



3. Солдатова О.П., Лёзин И.А. Исследование качества решения задачи классификации нейронными нечёткими продукционными сетями на основе модели вывода Мамдани–Заде // Вестник Самарского государственного технического университета. Серия «Технические науки», № 2(35), 2014. С. 136-148.

4. Understanding binary cross-entropy / log loss: a visual explanation [Электронный ресурс] // Towards Data Science: [сайт]. - URL: <https://towardsdatascience.com/understanding-binary-cross-entropy-log-loss-a-visual-explanation-a3ac6025181a> (дата обращения: 02.04.2020).

Е.В. Орлова

СРАВНИТЕЛЬНЫЙ АНАЛИЗ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ ПРИ РЕШЕНИИ ЗАДАЧИ КЛАСТЕРИЗАЦИИ В КРЕДИТНОМ СКОРИНГЕ

(Уфимский государственный авиационный технический университет)

Аннотация. Осуществляется сравнительный анализ методов машинного обучения при формировании однородных кластеров кредиторов в задаче кредитного скоринга. Показана большая устойчивость метода k -средних по сравнению с методов кластеризации, основанным на самоорганизующихся картах Кохонена.

Ключевые слова: машинное обучение, метод k -means, метод ANN.

Для проведения мониторинга платежеспособности клиентов кредитные организации используют скоринговые модели, с помощью которых на основе анализа предыдущих кредитных историй составляется рейтинг заемщиков и определяется вероятность возврата кредита в срок потенциальным заемщиком. Одной из основных проблем при этом и как как составная часть модели кредитного скоринга кредитная организация должна разделить своих клиентов на однородные группы, имеющих одинаковые кредитные портреты, с последующим определением рисковой составляющей кредиторов в каждой однородной группе, а также дальнейшего выявления и моделирования влияния существенный факторов на возможные риски неплатежей у клиентов каждой группы [1-4]. Исследование и сравнительный анализ методов кластеризации, пригодных для решения поставленной задачи показал следующие результаты.

Модельные эксперименты проводятся с использованием статистической информации о кредитных историях клиентов кредитной организации с использование программы статистической обработки данных Statistica 7.0. Особенности изучаемых данных состоят в их разнородности и многомерности. Исследуемая модельная выборка включает данные по 38 клиентам и состоит из следующих показателей, характеризующих заемщиков: срок кредита (*month*), сумма кредита (*value*), пол (0 – женщина, 1 – мужчина), возраст (*age*), количество детей (*children*), средний доход (*income*), а также каждому заемщику присвоим переменную *problems*, характеризующую наличие или



отсутствие проблем с возвратом кредита (0 – проблема отсутствует, 1 – проблема существует).

Так как в задаче кластеризации количество кластеров заранее неизвестно и выборка клиентов достаточно разнородная, то для получения более обоснованного разбиения на кластеры используем для сравнения другой метод машинного обучения – Automated Neural Networks, ANN и проведем сравнительную оценку эффективности двух методов для получения кластеров клиентов – методы k-средних и Automated Neural Networks метод.

Для получения количества групп (кластеров) будем использовать самоорганизующиеся карты Кохонена (Self-organizing map) – нейронную сеть с обучением без учителя. Разделим исходную совокупность на три подвыборки – первая – train sample в объеме 70 % от совокупного объема выборки, которая используется для обучения нейронной сети и подстройки ее весов. Вторая подвыборка – тестовая, ее объем составляет 15 %, используется для проверки правильности обучения и переобучения. Третья подвыборка – валидационная, служит для оценки точности нейронной сети на «новых данных». Сначала топологическая размерность сети задается 5*5 нейронов. В результате моделирования показано, что можно четко выделить четыре класса клиентов, так как именно четыре нейрона описали большую часть исходных данных.

Результаты кластеризации, полученные на основе метода Automated Neural Networks, представлены в таблице 1.

Таблица 1. Описательные статистики сформированных кластеров (Automated Neural Networks)

Клас-тер	Количес-тво клиентов	Номер клиента	Описательные статистики переменной month		Описательные статистики переменной value	
			Среднее значение	Среднеквадратическое отклонение	Среднее значение	Среднеквадратическое отклонение
1	7	22, 23, 26-29, 38	33,4	15,44	327142,9	76313,9
2	2	20, 21	6	0	175000	35355,3
3	11	17-19, 25, 31-37	40,4	20,9	69545,5	25734,7
4	18	1-16, 24, 30	14,6	12,26	22222,2	8292,8

Статистические характеристики – среднее значение, стандартное отклонение для каждого класса по каждому из исследуемых признаков показаны, а также количество объектов в каждом классе, полученные по методу k-средних, показаны в таблице 2.



