

ИДЕНТИФИКАЦИЯ НА АНОМАЛИИ В СИСТЕМИ ЗА ИНТЕЛИГЕНТНО ЗЕМЕДЕЛИЕ

Станимир Стоянов, Венета Табакова-Комсалова,
Ася Стоянова-Дойчева, Емил Дойчев

***Резюме.** В последните години, разпознаването на аномалии в различни ситуации се оформя като област на нарастващ теоретичен и практически интерес. В статията се представят идея и концепция за разработване на услуга за идентифициране и локализиране на възможни аномалии в земеделски процеси. Ще бъде представена възможна класификация на възможните типове аномалии, характерни за земеделието. Услугата ще използва методи от дълбокото машинно учене (deep learning) и обясним изкуствен интелект (XAI).*

Ключови думи: телигентно земеделие, откриване на аномалии, дълбоко машинно учене, обясним изкуствен интелект.

Въведение

Откриването на аномалии е активно изследователско поле, широко прилагано за много търговски и критични приложения, включително здравеопазване, откриване на измами, промишлена предсказуема поддръжка и киберсигурност. Това е процес на откриване на отклонения, необичайни модели и необичайни наблюдения, които се отклоняват от установеното нормално поведение и очакваните характеристики в системи, набори от данни или околни среди. Откриването на аномалии все повече се обособява като самостоятелен клон на машинното учене.

Airbus [1] използва такива техники за прогнозиране на повреди в реактивни двигатели и откриване на аномалии в телеметричните данни, предавани от международната космическа станция. Компаниите за кредитни карти го използват за откриване на измами с кредитни карти [2, 3]. Целта на откриването на аномалия е да се идентифицират отклонения в данните – проби, които не са „нормални“ в сравнение с други. В случай на измами с кредитни карти допускането е, че ако транзакциите са подложени на алгоритъм за откриване на аномалии, измамните транзакции ще се покажат като аномални, докато законните транзакции

няма. AI Anomaly Detector на Майкрософт [4] се използва за откриване на аномалии във времеви серии в различни приложения. Услугата разглежда времеви серии от различни типове данни и избира най-добрия алгоритъм за откриване на аномалии за да осигури висока точност.

Дефинирането и специфицирането на аномалии не е лесна задача, тъй като границите между нормалното и необичайното поведение могат да бъдат специфични за домейна и да подлежат на нива на толерантност към риска, определени от бизнеса, организацията и индустрията. Например, неправилен сърдечен ритъм от данни от времеви серии на електрокардиограма (ЕКГ) може да сигнализира за риск от сърдечно-съдови заболявания, докато колебанията в цените на акциите могат да се считат за нормални въз основа на пазарното търсене. По този начин няма универсална дефиниция на аномалия и универсално решение за откриване на аномалия.

Откриването на аномалии с методи на изкуствения интелект (и по специално на дълбокото учене) показва обещаващи резултати при справяне с предизвикателствата с рядкото естество на аномалии, сложно моделиране на високоразмерни данни и идентифициране на нови аномалии класове. Основният интерес към откриването на аномалии често е фокусиран върху изолирането на нежелани случаи на данни, като продуктови дефекти и рискове за безопасността. Други интереси включват подобряване на производителността на модела чрез премахване на шумни данни или неуместни отклонения и идентифициране на нововъзникващи тенденции от набора от данни за конкурентно предимство.

Съществено предизвикателство е да се намерят подходящи методи за откриване на аномалии. Поради голямото разнообразие от аномалии универсален метод едва ли може да бъде разработен. По-надежден подход е да се търсят такива методи за конкретна приложна област. Съществуват различни подходи за откриване на конкретни аномалии. В съвременето, особени надежди се възлагат на методи, използващи интелигентни компоненти. Обикновено такива подходи използват методи, базирани на машинно учене. Могат да се използват методи за контролирано (вкл. невронни мрежи) и неконтролирано учене, усилено учене и други такива.

Друго предизвикателство е породено от необходимостта за получаване, комбиниране и анализ на данни от различни и разнообразни източници, като например сензорни мрежи, разпознаване на обекти в снимков

материал, провеждане на анкети с потребители или пациенти.

Основна цел на предлагания проект е да се създадат модел и архитектура на интелигентна услуга за откриване, идентифициране и локализиране на възможни аномалии в критични за интелигентното земеделие и животновъдство процеси.

