

МИХАЙЛІВ АНДРІЙ

Національний університет «Львівська політехніка»

<https://orcid.org/0009-0008-6348-4002>e-mail: andrii.p.mykhailiv@lpnu.ua

ПОРІВНЯЛЬНИЙ АНАЛІЗ РОБІТ ІЗ ДОСЛІДЖЕННЯ ПОВЕДІНКИ ТВАРИН ЗА ЇХНІМИ ЗВУКАМИ

У статті порівнюються праці сучасних дослідників у галузі вивчення поведінки тварин на основі звуків, що ті видають. Дослідження вокалізації є важливим для вивчення їхньої комунікації та взаємодії, впливу географії на фонетику тварин та контекстної залежності вокалізації. Автор аналізує методи розпізнавання звуків та їхньої інтерпретації, зокрема вибирав ті наукові роботи, які використовували алгоритми штучного інтелекту. Наявні дослідження показали, що навіть для різних видів тварин можуть застосовуватися схожі методи з адаптацією під кожен вид. Вчені використовували методи класифікації на основі прихованих моделей маркова (HMM) та латентного простору, що дозволило отримати кращі результати в реальному часі. Ці дослідження підходять для вивчення птахів чи пошуку контекстів у мовленні тварин. Найкраще ця технологія підходить для бенгальських амадин, а похідні модифікації можуть застосовуватися для інших видів співочих птахів. Також вчені застосовували згорткові нейронні мережі (CNN) та мел-частотні кепстральні коефіцієнти (MFCCs) у глибокому навчанні для визначення значень конкретних звукових складів. Найчастіше це підходить для тварин, які використовують спеціальні звуки в конкретні моменти. Ссавці частіше за інших використовують вокалізацію саме так. Найкращі дослідження з технологію глибокого навчання на даний момент присвячені мові котів, адже вони часто використовують вокалізацію, як елемент спілкування між собою чи людьми. Результати цієї роботи можуть бути корисними для дослідників як зоолінгвістики, так і інших галузей, що вивчають різні види, яким притаманна складна поведінка. Адже за звуками можна отримати багато інформації про стан, самопочуття та контекст використання тваринами свого голосу. Загалом нові дослідження зосереджені на мові тварин можуть сприяти розвитку ринку товарів для них.

Ключові слова: штучний інтелект, вокалізація тварин, мова тварин, глибоке навчання, нейронні мережі.

MYKHAILIV ANDRII

Lviv Polytechnic National University

COMPARATIVE ANALYSIS OF RESEARCHES WORKS OF ANIMALS BEHAVIOR BY THEIR SOUNDS

The article compares the works of contemporary researchers in the field of studying animal behaviour based on the sounds they emit. Research on vocalization is important to study their communication and interaction, the impact of geography on animal phonetics, and the contextual dependence of vocalization. The author analyses methods of sound recognition and its interpretation, particularly selecting those scientific works that utilized artificial intelligence algorithms. Existing researches have shown that similar methods apply to different animal species but require adaptations for each. Scientists used classification methods based on hidden Markov models (HMM) and latent space, which allowed to get better real-time results. These researches are suitable for studying birds or searching for contexts in animal speech. The technology is best for Bengal finches, and derived modifications apply to other species of songbirds. Scientists also used convolutional neural networks (CNNs) and Mel-frequency cepstral coefficients (MFCCs) in deep learning to identify the meanings of specific sound syllables. This is most suitable for animals that use specific sounds at particular moments. Mammals use vocalization in this way more often than others. The best research on deep learning technology at the moment is dedicated to cat language, as they often use vocalization as an element of communication among themselves or with humans. The results of this work can be useful for researchers in zoolinguistics and other fields that study various species with complex behaviour. After all, sounds can provide a lot of information about the state, well-being, and context in which animals use their voice. Overall, new research focused on animal language can contribute to the development of the market for animal products.

Keywords: artificial intelligence, animal vocalization, animal language, deep learning, neural networks.

