварительных исследований было показано, что новая модель дает более точные результаты. Тем не менее требуется продолжить исследования, в частности испытать ее на реальных данных, а также предложить методы синтеза такой модели как в части структуры, так и параметров.

Основной целью проекта является является создание правдоподобной модели поведения, позволяющей косвенно оценить параметры поведения респондента на основе неточной, неполной, нечисловой информации об эпизодах поведения.

Исследование направлено на развитие алгоритмического обеспечения систем поддержки и принятия решений в условиях неполной, неточной, нечеткой и нечисловой информации, в частности, на разработку моделей, методов и алгоритмов для выявления характеристик процессов в социуме при невозможности организовать классические формы длительного наблюдения и многофакторного измерения параметров процесса, но имеются сведения, полученные от экспертов, предположения о классах и семействах таких процессов, а также ограниченное число измеряемых особенностей такого процесса.

Отдельной целью является формирование исследовательский инструментария для наук социогуманитарного цикла, который позволит с приемлемыми затратами и точностью производить оценку различных параметров изучаемого поведения на основе сведений об ограниченном числе его эпизодов.

В рамках выдвинутых целей проекта на текущий год (первый год проекта, 2020) были поставлены следующие задачи.

— Изучение и анализ моделей социально-значимого поведения на основе байесовских сетей доверия.
— Сбор данных об интенсивности поведения для работы с моделями социально-значимого поведения.
— Реализация части программного комплекса для работы с моделями социально-значимого поведения.

Отчет составлен по результатам выполнения НИР, поддержанной грантом РФФИ № 19-37-90120 (Приложение А).

1 Байесовские сети доверия для моделирования социально-значимого поведения

Предложена новая модель оценки социально-значимого поведения, корректно учитывающая величину интервала между моментами последнего эпизода и завершения исследуемого периода. Проведена ее апробация и сравнение с уже существующими моделями. Показано, что предложенная модель показывает хорошие результаты качества предсказания и лучшие результаты в сравнении с другими моделями, (Приложение В, Г).

2 Данные об интенсивности поведения

Разработано программное обеспечение для сбора данных о постинге из социальных сетей Вконтакте и Instagram. С помощью этих программ были собраны 2 датасета с информацией о постинге за январь 2020 года (6556 записей о пользователях) и март 2020 года (5338 записей о пользователях) в социальной сети Вконтакте (Приложение В, Г) и 1 датасет с информацией о постинге за июнь 2020 года (2800 записей о пользователях) в сети Instagram. На языке R написана программа, автоматически синтезирующая данные, которые могут быть использованы для работы с моделями социально-значимого поведения.

3 Реализация программного комплекса для работы с моделями социально-значимого поведения

Реализована программная часть для работы с моделью социально-значимого поведения со «следующим» эпизодом: реализовано обучение и работа модели в предсказании интенсивности поведения на автоматически синтезированных данных и данных, введенных пользователем в форматах csv и электронных таблиц Excel, при дискретизации непрерывных величин, определенной пользователем.

ЗАКЛЮЧЕНИЕ

Полученные в данном отчете результаты расширяют, развивают и открывают новые направления исследований в изучении моделей социально-значимого поведения и готовят базу для использования этих моделей в различных исследованиях психологического, социологического, эпидемиологического характера в тех случаях, когда необходимо получить сведения об интенсивности поведения исследуемых на основе их интервью или заполнения опросников, то есть в условиях дефицита данных.

Предложена модель социвльно-значимого поведения со «следующим» эпизодом, корректно учитывающая величину интервала между моментами последнего эпизода и завершения исследуемого периода. Проведена ее апробация и сравнение с другими моделями социально-значимого поведения.

Реализованы программы для сбора данных о постинге из социальных сетей Вконтакте и Instagram, с помощью них собраны датасеты для работы с изучаемыми моделями.

