

УДК 004.032.26

ПОДГОТОВКА МАТЕРИАЛОВ ДЛЯ МАШИННОГО ОБУЧЕНИЯ С ПОМОЩЬЮ ВИРТУАЛЬНОЙ СТУДИИ



Коркин Л.Р.

ассистент кафедры Инженерной психологии и эргономики БГУИР, магистр техники и технологии, аспирант.



М.В.Коркина

младший научный сотрудник производственно-испытательной лаборатории РУП «Институт мясо-молочной промышленности»



Саевич К.Ф.

профессор кафедры физикохимии материалов и производственных технологий БГЭУ, доктор биологических наук, профессор



К.Д. Яшин

заведующий кафедрой инженерной психологии и эргономики БГУИР, кандидат технических наук, доцент

РУП «Институт мясо-молочной промышленности» (сокращённое название), Республика Беларусь, Минск.

Белорусский государственный университет информатики и радиоэлектроники», Республика Беларусь, Минск.

Л. Р. Коркин

В 2018 году окончил БГУИР, специальность «Инженерно-психологическое обеспечение информационных технологий». Квалификация инженер-системотехник. В 2020 году окончил магистратуру по специальности «Управление безопасностью производственных процессов» с присвоением академической степени магистра техники и технологии. Читаемые курсы: Безопасность жизнедеятельности человека, интерфейсы информационных систем, схемотехника, активные элементы систем безопасности.

М. В. Коркина

В 2017 году окончила Белорусский государственный университет, специальность «Микробиология». Квалификация биолог-микробиолог. Младший научный сотрудник производственно-испытательной лаборатории РУП «Институт мясо-молочной промышленности».

К. Ф. Саевич

В 1984 защитил диссертацию на соискание учёной степени кандидата сельскохозяйственных наук, в 1992 – доктора биологических наук, с 1998 – профессор. Области научных интересов: Экология, Рациональное природопользование, Безопасность жизнедеятельности. Достигнутые научные результаты: изобретений – 2, монографий – 6, брошюр – 20, учебных пособий – 1, научных публикаций

более 150, учебно-методических работ более 20, участие в научных семинарах и конференциях: Международных более 20, Республиканских более 15. Всего опубликовано более 220 работ.

К. Д. Яшин

Руководитель научно-исследовательской группы НИГ – 7.1 «Системы и приборы экологического мониторинга в управлении безопасностью жизнедеятельности».

Аннотация. В настоящее время для идентификации микроорганизмов по морфологическим признакам в учреждениях контроля качества продуктов питания используют определитель бактерий Берджи (1997 г, в 2-х томах). Этот метод трудоемкий и требует большого внимания и точности выполнения. Таким образом, разработка автоматизированной системы распознавания снимков колоний микроорганизмов на чашках Петри для последующего подсчета и предварительной идентификации колоний является актуальной задачей.

Ключевые слова: Машинное обучение, искусственный интеллект, микробиология, колония микроорганизмов.

Введение

Чтобы превратить распознавание объектов на изображении из идеи в реальность необходимо использовать такой вид машинного обучения, как обучение нейросети которое всегда начинается с подготовки материалов, а в нашем случае это размеченные изображения (снимки колоний микроорганизмов (СКМ))колоний, которые нужно будет распознавать. С их помощью мы будем обучать нашу систему распознавания по СКМ (СРСКМ). Для этого нужно – создать систему, не только распознающую колоний микроорганизмов (КМ) по изображениям, но и производить СКМ необходимого качества на разных этапах развития КМ. Процесс усложняется тем, что существует большое количество различных КМ, многие из которых морфологически очень похожи. Сейчас для того, чтобы человек научился различать КМ, выпускаются специальные каталоги или плакаты, на которых показано, чем одна колония отличается от другой. С их помощью также можно будет обучать систему распознаванию колоний. Это процесс усложняется еще и тем, что один и тот же вид микроорганизмов растёт по-разному в зависимости от условий культивирования (температуры, pH, доступа кислорода, источников углерода и др.). Поэтому в настоящее время идентификация микроорганизмов по морфологическим признакам проводят с помощью определителя бактерий Берджи.

Методы и материалы

В процессе обучения СРСКМ необходимо учитывать большое количество различных факторов, таких как разнообразность материалов, качество самих изображений и др. При реализации проекта планируется использовать возможность идентифицировать микроорганизмы по СКМ, выращенных на чашках Петри (рисунок 1) в разных питательных средах и при различном освещении. Следует учитывать, что наряду с качественными изображениями на практике придётся использовать также СКМ более низкого качества.

