**ПРИМЕНЕНИЕ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ В ЗАДАЧАХ
ПРОДУКТИВНОГО ЖИВОТНОВОДСТВА**

Михальский А.И., Новосельцева Ж.А.

*Институт проблем управления им. В.А. Трапезникова РАН,
Москва, Российская Федерация*

Применение точных методов прогноза индивидуальной продуктивной ценности, фертильности, продолжительности продуктивной жизни коров и оптимизации управления стадом являются перспективными направлениями для повышения эффективности современного продуктивного животноводства. Практическое применение систем электронных микродатчиков с беспроводной передачей данных, обеспечивающее получение в реальном времени больших объёмов данных, вызвало потребность в эффективных технологиях анализа и компьютерной обработки получаемой информации. В обзоре рассматривается ряд прикладных задач в области животноводства, в которых нашли успешное применение методы машинной классификации и обучения. В арсенале машинного обучения имеются методы выделения наиболее значимых факторов, установления скрытых зависимостей, использования априорной информации и информации из сторонних источников. В настоящее время на основе методов классификации и восстановления зависимостей разработаны алгоритмы автоматической идентификации отдельных животных, их поведения и состояния на выпасе, предсказания физических и физиологических показателей, прогноза селекционной ценности и др. Основные разделы обзора: индивидуальная идентификация, классификация поведения и обнаружение теплового стресса; выявление поведенческих и физиологических характеристик; прогноз продуктивных показателей; обоснование решений по выбраковке коров; диагностика и оценка эффективности лечения мастита и респираторных заболеваний; оценка эффективности осеменения; селекция по показателям продуктивности и геномным оценкам; селекция по показателям конверсии энергии корма в продукцию; селекция по показателям фертильности и дожития. Широкий спектр и экономическая эффективность решаемых задач свидетельствуют о высоком потенциале методов анализа данных и машинного обучения при решении практических задач животноводства.

Ключевые слова: продуктивное животноводство, анализ данных, многофакторные биосистемы, классификация, диагностика, машинное обучение

Проблемы биологии продуктивных животных, 2018, 4: 98-109

Введение

Применение вычислительных методов для прогноза продолжительности продуктивной жизни животных, оценки индивидуальной продуктивной ценности и оптимизации управления стадом являются перспективными направлениями в повышении эффективности продуктивного животноводства (Кузнецов, 1996; Михайленко, 2015; Черепанов и др., 2017). При решении этих задач немаловажным является учёт многофакторности, отсутствие математических моделей, в явном виде описывающих изучаемые явления, а также неполнота доступных данных. Для поиска решения таких задач могут быть использованы средства математической статистики, численных методов, методов оптимизации, теории вероятностей, теории графов, различные техники работы с данными в цифровой форме. Ситуация, когда отсутствует модель явления, но регистрируются некоторые факторы и

конечный результат, широко распространена в задачах, решаемых современными методами искусственного интеллекта. В настоящее время в этой области широкое распространение получили так называемые методы машинного обучения.*

При этом по совокупности наблюдений (обучающая выборка), состоящей из нескольких наборов факторов (признаков) и соответствующих им значений целевой переменной, вычисляется возможное значение целевой переменной при новом, не входящим в обучающую выборку, наборе факторов.

Строго говоря, при этом не строится математическая модель рассматриваемого явления, например, зависимости удои от показателей производителя, а решается задача предсказания удои при имеющихся данных. В арсенале машинного обучения имеются методы оценки надёжности такого предсказания в зависимости от числа наблюдений (пар «набор факторов - значение целевой переменной») и характера целевой переменной (качественная или количественная величина). В случае, если целевая переменная имеет качественный характер, например годовой надой либо выше 25 тыс. кг., либо ниже, то говорят о задаче классификации. Если целевая переменная выражается числом, например длительность продуктивной жизни, выраженная числом лактаций, то это задача регрессии.

По данным ресурса Google Scholar, ежегодно, начиная с 2014 года, публикуется около 6 тысяч научных статей на английском языке, посвящённых использованию методов анализа данных и машинного обучения в продуктивном животноводстве. При этом рассматриваются различные проблемы, возникающие в скотоводстве, начиная от автоматической идентификации животного и классификации характера его поведения, предсказания продуктивных характеристик до использования современных методов генетического анализа для селекции по показателям продуктивности, фертильности и дожития.

Настоящий обзор посвящен описанию некоторых применений методов машинного обучения в животноводстве при решении практических задач. Ряд публикаций содержит сравнение эффективности использования различных методов машинного обучения, что позволит читателю составить суждение о целесообразности использования того или иного подхода.

Краткое описание методов машинного обучения будет представлено в отдельной публикации под названием «Методы машинного обучения для решения задач животноводства», где приводится схематическое описание методов, упомянутых в настоящем обзоре, с минимально необходимыми формальными математическими сведениями. Описываются методы построения регрессионных зависимостей и методы классификации с различными модификациями. Основная задача данной статьи – обосновать мотивацию применения конкретных методов, показать возможные области их применения и различия между ними. Заинтересованный читатель сможет получить более глубокие знания в области машинного обучения, воспользовавшись библиографическими ссылками, приводимыми в статье.

Индивидуальная идентификация, классификация поведения и обнаружение теплового стресса

Автоматическая процедура идентификации дойных коров на основе распознавания изображений с использованием методов машинного обучения предложена в работе (Li et al., 2017). Сравнение 4-х методов классификации (линейный и квадратичный дискриминантный анализ, искусственные нейронные сети и метод опорных векторов) показало, что наилучшие результаты обеспечивают квадратичный дискриминантный анализ и метод опорных векторов.