Таблица 2. Описательные статистики сформированных кластеров
(метод k-средних)

Клас-тер	Количес-тво клиентов	Номер клиента	Описательные статистики переменной month		Описательные статистики переменной value	
			Среднее значение	Среднеквадратическое отклонение	Среднее значение	Среднеквадратическое отклонение
1	7	1-16, 24, 30	33,4	15,44	327142,9	76313,9
2	6	17-21, 37	16	15,95	125000	41833,0
3	7	25, 31-36	51,4	13,35	52142,9	11495,3
4	18	22, 23, 26-29, 38	14,6	12,26	22222,2	8292,8

Сравнительный анализ результатов кластеризации (табл. 1 и 2) демонстрирует, что первый и четвертый кластеры имеют одинаковый состав заемщиков и одинаковые значения всех статистических показателей – средние, отклонения, коэффициенты вариации. Третий кластер по составу заемщиков похож, но вторые кластеры в кластеризации на базе метода k-means объединил шестерых заемщиков, а Automated Neural Networks метод – всего двоих. Кроме этого коэффициент вариации показателя объем кредитования в третьем кластере заемщиков показывает их значительную неоднородность в случае использования Automated Neural Networks метода.

Для более обоснованного выбора одного из методов кластеризации и получения однородных групп заемщиков был проведен дополнительно анализ устойчивости используемых методов. Для этого в выборку данных была введена дополнительная бинарная переменная, отражающая проблемность заемщика. Далее сформированы с учетом этого нового фактора новые разбиения на кластеры. Кластеризация на основе метода k-means показала большую стабильность групп с позиции их одинакового состава до и после введения нового фактора, чем кластеризации на основе Automated Neural Networks.

Поэтому для решения поставленной нами задачи кластеризации клиентов в задаче кредитного скоринга и формирования их однородных групп выбирается метод k-means.

Литература

1. Орлова Е.В. Оценка кредитного риска на основе методов многомерного анализа // Компьютерные исследования и моделирование. 2013. Том 5. № 5. С. 893-901.
2. Orlova E.V. Economic Efficiency of the Mechanism for Credit Risk Management. Proceedings of the Workshop on Computer Modelling in Decision Making (CMDM 2017), CEUR Workshop Proceedings. 2017, 2018, 139–150.
3. Orlova E.V. Mechanism for Credit Risk Management. Proceedings of the 30th International Business Information Management Association Conference



(IBIMA). Vision 2020: Sustainable Economic development, Innovation Management, and Global Growth, Madrid. 2017, 827-837.

4. Orlova E.V. Decision-Making Techniques for Credit Resource Management Using Machine Learning and Optimization // Information. 2020. Vol. 11(3). DOI: 10.3390/info11030144

Н.А. Паренский, Р.А. Парингер

ИССЛЕДОВАНИЕ АЛГОРИТМОВ ДЕТЕКТИРОВАНИЯ ДВИЖЕНИЯ В ВИДЕОПОТОКЕ

(Самарский университет)

Аннотация. Данная статья посвящена исследованию различных алгоритмов обнаружения движения в видеопотоке. Сравнение алгоритмов в трех разрешениях на компактных вычислительных модулях Raspberry Pi 3 Model B и Nvidia Jetson nano по количеству кадров в секунду. Выбор наилучшего алгоритма по отношению качества видео к кадрам в секунду.

Введение

Задача детектирования движущихся объектов является актуальной в современном мире. Её применение можно найти в таких задачах как отслеживание людей в видеопотоке, жестов человека, мимики человека, отслеживание автомобилей в движении, охранных системах, в видеосистемах для госорганов, развлекательно по, робототехнике. Систему видеонаблюдения будет моделировать на компактных вычислительных устройствах RaspberryPi 3 Model B и Nvidia Jetson nano. Задаче является изучить возможности двух данных устройств для построения системы видеонаблюдения с детектирование движущихся объектов с разрешением видео в трех форматах 1920x1080, 1280x720, 640x480 и прийти к выводу какой из заданных разрешений позволит получить плавный видео поток с детектирование движения.

Методы детектирования движения

Метод вычитания фона [1]. Один из трех способов детектирования движения. В основе заложено вычитание фонового кадра, который является первым кадром в видеопотоке, из следующих кадров. Тем самым мы разделяем кадр на передний план, где происходит само движение и задний, то, что является первоначальным кадром. Данный метод очень чувствителен к различным изменениям в видеопотоке. Одним из примеров будет являться освещение, которое будет изменяться с течением времени, что приведет к тому, что все изображение будет как бы в движении. Поэтому данный метод подходит для мест с одинаковым постоянным окружением.

Метод усредненного фона [1]. Данный метод является улучшенной версией метода вычитания фона. Основным действием является вычитание первоначального кадра из последующих, но за исключением, что