

РАЗРАБОТКА АВТОМАТИЧЕСКОГО ИДЕНТИФИКАТОРА НА ОСНОВЕ МИКРОСТРУКТУРНЫХ ДАННЫХ ПЕРА ПТИЦЫ

Силаева О. Л., Вараксин А. Н.

Институт проблем экологии и эволюции им. А. Н. Северцова, Москва,

silaeva.o@gmail.com;

Научно-исследовательский центр распознавания образов, Москва,

prr@mail.ru

Введение

Исследования морфологических особенностей пера птицы востребованы множеством практических и теоретических сфер деятельности человека, непосредственно связанных с орнитологией, а также удалённых от неё. В первую очередь результаты практической птерилогграфии могут быть использованы в авиационной орнитологии. Определение таксономической и популяционной принадлежности птицы, а также её возраста и пола необходимы для анализа столкновений воздушных судов с птицами и рекомендаций по их предотвращению в дальнейшем. Создание автоматического определителя обусловлено необходимостью идентификации птицы по структуре пера.

Эти данные помогают выяснить орнитологическую обстановку на аэродроме или на пути следования воздушного судна, а также параметры риска, с которыми может быть связано наибольшее число столкновений: сезон, время суток, высота полета, траектория движения. Данная информация поможет уточнить список опасных для авиации видов с учётом данного района, а значит, будет способствовать правильному выбору и/или разработке местных средств управления поведением птиц. Сведения о видовой принадлежности птицы и количестве особей, столкнувшихся с воздушным судном, позволят конструировать защитные устройства для двигателей. Известно, что большинство тяжёлых лётных происшествий связано с попаданием птиц в двигатели.

Важно также знать, произошло столкновение с местным или с мигрирующим видом. Подробные данные о птице нужны ещё и для того, чтобы страховые компании смогли ответить на вопрос, на чьей территории произошло столкновение.

В авиации США и Европы уже в течение десятилетий действуют хорошо налаженные информационные системы, оперирующие базами данных по структуре пера птицы (Prast et al., 1992, 1996, 1998; Shamoun, Yom-Tov, 1995, 1996; Busching, 1997; Wattel, 2000). Все останки биологического происхождения направляются на исследование, которым занимаются профессиональные эксперты, так как методика определения вида, участвовавшего в столкновении, коренным образом отличается от определения того же вида в природе. Имеются подробные списки видов птиц, опасных для воздушных судов в определённом районе. Ни одно лётное происшествие с участием птиц независимо от его последствий не остаётся неисследованным.

Для авиационных служащих подготовлены информационные письма, популярные статьи и брошюры с инструкциями о том, как собрать материал, как правильно его упаковать, при необходимости сохранить и куда его послать на экспертизу. Распространяются анкеты, которые предлагается заполнить и приложить к отсылаемым на экспертизу фрагментам птицы. В анкете запрашиваются все необходимые данные по лётному происшествию. Эти материалы выложены на постоянно обновляющихся сайтах Интернета: http://www.faa.gov/airports_airtraffic/airports/airport_safety. Сопроводительную анкету можно заполнить виртуально и отправить в экспертный центр по электронной почте (Dove, 2002).

Назрела насущная необходимость организации аналогичной системы в России и создания базы данных по фрагментам птиц, участвовавших в столкновении. Мы надеемся, что наш автоматический определитель будет частью такой системы.

Материал и метод

В данной работе для изучения морфометрических характеристик микроструктуры пера используются кластерный анализ, метод главных компонент, а также дисперсионный анализ. Предлагается кардинально новый метод идентификации — разработка электронного определителя (классификатора) нового поколения с автоматической идентификацией таксона птицы. Методика разрабатывается совместно сотрудниками Лаборатории экологии и управления поведением птиц ИПЭЭ РАН и специалистами Научно-исследовательского центра распознавания образов. Впервые для изучения перьевого материала были применены методы анализа многомерных данных (Силаева, Вараксин, Ильичёв, 2010).

Для выявления межгруппового и внутригруппового рассеивания объектов была создана микроструктурная выборка, включившая в себя признаки 19 видов (более 120 особей), принадлежащих отрядам Воробьинообразных, Курообразных и Голубеобразных. Типичный пуховой луч¹ из базальной частей опахал покровного пера птицы, описали набором цифровых признаков. Были выбраны следующие 6 промеров-признаков:

1. сузл — среднее количество узлов (плотность) на 1 мм луча;
2. lмежд — длина междоузлия, мк;
3. wмежд — ширина междоузлия, мк;
4. lузл — длина узла, мк;
5. wузл — ширина узла, мк;
6. lлуча — длина луча, мм.