---

*Машинное обучение (*machine learning*) — класс методов искусственного интеллекта, характерной чертой которых является не прямое решение задачи, а обучение в процессе применения решений множества схожих задач (см., напр.: Мерков, 2011).

В работе (Nadimi et al., 2012) описывается система мониторинга и классификации поведения коров, в которой используются мобильные беспроводные датчики стандарта ZigBee, а данные о поведенческих параметрах обрабатываются с помощью многослойных искусственных нейронных сетей, обученных на основе методов обратного распространения ошибки. Система производит классификацию пяти режимов поведения (пастьба, лежание, ходьба, стояние и «другое»). В среднем успешный результат достигается в 76.2% случаев.

Разработана методология классификации поведения на основании информации от датчиков движения и GPS-датчиков, размещенных на электронных воротниках пасущихся коров (Gonzalez et al., 2015). Классифицируются 5 видов активности: добывание корма, жевание, ходьба, отдых и «другие». Величины порогов в дереве решений для каждого животного определяются с использованием смешанных распределений. Алгоритм правильно классифицирует 85.5 и 90.5% всех данных для обучающей и тестовой баз данных, причем наиболее точно определяются режимы добывания корма и жевания.

Электронные воротники с трехмерными акселерометрами служат источниками данных при классификации перечисленных выше пяти режимов поведения пасущихся коров (Dutta et al., 2015). На первой стадии анализа данных была использована новая гибридная методика кластеризации без учителя, включающая алгоритмы анализа главных компонент и самоорганизующихся отображений (карты Кохонена), что позволило изучить естественную структуру сенсорных данных. Далее проводилось сравнение различных методов машинного обучения с учителем (при учете реальных данных полевых наблюдений), в частности таких распространённых методов классификации, как бинарные деревья, линейный дискриминантный анализ, наивный байесовский метод, метод k ближайших соседей и адаптивная нейронная сеть с размытой логикой. Наивысшую среднюю точность классификации 96% удалось получить при использовании классификации bagging ensemble с деревом в роли учителя.

В работе (Smith et al., 2016) предлагается новый подход при классификации пяти режимов поведения (добывание корма, жевание, ходьба, отдых и «другие»), при котором многоклассовая классификация заменяется бинарной «один против всех остальных», т.е. проводится независимое обучение пяти бинарных классификаторов. В качестве источника данных использовался модуль инерциальных измерений на воротнике коровы. Наилучшие результаты классификации были получены для режима добывания корма ($F=0.98$). Наиболее информативными входными характеристиками при обучении всех классификаторов оказались интенсивность движения и наклон головы. Авторы рекомендуют для повышения точности классификации дополнительно использовать блок GPS, датчик давления и микрофон.

Рассмотрена возможность повысить точность классификации трёх режимов поведения пасущихся коров (добывание корма, отдых и ходьба) в результате совместного использования классификатора и марковской модели, описывающей вероятность перехода между режимами при фиксированной временной сегментации данных (Williams et al., 2017).

Проанализировано решение проблемы обнаружения теплового стресса для пасущихся коров с использованием методов машинного обучения без учителя (Shahriar et al., 2016). К сегментированной временной последовательности данных с акселерометров, закреплённых на воротнике коровы, применяется алгоритм кластеризации по методу k -means, на основании которого определяется индекс активности животного – высокий, средний или низкий. Переход на высокий уровень активности ассоциируется с тепловым стрессом; при этом была достигнута точность 82-100%.

Обнаруживать тепловой стресс у коров, находящихся на выпасе, по данным измерения ректальной температуры предлагается в работе (de Sousa et al., 2018). Для оценки ректальной температуры используется неинвазивный метод, основанный на применении инфракрасной термографии, обеспечивающей измерения температуры кожи в 4-х точках, и нейронной сети, обученной с учителем. Результаты экспериментов показали хорошие предикторные возможности предложенного метода ($R^2=0.72$) по сравнению с традиционным методом,

использующим индекс температуры-влажности ($R^2=0.57$).

Выявление поведенческих и физиологических характеристик

Одна из первых работ, посвященных применению методов машинного обучения в молочном животноводстве (Mitchell et al., 1996), содержит описание использования дерева решений и FOIL для обнаружения точки по нормализованным отклонениям суточного удоя. Лучшую точность обнаружения (69%) продемонстрировал метод дерева решений. Однако неприемлемо высокий уровень ложных положительных выводов (73%) свидетельствовал о необходимости увеличения набора входных переменных.

Предложен способ предсказания изменений индивидуальной массы тела животного по данным о прошлой эволюции стада на основании классификации с использованием метода опорных векторов (Alonso et al., 2015). Показано преимущество описываемого подхода по сравнению с построением индивидуальной регрессии по предыдущим значениям живой массы при малом объеме выборки. По утверждению авторов, даже одного значения массы тела достаточно для обеспечения приемлемого качества предсказания.

Для предсказания траектории роста коров (изменений живой массы) использованы методы градиентного бустинга и случайного леса (Flores et al., 2017). Кроме того, применение иерархической кластеризации помогает сформировать более однородные группы по продуктивности.

Описано предсказание параметров рубцовой ферментации на основе анализа жирнокислотного состава молочного жира с использованием искусственных нейронных сетей и различных методов их обучения (Craninx et al., 2008).