Защо предлагаме целевата приложна област да бъде интелигентно земеделие и животновъдство? Земеделието и животновъдството са важен отрасъл за България. Решенията в тази област обикновено се вземат на основата на непълни и непрецизни данни, което води до неефективни резултати. С повишаване на прецизността на получаваната информация се очаква значително повишаване на ефективността на вземаните решения. По-специално, навременното и прецизното откриване на аномалиите в селското стопанство има съществено значение за повишаване на селскостопанската продукция. Разкриването на възможни аномалии може да помогне за навременна реакция при заболявания на земеделските култури и селскостопански животни, при появяване на вредители, за подпомагане оптималното протичане на различни биологични и физиологични процеси. Освен това, навременното откриване и идентифициране на аномалии може да помогне за минимизиране на товарването на околната среда с различни химически вещества, използвани в земеделието. Интелигентното селско стопанство трябва да отговори на предизвикателства като климатични промени, екстремни метеорологични условия и въздействието на интензивните земеделски практики върху околната среда. Ще помогне на фермерите да намалят отпадъците и да увеличат производителността. Интелигентното земеделие е по същество чиста и устойчива високотехнологична система за добиване на храни.

Селското стопанство със сигурност е най-важният производител на храна в световен мащаб. В същото време той изразходва повече от 70% от използваната вода и допринася значително за замърсяването на водата. Поливното земеделие е изправено пред нарастваща конкуренция в световен мащаб за достъп до надеждни, евтини и висококачествени водни ресурси. Въпреки това напояването, като основен инструмент и определящ фактор за въздействието върху селскостопанското производство и ресурсите на околната среда, играе критична роля за продоволствената сигурност и устойчивостта на околната среда. От една страна, иновативните технологии и практики за напояване могат да подобрят използваемостта на наличните водни ресурси и ефективността на селс-

костопанското производство като цяло. От друга страна, те биха смекчили търсенето на вода и проблемите с качеството на водата. Население се увеличава с бързи темпове, което води до огромен натиск върху ограничените водни ресурси. Само около 3% от общата вода на земята е прясна вода, която може да се използва за различни приложения, включително напояване. Следователно, ефективното управление на водата за напояване е от решаващо значение за оцеляването на хората. В търсене на решения на тези проблеми се появи областта на интелигентното земеделие. В момента интелигентното земеделие стана обект на нарастващ научен и практически интерес. Изкуственият интелект (AI), усилен с концепции като Cyber-Physical System (CPS) и Cyber-Physical-Social System (CPSS) и технологии като Internet of Things (IoT) и Big Data, се очертава като мощен инструмент за внедряване на приложения за интелигентно селско стопанство. Възможен подход за интегриране на такива разнообразни технологии е създаването на специализирани платформи. От особено важно значение е също изборът на подходяща технология за разработване на необходимите софтуерни средства. Прецизността на наблюденията и отчитането на параметрите на жизнената среда на земеделските култури става все по-необходима в условията на постоянно намаляване на ресурсите като вода и енергия, за да се гарантира максимална ефективност на тяхното използване. Модерната високотехнологична селскостопанската система трябва да позволява активен контрол, осъществяван от различни видове сензори и задвижващи контролни механизми, за да се създаде благоприятен климат за растежа на културите. Друг съществен аспект на интелигентното земеделие е необходимостта от моделиране на процеси. За осъществяване на тази цел възможностите за откриване и идентифициране на аномалии в селскостопанските процеси, като например вегетация на земеделските култури, поведение на селскостопанските животни стават основополагащи.

Платформата ЗЕМЕЛА

ЗЕМЕЛА [5] е регионална платформа, където фокусът се поставя върху максимално експлоатиране на спецификата на наблюдавания земеделски район, като основен източник за данни са сензорни мрежи. По тази причина е съществено платформата да има възможности за динамично адаптиране към околната среда. Отчитайки тези специфика сме избрали агентно-ориентиран подход за реализиране на платформата, като нейното ядро е персонален асистент, имплементиран като интели-

гентен рационален агент. Регионалният характер предполага ефективно използване на платформата за по-малки площи насаждения. В този смисъл ЗЕМЕЛА е предназначена основно за подпомагане на малки и средни земеделски стопани, отглеждащи зеленчуци и плодове. В България производителите на зеленчуци и плодове (за разлика от производителите на пшеница, царевица и слънчоглед) са предимно малки и средни стопани, стопанисващи сравнително малки площи и това производство е недостатъчно финансово подпомогано от държавата. Очакваме въвеждането на системи като платформата да допринесе за повишаване на добивите и подобряване ефективността на този вид производство.