Вступ

Люди, які доглядають чи розводять тварин, часто спостерігають, що вони, як і люди видають різні звуки з метою повідомити щось або реагують на подразник таким чином, щоб його можна було почути. Як наслідок, вивчення вокалізації різних видів тварин може допомогти у розв'язанні проблем з їхнього вивчення, утримання, розведення тощо. Загалом, знаходження способу чи способів для отримання інформації із мовлення різних видів принесли б вигоду як і людям, так і фауні.

Основним завданням даної роботи є знаходження і порівняння досліджень та методів з розпізнавання звуків різних тварин та їхньої інтерпретації. Було проведено пошук робіт із вокалізації тварин, особливо досліджень, що застосовують алгоритми штучного інтелекту у своїй роботі.

Аналіз робіт попередників, їх переваг та недоліків, визначає напрямки розвитку зоолінгвістики, її успіхи та перспективи. Більше того, сучасні дослідники вважають, що за пару років люди зможуть повністю перекласти мовлення деяких тварин [1], а саме: коти, собаки, ворони, співочі птахи, кити та кажани. Також найчастіше вчені використовують методи глибокого навчання, але для його ефективної роботи також потрібні сучасні методи класифікації, маркування та очищення звуків від шумів.

Висновки порівняння дозволять практичні методи для вирішення проблем з утриманням та розведенням тварин. Власники домашніх улюбленців частіше реагуватимуть на проблеми хатніх тварин, а

фермери зможуть запобігати хворобам чи травмам худоби. Також дослідники з інших галузей зможуть використовувати аналіз звуків диких тварин для вивчення біосфери.

Матеріали і методи вивчення вокалізації тварин

1. Спроби вивчення тварин на основі звуків універсальними методами. Одне з новіших каліфорнійських досліджень [2] шукає універсальний метод для вивчення звуків тварин, що користуються голосом. Використання латентної проєкції підходить для аналізу часових патернів у контексті, щоб класифікувати звуки.

Принцип такої проєкції полягає в перетворенні високорозмірних даних в меншу кількість вимірів, зберігаючи їхню структуру та варіативність. Кожна точка у такому просторі може бути відображена у низьковимірному, де виміри латентного простору можуть розглядатися як ознаки набору даних. Це важливо, оскільки звукові сигнали тварин мають складну періодичну та спектральну структуру, яку можна описати меншою кількістю динамічних вимірів.

Вчені проаналізували 19 наборів даних, що складаються з 400 годин вокалізацій та понад 3 000 000 окремих одиниць від 29 унікальних видів. Кожний набір даних був розділений на окремі одиниці (наприклад, склади чи ноти), або на основі меж сегментації, наданих набором даних (де це було можливо), або за допомогою алгоритму сегментації з динамічним порогом, який розділяє склади вокалізацій між виявленими паузами у вокальному потоці. З кожною одиницею дослідники обчислювали спектрографічне представлення даних.

Дослідники використовували два типи для даних – спектрограми та перцептивні акустичні ознаки (PAF) та тестували два методи вибору наборів ознак оптимізації даних: T-розподіленого вкладення стохастичної близькості (t-SNE) та рівномірну апроксимацію та проєкцію різноманіття (UMAP). Перший алгоритм спочатку створює розподіл імовірностей по парах багатовимірних об'єктів таким чином, що подібні об'єкти мають високу ймовірність бути вибраними, а далі визначає подібний розподіл імовірностей для точок у карті низьковимірному простору та мінімізує їхні розбіжності. Інший алгоритм для зниження розмірності спочатку виконує побудову зваженого графа, з'єднуючи ребрами лише об'єкти, які є найближчими сусідами, а потім алгоритм створює граф у низьковимірному просторі та наближає його до вихідного, мінімізуючи суму для кожного ребра з множин. Результатом порівняння цих алгоритмів став вибір алгоритму UMAP, адже була доведена його статистична значущість і надійність даних.