Отраженные в отчете теоретические результаты были представлены на международных конференциях разного уровня (2 публикации с индексацией РИНЦ) (Приложение Б). Кроме того, теоретические и алгоритмические разработки были частично реализованы в ряде программ — компонент комплекса программ, с помощью которого производятся вычислительные эксперименты. Также были подготовлены две статьи для публикации в рецензируемых журналах, одна из них отправлена в редакцию журнала «Информационно-управляющие системы».

Таким образом, задачи первого года исследования выполнены, а цель первого года выполнения проекта — достигнута.

ПРИЛОЖЕНИЕ А
Перечень грантов, заказанных НИР, контрактов, хоздоговоров, которыми поддерживались исследования по данной НИР

1. Грант РФФИ, проект № 19-37-90120 — «Машинное обучение и структурные особенности байесовской сети доверия со скрытыми переменными как модели социально-значимого поведения», руководитель А.Л. Тулупьев.

ПРИЛОЖЕНИЕ Б
Список публикаций в рамках проекта

1. *Торопова А.В., Тулупьева Т.В.* Модели оценки интенсивности поведения на примере постинга в социальной сети // VIII Международная научно-практическая конференция «Нечеткие системы, мягкие вычисления и интеллектуальные технологии» НСМВИТ–2020. Труды конференции в 2-х томах. Смоленск. 29 июня -– 1 июля 2020 г. Смоленск: Универсум. Том 2. С. 164–172. (РИНЦ).
2. *Торопова А.В., Тулупьева Т.В.* Байесовская сеть доверия как модель оценки интенсивности поведения на примере постинга в социальной сети // XXIII Международная конференция по мягким вычислениям и измерениям (SCM–2020). Сборник докладов. Санкт-Петербург. 27–29 мая 2020 г. СПб.: СПбГЭТУ «ЛЭТИ». С. 20–22. (РИНЦ).

ПРИЛОЖЕНИЕ В
Сравнение модели со «следующим» эпизодом и модели со «следующим» эпизодом и скрытыми переменными

Приложение подготовлено на основе работы [18].

На рис. В.1 представлена модель в виде байесовской сети доверия [19]. Вершина λ характеризует интенсивность поведения, t_12 — интервал между последним и предпоследним эпизодами поведения, t_23 — интервал между предпоследним и третьим с конца эпизодами поведения за исследуемый период, t_min и t_max — минимальный и максимальный интервалы между эпизодами за исследуемый период, t_next — интервал между последним эпизодом за исследуемый период и первым эпизодом после окончания исследуемого периода, n – количество эпизодов за исследуемый период.

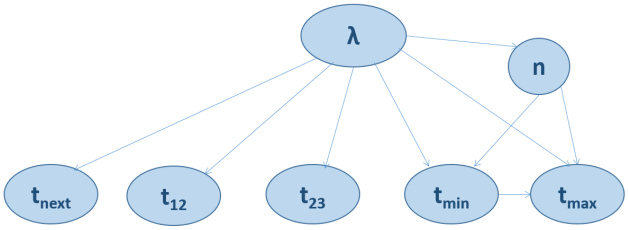

Рис. В.1 – Модель оценки интенсивности поведения

В работе [17] была представлена модель социально-значимого поведения со скрытыми переменными. Данная модель учитывала то, что сведения, поступившие от респондентов, могут быть неверны. Это может быть связано с тем, что в некоторых случаях респонденты, желая получить социальное одобрение, могут специально исказить реальные значения, а также с тем, что, отвечая по памяти, респонденты могут непреднамеренно ошибиться.

На рис. В.2 представлена модель интенсивности поведения со скрытыми переменными. Вершины с ноликом (t^0_12, t^0_23, t^0_min, t^0_max) — это сведения о соответству-

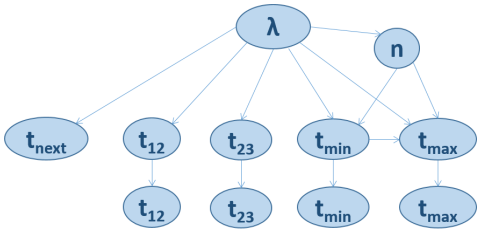

Рис. В.2 – Модель оценки интенсивности поведения со скрытыми переменными

ющих интервалах, предоставленные респондентами, остальные вершины описываются также как для модели выше. В данном случае мы не включаем вершину t^0_next, так как респондент не может дать сведений об этом эпизоде, если он еще не произошел. Таким образом t_next, t_12, t_23, t_min, t_max и n являются скрытыми переменными, характеризующими реальные данные об интенсивности поведения.