От каждой особи было взято по одному перу. Лучи для измерений выбирали произвольно из медиальных участков двух–трёх боронок.

Проводили по 10 измерений каждого признака каждой особи (рис. 1).

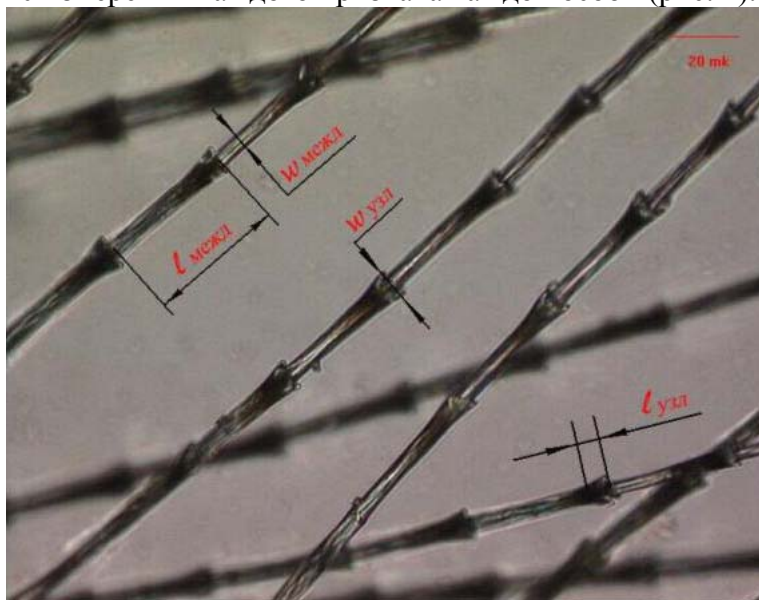


Рис. 1. Промеры пуховых лучей ворона

¹ Пуховой луч — бородка второго порядка пуховой части опахала, которая состоит из основания и длинного узкого опахальца.

Длину луча получали в результате суммирования длин отрезков ломаной линии, наложенных на исследуемый луч. Длину междоузлия измеряли вместе с одним узлом, т. е. измерялось расстояние между дистальными зубцами соседних узлов. Длину узла измеряли от начала расширения междоузлия до вершины среднего зубца (рис. 1).

Результаты и обсуждение

Две трети признаков-промеров были выбраны произвольно и лишь одна треть на основании литературных данных. В работах коллектива авторов — создателей ключевых идентификаторов, а также электронного определителя BRIS (Brom, 1992, Prast, Shamoun, Bierhuizen et al., 1996) наиболее информативными считаются плотность узлов (у нас это 1-ый признак) и длина луча (у нас это 6-ой признак). Кроме этих признаков мы выбрали ещё четыре перечисленные выше признака, которые в упомянутых литературных источниках почти не используются.

После выбора признаков и формирования базы данных появилась необходимость оценки информативности признаков. Информативность признака — это его способность принимать одни значения на данных, принадлежащих одному классу объектов, и совершенно другие на данных другого класса. Формальная процедура оценки информативности базируется на однофакторном дисперсионном анализе и имеет следующий вид:

$$I_1 = \frac{Mr_1}{Vr_1}, \quad (1)$$

где I_1 — информативность 1-го признака, Mr_1 — межгрупповое рассеивание и Vr_1 — внутригрупповое рассеивание.

Информативность всех шести признаков, вычисленная по формуле (1), приведена на рисунке 2.