Було виявлено, що залежно від подачі набору даних, проєктованих в латентний простір, їхні ознаки можуть враховувати біологічно або психологічно значущі фактори. Відповідно, вчені використовували записи тварин на самоті, з кимось, у популяції та з урахуванням філогенезу. Сформовані латентні проєкції охоплюють ряд складних ознак, включаючи ідентифікацію окремих осіб, ідентифікацію видів, мовні ознаки, категорії складів та вплив географії. Проєкції показують, що ідентичність особини може бути визначена за всіма цими ознаками. Також метод показує якими видами краще розрізняти кликання іншої конкретної особи, як-от макаки показують більшу ймовірність відгукнутися на клич іншої особини ніж у мавп мармозеток.

Одним з видів для аналізу, що мав добре описану вокалізацію нот, був болотний горобець. Дослідження включало проєктування спектрограм у простір UMAP та порівняння їх з відомими ознаками. Для кластеризації використовувалися методи, такі як HDBSCAN та модель змішаних гаусівських розподілів (GMM). Результати показали перетин між методами кластеризації, що вказує на схожі стилі представлення.

Шляхом проєктування складів у латентний простір, дослідження порівнювало мітки, промарковані вручну та алгоритмом для наборів даних пісень бенгальських амадин. Аналіз показав, що різні мітки HDBSCAN в межах однієї категорії промаркованої вручну можуть мати відмінні ролі в синтаксичній структурі пісень, демонструючи складність послідовної організації, яку зафіксовано алгоритмічно. Дослідження також підкреслює відмінності в послідовній організації, підкреслюючи важливість використання обидвох методів для розуміння синтаксису пісень.

На жаль не всі звуки виявляють дискретні кластери в латентному просторі. У деяких випадках вокалізації важко розділити за тривалістю. Тому треба уважно обирати ознаки для сегментації.

Постійні траєкторії в латентному просторі дозволяють зафіксувати короткочасні та довгочасні структури звуків без необхідності сегментації або позначення вокальних елементів. Така обробка дозволяє порівняти репрезентації з тими, що обробляють вокальні сегменти як окремі точки. Цей підхід дозволяє глибше розуміти вплив тимчасової дискретизації та сегментації категоріями на аналіз звуків.

Загалом автори дослідження показали, що складні ознаки вокальних сигналів, такі як ідентифікація особи, ідентифікація виду, вплив географії на популяцію, фонетика та кластери на основі схожості, можуть бути зафіксовані у латентних просторах представлення. Також показали, що птахи-співачі мають тенденцію виробляти сигнали, які дискретно кластеризуються у латентному просторі, тоді як вокалізації ссавців рівномірніше розподілені.

2. Вивчення птахів на основі їхнього співу. Одне з перших відповідних досліджень [3] використовувало спеціальний записувальний механізм для вивчення поведінки птахів на основі вокалізації та оцінювало ефективність методу у майбутньому. Шляхом автоматичного визначення звуків, пов'язаних з птахами, дослідники вивчали не лише поведінку, але й різні аспекти середовища.

Вчені вивчали контекст вокалізації у реальному часі з урахуванням поточної діяльності диких галок. Підтвердження звукових даних відбувалося через відео записи, промарковані людьми. Експериментально

порівнювали дві різні парадигми визначення звуку: класифікація та виявлення подій, а також комбінації та варіанти можливого визначення різних категорій поведінки птахів.

Для збору даних використали 7 особин, що трималися в неволі, та 6 на волі. Птахів ідентифікували індивідуально за допомогою відео та кольорових кілець. Додаткові мітки були додані для типової поведінки птахів на волі, таких як процес розмноження та крики пташині в середині гнізда.