Для апробации этих моделей были использованы данные, собранные в социальной сети Вконтакте, а также синтезированные на их основе данные в качестве «неточных» ответов респондентов.

Для сбора данных из Вконтакте [**20**] была написана программа на языке C#. Также для этого был использован метод wall.get, предоставляемый API Вконтакте [**21**]. Он дает возможность получить информацию о последних 100 записях пользователя. Этого достаточно, если в качестве исследуемого периода рассмотреть один месяц. Кроме того, у этого метода есть ограничение на 5000 запросов в день.

Обрабатывались аккаунты пользователей, предоставивших надлежащее разрешение. Программа извлекает время последних трех постов за исследуемый период, время первого поста, сделанного по истечении исследуемого периода, минимальный и максимальный интервалы между публикацией постов за исследуемый период, а также количество постов за исследуемый период. Эта информация временно сохраняется в файле Excel для дальнейшей проверки моделей.

В качестве исследуемого периода был взят декабрь 2019-го года. Выборка пользователей проводилась случайным образом. Данные о пользователях с закрытыми профилями и о тех пользователях, у которых не оказалось достаточного числа постов, не учитывались. Таким образом был собран датасет, содержащий 6556 записей.

Данные об ответах респондентов были синтезированы автоматически с помощью добавления шума следующим образом: рассчитывалось расстояние между последним и предпоследним эпизодами в днях и добавлялась случайная величина таким образом, чтобы это расстояние изменилось не более чем в полтора раза, к расстоянию между предпоследним и третьим с конца эпизоду добавлялась случайная величина так, чтобы расстояние изменилось не более, чем в два раза. К минимальному и максимальному интервалам был добавлен нормальный шум.

Все вычисления в этом и следующем разделах были выполнены на языке R [**22**] с использованием пакета bnlearn [**23**], обеспечивающего работу с байесовскими сетями доверия.

Для работы с байесовской сетью доверия требуется дискретизация всех непрерывных данных. Поэтому значения переменных, связанных со временем (в качестве единицы измерения используем день), то есть t_next, t_12, t_23, t^0_12, t^0_23, t_min, t_max, t^0_min, t^0_max были разбиты на интервалы $t_1 = (0; 0.1)$, $t_2 = [0.1; 0.5)$, $t_3 = [0.5; 1)$, $t_4 = [1; 7)$, $t_5 = [7; 10)$, $t_6 = [10; 20)$, $t_7 = [20; \infty)$; значения переменной λ (интенсивность измеряем как количество постов деленное на количество дней в месяце) — на интервалы $\lambda_1 = (0; 0.1)$, $\lambda_2 = [0.1; 0.2)$, $\lambda_3 = [0.2; 0,3)$, $\lambda_4 = [0.3; 0,5)$, $\lambda_5 = [0.5; 1)$, $\lambda_6 = [1; \infty)$.

4556 записей было использовано для машинного обучения параметров моделей. То есть для всех пар вершин сети, соединенных дугой, были построены таблицы условных вероятностей.

Для тестирования модели было использовано 2000 записей. В качестве вводных данных в модели передавались синтезированные ответы респондентов.

После получения оценок интенсивности, предсказанных моделями, можно их сравнить с известными интенсивностями публикации постов пользователями. Таблица B.1 представляет собой матрицу неточностей модели оценки интенсивности поведения, а таблица B.2 —- матрицу неточностей модели оценки интенсивности поведения со скрытыми переменными. Строки представляют собой реальные интенсивности, а столбцы – интенсивности, предсказанные моделью.