Ядрото на платформата е персонален интелигентен асистент, разработен като VDI агент [6]. Земеделските процеси се моделират (представят) като събития. Разработен е специален събитийен модел, който е в процес на постоянно разширяване и прецизиране. Основната функция на персоналния асистент е идентифициране и локализиране на аномалии, специфични за земеделието. Такива могат да бъдат отклонения в различните фази на вегетацията на зеленчуците, признаци на напад от вредители и симптоми на заболявания. Поради регионалния характер на платформата основополагаща става задачата за адекватната адаптация на персоналния асистент за мониторинг в съответствие с особеностите на вегетацията на наблюдавания вид зеленчук и работа в конкретните специфични условия на района. За постигане на тази цел особено внимание се обръща на наблюденията и проведените експерименти, резултатите от които са обобщени в тази статия.

Събитийен модел на платформата ЗЕМЕЛА

Теоретична основа на ЗЕМЕЛА е събитийен модел [7], който класифицира възможните събития в следните групи:

- *Базови събития* – атомарни събития, които могат да инициират различни събития в платформата, обслужваща други събития. Типични примери за основни събития са дата, час, екстремни стойности на сензори, местоположение.
- *Системните събития* – събития в инфраструктурата на платформата, като например генериране/премахване на динамични компоненти или изпращане/получаване на съобщения.
- *Домейн-събития* – събития, типични за интересуващата ни област. Тези събития се разделят на две други групи – планирани и

инцидентни. Планираните събития са тези, които имат времево събитие като параметър и идентификацията на събитието се насочва от това времево събитие. Инцидентните събития не могат да бъдат предварително планирани – те се случват в резултат на случайност.

За адаптиране на платформата за идентифициране на аномалии инцидентните събития ще бъдат разширени с нов подклас „аномалии“. Аномалиите в земеделието могат да се проявяват в различни форми. Най-общо различаваме следните три вида:

- *Точкови аномалии* – съществуват в набор от данни, където отделна точка от данни е извън границата на нормалното разпределение.
- *Колективни аномалии* – възникват само когато група от свързани записи на данни или последователности от наблюдения се появяват колективно и значително се различават от останалия набор от данни.
- *Контекстуални аномалии* – разглеждат се спрямо модални атрибути като такива за време, пространство или условности.

Подходи за идентифициране на аномалии

В сегашната версия на персоналния асистент е имплементиран подход за идентифициране на аномалии, който използва сравняване на шаблони. Предлагаме той да бъде разширен с два нови метода. Първият е базиран на PCA (Principal Component Analysis) подход за неконтролирано машинно учене, който се използва за редуциране на размерността на използваните набори от данни. Намаляването на размерността помага да се справим с едно от най-често срещаните предизвикателства в машинното учене – невъзможност за ефективно и ефикасно обучение върху данните поради големия размер на пространството на техните характеристики. За идентифициране на аномалии с помощта на PCA предлагаме алгоритъм, включващ стъпките, които ще бъдат представени накратко.

Подготовка на данните. Трябва да подготвим набора от данни за наблюдавания процес (в случая вегетация на селскостопанската култура), като генерираме матрицата на характеристиките и масива от етикети. Разделяме набора от данните на две части – данни за обучение и за тестване. Няма да използваме етикетите, за да извършваме открива-

не на аномалии, но ще използваме етикетите, за да помогнем да оценим системите за откриване на аномалии, които изграждаме.

Дефиниране на функция за оценка на аномалии. След това трябва да дефинираме функция, която е в състояние да оценява каква е вероятността за наличието на аномалия в съответната фаза на вегетацията. Колкото по-големи са отклоненията от очакваните стойности на параметрите, характеризиращи нормалното състояние на растението, толкова по-вероятно е наличието на аномалия. Алгоритмите намаляват размерността на данните, докато се опитват да минимизират грешката при реконструкцията. С други думи, тези алгоритми се опитват да уловят най-важната информация за оригиналните характеристики по такъв начин, че да могат да възстановят възможно най-добре оригиналния набор от характеристики (свойства) от редуцирания набор от характеристики. В контекста на нашия набор от данни, алгоритмите ще имат най-голямата грешка при реконструкция на тези случаи, които са най-трудни за моделиране – с други думи, тези, които се случват най-рядко и могат да се приемат за аномални.

Дефиниране на показатели за оценка. Въпреки че няма да използваме етикети, за да създадем решения за неконтролирано откриване на аномалии, ще използваме такива за оценка на решенията, които разработваме. Етикетите ще ни помогнат да разберем колко добри са тези решения при улавяне на известни шаблони на аномалии.