Частота дискретизації запису балу 22050 Гц у форматі WAV без стиснення. Використано два методи класифікації – класифікатора звуків на основі НММ (прихованої моделі Маркова) і латентний простір. У першому випадку вчені сегментували вхідне аудіо на п'яти секундні кліпи, обчислювали мел-спектрограми та застосовували медіанне зменшення шумів. В іншому створювали систему виявлення подій на основі латентного аналізу компонент, що розкладав аудіоспектрограми на звукові активності або контексти. Крім того, у дослідженні розглядалася різна обробка відсутніх даних залежно від методу класифікації.

Загалом система на основі класифікатора показувала результати, що краще відповідають спостереженням в реальному часі. З іншого боку, латентний простір мав перевагу у спостереженнях короткочасних вокальних взаємодій. Також для розвитку результатів дослідження потрібна більша кількість підтверджених даних.

У наступній статті [4] намагаються знайти альтернативний підхід у роботі з птахами та досліджують контекстну залежність у їхніх піснях за допомогою нейронних мереж. Вплив минулих подій на поведінку будь-якої тварини є важливим напрямком в біології. Стаття вказує на те, що пісні птахів є репрезентативними для вивчення контекстної залежності в послідовних сигналах, так само як і людська мова. Дослідження оцінює контекст у піснях бенгальських та зебрових амадин. Ці птахи були обрані, бо вони утворювати складні конструкції зі складів.

Для аналізу було використано нейромережеву мовну модель на основі архітектури Transformer, яка була навчена на 7 779 послідовностях звукових складів. Результати показали, що птахи можуть використовувати до восьми попередньо вимовлених складів під час співу, що свідчить про тривалу контекстну залежність. Для класифікації складів було використано метод кластеризації на основі ABCD-VAE, що ефективно аналізувало контекст у послідовностях складів. Дослідження вказує на можливість застосування подібних методів для аналізу контекстної залежності в інших послідовних даних, таких як танці чи жести. Оцінка кластеризації складів у дослідженні проводилася за допомогою двох метрик: коефіцієнта Каппи Коена й однорідності. Перша є показником рівня узгодження між двома класифікаціями, тоді як інша оцінює, наскільки добре передбачені кластери відповідають класам істинності.

Кластери складів у вигляді дискретних символів аналізує нейромережева мовна модель для оцінки ймовірності послідовностей залежно від індивідуальної інформації. Ефективна довжина контексту (ECL) моделі визначена шляхом оцінки того, наскільки ефективно використовувався контекст для передбачення наступних складів. Використовувався бутстрепінг для отримання більш надійної оцінки статистично ефективної довжини контексту (SECL). SECL обчислювалася шляхом оцінки того самого набору складів при різних довжинах обрізаних контекстів.

Результати дослідження підкреслюють важливість метрик кластеризації, а саме їхню однорідність та примусове вирівнювання даних, для оцінки передбачених моделлю категорій, а також ефективність методу ABCD-VAE для кластеризації та нормалізації мовця в аналізі пісень птахів. Проте, вчені також наголосили, що метод може програвати у точності іншим, якщо вони будуть створені окремо під конкретно обраний вид тварин. Наприклад, вдалося виділити категорії бенгальських амадин, але для категорій зебрових амадин виявили значну неточність у результатах.

Також результати дослідження вказують на те, що об'єднані склади з кількома елементами ускладнюють пошук категорій, але це залежить від виду. Дослідження підкреслило важливість контексту у передбаченні майбутніх складів. Автори впевнені, що нейронні мовні моделі будуть універсальною парадигмою для розуміння впливу минулих дій птахів на майбутнє прийняття рішень.

Отже, вивчаючи птахів, вчені орієнтуються на контекст їхньої поведінки, оскільки багато видів здатні створювати складні звукові конструкції зі складів. Але треба враховувати мету вивчення вокалізації птахів, щоб обрати правильні методи дослідження.

3. Вивчення поведінки сільськогосподарських тварин. Найчастіше сільському господарству треба розуміти тварину для економії витрат на утримання та збереження здоров'я. Наприклад, метою дослідження [5] є розробка ефективною системи класифікації звуків курей-несучок і худоби для їхнього моніторингу за допомогою згорткових нейронних мереж (CNN).