В данном случае задача представляет собой задачу классификации по шести классам, поэтому стоит рассмотреть такую характеристику как средняя точность (0,799 и 0,797).

Таблица В.1 – Матрица неточностей модели оценки интенсивности поведения

	λ_1	λ_2	λ_3	λ_4	λ_5	λ_6
λ_1	120	121	18	19	26	0
λ_2	68	335	58	84	67	4
λ_3	9	125	50	111	67	6
λ_4	4	63	51	101	113	15
λ_5	0	16	10	50	164	24
λ_6	1	0	0	4	72	23

Таблица В.2 – Матрица неточностей модели оценки интенсивности поведения со скрытыми переменными

	λ_1	λ_2	λ_3	λ_4	λ_5	λ_6
λ_1	110	117	20	21	24	7
λ_2	71	319	73	67	73	8
λ_3	16	112	67	86	79	7
λ_4	1	60	65	110	97	14
λ_5	3	19	14	50	149	29
λ_6	0	2	3	10	61	24

Из матриц неточностей видно, что основная часть значений находится на диагонали или в смежных с ней ячейках, это значит, что даже при ошибке классификации полученные значения скорее всего находятся в соседних классах.

В таблице В.3 сравниваются точность (accuracy), средняя точность, точность (precision) и полнота (recall), основные метрики качества. Как видно, разница в резуль-

Таблица В.3 – Матрица неточностей модели оценки интенсивности поведения со скрытыми переменными

	Точность (accuracy)	Ср. Точность	Точность (precision)	Полнота
Модель оценки интенсивности поведения	0.397	0.799	0.397	0.369
Модель оценки интенсивности поведения со скрытыми переменными	0.391	0.797	0.392	0.366

татах довольно незначительная, однако модель со скрытыми переменными показала результаты немного хуже, возможно, это связано с усложнением модели (было добавлено 4 новые вершины и связи с ними).

ПРИЛОЖЕНИЕ Г
Апробация модели социально-значимого поведения со «следующим» эпизодом на данных из Вконтакте

Приложение подготовлено на основе работы [24].

В [25, 26] была представлена модель социально-значимого поведения, рассчитывающая оценку интенсивности поведения на основании данных о последних трех эпизодах поведения. В [27] также были предложены такие модели с обученной структурой. Основное отличие предлагаемой модели состоит в том, что вместо интервала между последним эпизодом поведения и временем интервью, рассматривается интервал между последним эпизодом поведения и следующим. Например, наш период исследования — это 2019-ый год, последний эпизод исследуемого поведения произошел 30-го декабря, а первый эпизод по окончании исследуемого периода — 5-е января, тогда мы берем интервал от 30-го декабря до 5 января. Дело в том, что интервью не является эпизодом исследуемого поведения, и данные о моменте интервью не содержат информации о поведении человека и могут исказить оценку интенсивности поведения.

На рис. Г.1 представлена модель оценки интенсивности поведения в виде байесовской сети доверия [19]. Вершина λ характеризует интенсивность поведения, t_12 — интервал между последним и предпоследним эпизодами поведения, t_23 -– интервал между предпоследними предпредпоследним эпизодами поведения за исследуемый период, t_min и t_max — минимальный и максимальный интервалы между эпизодами за исследуемый период, n -– количество эпизодов за исследуемый период, а t_next — интервал между последним эпизодом из исследуемого периода и первым эпизодом по окончании исследуемого периода.