Изменение поведенческих характеристик коровы за некоторое время до начала отёла позволяет с достаточной точностью предсказать это событие. Рассматривается применение для этой цели методов машинного обучения (случайный лес, линейный дискриминантный анализ и искусственные нейронные сети) (Borchers et al., 2017). В частности, нейронная сеть обеспечила 100%-ю чувствительность и 86.8%-ю специфичность предсказания за день до отёла и 82.8%-ю чувствительность и 80.4%-ю специфичность предсказания за 8 часов до отёла.

Прогноз продуктивных показателей

В работе (Salehi et al., 2000) описана методика создания нейронной сети, использующей нечёткую логику для предсказания степени принадлежности величины надоя молока к одной из 5 категорий – «очень низкая», «низкая», «средняя», «высокая», «очень высокая». Предсказание основывалось на информации о средней по стаду величине 305-дневного надоя, номеру лактации, количеству дней лактации, стандартной величине суточного удоя и отклонению от неё в тестовый день.

Исследованию точности предсказания объёмов производства молока на основе фермерских данных при использовании искусственной нейронной сети прямого распространения с полиномиальной постобработкой посвящена работа (Sanzogni, Kerr, 2001). Показано, что именно постобработка позволяет получить более высокую точность предсказания по сравнению со стандартным методом множественной линейной регрессии.

Описано построение дерева решений для обнаружения и фильтрации выбросов данных по отдельным лактациям при анализе усреднённых (по рассматриваемой группе) лактационных кривых (Pietersma et al., 2003). Качество классификации определялось на основе 10-кратной кросс-валидации, и доля правильных решений о необходимости удаления «выпадающих» лактаций была доведена до 92%.)

Даны оценка возможностей применения искусственных нейронных сетей для предсказания величины надоя молока и сравнение точности данного метода с регрессионными моделями (Grzesiak et al., 2006). В качестве показателей точности используются коэффициент детерминации, относительная ошибки аппроксимации и корень из среднеквадратичной

ошибки. Показано преимущество нейронных сетей в предсказании как суточного удоя, так и надоя в тестовые дни.

Для генетического анализа данных по суточному выходу молочного белка используется стохастическая регрессия, в которую входят ковариации, зависящие от текущего времени лактации и от индекса температуры-влажности. (Brugemann et al., 2011). В результате исследования было установлено, что имеется тенденция небольшого увеличения наследуемости и дополнительных генетических вариаций стохастической регрессии для диапазона значений индекса температуры-влажности, соответствующего зоне комфорта для коровы.

Описана система оптимизации предсказания производства молочного сырья (Zhang et al., 2016). Система обеспечивает автоматическую генерацию оптимальной конфигурации, статистического анализа и прогноза производства молока в каждой из 9 включённых в нее моделей предсказания, используя данные о животном, стаде и надоях молока. Модели предсказания подразделяются на три категории – подгонка кривой, регрессия и авто-регрессия. Точность предсказания определяется величиной относительной ошибки или коэффициента детерминации и зависит от горизонта предсказания. При большом (365 дней) и среднем (30 дней) горизонте наилучшие результаты продемонстрировала модель, основанная на методе подгонки поверхности, а при малом горизонте (10 дней) – нелинейная авто-регрессионная модель с внешним входом.

Обоснование решений по выбраковке коров

В одной из первых работ по методам машинного обучения в животноводстве рассматривалось их применение для принятия решений по выбраковке на основе учёта показателей продуктивной и селекционной ценности животного в сравнении с их средними значениями по стаду (McQueen et al., 1995). Дерево решений строится для исходных данных, полученных из базы данных среднего размера, подвергнутых существенной предобработке. В каждом узле дерева выдаётся решение об оставлении в стаде или выбраковке, причём коровы моложе двух лет оставляются безусловно из-за отсутствия информации об их продуктивном потенциале. Следующее решение основывается на величине селекционного показателя, который вычисляется по показателям потомства. Объём полученного молока обуславливает окончательное решение. Дерево было обучено на 30% случаев и дало правильную классификацию в 95% оставшихся случаев.

Исследована связь нарушений репродуктивного и метаболического здоровья на первой лактации с интенсивностью выбраковки и продуктивными характеристиками коров (Dhakal et al., 2015) с применением байесовских рекурсивных моделей с гауссовскими порогами (a threshold-Gaussian structural equation model). Установлено наличие сильной причинной связи рассматриваемых нарушений с интенсивностью выбраковки и слабое влияние этих нарушений на продукцию молока для тех коров, которые избежали выбраковки.

В работе (Adamchuk et al., 2016) проводится сравнение эффективности использования искусственных нейронных сетей и классификационных деревьев с применением линейного дискриминантного анализа и других решающих функций при формировании решений о выбраковке коров из стада на основе рутинно фиксируемых данных. Авторы приходят к выводу, что ни один из исследованных методов не гарантирует необходимой точности и для её повышения требуется дополнительно использовать данные от систем непрерывного мониторинга состояния здоровья животного и условий окружающей среды (индекс температуры-влажности).

Диагностика и оценка эффективности лечения мастита и респираторных заболеваний

Исследованию возможностей построения и обучения деревьев решений для обнаружения клинического мастита при автоматическом доении посвящена работа (Kamphuis

et al., 2010). В качестве входных использовались сенсорные данные, включающие электрическую проводимость и цвет. Детектирующие способности деревьев решений оценивались по значениям чувствительности и специфичности в результате 10-кратной кросс-валидации.