Комунікації між групами тварин або окремими особинами надають інформацію про різні ознаки: вік, стать, соціальна ієрархія та самопочуття. Моніторинг вокалізації може повідомити про благополуччя тварин, відчуття болю, статеву зрілість, бажання відокремлення від потомства, спрагу тощо. Завдяки методам глибокого навчання фермерам стало вигідно визначати поведінку тварин за звуками на основі CNN і мел-частотних кепстральних коефіцієнтів (MFCCs). У дослідженні використали WAV аудіозаписи з бітовою швидкістю 16 біт та частотою дискретизації 44,1 кГц. Також було проведено відеозапис для визначення контексту звуків.

Для класифікації мовлення використали дві моделі CNN – на основі 2D ConVnet (C2D) та 1D ConVnet (C1D) з мережею довготривалої короткочасної пам'яті (LSTM). Було виявлено, що модель C1D з LSTM показала кращу продуктивність через її здатність зафіксувати зміни з часом у індивідуальних звуках. Але

обидві моделі залежали від наявності достатньої кількості тренувальних даних для всіх категорій, і відзначено необхідність методів аугментації даних у глибокому навчанні для ситуацій, коли важко отримати голосові дані, пов'язані з конкретною поведінкою.

Наступне дослідження [6] також виявило, що вокалізація худоби містить інформацію, пов'язану з їхнім добробутом та поведінкою.

Системи Precision Livestock Farming (PLF) використовуються фермерами для постійного моніторингу тварин. Ці системи містять різноманітні сенсори: мікрофони, камери, акселерометри та GPS, для постійного збору даних. Моделі машинного навчання навчаються на цих даних для передбачення станів тварини. В даному випадку PLF спрямовані на автоматизоване виявлення та класифікацію подій на основі звуків. Виділені ознаки на основі MFCCs та дискретного вейвлетного перетворення (DWT), були запропоновані для аналізу вокалізацій. Метод екстракції ознак на основі DWT показав обчислювальні переваги порівняно з MFCCs: є швидшим та більш послідовним, що важливо для систем з обмеженими ресурсами.

Попри те, що набір даних мав високий рівень шуму та низькочастотні компоненти, що могли вплинути на результати класифікації, алгоритм продемонстрував стійкість до шуму. В результаті дане дослідження презентувало багатоцільовий алгоритм класифікації вокалізацій тварин на фермах, використовуючи бінарну модель C-SVM для класифікації. Алгоритм продемонстрував високу точність у виявленні вокалізацій овець, худоби та собак-охоронців на відкритих фермерських майданчиках.

Удосконалення [7] попереднього дослідження займається фільтрацією фонового шуму в заганах із худобою за допомогою короткочасного перетворення Фур'є (STFT), що значно покращує точність систем моніторингу звуків. Очищення даних покращило точність застосованих методів глибокого навчання майже на 3%. Категоризовані звуки від худоби були розділені на чотири класи, що призвело до кінцевої точності моделі на рівні 81,96%.

Це дослідження показало успішність моделей глибокого навчання у класифікації звуків тварин. Також підкреслено потенціал систем на основі поєднання CNN для покращення результатів класифікації.

Загалом можна стверджувати, що зараз методи дослідження поведінки сільськогосподарських тварин за їхніми звуками базується на проектах із застосуванням глибокого навчання. Удосконалення цих методів або якісніша подача звукових даних є головною метою для науковців у цій галузі.