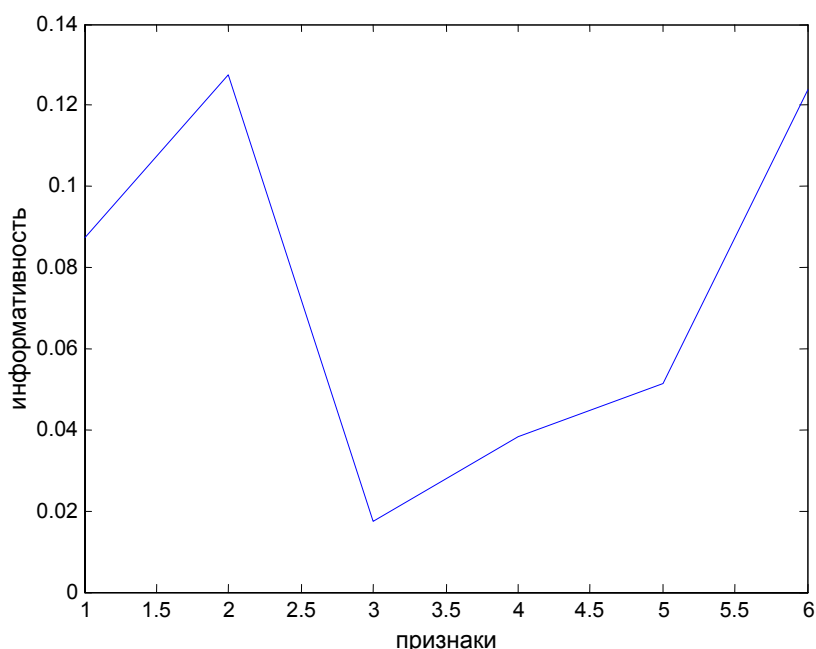


Рис. 2. Информативность признаков

Полученные оценки информативности признаков показали, что все признаки обладают сравнительно низкой информативностью. Вместе с тем, наиболее информативными оказались 2-ой и 6-ой признаки. А первый признак, широко используемый в

традиционных идентификаторах и определителях, является далеко не самым информативным. Очевидно, что при такой низкой информативности признаков создание классификатора на основе хорошо отработанных Байесовских методов не представляется возможным. Наиболее правильным был бы подход, основанный на синтезе сложного признака, куда входили бы имеющиеся признаки пропорционально их информативности.

Исследуем вариативность 1-го признака и его статистические характеристики, в качестве которых рассмотрено математическое ожидание 1-го признака, а также математические ожидания на интервалах видов (табл. 1А). Большинство видов хорошо отделяются друг от друга, их средние значения находятся на некотором расстоянии от линии математического ожидания и друг от друга. В данном случае математическое ожидание на интервале вида может быть интерпретировано как некоторое расстояние центроида от линии математического ожидания всего признака. Центроиды обыкновенной чечётки и обыкновенного поползня очень близки друг к другу, а центроиды родственных видов певчего и чёрного дроздов являются, кроме того, центроидами всего кластера сформированного первым признаком (табл. 1А).

Аналогично были сделаны оценки вариативности остальных признаков. На линии математического ожидания второго признака лежат два вида — серая ворона и дрозд-рябинник (табл. 1Б). При оценке вариативности 3-го признака обнаружили три вида, а именно, грач, дрозд-рябинник и сизый голубь, центроиды которых являются центрами тяжести всего кластера, сформированного третьим признаком. Кроме того, центроиды певчего и чёрного дроздов близки между собой (табл. 1В). Вариативность 4-го признака: центроиды двух близких видов сойки и кедровки не различимы между собой, а центроид рябинника сливается с системой (табл. 1Г). Вариативность 5-го признака: снова очень близки между собой центроиды обыкновенной чечётки и обыкновенного поползня. Близкие виды (серая куропатка и рябчик) плохо различимы между собой (табл. 1Д). Вариативность 6-го признака: плохо различаются между собой центроиды домового воробья, обыкновенной чечётки и обыкновенного поползня. На линии математического ожидания находятся центроиды ворона и чёрного дрозда (табл. 1Е).

В результате получаем несколько «видов-близнецов». Обыкновенная чечётка и обыкновенный поползень различимы по трём признакам, описывающим длину и ширину междоузлия, а также длину узла. Певчий и чёрный дрозды различимы также по трём признакам — длине и ширине узла, а также длине междоузлия. Остальные «виды-близнецы» оказываются различимыми по пяти признакам.

Весь комплекс признаков-промеров можно определить как систему, где каждый элемент дополняет другие элементы. Системный характер признаков позволяет распознать и такие сложные для идентификации «виды-близнецы» как обыкновенная чечётка и обыкновенный поползень. Всего отмечено 13 трудно различимых видов. Пять видов не сливаются между собой ни в одном признаке и не являются центрами тяжести кластеров, сформированных признаками, т. е. они различимы по всем признакам и по каждому в отдельности. На основании данных по вариативности каждого признака-промера можно сделать вывод относительно информативности соответствующего признака для каждого вида, имеющегося в информативной выборке.