При дальнейшем развитии описанного выше метода предложено дополнительно использовать наивную байесовскую систему, обученную определять индивидуальную (для данного животного) априорную вероятность наличия клинического мастита (Steenefeld et al., 2010). На конкретных примерах показано, что при таком дополнении доля успешной работы метода может достигать до 84%, в то время как предположение о равенстве индивидуальных априорных вероятностей обеспечивает долю успешных решений в 21%.

Проводилось сравнение методов обнаружения ранней стадии мастита на основе автоматически фиксируемых данных (электрическая проводимость, соматические клетки и активность лактатдегидрогеназы, суточный удой), предусматривающих применение нейронных сетей и обобщенных аддитивных моделей (Ankinakattea et al., 2013). Качество обучения оценивалось в терминах чувствительности и специфичности. Исследования показали, что нейронные сети и обобщенные аддитивные модели обладают аналогичными возможностями обнаружения мастита; при фиксированной специфичности, равной 80%, чувствительность была близка к 75%.

Разработка дерева решений для оценки экономических аспектов применения различных стратегий лечения клинического мастита легкой и средней тяжести описана в статье (Pinzon-Sanchez et al., 2011). Оптимальной с экономической точки зрения стратегией оказалось применение антимикробной терапии в течение двух дней при лечении клинического мастита, вызываемого грамположительными патогенами, и отказ от такой терапии в случаях грамотрицательных патогенов.

Проведена оценка различных методов классификации для индивидуального предсказания исхода лечения респираторных заболеваний, полученного в день начала лечения коров (Amrineab et al., 2014). Рассматривались 9 алгоритмов классификации, репрезентативно представляющих 5 основных групп: деревья решений, байесовские методы, мета-классификаторы, функциональные/нейронные сети и статистические методы. Было показано, что точность предсказания для каждого классификатора варьируется в зависимости от базы данных (например, для оптимально обученного классификатора диапазон изменения точности может составлять от 63 до 95%). Авторы приходят к выводу, что адекватное соответствие типа классификатора и доступных данных позволит получить предсказание с необходимой точностью.

Оценка эффективности осеменения

Рассматривается возможность практического выявления коров с проблемами искусственного осеменения при применении некоторых традиционных статистических методов и методов машинного обучения (Grzesiak et al., 2010). К «беспроblemному» классу были отнесены коровы, которым для зачатия требовалось не более двух сеансов осеменения, в противном случае коровы считались «проблемными». Лучшие результаты классификации показали искусственные нейронные сети (ANN) и сплайны многовариантной адаптивной регрессии (MARS). Среди факторов, оказывающих наиболее сильное влияние на показатель эффективности осеменения, оказались средняя величина межжотельного интервала и индекс физического состояния животного, а также (в меньшей степени) номер лактации, длительность беременности, пол теленка в предыдущем отёле и возраст коровы.

Результаты сравнения 5 алгоритмов машинного обучения (наивный байес, байесовская сеть, дерево решений, бутстрэпная агрегация, случайный лес) при предсказании эффективности осеменения (беременность – отсутствие беременности) на основе фенотипических и генотипических данных приведены в работе (Shahinfar et al., 2014). Наибольшая точность классификации – 72.3 и 73.6% для первотельных и многотельных коров

соответственно была достигнута с использованием алгоритма случайного леса; при этом площадь (AUC) под кривой ошибок (ROC) равнялась 0.756 и 0.736. Наиболее информативными входными переменными оказались средний по стаду процент успешного осеменения, случаи кетоза, число предыдущих неудачных осеменений, число дней лактации и наличие мастита.

Дальнейшие исследования в этой области нашли отражение в работе, в которой рассматривались 8 алгоритмов машинного обучения (C4.5 дерево решений, наивный байесовский метод, байесовская сеть, логистическая регрессия, метод опорных векторов, регрессия по методу наименьших квадратов, случайный лес, вращающийся лес) (Hempstalk et al., 2015). Предсказание успешности осеменения осуществлялось на основе данных, характеризующих стадо в целом и конкретную корову. Лучшие характеристики предсказания дало применение логистической регрессии (AUC=0.675). Проверка возможности улучшения показателей при добавлении данных спектрального анализа состава молока дала отрицательный результат.

Селекция по показателям продуктивности и геномным оценкам

Применение метода главных компонент в модели BLUP при вычислении геномных селекционных характеристик для продуктивных показателей (объемы производства молока, жира и протеина, процентное содержание жира и белка) рассматривалось в работе (Pintus et al., 2012). Показано, что этот метод обеспечивает такую же точность предсказания, как и стандартные методы BLUP и Байес-А, но позволяет на 90% сократить количество независимых переменных и существенно снизить время вычислений.

Для оценки пяти селекционных показателей (надой молока, выход жира и белка, процентное содержание жира и глубина вымени) были применены две байесовские линейные регрессионные модели (Байес-А и Байес-Lasso), бустинг и геномный BLUP (Jimenez-Montero et al., 2013). Включение геномной информации обеспечило повышение точности предсказания указанных показателей на 0.04-0.42 единиц корреляции Пирсона.

Проводилось сравнение двух байесовских методов (Байес-Ст и Байес-LASSO) с методами BLUP (для породы) и геномный BLUP, парциальной регрессией по методу наименьших квадратов и её «разреженным» вариантом при предсказании надоя молока, процентного содержания жира и вероятности зачатия (Colombani et al., 2013). Показано преимущество байесовских методов.

Получение оценок генетических параметров для надоев молока в месячные тестовые дни первой лактации с применением случайной регрессии описывается в статье (Savagnago et al., 2013). В модели учитывались фиксированный эффект групповых характеристик и случайные генетический и негенетический эффекты индивидуальных характеристик. Для вычисления генетических коэффициентов по надоям молока и скорости спада лактационной кривой использовались индексы собственных векторов ковариационной матрицы.