4. Вивчення спілкування домашніх котів. Одне з перших пілотних досліджень [8] присвячене комунікації котів з людиною та показує, що основна частота спілкування (позначено f_0) слугує корелятом паралінгвістичної інформації та біологічних кодів у міжвидовій комунікації між людьми та домашніми котами. В дослідженні використовували записи з попередніх робіт з аналізу котячого мовлення. Через порівняння показників f_0 між котами та в міжвидовій комунікації для обох видів було досліджено діапазон f_0 та стандартне відхилення. Використана паралінгвістична інформація в мовленні включає його варіації через такі фактори, як спікер, аудиторія та фізичне середовище, які можна класифікувати на різні типи якості. Експресивна та органічна інформація в мовленні визначена як біологічні коди, що включають варіації у f_0 та якості голосу, де висока f_0 асоціюється з "малими" значеннями, а низька f_0 – з "великими" значеннями.

Дослідження показало, що у котів є різноманітний вокальний репертуар для вираження свого психічного стану, потреб і намірів. Незважаючи на дослідження емоційних та контекстуальних підказок у вокальних взаємодіях між людьми та котами, залишається невідомим, наскільки коти та люди можуть ідентифікувати біологічні коди та паралінгвістичну інформацію у міжвидовій розмові.

Загалом результати показали, що диктори - жінки і чоловіки - мали вищі значення f_0 у мовленні до котів, ніж до людини. Також всі коти мали вище середнє значення f_0 у вокалізаціях, спрямованих на людину, але вищий діапазон f_0 та стандартне відхилення у спілкування з котами, ніж з людиною. Це свідчить про те, що власники кішок використовують паралінгвістичну інформацію та частотні коди для демонстрації ніжності, щастя та єдності зі своїми котами. Однак досі невідомо, чи сприймають кішки це за важливу інформацію, але вони використовують вищий f_0 з людьми, ніж між собою – це свідчить про свідоме підвищення тональності в міжвидовій взаємодії.

Наступне дослідження [9] іде далі та вивчає комунікацію для спроби перекладу емоційного стану та контексту котячого нявкоту. Були записані голосові сигнали представників порід мейн-кун та європейська короткошерста в трьох різних контекстах: очікування їжі, ізоляція в незнайомому середовищі та вичісування. Загалом вчені отримали 448 аудіозаписів. Шляхом вилучення акустичних параметрів через MFCCs та характеристик темпоральної модуляції, дослідники мали на меті вивчити контекст використання звуків за допомогою класифікаційної схеми на основі направленої ациклічного графа (DAG). Проведені експерименти підтвердили ефективність цього підходу порівняно з іншими рішеннями класифікації.

Технологію DAG-HMM використали для розбиття проблеми класифікації на серію завдань класифікації двох класів. Множина S_m для методу містила можливі комбінації двох вузлів, де кожен відповідав певному завданню класифікації, з дугами, що мали один напрямок і відсутні петлі. Топологічне упорядкування було важливим для продуктивності, всі завдання були впорядковані на основі рівня класифікації для пріоритетного виконання простіших перед складнішими.

Порівняння запропонованої структури зі звичайними системами класифікації звуків підтвердило, що MFCCs та часові модуляційні ознаки захоплюють відмінні властивості аудіосигналів, що призводить до різноманітних помилкових класифікацій, тому їх використовували одночасно.

В наступному дослідженні [10], що базується на вище згаданих, науковці поставили собі за мету

узагальнити різноманітні типи вокалізацій домашніх кішок, вивчити контекст використання та вплив генетики і середовища на ці вокалізації. Вчені все ще закликають до проведення більшої кількості досліджень у цій області для поліпшення розуміння та покращення добробуту кішок. У роботі вони детально навели 20 типів звуків, які, на їхню думку, є найбільш відмінними у мовленні котів.