Несмотря на то, что основная задача предложенной модели заключается в оценке интенсивности поведения в условиях недостаточности данных, проведем апробацию мо-

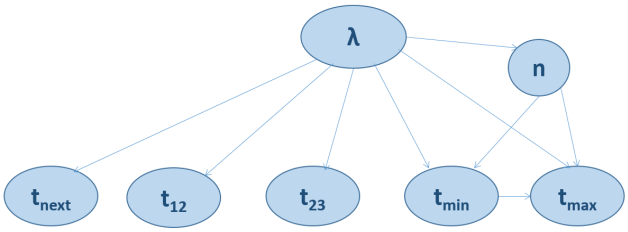

Рис. Г.1 – Модель оценки интенсивности поведения

дели. То есть нам нужны такие данные при которых может быть получена максимально точная реальная оценка интенсивности поведения. Для этого подходят данные о постинге в социальной сети. Мы взяли данные из Вконтакте [**20**], самой популярной социальной сети в России [**28**]. Каждый пользователь этой сети имеет так называемую «стену», на которой он может публиковать свои посты, делать репосты записей других пользователей, а также на которой могут оставлять записи другие пользователи. У пользователя может быть закрытый тип профиля, в таком случае его посты могут видеть только «друзья». Такие аккаунты не были включены в анализ.

Поскольку подобных готовых датасетов не удалось найти в отрытом доступе, была разработана специальная программа для его сбора. API Вконтакте предоставляет метод wall.get [**21**], с помощью которого можно получить последние 100 записей пользователя. Этого достаточно, если рассматривать в качестве исследуемого периода один месяц. Также использование этого метода ограничено 5000 запросами в сутки.

Программа для сбора мета-информации о постах из Вконтакте была написана на языке C#. Обрабатывались аккаунты пользователей, которые дали соответствующие разрешение. Программа извлекает время последних трех постов за исследуемый период, время первого поста, сделанного по окончании исследуемого периода, минимальный и максимальный интервалы между временем публикации постов за исследуемый период и количество постов за исследуемый период. Полученные данные временно сохраняются в файле Excel для проверки математической модели.

В качестве исследуемого периода был взят декабрь 2019-го года. Пользователи выбирались случайным образом. Данные о пользователях с закрытыми профилями и о тех пользователях, у которых не оказалось достаточного числа постов не учитывались. Таким образом было собрано 6556 записей о пользователях.

Для использования байесовской сети доверия непрерывные данные нужно дискретизировать. Была использована следующая дискретизация: значения всех переменных t (в качестве единицы измерения используем день) были разбиты на интервалы $t_1 = (0; 0.1)$, $t_2 = [0.1; 0.5)$, $t_3 = [0.5; 1)$, $t_4 = [1; 7)$, $t_5 = [7; 10)$, $t_6 = [10; 20)$, $t_7 = [20; \infty)$; для переменной λ (интенсивность измеряем как количество постов деленное на количество дней в месяце) — на интервалы $\lambda_1 = (0; 0.1)$, $\lambda_2 = [0.1; 0.2)$, $\lambda_3 = [0.2; 0,3)$, $\lambda_4 = [0.3; 0,5)$, $\lambda_5 = [0.5; 1)$, $\lambda_6 = [1; 2)$, $\lambda_7 = [2;)$.

Все вычисления в этом разделе были выполнены на языке R [**22**] с использованием пакета bnlearn [**23**] для работы с байесовскими сетями доверия.

4556 записей было использовано для машинного обучения параметров модели, то есть для каждой пары вершин сети, связанных дугой, были построены таблицы условных вероятностей.

2000 записей было использовано для тестирования модели. В качестве вводных данных использовались значения t_12, t_23, t_min и t_max.

После получения оценок интенсивности, предсказанных моделью, можно их сравнить с известными интенсивностями публикации постов пользователями. Таблица Г.1 представляет собой матрицу неточностей, где строки представляют собой реальные интенсивности, а столбцы -– интенсивности, предсказанные моделью.

Таблица Г.1 – Матрица неточностей модели оценки интенсивности поведения

	λ_1	λ_2	λ_3	λ_4	λ_5	λ_6	
λ_1	116	127	18	9	34	1	0
λ_2	65	342	52	83	72	2	0
λ_3	7	111	78	102	65	5	0
λ_4	2	51	53	112	123	6	0
λ_5	0	10	8	40	191	15	0
λ_6	0	1	0	3	70	17	0
λ_7	0	0	0	1	7	1	0

Точность (accuracy) равна 0,428, но в данном случае задачей является классификация по семи классам, поэтому имеет смысл оценить среднюю точность (average accuracy), она равна 0,837. В таблице Г.2 представлены точность (precision), полнота (recall) и F-1, основные метрики качества по классам.