При геномном предсказании продуктивных показателей для относительно малой генотипированной популяции быков использовались различные методы машинного обучения – Байесовская регрессия (Байес-С), геномный BLUP, одношаговый геномный BLUP и взвешенный одношаговый геномный BLUP (Lourenco et al., 2014). Геномные селекционные оценки, полученные с помощью одношагового геномного BLUP, обладали наименьшим смещением, и этот метод, по мнению авторов, кажется наиболее адекватным для решения поставленной задачи, поскольку позволяет без предобработки включать в расчёты геномную информацию по коровам.

При оценке «трудных для предсказания» продуктивных показателей, таких как процент жира в молоке и технологических характеристик (производство свежего сыра и выход белка) исследовались возможности байесовских моделей (байесовская гребневая регрессия, Байес-А, Байес-В) на основе данных инфракрасной спектроскопии (Ferragina et al., 2015). Сравнение байесовских методов, в частности Байес-А и Байес-В, с используемым в

настоящее время парциальным методом наименьших квадратов показало их преимущество по величине коэффициента детерминации. Применение одношагового геномного BLUP позволило улучшить надежность и уменьшить смещение геномных предсказаний выхода молочного белка (Ma et al., 2015).

Возможность повышения точности геномного предсказания продуктивных показателей (надой молока, выход жира и белка) в результате увеличения обучающей выборки коров рассматривалась в работе (Boison et al., 2017). При этом надёжность предсказания увеличивалась с 0.46 до 0.51 (жир и белок) и с 0.56 до 0.65 (молоко).

На массиве производственных данных исследовано влияние понижения размерности подмножеств SNP (Single Nucleotide Polimorphism – мононуклеотидный полиморфизм), используемых в качестве селекционных маркеров для прогнозирования изменений живой массы животного, на точность прогноза при применении методов случайного леса, градиентного бустинга и экстремального градиентного бустинга (Li et al., 2018).

Селекция по показателям конверсии энергии корма в продукцию

В качестве основного показателя эффективности усвоения энергии корма и её использования для производства молочной и мясной продукции, применяется величина RFI (Residual Feed Intake – остаточное потребление корма), которая служит мерой парциальной эффективности, не зависящей от уровня продуктивности на единицу массы тела. Этот показатель является полезным селекционным критерием для повышения эффективности использования корма.

Рассматривалось получение геномных селекционных величин для RFI и живой массы в возрасте 250 дней с использованием метода геномного BLUP и 2-х байесовских методов (Байес А и мультибайес) (Pryce et al., 2012). Точность селекционных оценок составила в среднем по австралийской и новозеландской популяций 0.37 и 0.31 для RFI и 0.4 и 0.25 для живой массы коров соответственно.

На основе метода случайного леса оценивалась связь между SNP-маркерами и фенотипическими значениями RFI, которые определялись как разность между индивидуальным и средним по когорте потреблением корма (Yao et al., 2013). В результате было проведено ранжирование SNP по степени влияния на RFI с учётом основного эффекта и взаимодействия между факторами. При этом когорты подбирались по году и сезону отёла, году и сезону измерения, возрасту при отёле, количеству дней лактации, надоям молока, массе тела и её изменений.

Оригинальная методика предсказания RFI для молочных коров на первой лактации предложена в работе (Manafiazar et al., 2013). На основе комбинации логарифмического теста максимального правдоподобия и байесовского информационного критерия производилась настройка регрессии по полиномам Лежандра от фиксированных и случайных факторов различного порядка (1 - 5) для моделирования нелинейных профилей энергетических затрат на поддержание метаболической массы тела и продукцию молока. Чтобы получить оценку RFI, строится линейная регрессия реального поступления энергии на полученные модельные характеристики. В результате было показано, что предсказание RFI может быть осуществлено с приемлемым коэффициентом детерминации (0.68) и использовано в соответствующих селекционных программах.

Селекция по показателям фертильности и дожития

Генетическая оценка процента удачных зачатий, проведенная на основе одношагового геномного BLUP (Aguilar et al., 2011), дала почти вдвое большее значение коэффициента детерминации, чем регулярный BLUP, причем возрастание вычислительного времени составило только 33%.

Генетическое предсказание фертильности быков по проценту удачных зачатий у дочерей проводилось на основе одно- или многоядерных моделей, использующих линейные

или гауссовские ядра (Abdollahi-Arpanahi et al., 2017), при этом модели с гауссовскими ядрами обеспечили более высокую точность предсказания.

Разрабатываются вероятностные регрессионные модели для генетического предсказания кумулятивной интенсивности выбытия коров из стада в течение лактации при бинарном представлении их присутствия в стаде в тестовые дни (Sasaki et al., 2015). Показана возможность использования этих характеристик в качестве селекционных индексов для предсказания генетической компоненты дожития.

Статья (van Pelt et al., 2015) посвящена рассмотрению вопроса о том, следует ли рассматривать величину дожития в качестве одной генетической характеристики на протяжении всей жизни коровы. Использование обученной случайной регрессионной модели с полиномами Лежандра второго порядка при разбиении времени предстоящей жизни на 6-месячные (или меньшие) интервалы дало положительный ответ на этот вопрос.