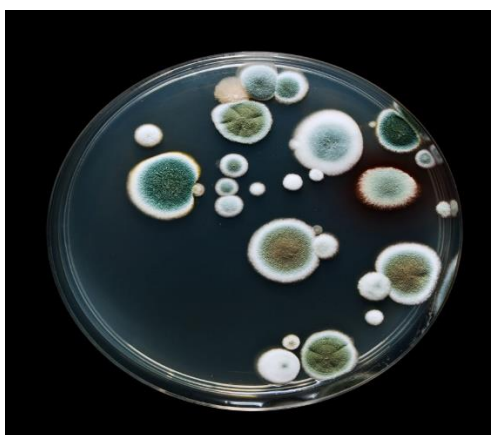


Рисунок 1. Снимок колоний плесневых грибов, выросших на чашке Петри

Как известно машинное обучение строится на трех основных принципах: данные, признаки, алгоритмы.

Данные (Data) – это базовая информация. Сюда входят необходимые выборки данных, работе с которыми требуется обучить систему. Чем больше данных загружается в СРСКМ, тем точнее и лучше она будет работать. Данные напрямую зависят от задачи, которая стоит перед нейронной сетью.

Например, чтобы научить почту отфильтровывать спам от важных писем, необходимы примеры. И чем больше их выборка, тем лучше. Таким образом система учится воспринимать конкретные слова, как признаки спама и отправлять такие письма в отдельную папку. А в нашем случае это особенности параметров изучаемых колоний.

Выборка данных и ее обработка занимает большой промежуток времени. Что повысит точность оценки и прогнозирования, которую мы получаем в итоге. При этом очень важно предоставлять максимум разнообразной информации, поскольку система может увидеть связь там, где человек ее не заметит.

Изучение признаков (свойства, метрики, фичи, характеристики, features и другие признаки) желательно проводить в тесном сотрудничестве с конечным пользователем. Для работы необходимо определить ключевые потребности и совместно решить какие именно характеристики и свойства должна отслеживать система в результате обучения.

Поскольку правильность определения свойств напрямую влияет на результат, который необходимо получить, их отбор занимает зачастую больше времени, чем сам процесс обучения. Здесь главное – не ограничивать набор характеристик, так как можно исказить восприятие системы. А вместе с ним и конечный результат.

Исходя из анализа свойств этих компонентов можно определить признаки и в зависимости от задач установить основные свойства.

Необходимо также разработать алгоритм, который представляет систему последовательных операций и методов для решения определенной задачи. Эту задачу обычно решает группа программистов. Именно от выбранного метода напрямую зависит скорость и точность результата обработки исходных данных.

Если изначально будет некорректно обработаны имеющиеся данные и признаки, или конечный пользователь проигнорирует рекомендации, тогда необходимо пересмотреть и доработать алгоритм работы системы или произвести оптимизацию самой системы.

Очень важно подготовить изображения всех вероятных микроорганизмов с разными фонами, разным освещением, и в разных положениях. Каждое изображение КМ должно быть размечено, выделено, чтобы СРСКМ понимала, где на СКМ объекты, которые нужно научиться распознавать. Если делать это всё вручную, понадобится очень много времени.

Разметка данных на изображениях или видео необходима для загрузки моделей глубинного обучения. Так система учиться лучше узнавать и распознавать объекты.

В нашей работе было принято решение создать виртуальную студию на основе Blender, с помощью которой можно будет сгенерировать размеченные изображения для обучения системы.

В начале работы необходимо сделать по одному снимку колоний каждого исследуемого вида микроорганизмов и подобрать некоторое количество разных фоновых изображений. С помощью скрипта, подключенного к Blender, подставить изображение колоний микроорганизмов на каждый фон. В развернутой нами студии есть виртуальная камера, которая позволяет создать анимацию при изменении положения. В процессе анимации скрипт делает рендеры каждого кадра, и у нас появляется база изображений (материалов) для дальнейшего обучения. Этот процесс получения изображений с помощью 3D-программы называется рендерингом, (Рендер – это изображение, которое получается в результате прорисовки трехмерной модели в изображение).

В процессе рендеринга скрипт делает автоматическую разметку материалов (боксинг). В результате получается база рендеров и соответствующие каждому рендеру xml файлы с указанием точных координат объекта на изображении, а в нашем случае КМ рисунок 2. Использование рендеров автоматизирует и существенно ускоряет процесс подготовки необходимых данных для обучения СРСКМ.