Таблица Г.2 – Матрица неточностей модели оценки интенсивности поведения

	Точность	Полнота	F-1
λ_1	0.611	0.38	0.469
λ_2	0.533	0.555	0.544
λ_3	0.373	0.212	0.27
λ_4	0.32	0.323	0.321
λ_5	0.34	0.723	0.46
λ_6	0.362	0.187	0.246
λ_7	NaN	0	NaN

Список используемой литературы

1. *Lo F.-Y., Yu T.H.-K., Chen H.-H.* Purchasing intention and behavior in the sharing economy: Mediating effects of APP assessments // Journal of Business Research, 121, pp. 93-102. 2020. DOI: 10.1016/j.jbusres.2020.08.017.
2. *Lee H.W., Kim M.Y.* Structural modeling of dissatisfaction, complaint behavior, and revisiting intentions in hairdressing services // Fashion and Textiles, 7 (1). 2020. DOI: 10.1186/s40691-019-0191-3.
3. *Motoki K., Suzuki S., Kawashima R., Sugiura M.* A Combination of Self-Reported Data and Social-Related Neural Measures Forecasts Viral Marketing Success on Social Media // Journal of Interactive Marketing, 52, pp. 99-117. 2020. DOI: 10.1016/j.intmar.2020.06.003.
4. *Vithayathil J., Dadgar M., Osiri J.K.* Social media use and consumer shopping preferences // International Journal of Information Management, 54, № 102117. 2020. DOI: 10.1016/j.ijinfomgt.2020.102117.
5. *Khoshhal K., Nunes U., Dias J.* Probabilistic Social Behavior Analysis by Exploring Body Motion-Based Patterns // IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, V. 38, N. 8, 2016.
6. *Oliveira M., Pinheiro D., Macedo M., Bastos-Filho C., Menezes R.* Uncovering the social interaction network in swarm intelligence algorithms // Applied Network Science, 5 (1), № 24. 2020. DOI: 10.1007/s41109-020-00260-8.
7. *Davidson I., Gourru A., Velcin J., Wu Y.* Behavioral differences: insights, explanations and comparisons of French and US Twitter usage during elections // Social Network Analysis and Mining, 10 (1), № 6, 2020. DOI: 10.1007/s13278-019-0611-9.
8. *Козлов А.А., Станько Э.П., Игумнов С.А.* Рискованные формы поведения и уровень социального функционирования ВИЧ-позитивных потребителей инъекционных наркотиков: прогностические модели // Медицинская психология в России. Т. 10. 1 (48). 2018.