Заключение

Развитие технических средств накопления, хранения и обработки данных привело к широкому внедрению в практику методов анализа данных и машинного обучения. Эти методы применяются в управлении городским хозяйством, транспортом, на производстве и в научных исследованиях. Особенность методов машинного обучения заключается в возможности учёта многих факторов при относительно небольшом числе наблюдений. Например, при исследованиях генома число исследуемых генов может достигать до десятков тысяч, а число образцов ткани ограничивается сотнями. В арсенале машинного обучения имеются методы выделения наиболее значимых факторов, установления скрытых зависимостей, использования априорной информации и информации из сторонних источников.

Все эти методы и подходы нашли применение в современном продуктивном животноводстве. Многие из них были модифицированы, чтобы удовлетворить специфическим запросам отрасли. В настоящее время на основе методов классификации и восстановления регрессионных зависимостей разработаны алгоритмы автоматической идентификации отдельных животных, их поведения на выпасе, предсказания физиологических характеристик. Важное значение для управления стадом имеет прогноз продуктивных показателей и научно обоснованных признаков для выбраковки животных. Проблемы сохранения здоровья животных, диагностика и лечение мастита также решаются с помощью формальных методов анализа данных.

Большой экономический эффект дают методы оценки эффективности осеменения, селекции по показателям продуктивности, эффективности использования корма, фертильности и продолжительности продуктивной жизни, использующие сочетания современных методов генетического анализа и машинного обучения. Как и во всех отраслях деятельности человека, применение современных методов анализа и обработки данных с привлечением вычислительных средств, в том числе методов машинного обучения, позволяет повысить эффективность продуктивного животноводства.

REFERENCES

1. Abdollahi-Arpanahi R., Morota G., Peñagaricano F. Predicting bull fertility using genomic data and biological information. *J. Dairy Sci.* 2017, 100(1): 9656-9666.
2. Adamczyk K., Zaborski D., Grzesiak W., Makulska J., Jagusiak W. Recognition of culling reasons in Polish dairy cows using data mining methods. *Computers and electronics in agriculture.* 2016, 127: 26-37.
3. Aguilar I., Misztal I., Tsuruta S., Wiggans G.R., Lawlor T.J. Multiple trait genomic evaluation of conception rate in Holsteins. *J. Dairy Sci.* 2011, 94(5): 2621-2624.
4. Alonso J., Villa A., Bahamonde A. Improved estimation of bovine weight trajectories using Support Vector Machine Classification. *Computers and electronics in agriculture.* 2015, 110: 36-41.

5. Amrineab D.E., Whiteb B.J., Larsonb R.L. Comparison of classification algorithms to predict outcomes of feedlot cattle identified and treated for bovine respiratory disease. *Computers and electronics in agriculture*. 2014, 105: 9-19.
6. Ankinakattea S., Norberga E., Løvendahla P., Edwardsa D., Højsgaardb S. Predicting mastitis in dairy cows using neural networks and generalized additive models: A comparison. *Computers and electronics in agriculture*. 2013, 99: 1-6.
7. Boison S.A., Utsunomiya A.T.H., Santos D.J.A., Neves H.H.R., da Silva M.V.G.B. Accuracy of genomic predictions in Gyr (*Bos indicus*) dairy cattle. *J. Dairy Sci.* 2017, 100(7): 5479-5490.
8. Borchers M.R., Chang Y.M., Proudfoot K.L., Wadsworth B.A., Stone A.E., Bewley J.M. Machine-learning-based calving prediction from activity, lying, and ruminating behaviors in dairy cattle. *J. Dairy Sci.* 2017, 100(7): 5664-5674.
9. Brügemann K., Gernand E., von Borstel U.U., König S. Genetic analyses of protein yield in dairy cows applying random regression models with time-dependent and temperature x humidity-dependent covariate. *J. Dairy Sci.* 2011, 94(8): 4129-4139.
10. Cherepanov G.G., Kharitonov E.L., Makar Z.N., Mikhal'skii A.I., Novosel'tseva Zh.A. [An analysis of possible approaches to overcome the antagonism between the level of productivity and the viability of the breeding stock by using intensive technologies]. *Problemy biologii produktivnykh zhivotnykh - Problems of Productive Animal Biology*. 2017, 1: 5-27. (In Russian)
11. Colombani C., Legarra A., Fritz S., Guillaume F., Croiseau P., Ducrocq V., Robert-Granié C. Application of Bayesian least absolute shrinkage and selection operator (LASSO) and BayesC π methods for genomic selection in French Holstein and Montbéliarde breeds. *J. Dairy Sci.* 2013, 96(1): 575-591.
12. Craninx M., Fievez V., Vlaeminck B., De Baets B. Artificial neural network models of the rumen fermentation pattern in dairy cattle. *Computers and electronics in agriculture*. 2008, 60(2): 226-238.
13. De Sousa R.V., da Silva Rodrigues A.V., de Abreu M.G., Tabile R.A., Martello L.S. Predictive model based on artificial neural network for assessing beef cattle thermal stress using weather and physiological variables. *Computers and electronics in agriculture*. 2018, 144: 37-43.
14. Dhakal K., Tiezzi F., Clay J.S., Maltecca C. Inferring causal relationships between reproductive and metabolic health disorders and production traits in first-lactation US Holsteins using recursive models. *J. Dairy Sci.* 2015, 98(4): 2713-2726.
15. Dutta R., Smith D., Rawnsley R., Bishop-Hurley G., Hills J., Timms G., Heanry D. Dynamic cattle behavioural classification using supervised ensemble classifiers. *Computers and electronics in agriculture*. 2015, 111: 18-28.
16. Ferragina A., de los Campos G., Vazquez A.I., Cecchinato A., Bittante G. Bayesian regression models outperform partial least squares methods for predicting milk components and technological properties using infrared spectral data. *J. Dairy Sci.* 2015, 98(11): 8133-8151.
17. Flores H., Meneses C., Villalobos J.R., Sanchez O. Improvement of feedlot operations through statistical learning and business analytics tools. *Computers and electronics in agriculture*. 2017, 143: 273-285.
18. González L.A., Bishop-Hurley G.J., Handcock R.N., Crossman C. Behavioral classification of data from collars containing motion sensors in grazing cattle. *Computers and electronics in agriculture*. 2015, 110: 91-102.
19. Grzesiak W., Błaszczak P., Lacroix R. Methods of predicting milk yield in dairy cows –Predictive capabilities of Wood's lactation curve and artificial neural networks (ANNs). *Computers and electronics in agriculture*. 2006, 54(2): P.69-83.
20. Grzesiak W., Zaborski D., Sablik P., Żukiewicz A., Dybus A., Szatkowska I. Detection of cows with insemination problems using selected classification models. *Computers and electronics in agriculture*. 2010, 74(2): 265-273.
21. Hempstalk K., McParland S., Berry D.P. Machine learning algorithms for the prediction of conception success to a given insemination in lactating dairy cows. *J. Dairy Sci.* 2015, 98(8): 5262-5273.
22. Jiménez-Montero J.A., González-Recio O., Alenda R. Comparison of methods for the implementation of genome-assisted evaluation of Spanish dairy cattle. *J. Dairy Sci.* 2013, 96(1): 625-634.
23. Kamphuis C., Mollenhorst H., Feelders A., Pietersma D., Hogeveen H. Decision-tree induction to detect clinical mastitis with automatic milking. *Computers and electronics in agriculture*. 2010, 70(1): 60-68.
24. Kuznetsov V.M. [Development of optimal breeding programs in dairy cattle breeding]. *Zootekhniya - Zootechnics*. 1996, 1: 5-13. (In Russian)