Рисунок 2. Подготовленные и размеченные СКМ

Все полученные материалы виртуальная студия сразу распределяет в две папки test и train. Папка train будет использоваться для обучения СРСКМ, а в папку test попадает каждый пятый рендер, что будет базой для самотестирования системы.

Во время обучения система делает так называемые чекпоинты – проверяет насколько правильно идет процесс. Для этого нужен материал похожий на изучаемый. Такой метод обучения универсален для любых объектов, которые должна будет научиться распознавать система. Этот метод называется обучением СРСКМ на основе синтетических материалов.

Для выполнения этой работы нами планируется подготовить более сорока тысяч изображений колоний различных видов микроорганизмов, выросших на чашках петри. Часть СКМ будет выполнено вручную, но большинство – сгенерировано в виртуальной студии в полуавтоматическом режиме.

Первый этап обучения планируется длиться не менее 3-5 месяцев. За это время СРСКМ должна пройти не менее трёх миллионов шагов. Результат может быть неудовлетворительный, так как система будет распознавать и идентифицировать колонии микроорганизмов только со снимков, сделанных в идеальных условиях, а тень, засвет, искажение приведут к критической потере точности идентификации. Поэтому на следующем этапе обучения СРСКМ необходимо учитывать особенности получения снимков.

Исходя из того, что пользователь будет в основном делать снимки при помощи электронного микроскопа, а в лаборатории свои особенности помещения чем в фотостудии, где возможны условия для изготовления СКМ высокого качества. Поэтому при подготовке и обучении системы необходимо учесть такие факторы, как:

- неудачный ракурс;
- слишком яркое или тусклое освещение;
- плохой фон;
- фото с перспективой.

В связи с этим планируется второй этап обучения СРСКМ, который займет еще несколько месяцев. В результате будет подготовлено еще целый ряд материалов для обучения, а также будет изменен подход. В результате система должна пройти ещё 3000000 шагов и в итоге выдать результат, который должен оказаться гораздо менее чувствительный к качеству входящего изображения, а значит и к качеству СКМ.

Автоматическое описание снимков микроорганизмов, выросших на чашках Петри, с помощью системы позволит сортировать и классифицировать СКМ колоний микроорганизмов намного быстрее и проще. Система должна буквально «читать» визуальное содержимое и «объяснять» полученные данные в виде текста и таблицы. Таким образом смысл изображения становится более понятным для пользователя.

Заключение.

В результате выполнения этой работы станет возможным использование указанных

современных методов интерпретации изучаемых СКМ колоний микроорганизмов в виде таблицы, в которой будет содержаться качественное и количественные характеристики состава микробиоты, что повысит качество оценки безопасности продуктов питания для населения и продовольственную безопасность в целом, устранив субъективизм, как «человеческий фактор» в этом важном деле.

Список литературы

- [1] Коркин Л.Р. Использование системы распознавания снимков колоний микроорганизмов на чашках Петри для оценки качества продуктов питания. 57-я Научная Конференция Аспирантов, Магистрантов и Студентов БГУИР; Минск. – 2021.
- [2] Хоулт Дж., Криг Н. Определитель бактерий Берджи. В 2-х т, 1997 год.
- [3] Тимур Казанцев. Искусственный интеллект и Машинное обучение. Основы программирования на Python. 2020.
- [4] Тонни Гэддис. Начинаем программировать на Python, 4-е издание, 2019 год.
- [5] Глубокое обучение на Python Автор: Шолле Франсуа Год: 2018 Издательство: Питер ISBN 978-5-4461-0770-4 Страниц: 400.
- [6] Применение нейронных сетей для задач классификации [Электронный ресурс]. – Режим доступа: <https://basegroup.ru/community/articles/classification> (дата обращения 08.05.2018).

PREPARING MACHINE LEARNING MATERIALS WITH A VIRTUAL STUDIO

L. KORKIN

*Assistant of the Department of Engineering
Psychology and Ergonomics of BSUIR, Master of
Engineering and Technology*

K. SAEVICH

Doctor of Biological Sciences, Professor

M. KORKINA

*Researcher of The Production
and Testing Laboratory,
Inst. for the Meat and Dairy Industry*

K. YASHIN

*PhD, Head of the Department of
Human Engineering and Ergonomics, BSUIR*

*Belarusian State University of Informatics and Radioelectronics, Republic of Belarus, Minsk
Inst. for the Meat and Dairy Industry (это сокращённое название), Republic of Belarus, Minsk*

Abstract. At present, Bergey's determinant of bacteria (1997, in 2 volumes) is used to identify microorganisms by morphological characteristics in food quality control. —

Keywords. Machine learning, artificial intelligence, microbiology, microbial colony.