В підсумку вчені отримали 5 різних дружніх категорій: Бурмотіння (Murmur), Вібрації (Trills), Клекотіння (Gurgle), Муркотіння (Purr) і Писк (Squeak); 6 агоністичних: Бурчання (Snarl), Виття (Howl), Гарчання (Growl), Стогін (Moan), Харчання (Spit) і Шипіння (Hiss); 4 способи сказати хоча: Пілікання (Tweedle), Щебетання (Tweet), Балакання (Chatter), цей звук спрямований на здобич, яку тварина не може вполювати, і Стрекотання (Chirp), який використовується для пошуку іншого kota та якому самиці навчають кошенят. Наступні звуки домашні коти застосовують як при різних контекстах, так і при розмноженні: Котячий «концерт» (Caterwaul or mowl) коцури використовують для приваблення кішок, а ті натомість для попередження; Ричання (Yowl) є і легким попередженням і звуком під час спарювання; а також самки можуть видавати особливий Крик (Copulatory cry) під час спарювання, як реакцію на статевий акт; звук Болісного вереску (Pain shriek), що трапляється у розпалі бою; а стандартне Нявкання (Meow) є найбільш багатофункціональною вокалізацією, фонетика якого може розшифруватися детальніше.

Найновіше дослідження [11] у цій темі присвячене класифікації мови котів на основі Wav2Vec2.0. Це модель розпізнавання мовлення, запропонована дослідниками з Facebook AI Research.

Дослідники наголошують, що пристрої для перекладу звуків тварин мають обмежену точність, натомість кращий аналіз поведінки тварин через аудіо може сприяти розвитку ринку товарів для домашніх тварин. Автори ставлять за мету покращення підходу з [7] за допомогою глибокого навчання для побудови кращих моделей. Використовується попередньо навчена модель для вилучення аудіо ознак зі звуків котів, використовуючи власний кодер моделі Transformer з механізмом самоуваги для кращого захоплення довгострокових залежностей та покращення здатності узагальнення.

Модель для попереднього тренування складається з CNN для видобування ознак з аудіосигналу, модуля квантизації для перетворення неперервних ознак у дискретні, кодера Transformer для обробки квантованих ознак, та допоміжних функцій.

Сам Transformer детально описаний, він створює вектор ознак, що є ключовим для подальшого аналізу та обробки вхідних даних. Під час кодування цей вектор проходить через різні шари мережі, включаючи механізм багаторазової уваги та нейронну мережу зворотного зв'язку, для отримання високоякісного представлення даних.

Дослідження пропонує модель класифікації звуків вокальної поведінки тварин з нуля під назвою W2VCM через поєднання моделі Wav2Vec2.0 з набором даних CatMeows. Автори стверджують, що змогли на порівнянні із C1D, C2D, LSTM, DAG-HMM й іншими алгоритмами довести переваги їхнього результату, але зауважують, що вони страждають від нестачі даних.

Отже, роботи з вивчення домашніх котів зазвичай застосовували один метод. Уважно розглянувши доповнення старих досліджень новими, можна сказати, що вчені довели наявність котячої фонетики, яку можна детально вивчати та перекладати.

Висновки

Пророблений порівняльний аналіз досліджень про поведінку тварин за звуками підкреслює важливість використання сучасних моделей та методів класифікації для розуміння поведінки тварин на основі вокалізації. Такі дослідження грають важливу роль у догляді та вивченні різних тварин, показують їхню взаємодію із людиною та вплив факторів життя на мовлення будь-якого виду.

Першочергово такі дослідження дали зрозуміти важливість наявності достатньої кількості тренувальних даних для всіх категорій, підкреслюючи це як ключовий фактор успішної класифікації.

Також порівняння показало, які алгоритми та коли краще застосовувати. Наприклад, використання моделей CNN на основі MFCCs для класифікації мовлення у великої рогатої худоби та курей показало, що модель LSTM має кращу продуктивність завдяки здатності фіксувати зміни у звуках з часом. На основі цих алгоритмів були створені моделі для розпізнавання мовлення інших видів ссавців, до прикладу котів.

Результатом цього дослідження є демонстрація того, що навіть для різних видів тварин зазвичай потрібні однакові методи, але з адаптацією під кожен конкретний вид. Сам факт наявності такої великої кількості робіт за останні роки демонструє зацікавленість суспільства у розвитку зоолінгвістики. А головне, це має як і наукові, так і економічні перспективи розвитку, тому дослідження на цю тему будуть актуальні ще не один рік.