25. Li B., Zhang N., Wang Y.-G., George A.W., Reverter A., Li Y. Genomic prediction of breeding values using a subset of SNPs identified by three machine learning methods. *Frontiers in Genetics*. 2018, 9: 237. doi: 10.3389/fgene.2018.00237.
26. Li W., Ji Z., Wang L., Sun C., Yang X. Automatic individual identification of Holstein dairy cows using tailhead images. *Computers and electronics in agriculture*. 2017, 142(B): 622-631.
27. Lourenco D.A.L., Misztal I., Tsuruta S., Aguilar I., Ezra E., Ron M., Shirak A., Weller J.I. Methods for genomic evaluation of a relatively small genotyped dairy population and effect of genotyped cow information in multiparity analyses. *J. Dairy Sci*. 2014, 97(3): 1742-1752.
28. Ma P., Lund M.S., Nielsen U.S., Aamand G.P., Su G. Single-step genomic model improved reliability and reduced the bias of genomic predictions in Danish Jersey. *J. Dairy Sci*. 2015, 98(12): 9026-9034.
29. Manafiazar G.G., McFadden T.T., Goonewardene L.L., Okine E.E., Basarab J.J., Li P.P., Wang Z. Z.. Prediction of residual feed intake for first-lactation dairy cows using orthogonal polynomial random regression. *J. Dairy Sci*. 2013, 96(12): 7991-8001.
30. McQueen R.J., Garner S.R., Nevill-Manning C.G., Witten I.H. Applying machine learning to agricultural data. *Computers and electronics in agriculture*. 1995, 12(4): 275-293.
31. Merkov A.B. *Raspoznavanie obrazov. Vvedenie v metody statisticheskogo obucheniya*. (Pattern recognition. Introduction to statistical learning methods). Moscow: Editorial URSS. 2011, 250 p. (In Russian)
32. Mikhailenko I.M. [Life cycle management of lactating cows on the basis of probabilistic-statistical and dynamic models]. *Sel'skokhozyaistvennaya biologiya - Agricultural Biology*. 2015, 50(4): 467-475. (In Russian)
33. Mitchell R.S., Sherlock R.A., Smith L.A. An investigation into the use of machine learning for determining oestrus in cows. *Computers and electronics in agriculture*. 1996, 15(3): 195-213.
34. Nadimi E.S., Jørgensen R.N., Blanes-Vidal V., Christensen S. Monitoring and classifying animal behavior using ZigBee-based mobile ad hoc wireless sensor networks and artificial neural networks. *Computers and electronics in agriculture*. 2012, 82: 44-54.
35. Pietersma D., Lacroix R., Lefebvre R.D., Mwade K. Induction and evaluation of decision trees for lactation curve analysis. *Computers and electronics in agriculture*. 2003, 38(1): 19-32.
36. Pintus M.A., Gaspa G., Nicolazzi E.L., Vicario D., Rossoni A., Ajmone-Marsan P., Nardone A., Dimauro C., Macciotta N.P.P. Prediction of genomic breeding values for dairy traits in Italian Brown and Simmental bulls using a principal component approach. *J. Dairy Sci*. 2012, 95(6): 3390-3400.
37. Pinzón-Sánchez C., Cabrera V.E., Ruegg P.L. Decision tree analysis of treatment strategies for mild and moderate cases of clinical mastitis occurring in early lactation. *J. Dairy Sci*. 2011, 94(4): 1873-1892.
38. Pryce J.E., Arias J., Bowman P.J., Davis S.R., Macdonald K.A., Waghorn G.C., Wales W.J., Williams Y.J., Spelman R.J., Hayes B.J. Accuracy of genomic predictions of residual feed intake and 250-day body weight in growing heifers using 625,000 single nucleotide polymorphism markers *J. Dairy Sci*. 2012, 95(4): 2108-2119.
39. Salehi F., Lacroix R., Wade K.M. Development of neuro-fuzzifiers for qualitative analyses of milk yield. *Computers and electronics in agriculture*. 2000, 28(3): 171-186.
40. Sanzogni L., Kerr D. Milk production estimates using feed forward artificial neural networks. *Computers and electronics in agriculture*. 2001, 32(1): 21-30.
41. Sasaki O., Aihara M., Nishiura A., Takeda H., Satoh M. Genetic analysis of the cumulative pseudo-survival rate during lactation of Holstein cattle in Japan by using random regression models. *J. Dairy Sci*. 2015, 98(8): 5781-5795.
42. Savegnago R.P., Rosa G.J.M., Valente B.D., Herrera L.G.G., Carneiro R.L.R., Sesana R.C., Faro L.E., Munari D.P. Estimates of genetic parameters and eigenvector indices for milk production of Holstein cows. *J. Dairy Sci*. 2013, 96(11): 7284-7293.
43. Shahinfar S., Page D., Guenther J., Cabrera V., Fricke P., Weigel K. Prediction of insemination outcomes in Holstein dairy cattle using alternative machine learning algorithms. *J. Dairy Sci*. 2014, 97(2): 731-742.
44. Shahriar S., Smith D., Rahman A., Freeman M., Hills J., Rawnsley R., Henry D., Bishop-Hurley G. Detecting heat events in dairy cows using accelerometers and unsupervised learning. *Computers and electronics in agriculture*. 2016, 128: 20-26.
45. Smith D., Rahman A., Bishop-Hurley G.J., Hills J., Shahriar S., Henry D., Rawnsley R. Behavior classification of cows fitted with motion collars: Decomposing multi-class classification into a set of binary problems. *Computers and electronics in agriculture*. 2016, 131: 40-50.