9. *Nowak B., Brzoska P., Piotrowski J., Sedikides C., Żemojtel-Piotrowska M., Jonason P.K.* Adaptive and maladaptive behavior during the COVID-19 pandemic: The roles of Dark Triad traits, collective narcissism, and health beliefs // Personality and Individual Differences, 167, с№ 110232. 2020. DOI: 10.1016/j.paid.2020.110232.
10. *Абрамов М. В., Тулупьева Т. В., Тулупьев А. Л.* Социоинженерные атаки: социальные сети и оценки защищенности пользователей. СПб.: ГУАП, 2018, 266 с. ISBN 978-5-8088-1377-5.
11. *Khlobystova A.O., Abramov M.V., Tulupyev A.L.* Soft Estimates for Social Engineering Attack Propagation Probabilities Depending on Interaction Rates Among Instagram Users // International Symposium on Intelligent and Distributed Computing. Springer, Cham. 2019.
12. *Khlobystova A.O. , Abramov M.V., Tulupyev A.L., Zolotin A.A.* Search for the shortest trajectory of a social engineering attack between a pair of users in a graph with transition probabilities // Information and Control Systems. 2018. no. 6.
13. *Ярушкина Н. Г., Афанасьева Т. В., Перфильева И. Г.* Интеллектуальный анализ временных рядов: Учебное пособие. Ульяновск: УлГТУ, 2010. 320 с.
14. *Brommer J.E., Alho J.S., Biard C., Chapman J.R., Charmantier A., Dreiss A., Hartley I.R., Hjernquist M.B., Kempenaers B., Komdeur J., Laaksonen T., Lehtonen P.K., Lubjuhn T., Patrick S.C., Rosivall B., Tinbergen J.M., Van Der Velde M., Van Oers K., Wilk T., Winkel W.* Passerine extrapair mating dynamics: A Bayesian modeling approach comparing four species // American Naturalist, 176 (2), pp. 178-187. 2010. DOI: 10.1086/653660
15. *Cruyff M., Bockenholt U., Hout A. van den , Heijden P. van der* Accounting for Self-Protective Responses in Randomized Response Data from a Social Security Survey Using the Zero-Inflated Poisson Model // The Annals of Applied Statistics, 2008, Vol. 2, No. 1. P. 316-–331.
16. *Суворова А. В., Тулупьев А. Л.* Синтез структур байесовской сети доверия для оценки характеристик рискованного поведения. Информационно-управляющие системы, 2018, № 1, с. 116-–122. doi:10.15217/issn1684-8853.2018.1.116.

17. *Toropova A.V., Tulupyeva T.V.* Synthesis and learning of socially significant behavior model with hidden variables // Advances in Intelligent Systems and Computing. 2019. T. 875. С. 76–84.

18. *Торопова А.В., Тулупьева Т.В.* Модели оценки интенсивности поведения на примере постинга в социальной сети // VIII Международная научно-практическая конференция «Нечеткие системы, мягкие вычисления и интеллектуальные технологии» НСМВИТ–2020. Труды конференции в 2-х томах. Смоленск. 29 июня-–1 июля 2020 г. Смоленск: Универсум. Том 2. С. 164–172.

19. *Тулупьев А.Л., Николенко С.И., Сироткин А.В.* Основы теории байесовских сетей: учебник. СПб.:Изд-во С.Петерб. ун-та, 2019. 399 с.

20. Вконтакте. URL: http://www.vk.com (дата обращения: 25.08.20).

21. Вконтакте. Описание методов API. URL: https://vk.com/dev/methods (дата обращения: 25.08.20).

22. R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. URL: http://www.R-project.org (дата обращения: 25.08.20).

23. *Scutari M.* Learning Bayesian Networks with the Bnlearn R Package. arXiv preprint. arXiv:0908.3817. 2009.

24. *Торопова А.В., Тулупьева Т.В.* Байесовская сеть доверия как модель оценки интенсивности поведения на примере постинга в социальной сети // XXIII Международная конференция по мягким вычислениям и измерениям (SCM–2020). Сборник докладов. Санкт-Петербург. 27–29 мая 2020 г. СПб.: СПбГЭТУ «ЛЭТИ». С. 20–22.

25. *Суворова А.В.* Моделирование социально-значимого поведения по сверхмалой неполной совокупности наблюдений // Информационно-измерительные и управляющие системы. 2013. №9, т. 11. С. 34—38.

26. *Суворова А.В., Тулупьев А.Л., Сироткин А.В.* Байесовские сети доверия в задачах оценивания интенсивности рискованного поведения // Нечеткие системы и мягкие вычисления. 2014. Т. 9, № 2. С. 115—129.

27. *Suvorova A.V.* Models for respondents' behavior rate estimate: bayesian network structure synthesis // Proceedings of 2017 XX IEEE International Conference on Soft Computing And Measurements (SCM). 2017. pp. 87–89.

28. SimilarWeb. URL: https://www.similarweb.com/fr/top-websites/russian-federation/ (дата обращения: 25.08.20).