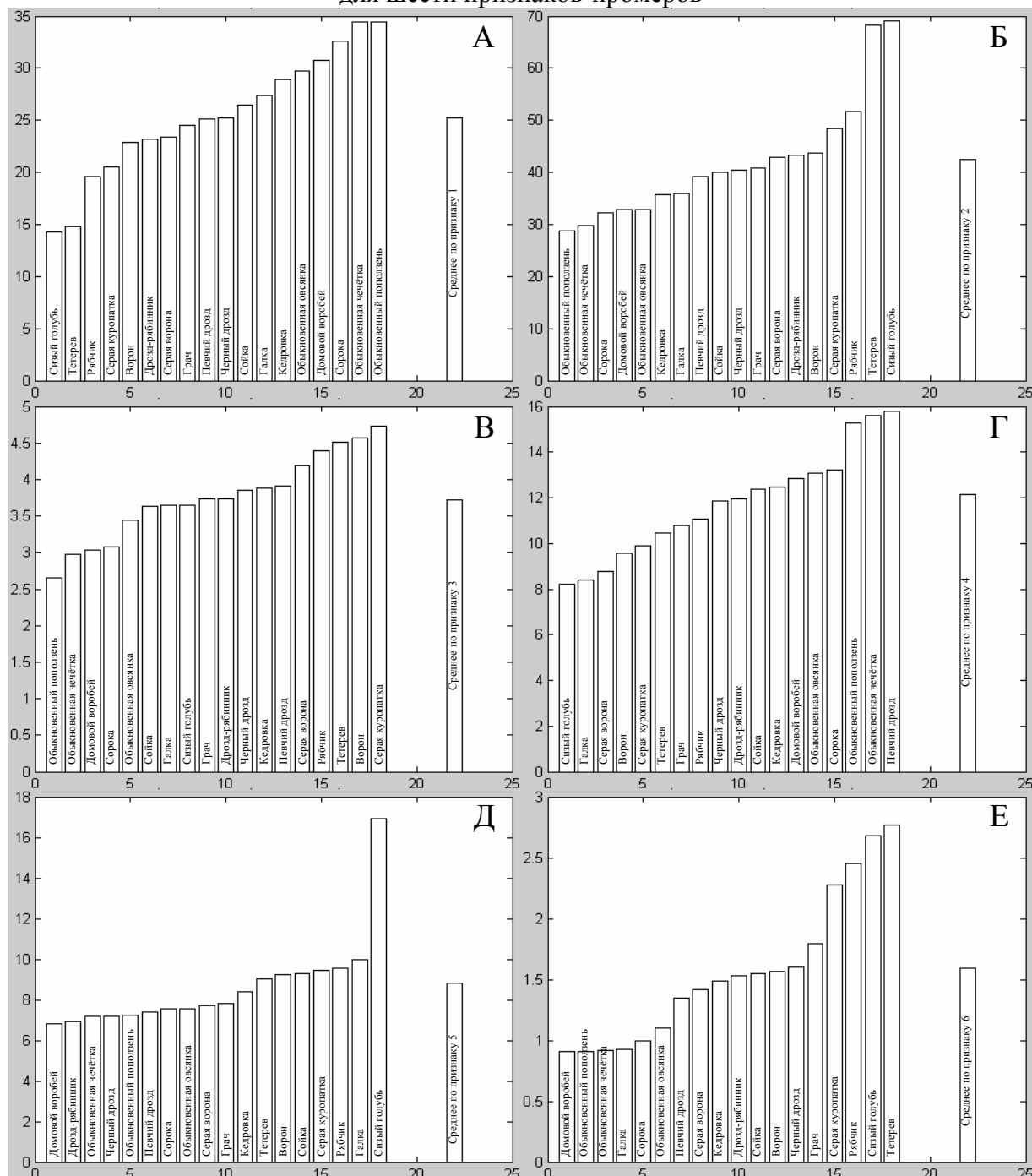
Результатом этих исследований стало наше значительное продвижение на пути создания электронного определителя: мы научились определять информативность признаков. В настоящий момент мы разрабатываем подходы к решению проблемы синтеза сложного признака. Определенный интерес представляют также методы, основанные на иерархической классификации.

Параллельно с данной системой распознавания, которая базируется на шести микроструктурных промерах, нами разрабатывается макроструктурная система идентификации видов (Силаева, Вараксин, Ильичёв, 2010; Силаева, Ильичёв, Чернова,

Фадеева, 2010). При возникновении сложностей определения вида в рамках одной системы можно будет воспользоваться и другой системой распознавания.

Авторы выражают благодарность сотрудникам ИПЭЭ РАН Г. В. Гуменюк — за помощь в измерениях микроструктурных деталей пера, В. А. Никулину — за техническую помощь при подготовке рукописи, а также всем, кто помогал в сборе материала.

Таблица 1. Упорядоченные средние значения вариативности для шести признаков-промеров



По шкале ординат — математическое ожидание; по шкале абсцисс — виды птиц. А — вариативность 1-го признака; Б — вариативность 2-го признака; В — вариативность 3-го признака; Г — вариативность 4-го признака; Д — вариативность 5-го признака; Е — вариативность 6-го признака.