Дискусія та подальші дослідження

Дана стаття надає цінне порівняння досліджень з поведінки тварин. Один з головних плюсів проробленої роботи полягає в підкресленні важливості використання великих баз даних для розуміння поведінки на основі вокалізації. Такі роботи грають велику роль у догляді та вивченні тварин у інших сферах наук присвячених фауні.

Проте, сам факт обмеженої кількості тренувальних даних для деяких категорій звуків впливає на точність класифікації у дослідженнях, що ставить під сумнів цінність цих досліджень у майбутньому. Для нових робіт у цій області важливо розвивати методи глибокого навчання та поєднання їх з іншими підходами

для досягнення більш точних результатів. Стратегія комбінування методів та виокремлення ключових факторів для кожного окремого виду показує себе найбільш перспективною.

Загалом, порівняльний аналіз робіт із дослідження поведінки тварин за звуками є надзвичайно актуальним у наш час і відкриває багато шляхів для розвитку науки.

References

1. Cat got your tongue? How AI could be on cusp of breakthrough that'd allow people and ANIMALS to talk to each other in «12 to 36 months». Daily Mail Online. <https://www.dailymail.co.uk/sciencetech/article-13113685/Cat-got-tongue-AI-cusp-breakthrough-thatd-allow-people-ANIMALS-talk-12-36-months.html>.
2. T. Sainburg, M. Thielk, i T. Q. Gentner Finding, visualizing, and quantifying latent structure across diverse animal vocal repertoires. *PLOS Comput. Biol.* 16, №10, 2020.
3. D. Stowell, E. Benetos, i L. F. Gill On-Bird Sound Recordings: Automatic Acoustic Recognition of Activities and Contexts. *IEEE/ACM Trans. Audio Speech Lang. Process.* 25, №6, 2017. P. 1193–1206.
4. T. Morita, H. Koda, K. Okanoya, i R. O. Tachibana Measuring context dependency in birdsong using artificial neural networks. *PLOS Comput. Biol.* 17, № 12, 2021.
5. D.-H. Jung, N. Y. Kim, S. H. Moon, H. S. Kim, T. S. Lee, J.-S. Yang, J. Y. Lee, X. Han, i S. H. Park Classification of Vocalization Recordings of Laying Hens and Cattle Using Convolutional Neural Network Models. *J. Biosyst. Eng.* 46, №3, 2021. P. 217–224.
6. J. C. Bishop, G. Falzon, M. Trotter, P. Kwan, i P. D. Meek Livestock vocalisation classification in farm soundscapes. *Comput. Electron. Agric.* 162, 2019. P. 531–542.
7. D.-H. Jung, N. Y. Kim, S. H. Moon, C. Jhin, H.-J. Kim, J.-S. Yang, H. S. Kim, T. S. Lee, J. Y. Lee, i S. H. Park Deep Learning-Based Cattle Vocal Classification Model and Real-Time Livestock Monitoring System with Noise Filterin. *Animals.* 11, №2, 2021. P. 357.
8. S. Schötz Paralinguistic information and biological codes in intra- and interspecific vocal communication: A pilot study of humans and domestic cats. 2019.
9. L. A. Ntalampiras, S. Ludovico, G. Presti, E. P. Prato Previde, M. Battini, S. Cannas, C. Palestrini, i S. Mattiello Automatic Classification of Cat Vocalizations Emitted in Different Contexts. *Animals.* 9, №8, 2019. P. 543.
10. C. Tavernier, S. Ahmed, K. A. Houpt, i S. C. Yeon Feline vocal communication. 21, №1, 2020.
11. Z. Kuang, S. Jiang, H. Huang, Y. Liu, i X. Li Research on end-to-end animal behavior speech classification based on Wav2Vec2.0. in 3rd International Conference on Artificial Intelligence, Automation, and High-Performance Computing (AIAHPC 2023), Wuhan, China. 2023. P. 44.