Откриване на PCA аномалии. Сега ще използваме PCA за да научим основната структура на набора от данни за отделните фази на вегетацията на наблюдаваните растения. След като научим тази структура, ще използваме обученния модел, за да реконструираме протичането на фазите на вегетацията и след това да изчислим колко различни са реконструиранията фази от оригиналните. Фазите, които PCA възстановява най-различно са най-ненормалните и най-вероятно са аномални.

Вторият метод се основава на принципи от XAI (eXplainable Artificial Intelligence). XAI техниките се прилагат към различни модалности на данни, включително изображения, текст и таблични данни. Тъй като аномалиите са редки по природа, много модели за дълбоко учене научават представянето на характеристиките от нормални наблюдения и откриват аномалии чрез измерване на отклонения от известното поведение. За нашите цели ще използваме XAI метод за откриване на аномалии във времеви серии – обикновено това са данни получавани от сензорни

мрежи.

Заклучение

Настоящата разработка представя идеята за създаване на услуга за идентифициране на възможни аномалии в земеделски процеси, поспециално при вегетацията на растенията. Следващите стъпки са да се проведат експерименти и с предлаганите два нови метода и да се имплементират техни прототипи.

Благодарности

Това изследване е подкрепено от Националната програма за научни изследвания „Интелигентно растениевъдство“, Договор за безвъзмездна финансова помощ № D01-65/19.03.2021 г., одобрен с Решение на Министерския съвет № 866/26.11.2020 г., както и от проект BG05M2OP001-1.001-0003: „Център за върхови постижения по Информатика и информационни и комуникационни технологии“, финансиран по Оперативна програма „Наука и образование за интелигентен растеж“.

Литература

- [1] P. Grashorn, J. Hansen, M. Rummens, How Airbus Detects Anomalies in ISS Telemetry Data Using TFX, April 09, 2020, <https://blog.tensorflow.org/2020/04/how-airbus-detects-anomalies-iss-telemetry-data-tfx.html>.
- [2] F. Tanant, Fraud Detection with Machine Learning & AI, <https://seon.io/resources/fraud-detection-with-machine-learning/>; Sanket Sarwade, Anomaly Detection in Credit Card Fraud, May 12th, 2023, <https://www.analyticsvidhya.com/blog/2023/05/anomaly-detection-in-credit-card-fraud/>.
- [3] Sanket Sarwade, Anomaly Detection in Credit Card Fraud, May 12th, 2023, <https://www.analyticsvidhya.com/blog/2023/05/anomaly-detection-in-credit-card-fraud/>.
- [4] AI Anomaly Detector, <https://azure.microsoft.com/en-us/products/ai-services/ai-anomaly-detector>.
- [5] S. Stoyanov, J. Todorov, I. Stoyanov, V. Tabakova-Komsalova, L. Dukovska, ZEMELA – An Intelligent Agriculture Platform, *Big Data, Knowledge and Control Systems Engineering – BdKCSE'2021*, 28–29 October 2021, Sofia, Bulgaria.

- [6] I. Stoyanov, A. Stoyanova-Doycheva, I. Krasteva, Z. Uhr, A Personal Assistant Supporting Agriculture Operators, *2020 IEEE 10th International Conference on Intelligent Systems*, 24-26 August, Varna.
- [7] S. Stoyanov, A. Stoyanova-Doycheva, V. Ivanova, V. Tabakova-Komsalova, An Event Model for Smart Agriculture, *2021 IEEE International Conference Automatics and Informatics (ICAI)*, 30 September – 2 October 2021. Varna

Станимир Стоянов¹, Венета Табакова-Комсалова²,
Ася Стоянова-Дойчева³, Емил Дойчев⁴
^{1,2,3,4} Пловдивски университет „Паисий Хилендарски“
Факултет по математика и информатика,
бул. „България“ № 236, Пловдив, България
Автор за кореспонденция: stani@uni-plovdiv.bg

IDENTIFICATION OF ANOMALIES IN INTELLIGENT AGRICULTURE SYSTEMS

Stanimir Stoyanov, Veneta Tabakova-Komsalova, Asya
Stoyanova-Doycheva, Emil Doichev

Abstract. *In recent years, anomaly detection in various situations has emerged as an area of increasing theoretical and practical interest. This paper presents the idea and the concept for developing a service for identifying and locating possible anomalies in agricultural processes. A classification of the possible types of anomalies that are typical for agriculture is also presented. The service uses methods from the field of deep learning and explainable artificial intelligence (XAI).*

Key words: Smart agriculture, Anomaly detection, Deep learning, Explainable artificial intelligence.