46. Steeneveld W., van der Gaag L.C., Barkema H.W., Hogeveen H. Simplify the interpretation of alert lists for clinical mastitis in automatic milking systems. *Computers and electronics in agriculture*. 2010, 71(1): 50-56.
47. van Pelt M.L., Meuwissen T.H.E., de Jong G., Veerkamp R.F. Genetic analysis of longevity in Dutch dairy cattle using random regression. *J. Dairy Sci.* 2015, 98(6): 4117-4130.
48. Williams M.L., James W.P., Rose M.T. Fixed-time data segmentation and behavior classification of pasture-based cattle: Enhancing performance using a hidden Markov model. *Computers and electronics in agriculture*. 2017, 142(B): 585-596.
49. Yao C., Spurlock D.M., Armentano L.E., Page Jr C.D., VandeHaar M.J., Bickhart D.M., Weigel K.A. Random Forests approach for identifying additive and epistatic single nucleotide polymorphisms associated with residual feed intake in dairy cattle. *J. Dairy Sci.* 2013, 96(10): 6716-6729.
50. Zhang F., Murphy M.D., Shaloo L., Ruelle E., Upton J. An automatic model configuration and optimization system for milk production forecasting. *Computers and electronics in agriculture*. 2016, 128: 100-111.

Application of machine learning methods in solving problems of productive animal husbandry

Mikhalskii A.I., Novoseltseva Zh.A.

*Trapeznikov Institute of Control Sciences RAS,
Moscow, Russian Federation, ipuran@yandex.ru*

ABSTRACT. Application of precise methods for forecasting individual productive values, fertility, the length of productive life of cows, and optimizing herd management are promising areas for improving the efficiency of modern productive animal husbandry. Practical application of electronic micro-sensors systems with wireless data transmission, which provides real-time data acquisition, caused the need for effective analysis and computer processing of the received information. The review considers a number of applied problems in the field of animal husbandry, in which the methods of machine classification and learning have found a successful application. In the arsenal of machine learning there are methods for identifying the most significant factors, establishing hidden dependencies, using a priori information and information from outside sources. At present, algorithms for automatic identification of individual animals, their behavior and state on grazing, prediction of physical and physiological parameters, prediction of breeding value, etc. have been developed based on methods of classification and restoration of dependencies. Main sections of the review: individual identification, classification of behavior and detection of heat stress; the identification of behavioral and physiological characteristics; prognosis of productive indicators; substantiation of decisions on culling; diagnosis and evaluation of the effectiveness of treatment of mastitis and respiratory diseases; assessment of the effectiveness of insemination; selection by performance indicators and genomic estimates; selection based on the conversion of feed energy into production; selection on fertility and survival indicators. The wide range and cost-effectiveness of the problems being solved testify to the high potential of machine learning and methods of data analysis in solving practical problems of animal husbandry.

Keywords: productive livestock, data analysis, multifactorial biosystems, classification, diagnostics, machine learning

Problemy biologii produktivnykh zhivotnykh - Problems of Productive Animal Biology, 2018, 4: 98-109

Поступило в редакцию: 09.09.2018

Получено после доработки: 21.09.2018

Михальский Анатолий Иванович, д.б.н., к.т.н., г.н.с., 8(915)199-55-26,
ipuran@yandex.ru; mprostok@yandex.ru;

Новосельцева Жанна Анатольевна, к.т.н., с.н.с., 8(495)334-88-91, novoselc.janna@yandex.ru